



UNIVERSIDADE DA BEIRA INTERIOR
Engenharias

Fatores relevantes para a escolha modal em áreas urbanas

Jorge Humberto Gaspar Gonçalves

Tese para obtenção do Grau de Doutor em
Engenharia Civil
(3º ciclo de estudos)

Orientador: Prof. Doutor Álvaro Jorge da Maia Seco
Coorientador: Prof. Doutor Michael Joseph Mathias

Covilhã, Dezembro de 2012

Dedicatória

Dedico este trabalho em especial à minha família, aos meus pais, Manuel e Licínia, pelos princípios e exemplos que sempre me incutiram, às minhas filhotas queridas, Daniela e Camila, pelos momentos de descontração que me proporcionaram nos períodos mais problemáticos, e aos amigos que nos últimos anos me mostraram o valor e a importância de ter amigos do coração. A todos as minhas sinceras desculpas pela menor atenção que lhe pude dar.

Um beijo muito especial para a Xana que me permitiu a estabilidade emocional sem a qual provavelmente não teria concluído mais este passo...

Agradecimentos

Um obrigado muito especial e sentido ao Orientador Científico Prof. Doutor Álvaro Seco pelo exemplo, disponibilidade, paciência e inteligência que me transmitiu nestes anos de colaboração e nas inúmeras conversas e troca de ideias que fomos tendo ao longo dos anos. Um sincero pedido de desculpa à sua família pelas tardes de fim-de-semana e de férias em que os privei do seu ente querido...

Gostaria também de agradecer aos colegas da UBI, cujo apoio voluntário e desinteressado assumindo o meu serviço de docência, foi fundamental para que este trabalho fosse concluído. Uma palavra de agradecimento muito especial ao Eng.º. José Riscado, ao Prof. Doutor Martin Höck, ao Prof. Doutor Pedro Almeida, ao Prof. Doutor Miguel Nepomuceno, ao Arq. Fernando Diniz, e particularmente à minha querida amiga ao longo destes já muitos anos - Prof.^a Doutora Bertha Santos, por terem sido uma prova de que é nos piores momentos que se reconhecem os poucos, mas muito bons amigos, que realmente temos.

Agradeço igualmente ao Coorientador Prof. Doutor Michael Mathias, pela sua disponibilidade e ajuda com as burocracias inerentes a este tipo de processos.

Gostaria igualmente de agradecer às restantes pessoas cuja disponibilidade para me ajudar foi essencial para concluir este trabalho, nomeadamente:

- Ao Prof. Doutor Vítor Cavaleiro pela disponibilização de verbas para aquisição do *software* *NLOGIT*® e apoio em todos os aspetos burocráticos.
- Ao Prof. Doutor Carlos Rodrigues e Eng.º. Paulo Ribeiro, pelo apoio na obtenção dos dados referentes ao estacionamento no Porto e às redes de transportes coletivos a Norte da AMP em 2005 para o SIG.
- Ao Técnico Eduardo Ribeiro pelo auxílio com a cartografia digital da AMP para o SIG.
- Ao Eng.º. Luís Bento e ao Sr. José David da CMP pela disponibilidade demonstrada para fornecer os dados relativos às alterações no estacionamento municipal no Porto entre 2000 e 2005.
- Ao Eng.º. João Aires da STCP pela disponibilidade demonstrada para fornecer os dados da rede dos STCP em 2001.
- Ao Eng.º. Jorge Moreira da Companhia de Parques de Estacionamento - CPE, S.A., ao Eng.º. Manuel Sanchez da empresa EMPARQUE S.A., e ao Eng.º. Marco Martins da empresa Sociedade de Parques de Estacionamento - SPEL S.A., pelos dados prontamente fornecidos sobre os parques explorados no Porto.
- Ao Prof. Doutor João Abreu e Silva pela sua disponibilidade em me enviar a sua tese de doutoramento.
- À Prof. Doutora Rosário Macário pelas críticas construtivas e pela sua disponibilidade em me enviar documentos sobre a Avaliação da Qualidade dos Transportes Coletivos.
- Ao Prof. Marco Dias pela autorização para utilizar as suas folhas de cálculo de sequências de Halton.

Gostaria igualmente de agradecer aos Profs. Doutores David Hensher, Juan de Dios Ortúzar, William Green, Kenneth Train, Donald Shoup, Joan Walker e Fred Mannering, pela amabilidade e disponibilidade que manifestaram no esclarecimento de dúvidas sobre o tema e no envio dos seus trabalhos.

Resumo

Os problemas de mobilidade urbana verificados na maioria das cidades de média e grande dimensão a nível mundial intensificaram-se nas últimas décadas devido ao desenvolvimento económico verificado na maioria dos países. Este desenvolvimento potenciou alterações nos hábitos de mobilidade, que resultaram, na grande maioria dos casos, num uso excessivo do automóvel particular, modo extraordinariamente apelativo não apenas pelo conforto e liberdade que oferece, como pelo *status* que lhe está associado, e a um decréscimo acentuado e progressivo de modos de transportes alternativos mais sustentáveis. É consensualmente assumido internacionalmente que apenas uma conjugação coerente de políticas que coordenem, mas também condicionem e complementem o uso do automóvel particular com as alternativas mais sustentáveis, poderá inverter este panorama, e mitigar os problemas verificados sem, no entanto, reduzir ou prejudicar a mobilidade e a qualidade de vida dos cidadãos, essenciais para a continuação do desenvolvimento económico e social das cidades. Esta lógica passa por políticas que além de melhorarem a atratividade dos modos alternativos reduzam a facilidade e conveniência de uso do automóvel particular. Apesar da grande quantidade de estudos internacionais sobre a relevância de parte deste tipo de fatores no processo de tomada de decisão na escolha do modo de transporte, tal não acontece relativamente ao desenvolvimento de modelos objetivos e analíticos, capazes de, de forma completamente integrada e detalhada, determinar o impacto sobre os padrões de escolha modal da conjugação das condições socioeconómicas dos cidadãos e das políticas gerais de ordenamento do território e mobilidade nesse processo. A possibilidade de quantificar este tipo de impactos apresenta-se como um fator que poderá tornar as decisões políticas sobre investimentos nos sistemas de transportes de uma cidade mais eficientes.

O presente estudo pretende contribuir para esse mesmo conhecimento, utilizando como caso de estudo a Área Metropolitana do Porto, com dados socioeconómicos da população e sobre a estrutura urbana e da mobilidade no ano 2000. Assim, através da especificação de modelações de escolha discreta utilizando formulações Logit Multinomiais, realizaram-se extensas análises exploratórias para identificação dos atributos mais relevantes para a escolha dos modos de transportes mais sustentáveis, assim como dos fatores que manifestamente influenciam a atratividade do automóvel particular. Atenção foi também dedicada à definição das melhores modelações das inter-relações entre variáveis explicativas e a variável dependente, repartição modal, com especial atenção ao estudo de potenciais fenómenos de interação e interdependência de efeitos provocados por diferentes variáveis socioeconómicas caracterizadoras do universo de utilizadores do sistema. Adicionalmente foram realizadas algumas análises exploratórias para testar a estabilidade das tendências atuais de evolução dos padrões de escolha modal e do potencial revelado pelas medidas de política de ordenamento do território e de mobilidade sustentável mais relevantes, particularmente se aplicadas de forma integrada e coordenada. O estudo pretende, assim,

contribuir para um conhecimento, cientificamente comprovado, dos processos de escolha modal, nomeadamente mas não especificamente, para a realidade nacional.

Palavras-chave

Políticas de mobilidade sustentável; repartição modal; serviços de transportes coletivos; políticas e gestão de estacionamento; escolhas modais; modelos de escolha discreta; modelos Logit Multinomiais; variáveis socioeconómicas; variáveis políticas.

Abstract

The problems of urban mobility observed in most cities of medium and large scale worldwide have intensified in recent decades due to economic development observed in almost all countries. This development has leveraged changes in mobility habits, which, in most cases, resulted in excessive use of private cars, so extraordinarily appealing not only because of the comfort and freedom they offer, but also due to the social status it provides as well as progressive and distinct decrease in the use of alternative more sustainable transport modes. It is widely and internationally assumed that only a coherent combination of policies that coordinate, but constrain the use of the private car usage and complement it with more sustainable alternatives, may reverse this panorama and mitigate the problems faced without, however, reduce or impair citizens' mobility and quality of life, essential to the cities' economic growth and social development. This logic involves policies that, besides improving alternative modes attractiveness, also reduce the ease of use and attractiveness of the private car. Despite the large amount of international studies on the mode choice decision process, there is a lack in the development of analytical and objective models, capable of fully integrating the impact on mode choice patterns of citizens' socio-economic conditions and general urban and mobility planning policies in this process. The possibility of quantifying such impacts presents itself as a key factor to enable the improvement of the efficiency of political decisions on urban transport systems investments.

The present study aims to contribute to such knowledge, using as a case study the Oporto Metropolitan Area, with pre-existing socioeconomic data of its population and on the urban structure and mobility in 2000. Thus, discrete choice models using Multinomial Logit formulations were specified and extensive exploratory analysis were carried out to identify the most relevant attributes related with attraction to more sustainable transport modes as well as factors that clearly reduce the attractiveness of the private car. Attention was also devoted to the specification of the best models to describe interrelationships between explanatory variables and the dependent variable, modal split, with special attention to the study of potential interaction and interdependence effects caused by different socio-economic variables characterizing the system's user's universe. Additionally some exploratory analyses were performed to test the stability of the current trends on modal choice patterns and on the potential impact of relevant urban planning and sustainable mobility policy measures, particularly if applied in an integrated and coordinated way. The study, thus, seeks to improve scientific knowledge of transport users' modal choices processes, namely, but not specifically, regarding the national reality.

Keywords

Sustainable mobility policies; modal split; public transport service quality; parking policies and management; mode choice; discrete choice models; Multinomial Logit; socioeconomic variables; political variables.

Índice

Capítulo I

Introdução

1. Enquadramento do Tema	1
2. Identificação do Problema e Definição de Objetivos	1
3. Abordagem Metodológica	3
4. Estrutura da Tese	6

Capítulo II

Enquadramento

1. Introdução	11
2. A Interação entre os Sistemas de Transporte e o Desenvolvimento das Áreas Urbanas	12
3. Tipologias dos Sistemas de Transporte	15
3.1. Transporte Privado	15
3.2. Transportes de Aluguer	18
3.3. Transportes Coletivos	18
3.4. Transportes Coletivos em Meio Urbano	19
4. A Necessidade de Sistemas de Transporte Coletivos	20
5. A Mobilidade por Transporte Coletivo em Meio Urbano	22
5.1. A Qualidade dos Transportes Coletivos em Meio Urbano	26
5.2. Avaliação da Acessibilidade por Transportes Coletivos em Meio Urbano	28
5.3. Restrições à Livre Utilização do Automóvel Particular	29
6. Modelação da Mobilidade dos cidadãos	31
6.1. Introdução	31
6.2. Planeamento e Modelos de Transportes	31
6.3. Princípios Gerais de Modelação	35
6.3.1. Variáveis Endógenas e Exógenas	36
6.3.2. Modelação Agregada ou Desagregada	36
6.3.3. Recolha de Dados no Tempo	38
6.3.4. Preferências Reveladas e Declaradas	38
6.3.5. Modelação Contínua e Discreta	39
6.3.6. Amostragem	40
6.3.7. Estratégias de Amostragem	41
6.3.8. Tamanho da Amostra	42
6.3.9. Segmentação da População	44
6.3.10. Correção e Extrapolação/Expansão dos dados	45

6.3.11. Calibração/Estimação e Validação de Modelos	46
6.4. Estrutura Clássica dos Modelos de Transporte	47
6.5. Modelos de Escolha/Repartição Modal	50
6.5.1. Evolução Histórica dos Modelos de Escolha/Repartição Modal	50
6.5.1.1. Modelos Agregados	50
6.5.1.2. Modelos Desagregados	52
6.5.1.3. Modelos de Procura Direta	53
6.5.1.4. Modelos de Escolha Discreta	54
6.6. Modelos Baseados em Atividades	59
6.6.1. Definição de Jornada	61
6.6.2. Técnicas de Microsimulação	63
6.7. Diferenças e Similaridades entre os Modelos Baseados em Atividades e Modelos de Quatro Passos	64
7. Síntese	66

Capítulo III

Modelos de Escolha Discreta

1. Introdução	69
2. Fundamentação Teórica dos Modelos de Escolha Discreta	70
2.1. Teoria de Escolhas Individuais	70
2.2. Teoria da Maximização da Utilidade Aleatória	72
2.3. Formulação do Processo Comportamental	73
2.4. Generalidades dos Modelos de Escolha Discreta	76
2.4.1. Diferença entre Utilidades	76
2.4.2. Constantes Específicas das Alternativas	77
2.4.3. Variáveis Sociodemográficas ou Socioeconómicas	78
2.4.4. Número de Termos para os Erros Independentes	78
2.4.5. Escala da Utilidade	78
2.4.5.1. Normalização da Escala da Utilidade assumindo Erros Independentes e Identicamente Distribuídos (IID)	79
2.4.5.2. Normalização da Escala da Utilidade assumindo Erros com Diferentes Variâncias - Heteroscedaticidade	80
2.4.5.3. Normalização da Escala da Utilidade assumindo Erros Correlacionados	81
2.4.6. Estimação dos Coeficientes dos Modelos	83
2.4.7. Agregação de Dados e Transmissibilidade	86
3. Modelos de Escolha Discreta mais Comuns	94
3.1. Logit Multinomial (ou Multimodal)	94
3.1.1. Pressupostos Subjacentes à Formulação Logit	94
3.1.1.1. Independência entre Alternativas Irrelevantes - IIA	94

3.1.1.2. Erros Independentes e Igualmente Distribuídos - IID	98
3.1.2. Vantagens e Desvantagens dos Logit Simples	103
3.1.3. Erros de Especificação	104
3.1.4. Influência da Amostragem na Estimação dos Modelos	106
3.1.4.1. Amostras Exógenas	106
3.1.4.2. Amostras Baseadas em Escolhas	108
3.1.5. Precisão da Estimação Estatística dos Parâmetros do Modelo Logit	109
3.1.6. Ajustamento Global e Teste de Hipóteses	109
3.2. Modelos de Valores Extremos Generalizados - GEV	113
3.3. Probit Multinomial	116
3.3.1. Formulação Base	116
3.3.2. Normalização	119
3.3.3. Variações de Gosto	119
3.3.4. Padrões de Substituição/Falha no Pressuposto IIA	121
3.3.5. Dados Recolhidos ao Longo do Tempo por Preferências Declaradas	121
3.3.6. Simulação das Probabilidades de Escolha	121
3.4. Logit Misto	122
3.4.1. Formulação Base	122
3.4.2. Coeficientes Aleatórios	124
3.4.3. Componentes dos Erros	126
3.4.4. Padrões de Substituição	126
3.4.5. Inclusão de Dados Recolhidos por Escolhas Repetidas - Dados de Painéis	127
3.4.6. Especificidades de Identificação do Modelo Logit Misto	128
3.4.7. Simulação	130
4. Integração de Monte Carlo e Sequências de Extração	131
4.1. Integração de Monte Carlo	131
4.2. Sequências Não Aleatórias	133
5. Aplicação de Modelos de Escolha Discreta em Estudos Reais de Transporte	135
6. Síntese	143

Capítulo IV

Informação de Base

1. Introdução	147
2. Caracterização da AMP	148
2.1. População e Território Estudado	148
2.2. Rede de Transportes	149
2.2.1. Rede Viária	149
2.2.2. Rede de Transportes Coletivos	150

2.2.3. Oferta de Estacionamento	150
3. Inquérito à Mobilidade da População Residente em 2000	153
3.1. Caracterização da Informação Original Recolhida na Base de Dados	153
3.2. Condicionantes na Utilização da Base dos Dados do INE	157
3.3. Outras Condicionantes à Utilização da Base dos Dados Seleccionada	158
4. Opções Assumidas no Tratamento da Informação	160
4.1. Distância Percorrida na Viagem	160
4.2. Frequências Horárias de BUS	160
4.3. Velocidade Comercial dos BUS	162
4.4. Tempos de Viagem Apreendidos	162
5. Definição e Justificação das Variáveis Utilizadas na Análise	163
5.1. Variáveis Socioeconómicas	163
5.1.1. Rendimento Líquido Mensal do Agregado	163
5.1.2. Rendimento Líquido Mensal do Agregado <i>per Capita</i>	165
5.1.3. Sexo do Inquirido	166
5.1.4. Idade do Inquirido	167
5.1.5. Nível de Instrução	169
5.1.6. Número de Automóveis Disponíveis Diariamente <i>per Capita</i> no Agregado	171
5.1.7. Disponibilidade Diária de Moto no Agregado	174
5.1.8. Licença de Condução de Automóvel	174
5.2. Variáveis Caracterizadoras da Viagem e da Organização das Atividades Diárias	175
5.2.1. Modo de Transporte Utilizado	175
5.2.2. Distância entre as Zonas de Geração-Atração	177
5.2.3. Duração Total da Viagem Apreendida	179
5.2.4. Motivo da Viagem	184
5.2.5. Duração das Atividades	186
5.2.6. Número Total de Viagens Diárias	187
5.3. Variáveis Caracterizadoras das Políticas de Mobilidade	188
5.3.1. Variáveis Caracterizadoras das Políticas de Estacionamento	188
5.3.1.1. Densidade de Oferta face à Procura de Estacionamento	188
5.3.1.2. Proporção de Estacionamento Pago	190
5.3.1.3. Combinação das Variáveis Binárias sobre Políticas de Estacionamento	191
5.3.2. Variáveis Caracterizadoras das Políticas de Transporte Coletivo	193
5.3.2.1. Razão Tempo de Viagem por Bus/Tempo de Viagem por Auto	193
5.3.2.2. Frequência Média Horária de BUS	194
5.3.2.3. Velocidade comercial equivalente dos BUS	200
6. Amostras Disponíveis	201

7. Síntese	203

Capítulo V

Modelações Preliminares

1. Introdução	207
2. Enquadramento Base	208
2.1. Tipologias de Especificações Mais Comuns	209
2.2. Especificação Discreta Alternativa	210
3. Análise Isolada de Variáveis Socioeconómicas	211
3.1. Rendimento Líquido Mensal do Agregado	211
3.1.1. Análise Comportamental de Diferentes Especificações	211
3.1.2. Avaliação de Comportamentos Não-lineares	213
3.1.3. Rendimento Líquido Mensal <i>versus</i> Rendimento Líquido Mensal <i>per capita</i>	216
3.2. Sexo do Inquirido	217
3.3. Escalão Etário	218
3.3.1. Análise Comportamental de Diferentes Especificações	218
3.3.2. Avaliação de Comportamentos Não Lineares	219
3.4. Nível de Instrução	220
3.4.1. Análise Comportamental de Diferentes Especificações	220
3.4.2. Avaliação de Comportamentos Não Lineares	222
3.5. Número de Automóveis Disponíveis no Agregado <i>per capita</i>	223
3.5.1. Avaliação Comportamental de Diferentes Especificações	223
3.5.2. Avaliação de Comportamentos Não Lineares	224
3.6. Disponibilidade de Moto no Agregado	226
3.7. Licença de Condução	227
4. Análise Isolada de Variáveis Relacionadas com a Posição Relativa entre a Zona de Geração e a de Atração	227
4.1. Afastamento entre a Geração e a Atração	227
4.1.1. Análise Comportamental de Diferentes Especificações	227
4.1.2. Avaliação de Comportamentos Não Lineares	230
4.2. Duração Total da Viagem por Modo	232
4.2.1. Análise Comportamental de Diferentes Especificações	232
4.2.2. Avaliação de uma Especificação Alternativa	233
5. Análise Isolada de Variáveis Relacionadas com a Organização das Atividades Diárias	234
5.1. Motivo da Viagem	234
5.1.1. Análise Comportamental de Diferentes Especificações	234
5.1.2. Avaliação de Comportamentos Não Lineares	245
5.2. Duração da Atividade no Destino da Viagem	236

5.2.1. Análise Comportamental de Diferentes Especificações	236
5.2.2. Avaliação de Comportamentos Não Lineares	238
5.2.3. Avaliação de uma Especificação Alternativa	240
5.3. Número de Viagens no Dia	240
5.3.1. Análise Comportamental de Diferentes Especificações	240
5.3.2. Avaliação de Comportamentos Não Lineares	242
6. Variáveis Relativas a Efeitos Isolados Relacionados com Políticas de Estacionamento	243
6.1. Rácio “Oferta de Estacionamento / Total de Viagens Atraídas”	243
6.1.1. Análise Comportamental de Diferentes Especificações	243
6.1.2. Avaliação de Comportamentos Não Lineares	245
6.2. Rácio “Oferta de Estacionamento Paga / Total”	246
6.2.1. Análise Comportamental de Diferentes Especificações	246
6.2.2. Avaliação de Comportamentos Não Lineares	248
7. Análise Isolada de Variáveis Relacionadas com o Serviço de Transportes Coletivos	249
7.1. Rácio “Tempo Viagem BUS / Auto”	249
7.1.1. Análise Comportamental de Diferentes Especificações	249
7.1.2. Avaliação de Comportamentos Não Lineares	251
7.2. Frequências Horárias Médias dos Serviços BUS	253
7.2.1. Análise Comportamental de Diferentes Especificações	253
7.2.2. Avaliação de Comportamentos Não Lineares	255
7.2.3. Outras Especificações da Frequência Média Horária de BUS	257
7.3. Velocidades Comerciais Equivalentes dos Serviços BUS	259
7.3.1. Análise Comportamental de Diferentes Especificações	259
7.3.2. Avaliação de Comportamentos Não Lineares	261
8. Síntese das Análises Isoladas	262
9. Avaliação de Questões de Correlação/Complementaridade entre Variáveis	263
9.1. Correlação entre a taxa de motorização familiar e a existência de licença de condução	264
9.2. Correlação entre o Escalão Etário e o Nível de Instrução	266
9.3. Correlação entre as Variáveis sobre Estacionamento	268
9.4. Correlação entre a Velocidade Comercial Equivalente de BUS e a Distância Mais Curta entre Centróides ou o Tempo Médio de Viagem	272
10. Estabilidade Modal ao Longo das Viagens Diárias	274
10.1. Modo Escolhido na 1ª Viagem do Dia versus Modo da Viagem Atual	274
10.2. Estabilidade do Modo Escolhido ao Longo das Atividades do Dia	275
11. Síntese	289

Capítulo VI

Modelações Combinados

1. Enquadramento das Análises	291
2. Modelos Multinomiais Logit com Combinação de Variáveis	292
2.1. Modelo Base	292
2.2. Modelo de Referência com Inclusão de mais Variáveis Socioeconómicas	294
2.3. Modelo de Referência com Inclusão da Variável Genérica Duração Total da Viagem	301
2.4. Modelo de Referência com Inclusão das Variáveis Relacionadas com a Organização das Atividades ao Longo do Dia	303
2.4.1. Duração da Estadia no Destino	303
2.4.2. Número Total de Viagens Diárias	307
2.4.3. Motivo da Viagem	309
2.4.3.1. Impacto no Modelo dos Motivos de Viagem	309
2.4.3.2. Correlação entre Motivo da Viagem e Duração da Estadia	311
2.5. Modelo de Referência Acrescido de Variáveis Representativas da Oferta de Estacionamento na Zona de Atração	315
2.5.1. Eliminação dos Efeitos Resultantes de Diferenças de Dimensão nas Bases de Dados	315
2.5.2. Modelo de Referência Combinado com Variáveis Binárias sobre a Oferta de Estacionamento	316
2.5.3. Modelo de Referência Combinado com Variáveis Contínuas sobre a Oferta de Estacionamento	322
2.6. Modelo de Referência com Inclusão das Variáveis Relacionadas com Transportes Coletivos	322
2.6.1. Eliminação dos Efeitos Resultantes de Diferenças de Dimensão nas Bases de Dados	322
2.6.2. Modelos de Referência Combinando apenas Variáveis Representando Oferta de Transporte Coletivo	325
2.7. Modelo de Referência com Inclusão das Variáveis Relacionadas com Transportes Coletivos e Estacionamentos	327
2.7.1. Modelo Base com Combinação de Variáveis Binárias	327
2.7.2. Teste à utilização de Variáveis de Estacionamento Contínuas	329
2.7.3. Teste de Modelo sem Variáveis Socioeconómicas menos Relevantes	331
2.7.4. Redução da Base de Dados aos Modos Alternativos mais Relevantes	334
2.7.5. Exclusão do Motivo da Viagem	336
2.7.6. Inclusão da Duração Média da Viagem por Modo	337
3. Resultados Globais Fundamentais	340
3.1. Grau de (In)dependência dos Efeitos dos Diferentes Fatores	340

3.2. Modelos Integrados Fundamentais	342
3.2.1. Modelo sem Variáveis sobre Políticas de Mobilidades	344
3.2.2. Modelo com Variáveis sobre Políticas de Mobilidades	348
4. Avaliação da Consistência das Simulações dos Modelos através dos Resultados da Base de Dados	351
5. Análises Adicionais Relativamente à Não Linearidade e Interdependência das Variáveis Explicativas/Independentes	353
5.1. Tendências Verificadas na Base de Dados Relativamente às Principais Variáveis Sociais: Sexo vs. Licença de Condução vs. Idade vs. Taxa de Motorização	353
5.2. Modelos Socioeconómicos Adicionais	360
6. Comparação dos Resultados com os de Estudos Internacionais	369
7. Conclusões	378
 Capítulo VII	
Simulação de Cenários Futuros	
1. Introdução	383
2. Análise da Potencial Influência das Características Socioeconómicas dos Indivíduos - Perfis Padrão	384
3. Análise de Sensibilidade dos Modelos a Alterações Sociais e a Mudanças de Políticas de Mobilidade	390
4. Simulação de Repartições Modais Globais para Cenários Resultantes de Políticas de Mobilidade Específicas	399
5. Relação entre o Rendimento Mensal Médio <i>per Capita</i> e as Variáveis Explicativas mais Relevantes	402
5.1. Relação Taxa de Motorização <i>versus</i> Rendimento Familiar	411
5.2. Relação Licença de Condução <i>versus</i> Sexo <i>versus</i> Rendimento Familiar	412
6. Conclusões	413
 Capítulo VIII	
Conclusões	417
 Capítulo IX	
Referências Bibliográficas	433
 Anexos	
Anexo 1 do Capítulo II	
1. Amostragem Aleatória Simples	443
2. Amostragem Estratificada Generalizada	443
3. Amostragem Exógena	443

4. Amostragem Baseada em Escolhas	443
5. Amostragem Enriquecida	444
6. Amostragem Dupla	444
Anexo 1 do Capítulo III	
Definição das expressões das variâncias e covariâncias entre erros	445
Anexo 2 do Capítulo III	
Tratamento de funções quadráticas, definição do tamanho do incremento e avaliação da concavidade no procedimento de <i>Newton-Raphson</i> para a maximização da função de verosimilhança	447
Anexo 3 do Capítulo III	
Definição de elasticidades agregadas	449
Anexo 4 do Capítulo III	
Variações dos Modelos GEV	451
1. Decomposição em Dois Logit Padrão	451
2. Logit Hierárquico Sobreposto	452
3. Logit de Pares Combinatórios (PCL)	453
4. Logit Hierárquico Generalizado (GNL)	454
5. Logit Heteroscedástico	455
6. Formulação Geral para Modelos GEV	456
Anexo 5 do Capítulo III	
Metodologia genérica para a definição da matriz M de diferença de erros	459
Anexo 6 do Capítulo III	
Metodologia para a normalização de modelos Probit	461
Anexo 7 do Capítulo III	
Metodologia para a definição de padrões de substituição com e sem restrições	465
1. Covariância completa ou Padrões de substituição sem restrições	465
2. Covariância estruturada ou Padrões de substituição restringidos	465
Anexo 8 do Capítulo III	
Metodologia a inclusão de dados recolhidos ao longo do tempo por preferências declaradas	467
Anexo 9 do Capítulo III	
Exemplo do método de simulação GHK simples e com estimação da máxima verosimilhança para a probabilidade de escolha	469
1. Simulador GHK	469
2. Simulador GHK com Estimação da Máxima Verosimilhança	473
Anexo 10 do Capítulo III	
Caracterização das distribuições mais comuns para a estimação de modelos Logit Mistos	476
1. Distribuição Uniforme	476

2. Distribuição Triangular	477
3. Distribuição Lognormal	478
4. Imposição de Restrições às Distribuições	479
5. Comparação de Distribuições Empíricas com Distribuições Analíticas	480
Anexo 11 do Capítulo III	
Procedimentos empíricos para a identificação de modelos Logit Mistos	482
1. Bases para a Identificação	483
1.1. Condição de Ordem da Matriz	484
1.2. Condição da Característica da Matriz	485
1.3. Condição da Definição Positiva	485
2. Restrições a Aplicar para a Identificação de Modelos Logit Mistos	487
3. Outros Aspetos Relevantes na Especificação do Modelo Logit Misto	490
3.1. Seleção dos Parâmetros Aleatórios	490
3.2. Seleção da Distribuição dos Parâmetros Aleatórios	491
3.3. Seleção do Número de Extrações para as Simulações	491
3.4. Heterogeneidade de preferência em redor da média de um parâmetro aleatório	492
3.5. Inclusão de Situações de Escolha Correlacionadas	494
3.6. Correlação entre Parâmetros	495
3.7. Disponibilidade para Pagar (DpP)	495
3.8. Existência de Variáveis Endógenas	496
Anexo 12 do Capítulo III	
Caracterização das Sequências de Halton	499
Anexo 13 do Capítulo III	
Exemplo da Determinação da Característica de uma Matriz de Covariâncias entre Diferenças de Utilidade de um Modelo Heteroscedástico com quatro Alternativas	506
Anexo 1 do Capítulo IV	
Freguesias dos municípios da NUT III - Grande Porto consideradas no zonamento para o caso de estudo	509
Anexo 2 do Capítulo IV	
Rede de Transportes Coletivos utilizada na análise do caso de estudo	513
Anexo 3 do Capítulo IV	
Oferta total de estacionamento nas subseções estatísticas do concelho do Porto em 2000	517
Anexo 1 do Capítulo VI	
Resultados combinando as variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >1km e ≤2km	521

Anexo 2 do Capítulo VI

Resultados combinando as variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA > 2km

525

Lista de Figuras

Capítulo II

Figura II. 1 - Interações dos diferentes vetores de intervenção para uma Mobilidade Sustentável em meio urbano	25
Figura II. 2 - Fatores que podem influenciar a procura de tráfego (adaptado de Ortúzar e Willumsen, 2001)	33
Figura II. 3 - Estrutura clássica do modelo de transporte de quatro passos (retirado de Ortúzar e Willumsen, 2001)	48
Figura II. 4 - Exemplo de uma estrutura de um modelo baseado em atividades (adaptado de Vovsha <i>et al.</i> , 2008b)	65

Capítulo III

Figura III. 1- Ilustração da influência da inclinação e da curvatura no processo incremental (traduzido de Train (2003))	86
Figura III. 2 - Diferença entre a probabilidade média e a probabilidade calculada para o ponto médio da utilidade representativa ou mensurável (adaptado de Train (2003) e Washington <i>et al.</i> (2003))	87
Figura III. 3- Representação gráfica das densidades de probabilidades da expressão III.55 com distribuição de valores extremos do tipo I ou de Gumbel para diferentes valores dos parâmetros de localização e de escala	100
Figura III. 4 - Representação gráfica da distribuição cumulativa da expressão III.54 com distribuição de valores extremos do tipo I ou de Gumbel para diferentes valores dos parâmetros de localização e de escala	100
Figura III. 5 - Relação entre R^2 de modelos lineares e R^2 não lineares (pseudo- R^2) (adaptado de Hensher <i>et al.</i> , 2005)	110
Figura III. 6- Possível diagrama em árvore para um modelo de escolha modal Logit Hierárquico	115
Figura III. 7 - Exemplo de cálculo do integral bidimensional (Área) de uma função	132

Capítulo IV

Figura IV. 1 - Rede viária principal na AMP em 2000 (Microsoft Corporation, 2001)	149
Figura IV. 2 - Dados estatísticos e distribuição da variável R através do <i>software</i> SPSS®	164
Figura IV. 3 - Dados estatísticos e distribuição da variável R_c através do <i>software</i> SPSS®	166
Figura IV. 4 - Dados estatísticos e distribuição da variável I_{db} através do <i>software</i> SPSS®	168

Figura IV. 5 - Dados estatísticos e distribuição da variável <i>In</i> através do <i>software</i> SPSS®	170
Figura IV. 6 - Dados estatísticos e distribuição da variável <i>NC</i> através do <i>software</i> SPSS®	172
Figura IV. 7 - Dados estatísticos e distribuição da variável <i>MTrp</i> através do <i>software</i> SPSS®	177
Figura IV. 8 - Dados estatísticos e distribuição da variável <i>D2</i> através do <i>software</i> SPSS®	178
Figura IV. 9 - Dados estatísticos e distribuição da variável <i>T1a</i> através do <i>software</i> SPSS®	179
Figura IV. 10 - Dados estatísticos e distribuição da variável <i>T2a</i> através do <i>software</i> SPSS®	180
Figura IV. 11 - Dados estatísticos e distribuição da variável <i>T3a</i> através do <i>software</i> SPSS®	180
Figura IV. 12 - Dados estatísticos e distribuição da variável <i>T4a</i> através do <i>software</i> SPSS®	180
Figura IV. 13 - Dados estatísticos e distribuição da variável <i>T5a</i> através do <i>software</i> SPSS®	181
Figura IV. 14 - Dados estatísticos e distribuição da variável <i>T6a</i> através do <i>software</i> SPSS®	181
Figura IV. 15 - Dados estatísticos e distribuição da variável <i>T1l</i> através do <i>software</i> SPSS®	182
Figura IV. 16 - Dados estatísticos e distribuição da variável <i>T2l</i> através do <i>software</i> SPSS®	182
Figura IV. 17 - Dados estatísticos e distribuição da variável <i>T3l</i> através do <i>software</i> SPSS®	182
Figura IV. 18 - Dados estatísticos e distribuição da variável <i>T4l</i> através do <i>software</i> SPSS®	183
Figura IV. 19 - Dados estatísticos e distribuição da variável <i>T5l</i> através do <i>software</i> SPSS®	183
Figura IV. 20 - Dados estatísticos e distribuição da variável <i>T6l</i> através do <i>software</i> SPSS®	183
Figura IV. 21 - Dados estatísticos e distribuição da variável <i>Mt</i> através do <i>software</i> SPSS®	185
Figura IV. 22 - Dados estatísticos e distribuição da variável <i>De</i> através do <i>software</i> SPSS®	186
Figura IV. 23 - Dados estatísticos e distribuição da variável <i>nV</i> através do <i>software</i> SPSS®	187
Figura IV. 24 - Dados estatísticos e distribuição da variável <i>OV</i> através do <i>software</i> SPSS®	189
Figura IV. 25 - Dados estatísticos e distribuição da variável <i>PO</i> através do <i>software</i>	190

SPSS®

Figura IV. 26 - Dados estatísticos e distribuição da variável TBA através do <i>software</i> SPSS®	193
Figura IV. 27 - Dados estatísticos e distribuição da variável Fm através do <i>software</i> SPSS®	195
Figura IV. 28 - Dados estatísticos e distribuição da variável Fa através do <i>software</i> SPSS®	195
Figura IV. 29 - Dados estatísticos e distribuição da variável Ft através do <i>software</i> SPSS®	196
Figura IV. 30 - Dados estatísticos e distribuição da variável Ftl através do <i>software</i> SPSS®	196
Figura IV. 31 - Dados estatísticos e distribuição da variável Fm0 através do <i>software</i> SPSS®	196
Figura IV. 32 - Dados estatísticos e distribuição da variável Fa0 através do <i>software</i> SPSS®	197
Figura IV. 33 - Dados estatísticos e distribuição da variável Ft0 através do <i>software</i> SPSS®	197
Figura IV. 34 - Dados estatísticos e distribuição da variável Ftl0 através do <i>software</i> SPSS®	197
Figura IV. 35 - Dados estatísticos e distribuição da variável Fm1 através do <i>software</i> SPSS®	198
Figura IV. 36 - Dados estatísticos e distribuição da variável Fa1 através do <i>software</i> SPSS®	198
Figura IV. 37 - Dados estatísticos e distribuição da variável Ft1 através do <i>software</i> SPSS®	198
Figura IV. 38 - Dados estatísticos e distribuição da variável Ftl1 através do <i>software</i> SPSS®	199
Figura IV. 39 - Dados estatísticos e distribuição da variável V2 através do <i>software</i> SPSS®	200

Capítulo V

Figura V. 1 - Utilidades parciais para os diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável binária para cada categoria	213
Figura V. 2 - Utilidades parciais dos diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável discreta por categorias	214
Figura V. 3 - Utilidades parciais dos diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável discreta com o valor máximo de cada categoria	214

Figura V. 4 - Esquema representativo de relações lineares entre parâmetros de variáveis com combinação de discretas	215
Figura V. 5 - Utilidades parciais para os diferentes modos de transporte considerando o escalão etário como variável binária para cada categoria	220
Figura V. 6 - Utilidades parciais para os diferentes modos de transporte considerando o nível de instrução como variável binária para cada categoria	222
Figura V. 7 - Utilidades parciais para os diferentes modos de transporte considerando a taxa de motorização familiar especificada como variável binária para cada categoria	225
Figura V. 8 - Utilidades parciais para os diferentes modos de transporte considerando a taxa de motorização familiar especificada como variável contínua	225
Figura V. 9 - Utilidades parciais para os diferentes modos de transporte considerando a distância mais curta entre centróides em km como variável não continua	230
Figura V. 10 - Utilidades parciais para os diferentes modos de transporte considerando a distância mais curta entre centróides em km como variável contínua	231
Figura V. 11 - Utilidades parciais para os diferentes modos de transporte considerando a combinação das especificações continua e binárias da distância mais curta entre centróides em km	231
Figura V. 12 - Utilidades parciais para os diferentes modos de transporte considerando a combinação das especificações continua e binárias (apenas $Dx1 \leq 1\text{km}$) da distância mais curta entre centróides em km	232
Figura V. 13 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para o motivo da viagem	236
Figura V. 1 - Utilidades parciais para os diferentes modos de transporte considerando a especificação binária da duração da estadia no destino em minutos	238
Figura V. 2 - Utilidades parciais para os diferentes modos de transporte considerando a especificação contínua da duração da estadia no destino em minutos	239
Figura V. 3 - Utilidades parciais para os diferentes modos de transporte considerando a especificação não linear para o número de viagens diárias de um individuo	242
Figura V. 4 - Utilidades parciais para os diferentes modos de transporte considerando a especificação linear para o número de viagens diárias de um individuo	243
Figura V. 5 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão oferta de lugares versus viagens totais extrapoladas	245
Figura V. 6 - Utilidades parciais para os diferentes modos de transporte considerando uma variável continua para a razão oferta de lugares versus viagens totais extrapoladas	246
Figura V. 20 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão oferta de lugares pagos versus oferta total de estacionamento na zona de atração	248
Figura V. 21 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a razão oferta de lugares pagos versus oferta total de	249

estacionamentos na zona de atracção

Figura V. 22 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão entre a Duração Média da Viagem por BUS e AUTO entre as zonas de GA 251

Figura V. 23 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a razão entre o Duração Média da Viagem por BUS e AUTO entre as zonas de GA 252

Figura V. 24 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a frequência média horária de TC potencialmente à disposição entre as 7h30 e as 19h30 256

Figura V. 25 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a frequência média horária de TC potencialmente à disposição entre as 7h30 e as 19h30 256

Figura V. 26 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a velocidade comercial equivalente para BUS em km/h 261

Figura V. 27 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a velocidade comercial equivalente para BUS em km/h 262

Figura V. 7 - Análise do modo de cada viagem em relação ao modo da primeira viagem do dia 275

Figura V. 8 - Estabilidade do uso do modo Auto ao longo do dia de indivíduos do Tipo I 276

Figura V. 30 - Estabilidade do uso do modo Auto ao longo do dia para indivíduos do Tipo II 276

Figura V. 31- Estabilidade do uso do modo Auto ao longo do dia para indivíduos do Tipo III 277

Figura V. 32 - Análise da estabilidade do modo ao longo do dia para indivíduos Tipo I, atendendo ao modo utilizado na primeira viagem do dia 281

Figura V. 33- Análise da estabilidade do modo ao longo do dia para indivíduos Tipo II, atendendo ao modo utilizado na primeira viagem do dia 283

Figura V. 34 - Análise da estabilidade do modo ao longo do dia para indivíduos Tipo III, atendendo ao modo utilizado na primeira viagem do dia 285

Capítulo VI

Figura VI. 1 - Evolução do ajustamento global dos modelos MNL com a introdução das variáveis mais relevantes para a análise com 6 modos 341

Figura VI. 2 - Evolução do ajustamento global dos modelos MNL com a introdução das variáveis mais relevantes para a análise com os 3 modos mais relevantes 342

Figura VI. 3 - Diferenças entre a repartição modal verificadas na amostra e as estimativas obtidas pelo modelo MNL7.57 352

Figura VI. 4 - Diferenças entre a repartição modal verificadas na amostra e as estimativas obtidas pelo modelo MNL12.16	352
Figura VI. 5 - Gráficos com resultados da combinação das variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >1km e ≤2km	355
Figura VI. 6 - Gráficos com resultados da combinação das variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >2km	357
Figura VI. 7 - Gráficos com resultados do modelo MNL99.3a_BWA_NCA_b e repartições modais correspondentes dos modos Bus e Walk para indivíduos sem licença de condução	364
Figura VI. 8 - Gráficos com resultados do modelo MNL99.3a_BWA_NCA_b e repartições modais correspondentes dos modos Bus e Walk para indivíduos com licença de condução	365
Figura VI. 9 - Gráficos dos parâmetros do modelo MNL99.9a_BWA_NCA_b com especificação binária de combinações de variáveis socioeconómicas	367

Capítulo VII

Figura VII. 1 - Utilidades parciais para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores extremos (MNL3.139)	385
Figura VII. 2 - Repartições modais estimadas pelo modelo MNL3.139 para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores extremos	385
Figura VII. 3 - Comparação das Repartições modais reais da BD com as estimadas pelo modelo MNL3.139 para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores intermédios	389
Figura VII. 4 - Utilidades parciais para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores intermédios (MNL3.139).	390
Figura VII. 5 - Análise de sensibilidade de variáveis com o modelo MNL7.57	395
Figura VII. 6 - Análise de sensibilidade de variáveis com o modelo MNL12.16	397
Figura VII. 7 - Resultados da aplicação de cada cenário ao modelo MNL12.16	401
Figura VII. 8 - Variações nos resultados para cada cenário com o modelo MNL12.16, relativamente à repartição modal verificada na base de dados	401
Figura VII. 9 - Evolução da taxa de motorização familiar <i>per capita</i> média com o rendimento mensal do agregado <i>per capita</i>	403
Figura VII. 10 - Evolução da taxa de motorização familiar por adulto média com o rendimento mensal do agregado <i>per capita</i>	405

Figura VII. 11 - Evolução da taxa de motorização familiar de moto <i>per capita</i> média com o rendimento mensal do agregado <i>per capita</i>	407
Figura VII. 12 - Evolução da taxa de motorização familiar de moto por adulto média com o rendimento mensal do agregado <i>per capita</i>	409
Figura VII. 13 - Evolução da percentagem de encartados por sexo com o rendimento familiar mensal <i>per capita</i>	412

Anexos

Figura A. 1 - Exemplo da probabilidade da alternativa 1 no espaço dos erros independentes η_1 e η_2 (adaptado de Train, 2003)	471
Figura A. 2 - Probabilidade de η_2 estar no limite correto atendendo a η_2^r (adaptado de Train, 2003)	472
Figura A. 3 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição uniforme, para várias médias e desvios padrão	476
Figura A. 4 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição triangular (simétricas), para várias médias/modas e desvios padrão	477
Figura A. 5 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição Lognormal, para várias médias e desvios padrão	479
Figura A. 6 - Exemplo de 1000 extrações de sequências de Halton de base 2 e base 3 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009)	500
Figura A. 7 - Exemplo de 1000 extrações de sequências aleatórias de distribuição uniforme [0,1] utilizando o Microsoft Office Excel© 2003	501
Figura A. 8 - Exemplo de correlações em 1000 extrações de sequências de Halton de base 10 e base 11 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009)	502
Figura A. 9 - Exemplo de correlações em 1000 extrações de sequências de Halton de base 20 e base 21 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009)	502
Figura A. 10 - Exemplo de 1000 extrações de sequências de Halton aleatórias de base 2 e base 3 deslocadas respetivamente 0,4177 e 0,5832 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009)	504
Figura A. 11 - Exemplo de 1000 extrações de sequências de Halton aleatórias de base 2 e base 3 deslocadas respetivamente 0,3805 e 0,0943 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009)	504

Figura A. 12 - Freguesias dos municípios da NUT III - Grande Porto consideradas no caso de estudo	511
Figura A. 13 - Rede de Transportes Coletivos utilizada na análise do caso de estudo	515
Figura A. 14 - Oferta total de estacionamento nas freguesias do município do Porto em 2000 (baseado em Rodrigues <i>et al.</i> , 2005)	519

Lista de Quadros

Capítulo II

Quadro II. 1 - Tamanhos de amostras atendendo ao nível de confiança e margens de erro pretendidas para $p=q=0,5$ (ITE, 1999)	44
Quadro II. 2 - Tamanhos de amostras atendendo ao nível de confiança e margens de erro pretendidas para $p=0,22$, $q=0,78$ (ITE, 1999)	44

Capítulo III

Quadro III. 1 - Estudos internacionais sobre escolhas modais com modelos de escolha discreta	137

Capítulo IV

Quadro IV. 1 - Municípios inseridos na AMP (INE, 2004)	148
Quadro IV. 2 - Caracterização dos Municípios da NUT III - Grande Porto com dados de 2000 (INE, 2004)	148
Quadro IV. 3 - Tipologias de oferta de estacionamento consideradas (baseado em Rodrigues <i>et al.</i> , 2005)	151
Quadro IV. 4 - Oferta de estacionamento no município do Porto ajustados para o ano 2000	152
Quadro IV. 5 - Variáveis recolhidas para caracterizar os Agregados Domésticos Privado (ADP) (INE, 2002)	154
Quadro IV. 6 - Variáveis recolhidas para caracterizar cada indivíduo (INE, 2002)	155
Quadro IV. 7 - Variáveis recolhidas para caracterizar as viagens (INE, 2002)	156
Quadro IV. 8 - Variáveis recolhidas para caracterizar cada deslocação (INE, 2002)	157
Quadro IV. 9 - Variáveis recolhidas para caracterizar cada deslocação (INE, 2002)	158
Quadro IV. 10 - Repartição modal verificada para cada categoria binária do rendimento líquido mensal do agregado	165
Quadro IV. 11 - Repartição modal verificada para cada categoria binária do rendimento líquido mensal do agregado <i>per capita</i>	166
Quadro IV. 12 - Repartição da variável Sexo do inquirido	167
Quadro IV. 13 - Repartição modal da variável Sexo do inquirido	167
Quadro IV. 14 - Codificação do escalão etário original na base de dados	168
Quadro IV. 15 - Repartição das idades dos inquiridos na amostra	169

Quadro IV. 16 - Repartição modal segundo os escalões etários	169
Quadro IV. 17 - Codificação do nível de instrução original na base de dados	170
Quadro IV. 18 - Repartição do nível de instrução dos inquiridos na amostra	171
Quadro IV. 19 - Repartição modal segundo os níveis de instrução	171
Quadro IV. 20 - Repartição das variáveis binárias consideradas sobre a taxa de motorização familiar da base de dados	173
Quadro IV. 21 - Repartição modal verificada para cada categoria da taxa de motorização familiar na base de dados	173
Quadro IV. 22 - Repartição da variável disponibilidade diária de Moto (Mot) no agregado na base de dados	174
Quadro IV. 23 - Repartição modal verificada para a disponibilidade diária de Moto na base de dados	174
Quadro IV. 24 - Repartição da variável Lc na base de dados	175
Quadro IV. 25 - Modos de transportes considerados na base de dados original (INE, 2002)	175
Quadro IV. 26 - Modos de transportes agregados	176
Quadro IV. 27 - Repartição do modo de transporte utilizado na viagem em estudo na amostra	177
Quadro IV.28 - Variáveis binárias para a distância mais curta entre cent	178
Quadro IV. 29 - Repartição modal para as diferentes distâncias entre centróides	179
Quadro IV. 30 - Codificação dos motivos de viagem na base de dados original	184
Quadro IV. 31 - Repartição do motivo da viagem na amostra	185
Quadro IV. 32 - Repartição modal segundo o motivo da viagem	185
Quadro IV. 33 - Variáveis binárias para o número total de viagens diárias	186
Quadro IV. 34 - Repartição modal para as diferentes categorias duração da atividade	187
Quadro IV. 35 - Variáveis binárias para o número total de viagens diárias	188
Quadro IV. 36 - Repartição modal para as diferentes categorias do total de viagens diárias	188
Quadro IV. 37 - Variáveis binárias para a densidade de oferta de estacionamento face à procura de estacionamento	189
Quadro IV. 38 - Repartição modal para as diferentes categorias da densidade de oferta de estacionamento face à procura de estacionamento vago	189
Quadro IV. 39 - Variáveis binárias para a proporção de estacionamento pago	190
Quadro IV. 40 - Repartição modal para as diferentes categorias da proporção de estacionamentos pagos	191

Quadro IV. 41 - Amostras para a possível combinação de variáveis binárias sobre políticas de estacionamento	191
Quadro IV. 42 - Amostras de pares GA para a possível combinação de variáveis binárias sobre políticas de estacionamento	191
Quadro IV. 43 - Repartições modais para possível combinação de variáveis binárias sobre políticas de estacionamento	192
Quadro IV. 44 - Amostras das variáveis binárias de combinação de variáveis de políticas de estacionamento	192
Quadro IV. 45 - Amostras das variáveis binárias da razão tempo de viagem BUS/AUTO	194
Quadro IV. 46 - Repartição modal para as diferentes gamas de relações tempos de viagem Bus/AUTO	194
Quadro IV. 47 - Amostras das variáveis binárias sobre a frequência média horária diária sem exclusão das áreas muito próximas	199
Quadro IV. 48 - Repartição modal para as diferentes gamas de frequências médias horárias diárias	199
Quadro IV. 49 - Amostras das variáveis binárias da velocidade comercial equivalente de BUS	201
Quadro IV. 50 - Repartição modal para as diferentes gamas de velocidades comerciais equivalentes de BUS	201
Quadro IV. 51 - Amostras disponíveis atendendo às diferentes combinações de variáveis	202

Capítulo V

Quadro V. 1 - Resultados dos modelos preliminares para o reajustamento dos valores discretos lineares pelos valores médios dos parâmetros de cada categoria, para o rendimento mensal do agregado	211
Quadro V. 2 - Resultados dos modelos preliminares para o reajustamento dos valores discretos lineares pelos valores médios dos parâmetros de cada categoria, para o número de veículos disponíveis diariamente <i>per capita</i> do agregado	211
Quadro V. 3 - Resultados dos modelos preliminares da influência do rendimento líquido mensal do agregado	212
Quadro V. 4 - Resultados do teste de <i>Wald</i> para o modelo MNL3.11 para avaliar a linearidade dos parâmetros da combinação de binárias para o rendimento mensal do agregado	216
Quadro V. 5 - Resultados dos modelos preliminares comparativos entre o rendimento líquido mensal do agregado e o rendimento líquido mensal do agregado <i>per capita</i>	217
Quadro V. 6 - Resultados dos modelos preliminares da influência do Sexo do inquirido	218

Quadro V. 7 - Resultados dos modelos preliminares da influência do escalão etário	219
Quadro V. 8 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	219
Quadro V. 9 - Resultados dos modelos preliminares da influência do nível de instrução	221
Quadro V. 10 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	221
Quadro V. 11 - Resultados dos modelos preliminares da influência da taxa de motorização familiar	223
Quadro V. 12 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	224
Quadro V. 13 - Resultados do teste de <i>Wald</i> para o modelo MNL3.29 para avaliar a linearidade dos parâmetros da combinação de binárias para a taxa de motorização familiar	226
Quadro V. 14 - Resultados dos modelos preliminares da influência da Licença de Condução	227
Quadro V. 15 - Resultados dos modelos preliminares da influência da distância mais curta em km entre centróides das zonas de geração e de atracção das viagens	228
Quadro V. 16 - Resultados do teste estatístico da razão das MVL com o modelo com especificação contínua como base	229
Quadro V. 17 - Resultados do teste estatístico da razão das MVL com modelo com especificação com combinação de binárias como base	229
Quadro V. 18 - Resultados dos modelos preliminares da influência do tempo total de viagem em minutos entre a zona de geração e de atracção das viagens	233
Quadro V. 19 - Resultados dos modelos preliminares da influência do tempo total de viagem em minutos entre a zona de geração e de atracção das viagens e do logaritmo do mesmo	234
Quadro V. 20 - Resultados dos modelos preliminares da influência do motivo da viagem	234
Quadro V. 21 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	235
Quadro V. 22 - Resultados dos modelos preliminares da influência da duração em minutos da atividade no destino da viagem	237
Quadro V. 23 - Resultados do teste estatístico da razão das MVL com modelo com especificação contínua como base	237
Quadro V. 24 - Resultados do teste de <i>Wald</i> para o modelo MNL3.120a para avaliar a linearidade dos parâmetros da combinação de binárias para a duração da ati	238
Quadro V. 25 - Resultados dos modelos preliminares com diferentes especificações para a variável duração da estadia	240

Quadro V. 26 - Resultados dos modelos preliminares da influência do numero total de viagens num dia	241
Quadro V. 27 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	241
Quadro V. 28 - Resultados dos modelos preliminares da influência da proporção de estacionamento relativamente ao total de viagens para uma zona de atracão	244
Quadro V. 29 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	244
Quadro V. 30 - Resultados dos modelos preliminares da influência da proporção de estacionamento pagos na zona de atracão	247
Quadro V. 31 - Resultados do teste estatístico da razão das MVL com modelo com especificação contínua como base	247
Quadro V. 32 - Resultados dos modelos preliminares da influência da dos tempos médios de viagem de Bus e de Automóvel	250
Quadro V. 33 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	251
Quadro V. 34 - Resultados do teste de <i>Wald</i> para o modelo MNL3.91 para avaliar a linearidade dos parâmetros da combinação de binárias para razão de tempos de viagem BUS/AUTO	252
Quadro V. 35 - Resultados dos modelos preliminares da influência da frequência média horária diária dos BUS	254
Quadro V. 36 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	255
Quadro V. 37 - Resultados do teste de <i>Wald</i> para o modelo MNL3.137 para avaliar a linearidade dos parâmetros da combinação de binárias para a frequência média diária	257
Quadro V. 38 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego da manhã	258
Quadro V. 39 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego do Meio da manhã	258
Quadro V. 40 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego da tarde	258
Quadro V. 41 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período laboral	258
Quadro V. 42 - Resultados dos modelos Preliminares da influência da velocidade comercial equivalente do BUS	259

Quadro V. 43 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base	261
Quadro V. 44 - Quadro síntese dos resultados dos modelos preliminares com as diferentes especificações de cada variável analisada	263
Quadro V. 45 - Possíveis combinações da disponibilidade de automóvel e de carta de condução e possíveis repercussões na atitude perante o automóvel	264
Quadro V. 46 - Resultados do modelo MNL exploratório combinando o número de automóveis diariamente à disposição do agregado <i>per capita</i> e a existência de carta de condução	265
Quadro V. 47 - Resultados da análise de correlação entre as variáveis automóveis diariamente à disposição do agregado <i>per capita</i> e licença de condução através do <i>software SPSS</i> ®	266
Quadro V. 48 - Resultados da análise de correlação entre as variáveis grupo etário e nível de instrução através do <i>software SPSS</i> ®	267
Quadro V. 49 - Resultados do modelo MNL exploratório combinando o grupo etário com o nível de instrução	267
Quadro V. 50 - Resultados da análise de correlação entre as variáveis sobre estacionamento através do <i>software SPSS</i> ®	268
Quadro V. 51 - Resultados do modelo MNL exploratório combinando as variáveis sobre estacionamento	269
Quadro V. 52 - Resultados dos modelos exploratórios combinando variáveis sobre estacionamento	270
Quadro V. 53 - Resultados da análise de correlação entre as variáveis contínuas velocidade comercial equivalente de BUS (V2) e distância mais curta entre centróides (D2) através do <i>software SPSS</i> ®	272
Quadro V. 54 - Resultados da análise de correlação entre a variável contínua velocidade comercial equivalente de BUS (V2) e as binárias da Duração Média da Viagem (Til) através do <i>software SPSS</i> ®	272
Quadro V. 55 - Resultados dos modelos exploratórios com combinação das variáveis velocidade comercial equivalente do BUS e distância mais curta entre centróides	273
Quadro V. 56 - Resultados dos modelos exploratórios com combinação das variáveis velocidade comercial equivalente do BUS, distância mais curta entre centróides e tempo médio de viagem	273
Quadro V. 57 - Amostras para os diferentes tipos de viagens analisadas para avaliar a influência do modo escolhido na 1ª viagem diária nas restantes viagens diárias	278
Quadro V. 58 - Análise dos modos utilizados nas viagens intermédias e finais do dia comparativamente ao modo utilizado na primeira viagem do dia para diferentes distâncias	279
Quadro V. 59 - Síntese das variáveis e especificações sugeridas para os modelos combinados	289

Capítulo VI

Quadro VI.1 - Resultados do modelo exploratório de base MNL7.40	292
Quadro VI.2 - Resultados do modelo exploratório base e comparação com os resultados acrescentando a variável Dx1	294
Quadro VI.3 - Comparação do modelo de referência MNL7.40 com o modelo com a adição das restantes variáveis socioeconômicas	295
Quadro VI.4 - Combinação dos parâmetros das variáveis <i>Idi</i> e <i>Ini</i> do modelo MNL7.37	297
Quadro VI.5 - Comparação do modelo MNL7.48 (sem rendimento) com o modelo de referência com a adição de todas as variáveis socioeconômicas	299
Quadro VI.6 - Comparação do modelo de referência MNL7.49 (sem instrução) com o modelo com a adição de todas as variáveis socioeconômicas	300
Quadro VI.7 - Comparação dos resultados do modelo de referência MNL7.48 com o mesmo modelo com a base de dados reduzida MNL7.48a	302
Quadro VI.8 - Comparação dos resultados do modelo de referência MNL7.48a com a inclusão do tempo de viagem MNL7.51	303
Quadro VI.9 - Combinação dos parâmetros das variáveis <i>Idi</i> e <i>Ini</i> do modelo MNL7.51	304
Quadro VI. 10 - Comparação dos resultados do modelo de referência com inclusão do tempo de viagem com o modelo de referência incluindo a duração da estadia no destino	305
Quadro VI. 11 - Comparação dos resultados do modelo de referência com inclusão do tempo de viagem com o modelo de referência incluindo a duração da estadia no destino	306
Quadro VI. 12 - Comparação dos resultados do modelo de referência com duração da estadia com o mesmo incluindo a duração da estadia no destino e o número de viagens diárias	308
Quadro VI. 13 - Comparação dos resultados do modelo de referência MNL7.54 com o modelo de referência incluindo o motivo da viagem	310
Quadro VI. 14 - Resultados de correlações entre as variáveis binárias sobre o motivo da viagem e sobre a duração das estadias através o <i>software</i> SPSS®	311
Quadro VI. 15 - Comparação dos resultados do modelo de referência MNL7.55 com o modelo sem duração da estadia no destino	313
Quadro VI. 16 - Comparação dos resultados dos modelos sem duração da estadia no destino (MNL7.56a) e sem motivo da viagem (MNL7.54)	314
Quadro VI. 17 - Modelo de referência MNL7.56a comparado com o modelo sem a variável tempo total médio de viagem por modo	316
Quadro VI. 18 - Comparação dos parâmetros do modelo de referência MNL7.57a com o mesmo modelo sem a variável disponibilidade diária de Moto no agregado	317
Quadro VI. 19 - Comparação dos parâmetros do modelo de referência MNL7.58 com	319

o mesmo modelo com base de dados reduzida devido às variáveis sobre estacionamento	
Quadro VI. 20 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58a com a introdução das variáveis sobre estacionamentos	320
Quadro VI. 21 - Avaliação das zonas de origem e de destino das viagens pedonais	321
Quadro VI. 22 - Comparação dos resultados dos modelos de referência com a introdução das variáveis sobre estacionamentos contínuas e binárias	323
Quadro VI. 23 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58a com a redução da base de dados aos inquéritos com dados sobre os TC e estacionamentos	324
Quadro VI. 24 - Análise dos resultados do modelo de referência MNL7.58b com a introdução das variáveis sobre TC	326
Quadro VI. 25 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58b com a introdução simultânea das variáveis sobre TC e estacionamento (combinação de binárias)	328
Quadro VI. 26 - Comparação dos resultados do modelo de referência com a introdução das variáveis sobre TC e estacionamento contínuas e combinação de binárias	330
Quadro VI. 27 - Comparação do modelo de referência MNL12.16 com o modelo sem as variáveis socioeconómicas consideradas menos relevantes	332
Quadro VI. 28 - Comparação do modelo de referência com o modelo apenas com as variáveis socioeconómicas consideradas mais relevantes utilizando a base de dados reduzida às alternativas mais relevantes	334
Quadro VI. 29 - Avaliação do impacto da variável motivo da viagem no modelo de referência com as variáveis socioeconómicas essenciais e com a base de dados reduzida às alternativas mais relevantes	336
Quadro VI. 30 - Avaliação do impacto no modelo de referência MNL12.17BWA da redução da base de dados	338
Quadro VI. 31 - Comparação dos resultados do modelo de referência MNL 12.17BWA_a com a introdução do tempo médio de viagem	339
Quadro VI. 32 - Modelo de referência MNL7.57 para a modelação sem políticas de mobilidade	345
Quadro VI. 33 - Modelo de referência MNL12.16BWA para a modelação com políticas de mobilidade	348
Quadro VI. 34 - Variáveis binárias de combinação de variáveis socioeconómicas para o modelo MNL99.3a	361
Quadro VI. 35 - Variáveis binárias de combinação de variáveis socioeconómicas com taxas de motorização familiar por adulto para o modelo MNL99.9b	361
Quadro VI. 36 - Modelos de combinações de variáveis socioeconómicos	362

MNL99.3a_BWA_NCA_b	
Quadro VI. 37 - Modelos de combinações de variáveis socioeconômicos	366
MNL99.9b_BWA_NCA_b	
Quadro VI. 38 - Comparação dos ajustamentos dos modelos com diferentes especificações para as variáveis binárias de combinação de variáveis socioeconômicas	368
Quadro VI. 39 - Quadro com valores dos parâmetros obtidos de modelos discretos em estudo internacionais	371
Quadro VI. 40 - Quadro resumo das gamas de valores dos parâmetros analisados obtidos de modelos discretos em estudo internacionais	375
Quadro VI. 41 - Quadro comparativo das gamas de valores dos parâmetros analisados obtidos de modelos discretos em estudo internacionais e nos modelos sugeridos no caso de estudo	376
 Capítulo VII	
Quadro VII. 1 - Perfis tipo extremos de utilizadores para os modos mais relevantes	385
Quadro VII. 2 - Perfis tipo de utilizadores intermédios	388
Quadro VII. 3 - Gamas de variações utilizadas nas análises de sensibilidade dos modelos	392
Quadro VII. 4 - Alterações das variáveis a considerar nos hipotéticos cenários futuros	400
 Anexos	
Quadro A. 1 - Resultados combinando as variáveis carta de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >1km e ≤2km	523
Quadro A. 2 - Resultados combinando as variáveis carta de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA > 2km	527

Lista de Variáveis

D2	Variável contínua para a distância mais curta entre os centróides espaciais da cada zona de tráfego de geração e de atração, em km
DX1	Variável binária para distâncias mais curtas entre os centróides espaciais $\leq 1\text{km}$
DX2	Variável binária para distâncias mais curtas entre os centróides espaciais $> 1\text{km}$ e $\leq 2\text{km}$
DX3	Variável binária para distâncias mais curtas entre os centróides espaciais $> 2\text{km}$ e $\leq 5\text{km}$
DX4	Variável binária para distâncias mais curtas entre os centróides espaciais $> 5\text{km}$ e $\leq 10\text{km}$
DX5	Variável binária para distâncias mais curtas entre os centróides espaciais $> 10\text{km}$
De	Variável contínua para a duração de estadia/atividade, em minutos
De1	Variável binária duração da estadia $\leq 60\text{min}$ (curta duração)
De2	Variável binária duração da estadia $> 60\text{min}$ & $\leq 120\text{min}$ (média duração)
De3	Variável binária duração da estadia $> 120\text{min}$ & $\leq 240\text{min}$ (média duração)
De4	Variável binária duração da estadia $> 240\text{min}$ & $\leq 480\text{min}$ (longa duração)
De5	Variável binária duração da estadia $> 480\text{min}$ (muito longa duração)
Fa	Variável contínua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 10:30 e as 12:30, em TC/h
Fa0	Variável contínua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 10:30 e as 12:30 sem zeros, em TC/h
Fa1	Variável contínua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 10:30 e as 12:30 com valores em pares GA muito próximas, em TC/h
Fm	Variável contínua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 9:30, em TC/h

Fm0	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 9:30 sem zero, em TC/h
Fm1	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 9:30 com valores em pares GA muito próximas, em TC/h
Fl1	Variável binária Frequência média horária potencial de TC 7:30-19:30 com valores em pares GA muito próximas ≤ 2 TC/h
Fl2	Variável binária Frequência média horária potencial de TC 7:30-19:30 com valores em pares GA muito próximas > 2 & ≤ 4 TC/h
Fl3	Variável binária Frequência média horária potencial de TC 7:30-19:30 com valores em pares GA muito próximas > 4 & ≤ 12 TC/h
Fl4	Variável binária Frequência média horária potencial de TC 7:30-19:30 com valores em pares GA muito próximas > 12 TC/h
Ft	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 17:00 e as 19:30, em TC/h
Ft0	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 17:00 e as 19:30 sem zeros, em TC/h
Ft1	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 17:00 e as 19:30 com valores em pares GA muito próximas, em TC/h
Ftl	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 19:30, em TC/h
Ftl0	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 19:30 sem zeros, em TC/h
Ftl1	Variável continua para a frequência média horária de TC potencialmente à disposição entre a zona de geração e de atração entre as 7:30 e as 19:30 com valores em pares GA muito próximas, em TC/h
FNumInd	Número de indivíduos do agregado doméstico particular
ldb	Variável discreta sobre o escalão etário do indivíduo
ld1	Variável binária para o escalão etário 1:0-9 anos
ld2	Variável binária para o escalão etário 2:10-19 anos

Id3	Variável binária para o escalão etário 3:20-34 anos
Id4	Variável binária para o escalão etário 4:35-64 anos
Id5	Variável binária para o escalão etário 5:+65 anos
In	Variável discreta sobre o nível de instrução do inquirido
In1	Variável binária para o nível de instrução 1 - Analfabeto
In2	Variável binária para o nível de instrução 2 - Ensino Básico
In3	Variável binária para o nível de instrução 3 - Ensino Secundário
In4	Variável binária para o nível de instrução 4 - Ensino Superior
Lc	Variável binária de disponibilidade de Licença/Carta de Condução (0- Não tem; 1- Tem)
Mot	Variável binária sobre a disponibilidade diária de motociclos no agregado
Mt	Variável discreta sobre o motivo da viagem
Tr	Variável binária para viagens para trabalho/Emprego
Rg	Variável binária para Viagens de Regresso ao Domicilio
Es	Variável binária para viagens para a Escola
CS	Variável binária para viagens para Compras e/ou Serviços
Lz	Variável binária para viagens para Lazer
Mtrp	Variável discreta sobre o Modo de Transporte utilizado (Variável dependente: BW - Bus+Walk; B - Bus; OB - Other Bus; M - Moto; W - Walk; A - Auto)
NC	Variável continua taxa de motorização familiar ou N° total de automóveis ligeiros à disposição diária do agregado per capita, em veíc./residente
NC0	Variável binária para automóveis ligeiros disponíveis per capita = 0 veic./residente

NC1	Variável binária para automóveis ligeiros disponíveis per capita >0 & $\leq 0,25$ veic./residente
NC2	Variável binária para automóveis ligeiros disponíveis per capita $>0,25$ & $\leq 0,5$ veic./residente
NC3	Variável binária para automóveis ligeiros disponíveis per capita $>0,5$ & $\leq 0,75$ veic./residente
NC4a	Variável binária para automóveis ligeiros disponíveis per capita $>0,75$ veic./residente
Nca	Variável continua Nº total de automóveis ligeiros à disposição diária do agregado por adulto, veic./adulto
Na0	Variável binária para automóveis ligeiros disponíveis por adulto = 0 veic./adulto
Na1	Variável binária para automóveis ligeiros disponíveis por adulto > 0 & $<0,5$ veic./adulto
Na2	Variável binária para automóveis ligeiros disponíveis por adulto $\geq 0,5$ & $<1,0$ veic./adulto
Na3	Variável binária para automóveis ligeiros disponíveis por adulto ≥ 1 veic./adulto
nV	Variável continua para o número total de viagens de cada indivíduo por dia, viag./dia
nV1	Variável binária para número total de viagens de cada indivíduo por dia ≤ 2 viag./dia
nV2	Variável binária para número total de viagens de cada indivíduo por dia >2 & ≤ 4 viag./dia
nV3a	Variável binária para número total de viagens de cada indivíduo por dia >4 viag./dia
OV	Variável continua razão entre a oferta total de estacionamento na zona de atração e o número total de viagens extrapoladas feitas para a zona de atração em 2000
OV1	Variável binária razão Oferta total de estacionamento/procura total de Viagens para zona de atração $\leq 0,075$
OV2	Variável binária razão Oferta total de estacionamento/procura total de Viagens para zona de atração $>0,075$ & $\leq 0,1$
OV3	Variável binária razão Oferta total de estacionamento/procura total de Viagens para zona de atração $>0,1$ & $\leq 0,15$
OV4	Variável binária razão Oferta total de estacionamento/procura total de Viagens para zona de atração $>0,15$
PO	Variável continua razão entre o número de estacionamento pagos na zona de atração e o número total de estacionamento existentes na zona de atração em 2000 , ou proporção de estacionamento pagos

PO1	Variável binária razão estacionamento Pagos/Oferta total de estacionamento na zona de atração $\leq 0,25$
PO2	Variável binária razão estacionamento Pagos/Oferta total de estacionamento na zona de atração $> 0,25$ & $\leq 0,5$
PO3	Variável binária razão estacionamento Pagos/Oferta total de estacionamento na zona de atração $> 0,5$
R	Variável discreta rendimento líquido do agregado familiar em euros (valores máximos de cada escalão 1-500€, 2-1000€, 3-1500€, 4-2000€, 5-2500€)
Rb	Variável discreta com o escalão de rendimento líquido mensal do agregado familiar em euros (1:0-500€, 2:501-1000€, 3:1001-1500€, 4:1501-2000€, 5:>2000€)
R1	Variável binária para rendimento líquido do agregado familiar 0-500€
R2	Variável binária para rendimento líquido do agregado familiar 501-1000€
R3	Variável binária para rendimento líquido do agregado familiar 1001-1500€
R4	Variável binária para rendimento líquido do agregado familiar 1501-2000€
R5	Variável binária para rendimento líquido do agregado familiar >2000€
RC	Variável continua rendimentos líquidos mensais do agregado per capita, em Euros/residente/mês
RC1	Variável binária para rendimentos líquidos mensais do agregado per capita ≤ 250 €/residente/mês
RC2	Variável binária para rendimentos líquidos mensais do agregado per capita > 250 & ≤ 500 €/residente/mês
RC3a	Variável binária para rendimentos líquidos mensais do agregado per capita > 500 €/residente /mês
Sx	Variável binária sobre o sexo do Inquirido (M=0;F=1)
TBA	Variável continua Razão tempos médios de viagem em Bus/AUTO entre as zonas de Geração e Atração
TB1	Variável binária para razão entre a Duração Média da Viagem em BUS e o Duração Média da Viagem em AUTO ≤ 1
TB2	Variável binária para razão entre a Duração Média da em BUS e o Duração Média da Viagem em AUTO > 1 & $\leq 1,5$
TB3	Variável binária para razão entre a Duração Média da Viagem em BUS e o Duração Média da Viagem em AUTO $> 1,5$ & $\leq 2,0$

TB4	Variável binária para razão entre a Duração Média da Viagem em BUS e o Duração Média da Viagem em AUTO >2,0
Til	Variável contínua genérica Logaritmo da duração média apreendida da viagem por modo (em minutos) entre GA com a duração apreendida real quando é o modo escolhido
Tia	Variável contínua genérica duração média apreendida da viagem por modo (em minutos) entre GA com a duração apreendida real quando é o modo escolhido
V2	Variável contínua para a velocidade comercial equivalente por Bus, em km/h
V21	Variável binária para velocidade comercial equivalente por Bus ≤ 10 km/h
V22	Variável binária para velocidade comercial equivalente por Bus $10 < \leq 20$ km/h
V23	Variável binária para velocidade comercial equivalente por Bus ≥ 20 km/h

Lista de Acrónimos

A	Automóvel particular - Auto
ADP	Agregado Doméstico Privado
AMP	Área Metropolitana do Porto
AMTP	Autoridade Metropolitana de Transportes do Porto
B	Autocarro com carreira regular - Bus
BD	Base de dados
BRT	Bus Rapid Transit
BW	Combinação de autocarro regular com andar a pé - Bus+Walk
BWA	Bus - Walk - Auto
CMP	Câmara Municipal do Porto
DpP	Disponibilidade para Pagar
FEUP	Faculdade de Engenharia da Universidade do Porto
FMVL	Função de Máxima Verosimilhança Logarítmica
GA	Geração-Atração
GP	Grande Porto
IIA	Independente de Alternativas Irrelevantes
IID	Independentes e Identicamente Distribuídos
IMTT	Instituto da Mobilidade e dos Transportes Terrestres
INE	Instituto Nacional de Estatística
IT	Índice de Transferência
LL	<i>Log-likelihood</i>
LRT	Light Rail Transit
M	Motas e motociclos - Moto
MC	Monte Carlo
MNL	Modelo Logit Multinomial
MVL	Máxima Verosimilhança Logarítmica
MXL	Modelo Logit Misto
NCa_b	Base de dados com informação completa do número de adultos do agregado familiar e das suas das viagens diárias
OB	Autocarros sem carreira regular - Other Bus
OD	Origem-Destino
OSAS	Cenário Otimista Social e Ambientalmente Sustentável
OSAT	Cenário Otimista Social e Ambientalmente Tradicional

PSAS	Cenário Pessimista Social e Ambientalmente Sustentável
PSAT	Cenário Pessimista Social e Ambientalmente Tradicional
QMC	Quasi-Monte Carlo
SIG	Sistemas de Informação Geográfica
SLL	Verosimilhança Logarítmica Simulada
TC	Transportes Coletivos
TET	Teste Estatístico de Transmissibilidade
TI	Transporte Individual
TLC	Teorema do Limite Central
W	Pedonal ou pedestre - Walk

Capítulo I

Introdução

1. Enquadramento do Tema

O objeto de estudo do presente trabalho foi a avaliação do potencial que diferentes alterações ao perfil socioeconómico dos Portugueses, bem como de diferentes estratégias inseridas em políticas urbanas e de mobilidade sustentável, poderão ter para influenciar o processo de tomada de decisão na escolha do modo de transporte a utilizar nas viagens diárias. A este último nível pretendeu-se especificamente quantificar o potencial impacto que diferentes estratégias de gestão da mobilidade, nomeadamente de incentivo ao uso de modos sustentáveis e de restrições ao livre uso do automóvel particular, bem como de políticas que promovam a densificação urbana, poderão ter na alteração de hábitos, de forma a promover uma mobilidade mais sustentável.

Atendendo à evolução que se tem verificado na maioria das cidades de média e grande dimensão a nível mundial nas últimas décadas, nomeadamente com o aumento do uso do automóvel particular e com o decréscimo do uso dos transportes mais sustentáveis, principalmente os Transportes Coletivos (TC), existe a necessidade de analisar a eficácia de diferentes estratégias na tentativa de inverter esta tendência. A experiência internacional tem demonstrado que uma conciliação de políticas que conjuguem a utilização do automóvel particular com os transportes mais sustentáveis é a estratégia com melhores condições de promover uma mobilidade sustentável a todos os cidadãos (Vuchic, 1999). A comodidade no uso do automóvel particular dificilmente poderá ser igualada por outros modos de transporte. A sua disponibilidade para certos tipos de viagens e entre determinadas zonas torna-o em alguns casos a única opção verdadeiramente disponível, desaconselhando a aplicação de restrições severas indiscriminadas ao seu uso. Ainda assim, viagens há cujas características intrínsecas as tornam propícias a ser realizadas por modos mais sustentáveis e que em diversas cidades, nomeadamente europeias, revelaram ser passíveis de ser captadas com sucesso.

2. Identificação de Problemas e Definição dos Objetivos

O excesso de dependência em relação ao automóvel particular e a tradicional pouca atratividade dos modos mais sustentáveis deve-se principalmente à descoordenação entre políticas de ordenamento do território e as políticas de transporte/mobilidade. As autoridades nacionais, regionais e locais devem encetar esforços no sentido de coordenarem as diferentes políticas de desenvolvimento, não só horizontalmente, entre os diferentes

sistemas coexistentes nas áreas urbanas, como verticalmente, de forma a que permitam a integração das políticas/sistemas locais nas políticas/sistemas regionais e nacionais. Esta coordenação permitirá não só uma melhor eficiência na aplicação dos escassos recursos humanos e financeiros, como uma melhor eficácia na implementação de um efetivo desenvolvimento sustentável.

No que diz respeito às políticas de mobilidade, parece claro que o rumo a seguir passa por políticas que incentivem o uso dos sistemas de transportes mais sustentáveis, em detrimento do automóvel particular. O transporte coletivo, ou de massas, apresenta-se como uma solução mais sustentável a longo prazo, económica e socialmente, mas apenas nas zonas com maiores densidades populacionais e com uma grande diversidade de usos do solo, de forma a tornar os elevados investimentos iniciais justificáveis. Em paralelo devem ser implementadas medidas que integrem nas viagens por automóvel particular os reais custos económicos e sociais que a opção pelo mesmo acarreta para a sociedade em geral. Estas políticas, no entanto, devem ser cuidadosamente estudadas e integradas de modo a não diminuir a mobilidade de certos grupos de cidadãos, ou a partir de certas zonas urbanas, pela simples imposição de restrições sem a oferta de alternativas viáveis. Estas políticas devem assim permitir aos cidadãos ter acesso a bens e serviços considerados básicos para a sua subsistência, e permitir o desenvolvimento sustentável, económica e socialmente, dessas áreas urbanas.

Assim, a tendência das políticas urbanas deverá potenciar o aumento da densidade de uso do solo nas áreas centrais, propícias a serem servidos por TC, limitando o alastramento dos subúrbios que implicam um maior uso do automóvel particular e dificultam a implementação de um sistema de TC eficiente. Estas políticas de ordenamento do território deverão ser acompanhadas de políticas de mobilidade urbana que tornem os modos alternativos mais atrativos para as viagens diárias, coordenadas com medidas que, onde possível e desejável, tornem o uso do automóvel particular mais caro e difícil, ou que pelo menos diminuam a sua conveniência comparativamente às alternativas.

É, no entanto, aparente que em muitas cidades de muitos países se verifica presentemente uma situação em que existe um grande fosso entre o funcionamento desejado e o funcionamento real do sistema de transportes. Verifica-se em muitos casos não apenas uma incapacidade de inverter tendências passadas, mas também um agravar das mesmas. Esta é uma situação comum à generalidade das cidades Portuguesas. Por outro lado é inquestionável que existem muitos outros casos de sucesso internacional na implementação de políticas mais sustentáveis de mobilidade, onde se verifica que modos alternativos ao automóvel podem ser verdadeiramente competitivos na captação de múltiplos tipos de deslocações urbanas.

Atendendo a esta constatação, torna-se extremamente importante perceber de que forma diferentes fatores internos e externos ao sistema de transportes, de forma isolada ou

integrada, dificultam ou potenciam a concretização de novos paradigmas de mobilidade, onde exista uma menor dependência relativamente ao transporte individual motorizado. Um primeiro aspeto que necessita de ser avaliado de forma a compreender o potencial impacto relativamente aos padrões de mobilidade, prende-se com a topologia urbana, e, particularmente com a sua densidade de ocupação e de oportunidades. Mais especificamente ligada às questões da mobilidade, é importante que se possa compreender, de forma mais objetiva e quantificada, a influência que a qualidade dos sistemas de transportes alternativos ao automóvel particular, especialmente os TC, tem ou poderá ter, na escolha do modo de transporte em viagens diárias. Em paralelo, justifica-se também avaliar o impacto que medidas que restrinjam o livre uso do automóvel particular possam ter nas opções modais. Finalmente, não pode deixar de se estudar o impacto potencial que as dinâmicas de evolução socioeconómicas que se verificam presentemente, e as que podem vir a ocorrer no futuro tanto próximo como mais distante, podem vir a materializar, e o condicionamento que tenderão a impor à adoção e concretização de novas políticas de transportes.

Apesar de existirem estudos internacionais que incluem separadamente partes destas análises, não se conhecem modelos com a sua combinação na globalidade e os seus resultados são específicos para os cenários analisados, não sendo de aplicação generalizada direta.

Com o presente estudo pretende-se assim, contribuir para quantificar os possíveis impactos de estratégias de densificação urbana, bem como de incentivo ao uso do TC e de restrição ao livre uso do automóvel particular, no processo de tomada de decisão da escolha modal, considerando aspetos relacionados com as características pessoais/sociais de cada indivíduo com as características de cada viagem, conjugadas com características do sistema de mobilidade urbana disponível. Mais concretamente pretende-se contribuir para a identificação da melhor forma de incluir todos estes aspetos em modelações matemáticas que possam simular comportamentos reais, atuais e futuros, e que o façam com uma complexidade que não os tornem demasiado onerosos de estabelecer.

Para a sua definição pretende-se quantificar quais os diferentes fatores mais determinantes bem como avaliar a sua melhor especificação para uma descrição mais realista. Espera-se que os resultados do estudo possam vir a servir de base para a definição de linhas orientadoras aos decisores políticos na afetação de recursos e no estabelecimento de estratégias para os diferentes subsistemas que compõem o complexo sistema urbano, de forma mais eficiente e potencialmente com resultados mais concretos.

3. Abordagem Metodológica

Para avaliar o impacto de diferentes estratégias de organização do território e da mobilidade, bem como das características socioeconómicas de cada indivíduo nas suas escolhas modais diárias, optou-se por estudar uma possível formulação matemática que permitisse essa

quantificação. Apesar de atualmente os modelos baseados em atividades permitirem uma modelação teoricamente mais realista, considerando o encadeamento da totalidade das atividades diárias de um indivíduo, a sua relativa complexidade de processamento e de aquisição de dados tem levantado algumas críticas e levado a alguma dificuldade de generalização da sua aplicação. Estas críticas fundamentam-se na relação entre o investimento necessário ao seu desenvolvimento em larga escala e as contrapartidas de precisão das previsões comparativamente aos modelos tradicionais. No presente estudo, a limitação da forma de levantamento dos dados pré-existentes utilizados, inviabilizou a possibilidade de utilizar um modelo baseado em atividades.

Assim, optou-se pelo estudo dos modelos mais tradicionais. Nos finais da década de 70 (Ortúzar e Willumsen, 2001; Train, 2003) surgiram os modelos de escolha discreta que, com uma base teórica consecutivamente consolidada, se apresentam como ferramentas potentes para a modelação das escolhas de indivíduos. Estes modelos, além de se basearem em teorias de comportamento, permitem modelar escolhas com base em variáveis contínuas mas também discretas, isto é, seleccionadas de um conjunto finito de opções. Nos modelos de escolha discreta assume-se que cada alternativa tem associada uma utilidade bruta para um determinado indivíduo, e que o analista não possui a informação completa sobre os elementos considerados para a escolha de uma alternativa. A opção escolhida será a que maximizar a utilidade apreendida pelo utilizador. Assim, o analista assume que a utilidade é representada por dois componentes: uma mensurável, função de atributos mensuráveis e outra parte aleatória, que reflete o gosto pessoal de cada indivíduo, agregado a eventuais erros de observação ou medição. A probabilidade de um indivíduo escolher uma determinada opção será o valor esperado de uma função probabilística considerando todos os possíveis valores dos fatores não observáveis. A forma como o analista considera esta parte aleatória define o tipo de modelo a utilizar, uma vez que para calcular esta probabilidade, o respetivo integral pode ser avaliado de três maneiras: através de expressões de forma completamente fixa (por exemplo os Logit multinomiais), simulação completa (por exemplo os Probit) e através de simulação e forma fixa parciais (por exemplo os Logit mistos).

Dos diversos modelos de escolha discreta desenvolvidos desde os anos 70, o modelo mais conhecido e vastamente utilizado é o Logit Multinomial. Dos modelos aplicados a casos reais referidos na bibliografia até 2010 a sua esmagadora maioria foi realizada através de Logit Multinomiais. A sua designação deriva da utilização de uma curva logística (ou função sigmoide) para representar a probabilidade de escolha de uma alternativa entre as restantes. Entre as suas limitações mais relevantes destaca-se a assunção de padrões de substituição proporcionais e a independência entre alternativas (IIA), e a assunção de erros independentes e igualmente distribuídos (IID), quer entre alternativas quer entre observações. Assim, apesar de contestado para problemas especialmente complexos do ponto de vista conceptual devido às limitações referidas, o seu uso para extensas modelações exploratórias envolvendo o teste

de formulações muito variadas e de um número elevado de variáveis, é justificado atendendo à facilidade de cálculo e de interpretação de resultados (Train, 2003; Hensher *et al*, 2005), tendo sido selecionada como estrutura de modelação base para o desenvolvimento do presente projeto. Tal procedimento permite, de forma estruturada, a identificação das variáveis explicativas mais relevantes, da sua melhor especificação e forma de combinação, para posteriormente integrar formulações matemáticas mais complexas capazes de mitigar algumas das suas limitações.

No presente projeto optou-se pelo desenvolvimento de um conjunto extensivo e sistemático de análises exploratórias que visam avaliar o contributo que as diferentes variáveis consideradas podem dar para a compreensão do processo de decisão das escolhas modais. Este procedimento permite a caracterização do impacto individual de cada variável analisada, bem como das suas diferentes possíveis especificações e avaliar a eventual redundância parcial na capacidade explicativa e eventuais inter-relações verificadas quando utilizadas de forma combinada. Para realizar estas extensas análises exploratórias foi utilizada a informação recolhida pelo Instituto Nacional de Estatística (INE) no Inquérito à Mobilidade na Área Metropolitana do Porto (AMP) no ano 2000. Aos dados recolhidos nesse estudo foi adicionada informação considerada relevante relativamente à oferta de TC através de um Sistema de Informação Geográfica (SIG), bem como das diferentes tipologias e densidades de oferta de estacionamento existentes na área urbana do Porto. Dificuldades na aquisição de dados sobre os sistemas de mobilidade limitaram a sua análise à área norte da NUT III do Grande Porto. Com esta informação foi possível organizar uma base de dados para a estimação dos modelos recorrendo ao *software* de análise econométrica *NLogit 4*^o.

Atendendo à bibliografia internacional, e à informação disponível, foi selecionado um conjunto de diferentes variáveis com potenciais características para permitirem explicar o processo de tomada de decisão na escolha modal. A metodologia adotada para explorar quais as variáveis independentes com maior potencial, e quais as suas melhores especificações, foi iniciada com a introdução de cada uma destas isoladamente em modelos Logit Multinomiais, avaliando o seu ajustamento global e a coerência tanto de sinais como do comportamento dos parâmetros.

Após a definição de um conjunto de variáveis mais proeminentes, foi testada a viabilidade da sua combinação num único modelo Logit Multinomial. Este teste foi realizado de forma iterativa, com sucessivas alterações de combinações e analisando as melhorias no ajustamento global e a eventual manutenção da coerência dos parâmetros. Este processo permitiu a obtenção de um conjunto de variáveis que combinadas apresentam potencial para explicar, de forma muito significativa e de forma mais integrada e abrangente do que é normal, o processo de escolha modal. Os resultados dos seus parâmetros foram posteriormente comparados com resultados de estudos internacionais para avaliar a sua coerência e validade.

A possível dificuldade dos modelos em revelar eventuais relações de interdependências entre variáveis, nomeadamente as relacionadas com aspetos sociais, levou à exploração das mesmas na base de dados utilizada. A verificação de algumas tendências interessantes motivou o estudo da sua introdução nos modelos através de variáveis binárias para cada uma das combinações consideradas relevantes. Estas combinações caracterizam grupos sociais com comportamentos similares em termos de valorização das diferentes alternativas, e a sua definição permite a introdução de variáveis aparentemente não correlacionadas adequadas à modelação utilizada. Esta metodologia permite a mitigação de problemas de correlação e de falta de representatividade para interações mais complexas entre as diferentes variáveis sociais.

Com conhecimento das variáveis mais relevantes e dos modelos mais estabilizados e com melhores resultados em termos de ajustamento global e de coerência de comportamento de parâmetros, foram realizadas simulações para avaliar as disparidades entre as previsões dos modelos e as repartições modais efetivamente verificadas na base de dados utilizada para a modelação. Estas simulações foram complementadas com análises de sensibilidade a diferentes alterações de cada uma das variáveis explicativas isoladamente. Após este estudo aplicaram-se os modelos a hipotéticas condições definidoras de perfis-tipo que caracterizam determinado tipo de cidadãos, estimaram-se as suas hipotéticas repartições modais e avaliaram-se as implicações dessas alterações.

Este processo concluiu-se com o teste de possíveis cenários de evolução futura das condições socioeconómicas dos residentes no espaço urbano, bem como da implementação de possíveis diferentes políticas de ordenamento e mobilidade urbana aplicados aos modelos sugeridos. Este procedimento permite testar o potencial que as diferentes variáveis explicativas identificadas apresentam para, de forma conjunta, influenciarem os padrões de repartição modal que presentemente se verificam. Em concreto procurou-se, através de análises exploratórias avaliar o grau de irreversibilidade e inevitabilidade dos padrões de mobilidade atualmente verificados em espaços urbanos. Paralelamente procurou-se também identificar as dinâmicas socioeconómicas que presentemente se verificam e se perspetivam para o futuro, e que maior potencial revelam para influenciar esses padrões, bem como as medidas políticas que se mostram com algum potencial para condicionar ou mesmo inverter as tendências atuais.

4. Estrutura da Tese

Para a execução dos estudos referidos, no capítulo II é feito um enquadramento dos problemas de mobilidade atuais e das filosofias de intervenção com maior potencial para mitigar os mesmos, promovendo uma mobilidade equitativa e sustentável. São apresentados os aspetos gerais a considerar na modelação da mobilidade dos cidadãos, focando as

características genéricas dos modelos de escolha discreta como parte dos modelos de quatro passos.

No capítulo III são apresentadas as fundamentações teóricas gerais dos modelos de escolha discreta, apresentando-se também os seus modelos mais comuns. Para cada um dos diferentes tipos de modelos são apresentadas as suas fundamentações teóricas específicas, vantagens e desvantagens, formulações, processos de estimação, e formas de avaliação/interpretação dos resultados. Como elemento de referência e ponto de partida dos processos de modelação desenvolvidos, termina-se este capítulo com a apresentação dos principais resultados obtidos em estudos internacionais de modelação de escolhas modais.

O capítulo IV apresenta a caracterização do caso de estudo: área de estudo, padrões e níveis de mobilidade, definidos através da caracterização socioeconómica dos viajantes, e sistemas de mobilidade envolvidos. É ainda feita uma descrição cuidada das dificuldades verificadas na aquisição e tratamento dos dados, procurando-se desta forma definir um conjunto de procedimentos que garantem uma organização eficiente de processos de caracterização de sistemas de transportes urbanos complexos. São, especificamente apresentadas as simplificações e opções tomadas para a obtenção de uma base de dados relevante com dados credíveis disponíveis para grande parte dos registos. É explicada a forma como foram considerados os dados sobre os sistemas de mobilidade, e como foram realizadas as análises dos mesmos. Descrevem-se detalhadamente, justificando o seu potencial, as diferentes variáveis consideradas para a análise, seleccionadas entre as sugeridas em estudos internacionais específicos sobre transportes ou tradicionalmente consideradas para a modelação de transportes. Finalmente, faz-se também referência às dificuldades verificadas no tratamento dos dados pré-existentes e na recolha de dados sobre os sistemas de mobilidade, apresentando-se recomendações sobre a tipologia e características fundamentais desejáveis em bases de dados, de forma a mitigar as dificuldades enunciadas. Estas recomendações deverão possibilitar a realização de estudos técnico-científicos o mais rigorosos e produtivos possível.

No capítulo V são apresentados os resultados de modelações preliminares com cada uma das variáveis avaliadas de forma isolada das restantes. São testadas diferentes especificações e avalia-se a linearidade do seu comportamento.

Para algumas variáveis, que revelaram resultados menos espectáveis ao nível do tipo e intensidade do seu efeito, quando analisadas de forma isolada, são apresentados os resultados de análises combinatórias preliminares com outras variáveis. Este procedimento visou a identificação de eventuais situações de correlação ou complementaridade especiais entre determinadas variáveis que justificassem esses resultados não esperados.

É igualmente realizada uma análise exploratória na base de dados sobre a estabilidade do modo de transporte utilizado ao longo das viagens do dia, procurando perceber-se até que ponto o efeito de encadeamento do processo de escolha modal ao longo das viagens diárias, deve ser um aspeto relevante a ter em consideração. Esta análise deverá permitir em estudos futuros uma ponderação melhor fundamentada sobre o(s) elo(s) mais relevante(s) nessa cadeia de decisões no sentido de influenciar uma potencial e desejável alteração de comportamentos.

No capítulo VI são estimados modelos com combinações de variáveis partindo de modelações base que incluem apenas as variáveis até aí declaradamente mais relevantes. Estas combinações são realizadas de modo iterativo na tentativa de obter melhorias no ajustamento global dos modelos e para melhor poder identificar e incorporar eventuais interdependências e efeitos integrados em diferentes variáveis. Atendendo aos resultados e às restrições que impõem à base de dados passível de ser utilizada, são apresentados os modelos mais promissores para a definição das variáveis mais relevantes e das suas hipotéticas influências nas escolhas modais.

O capítulo finaliza com um conjunto de resultados de análises exploratórias complementares correspondentes a um estudo mais detalhado de eventuais relações de interdependência entre as variáveis socioeconómicas mais relevantes, potenciais caracterizadoras de grupos com comportamentos distintos em termos de valorização das opções modais. Atendendo aos resultados obtidos, é igualmente apresentado um estudo de viabilidade da sua integrabilidade nos modelos definidos através da definição de variáveis binárias representativas das combinações caracterizadoras de cada grupo, com o objetivo de expurgar eventuais correlações e tornar as variáveis independentes entre si, melhorando a adequabilidade dos modelos utilizados.

O capítulo VII apresenta os resultados da aplicação dos modelos anteriormente desenvolvidos na produção de simulações e previsões para o futuro. Atendendo à análise dos valores dos parâmetros obtidos para as diferentes variáveis, testam-se quais as combinações que potenciam cada um dos modos mais relevantes, definindo-se hipotéticos perfis-tipo de utilizadores preferenciais dos mesmos.

Para avaliar a precisão das simulações dos modelos desenvolvidos, são definidos perfis socioeconómicos com representatividade na base de dados, e comparam-se as repartições modais existentes com as estimadas pelos modelos. Após o estudo da precisão das simulações, realizam-se previsões considerando variações dos valores das diferentes variáveis utilizadas. Numa primeira análise essas variações, dentro de gamas plausíveis, são introduzidas na base de dados para cada uma das variáveis isoladamente. Esta análise permite avaliar a sensibilidade dos modelos a possíveis variações na base de dados, estimando o previsível impacto de cada variável.

Numa segunda fase, estas alterações nas variáveis na base de dados, são realizadas combinando diversas variáveis. Estas alterações são definidas de forma a representarem plausíveis evoluções das variáveis, caracterizadoras e específicas de hipotéticos cenários futuros.

O capítulo finaliza com a avaliação qualitativa de eventuais relações entre as variáveis que se revelaram mais determinantes para as escolhas modais atuais e as variáveis caracterizadoras do nível socioeconómico do agregado familiar. A avaliação da evolução das últimas é geralmente mais inteligível devido a sua ampla documentação e comprovação por entidades idóneas, possibilitando a estimação da evolução das variáveis mais relevantes, apesar de indiretamente, de forma potencialmente mais simples e célere.

O estudo termina com o Capítulo VIII onde são apresentadas as conclusões mais relevantes, as tendências verificadas e melhor fundamentadas e onde se sugerem possíveis melhorias na aquisição e especificação dos dados de base. São igualmente apresentados possíveis evoluções do presente estudo e eventuais futuras aplicações da metodologia desenvolvida.

Capítulo II

Enquadramento

1. Introdução

Na maioria dos países europeus, incluindo Portugal, tem-se assistido nas últimas décadas a um acentuado aumento da posse e uso do automóvel particular. Este facto tem levado a um aumento dos acidentes rodoviários, da ocupação do solo e, principalmente, a uma diminuição da qualidade de vida dos cidadãos na maioria das cidades de média e grande dimensão, devido ao aumento dos congestionamentos rodoviários, do estacionamento ilegal, e do aumento da poluição atmosférica e sonora, que causam.

Estes problemas surgiram nos países mais desenvolvidos mas, com o crescimento acelerado das taxas de motorização e utilização do automóvel particular que se tem registado, têm-se alargado à maioria dos países em desenvolvimento e surgirão naturalmente nos países menos desenvolvidos se não se tomarem as devidas cautelas.

Grande parte destes automóveis particulares é motorizado através de combustíveis fósseis, implicando este aumento do parque automóvel um agravamento da dependência em relação aos produtos petrolíferos da maioria das economias mundiais, condicionando consideravelmente o desenvolvimento económico e social das mesmas.

Os problemas referidos, existentes na maioria das cidades de média e grande dimensão atualmente em todo o mundo, desenvolveram-se mais rapidamente nas cidades europeias e japonesas, do que nas norte-americanas. Esta situação deveu-se ao facto de as mesmas terem originado na sua maioria de cidades mais antigas, pré era automóvel, desenvolvidas a partir de infraestruturas construídas à dimensão humana e não do transporte individual (TI) motorizado. No entanto, esta realidade só atrasou o seu aparecimento nas cidades norte-americanas, desenvolvidas a maioria já no séc. XX e atendendo à dimensão do automóvel, estando atualmente também plenamente implantados em muitas cidades norte-americanas, verificando-se também o rápido e intenso agravamento nas cidades dos países emergentes.

Esta constatação vem comprovar que o desenvolvimento sustentável das metrópoles modernas só ocorrerá se as mesmas forem encaradas como sistemas complexos com miríades interações entre todas as atividades, serviços e infraestruturas nelas inseridas.

Um dos sistemas mais importantes, que influencia de forma relevante os restantes é o sistema de transportes. Beneficiar uma infraestrutura, um modo específico, ou um determinado modo de vida, implica impor desequilíbrios que podem prejudicar grupos de cidadãos e/ou determinadas zonas urbanas, levando a uma diminuição geral da qualidade de vida urbana. Os sistemas de transporte devem, portanto, ser encarados como sistemas multimodais únicos e complexos.

Assim, no presente capítulo pretende-se salientar a relevância de um planeamento coordenado e integrado de todos os subsistemas que coexistem nas cidades modernas. São caracterizados os modos de transportes mais comuns e referem-se as filosofias mais recentes com maior potencial para permitir um desenvolvimento sustentável das áreas urbanas. Para a alteração dos padrões de mobilidade atuais para modos mais sustentáveis é necessário entender como as escolhas modais são realizadas. Neste sentido são apresentados os aspetos gerais a atender na modelação de transportes, caracterizando-se de forma geral as famílias de modelos mais relevantes para a análise das escolhas modais.

A descrição do estado da arte ao nível das metodologias de análise e das tipologias de modelações de referência, é acompanhada de uma análise crítica relativamente ao seu potencial e condições de aplicabilidade, salientando-se as suas limitações e potencial de evolução.

2. A Interação entre os Sistemas de Transporte e o Desenvolvimento das Áreas Urbanas

O crescimento dinâmico das cidades e as suas alterações em termos de área/forma, impõem mudanças significativas nos seus sistemas de transporte. As tipologias de soluções adotadas por sua vez, influenciam o modo como as cidades crescem, as suas características funcionais e o ambiente natural e construído oferecido aos seus cidadãos. Os problemas referidos surgiram em grande parte devido à falta de coordenação entre as políticas de mobilidade e o planeamento urbano realizado no passado, que não atenderam a estas relações entre as cidades e os seus sistemas de transporte a longo prazo. Muitas das decisões tomadas durante a formulação e implementação das políticas municipais, intermunicipais, regionais e até nacionais, foram muitas vezes focando apenas a construção e melhoramento de infraestruturas para o TI sem atenderem às consequências a longo prazo nos sistemas globais de transporte. Décadas de políticas vocacionadas para o TI, criaram hábitos, rotinas e estilos de vida que mesmo com os grandes investimentos em alternativas ao TI verificados atualmente, não se conseguem quebrar facilmente, ou mesmo pacificamente.

Uma qualidade de vida aceitável em cidades, enquanto aglomerados humanos organizados, só é possível se qualquer um dos seus cidadãos tiver *mobilidade*, isto é, a capacidade de se deslocar entre uma origem e um destino dentro da área urbana em modos de transporte

privados ou públicos. Esta mobilidade é geralmente dificultada quando há grandes distâncias a ultrapassar, quando existem condições climatéricas adversas, quando há grandes rampas a vencer, mas acima de tudo, quando os meios para se deslocar estão indisponíveis, os custos são demasiado elevados ou existem outros fatores de exclusão (Grava, 2003).

Na maioria dos países podem definir-se três tipos de políticas de transporte (Vuchic, 1999):

- *Restringir o uso do TI à capacidade das cidades*: Esta política baseia-se no conceito de que as cidades têm grande valor social e histórico, e que nenhum destes deve ser sacrificado para permitir viagens ilimitadas por TI nas suas áreas urbanas;
- *Reconstruir/adaptar as cidades para permitir o máximo de viagens por TI*: Para atingir este objetivo as cidades devem ser redesenhadas e construídas de forma diferente, alterando de forma drástica o carácter físico e social das áreas urbanas;
- *Desenvolvimento equilibrado*: Deve realizar-se um planeamento urbano incluindo na definição das políticas de desenvolvimento urbano uma coordenação com os sistemas de transportes considerando-os multimodais, onde cada modo de transporte tem o seu papel a desempenhar numa lógica de complementaridade mútua.

O primeiro tipo de política, restrições às viagens por automóvel particular, tem dados alguns resultados positivos em alguns centros históricos e em partes de cidades com diferentes dimensões, principalmente na Europa. Tem a vantagem de assegurar o carácter humano das cidades evitando os efeitos negativos das viagens por modos motorizados e dos congestionamentos de tráfego. No entanto, as consequências que resultam da sua implementação por defeito - sem nenhum ajustamentos nas vias urbanas e no sistema de transportes - descoordenadas de uma política sustentável, leva a problemas de congestionamento crónicos e à ineficácia dos diversos sistemas de transporte. Restrições excessivas ao uso do TI motorizado às áreas centrais podem privar as cidades dos benefícios resultantes da alta mobilidade individual que permitem, nomeadamente a numerosas atividades de negócio, comércio e lazer.

O segundo tipo de políticas, adaptação das cidades ao crescente tráfego de TI motorizados, surge da visão simplista de aumentar a capacidade das suas infraestruturas para acabar com os crescentes congestionamentos rodoviários e as suas consequências no ambiente urbano. À medida que os volumes de tráfego vão atingindo a capacidade das infraestruturas urbanas, as autoridades tentam resolver o problema aumentando as vias, as ruas, os estacionamento, e principalmente implementando redes rodoviárias de alta capacidade, de modo a servir os grandes volumes de tráfego de atravessamento, por exemplo, através de vias circulares interiores e exteriores. Os impactos negativos destas políticas sobre a forma física das cidades e da qualidade de vida que oferecem aos seus cidadãos, são considerados secundários

relativamente à conveniência de viajar por TI motorizado. Os inconvenientes para aqueles que não o podem, ou não querem, utilizar também são negligenciados.

Este tipo de abordagem foi seguido em muitas cidades que cresceram de forma acentuada nas últimas décadas, encontrando-se bons exemplos destas políticas em muitas cidades norte americanas. Este tipo de políticas tem levado ao aumento da área das cidades, tornando a utilização do TI cada vez mais imprescindível, e tornando a utilização de modos mais sustentáveis, transportes coletivos (TC), bicicleta ou a pé, economicamente inviáveis ou fisicamente impossíveis. A distância entre destinos, a dispersão dos habitantes por grandes áreas, a necessidade de atravessar vias com tráfego elevado ou grandes parques de estacionamento, têm limitado a mobilidade de cidadãos adolescentes, de pessoas idosas ou com dificuldades motoras, reduzindo o papel dos TC a um mero serviço social para os que não podem, ou não querem, utilizar o automóvel.

Curiosamente, este tipo de políticas tem falhado no seu principal objetivo de diminuir/acabar com os congestionamentos rodoviários. A construção de centenas de quilómetros de vias coletoras ou autoestradas urbanas, de grandes parques de estacionamento que dominam o ambiente visual de áreas centrais e suburbanas, não reduziu significativamente os congestionamentos rodoviários comparativamente a outras cidades que não adotaram a mesma orientação em termos de política de mobilidade/transporte. Pelo contrário, porções significativas da sua população - os que não possuem automóvel particular - sofreram reduções na sua mobilidade relativamente à situação anterior às alterações resultantes da implementação deste tipo de políticas, e a qualidade de vida urbana em termos ambientais tem diminuído consideravelmente.

O terceiro tipo de políticas, para um desenvolvimento equilibrado, baseia-se no conceito de que uma cidade é um sistema complexo com miríades de atividades e serviços, entre os quais se encontra o sistema de transportes. O seu funcionamento ótimo só poderá ser alcançado se todos os seus serviços e funções estiverem coordenados. Assim, o sistema de transportes deve servir de forma eficiente os seus cidadãos e interagir com as restantes funções e serviços existentes nas áreas urbanas. Não deve ser demasiado restritivo nem deve dominar sobre outras funções ou modos de vida. Haverá viagens que dependerão e serão melhor servidas inevitavelmente por TI, mas também haverá muitas que poderão ser perfeitamente servidas por modos alternativos mais sustentáveis se as condições que oferecerem forem atrativas. Será particularmente nas áreas de maior densidade de atividades, residencial, comércio ou serviços, que os modos alternativos de maior capacidade e ambientalmente mais sustentáveis serão uma alternativa incontestavelmente melhor face ao TI motorizado.

Será este tipo de política de transporte/mobilidade o único capaz de alcançar um sistema de transportes eficiente e um ambiente urbano de alta qualidade para todos os cidadãos.

Abreu Silva (2007), num estudo sobre a modelação da influência dos padrões de uso do solo na mobilidade sustentável em áreas urbanas, verificou que a acessibilidade proporcionada pelas infraestruturas de transporte é uma condição necessária para a indução do crescimento urbano ou a revitalização das áreas deprimidas, embora não seja condição suficiente. Por outro lado estes efeitos tendem a diminuir com o aumento do nível global de acessibilidade, ou seja, níveis de acessibilidade elevados (principalmente através do automóvel particular) tornam-nos irrelevantes nas decisões sobre a localização de muitas atividades. Refere ainda que as políticas de ordenamento do território não devem ser encaradas como um modo direto de reduzir a utilização do automóvel particular, mas antes como um modo de aumentar a atratividade relativa dos TC e modos não motorizados. O modelo desenvolvido permitiu verificar que residentes em áreas urbanas densas centrais, compactas e com importantes níveis de mistura de empregos e habitantes, tenderão a utilizar mais o TC e modos não motorizados e a possuir menores níveis de motorização, enquanto os residentes em zonas suburbanas tenderão a possuir uma taxa de motorização mais elevada e a utilizar o automóvel mais frequentemente. Atendendo aos resultados, sugere as seguintes orientações (por vezes difíceis de implementar ou inaceitáveis por determinadas autarquias) em termos de configuração espacial e políticas de ordenamento do território:

- Concentração da população e do emprego nas zonas mais centrais e com padrões de urbanização mais tradicionais;
- Permitir o surgimento de polos urbanos importantes em zonas exteriores organizados em torno de áreas bem servidas de transporte coletivo;
- Permitir a intensificação do uso dos solos apenas nas zonas bem servidas por transporte coletivo, reduzindo as áreas de expansão nas restantes zonas, e repensar a política de infraestruturas rodoviárias de modo a que a construção de novas vias não crie pressões contrárias às políticas de contenção urbana.

3. Tipologias dos Sistemas de Transporte

A variedade de sistemas de transporte é extensa, abrangendo desde os mais simples sistemas pedonais, baratos e ideais para viagens de curta extensão, passando por sistemas mais complexos como os BRT (Bus Rapid Transit) ou LRT (Light Rail Transit) que podem operar quer em vias próprios quer partilhando os arruamentos urbanos, até sistemas mais pesados como os metros subterrâneos, com grande capacidade de transporte mas com necessidades financeiras e técnicas muito exigentes.

3.1. Transporte Privado

Inclui os modos pedestre, o ciclista, os motociclos, o automóvel particular, e outros similares. Oferecem aos utilizadores uma grande liberdade de movimento no tempo e no espaço. No entanto, têm características e impactos ambientais muito distintos entre si.

O *tráfego pedestre/pedonal*, representa de longe o melhor (mais eficiente) modo de transporte para viagens de curta distância, sendo o mais conveniente, barato e geralmente o mais rápido para deslocações planas até 400 a 500 metros (Vuchic, 1999), utilizando-se geralmente esta distância como a máxima admissível para o distanciamento entre paragens de carreiras urbanas regulares de autocarros (White, 2002; Tyler, 2002; Grava, 2003; e Weatherall, 1997). É geralmente aceite que distâncias pedonais planas até 800 metros reduzem em 25 a 50% a percentagem de pedestres e que 80% das viagens a pé se efetua a distâncias inferiores a 1 km (Grava, 2003; Olzewski *et al.*, 1999). Geralmente esta distância é maior em grandes centros urbanos com áreas muito atrativas e uma grande variedade de atividades ao dispor do peão, e menor em pequenos centros urbanos. As suas maiores limitações são a baixa velocidade, o esforço humano necessário principalmente em terreno acidentado, os obstáculos à sua circulação quando o tráfego pedestre é negligenciado em relação aos restantes modos, e a sua inconveniência em condições climatéricas adversas. Devido às suas superiores características para viagens curtas, de grande densidade e diversidade, o modo pedestre é de longe o modo mais eficiente e conveniente para áreas centrais (“baixas” das cidades), terminais de transporte, praças públicas, centros comerciais, ruas urbanas comerciais, campus universitários, etc. Nestas zonas as infraestruturas devem ser cuidadosamente planeadas de modo a torná-lo atrativo e seguro, dando-lhe preferência sobre o tráfego motorizado ou segregando-os completamente. Cidades com alta criminalidade nas ruas, com más infraestruturas pedestres sem proteção contra o tráfego motorizado e condições climatéricas adversas, e poucas atrações ao longo dos trajetos pedestres, não podem ser consideradas atrativas ou com escala humana.

O *tráfego ciclista*, é o modo de transporte veicular mais económico. É muito menos confortável que o automóvel, necessita de esforço humano e depende em grande parte das condições climatéricas e dos declives a vencer, mas é o modo ideal para quem pretende realizar exercício físico e ideal para viagens curtas em cidades, subúrbios, espaços de lazer, campus universitários, etc.. Devido aos impactos ambientais bastante mais reduzidos que os do automóvel e ao seu baixo custo, são um modo de transporte a promover, verificando-se mesmo a nível nacional a sua integração nos planos diretores municipais mais recentes (por exemplo, Marinha Grande, Aveiro, Leiria, Figueira da Foz).

Os restantes modos de transporte privados são motorizados. Destes, o mais comum, principalmente nos países desenvolvidos, é o *automóvel particular*. As suas principais vantagens são a grande conveniência e independência na hora e direção das viagens, a sua “rapidez” e o seu conforto pessoal. Estas características tornam-no muito atrativo para utilizadores individuais, famílias ou pequenos grupos. Como sistema de transporte em meio urbano tem grandes limitações, uma vez que causa congestionamentos com volumes de tráfego (principalmente de pessoas) relativamente baixos, tem elevados custos sociais devido a esses mesmos congestionamentos e aos seus impactos ambientais, tem um consumo

energético médio por pessoa-quilómetro viajado muito mais elevado que qualquer outro modo alternativo, e tem um elevado custo social e pessoal devido aos frequentes acidentes rodoviários. O impacto mais importante, e muitas vezes ignorado, da dependência excessiva em relação ao automóvel particular são as suas repercussões na forma e carácter das áreas urbanas, nomeadamente no empobrecimento das relações sociais, na deterioração de áreas históricas e construídas à escala humana, no aumento da segregação de grupos sociais ou minorias étnicas, na maior poluição atmosférica, sonora e visual, entre outros.

Apesar da maioria dos países desenvolvidos terem taxas de motorização da população bastante elevadas (em Portugal em 2001 existiam cerca de 333 veículos ligeiros de passageiros matriculados por 1000 habitantes segundo dados do Instituto Nacional de Estatística (INE, 2002a e 2002b) e European Commission (2011)), existem percentagens consideráveis de cidadãos que não possuem essa opção (não-condutores). Neste grupo de cidadãos incluem-se os que não possuem TI privado e os que optam por não o ter ou utilizar. Na primeira categoria incluem-se partes da população que não tem rendimentos familiares suficientes para possuir automóvel, jovens e pessoas idosas, ou simplesmente todos os que não estão qualificados para conduzir quer pelas circunstâncias quer por opção, enquanto na segunda se incluem os cidadãos que estão qualificados, e por vezes até têm automóvel, mas que tomam a opção de não o utilizar, principalmente em áreas centrais de grande tráfego ou com grandes congestionamentos.

Entre os não-condutores, existem alguns que utilizam o automóvel indiretamente, por exemplo as crianças transportadas pelos pais para as instituições escolares, que sobrecarregam o sistema de transporte utilizando o automóvel de forma muito pouco eficiente. Esta sobrecarga é agravada com os idosos que apesar de terem as suas capacidades reduzidas necessitam de utilizar o automóvel (por exemplo táxis) pela inexistência de modos alternativos convenientes.

Outro tipo de não-condutores bastante significativo são os cidadãos com baixos rendimentos familiares que não têm possibilidade de adquirir um automóvel para as suas deslocações diárias. Segundo Vuchic (1999), este grupo de cidadãos representa entre 15 a 25% da população das cidades. Geralmente podem utilizar os TC (quando oferecem serviços adequados) ou andar a pé para a maioria dos destinos. Em áreas rurais são grupos virtualmente sem mobilidade. Este grupo é bastante relevante para as políticas de mobilidade, devido ao seu elevado potencial para a utilização do automóvel mediante a melhoria das suas condições socioeconómicas futuras.

Os não-condutores por opção, muitos licenciados a conduzir, vivem geralmente nas cidades e consideram mais conveniente e económico não possuir um automóvel, preferindo andar a pé, utilizar TC ou táxis, ou ocasionalmente os automóveis de aluguer. As suas motivações podem

dever-se a fatores sociais, ambientais ou éticos, mas serão um grupo que não terá tendência a mudar de modo sendo portanto utilizadores fiéis às alternativas.

O tamanho de cada grupo de cidadãos referidos depende não só da sua idade, distribuição espacial, situação económica, ou outra característica socioeconómica, mas também da disponibilidade e qualidade dos modos alternativos de transporte. Atualmente na maioria das cidades mundiais os grupos de cidadãos que optam por não possuir automóvel particular e utilizar modos alternativos são extremamente reduzidos devido à reduzida conveniência e diferença de mobilidade sentida na utilização de TC ou andando a pé, em cidades maioritariamente adaptadas ao automóvel particular.

3.2. Transportes de Aluguer

Esta categoria inclui vários modos de transporte que pelas suas características se encontram entre o automóvel particular e os TC. Estes modos caracterizam-se fundamentalmente não só por serem transportes públicos que transportam quantidades de passageiros geralmente muito menores que os TC normais, mas também por prestarem um serviço mais caro e personalizado, de qualidade próxima da do automóvel particular. Entre este tipo de transporte podem incluir-se os táxis e outros serviços de miniautocarros que satisfazem deslocações muito específicas no tempo e no espaço. Não têm percursos nem horários pré-estabelecidos, pelo que se adaptam facilmente a alterações na procura. Têm um papel complementar em relação ao automóvel particular e aos TC, oferecendo em cidades de grande dimensão um serviço personalizado mais valorizado e geralmente mais conveniente que o automóvel particular, enquanto em áreas de baixa densidade de ocupação os sistemas de miniautocarros que respondam a chamadas proporcionam um sistema mais barato que os sistemas regulares. Em muitas cidades os transportes de aluguer são o modo mais eficiente (e por vezes os únicos existentes) para servir pessoas de mobilidade reduzida.

3.3. Transportes Coletivos

Outrora o modo de transporte de eleição para a movimentação de passageiros em áreas urbanas, o TC tem vindo a perder sucessivamente quotas de mercado para o automóvel particular, com o aumento da posse e uso dos mesmos devido à melhoria dos rendimentos dos agregados familiares. Apesar do seu uso ter decrescido sistematicamente ao longo do séc. XX, a sua importância nos sistemas de transporte urbanos continua incontestável (apesar de muitas vezes pouco reconhecida), principalmente nas áreas centrais. Representam um serviço básico que diferencia as áreas urbanas das áreas rurais, oferecendo um serviço público disponível à maioria da população (se devidamente implementado), com grande capacidade, menores necessidades de espaço, e impactos negativos por passageiro transportado significativamente menores comparativamente aos impostos pelo automóvel particular. Apresenta-se como uma solução melhor para áreas urbanas de média e grande dimensão, caracterizadas por uma grande densidade de geração/atração de viagens, permitindo um

funcionamento mais eficiente das mesmas com toda a diversidade de atividades e densidades que as distinguem.

Apesar de Vuchic (1981), considerar que o sistema de TC só são economicamente viáveis em cidades com mais de 50 000 habitantes, a realidade é que em muitas cidades de menor dimensão estes cumprem uma função social, permitindo alguma mobilidade a partes significativas da sua população. Giannopoulos (1990) refere que quase todas as cidades com menos de 300 000 habitantes têm predominantemente apenas uma rede de transportes públicos, normalmente autocarros. Dos diferentes papéis sociais que os TC desempenham, salientam-se o transporte de crianças para as escolas; a distribuição dos passageiros que utilizam terminais rodoviários, ferroviários ou aéreos, periféricos; o transporte de passageiros em viagens pendulares casa-trabalho; o transporte de cidadãos idosos em áreas centrais; entre outros.

Em suma, os TC podem ser divididos em duas grandes funções:

- *Servir passageiros cativos*: Transportar os cidadãos que não têm acesso, não podem conduzir ou não querem utilizar o automóvel particular, ou seja, realizar um serviço social para os cidadãos que não têm um modo alternativo conveniente para se deslocar. Apesar de esta função ser necessária em toda a área urbana, é essencial nas zonas com muitas crianças, idosos, cidadãos de baixos rendimentos ou outros que não utilizem o automóvel particular;
- *Servir de alternativa conveniente e competitiva para os passageiros não cativos (cuja opção modal pode ser alterada)*: Ou seja, para os cidadãos que apesar de terem acesso ao automóvel particular, o consideram tanto ou menos atrativo que os TC alternativos, em termos de tempo de viagem, custo, conveniência, etc..

Vuchic (1999) considera que os TC representam o único modo de transporte que possibilita a existência de grandes cidades que funcionem de modo eficiente e que mantenham o seu carácter humano. Salienta que as cidades economicamente fortes, socialmente saudáveis e que permitem uma vivência sustentável, não são planeadas em função do automóvel particular mas possuem sistemas de transporte multimodais, que incluem extensos sistemas de TC que funcionam de modo eficiente, e infraestruturas para tráfego pedestre igualmente extensos e atrativos.

3.4. Transportes Coletivos em Meio Urbano

A seleção dos sistemas de TC urbanos mais apropriados para cada situação, depende da sua atratividade para os passageiros, da sua eficácia em relação à mobilidade que proporcionam e da sua eficiência económica. Para cada conjunto de condições de operacionalidade há a necessidade de considerar a hipótese de um ou mais modos que ofereçam a relação

investimento/desempenho mais apropriada, isto é, prestem o serviço necessário a custos aceitáveis.

As suas características básicas referem-se à tecnologia, ao tipo de operação e à infraestrutura que utilizam. Estas características são até certo ponto interdependentes. A tecnologia depende em grande parte do tipo de via sobre o qual o sistema vai circular, isto é, em vias de tráfego de arruamentos urbanos, os autocarros normais são a tecnologia dominante, mas à medida que se deseje segregar mais os vários modos de transporte (aumento da capacidade e da segurança), os veículos que utilizam carris passam a prevalecer, como por exemplo os metros ligeiros de superfície. Esta realidade deve-se ao facto de o aumento dos custos de construção e operação de infraestruturas segregadas só se justificar para sistemas com grande capacidade capazes de responder a níveis muito elevados de procura.

4. A Necessidade de Sistemas de Transporte Coletivos

Nos países desenvolvidos e industrializados as áreas urbanas albergam mais de dois terços da sua população (Vuchic, 1999). Assim, os problemas urbanos: económicos, ambientais, de segurança, de saúde ou de segregação social/cultural; afetam direta e indiretamente a grande maioria da população, pelo que a sua resolução é de importância capital para a vitalidade, prosperidade e competitividade económica de cada país.

O alastramento das cidades levou a uma grande separação física de diferentes grupos económicos e étnicos da população, entre as áreas centrais e os subúrbios ou entre diferentes zonas de residência e de trabalho, representando um obstáculo à resolução de problemas de ordem social e económica. Ao contrário do que acontece em vilas e aldeias, a maioria dos destinos dos cidadãos em movimento em cidades não se encontra numa distância conveniente para aceder a pé (até 1 a 2km). A maioria dos sistemas de transporte atuais agrava este problema, incentivando ao uso do TI em detrimento dos modos ambientalmente mais sustentáveis, excluindo parte das populações urbanas que não podem, ou não querem, utilizar o automóvel particular. A sua resolução passa por implementar um sistema de transporte que promova a *mobilidade* de todos os seus cidadãos. Neste contexto, a *mobilidade* será a possibilidade de qualquer cidadão alcançar qualquer bem, serviço ou atividade, que se forem de particular importância para a sociedade (essenciais) traduzem a sua *mobilidade básica*. Pode-se assim dizer, que a mobilidade traduz a qualidade e eficiência operacional de uma área urbana (Grava, 2003).

É internacionalmente aceite que sistemas de TC de elevada qualidade, que permitam uma boa mobilidade dos cidadãos, geralmente contribuem para o aparecimento de usos do solo mistos - por exemplo, de comércio e serviços ou de comércio e residenciais - estimulando a criação de centros de grande atividade e de empreendimentos residenciais mais diversificados. Enquanto certos usos do solo como o industrial e de lazer, têm maior

viabilidade em áreas de baixa densidade, outros há cuja viabilidade é otimizada nas áreas de maior densidade, tais como a prestação de serviços, negócios, institucionais e financeiros ou estabelecimentos de ensino. Eventos culturais e sociais, como concertos, exposições, espetáculos culturais e desportivos ou convenções, são outros usos que geralmente impõem altas concentrações de pessoas. Para permitir a existência de todas estas atividades dentro da área de influência de uma cidade, essenciais para a sua vitalidade, os sistemas de transportes urbanos devem conseguir servir eficientemente uma grande variedade de densidades de usos do solo e de volumes de tráfego. Nas zonas com maior densidade, são os TC, principalmente os de maior capacidade, os que melhor resposta dão às diferentes necessidades de mobilidade dos cidadãos.

A não inclusão nos custos para realizar viagens por TI das despesas decorrentes da implementação das infraestruturas rodoviárias que utiliza (pontes, viadutos, arruamentos urbanos, repavimentações, entre outros), das externalidades ambientais, ou da repercussão em termos de dependência energética, leva à ilusão de que o automóvel é sempre mais competitivo, incentivando ao uso do mesmo em detrimento de outros social e ambientalmente mais sustentáveis. Este problema fomenta paralelamente a tendência para subvalorizar o valor económico das áreas suburbanas em relação às áreas centrais, potenciando o alastramento das áreas urbanas devido a, em geral, ser mais económica a deslocação de empreendimentos das áreas centrais para áreas suburbanas mais afastadas do que reconverter as infraestruturas existentes nas primeiras. A população que eventualmente permaneça nestas áreas centrais enfrenta um aumento da insegurança, uma diminuição nas oportunidades de emprego e da possibilidade de manter um nível de vida aceitável, acelerando ainda mais o declínio destas áreas. Este alastramento leva igualmente a que a implementação de um sistema de TC de qualidade seja economicamente inviável completando o ciclo vicioso, promovendo o aumento do uso do transporte particular e agravando ainda mais os problemas referidos.

Assim, pode-se dizer que os grandes aglomerados económicos que as cidades representam, são o espelho da economia e da qualidade de vida dos respetivos países e que a continuação das políticas urbanas do passado contribuirá para agravamento dos problemas verificados atualmente. Existe a necessidade de realizar um planeamento integrando todos os vetores que afetam a qualidade de vida e o desenvolvimento económico, nomeadamente a interação entre o uso do solo e o sistema de transportes. Nesta interação há que atender às necessidades de todos os intervenientes no desenvolvimento social e económico, e promover uma equidade de oportunidades para todos os cidadãos, estipulando metas e objetivos a alcançar e definindo táticas para as atingir da forma mais sustentável possível.

5. A Mobilidade por Transporte Coletivo em Meio Urbano

Todos os cidadãos têm direito a uma mobilidade mínima que lhes permita aceder às atividades básicas para o seu desenvolvimento económico e social. Estas atividades básicas incluem o acesso a (VTPI,2007):

- Serviços de emergência (polícia, bombeiros, ambulâncias, etc.);
- Serviços públicos;
- Serviços de Saúde;
- Alimentação e Vestuário;
- Educação e Emprego;
- Correio;
- Entrega de Mercadorias;
- Atividades sociais e recreativas.

Esta mobilidade básica, por questões de equidade social deve ser avaliada para situações não usuais ou para as mais problemáticas, de modo a permitir o acesso mesmo a cidadãos com maiores dificuldades financeiras e/ou deficiências motoras.

Apesar de conceitos similares, *Mobilidade e Acessibilidade*, têm vertentes distintas em termos de aplicação. Enquanto a *mobilidade* se refere à possibilidade e capacidade de deslocação de um dado local para outro a custos aceitáveis e com duração razoável, a *acessibilidade* traduz a facilidade com que um cidadão consegue realizar ou alcançar uma atividade económica, social ou informação (Meyer e Miller, 2001; Grava, 2003). O objetivo fundamental de uma política de transportes urbanos é, portanto, efetivamente garantir a mobilidade de todos os seus cidadãos, mas esta apenas é efetiva se permitir a acessibilidade em condições similares a todas as atividades existentes na sua área. Em geral numa área urbana podem disponibilizar-se várias infraestruturas para melhorar a mobilidade dos seus cidadãos e mesmo assim ter zonas/atividades às quais é difícil aceder.

Atendendo a tudo o que foi exposto anteriormente pode ainda alargar-se o objetivo das políticas de transportes em meio urbano à melhoria da mobilidade de todos os seus cidadãos de *forma sustentável*. E este objetivo só poderá ser alcançado através de sistemas que reduzam, onde possível e viável, a dependência em relação ao automóvel particular.

Segundo Wegener (1995), a abordagem tradicional desde os anos 50 tem sido o ciclo uso do solo - transportes, definido como:

- A distribuição dos usos do solo, tais como o residencial, industrial ou comercial, na área urbana determina a localização das atividades humanas de viver, trabalhar, comprar, educar ou divertir;
- A distribuição das atividades humanas no espaço requer interações espaciais ou viagens no sistema de transportes de forma a ultrapassar a distância entre os locais das atividades;
- A distribuição das infraestruturas no sistema de transportes cria oportunidades para interações espaciais que podem ser medidas como acessibilidade;
- A distribuição da acessibilidade no espaço, por sua vez, determina decisões de localização e resulta em alterações no sistema de uso do solo.

Apesar de coerente e aparentemente lógico, segundo Wegener (1995) este ciclo não foi totalmente corroborado pela realidade. Estudos realizados nos Estados Unidos comprovaram que nos anos 90 se verificava que (Wegener, 1995):

- Locais perto de estações ferroviárias em geral não tinham atraído mais empreendimentos ou valorizado mais o mercado imobiliário que locais mais remotos;
- Investimentos em TC não tinham levado a deslocamentos da população e de empregos das áreas suburbanas para as áreas centrais;
- As maiores proporções de empreendimentos residenciais suburbanos ocorreram em cidades sem vias rápidas.

Estas evidências demonstraram que o modelo de planeamento urbano mais comum: de que os transportes moldam a cidade e, portanto, as políticas de transporte são o modo correto de influenciar o desenvolvimento dos espaços urbanos, e que os transportes são função do uso do solo e, portanto, as políticas de uso do solo são a forma correta de influenciar os transportes; não correspondeu ao desenvolvimento real de muitas áreas urbanas americanas.

Este aparente enfraquecimento da interdependência entre o uso do solo e o sistema de transportes apresenta-se como um inconveniente numa época em que se tenta desesperadamente diminuir os impactos negativos, ambientais, sociais e económicos, da dependência do uso do automóvel particular. Segundo Wegener (1995) a principal causa da evolução referida resulta fundamentalmente das condições socioeconómicas existentes. O modelo proposto nos anos 50 traduzia a realidade da época em que a acessibilidade era um recurso escasso, com redes viárias deficitárias e poucos cidadãos com acesso ao automóvel particular. Atualmente a maioria das cidades mundiais apresenta grandes redes viárias e são na sua maioria facilmente acessíveis (pelo menos a quem tem recursos para isso), o que aliado ao facto de o uso das vias urbanas ser na maioria das situações gratuito e o preço dos combustíveis ainda não ser demasiado elevado, torna o custo de viajar por automóvel

particular atrativo. Estes factos, aliados aos baixos preços imobiliários das zonas suburbanas (comparativamente às áreas centrais), à tranquilidade, à qualidade ambiental, ao prestígio social e a proximidade à natureza, tornam os subúrbios das cidades cada vez mais atrativos levando ao alastramento das manchas urbanas. Outro facto que tem contribuído para esta evolução é a descoordenação verificada nas políticas de planeamento a nível regional e municipal, na falta de controlo do desenvolvimento das áreas urbanas fundamentalmente devido à competição entre municípios e a permissão da descentralização de empregos para novas áreas suburbanas.

Na maioria das áreas urbanas a nível mundial, o baixo custo (aparente) de utilização das infraestruturas de transporte aliadas a um planeamento deficiente e a um mercado imobiliário desregulado, levam a que políticas que pretendam influenciar a localização e os comportamentos em termos de viagens unicamente através de incentivos estejam condenadas ao insucesso (Vuchic, 1999). Esta realidade justifica o facto de grandes investimentos para melhorar os sistemas de TC por si só não levem ao surgimento de maiores densidades de atividades perto dos seus locais de embarque ou que empreendimentos orientados para o TC em zonas de uso do solo mistos não incentivem a redução do uso do automóvel particular.

Pode assim dizer-se que apenas uma combinação sinérgica de medidas de incentivo e de restrições poderão ter perspectivas de alterar o panorama em termos de escolhas modais. Da perspectiva dos incentivos, há que encorajar o desenvolvimento de empreendimentos de uso misto de alta densidade e promover modos de transporte ambientalmente sustentáveis, como os TC, o modo pedonal ou ciclista. Na perspectiva das restrições devem-se adotar políticas que impeçam o alastramento das áreas urbanas através de controlo mais rigoroso das densidades de usos do solo permitidas e de medidas que dificultem o livre uso do automóvel particular, através, por exemplo, de limites de velocidade, de restrições ao acesso a determinadas zonas a determinados utilizadores (por exemplo, funcionários condutores que necessitam de estacionamento de longa duração) ou em determinados períodos do dia, do custo dos estacionamentos ou através do custo dos combustíveis.

Esta combinação de políticas de incentivos e restrições já são prática comum em alguns países europeus, nomeadamente na Holanda, Grã-Bretanha e países Escandinavos, apresentando resultados de consumo de energia em transporte *per capita* em média cerca de um quarto do consumo Norte-Americano (Wegener, 1995).

Atendendo a tudo o que foi referido, propõe-se o esquema da figura II -1 para definir as interações existentes entre os principais vetores de atuação para garantir a mobilidade dos cidadãos em meio urbano. A mobilidade dos cidadãos nas áreas urbanas atuais pode dividir-se fundamentalmente em 2 grupos de modos de transportes - o automóvel particular e as suas várias alternativas mais sustentáveis. Mediante investimentos eficientes para melhorar o seu desempenho atendendo às expectativas dos cidadãos, os modos alternativos podem oferecer

uma mobilidade comparável à do automóvel particular. No entanto, esses investimentos deverão ser incluídos num pacote coerente de políticas que também diminuam a conveniência do automóvel particular até um nível competitivo para os modos alternativos. Considerando a figura II -1, os principais fatores que influenciam a mobilidade, principalmente no sentido de a tornar sustentável no futuro, são a densidade de uso do solo, a mobilidade permitida por automóvel particular e a mobilidade alcançada com transportes alternativos (pedonal, ciclista e coletivos).

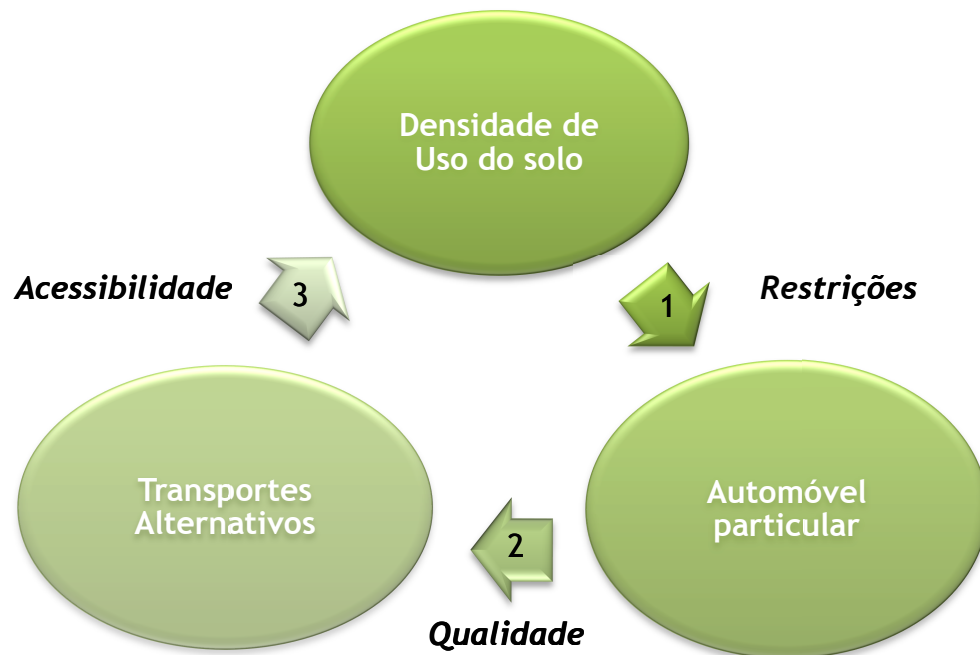


Figura II. 15 - Interações dos diferentes vetores de intervenção para uma Mobilidade Sustentável em meio urbano

Da experiência internacional (Vuchic,1999), a forma de desenvolvimento urbano que melhor qualidade de vida oferece aos seus cidadãos é a que permite/suporta grandes densidades de atividades e populacionais. Estas rentabilizam os grandes investimentos realizados nas diversas infraestruturas urbanas vitais, permitem uma maior variedade de atividades na vizinhança próxima da maioria dos seus cidadãos e rentabilizam o espaço/valor imobiliário dos terrenos dessas áreas. A concentração da população em áreas relativamente próximas permite que a maioria das atividades (residência, educação, saúde, emprego, lazer, comércio, etc.) possa estar ao alcance físico e económico da maioria da população. No entanto, a conveniência e a facilidade de uso do automóvel particular têm-no tornado um problema que necessita de ser resolvido. A mobilidade que permite aos seus utilizadores não é compensada pela poluição ambiental e sonora que provoca, e a sua grande necessidade de espaço para circular e estacionar, tornaram a sua utilização problemática, existindo a necessidade de restringir o seu livre uso. Por seu lado, estas restrições só serão aceites pela

população se modos de transporte alternativos lhe forem assegurados que lhe permitam uma mobilidade semelhante com uma qualidade equivalente. Nenhum cidadão aceitará restrições ao livre uso do seu veículo, se as alternativas que lhe ofereçam forem por exemplo autocarros velhos e sobrelotados, com horários desfasados das suas necessidades, com duração ou percursos mais extensos e os custos imediatos (“*out-of-pocket*”) semelhantes. As alternativas devem, portanto, ser implementadas tendo em consideração a qualidade da viagem por automóvel particular, tentando simulá-la, e permitir as mesmas ou melhores condições de acesso às diferentes atividades/zonas em toda a área urbana. Uma avaliação bem fundamentada de quais, entre as diversas metodologias disponíveis para fomentar estas alternativas, maior potencial apresentam para uma mudança de comportamentos, reveste-se de uma importância fulcral para aumentar a eficiência das decisões políticas e alcançar os comportamentos pretendidos.

Grandes investimentos para melhorar a qualidade dos serviços oferecidos aos cidadãos não se traduzem forçosamente numa melhoria da acessibilidade que oferecem pois se os cidadãos não conseguirem aceder facilmente aos seus locais de destino da forma desejada (custo, conforto e tempo despendido), as viagens por TC não terão a utilidade desejada e em geral os investimentos não terão o retorno financeiro e social esperado. Mas mesmo a melhoria da acessibilidade por modos alternativos, por si só não motiva a sua maior utilização em detrimento do automóvel particular. Exemplo claro desta realidade é a cidade de Lisboa onde se tem feito um grande investimento financeiro na melhoria das alternativas ao automóvel particular, nomeadamente através da expansão da rede do metropolitano, da modernização dos equipamentos circulantes nas linhas ferroviárias suburbanas, nomeadamente para Sintra, na travessia ferroviária da ponte 25 de Abril, ou na modernização dos equipamentos para o atravessamento fluvial do rio Tejo, aumentando significativamente as alternativas de mobilidade para a cidade. Este investimento, no entanto, não tem alcançado os resultados esperados de aumento do uso de transporte coletivos mas, pelo contrário, em alguns casos ainda se perdeu quotas de mercado para o automóvel particular. Existe, assim, a necessidade de, atendendo às densidades de uso do solo, à qualidade das alternativas e à acessibilidade que estas permitem às diferentes atividades essenciais para o dia-a-dia dos cidadãos, implementar restrições ao livre uso do automóvel particular que devem ser tanto mais severas quanto melhor for a qualidade da acessibilidade pelas alternativas.

5.1. A Qualidade dos Transportes Coletivos em Meio Urbano

Há apenas 20 anos atrás a qualidade não era um facto relevante para o sector dos TC. Somente no fim dos anos 80, princípio dos anos 90, é que se começou a dar maior relevância a este aspeto, existindo vários motivos, diferentes de país para país, que levaram ao seu desenvolvimento. A redução dos subsídios públicos aliada à necessidade de aumentar a eficiência produtiva, e a crescente introdução de competição através do pagamento em

função do serviço prestado (pelo menos em alguns países), estarão entre as causas mais relevantes (QUATTRO, 1998).

Em Portugal, no entanto, foi com o aumento dos problemas de redução da qualidade de vida em áreas urbanas devido ao crescente e excessivo uso do automóvel particular no acesso às mesmas, potenciada pela recente escalada dos preços dos combustíveis, que houve a tomada de consciência da necessidade de oferecer alternativas de mobilidade aos seus cidadãos.

De facto tem-se assistido na última década a grandes investimentos em sistemas de TC, principalmente nas grandes áreas metropolitanas, que no entanto não têm captado passageiros do modo esperado, tendo-se antes assistido a um acelerado processo de perda de capacidade de captação de passageiros por parte destes. A simples disponibilidade de alternativas, mesmo que credíveis, por si só não se tem mostrado aliciente à transferência modal por parte dos condutores atuais.

Assim as filosofias de intervenção mais recentes, tal como foi referido, sugerem a aplicação simultânea e coordenada de medidas que restrinjam, ou pelo menos tornem menos atrativo, o uso do automóvel particular quer através de portagens urbanas, quer de restrições de estacionamento, de restrições de acesso ou de discriminação positiva para os TC.

Estas medidas não são, no entanto, populares e a relutância em abandonar hábitos enraizados ou a revolta pela perda de direitos adquiridos só poderão ser ultrapassadas se essas medidas tiverem uma fundamentação justa e coerente, como seja a existência de alternativas viáveis e atrativas. Esta viabilidade e atratividade dependem, em grande parte, da sua qualidade e adequação às necessidades de mobilidade dos cidadãos, ou seja da sua *utilidade* para realizar as deslocações dentro da área urbana.

A *qualidade* é assim um dos pontos-chave na provisão de TC urbanos, que deverá receber mais atenção no futuro, tanto por parte das autoridades municipais como por parte dos operadores destes sistemas. Os TC são serviços que os cidadãos poderão decidir utilizar ou não, mas à semelhança de outros serviços e/ou indústrias terão que focar as atenções nos seus clientes, através de melhoramentos constantes tanto na sua organização como em estudos sobre a satisfação dos clientes.

Os TC são serviços complexos: oferecem serviços diários a todos os cidadãos que vivem em grandes áreas; estão organizados com uma série de infraestruturas de interligação a outros modos; e não se trata apenas de transportar pessoas e/ou bens, os sistemas de informação e de vendas fazem parte do serviço. Mais ainda, os cidadãos pretendem serviços livres de defeitos, enquanto esperam chegar ao seu destino na altura esperada e de uma forma organizada. Para cada viagem individual os cidadãos podem escolher entre diferentes modos

de transporte, cada um com as suas características específicas, vantagens e desvantagens, e com os seus custos.

Por outras palavras, os TC competem com outros modos de transporte, e só serão utilizados se conseguirem alcançar as expectativas dos cidadãos quando se deslocam e se se apresentarem como uma alternativa de mobilidade pelo menos tão boa como o automóvel particular. Isto é, se conseguirem prestar um serviço atrativo, acessível, fiável, que se adapte facilmente a alterações de mercado, de competidores ou do ambiente geral, tomem em consideração as orientações/opiniões dos clientes e respondam prontamente no sentido de minimizar eventuais roturas do serviço prestado, mas mesmo assim a um custo acessível para os clientes. Assim, mais uma vez, uma avaliação coerente e bem fundamentada da forma como estas expectativas podem ser alcançadas, nomeadamente através da análise dos fatores mais relevantes na sua apreensão, reveste-se de carácter fundamental para se puderem alcançar as alterações comportamentais desejadas.

O facto de a importância da qualidade do serviço ter sido reconhecida, não levou contudo a que o sector dos TC tivesse alcançado grandes avanços neste campo. Na maioria dos países não existe uma abordagem uniforme e o conceito de qualidade continua algo vago e teórico. Globalmente, pode dizer-se que os países mais a sul da Europa estão menos avançados que, por exemplo, a Escandinávia, a Alemanha, a Holanda ou a Grã-Bretanha, em conformidade com o seu desenvolvimento económico geral.

Outro problema a resolver é a dualidade entre qualidade do serviço (ou a qualidade que os cidadãos apreendem) e as condições de acesso permitida pelos serviços de TC e a facilidade de acesso aos mesmos. Um serviço de TC pode ser alvo de grandes investimentos em termos de qualidade e publicidade, mas se não atender às reais necessidades de mobilidade dos seus potenciais clientes, ou se não tiver a utilidade desejada, poderá não ter o retorno, económico e social, pretendido.

5.2. Avaliação da Acessibilidade por Transporte Coletivo em Meio Urbano

Como, em geral, a maioria dos cidadãos não realiza viagens por puro entretenimento mas para realizar determinadas atividades consideradas importantes, será o grau de sucesso com que alcança esse objetivo que definirá a utilidade dos serviços de transporte prestado. Será, portanto, a acessibilidade permitida pelo sistema de transporte a essas atividades vitais que determinará o seu maior ou menor uso por parte dos cidadãos. Esta acessibilidade é composta por duas áreas de intervenção: a acessibilidade ao sistema em si e a acessibilidade às diferentes áreas/atividades permitida pelo sistema.

A acessibilidade ao sistema tem fundamentalmente a ver com aspetos relacionados com a parte física de aceder e utilizar o sistema, desde distâncias aos locais de paragem, tipo de abrigos, entrada no veículo, circulação nas instalações e veículos, acesso a títulos de

embarque, entre outros. A acessibilidade permitida pelo sistema será a facilidade de alcançar quaisquer atividades desejadas num tempo razoável e a custos aceitáveis. No entanto, os sistemas de TC abrangem geralmente grandes áreas espaciais, pelo que não é fácil garantir os mesmos níveis de qualidade para toda a sua área de influência. Um sistema geralmente oferece serviços de qualidade superior somente para determinadas áreas (geralmente as que têm maior rentabilidade financeira) e serviços de qualidade inferior a outras, pelo que a acessibilidade oferecida será diferenciada de zona para zona. Assim, um dos problemas mais relevantes surge na definição de indicadores de acessibilidade em termos globais, e na sua determinação, nomeadamente nos seus principais caracterizadores (população potencial ou servida, qualidade do serviço, cobertura espacial, entre outros) e como inclui-los numa formulação matemática coerente. Na maioria das metodologias (CEP, 1976; Pons *et al.*, 1991; Foster *et al.*, 1993; Carson *et al.*, 1999; White *et al.*, 2000; Tyler, 2002) nota-se a tendência para associar a acessibilidade por transporte coletivo com o tempo de viagem neste modo (por vezes comparado com o tempo de viagem por TI) e com a população servida pela sua zona de inclusão. Ou seja, na maioria das metodologias estudadas, a acessibilidade depende não só da “qualidade” dos sistemas de transportes disponíveis, mas também da sua potencial disponibilidade pela população eventualmente interessada que se encontra disseminada pelo território.

Na avaliação da acessibilidade por TC, outro aspeto relevante tem a ver com medidas de discriminação positiva em relação ao automóvel particular. Estas medidas além de melhorarem a acessibilidade em si, permitem que comparativamente ao automóvel particular essa acessibilidade seja potenciada podendo, no caso de tornar a viagem significativamente mais curta (ou rápida) a preços semelhantes, torná-la numa alternativa bastante atrativa no processo de escolha do modo a utilizar. Existe, no entanto, ainda um relativo desconhecimento da real proporção dos efeitos deste tipo de medidas nas opções modais diárias. O seu estudo poderá, mais uma vez, melhorar substancialmente a sua eficiência isolada mas sobretudo a de medidas que também combinem restrições ao livre uso do automóvel.

5.3. Restrições à Livre Utilização do Automóvel Particular

Como foi referido, as estratégias que combinam pacotes de medidas que melhorem a acessibilidade por modos alternativos com restrições à livre utilização do automóvel particular, são as que maior potencial têm evidenciado para modificar hábitos e rotinas e levar a alterações modais que assegurem a sustentabilidade do desenvolvimento socioeconómico das áreas urbanas no futuro (Vuchic, 1999; COST 342, 2005).

Os principais fatores que fomentaram o aumento da utilização do automóvel particular, para além da melhoria das condições financeiras das famílias e do ambicionado estatuto social que a sua posse representa, foram os avultados investimentos que foram realizados nas últimas

décadas na melhoria da rede rodoviária conjugada com a falta de investimento e de incentivos à melhoria do serviço prestado pelos modos mais sustentáveis. A ilusão de que o uso do automóvel particular é relativamente mais económico (atendendo à conveniência, conforto e liberdade que permite) do que outros modos mais sustentáveis, provém da conjugação de diversas situações. Os avultados custos de implantação das estradas, dos parques de estacionamento e de todos os equipamentos e ações necessárias à sua utilização segura (manutenção do pavimento, das pinturas, dos sinais estáticos e luminosos, etc.) geralmente são pagos por toda a população, mesmo pela que não pode ou quer utilizar o automóvel particular. São diversos os custos das externalidades associadas à utilização excessiva do automóvel particular:

- Redução do espaço urbano disponível para outros usos (com maior retorno socioeconómico) devido ao espaço ocupado pelas vias e pelos estacionamentos;
- Assistência e reparação em caso de acidente;
- Redução de produtividade dos trabalhadores devido a atrasos na hora de chegada ou ao menor descanso motivado pela necessidade de grande antecedência na hora de saída de casa;
- Impacto visual e decréscimo no nível de mobilidade permitido aos peões, ciclistas e TC devido ao estacionamento ilegal e aos congestionamentos rodoviários;
- Redução da qualidade de vida dos cidadãos das áreas afetadas devido, por exemplo, à elevada poluição sonora e atmosférica que provocam;

Estes custos indiretos não são normalmente incluídos no custo efetivo das viagens por automóvel particular. Apenas os custos do combustível, de eventuais portagens e do estacionamento, são geralmente sentidos pelos utilizadores do TI em situações diárias pelo que são também os que estão geralmente presentes na decisão da escolha do modo de transporte.

Assim, parte das medidas para aumentar as restrições à livre utilização do automóvel particular nas áreas urbanas visa incluir na sua utilização estes custos atualmente pagos pela sociedade em geral, através do pagamento de portagens urbanas, do aumento do preço para estacionar (COST 342, 2005), através do aumento do custo do combustível ou pagamento de seguros segundo a quilometragem realizada. Outro grupo de medidas inclui restrições físicas ao livre acesso do automóvel particular, através do fecho de ruas ou proibição de certos movimentos, da limitação do número de lugares de estacionamento disponíveis ou a proibição de entrada de certos veículos em certas alturas do dia/semana (medida geralmente utilizada para diminuir a poluição atmosférica em alturas do ano/semana mais problemáticas).

6. Modelação da Mobilidade dos cidadãos

6.1. Introdução

Uma mobilidade urbana sustentável, como já foi referido, depende fundamentalmente do sucesso em transferir de forma racional algumas das viagens diárias dos cidadãos do automóvel particular para modos de transportes mais sustentáveis, como o modo pedonal, ciclista ou os TC. Essa transferência é fortemente condicionada pela existência de alternativas atrativas em termos de conforto, de custos, de conveniência, de disponibilidade, entre outras, verificando-se a necessidade de as conjugar com medidas que dificultem o livre uso do automóvel particular para que haja uma real alteração das opções modais. O nível de qualidade das alternativas e das restrições ao automóvel particular necessário para ultrapassar a inércia natural e os hábitos adquiridos, no entanto, ainda necessitam de melhor quantificação. Um dos objetivos deste estudo é avaliar a influência que a qualidade dos TC e das restrições ao uso do automóvel particular podem apresentar no processo de decisão da escolha do modo de transporte a utilizar no dia-a-dia. Essa influência é geralmente quantificada recorrendo a modelos matemáticos, teóricos e empíricos, cuja seleção e estruturação será apresentada nos pontos seguintes.

6.2. Planeamento e Modelos de Transportes

Desde os meados dos anos 70 que importantes desenvolvimentos técnicos se têm verificado na modelação de transportes (Ortúzar e Willumsen, 2001). No entanto, muitas destas inovações e aplicações têm tido pouca atenção fora do círculo académico. Após alguns anos de experimentação, é reconhecida atualmente a sua importância no apoio a um planeamento de transportes sustentável.

A importância da modelação de transportes surgiu da necessidade de tentar prever como os sistemas evoluíram, de modo a maximizar a eficiência dos avultados investimentos necessários para a implementação de grandes sistemas de transportes minimizando os seus efeitos colaterais (poluição, congestionamentos, acidentes, etc.). A impossibilidade de realizar experiências na implementação de uma ou outra intervenção no sistema de transportes sem graves implicações na comodidade da população envolvida, é outro facto que justifica o investimento na realização de modelos de previsão. Se no passado a modelação requeria grandes investimentos financeiros, humanos e muito tempo na realização de cálculo, o advento dos computadores pessoais e o aumento exponencial das suas capacidades que se verificou no final do século XX, facilitaram a sua implementação e aplicação a uma grande variedade de problemas.

Um *modelo* pode ser definido como “...uma representação simplificada de parte do mundo real, o sistema em estudo, que concentra certos elementos considerados importantes para a sua análise a partir de um determinado ponto de vista...” (Ortúzar e Willumsen, 2001).

A estruturação de um modelo que simule os sistemas de transporte apresenta-se complexa uma vez que a procura de transportes é altamente qualitativa e diferenciada. Existe uma variedade de necessidades de transporte específicas, diferenciadas pelo período do dia, dia da semana, propósito de viagem, tipo de carga, importância da velocidade e frequência, entre outras. Um sistema de transportes sem atributos que satisfaçam estas necessidades diferenciadas poderá ser completamente inútil. Estas características dificultam a análise e previsão da procura, uma vez que as unidades tradicionais de avaliação de desempenho, toneladas e pessoas-quilómetro transportadas, são demasiadas grosseiras podendo não revelar uma variedade imensa de requisitos e de serviços.

Outra característica importante na modelação de transportes é o facto da procura de transportes ser derivada, isto é, não é um fim em si mesma, uma vez que, excetuando as viagens de turismo, as pessoas se deslocam para satisfazer alguma necessidade (trabalho, compras, saúde, educação, etc.) no seu destino. Um bom sistema de transportes alarga as oportunidades de satisfazer estas necessidades, enquanto um sistema congestionado ou com poucas ligações restringe as opções e limita o desenvolvimento económico e social.

É a distribuição de atividades pelo espaço urbano que gera a procura de transportes, pelo que a abordagem tradicional subdivide a área em estudo em zonas designadas por zonas de tráfego interligadas por redes de transportes, por vezes simplificadas por linhas (arcos) e pontos (nós). O grau de simplificação depende da análise pretendida e das necessidades/capacidades dos programas informáticos a utilizar.

Outro aspeto a atender na modelação é a dinâmica existente entre a procura e a oferta de transportes. Em muitos casos, a oferta de transporte é suficiente para satisfazer a procura média mas insuficiente nos picos de procura, geralmente concentrados em curtos períodos ao longo do dia, especialmente nas áreas urbanas. Também esta especificidade da variabilidade no tempo levanta dificuldades na análise e previsão da procura no futuro.

Se considerarmos somente a oferta de transportes, também existem algumas especificidades que dificultam a sua análise, nomeadamente no que se refere aos seus intervenientes. Um sistema de transportes necessita de elementos fixos, a infraestrutura, e um número de elementos móveis, os veículos. É a sua combinação que, com um conjunto de regras para a sua utilização, permite o movimento de pessoas e bens. Na maioria das situações, a infraestrutura e os veículos não pertencem ou são geridos pela mesma entidade ou empresa (com exceção em algumas situações do modo ferroviário), gerando relações complexas entre autoridades governamentais (locais, regionais ou nacionais), empresas de construção, investidores, operadores de transportes, turistas e a população em geral.

Se existem situações em que o sistema pode ser implementado/intervencionado progressivamente no tempo respondendo a um aumento gradual da procura, outras há em que

a maior parte do investimento tem que ser realizado num curto prazo, acarretando grandes encargos e perturbando as atividades e a população na sua zona de implementação. Como investimentos no sistema de transporte são politicamente bem vistos pela população, mesmo que alienando grande número de residentes e de cidadãos que sofrem os congestionamentos das infraestruturas, é essencial que as decisões políticas sejam tomadas com base num planeamento cientificamente fundamentado, de modo a evitar decisões que apenas resolvam situações esporádicas e de curto prazo.

De um modo resumido, a figura II -2 apresenta as diferentes interações entre os fatores mais determinantes na definição da procura de transportes.

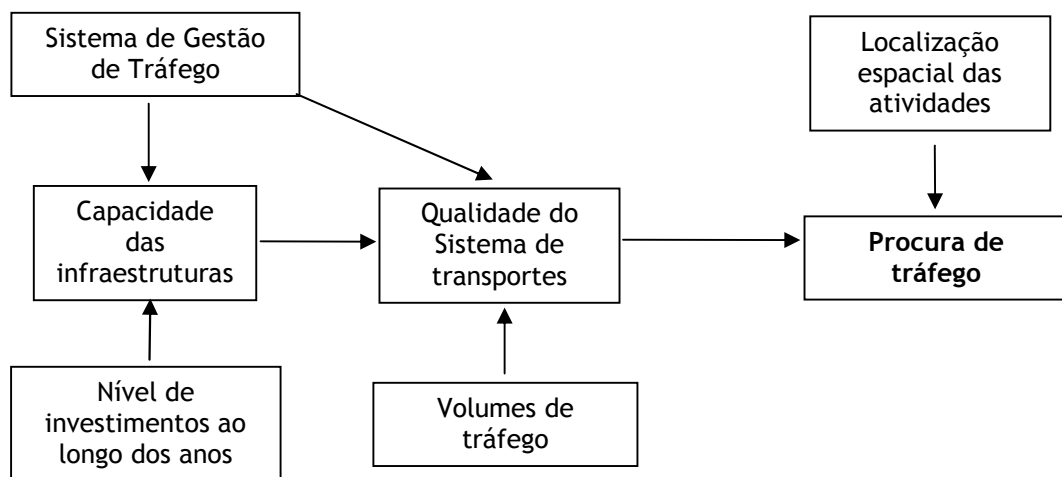


Figura II. 2 - Fatores que podem influenciar a procura de tráfego (adaptado de Ortúzar e Willumsen, 2001)

Para a maioria dos bens e serviços, é de esperar que o nível de procura de tráfego esteja dependente do nível de serviço (qualidade) proporcionado pelo sistema de transportes e pela distribuição espacial das atividades. No entanto, a qualidade do sistema de transportes pode ser aferida pelos volumes de tráfego que suporta atendendo à capacidade das infraestruturas existentes e do sistema de gestão adotado. O sistema de gestão pode igualmente pretender dar prioridade a certos modos de transporte quer por motivos de eficiência (TC, ciclistas), quer ambientais (veículos elétricos) ou de equidade social (pedestres), influenciando diretamente a qualidade do sistema, alterando ou não a capacidade das infraestruturas.

Deve atender-se, no entanto, que não é possível caracterizar todos os problemas de transportes de uma forma única e universal, uma vez que os mesmos são dependentes do contexto, bem como as possíveis soluções. Os modelos podem mesmo assim contribuir para uma identificação dos problemas e das suas possíveis soluções, de forma mais sustentada e coerente.

As abordagens metodológicas, utilizadas na modelação para o apoio a tomadas de decisão em problemas de transportes, devem atender a algumas características de modo a fornecerem resultados fiáveis e irem de encontro às expectativas dos decisores. Algumas dessas características são (Ortúzar e Willumsen, 2001):

- *O contexto da tomada de decisão*: Envolve a adoção de uma determinada perspetiva e a escolha da cobertura de um conjunto de interesses a atender. Na perspetiva deve definir-se o tipo de decisões a tomar, ou seja se são de nível estratégico, tático ou operacional. O âmbito de aplicação envolve o tipo de análise a realizar, se são intervenções pontuais para resolução de problemas específicos, ou análises mais gerais que incluam a interação com outros sistemas (por exemplo o uso dos solos), se além da procura a oferta também deve ser considerada, se o desempenho dos operadores também deve ser ponderado ou só a capacidade da infraestrutura, entre outros. Outra situação a considerar refere-se à opção entre uma única solução ótima ou à inclusão de vários cenários de modo a atender às necessidades de diferentes grupos de cidadãos. O contexto permitirá, portanto, definir as necessidades em termos de modelo a utilizar, variáveis a incluir, e pré-existências ou externalidades a considerar.
- *Precisão requerida*: É um facto que deriva do ponto anterior e que depende de outros fatores ligados a disponibilidade de dados e ao estado da arte na modelação. Geralmente trata-se de saber qual a precisão necessária (geralmente mínima) para se poder escolher entre uma opção boa e uma menos boa. Em situações óbvias, a precisão necessária é menor que em situações mais complexas, mas o bom senso tem sido a causa de várias más decisões de transportes no passado.
- *Disponibilidade de dados adequados*: A sua disponibilidade, estabilidade e dificuldade de previsão no futuro, influencia decisivamente o tipo de modelação a utilizar, podendo requerer elevados investimentos na aquisição que podem não ter o retorno esperado em termos de precisão dos resultados. Em muitos casos os dados pré-existentes são diminutos ou desadequados para o estudo em questão, a sua fiabilidade é duvidosa ou a instabilidade dos sistemas pode não permitir uma confiança nas suas previsões futuras. É um fator chave na decisão sobre o tipo de modelação a utilizar.
- *Estado da arte em termos de modelação*: Existe a necessidade de avaliar quais as técnicas mais recentes para determinados tipos de intervenção, atendendo à maneira como integram a riqueza/variabilidade de comportamentos, possibilidade de rastreio matemático e computacional e a disponibilidade de algoritmos que permitam chegar a boas soluções. Na prática, há que atender que todos os modelos assumem que certas variáveis são exógenas à modelação ou difíceis de incluir, assumindo-se ainda que muitas são irrelevantes para a análise, muito difíceis de prever no futuro ou que terão poucas alterações ou influência no sistema. Uma concreta definição das variáveis a utilizar ou a excluir pode auxiliar a decisão sobre a adequação do modelo para um

determinado problema. Hensher *et al.* (2005) sugerem que uma parte significativa do tempo despendido na modelação seja utilizada na avaliação das melhores variáveis e na sua especificação mais correta, antes de passar à estimação do modelo em si.

- *Recursos disponíveis*: Inclui os recursos financeiros, os dados, o material computacional, as aptidões técnicas, entre outros. Geralmente a duração do estudo e a comunicação com os decisores e o público têm um papel fundamental no sucesso das medidas sugeridas. Os decisores geralmente tendem a impor prazos temporais demasiado reduzidos para a coerente análise de projetos que demorarão anos a decidir, a implementar, e cujos resultados só a longo prazo serão realmente confirmados. Se, no entanto, existir um bom nível de comunicação com os decisores e o público, diminui-se a probabilidade de criar expectativas irrealistas e permite um melhor conhecimento das vantagens e limitações da modelação, evitando a aceitação incondicional dos seus resultados ou, no outro extremo, a sua completa rejeição.
- *Necessidade de processamento da informação*: Com o advento dos computadores pessoais e o aumento exponencial das suas capacidades, parte dos problemas de processamento de dados foi colmatada, restando apenas a habilidade humana nas tarefas de recolha, codificação, introdução de dados e a interpretação dos resultados. Esta habilidade será tanto mais importante quanto mais detalhada estiver a informação.
- *Nível de aptidão do analista*: A formação é geralmente tão dispendiosa que se torna mais viável apostar em modelos já existentes e compreensíveis do que adquirir e aprender a utilizar modelos ligeiramente mais avançados, podendo ser encarado como um entrave ao progresso e à inovação. No entanto, deve considerar-se a possibilidade de despendar algum tempo no estudo de novas técnicas, consolidando a experiência ganha com modelos mais antigos.

6.3. Princípios Gerais de Modelação

Existem duas abordagens clássicas para o desenvolvimento de teorias ou formulações que tentem o entendimento ou a replicação das complexas interações entre seres humanos, intervenientes inevitáveis dos sistemas de transporte: o estilo dedutivo, onde se constrói um modelo comparando os seus resultados com observações reais; e o estilo indutivo, onde a partir da observação dos dados se tenta inferir leis gerais. A abordagem dedutiva tem tido melhores resultados nas ciências puras, enquanto a indutiva se tem apresentado melhor em ciências sociais (Ortúzar e Willumsen, 2001). No entanto, verifica-se que em ambas a disponibilidade de dados é essencial, restringindo as opções em termos de relevância e complexidade permitidas. Assim, justifica-se que antes de qualquer análise se ponderem algumas especificidades no que concerne aos dados de base de modo a minimizar problemas posteriores.

6.3.1. Variáveis Endógenas e Exógenas

A problemática dos dados disponíveis tem grandes implicações ao nível do tipo de variáveis a utilizar e da teoria mais correta a empregar. Os modelos permitem prever um determinado número de variáveis dependentes a partir de outras variáveis independentes ou explicatórias, necessitando-se de dados de ambas para testar o modelo. Uma das suposições mais frequentes na modelação, mas também mais difícil de definir, é a consideração de que as variáveis independentes são exógenas à situação de escolha, ou seja, que se espera que as variáveis variem autonomamente ou independentemente de outras variáveis não observadas do modelo (isto é, as variáveis que entram na utilidade representativa são independentes da componente não observada da utilidade, ou seja, dos erros) (Train, 2003; Washington *et al.*, 2003). O valor das variáveis exógenas é determinado por fatores externos ao modelo, sendo exemplos destas o sexo ou o rendimento familiar. As variáveis endógenas, pelo contrário, têm uma variação que é causada por outras variáveis exógenas ou endógenas incluídas no modelo, podendo estas ser outras variáveis independentes ou dependentes, como por exemplo o custo da viagem ou o conforto. Uma forma expedita de avaliar a endogeneidade de uma variável independente é analisar se o seu valor é alterado com variações das variáveis dependentes. Como exemplo pode referir-se o caso das variáveis rendimento familiar e preço da viagem num modelo de escolha modal. Alterações na repartição modal do TC não implicam alterações no rendimento familiar de um agregado, mas pode implicar alterações do custo da viagem. Se se verificar um aumento da repartição modal a favor dos TC existe a possibilidade de uma otimização dos custos de operação do operador permitindo baixar o seu custo unitário. Pelo contrário uma maior cota por TI pode implicar mais congestionamentos e escassez de estacionamento que normalmente tornam a viagem mais cara. Assim, o rendimento familiar é efetivamente exógeno enquanto o custo da viagem tende a ser endógeno. Washington *et al.* (2003) refere que a suposição de que duas variáveis independentes são exógenas não implica que estas não possam ter covariância, assumindo-se, no entanto, que esta correlação não tem relação de causalidade. O estatuto de ser endógeno ou exógeno num determinado modelo depende da especificação do mesmo e das relações de causalidade assumidas entre variáveis independentes, podendo uma variável exógena tornar-se endógena pela inclusão de fatores adicionais ou de relações de causalidade no modelo. Um modelo que falhe em atender a estas especificações estima parâmetros inconsistentes. Train (2009), Louviere *et al.* (2005) e Guevara *et al.* (2009) sugerem algumas técnicas para tratar o problema de endogeneidade.

6.3.2. Modelação Agregada ou Desagregada

O nível de desagregação dos dados é outro facto fundamental para que uma modelação possa dar resultados credíveis. Embora um maior nível de detalhe geralmente determine uma maior precisão e qualidade das previsões, também aumenta os custos da sua recolha/aquisição e processamento, bem como de outros aspetos da modelação. Na modelação de transportes, a agregação, nomeadamente atendendo a dados exógenos que não dizem respeito ao

comportamento de quem viaja (endógenas), tem levantado algumas críticas devido a não representar as escolhas de cada indivíduo, mas a média das escolhas de um determinado segmento da população ou zona de tráfego. Quando o modelo pretende representar o comportamento de um determinado grupo (modelos agregados ou de primeira geração), tal como condutores de uma determinada zona de tráfego, a agregação de dados exógenos é inevitável. Mas quando o mesmo tenta replicar o comportamento de indivíduos (modelos desagregados ou de segunda geração), é conveniente que a informação exógena seja recolhida e tratada separadamente para cada viajante através de inquéritos individuais no domicílio. É necessário, portanto, escolher entre uma modelação mais desagregada e dispendiosa e uma mais agregada e menos rigorosa.

Das diferentes possíveis variáveis a utilizar, as que maior interesse têm para a modelação são as designadas *variáveis políticas*, isto é, aquelas que se assume que estão sob controlo dos decisores, ou seja, que o modelador pode variar de modo a avaliar as diferentes políticas ou estratégias. A modelação da procura futura é essencial para a maioria dos estudos de transporte, quer na previsão da procura em si, quer na previsão da sensibilidade do sistema a alterações de variáveis sob controlo do analista, necessitando muitas vezes de apresentar as previsões de forma agregada representando o comportamento de toda a população afetada. Cabe ao analista que utilizou uma modelação desagregada encontrar a melhor forma de agregar os resultados de forma a fornecer dados fidedignos. Salienta-se que a decisão de agregar informação mais pormenorizada, por exemplo segundo zonas de tráfego ou segmentos da população, leva a uma perda de informação que posteriormente poderá ser oneroso de corrigir, pelo que o nível mínimo de desagregação da informação deve ser bem avaliado antes do levantamento/recolha de dados.

Segundo Ortúzar e Willumsen (2001), a modelação agregada, ou de 1ª geração, teve uma utilização generalizada até ao final dos anos 70, uma vez que necessitavam de pouca especialização do analista e propunham uma metodologia que abarcava todo o processo de modelação, desde a recolha de dados até à previsão do nível de ligações numa rede. Foram fortemente criticados devido à sua falta de flexibilidade e precisão e aos seus elevados custos. Geralmente era baseada em informações zonais.

A modelação desagregada, ou de 2ª geração, tornou-se popular a partir dos anos 80, sendo encarados como uma revolução na modelação de transportes apresentando várias vantagens em relação à abordagem tradicional. A diferença fundamental entre as abordagens reside no tratamento da descrição de comportamentos durante o processo de desenvolvimento do modelo, exigindo os modelos desagregados um maior conhecimento em termos de estatística e econometria do analista, principalmente na interpretação dos resultados. Ortúzar e Willumsen (2001) referem que em vez de uma revolução radical, os modelos desagregados se apresentaram como uma evolução dos agregados, existindo casos em que a forma dos modelos de previsão são equivalentes, sendo difícil de avaliar qual das abordagens é mais

conveniente e em que situações. A modelação desagregada é geralmente baseada em dados individuais ou do agregado doméstico.

6.3.3. Recolha de Dados no Tempo

A maioria dos modelos (agregados e desagregados) de planeamento de transportes até ao final dos anos 80 dependia de dados dos padrões de viagem revelados por segmentos transversais da população (de modo a obter amostras heterogéneas e não enviesadas) num determinado instante no tempo. Assumia-se que os modelos permitiam avaliar as alterações no sistema simplesmente através da derivação da função de procura em relação às variáveis alteradas, ou seja, que os modelos possibilitavam avaliar uma relação estímulo-reação de forma realista através de observações num único momento no tempo. Esta assunção seria razoável de existisse um número suficiente de pessoas que alterassem o seu comportamento em ambas as direcções (utilizar mais e menos um modo de transporte, por exemplo), e se as suas opções não dependessem também dos hábitos já adquiridos ou não se alterassem com o tempo. As evidências apontam no sentido de pessoas de grupos com as mesmas características num determinado período, poderem apresentar comportamentos distintos consoante os seus antecedentes (Ortúzar e Willumsen, 2001).

Apesar de ser possível ajustar os modelos de modo a corroborarem os dados obtidos, por exemplo no ano base, este ajuste poderá incluir erros devido a uma incorreta especificação do problema, podendo apresentar graves problemas na previsão de evoluções futuras. Este facto tem levado à tentativa de incluir informação longitudinal, ou seja, avaliando a evolução no tempo, de forma a tornar os modelos mais fidedignos. Esta pode ser obtida, por exemplo, através da recolha da informação antes e depois da intervenção, podendo a modelação testar e eventualmente rejeitar determinados cenários atendendo à resposta do sistema. No entanto, problemas técnicos inerentes à sua implementação limitaram a sua inclusão na modelação pelo menos até finais dos anos 90 (Ortúzar e Willumsen, 2001).

6.3.4. Preferências Reveladas e Declaradas

Outro aspeto importante na modelação dos sistemas de transporte tem a ver com a possibilidade de se avaliar o comportamento do sistema sob diferentes condições. Como a implementação efetiva de diferentes cenários é inviável no planeamento de transportes (quer física e economicamente, quer ao nível do incómodo para os restantes sistemas existentes nas áreas urbanas), a observação de comportamentos restringe-se às situações/condições que podem ser controladas.

Até meados dos anos 80 era quase imperativo modelar com recurso a informação observada, isto é, usando dados de preferências reveladas. No entanto, os custos de recolha da informação e a dificuldade de especificar a relação causa-efeito de determinadas variáveis nas alterações do sistema, especialmente as qualitativas, aliadas à restrição da sua utilização

a situações já implementadas e às alterações de comportamento com o tempo (referidas anteriormente), limitavam a sua aplicação prática.

Assim, no final dos anos 70, surgiram as técnicas de preferências declaradas, provenientes do âmbito de estudos de mercado, na tentativa de colmatar os problemas citados anteriormente. Estas técnicas estimam a procura analisando as respostas dos utilizadores do sistema a cenários hipotéticos, que podem ter condições bastante distintas das existentes na altura. A sua aplicação, no entanto, também foi criticada pelo facto de possibilitar respostas oportunistas, quer por opção própria (a resposta pode revelar o próprio interesse e não o bem social), quer por falta de informação (por lapso da metodologia ou por falta de toda a informação relevante para uma resposta coerente). Estudos mais recentes, ver por exemplo Hensher *et al.* (2003a), apresentam metodologias combinando a apresentação de vários cenários comparativamente diferentes à realidade vivida com métodos estatísticos para determinar a veracidade das declarações tornando o seu uso mais fiável. Atendendo ao apresentado será a combinação de preferências declaradas com as reveladas normalmente a metodologia mais apropriada, nomeadamente para situações novas.

6.3.5. Modelação Contínua e Discreta

Nas especificações a atender para a formulação do modelo matemático a implementar de forma a replicar o sistema real, deve considerar-se como um aspeto bastante relevante, a sua forma funcional. Na sua definição é necessário uma análise “custo” *versus* benefício avaliando se um modelo linear, mais simples, é suficiente para a modelação que se pretende, ou se é necessário recorrer a modelos não lineares mais complexos, eventualmente mais precisos mas também mais exigentes em termos de recursos e técnicas de calibração. Mesmo modelos mais complexos podem ser pouco eficazes se os pressupostos na base da sua definição não forem convenientemente avaliados, como por exemplo a independência entre alternativas ou a necessidade de considerar opções anteriores. Muitas das teorias que suportam os modelos implementados dependem do respeito por esses mesmos pressupostos para a apresentação de resultados minimamente credíveis.

Outra especificação importante refere-se à definição do tipo de variáveis a utilizar, nomeadamente quais variáveis utilizar e sob que forma. Este aspeto é fulcral, não só para que o modelo possa replicar minimamente a realidade mas também porque condiciona o tipo de modelação a utilizar. Normalmente o tipo de variáveis pode dividir-se em variáveis contínuas ou variáveis discretas. As primeiras são variáveis numéricas com valores dentro de um intervalo finito ou infinito de valores reais. Como exemplo destas variáveis podem indicar-se o custo da viagem em euros, o tempo de viagem em minutos, ou os rendimentos familiares *per capita* em euros. As variáveis discretas são valores numéricos (as variáveis contínuas podem ser transformadas em discretas dividindo o seu intervalo em categorias) ou não, que pertencem a um número restrito de possíveis valores. Quando as mesmas não são variáveis

numéricas, podem ser definidas por categorias, nominais sem ordem definida como sim/não, tem/não tem, preto/branco, ou ordinais, com valores cuja ordem tem significado na modelação, como a avaliação da qualidade de um serviço numa escala de 1 (muito mau) a 5 (muito bom). Deve ter-se em consideração que apesar de ter uma ordem relevante, as variáveis ordinais podem não ter uma proporcionalidade entre os seus valores, isto é, o facto de um inquirido avaliar o nível de serviço com nota 5, não quer dizer que considere o serviço prestado cinco vezes melhor que o nível de serviço avaliado com 1. A definição do tipo de variável disponível e/ou adequada é fundamental uma vez que a modelação de variáveis numéricas pode ser realizada através, por exemplo, de regressões lineares múltiplas, mas a existência de pelo menos uma variável (endógena ou exógena) discreta obriga a utilizar outros modelos, por exemplo da família dos LOGIT ou PROBIT.

6.3.6. Amostragem

A abordagem à temática da amostragem deve iniciar-se pela definição da população em estudo, questão que em muitas situações pode ser mais complexa do que inicialmente se imaginaria. Além da população residente numa área metropolitana, o conjunto de potenciais viajantes inclui possíveis visitantes e indivíduos que passem na cidade, ou mesmo a população de outras áreas adjacentes que se saiba interagir com a área de estudo. Tradicionalmente a delimitação da área de estudo é realizada com algum grau de arbitrariedade, adotando para a sua compartimentação as fronteiras das divisões administrativas, como freguesias, concelhos e distritos, ou outras subdivisões utilizadas por entidades que realizam censos (por exemplo o Instituto Nacional de Estatística - INE). As populações podem ser tratadas como finitas ou infinitas, apesar de nenhuma ser de facto infinita. Este conceito de população infinita relaciona o tamanho da população com o da amostra, se a sua dimensão for suficientemente grande comparativamente à amostra pode dizer-se que é infinita. Como a maioria das populações são muito maiores que as amostras utilizadas, por economia de tempo e recursos (Ben-Akiva e Lerman, 1985), apenas se falará das análises mais simples com populações infinitas.

Outro conceito importante é a unidade de amostragem, que deve representar exhaustivamente a população em análise e ser mutuamente exclusiva. Geralmente em análises de viagens urbanas é o agregado familiar a unidade escolhida, mas poderá haver situações que necessitem de outra, como por exemplo na análise de viagens de transporte de mercadorias utilizando as empresas de envio e receção de mercadorias como unidade. Outras situações ainda poderão necessitar de unidades de amostragem ao nível do indivíduo ou mercadoria.

Coletivamente a listagem de todas as unidades de amostragem designa-se de estrutura (“*frame*”) de amostragem, definidas frequentemente recorrendo a listas telefónicas, a registos de propriedade de automóveis, entre outros. A criação de estruturas de amostragem convenientes pode ser por vezes bastante difícil, por exemplo em países menos desenvolvidos

sem um registo de moradas ou residentes, tornando-se necessário a visita a todas as residências da área de estudo para a definição de uma estrutura fiável (Ben-Akiva e Lerman, 1985).

Decidida a estrutura e a unidade de amostragem, existe a necessidade de estipular regras para a seleção das observações/inquéritos a realizar, definindo assim a estratégia de amostragem. A classe de estratégias mais comum é designada de amostragem probabilística, que segundo Ben-Akiva e Lerman (1985) tem as seguintes características:

- Deve permitir definir um conjunto de possíveis resultados (conjunto de observações da população resultante da amostragem) do processo de amostragem;
- Deve permitir relacionar uma probabilidade, ou densidade de probabilidade, a cada possível resultado da estratégia de amostragem utilizada;
- Deve seleccionar-se um dos possíveis resultados aleatoriamente (de modo a que a probabilidade de escolha de cada amostra seja igual) com a probabilidade associada definida no ponto anterior;
- Deve dispor-se de um estimador bem definido (como por exemplo a média, a mediana, a moda, etc.) que utilize a amostra e atinja uma estimativa única para cada possível resultado.

As estratégias de amostragem podem ainda ser realizadas com ou sem reposição. Sem reposição implica que uma unidade de amostragem depois de selecionada não pode ser selecionada outra vez, é retirada das possibilidades de escolha. Nas situações com reposição, a unidade de amostragem selecionada pode voltar a ser selecionada, que apenas têm influência com populações assumidas como finitas uma vez que nas infinitas se considera que a distribuição de características no conjunto das unidades de amostragem não é afetada pela remoção de um qualquer número finito de amostras.

6.3.7. Estratégias de Amostragem

Existem várias estratégias de amostragem bastante comuns para estimar características das populações como os totais, as médias ou a variâncias. Estas estratégias são conceptualmente semelhantes às utilizadas na análise de escolhas discretas, diferenciando-se pelo facto de nas últimas se pretender inferir o vetor de parâmetros das variáveis independentes do modelo de escolha (β) ou a distribuição das características da população, no caso concreto a sua probabilidade de escolha - $p(x)$. O facto de se tentar deduzir parâmetros de modelos de escolha não lineares complica significativamente o processo de amostragem.

As estratégias mais comuns segundo Ben-Akiva e Lerman (1985), Washington *et al.* (2003) e Ortúzar e Willumsen (2001), cuja descrição se apresenta no Anexo 1 do Capítulo II, são:

- Amostragem aleatória simples;
- Amostragem estratificada generalizada;
- Amostragem exógena;
- Amostras baseadas em escolhas;
- Amostragem enriquecida;
- Amostragem dupla.

6.3.8. Tamanho da Amostra

Quando a recolha de dados não pode ser realizada utilizando toda a população, como por exemplo numa área metropolitana, por motivos de logística e de custos, é normal recorrer à recolha de uma amostra da mesma. Mesmo que se pudesse inquirir toda uma população, por exemplo numa pequena vila num determinado dia, continuaria a ser uma amostra das viagens daquela semana, daquele mês e daquele ano. O desafio na definição de uma amostragem reside na técnica de recolha a utilizar e no tamanho da mesma, de modo a permitir obter resultados razoáveis, modelos fiáveis e não enviesados, economizando nos recursos despendidos. A dimensão da amostra é um facto de primordial importância pois define a validade estatística dos dados recolhidos. É atualmente aceite que uma amostra de tamanho ótimo que permita a sua utilização para os seus diferentes usos é praticamente impossível de definir devido à quantidade de usos que poderá ter e dos intervenientes que poderá servir. A forma mais comum de se recolher dados é por meio de entrevistas domiciliárias. São a metodologia mais dispendiosa e morosa mas permitem a obtenção de maior informação e mais fidedigna, se convenientemente aplicada. Tradicionalmente são baseadas em grandes amostras aleatórias o que aliado ao facto de serem realizadas pessoalmente ou por telefone, as tornam muito dispendiosas. Estatisticamente a seleção do tamanho da amostra numa população (in)finita depende de três fatores: o nível de precisão, o nível de confiança e grau de variabilidade. O nível de precisão, ou erro da amostra (E), fornece a gama de valores entre os quais se estima que esteja a real proporção da população. O nível de confiança ou de risco ($Z(\alpha/2)$), incorpora o teorema do limite central, ou seja, assume que quando uma população é repetidamente selecionada, os valores médios dos atributos avaliados nessas amostragens são iguais ao valor real da população. Assume, ainda, que os valores obtidos na amostragem seguem uma distribuição normal em volta do valor real, existindo amostras acima dos valores reais e amostras abaixo.

Com conhecimento destes fatores Ortúzar e Willumsen (2001), apresentam a expressão II-1 para o cálculo do tamanho da amostra:

$$n = \frac{CV^2 \times Z_a^2}{E^2}$$

Exp. II-1

Onde:

- n: tamanho da amostra;
- CV: coeficiente de variação, ou seja a razão entre o desvio padrão (σ) e a média (μ), sendo uma medida adimensional da dispersão;
- Z_{α} : Variação do desvio-padrão para um intervalo de confiança de α , também designado de valor da variável normal reduzida;
- E: Nível de precisão expresso como uma proporção;

A expressão anterior necessita, no entanto, de valores de base que podem à partida ser desconhecidos para o analista, como o coeficiente de variação, justificando por exemplo o estudo de uma amostra piloto.

O ITE (1999) e Guimarães *et al.* (1997) sugerem a expressão II -2 para o cálculo da amostra necessária para uma população infinita de um estudo binomial (por exemplo, utilizador de TC ou não), dependendo do nível de confiança, da proporção de atributo desejado, e do erro das estimativas:

$$n = \frac{Z^2 \times p \times q}{E^2} \quad \text{Exp. II-2}$$

onde:

- Z: Número associado ao nível de confiança desejado ou variação do desvio-padrão para um determinado intervalo de confiança (valor da variável normal reduzida);
- p: proporção de inquiridos que respondem afirmativo, ou seja, que verificam o atributo;
- q = 1 - p;
- E: erro específico na estimativa ou nível de precisão expresso como uma proporção;
- n: tamanho da amostra para uma população infinita.

Numa distribuição normal 95% dos valores da amostra estão entre 2 desvios padrão do valor da população real - $Z(\alpha/2 = 0,025) = 1,960$. O grau de variabilidade (p), refere-se à distribuição dos atributos na população. Quanto mais heterogénea for a população, maiores serão as amostras necessárias para obter um determinado nível de precisão. Como um valor de 50% implica um nível de variabilidade maior do que 20% ou 80% (que indica que a população ou não tem ou tem interesse no atributo), geralmente é o utilizado para caracterizar a máxima variabilidade e é utilizado para determinar a amostra mais segura. Assim é corrente a utilização de uma proporção $p = q = 0,5$. Nestas condições o ITE (1999) sugere as amostras a utilizar apresentadas no Quadro II. 1.

Para uma repartição modal geral semelhante à existente na área metropolitana do Porto, 22% de TC ($p=0,22$), os valores passariam a ser os do Quadro II. 2.

Os valores de referência para muitos estudos de transporte para o intervalo de confiança e para a precisão pretendida são, respetivamente, 95% e 5%, verificando-se que amostras de 250 a 350 inquéritos são geralmente suficientes para uma amostra estatisticamente relevante.

Quadro II. 1 - Tamanhos de amostras atendendo ao nível de confiança e margens de erro pretendidas para $p=q=0,5$ (ITE, 1999)

	Nível de confiança		
	90%	95%	99%
	Valor de Z		
Margem de erro	1,645	1,960	2,575
$\pm 1\%$	6765	9604	16590
$\pm 2\%$	1692	2401	4148
$\pm 3\%$	752	1068	1844
$\pm 4\%$	423	601	1037
$\pm 5\%$	271	384	664
$\pm 10\%$	68	96	166

Quadro II. 2 - Tamanhos de amostras atendendo ao nível de confiança e margens de erro pretendidas para $p=0,22$, $q=0,78$ (ITE, 1999)

	Nível de confiança		
	90%	95%	99%
	Valor de Z		
Margem de erro	1,645	1,960	2,575
$\pm 1\%$	4644	6592	11378
$\pm 2\%$	1161	1648	2845
$\pm 3\%$	516	732	1264
$\pm 4\%$	290	412	711
$\pm 5\%$	186	264	455
$\pm 10\%$	46	66	114

6.3.9. Segmentação da População

Na escolha da amostra há ainda que ter em atenção que existem grupos de cidadãos com condicionantes diferentes que podem alterar o resultado da análise. Assim, existe a necessidade de os diferenciar e realizar a sua análise em separado, através de uma segmentação em grupos homogêneos com características semelhantes. Exemplos desta segmentação poderão ser os cidadãos que já utilizam os TC (conhecimento de causa) e os que não utilizam (mas cuja opinião pode levar à adoção de medidas que levem à alteração modal), ou os utilizadores de TC que não possuem automóvel particular (utilizadores cativos que poderão alterar a sua escolha se as condições se alterarem e a utilidade do serviço oferecido não for suficiente para manter a sua fidelidade) relativamente aos que possuem

alternativa (geralmente clientes fieis cuja motivação para o uso do TC interessa conhecer). Outras segmentações usuais são o tamanho do agregado familiar, a idade, o sexo, o rendimento familiar, etc.

6.3.10. Correção e Extrapolação/Expansão dos dados

Após a definição do tamanho da amostra e da recolha dos dados existe a necessidade de corrigir os dados obtidos, no sentido de garantir a representatividade da amostra em relação ao tamanho dos agregados e à distribuição em relação a sexo e idade. Esta correção é tanto mais importante quanto o espaço temporal entre o inquérito e os censos da população. Ortúzar e Willumsen (2001) sugerem uma correção iterativa bi-proporcional onde após poucas iterações se obtêm valores que podem ser utilizados como dados expandidos/extrapolados para toda a população em relação aos valores obtidos na amostra. Este método, no entanto, só deverá ser utilizado quando o espaço temporal entre a amostragem e o censo populacional não for muito grande, devido às possíveis alterações da população especialmente em países em desenvolvimento.

Quando o espaço temporal entre o inquérito e o censo populacional é grande, a metodologia sugerida anteriormente não é apropriada existindo a necessidade da sua expansão/extrapolação. Esta expansão/extrapolação é realizada através do cálculo de fatores definidos para cada zona de estudo como a razão entre o número total de agregados familiares nessa zona pelo número de agregados obtidos na amostra corrigida. Este procedimento justifica-se pelo facto de a importância de cada resposta poder ser diferente dependendo do peso da amostra em relação à população avaliada. Uma resposta de uma amostra representativa de 20% da população de uma zona não deve ter a mesma importância que uma resposta de uma amostra representativa de 50% da população de outra zona. Ortúzar e Willumsen (2001), sugerem a expressão II -3 para o cálculo do facto de expansão de cada zona:

$$F_i = \frac{\text{Pop} - \frac{\text{Pop} \times (\text{Neleg} + \frac{\text{Neleg} \times \text{Nresp}}{\text{Amostra}})}{\text{Amostra}}}{\text{Amostra} - \text{Neleg} - \text{Nresp}} \quad \text{Exp. II-3}$$

Onde:

- F_i : Fator de expansão para a zona i ;
- Pop : Número total de moradas na população original;
- Neleg: Agregados da amostra não elegíveis (demolidas, não residenciais, etc.)
- Nresp: Agregados da amostra onde não houve resposta;
- Amostra: Número total de agregados na amostra original.

Se não houver agregados não elegíveis, a expressão II -3 passa a ser:

$$F_i = \frac{\text{Pop}}{\text{Amostra} - \text{Nresp}} \quad \text{Exp. II-4}$$

E se em todas os agregados da amostra selecionada se obtiverem respostas (amostragem perfeita), a fórmula considera que a percentagem de agregados não elegíveis da amostra é proporcional em toda a população e retira-a à última para não enviesar os resultados.

Outras correções sugeridas por Ortúzar e Willumsen (2001), referem-se a situações em que o inquérito está incompleto ou incorreto, que geralmente não devem acontecer em entrevistas domiciliares devido à presença de um inquiridor experiente. Existem, ainda assim, dados de difícil aferição, tais como o rendimento familiar ou o historial diário das deslocações e suas motivações, podendo levar a subestimações de certos fatores importantes na tomada de decisão. Quando existem inquéritos sem resposta deve verificar-se a sua causa, diferenciando situações entre a verdadeira recusa de responder e a perda de amostra quer por não existir quem responda (habitação desocupada), quer por ausência momentânea.

6.3.11. Calibração/Estimação e Validação de Modelos

O passo final na definição de um modelo matemático que tente imitar a realidade, é a sua calibração/estimação e a sua validação com dados reais. A calibração e estimação são conceitos semelhantes apesar de na prática a sua aplicação ser ligeiramente diferente. A calibração historicamente era associada a engenheiros responsáveis pela primeira geração de modelos, onde tentavam otimizar modelos que melhor se adaptavam a dados observados, escolhendo empiricamente parâmetros que assumiam como não nulos. Esta calibração era realizada sem grande base científica apesar de apresentar resultados consistentes. A estimação está associada a engenheiros e econometristas, que apesar de também ajustarem empiricamente certos aspetos específicos dos modelos, usam uma base científica para encontrar os valores de parâmetros que melhor ajustam o modelo aos dados observados. A estimação permite igualmente aferir os parâmetros estatisticamente significativos, permitindo a eventual exclusão dos restantes. O termo estimação provém do facto de se utilizar os dados de amostras para estimar os valores de toda a população. O facto da estimação se basear em propriedades estatísticas bem definidas permite adotar intervalos de confiança em relação aos valores estimados e aos resultados previstos.

A maioria dos modelos de transporte tem sido construída baseada em dados de segmentos transversais da população, levando à tendência de assumir que a sua validação se resumia exclusivamente em avaliar a precisão dos seus resultados confirmados com os valores comportamentais observados no ano de base. Apesar de ser uma condição necessária, não é suficiente, existindo a necessidade de avaliar as previsões do modelo com estudos antes e depois da intervenção. A validação necessita assim de comparações entre as previsões dos

modelos e dados não utilizados na estimação dos seus parâmetros (Ortúzar e Willumsen, 2001). Esta validação condiciona as previsões obtidas através dos modelos, devido às variáveis possíveis de utilizar e à precisão possível de obter atendendo às simplificações, erros e incertezas. As previsões estão condicionadas a estimar as variáveis dependentes, dado um conjunto de variáveis independentes, e são normalmente utilizadas para testar vários cenários alternativos para um conjunto de possíveis valores das restantes variáveis. Podem ser utilizadas múltiplas vezes para explorar vários vetores de atuação, sendo de importância crucial que a sua especificação permita retrocessos fáceis no tempo, situação complicada em modelos de transporte globais que envolvem processo complexos de equilíbrio da oferta e da procura.

6.4. Estrutura Clássica dos Modelos de Transporte

Apesar das inovações nas técnicas de modelação dos últimos 40 anos, a estrutura clássica dos modelos de transporte tem-se mantido praticamente inalterada. A figura II.3 apresenta a sua forma geral, tradicionalmente designada de *Modelo dos Quatro Passos*.

A abordagem considera um zonamento e a rede de infraestruturas, a recolha e codificação de dados de planeamento, de calibração e de validação do modelo, incluindo dados demográficos no ano base de cada zona da área de estudo e os níveis de atividade económica, incluindo empregos, áreas comerciais, equipamentos de educação e de lazer, entre outros. Com estes dados estima-se um *modelo de geração de viagens*, ou seja, o número total de viagens geradas (saídas) e atraídas (entradas) para cada zona de tráfego da área de estudo.

No passo seguinte pretende-se atribuir estas viagens a determinados destinos, ou seja, fazer a *distribuição de viagens* pelas diferentes zonas produzindo uma matriz de distribuição de viagens.

Após esta fase pretende-se atribuir as viagens da matriz anterior aos diferentes modos de transporte possíveis de utilizar, resultando na modelação da *repartição modal* (tipicamente entre TI ou privado e TC ou público).

O passo final requer a *atribuição das viagens* por cada modo de transporte às diferentes infraestruturas existentes na rede de transportes entre as origens e os destinos da área de estudo.

O modelo clássico é portanto apresentado como uma sequência iterativa de quatro submodelos, apesar de se reconhecer que as decisões sobre a viagem podem não se efetuar segundo esta sequência.

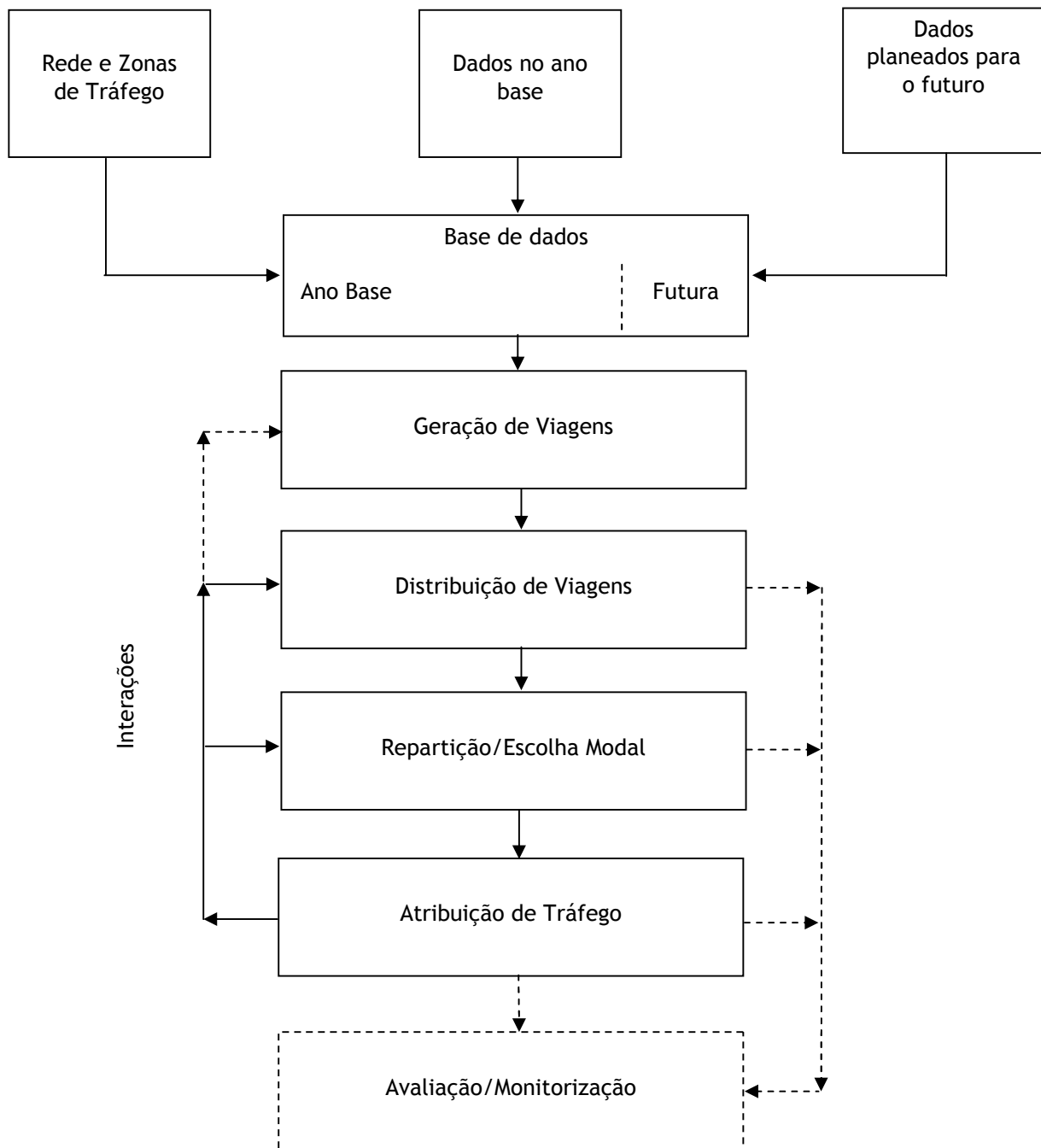


Figura II. 1 - Estrutura clássica do modelo de transporte de quatro passos (retirado de Ortúzar e Willumsen, 2001)

Uma visão mais contemporânea considera que as posições dos diferentes submodelos dependem das suas utilidades durante as opções de viagem. Perante, por exemplo, um congestionamento de tráfego, o condutor pode assumir várias simples alterações que podem ser de difícil quantificação nos modelos tradicionais (Ortúzar e Willumsen, 2001):

- O *trajeto* escolhido para evitar uma zona congestionada ou para aproveitar uma nova ligação, incluindo a escolha de um melhor local de estacionamento ou a possibilidade de combinação com viagens em TC;
- O *modo de transporte* utilizado para chegar ao destino, ou seja, o condutor pode simplesmente optar por mudar de modo de transporte se verificar que o seu uso é mais económico em termos de duração de viagem ou financeiramente;
- A *hora da viagem* que pode ser antecipada/retardada de modo a evitar as horas de ponta do tráfego;
- A alteração do *local de destino* para zonas menos congestionadas;
- A *frequência* da viagem alterando o dia para a realizar ou aproveitando para combinar com a necessidade de realizar outras atividades.

Outra dificuldade desta modelação advém da existência de transformações mais complexas a longo prazo, como as alterações no emprego, na residência, na atratividade das áreas comerciais, entre outros, que são condicionadas, pelo menos parcialmente, pela acessibilidade providenciada pelos próprios sistemas de transporte. Verifica-se igualmente na modelação tradicional da geração de viagens que esta é independente do nível de serviço prestado pelo sistema de transportes, assunção provavelmente irrealista, que só recentemente assistiu ao surgimento de técnicas capazes de a atenuar (Ortúzar e Willumsen, 2001).

Existem abordagens mais recentes que tentam integrar dois ou mais passos numa única modelação, por exemplo juntando a frequência de viagens, o seu destino e o modo de transporte utilizado num único modelo. Existem ainda abordagens que salientam a importância das atividades familiares e das viagens que envolvem, modelando as mesmas e às suas condicionantes atendendo por exemplo às atividades de curta duração, ao seu encadeamento ou às limitações em termos orçamentais e compatibilização de horários. Estas modelações impõem maiores dificuldades na sua decomposição nos quatro passos da metodologia clássica e apenas a partir de meados da primeira década do séc. XXI, é que deixaram de ser uma área de investigação para começarem a ter aplicação prática. No entanto, parece claro que estes modelos baseados nas atividades quotidianas ("*Activity-based models*"), poderão melhorar de forma significativa a modelação tradicional no futuro, principalmente devido a melhor compreensão que permitem do comportamento em termos de viagens (Vovsha e Freedman, 2008a).

Apesar destas novas abordagens, a larga experiência na aplicação do modelo clássico dos quatro passos, com todas as variantes metodológicas e aperfeiçoamentos implementados (principalmente na sua fundamentação científica) ao longo das últimas décadas, tornam-no

uma referência, que poderá mesmo servir de termo de comparação com as metodologias mais recentes e com perspectivas de viabilidade de implementação.

6.5. Modelos de Escolha/Repartição Modal

A escolha do modo de transporte é provavelmente um dos submodelos mais importantes na metodologia clássica para o planeamento de transportes. Apesar do domínio avassalador do TI em relação ao coletivo, principalmente nas áreas urbanas, o último desempenha um papel fundamental na definição das políticas de transporte devido à sua eficiência em termos de usos do espaço viário disponível e de recursos consumidos por passageiro transportado (desde que haja procura suficiente). Modos subterrâneos e outros ferroviários não partilham o espaço comum (exceto pontualmente nas intersecções) com o TI pelo que também não contribuem para o aumento do congestionamento rodoviário. É fisicamente inviável a disponibilização de espaço nas áreas urbanas para atender às necessidades de vias e de lugares de estacionamento para que todos os cidadãos que o desejem, possam usar o seu próprio TI. Existe assim, a necessidade de um planeamento do sistema de transporte que consiga atrair voluntariamente (ou quase) parte considerável das viagens diárias para o TC, permitindo uma melhor mobilidade para os utilizadores dos últimos mas também para os restantes que continuem a optar pelo TI. A escolha do modo de transporte afeta a eficiência geral das viagens urbanas, a quantidade de espaço urbano alocada às funções de transporte e a variedade de alternativas à disposição dos viajantes.

6.5.1. Evolução Histórica dos Modelos de Escolha/Repartição Modal

6.5.1.1. Modelos Agregados

No passado até à década de 50, principalmente nos Estados Unidos da América, era assumido que as características pessoais de quem realizava as viagens eram os fatores mais determinantes na escolha do modo de transporte, tentando-se realizar modelos de escolha modal imediatamente depois da geração de viagens. Esta metodologia, designada de modelos de escolha modal de fim de viagem (Ortúzar e Willumsen, 2001), permitia incluir as características individuais na estimação da repartição modal. No entanto, o destino da viagem, as suas características e do modo utilizado eram omissos. Esta assunção era consistente com a visão na época de que quanto maior fosse o rendimento maior seria a propensão dos cidadãos para viajar. A modulação na altura visava antever o aumento da procura de modo a planear os investimentos a realizar. Em termos de modelação da repartição modal, esta incidia sobretudo sobre fatores como o rendimento, a densidade habitacional ou a posse de automóveis, existindo casos em que a disponibilidade de TC era considerada através de índices de acessibilidade (provavelmente ao sistema de TC) (Ortúzar e Willumsen, 2001). Estes modelos apesar de obterem resultados aceitáveis em condições de serviços de TC semelhantes em toda a área de estudo e com poucos congestionamentos de tráfego, não permitiam a inclusão de decisões políticas para aumentar a mobilidade, tais

como: melhorias nos TC, restrições ao estacionamento, portagens, entre outros; ou seja, assumiam que estas políticas não influenciavam as escolhas modais.

Na Europa a modelação das escolhas modais foi historicamente dominada por modelos que aplicavam este submodelo após a distribuição de viagens (Ortúzar e Willumsen, 2001). Permitiam incluir não só as características das viagens mas também dos modos de transporte, mas dificultavam a inclusão de informação sobre as características individuais dos indivíduos. Geralmente realizavam a análise baseados em um ou dois fatores associados à viagem, que se consideravam os mais relevantes na escolha modal (tradicionalmente o custo ou a duração da viagem). Estes dados eram usados para calibrar empiricamente curvas em forma de “S”, designadas de “*diversion curves*”, em que a proporção de viagens num determinado modo atendendo a todas a viagens realizadas, estavam relacionadas com a diferença de custos ou tempo de viagem entre modos. O seu uso era limitado pelo facto de só poderem ser aplicados a matrizes de viagens com o modo realmente disponível, isto é, a matrizes de viagens de indivíduos com automóvel disponível, podendo no entanto também ser aplicado a escolhas entre diferentes modos de transporte coletivo. Outro problema devia-se à sua formulação agregada e fraca base teórica, que dificultavam a inclusão das características e restrições dos diferentes modos na modelação de comportamentos de agregados individuais, ignorando várias variáveis políticas como as portagens ou o custo do estacionamento, facto que limitava fortemente a credibilidade das suas previsões (Ortúzar e Willumsen, 2001).

A evolução do tipo de modelação anterior foram os modelos *sintéticos*. Assentes numa base teórica de maximização da entropia, os primeiros modelos sintéticos tentaram gerar modelos de distribuição e escolha modal simultâneos. A sua solução levou à seguinte formulação, onde a expressão II-5 é geralmente conhecida como o Modelo Gravitacional de Distribuição de viagens, e a expressão II-6 é designado por função de repartição modal LOGIT binomial (no exemplo, para dois modos):

$$T_{ij}^k = A_i O_i B_j D_j e^{(-BC_{ij}^k)} \quad \text{Exp. II-5}$$

$$P_{ij}^1 = \frac{T_{ij}^1}{T_{ij}} = \frac{e^{(-BC_{ij}^1)}}{e^{(-BC_{ij}^1)} + e^{(-BC_{ij}^2)}} \quad \text{Exp. II-6}$$

Onde:

- T_{ij}^k : número de viagens com origem na zona i e destino na zona j , através do modo k ;
- O_i : número de viagens com origem na zona i ;
- D_j : número de viagens com destino na zona j ;

- A_i, B_j : fatores de ajustamento calculados para as viagens com origens em i e destino em j ;
- β : parâmetro a estimar de modo a ajustar o modelo às especificidades de cada estudo;
- C_{ij}^k : função do custo generalizado da viagem entre i e j através do modo k . Geralmente é constituída por várias variáveis relacionadas com as características das viagens, como o custo, o tempo de viagem, o tempo a pé, o tempo de espera, o tempo de transbordo, etc. As suas importâncias relativas eram convertidas para uma mesma unidade (geralmente dinheiro ou tempo, designado de *custo do tempo*) adicionadas a uma variável/constante que engloba todos os restantes aspetos não quantificados;
- P_{ij}^1 : Proporção de viagens entre i e j realizadas através do modo 1.

A expressão II-6 caracteriza-se por gerar uma curva sigmoide em forma de S , similar às curvas da metodologia anterior, verificando-se que para custos equivalentes $C_1 = C_2$ uma proporção equitativa $P_1 = P_2 = 0,5$. À medida que o custo da alternativa 2 aumenta, a proporção de viagens realizadas pelo modo 1 aumenta, até ao ponto em que o custo de viajar na alternativa 2 é tão elevado que todas as viagens se realizam pelo modo 1 ($P_1 = 1$). Uma extensão do modelo a uma situação multinomial é facilmente realizada através da soma dos custos generalizados da viagem por cada um dos restantes modos em denominador. Nesta formulação, β tem uma função dupla de controlar a dispersão na escolha modal e na escolha entre destinos a diferentes distâncias da origem, tarefa demasiado complexa para um único parâmetro. Esta constatação levou vários estudos até final dos anos 70 a recorrer a outra parâmetro designado de *custo composto* (K_{ij}^n), entre a origem i e o destino j apreendido pelo indivíduo n . Este custo composto podia ser especificado de diferentes formas (Ortúzar e Willumsen, 2001).

Segundo Ortúzar e Willumsen (2001) estes modelos foram o estado da arte da modelação agregada da distribuição de viagens e de repartição modal, nomeadamente em áreas urbanas, até à década de 50. No entanto, foram substituídos por modelos desagregados que reagem melhor a fatores essenciais na escolha modal e utilizam de forma mais eficiente o esforço despendido na aquisição/recolha de dados.

6.5.1.2. Modelos desagregados

Os modelos desagregados que surgiram nas décadas de 60 e 70, designados de modelos de repartição multinomiais, envolvem a escolha de mais de dois modos de transporte. Das estruturas existentes a mais simples é a de n modos, caso geral da estrutura binomial, tendo sido a mais popular apesar de considerar todas as alternativas com importância igual. A sua contestação surgiu nos anos 70 através de um exemplo típico de quase todos os livros de modelação de transportes, o célebre problema da linha azul e da linha vermelha. Neste problema demonstra-se que a inclusão de uma nova alternativa correlacionada com uma já

existente (neste caso utilizou-se um exemplo exagerado de uma nova linha de TC apenas com uma cor diferente mas com a mesma qualidade do serviço), segundo o modelo diminuiria o uso das restantes proporcionalmente ao número de alternativas, nomeadamente do TI. No entanto, a prática demonstra que esta situação não se verifica e que a repartição modal do TI se se alterar, será sempre muito inferior às estimações do modelo.

A estrutura com a adição de modos, teve algumas aplicações durante os anos 60 e 70, mas foi possível verificar que as previsões do modelo dependiam de qual alternativa era adicionada e da estabilidade no tempo dos seus desempenhos utilizados nas previsões iniciais.

Outra possibilidade era a estrutura hierárquica (ou “*nested*”) onde as opções que eram semelhantes ou correlacionadas eram agrupadas numa primeira repartição, por exemplo de TC, separando-as dos modos não correlacionados, para posteriormente se realizar uma segunda repartição entre modos semelhantes. Esta estrutura era de uso generalizada na década de 60 e no princípio dos anos 70, mas tinha como defeito usar para função de custos compostos o valor mínimo do preço entre o autocarro e o comboio para cada par de zonas e uma atribuição *tudo-ou-nada* através de custos mínimo para definir a segunda repartição (Ortúzar e Willumsen, 2001).

A formulação geralmente utilizada para estimar as repartições modais nestas estruturas era o modelo LOGIT.

6.5.1.3. Modelos de Procura Direta

Uma alternativa à sequência metodológica tradicional de estimar submodelos relativamente bem conhecidos é o desenvolvimento direto de um modelo que junte a geração, distribuição e repartição modal numa única fórmula. A sua mais-valia, em relação à abordagem sequencial, é evitar a acumulação dos erros de cada passo, uma vez que tenta estimar simultaneamente os três submodelos. Os modelos de procura direta, tentam estimar uma equação que relacione a procura de viagens diretamente como as características do modo utilizado, da viagem realizada e do próprio indivíduo. As primeiras abordagens, no final dos anos 60, utilizavam uma função multiplicativa de caracterizadores socioeconómicos, das atividades e dos níveis de serviço dos modos que serviam determinado par de zonas origem/destino. Exemplos destes caracterizadores são a população, o rendimento, os tempos e os custos da viagem por modo, entre outros. Variações deste tipo de modelos sugeriam variáveis relativas, dividindo os seus valores pelos melhores encontrados entre pares de zonas. O volume de informação necessário para estimar este tipo de modelos, a sua base teórica insuficiente e o facto de utilizarem valores médios para determinados parâmetros limitaram a sua aplicação apenas a contextos interurbanos com zonas de grande dimensão (Ortúzar e Willumsen, 2001).

6.5.1.4. Modelos de Escolha Discreta

Uma limitação das metodologias apresentadas anteriormente surgia de apenas utilizarem características mensuráveis e contínuas como o tempo, o custo, o rendimento, a população, entre outros. A utilização de variáveis ordinais ou nominais era difícil de realizar e a teoria subjacente desadequada para a sua modelação, nomeadamente o seu ajuste através do método dos mínimos quadrados. A única possibilidade de modelar este tipo de variáveis através de métodos tradicionais requeria grupos de indivíduos com comportamento homogéneo ou o registo das suas escolhas em diversas ocasiões, de modo a obter frequências de observação possíveis de expressar em valores entre 0 e 1. Assim nos finais da década de 70 surgiram os modelos de escolha discreta, nomeadamente através dos trabalhos de Daniel MacFadden (Train, 2003), que através de uma base teórica sólida se apresentaram como ferramentas potentes para a modelação das escolhas de indivíduos em várias áreas de investigação, incluindo os transportes.

A designação de escolhas discretas deriva do facto de poder modelar escolhas com base em variáveis contínuas mas também discretas, isto é, seleccionadas de um conjunto finito de opções, tais como ter ou não carta de condução, realizar transbordo ou não, ter carro disponível ou não.

Os modelos de escolha discreta assumem que (Ortúzar e Willumsen, 2001):

“...A probabilidade de um indivíduo escolher uma determinada opção é uma função das suas características socioeconómicas e da atratividade relativa dessa opção...”

Para representar esta atratividade de uma opção é usual recorrer ao conceito de *Utilidade*, definida como o que um indivíduo procura maximizar. As diferentes alternativas propostas por si só não produzem utilidade, mas esta pode ser aferida pelas suas características e as do indivíduo, através, por exemplo, de uma *utilidade observável* definida como uma combinação linear de várias variáveis. Na bibliografia são vários os autores (Ben-Akiva e Lerman, 1985; Louviere e Street, 2000a; Louviere *et al.*, 2000b; Ortúzar e Willumsen, 2001; Arsénio *et al.*, 2002; Washington *et al.*, 2003; Fricker *et al.*, 2004) que sugerem o uso de funções de utilidade para os modelos de escolha modal. A maioria sugere expressões do tipo:

$$U_{jq} = \sum_k V_{jkq} + \epsilon_{jq} \quad \text{Exp. II-7}$$

ou

$$U_{jq} = \sum_k (\beta_{kj} \times X_{jkq}) + \epsilon_{jq} \quad \text{Exp. II-8}$$

onde:

- U_{jq} : Utilidade do modo j avaliada pelo individuo q ;
- V_{jkq} : parte mensurável, observável, representativa ou sistêmica da utilidade do modo j pelo individuo q ;
- X_{jkq} : Atributo mensurável k na quantificação da utilidade do modo j pelo individuo q ;
- β_{kj} : Peso do atributo k , constante para todos os indivíduos q , mas que varia segundo o modo j ;
- ε_{jq} : parte aleatória, que reflete variáveis intrínsecas do individuo q em relação ao modo j , bem como eventuais erros de observação e de modelação, ou seja, é a diferença entre a utilidade real de um modo j para o individuo q e a utilidade que é possível de determinar pelo analista.

O ITE (1999) ainda discrimina a função utilidade em:

$$U_{jq} = \sum_k (a_{kj} \times LOS_{jkq}) + \sum_k (b_{jkq} \times SEC_{jkq}) + \varepsilon_{jq} \quad \text{Exp. II-9}$$

onde:

- U_{jq} : Utilidade do modo j avaliada pelo grupo socioeconómico q ;
- a_{kj} : Coeficientes de cada atributo k de qualificação do nível de serviço prestado pelo modo j , avaliado pelo grupo q ;
- LOS_{jkq} : Variável k que descreve o nível de serviço prestado pelo modo j avaliado pelo grupo q ;
- b_{jkq} : Coeficiente de cada atributo k descriptor das características socioeconómicas do grupo q em relação ao modo j ;
- SEC_{jkq} : Característica socioeconómica k do grupo q em relação ao modo j ;
- ε_{jq} : constante específica para o modo j que representa variações inexplicáveis no uso do mesmo pelo grupo q .

No caso de modelos de repartição/escolha modal estes devem ser sensíveis aos atributos que podem influenciar as escolhas individuais do modo de transporte a utilizar. Ortúzar e Willumsen (2001) sugerem os seguintes 3 grupos de fatores como os mais preponderantes na escolha do modo de transporte:

- Características do viajante/agregado:
 - Disponibilidade e posse de TI;
 - Habilitações para a condução;

- Estrutura familiar (solteiros, casados, casados com filhos), e a situação profissional (ativo, desempregado ou reformado);
- Rendimento familiar
- Necessidades excepcionais, tais como levar os filhos à escola ou trabalhar com veículo próprio;
- Densidade residencial, ou seja, o número de agregados domésticos por unidade de espaço.
- Características da viagem:
 - O motivo da viagem, nomeadamente no que concerne à necessidade de programação e de conhecimento das condições do sistema de transportes no local de destino;
 - Período do dia para realizar a viagem, que pode dificultar a utilização dos TC;
- Características do sistema de transporte:

Quantitativas:

- Tempo de viagem relativo: em circulação, em espera e os tempos de percursos a pé para cada modo;
- Custos financeiros relativos: tarifas, combustível e outros custos diretos como portagens;
- Disponibilidade e custo de estacionamento;

Qualitativos:

- Conforto e conveniência;
- Fiabilidade e regularidade;
- Proteção e segurança.

Desta lista, que poderia ser mais extensa, é possível verificar que nem todos os fatores são fáceis de quantificar, e que muitos são difíceis de traduzir por valores contínuos, necessitando-se de modelos que consigam analisar variáveis discretas.

Após a definição das variáveis a utilizar na modelação, o valor da sua utilidade deve ser comparado com as restantes opções e transformada numa probabilidade entre 0 (não ser escolhida) e 1 (100% de probabilidade de ser escolhida). Para esta transformação é usual utilizar formulações caracterizadas por graficamente serem representadas por uma curva em S. As mais comuns são a formulação da família dos LOGIT (função sigmoide logística) e PROBIT (função de distribuição acumulativa inversa).

Os modelos de escolha discreta têm como base teórica a Teoria da Utilidade Aleatória, e basicamente assume que (Ortúzar et. al., 2001; Ben-Akiva e Lerman, 1985; e Hensher *et al.*, 2005):

- Os indivíduos que pertencem a uma população homogênea agem racionalmente e possuem total informação, isto é, escolhem sempre a opção que *maximiza* a sua utilidade pessoal atendendo às restrições legais, sociais, física e orçamentais (temporais ou financeiras).
- Os indivíduos com determinadas características são confrontados com um conjunto pré-definido de alternativas disponíveis e que não existem outras restrições que afetam o seu processo de decisão.
- Cada alternativa tem associada uma utilidade bruta para um determinado indivíduo, e o analista não possui a informação completa sobre os elementos considerados para a escolha de uma alternativa. Assim, o analista assume que a utilidade é representada por dois componentes: uma mensurável, sistêmica ou representativa, função de atributos mensuráveis (os X_{jkq} na expressão II-8); outra parte aleatória que reflete a idiosincrasia e gosto pessoal de cada indivíduo, agregado a eventuais erros de observação ou medição (o ε_{jq} na expressão II-8).

Este ponto permite duas aparentes irracionalidades: que dois indivíduos com as mesmas características e perante as mesmas alternativas possam ter opções diferentes; e que o mesmo indivíduo possa não escolher a melhor opção, segundo as características consideradas pelo analista. Para que esta formulação esteja correta necessita-se de uma certa homogeneidade na população em estudo, isto é, que todos tenham as mesmas alternativas disponíveis e sejam confrontados com as mesmas restrições.

- Dependendo da distribuição probabilística dos erros ou resíduos ε_{jq} , diferentes modelos podem ser gerados.

Atendendo a este último ponto, a distribuição dos erros, existem várias possibilidades que condicionam os modelos a utilizar. Aliás, segundo Train (2003), o erro ε não é definido para uma determinada escolha em si, mas sim relativamente a uma representação que o analista faz de uma determinada situação de escolha. Como o analista não conhece estes erros, trata-os como valores aleatórios e define uma função de distribuição de probabilidades para realizar aferições estatísticas sobre as escolhas dos indivíduos.

Considera-se ainda que o processo de escolha da alternativa i pela população pode ser representado por um modelo com atributos x , cuja distribuição conjunta pode ser representada por $f(i, x)$. Na modelação de escolhas individuais é comum dividir esta distribuição conjunta no produto entre uma distribuição condicional e uma marginal (Ben-Akiva e Lerman, 1985):

$$f(i, x) = P(i | x) \cdot p(x)$$

Exp. II-10

Onde:

- $f(i, x)$: distribuição conjunta das escolhas i e dos atributos x ;
- $P(i | x)$: distribuição condicional;
- $p(x)$: distribuição marginal, que não representa qualquer teoria comportamental, mas que representa a probabilidade de encontrar uma observação com as características de x , a sua cota de mercado.

O que torna esta decomposição interessante e útil é o facto de $P(i | x)$ não ser simplesmente a probabilidade condicional mas representar o resultado de um modelo que se acredita descrever as escolhas individuais (i) mediante a utilização de uma função com parâmetros desconhecidos (β). A forma como cada observação é realizada condiciona assim a distribuição conjunta dos (i) e dos x (e portanto dos (β)), pelo que o processo de amostragem é de importância vital (Ortúzar e Willumsen, 2001).

Outro aspeto importante de modo a permitir a utilização de modelos de escolha discreta, é a definição das alternativas que devem apresentar três características (Train, 2003):

- As alternativas devem ser mutuamente exclusivas na perspetiva do decisor: a escolha de uma alternativa implica a exclusão das restantes alternativas, ou seja, o decisor apenas escolhe uma alternativa do conjunto de opções;
- O conjunto de alternativas tem que ser exaustivo, isto é, deve incluir todas as alternativas possíveis;
- O conjunto de alternativas deve ser finito.

As duas primeiras características não são restritivas, uma vez que uma definição apropriada das alternativas pode assegurar na maioria das situações que as mesmas sejam mutuamente exclusivas e exaustivas. Mesmo em situações de viagens multimodais utilizando dois modos complementares como o automóvel particular e o autocarro, uma opção adicional pode efetivamente ser “carro+bus”, sendo o conjunto expandido mutuamente exclusivo. Na eventualidade de uma opção ser “não escolher nenhum modo”, esta deve ser inserida no conjunto de alternativas de modo a que o mesmo seja efetivamente exaustivo. Uma abordagem alternativa, menos trabalhosa mas também menos precisa, é limitar a análise apenas às situações de que se dispõem dados excluindo os dados de situações com combinação de alternativas ou que não escolham nenhuma, permitindo que conjunto menor de alternativas resultante continue a ser mutuamente exclusivo e também exaustivo.

A terceira característica, no entanto, é restritiva uma vez que caracteriza este tipo de modelos e os distingue do campo de aplicação dos modelos de regressão, que utilizam variáveis dependentes contínuas que podem ter um valor infinito de resultados. Um conjunto infinito de alternativas inviabiliza a utilização de modelos de escolha discreta. Existe sempre a possibilidade de, como já foi referido, reduzir o conjunto infinito de alternativas a um conjunto finito transformando-o em intervalos de possíveis valores.

6.6. Modelos Baseados em Atividades

A maioria da investigação mais recente em termos de modelação de transportes tem incidido numa alternativa à abordagem clássica dos quatro passos. São designados de *modelos baseados em atividades* em que a procura de viagens é calculada a partir de um ponto de vista mais geral e holístico (que concebe a realidade como um todo) das atividades de um indivíduo (Davidson *et al.*, 2007). Apesar de na prática ainda ser muito comum o recurso à modelação agregada no processo de planeamento para a previsão de necessidades a atender pelos sistemas de transportes, nos últimos anos tem-se assistido ao crescente interesse por modelos mais realistas em termos comportamentais, baseados nas atividades. A relativa complexidade dos novos modelos (real ou apreendida) e consequente tempo de processamento efetivo e de implementação dos algoritmos tem levado a algumas críticas, fundamentalmente no que se refere à relação investimento necessário ao seu desenvolvimento em larga escala *versus* as contrapartidas de precisão de previsões comparativamente a modelos tradicionais.

A modelação baseada em atividades tem semelhanças com a modelação mais tradicional no que concerne às etapas da escolha do modo de transporte e do destino, e à afetação a infraestruturas rodoviárias ou a TC. No entanto, incluem ainda outros aspetos atualmente relevantes com a altura do dia escolhida para realizar a viagem, a frequência da escolha da deslocação, escolhas de estacionamento, escolhas de trabalhar em casa ou não, etc. (Davidson, 2008). As aplicações práticas já implementadas têm demonstrado a sua viabilidade apresentando modelos práticos (em termos de tempo e grau de pormenor necessários) que incorporam as partes mais relevantes das investigações sobre atividades e comportamento em viagens, pelo menos a nível regional (Davidson *et al.*, 2007). A base dos modelos de transporte convencionais amplamente divulgada e utilizada define a unidade básica de viagem como a deslocação que uma pessoa faz de uma origem para um destino com um determinado propósito por meio de um determinado modo de transporte (a pé, de carro particular, de autocarro, de comboio, etc.). Este conceito de deslocação, no entanto, é limitado quando as viagens são interdependentes e as decisões tomadas para um modo afetam os restantes. Um exemplo deste problema verifica-se quando o automóvel particular é preterido numa viagem casa-trabalho, não estando disponível na viagem de regresso, situação que na modelação tradicional pode ser difícil de aferir.

Modelos mais recentes evoluíram no sentido de ponderar para unidade básica de viagem o conceito de *jornada* ou cadeia de deslocações, em que a escolha do modo é feita atendendo a toda a jornada diária para que o modo utilizado no retorno esteja condicionado ao modo utilizado na viagem de ida. Este conceito de jornada evoluiu ainda para utilizar como unidade básica de viagem a *atividade*, entendida como o que se quer fazer com o tempo disponível, quer em casa quer no exterior, como trabalho, compras, educação, lazer, etc. Cada indivíduo ajusta a sua escolha de atividade atendendo ao que quer fazer com o seu tempo e à disponibilidade de locais para as realizar, isto é, tal como já foi referido, as pessoas não querem viajar (na maioria das situações) mas realizar as suas atividades e as viagens são empreendidas entre os locais dessas atividades. A sequência de atividades a empreender forma a unidade a modelar e a sequência de deslocações é o modo de a alcançar (Davidson, 2008).

Na modelação de transportes os princípios conceptuais subjacentes incluem a escolha entre alternativas e a tomada de decisão sobre qual escolher. Assume-se que os indivíduos podem:

- Escolher o destino para onde ir;
- Que modo de transporte utilizar;
- Que trajeto percorrer;
- Com que frequência realizar a atividade pretendida;
- Onde estacionar;
- Em que altura do dia realizar a deslocação;
- Tomar a decisão de antecipar ou retardar a deslocação para evitar situações problemáticas, por exemplo, congestionamentos rodoviários.

O modelo pretende prever as decisões dos indivíduos em cada uma das situações anteriores tendo por base o sistema de transportes, tentando prever a sua disposição para utilizarem o sistema com eventuais alterações que se pretendam introduzir.

Assim, em termos gerais podem distinguir-se três aspetos comuns a todas as abordagens de modelos baseados em atividades (Vovsha e Freedman, 2008a; Davidson *et al.*, 2007):

- Estrutura baseada em jornadas ("*tour*"), isto é, numa sucessão de deslocações que começam e acabam num local base (casa ou emprego) utilizadas como a principal unidade para a modelação de viagens. Esta estrutura preserva a consistência entre deslocações da mesma jornada relativamente a questões de viagem como o destino, o modo de transporte ou o período do dia. Mesmo as deslocações não baseadas em casa podem ser relacionadas com deslocações a partir de casa.

- Uma plataforma baseada em atividades que implica que a modelação das viagens seja realizada a partir de atividades diárias do agregado familiar ou de um indivíduo. Esta plataforma permite a consistência da dimensão tipológica, espacial e temporal dos padrões de atividades individuais, a substituição das atividades domésticas e fora de casa, a duração das atividades em coerência com os tempos de chegada e de partida, interações nos agregados familiares e outros aspetos considerados pertinentes para a análise das atividades.
- Uma técnica de modelação por micro-simulação, aplicada a nível totalmente desagregado de indivíduos ou agregados familiares, que converte as escolhas relacionadas com as atividades e viagens resultantes de modelos de probabilidades fracionadas (“*fractional-probability models*”) numa série de decisões encadeadas (“*crisp decisions*”) entre as escolhas discretas. Este método de implementação do modelo origina resultados mais realistas com a aparência de dados de estudos de viagens/atividades.

O uso combinado destes requisitos tem permitido que este tipo de modelação incorpore várias técnicas avançadas de estudos comportamentais não utilizadas pela modelação tradicional dos quatro passos.

6.6.1. Definição de Jornada

As técnicas convencionais de modelação da procura de viagens utilizam como unidade de viagem a deslocação, definida como a viagem entre dois locais - a origem e o destino. Os modelos baseados em viagens têm algumas limitações devido a consideração de que cada deslocação é independente de todas as outras, agravadas pelo facto de a abordagem com zonamentos agregados não permitir uma fácil análise do encadeamento das viagens (Vovsha e Freedman, 2008a).

A definição de uma jornada como um *encadeamento de viagens fechado*, isto é, em que a origem da primeira deslocação é o destino da última deslocação, pode ser encarada como um passo importante no desenvolvimento de novos modelos operacionais. Esta definição permite que a jornada tenha apenas um ponto fulcral, a origem, com um ou mais potenciais destinos. Na maioria das jornadas, os pontos de origem são a residência ou o local de trabalho e o facto de ser considerada fechada permite que tenha propriedades para a tornar numa unidade de modelação extremamente conveniente:

- A escolha do modo de transporte é geralmente dependente de toda a jornada a realizar, mesmo que nas deslocações intermédias haja alguma liberdade;
- A opção entre TI ou TC é quase exclusivamente dependente de toda a jornada a realizar, uma vez que a opção do TI particular tem que ser tomada na residência;

- O número e local de destino das viagens não baseadas em casa são função das viagens baseadas em casa antecedentes e precedentes;
- E a hora de partida das sucessivas deslocações de uma jornada está limitada à hora de chegada da deslocação anterior.

A maioria dos modelos operacionais baseados em jornadas distingue entre os destinos o mais importante como o *destino primário*, tratando os restantes como paragens secundárias a caminho ou no regresso do primário. Este conceito é fácil de perceber em viagens pendulares ou para a escola, mas em viagens para compras ou de lazer, o destino primário pode ser difícil de definir pois as mesmas podem ser de multi-propósito/destino. Se vários destinos tiverem o mesmo propósito ou duração, a distinção do destino primário pode ser realizada através da sua distância à origem, considerando o mais afastado. A metade da jornada até ao destino primário é designada de viagem direta ou saída e a metade de volta à origem de viagem de retorno ou entrada.

A hierarquia das unidades de viagem, jornada - meia jornada - deslocação, apresenta as seguintes vantagens em relação aos modelos tradicionais baseados em viagens:

- *Total consistência nas escolhas do modo de transporte utilizado, do destino, e do período do dia para realizar a viagem, para todas as viagens da mesma jornada:* Este tipo de modelação assegura que na escolha modal todas as opções disponíveis para cada deslocação estejam condicionadas ao modo escolhido atendendo a toda a jornada, particularmente no que se refere a disponibilidade de automóvel particular.

Assegura igualmente um padrão lógico dos locais de paragem intermédia, uma vez que a modelação tradicional tratava de uma percentagem elevada de viagens não baseadas em casa difíceis de modelar devido às fracas relações entre as suas origens ou destinos com variáveis de uso do solo e à dificuldade em segmentar as mesmas atendendo a fatores socioeconómicos (Vosvha *et al.*, 2008b). A modelação tradicional também estima a repartição modal destas viagens não baseadas em casa geralmente apenas atendendo ao nível de serviço da rede de transportes, considerando que são independentes das restantes deslocações da jornada. A modelação com base em atividades permite uma estimativa mais consistente das viagens não baseadas em casa considerando os desvios mais prováveis no trajeto entre a origem e o destino primário.

Em relação ao período temporal das viagens, tal como já foi referido, a modelação com base em atividades permite um correto encadeamento das viagens intermédias atendendo às suas horas de chegada e de partida, ou seja, consideram a duração da atividade.

- *A possibilidade de modelar o planeamento das atividades diárias:* A modelação com base em atividades permite a modelação consistente do número, tipo e horário das

jornadas e deslocações num dia ou em vários dias, considerando as restrições temporais, como a existência de apenas 24h, ou o encadeamento lógico de determinadas atividades de acordo com o seu propósito. Permite ainda determinar as decisões e condicionantes consideradas por cada indivíduo ou pelo agregado familiar para cada jornada.

Outro aspeto que diferencia a modelação com base em atividades da tradicional é a definição não de um modelo de escolha modal mas de dois com níveis hierárquicos distintos. Haverá um modelo de nível superior que definirá o modo de transporte ao nível da jornada e um inferior (condicionado ao superior) que definirá o modo de transporte ao nível da deslocação. Um exemplo deste tipo de situação será a escolha do automóvel particular para realizar a jornada casa-trabalho-casa, e a opção de ao almoço se deslocar a pé.

Nem todos os modelos baseados em jornadas são automaticamente modelos baseados em atividades, apesar de o inverso na prática de verificar. Os modelos baseados em atividades incorporam episódios de atividades e a sua duração, de forma a assegurar um comportamento realista. Em termos de estrutura do modelo, os padrões individuais de viagem gerados serão assim devidamente restringidos e representam uma sequência viável de atividades, deslocações e localizações.

Outro aspeto relevante é a sua capacidade para modelar as interações dentro dos agregados familiares e as atividades e viagens conjuntas. Uma parte significativa das deslocações e atividades são decididas de forma conjunta por vários membros do agregado familiar, no sentido de sincronizar no tempo e no espaço as necessidades dos diferentes membros, afetando de forma significativa os padrões e horários diários, geralmente como condicionantes adicionais.

6.6.2. Técnicas de Microsimulação

As técnicas de Microsimulação utilizam uma síntese da população e para cada componente do modelo selecionam escolhas discretas de determinadas distribuições de probabilidades, aplicando-se ao nível completamente desagregado de pessoas e agregados familiares. O nível de desagregação tem como finalidade não avaliar exaustivamente cada comportamento individual, mas sim permitir melhores previsões estatísticas de viagens agregadas. As principais vantagens da Microsimulação em relação à abordagem tradicional de probabilidades fraccionadas podem ser resumidas a (Vovsha e Freedman, 2008a):

- Permitem economizar no tempo de armazenamento e de manipulação dos dados durante o processamento do modelo. A Microsimulação necessita de menos cálculos que um modelo de viagens agregado devido à eliminação de frações das deslocações, geralmente as que não são escolhidas ou têm pouca relevância. Um modelo tradicional opera com uma matriz origem-destino completa dando grande relevância à segmentação geográfica em detrimento da segmentação de viagens. Os modelos baseados em

atividades permitem um equilíbrio entre estas duas vertentes da modelação. Vovsha *et al.* (2008b) refere como exemplo um modelo elaborado para a região metropolitana de Nova Iorque, com 20 milhões de habitantes e 23 milhões de jornadas diárias e onde uma modelação tradicional de quatro passos teria uma matriz de 4 000 origens x 4 000 destinos x 10 modos motorizados x 100 segmentos da população = 16 000 000 000 para cada um dos 6 propósitos de viagens com muitas células preenchidas com frações extremamente pequenas do total de viagens na área metropolitana. A impraticabilidade da análise computacional deste volume de informação foi contornada através de modelos baseados em atividades considerando apenas 8 milhões de jornadas como as mais significativas;

- Permitem uma modelação mais realista das cadeias de tomadas de decisão, atendendo a variáveis de situação e a restrições temporais e espaciais, nomeadamente o efeito de escolhas anteriores;
- A aplicação de modelos de Microsimulação permite que os parâmetros do modelo para cada indivíduo sejam estocásticos/aleatórios, podendo a heterogeneidade não observada do comportamento de viagem ter qualquer distribuição pretendida (tipo Logit misto), sem significativas implicações no tempo e recursos de processamento.

Vovsha *et al.* (2008b) apresenta a figura II-4 de forma a exemplificar a estrutura do modelo baseado em atividades desenvolvido para a área metropolitana de Nova Iorque.

6.7. Diferenças e Similaridades entre os Modelos Baseados em Atividades e Modelos de Quatro Passos

Comparativamente, os modelos baseados em atividades cobrem todos os aspetos que os modelos tradicionais de quatro passos consideram, acrescentando ainda vários detalhes e benefícios ao procedimento de modelação. Os modelos baseados em atividades substituem os tradicionais propósitos de viagem por modelos de padrões diários de atividades e viagens que integram inter-relações, compensações e efeitos de substituição entre as diferentes jornadas e atividades (Vovsha e Freedman, 2008a).

Em relação aos modelos de geração de viagens tradicionais, os modelos de padrões de atividades diárias são sensíveis ao ambiente e à acessibilidade assegurando a inclusão da procura de viagens induzida. Os modelos baseados em atividades modelam as frequências das jornadas/deslocações por propósito de uma forma integrada, em vez de utilizar um modelo para cada propósito. Permitem ainda diferenciar os destinos como primários, mais atrativos, ou secundários, associados a condições de acessibilidade favoráveis, permitindo uma análise integrada das viagens não baseadas em casa atendendo à origem da jornada e ao seu destino primário.

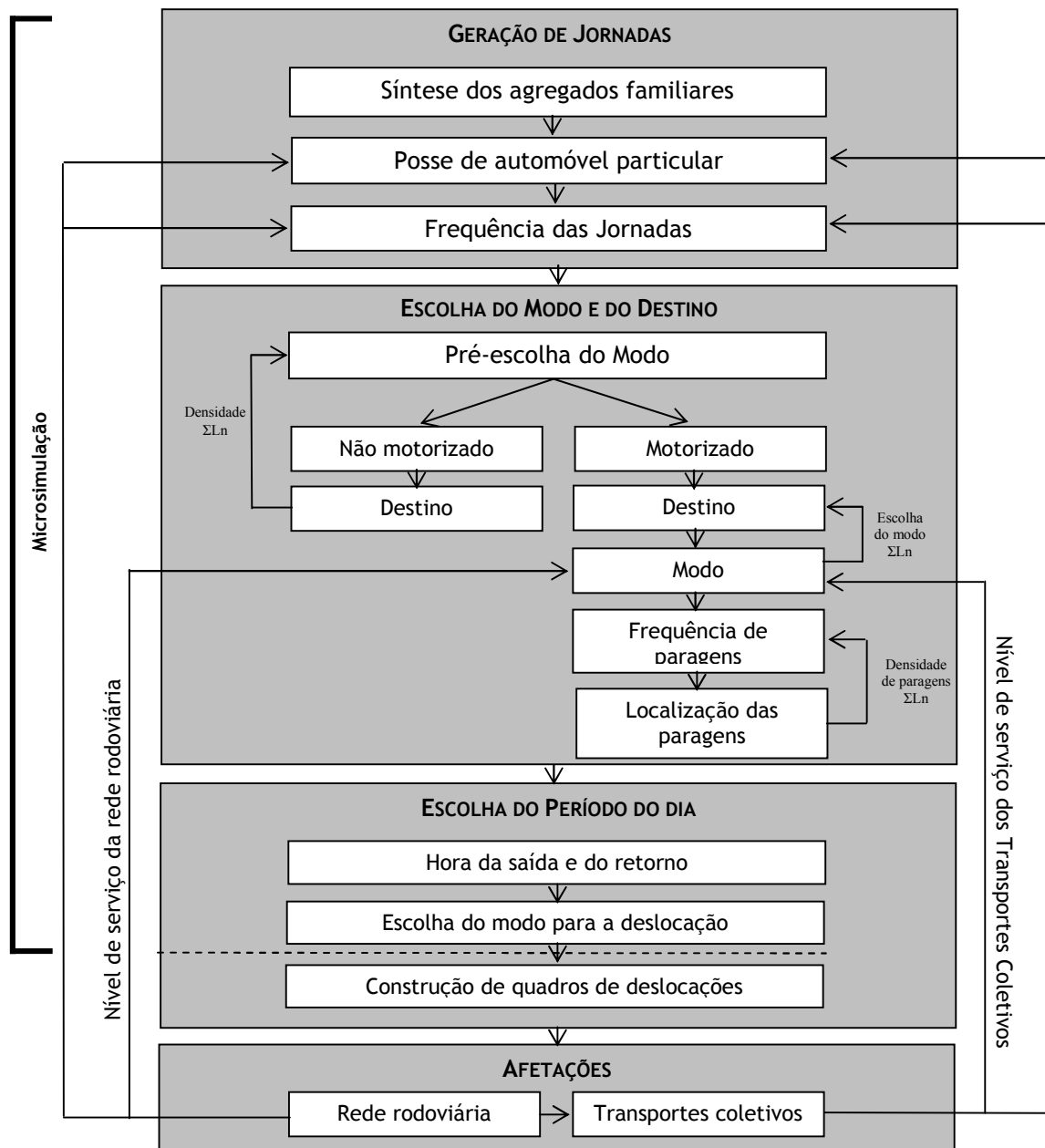


Figura II. 2 - Exemplo de uma estrutura de um modelo baseado em atividades (adaptado de Vovsha *et al.*, 2008b)

Na modelação tradicional a cada viagem é atribuída uma escolha modal levando a combinações pouco lógicas quando analisadas ao nível de jornada diária. A modelação baseada em atividades divide estas escolhas ao nível da jornada, mais importante, e ao nível da deslocação, permitindo uma correta modelação da decisão relacionada como o modo de transporte logo a partir do momento em que o indivíduo decide sair de casa.

Outra vantagem da modelação baseada em atividades é a total coerência das escolhas do período do dia para realizar as diferentes jornadas e das diferentes deslocações de cada

jornada, permitindo ao modelo refletir o impacto de qualquer alteração numa determinada altura do dia nas restantes deslocações ou viagens.

A modelação com base em atividades apresenta, no entanto, presentemente como grande desvantagem, a quantidade de dados e a forma de recolha. Para alcançar o pormenor pretendido, os dados têm que ser recolhidos numa espécie de diário de viagens, em que todas as deslocações têm de ser anotados com algum pormenor. Estudos com base em levantamentos de campo tradicionais não podem ser realizados através de modelos baseados em atividades, ou seja, atualmente há necessidade de recolher dados específicos para a situação em estudo. Outro aspeto a ter em atenção tem a ver com a disponibilidade de indivíduos para realizar estes levantamentos exaustivos de dados referidos, existindo a possibilidade de erros ou informação incompleta por falta de vontade em fornecer determinados pormenores das suas atividades diárias.

7. Síntese

Pelo que foi exposto, constata-se que a descoordenação entre políticas de ordenamento do território e as políticas de transporte/mobilidade geraram parte dos problemas existentes na maioria das áreas urbanas de média a grande dimensão, principalmente dos países mais desenvolvidos. Se estas consequências não forem consideradas também nas políticas dos países em desenvolvimento, parece inevitável o seu agravamento a médio ou longo prazo. Assim, as autoridades nacionais, regionais e locais, têm de encetar esforços no sentido de coordenarem as diferentes políticas de desenvolvimento, não só horizontalmente, entre os diferentes sistemas coexistentes nas áreas urbanas, como verticalmente, que permitam a integração das políticas/sistemas locais, nas políticas/sistemas regionais e nacionais. Esta coordenação permitirá não só uma melhor eficiência na aplicação dos escassos recursos humanos e financeiros, como uma melhor eficácia na implementação de um efetivo desenvolvimento sustentável.

No que diz respeito às políticas de mobilidade, parece claro que o rumo a seguir passa por políticas que incentivem o uso dos sistemas de transportes mais sustentáveis, em detrimento do automóvel particular. Estas políticas, no entanto, devem ser cuidadosamente estudadas e integradas de modo a não diminuir a mobilidade de certos grupos de cidadãos ou a partir de certas zonas urbanas, pela simples imposição de restrições sem a oferta de alternativas viáveis que permitam aos seus cidadãos ter acesso a bens e serviços considerados básicos para a sua subsistência.

A tendência atual deverá passar por aumentar a densidade de uso do solo nas áreas centrais, propícias a serem servidos por TC, limitando o alastramento dos subúrbios que implicam um maior uso do automóvel particular. Esta política de ordenamento do território deverá ser acompanhada de medidas que tornem os modos alternativos mais atrativos para as viagens

diárias, coordenadas com medidas que, onde possível e desejável, tornem o uso do automóvel particular mais caro e difícil ou pelo menos que diminuam a sua conveniência comparativamente às alternativas.

A dificuldade de testar o impacto de medidas promotoras de uma mobilidade sustentável em contextos reais, leva à necessidade de tentar modelar o comportamento dos cidadãos através de modelos matemáticos para simular atitudes atuais e prever comportamentos futuros. A variedade de possíveis modelações disponíveis para o planeamento de transportes, e mais especificamente para estudar os processos de tomada de decisão relativamente ao modo a utilizar, asseguram uma base teórica devidamente fundamentada que justifica o investimento de recursos para a sua correta definição, monitorização e eventual utilização no apoio a decisões políticas mais sustentáveis (e assim consubstanciadas). Da variedade de possíveis modelos disponíveis, será a formação técnica necessária para a sua implementação e uma correta avaliação da relação custo/investimento *versus* precisão das previsões e hipotéticos benefícios socioeconómicos resultantes das mesmas, que definirão qual o tipo de modelo a utilizar. A definição de um modelo matemático, devidamente fundamentado e integrando aspetos fundamentais tanto para os cidadãos como para os decisores políticos, poderá servir para justificar e auxiliar a aceitação de decisões que, embora polémicas, possam contribuir para o bem comum no futuro.

Apesar de os modelos baseados em atividades apresentarem fundamentos que os tornam propícios a uma modelação mais fidedigna das diversas opções modais ao longo do dia, têm também necessidade de dados mais detalhados para todas as atividades diárias. Esta necessidade limita a sua aplicação à possibilidade de executar levantamentos muito exaustivos e, eventualmente, específicos para essas modelações. Atendendo a estas exigências, a possibilidade de utilização de bases de dados pré-existent, opção normalmente muito interessante pelos menores custos que acarreta mas com algumas limitações ao nível da riqueza da informação, tenderá a limitar à partida as opções aos modelos mais tradicionais. Outro fator preponderante na seleção do tipo de modelo a utilizar está associado à relação custo de implementação *versus* precisão das estimativas, não se tendo encontrados estudos em quantidade suficiente que permitam justificar indubitavelmente os custos associados à modelação baseada em atividades comparativamente ao aumento de complexidade e a eventuais aumentos de precisão verificadas em relação aos modelos mais tradicionais.

Considerando o tradicional modelo dos quatro passos, a aplicação de diferentes tipologias de modelos, mais concretamente para a simulação do processo de decisão para as escolhas modais diárias, apesar de vasta, tem pecado por especificações demasiado restritas a situações específicas, tanto em termos de fatores como de realidades analisadas. Não foi possível encontrar estudos totalmente abrangentes que analisassem de forma integrada a enorme variedade de fatores, tanto sociais como relacionados com o ordenamento do

território ou com medidas de promoção de mobilidade sustentável, capazes de influenciar as escolhas modais. Também ao nível da identificação das tipologias de especificações mais adequadas para representar cada variável explicativa, e da deteção de potenciais tendências para a ocorrência de correlações, ainda que parciais, ou efeitos simbióticos entre variáveis explicativas, parece ainda haver lacunas significativas de conhecimento. Outro especto relevante intrínseco a todos os modelos de tomada de decisão, relaciona-se com a complexidade associada a aspetos qualitativos muito diferenciados. Estas dificuldades assentam não só na correta avaliação das especificidades dos sistemas de transportes existentes ou a implementar, como na forma como essa avaliação é feita pelos diferentes grupos sociais existentes na população alvo. A evolução dos modelos matemáticos para a simulação de comportamentos tem progredido significativamente nas últimas décadas, apresentando-se os modelos de escolha discreta como a família com fundamentação teórica sólida, maior aplicação em contextos reais e com resultados potencialmente satisfatórios para esforços de recolha de informação moderados.

Capítulo III

Modelos de Escolha Discreta

1. Introdução

Tendo presente o objetivo primordial do presente estudo de tentar analisar a forma como os cidadãos fazem as suas opções diárias na escolha do modo de transporte para as suas deslocações ao longo do dia, estimando o impacto das suas condições socioeconómicas e de diferentes políticas de mobilidade sustentável, da variedade de modelos matemáticos desenvolvidos nas últimas décadas, os modelos de escolha discreta, são dos que maior potencial evidenciam. Estes modelos, apesar de não permitirem analisar a influência do encadeamento dos motivos e das diferentes deslocações ao longo do dia, apresentam-se como uma ferramenta de modelação potente, com fundamentação teórica sólida, uma vasta aplicação a diferentes situações reais, e consistentes com as metodologias comuns de recolhas de dados. Atendendo à relação esforço de elaboração/precisão de estimativas, apresentam-se como uma excelente alternativa.

A base teórica dos modelos de escolha discreta torna-os ferramentas indispensáveis para várias áreas de investigação, especialmente as que necessitam de simular o modo como cada indivíduo toma as suas decisões. Se a regressão linear múltipla se apresenta como uma ferramenta com enorme potencial em muitos ramos da engenharia, a sua limitação a variáveis contínuas inviabiliza a sua utilização em situações em que o conjunto de possíveis respostas é bastante conciso ou mesmo limitado ao sim ou não. A família de modelos de escolha discreta apresenta formulações variadas, desde modelos baseados em pressupostos mais simples, a modelos mais complexos que permitem a introdução de informação de forma a simular uma miríade de possíveis cenários e ambientes.

No presente capítulo apresenta-se a fundamentação teórica subjacente aos modelos de escolha discreta em geral, caracterizando posteriormente as suas especificações mais comuns nomeadamente os Logit multinomiais, os Probit e os Logit Mistos. São apresentadas as suas formulações, as suas vantagens e desvantagens, e os diferentes aspetos a atender na sua implementação e seleção. Atendendo às características dos modelos mais complexos que necessitam de simulação de integrais, apresenta-se uma breve introdução à sua aplicação bem com aos desenvolvimentos mais recentes para tornar o seu cálculo mais célere. O capítulo é finalizado com a apresentação de alguns casos de estudo internacionais com

aplicação de modelos de escolha discreta a problemas de transporte, cujos resultados servirão de base e de termo de comparação com os obtidos no presente estudo.

2. Fundamentação Teórica dos Modelos de Escolha Discreta

2.1. Teoria de Escolhas Individuais

Geralmente interessa conhecer o comportamento de um grande número de indivíduos ou organizações expresso em termos de quantidades agregadas tais como a cota de mercado de um bem ou serviço. Este comportamento agregado, no entanto, resulta de decisões individuais pelo que a sua modelação está implicitamente correlacionada com todos os modelos de previsão de comportamento desagregados. É, portanto, fundamental conhecer alguns princípios da teoria de decisões individuais, úteis para a formulação empírica de modelos de escolha discreta, entre os quais se destacam (Ben-Akiva e Lerman, 1985):

- Descritivos: Os modelos devem postular o modo como um indivíduo se comporta e não como se deve comportar;
- Abstratos: Os modelos devem ser formalizados num sentido geral e não especificamente para circunstâncias particulares;
- Operacionais: Devem resultar em modelações com parâmetros mensuráveis e estimáveis.

No entanto, não existe uma única, e universalmente aceite, teoria de escolhas que satisfaça estes requisitos, existindo várias que diferem principalmente no nível de detalhe idealizado para o processo de escolha, que produz um determinado comportamento observado.

Pode-se especificar que uma teoria de escolhas é uma compilação de procedimentos que defina os seguintes elementos:

- O decisor;
- As alternativas disponíveis;
- Os atributos das alternativas;
- As regras de decisão.

Em relação ao decisor a unidade pode ser o indivíduo ou um grupo de pessoas, como uma família ou um agregado familiar, uma firma ou uma entidade governamental. A consideração de um grupo de pessoas como um único decisor permite a abstração parcial das complexas interações intrínsecas desse mesmo grupo. Cada decisor enfrenta diferentes situações de escolha e tem gostos bastante distintos, pelo que mesmo pretendendo previsões de procura agregadas o processo de decisão deve explicitamente atender à diversidade entre indivíduos. Deve ainda atender-se à possibilidade do processo de decisão poder divergir atendendo às opções resultantes de variações das interações entre indivíduos do grupo, como, por exemplo,

em casos onde a decisão final pode ser a do membro mais proeminente ou resultar da decisão democrática da maioria dos seus membros.

Em relação às alternativas disponíveis o ambiente definirá o *conjunto universal* de alternativas à disposição de um decisor, mas este apenas considerará um subconjunto que se designa por *conjunto de escolhas*. Este conjunto de escolhas será a parte do conjunto universal que o decisor tem efetivamente à sua disposição e dos quais este tem conhecimento no momento da tomada de decisão. A disponibilidade de uma alternativa dependerá de uma variedade de restrições como a disponibilidade física, os seus recursos financeiros, a disponibilidade temporal, as lacunas de informação, entre outros.

Os atributos/variáveis das alternativas permitem avaliar a atratividade de uma alternativa e, tal como já foi referido, podem ser discretos ordinais, discretos nominais ou numéricos/contínuos. O caso mais complexo implica a utilização de alternativas heterogêneas e onde os decisores têm diferentes conjuntos de alternativas disponíveis, avaliam diferentes atributos e valorizam de forma distinta os mesmos atributos das mesmas alternativas. Outro aspeto a atender tem a ver com a incerteza de resultados, podendo haver atributos cujo valor é conhecido com certeza, como o preço de aquisição, enquanto outros são incertos, como o custo de manutenção.

As regras de decisão são fundamentais para que se possa modelar uma decisão de entre um conjunto de duas ou mais alternativas. Estas regras tentam descrever mecanismos internos de processamento de informação disponível que permitem chegar a uma escolha única. Segundo Ben-Akiva e Lerman (1985) podem ser classificadas nas seguintes categorias:

- *Dominância*: Uma alternativa pode ser predominante em relação a outra se for melhor em pelo menos um atributo e não for pior nas restantes. É a regra menos controversa, mas pode levar a uma solução não única, pelo que pode ser útil para eliminar as piores alternativas do conjunto das disponíveis. Em situações reais existem muitos atributos relevantes para cada indivíduo e são raras as alternativas dominantes sobre todos os atributos. Outra situação a considerar é a complexidade devido a grandeza relativa entre atributos, em que apenas uma diferença significativa do valor de um atributo entre alternativas torna uma efetivamente melhor que outra. Um exemplo deste tipo de problemas é o tempo de viagem, cuja diferença de menos 5 minutos pode não ser suficiente para que um indivíduo considere o TC melhor que o TI;
- *Satisfação*: Cada atributo tem um nível que serve de critério de satisfação, definido o nível de aspiração, baseado nas expectativas do indivíduo em relação ao alcançável atendendo à informação que detém e às suas experiências passadas. Uma alternativa pode ser eliminada se pelo menos um dos seus atributos não alcançar o nível de satisfação mínimo, não sendo, no entanto, um critério suficiente para levar a uma

escolha por um indivíduo, podendo a sua combinação com outros critérios, como a dominância, ser mais decisiva;

- *Hierarquização de alternativas*: Se os atributos estiverem ordenados por nível de importância, o indivíduo escolhe a alternativa que tem melhor nível do atributo mais importante. Quando o atributo é qualitativo todas as alternativas que possuem esse atributo pertencem ao conjunto de alternativas viáveis, e a definição da sua escolha passa por avaliar a existência ou não do segundo atributo mais valorizado. Este processo pode estender-se por várias etapas até que apenas uma alternativa permaneça e seja a escolha do indivíduo. Uma variação desta regra pode ser a eliminação, em cada etapa, das alternativas com pior desempenho nos atributos mais valorizados. A combinação desta regra com a satisfação é designada por “eliminação por aspetos”;
- *Utilidade*: Este tipo de regras de decisão assume a comensurabilidade dos atributos, isto é, que a atratividade de uma alternativa pode ser expressa através de um vetor de atributos possíveis de reduzir a um valor escalar. Esta regra permite definir uma única função objetiva que expresse a atratividade de uma alternativa em termos dos seus atributos, que um indivíduo pretende maximizar através da sua escolha. A suposição de um índice único é baseada na noção de compensações que o decisor faz, explícita ou implicitamente, quando compara diferentes atributos, ou seja, o indivíduo escolherá a alternativa que tenha, na sua perspetiva, a melhor combinação dos atributos mais importantes. A função utilidade pode ser explicitada de diferentes formas, sendo fundamental a sua distinção entre utilidades ordinais ou cardinais. As utilidades ordinais são uma expressão matemática da ordenação preferencial das alternativas, e o seu valor numérico não tem significado excetuando no seu ordenamento, permitindo apenas análise de relações menor que, igual a ou maior que. A utilidade cardinal implica um valor numérico único e é mais restritiva que a ordinal, sendo utilizada em teoria de decisões com incerteza onde se assume que os indivíduos maximizam uma determinada medida da utilidade esperada.

2.2. Teoria da Maximização da Utilidade Aleatória

A abordagem da maximização da utilidade aleatória assume que um indivíduo escolhe sempre a alternativa que tem maior utilidade. Se se observarem inconsistências comportamentais assume-se que estas resultam de deficiências de observação por parte do analista, que não conhece as utilidades reais de cada alternativa, considerando-as como variáveis aleatórias. As fontes de aleatoriedade das utilidades podem ser divididas em quatro grupos distintos (Ben-Akiva e Lerman, 1985):

- *Atributos não analisados ou observados*: como o vetor de atributos que afetam uma decisão é incompleto, a função de utilidade inclui um elemento que em termos de observação é uma variável aleatória tornando a própria utilidade aleatória;

- *Variações de gosto não observadas*: a utilidade pode ter atributos não observados que variem de indivíduo para indivíduo, cuja variação desconhecida torna a utilidade, em termos de observação, uma variável aleatória;
- *Erros de medição/quantificação*: Como a função de utilidade tem atributos não mensuráveis a utilidade quantificável é uma estimativa imperfeita da utilidade real, pelo que se inclui na sua formulação um termo que incorpora os erros desconhecidos de medição, impondo à utilidade mensurável um elemento aleatório;
- *Variáveis instrumentais*: Alguns dos atributos que influenciam a escolha não são mensuráveis, ou pelo menos são difíceis de avaliar, pelo que a utilidade apenas incorpora atributos intrinsecamente relacionados com métodos ou medidas imperfeitas para os avaliar. Esta relação imperfeita inclui, de novo, um erro aleatório impondo à utilidade um elemento aleatório.

2.3. Formulação do Processo Comportamental

Considere-se que quando um cliente, pessoa, empresa, ou decisor, tem que tomar uma decisão, ou conjunto/sequência de decisões, ao longo do tempo entre um conjunto restrito de opções, que essa escolha é designada por y . Se o objetivo for entender os fatores que coletivamente determinam ou causam o processo comportamental que leva à escolha do indivíduo, existem fatores observáveis (x) e não observáveis (ε). Estes fatores relacionam-se com a escolha do indivíduo através de uma função $y = h(x, \varepsilon)$, designada de processo comportamental (Train, 2003) e seria determinística se conhecendo x e ε , se pudesse determinar plenamente a escolha do indivíduo. Como os ε não são conhecidos, a escolha não é determinística e não pode ser prevista com exatidão. Existe assim a necessidade de determinar a probabilidade que uma opção tem em ser escolhida. Os termos não observáveis ε são considerados aleatórios com densidade¹ $f(\varepsilon)$. A probabilidade de um indivíduo escolher uma determinada opção entre as alternativas disponíveis é a probabilidade dos fatores não observáveis serem tais que o processo comportamental resulte nessa escolha. Definindo uma função indicadora como:

$$I[h(x, \varepsilon) = y]$$

Exp. III- 1

Esta assume o valor 1 quando a expressão entre parêntesis for verdade e 0 quando for falsa, ou seja, a função indicadora terá valor 1 quando o valor de ε , combinado com os valores de x , induzir o indivíduo a escolher a opção y e terá valor 0 quando o valor de ε , combinado com os valores x , induzir o indivíduo a escolher outra opção que não y . A probabilidade de um

¹ - Densidade ou função de densidade de probabilidade, é a expressão utilizada para o cálculo de probabilidades e para especificar a distribuição de probabilidades, que para um espaço amostral descreve o conjunto de possíveis soluções e o método para o cálculo das probabilidades. Para uma variável aleatória discreta, a distribuição de probabilidades descreve os seus possíveis valores e uma formulação para a determinação de probabilidades (Montgomery e Runger, 2007).

indivíduo escolher a opção y será simplesmente o valor esperado desta função indicadora considerando todos os possíveis valores dos fatores não observáveis (Train, 2003):

$$P(y|x) = \text{prob}(I[h(x, \varepsilon) = y] = 1) = \int I[h(x, \varepsilon) = y] \cdot f(\varepsilon) \cdot d\varepsilon \quad \text{Exp. III- 2}$$

Para calcular esta probabilidade o integral pode ser avaliado de três maneiras: através de expressões de forma completamente fixa, simulação completa e através de simulação e forma fixa parciais.

Considerando que para certas especificações de h e f o integral pode ser expresso com forma fixa, a probabilidade de escolha pode ser calculada de forma precisa. Se se atender que um indivíduo apenas escolhe uma opção se esta lhe providenciar algum benefício real, ou seja, se tiver uma utilidade positiva, e considerando o que o analista pode efetivamente observar ($\beta'x$), a sua probabilidade de escolha será:

$$P = \int I[\beta'x + \varepsilon > 0] \cdot f(\varepsilon) \cdot d\varepsilon \quad \text{Exp. III- 3}$$

Assumindo que ε tem uma distribuição logística, de modo a que a sua função densidade $f(\varepsilon)$ e distribuição cumulativa² $F(\varepsilon)$ sejam:

$$f(\varepsilon) = \frac{e^{-\varepsilon}}{(1 + e^{-\varepsilon})^2} \quad \text{Exp. III- 4}$$

$$F(\varepsilon) = \frac{1}{1 + e^{-\varepsilon}} \quad \text{Exp. III- 5}$$

Então a probabilidade de um indivíduo escolher uma opção será:

$$\begin{aligned} P &= \int I[\beta'x + \varepsilon > 0] \cdot f(\varepsilon) \cdot d\varepsilon = \int I[\varepsilon > -\beta'x] \cdot f(\varepsilon) \cdot d\varepsilon = \\ &= \int_{\varepsilon=-\beta'x}^{\infty} f(\varepsilon) \cdot d\varepsilon = 1 - F(-\beta'x) = 1 - \frac{1}{1 + e^{\beta'x}} = \frac{e^{\beta'x}}{1 + e^{\beta'x}} \end{aligned} \quad \text{Exp. III- 6}$$

Podendo ser calculada exatamente através da expressão III.6. Vários modelos de escolha discreta usam expressões de forma fixa para determinar as probabilidades, nomeadamente o Logit Multinomial e o Logit Hierárquico (“nested”).

2 - Função de distribuição cumulativa de uma variável discreta aleatória X permite calcular a probabilidade de X ser menor que um determinado valor x , isto é:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

satisfazendo as seguintes propriedades: $0 \leq F(x) \leq 1$ e se $x \leq y$, então $F(x) \leq F(y)$ (Montgomery e Runger, 2007).

Em geral, no entanto, o integral de probabilidades não é possível de ser expresso numa forma fixa, uma vez que esta simplificação impõe restrições no modelo comportamental h e na distribuição dos termos aleatórios f , que podem tornar o modelo irrealista para muitas situações. Assim, em vez de resolver o integral analiticamente, outra opção é a aproximação por simulação, aplicável a quase todas as especificações de h e f . A simulação apoia-se no facto de a integração de uma densidade ser uma forma de avaliar a média (Train, 2003). Considere-se $t(\varepsilon)$ uma estatística baseada em ε com densidade $f(\varepsilon)$. Se se escolherem aleatoriamente numerosas vezes erros ε da sua distribuição f , e se se calcularem os seus $t(\varepsilon)$, o cálculo da sua média (neste caso simulada) seria uma estimativa não enviesada da sua média real, aproximando-se cada vez mais do valor correto à medida que mais valores forem sendo escolhidos. Apesar de fácil de entender esta simulação tem a desvantagem de as probabilidades de escolha poderem ser médias de outras funções estatísticas diferentes da média de uma função indicadora. No entanto, a metodologia de escolher valores aleatoriamente da sua densidade, calcular a sua estatística e realizar a média dos resultados continua válida. Entre os modelos estimados por simulação, o mais conhecido é o Probit.

Em muitas situações, no entanto, é possível aproveitar as vantagens de ambas as metodologias anteriores. Supondo que os termos aleatórios podem ser decompostos em duas partes ε_1 e ε_2 , com densidade conjunta $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$. A densidade conjunta pode ser expressa como o produto de uma densidade marginal e de uma densidade condicionada, isto é, $f(\varepsilon_1, \varepsilon_2) = f(\varepsilon_2 | \varepsilon_1) \cdot f(\varepsilon_1)$, que implica que a equação de probabilidades seja:

$$P(y | x) = \int I[h(x, \varepsilon) = y] \cdot f(\varepsilon) \cdot d(\varepsilon) =$$

$$= \int \left[\int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] \cdot f(\varepsilon_2 | \varepsilon_1) \cdot d\varepsilon_2 \right] \cdot f(\varepsilon_1) \cdot d\varepsilon_1 \quad \text{Exp. III- 7}$$

Se se considerar que a expressão entre parêntesis retos grandes pode ser substituída por uma forma fixa $g(\varepsilon_1)$, condicionada ao valor de ε_1 , tal que:

$$g(\varepsilon_1) \equiv \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] \cdot f(\varepsilon_2 | \varepsilon_1) \cdot d\varepsilon_2 \quad \text{Exp. III- 8}$$

Então a probabilidade de escolha de uma alternativa toma a forma:

$$P(y | x) = \int_{\varepsilon_1} g(\varepsilon_1) \cdot f(\varepsilon_1) \cdot d(\varepsilon_1) \quad \text{Exp. III- 9}$$

Este procedimento segundo Train (2003) designa-se de *partição conveniente dos erros*, e permite o cálculo exato do integral sobre ε_2 dado ε_1 enquanto o integral sobre ε_1 é simulado.

Caso não exista uma solução para o integral com forma fixa este também pode ser simulado considerando a média de g atendendo a densidade marginal de ε_i . As vantagens desta abordagem mista surgem do facto de integrais analíticos serem mais precisos e fáceis de calcular do que os integrais simulados, sendo uma opção bastante útil quando a decomposição de parte dos termos aleatórios e respetiva integração analítica for possível. Em exemplo proeminente deste tipo de abordagem é o modelo Logit Misto (“*Mixed Logit*”).

2.4. Generalidades dos Modelos de Escolha Discreta

Outro aspeto a atender na definição do processo comportamental de decisão, são os vários aspetos que afetam a especificação e estimação dos modelos de escolha discreta, que podem ser resumidos em duas afirmações (Train, 2003):

- “...Apenas as diferenças entre utilidades importam...”
- “...A escala da utilidade é arbitrária...”

2.4.1. Diferença entre Utilidades

A primeira afirmação implica que o nível absoluto da utilidade seja irrelevante tanto para o comportamento do decisor como para o modelo do analista. Se uma constante for adicionada à utilidade de todas as alternativas, a que maior utilidade tinha não altera a sua ordem, uma vez que todas aumentam na mesma proporção. Do mesmo modo, também não alteram a interpretação do analista, que apenas depende da diferença entre utilidades e não no seu valor absoluto, uma vez que a alternativa i de um conjunto de J alternativas só será escolhida pelo indivíduo n se a sua utilidade for maior que a das restantes, isto é:

$$U_{ni} > U_{nj} \quad \forall j \neq i \quad \text{Exp. III- 10}$$

Assim, a probabilidade de escolha será:

$$P_{ni} = \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i) = \text{Prob}(U_{ni} - U_{nj} > 0 \quad \forall j \neq i) \quad \text{Exp. III- 11}$$

que depende apenas da diferença entre as utilidades e não do seu valor absoluto. Mesmo decompondo a utilidade nas suas componentes mensuráveis e não mensuráveis:

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i) = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall j \neq i) = \\ &= \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad \text{Exp. III- 12} \end{aligned}$$

A probabilidade de escolha continua apenas a depender da diferença entre componentes e não no seu valor absoluto. Este facto tem várias implicações na identificação e especificação do modelo de escolha discreta, nomeadamente que apenas os parâmetros que podem ser

estimados/identificados revelam diferenças entre alternativas. Esta generalização pode ser verificada de diversas formas: na definição das constantes específicas das alternativas; na definição das variáveis sociodemográficas; e na definição do número de termos para os erros independentes.

2.4.2. Constantes Específicas das Alternativas

Se se especificar que a parte mensurável da utilidade tem uma formulação com parâmetros lineares com uma constante:

$$V_{nj} = \theta \cdot X_{nj} + k_j \forall j \quad \text{Exp. III- 13}$$

onde X_{nj} é um vetor de variáveis relacionadas com a alternativa j oferecida ao decisor n , θ são os coeficientes/pesos dos atributos e k_j é a constante específica da alternativa j . Esta constante específica de cada alternativa apreende o efeito médio de todos os fatores não incluídos no modelo, semelhante à constante dos modelos de regressão. Se esta constante for incluída a parte não mensurável da utilidade (ϵ) terá média igual a zero por definição (Train, 2003). Se (ϵ) tem uma média diferente de zero quando as constantes não são incluídas, então sua adição impõe que os erros remanescentes tenham média igual a zero, ou seja, se:

$$U_{nj} = \theta \cdot X_{nj} + \epsilon_{nj}^*, \text{ com } \mu(\epsilon_{nj}^*) = k_j \neq 0 \quad \text{Exp. III- 14}$$

então

$$U_{nj} = \theta \cdot X_{nj} + k_j + \epsilon_{nj}, \text{ com } \mu(\epsilon_{nj}) = 0 \quad \text{Exp. III- 15}$$

sendo portanto razoável incluir uma constante específica na parte mensurável da utilidade de cada alternativa. Como apenas a diferença entre utilidades interessa, apenas a diferença entre as constantes específicas é relevante e não o seu valor absoluto. Como o número de combinações de constantes específicas cuja diferença é semelhante, implicando a mesma probabilidade de escolha, é infinito, existe a necessidade de normalizar o nível absoluto das constantes, que pode ser feito normalizando uma das constantes com o valor zero. O resultado deste procedimento implica que a constante das alternativas não nulas sejam interpretadas como o efeito médio dos fatores não incluídos na utilidade em relação à alternativa normalizada a zero (por exemplo, em relação ao automóvel particular). A escolha da constante específica normalizada a zero é irrelevante e pode ser normalizada para qualquer outro valor, mas a utilização do valor zero simplifica o procedimento uma vez que simplesmente exclui a constante da alternativa normalizada do modelo.

2.4.3. Variáveis Sociodemográficas ou Socioeconómicas

O mesmo problema afeta o modo como as variáveis sociodemográficas entram no modelo. Atributos como o tempo e o custo da viagem geralmente variam entre alternativas, no entanto, os atributos do decisor não variam e só poderão ser incluídas no modelo se especificadas de modo a traduzir diferenças de utilidade entre alternativas. Um exemplo claro é o rendimento de um indivíduo, que apesar de ter um coeficiente sempre positivo (o aumento do rendimento é sempre avaliado favoravelmente) pode entre alternativas ter valores diferentes, pois o aumento do rendimento tem provavelmente efeitos diferentes dependendo do modo utilizado (um passageiro de transportes coletivos poderá passar a ter rendimentos suficientes que lhe permitam finalmente mudar para o tão desejado automóvel particular, enquanto que um condutor não passaria a andar de TC). Mais uma vez, os valores absolutos dos coeficientes deste atributo não podem ser estimados mas apenas a sua diferença, pelo que a sua normalização, tal como descrita anteriormente, facilita a aplicação e interpretação do modelo. Os coeficientes do rendimento das restantes alternativas não normalizadas a zero traduzem o efeito diferencial do mesmo em relação à alternativa normalizada, podendo ter valores positivos ou negativos. Outra maneira de atenuar o efeito referido pode ser a divisão dos custos pelo rendimento, implicando que o coeficiente passe a ser θ/R , decrescendo com o aumento do rendimento e refletindo o conceito de que o custo se torna menos importante nas decisões modais à medida que o rendimento aumenta. Quando os coeficientes são relacionados com os atributos não há necessidade de normalizar os primeiros, uma vez que as variáveis sociodemográficas afetam as diferenças em utilidade através da sua interação com os atributos (Train, 2003).

2.4.4. Número de Termos para os Erros Independentes

Considerando a expressão III.12, a probabilidade de escolha é um integral com J dimensões (ou seja, igual ao número de alternativas) em relação à densidade de J termos de erros, $\varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nJ})$. Este integral também pode ser simplificado para uma dimensão $J-1$, tendo em conta que apenas a diferença entre erros importa e não o seu valor absoluto. A densidade da diferença entre erros tem sempre menos uma dimensão que a densidade original dos erros, pelo que há necessidade de normalizar uma das dimensões dos erros originais. Esta normalização é específica para cada tipo de modelo com exceção dos modelos mais simples como os Logit multinomiais que realizam a sua normalização automaticamente com as suposições da sua distribuição. O analista deverá considerar assim que apenas a densidade de diferenças entre erros afeta a probabilidade de escolha e, portanto, pode ser identificada.

2.4.5. Escala da Utilidade

A adição de uma constante à utilidade de todas as alternativas não altera a escolha de um indivíduo, nem a multiplicação de utilidade de cada alternativa por uma constante. A alternativa com maior utilidade terá a mesma ordem independentemente da escala da utilidade. Assim, o analista deve normalizar a escala da utilidade e a forma mais comum é

através da normalização da variância dos seus erros. A escala da utilidade e a variância dos seus erros estão interligados uma vez que multiplicando uma utilidade por λ implica multiplicar a variância de cada erro por λ^2 , ou seja, $\text{Var}(\lambda \cdot \epsilon_{nj}) = \lambda^2 \cdot \text{Var}(\epsilon_{nj})$ (Train, 2003). Assim, normalizar a variância dos erros é equivalente a normalizar a escala da utilidade e depende da distribuição de probabilidades assumida para os erros.

2.4.5.1. Normalização da Escala da Utilidade assumindo Erros Independentes e Identicamente Distribuídos (IID)

Se se assumir que os erros são independentes e identicamente distribuídos (mesma variância) - IID, como por exemplo nos modelos Logit, a sua normalização é bastante simples. A normalização pode ser realizada com qualquer valor e, uma vez que todos os erros têm a mesma variância, a normalização ajusta a variância de todos os erros. Assumindo um modelo linear do tipo:

$$U_{nj}^0 = B \cdot X_{nj} + \epsilon_{nj}^0, \text{ com } \text{Var}(\epsilon_{nj}^0) = \sigma^2 \quad \text{Exp. III- 16}$$

A normalização da escala pode ser realizada assumindo a variância do erro igual a 1. O modelo normalizado passará a ter a seguinte forma equivalente:

$$U_{nj}^1 = \left(\frac{B}{\sigma} \right) \cdot X_{nj} + \epsilon_{nj}^1 = U_{nj}^1 = B' \cdot X_{nj} + \epsilon_{nj}^1, \text{ com } \text{Var}(\epsilon_{nj}^1) = 1 \quad \text{Exp. III- 17}$$

Os coeficientes originais B são divididos pelo desvio padrão (σ) da parte não mensurável da utilidade e refletem os efeitos das variáveis observáveis ou mensuráveis relativamente ao desvio padrão dos fatores não mensuráveis (Train, 2003). Este conceito aplica-se qualquer que seja o valor admitido para a variância dos termos não mensuráveis, erros ou parte aleatória. No modelo Logit tradicional é comum normalizá-lo com uma variância de valor $\pi^2/6$, aproximadamente 1,6:

$$U_{nj} = \beta' \cdot \sqrt{1,6} \cdot X_{nj} + \epsilon_{nj}, \text{ com } \text{Var}(\epsilon_{nj}) = 1,6 \quad \text{Exp. III- 18}$$

Os coeficientes ainda refletem a variância da parte não observada da utilidade mas estão incrementados num fator de $\sqrt{1,6}$.

Há, no entanto, que tomar algumas precauções na comparação entre modelos com escalas diferentes, uma vez que podem induzir a conclusões erradas e mesmo contraditórias. Apesar de dentro do mesmo modelo a normalização não afetar os resultados (todos aumentam ou diminuem proporcionalmente), na comparação de resultados obtidos a partir de modelos diferentes há necessidade de aplicar escalas de normalização semelhantes aos seus resultados, sob risco de os resultados não poderem ser imediatamente comparáveis.

O mesmo problema coloca-se na estimação de um mesmo modelo mas utilizando diferentes conjuntos de dados. A escala relativa das estimativas de dois conjuntos de dados diferentes reflete a relativa variância dos erros dos mesmos conjuntos de dados. Como os coeficientes originais são divididos pelo desvio padrão, coeficientes normalizados maiores significam desvios padrão dos erros e respectivas variâncias menores, e vice-versa. Esta constatação indica que quanto menor for o coeficiente normalizado de um atributo, e consecutivamente maiores o desvio padrão e a variância dos erros, maior a importância relativa dos atributos não observados nas escolhas dos indivíduos, e que quanto maior for o coeficiente normalizado, menores o desvio padrão e a variância dos erros, evidenciando que a importância relativa dos erros em relação ao atributo visado é menor (Train, 2003). Resumindo, quanto maior o valor dos coeficientes normalizados, maior a importância dos seus atributos, correlacionados em relação aos fatores não observados ou erros, na probabilidade de escolha.

2.4.5.2. Normalização da Escala da Utilidade assumindo Erros com Diferentes Variâncias - Heteroscedasticidade

Em certas situações a variância dos erros pode ser diferente para diversos segmentos da população, o que inviabiliza a definição do nível global da utilidade através da normalização dos erros de todos os segmentos. A abordagem sugerida por Train (2003) define o nível global da utilidade através da normalização da variância para um determinado segmento para posteriormente estimar a variância (e a escala) de cada um dos restantes segmentos em relação ao inicial. Um exemplo pode ser a de um modelo aplicado a duas cidades diferentes A e B, considerando o tempo de viagem (T) e o custo monetário (C) da mesma:

$$\begin{aligned} U_{nj} &= \alpha \cdot T_{nj} + \beta \cdot C_{nj} + \varepsilon_{nj}^A \quad \forall n \in A \\ U_{nj} &= \alpha \cdot T_{nj} + \beta \cdot C_{nj} + \varepsilon_{nj}^B \quad \forall n \in B \end{aligned} \quad \text{Exp. III- 19}$$

em que os erros têm variâncias diferentes, $\text{Var}(\varepsilon_{nj}^A) \neq \text{Var}(\varepsilon_{nj}^B)$. Se a razão entre variâncias for designada por:

$$k = \frac{\text{Var}(\varepsilon_{nj}^B)}{\text{Var}(\varepsilon_{nj}^A)} \quad \text{Exp. III- 20}$$

então pode dividir-se a utilidade dos viajantes para B por $\sqrt{k}$, que não afeta as suas escolhas uma vez que a escala da utilidade na interessa. No entanto, este passo permite reescrever as expressões da utilidade como:

$$U_{nj} = \alpha \cdot T_{nj} + \beta \cdot C_{nj} + \varepsilon_{nj} \forall n \in A$$

$$U_{nj} = \left(\frac{\alpha}{\sqrt{k}} \right) \cdot T_{nj} + \left(\frac{\beta}{\sqrt{k}} \right) \cdot C_{nj} + \varepsilon_{nj} \forall n \in B$$
Exp. III- 21

onde a variância ε_{nj} é igual para qualquer indivíduo n em ambas as cidades, já que (Train, 2003):

$$\text{Var}\left(\frac{\varepsilon_{nj}^B}{\sqrt{k}}\right) = \left(\frac{1}{k}\right) \cdot \text{Var}(\varepsilon_{nj}^B) = \left[\frac{\text{Var}(\varepsilon_{nj}^A)}{\text{Var}(\varepsilon_{nj}^B)}\right] \cdot \text{Var}(\varepsilon_{nj}^B) = \text{Var}(\varepsilon_{nj}^A)$$
Exp. III- 22

A escala da utilidade é definida normalizando a variância de ε_{nj} , e o parâmetro k , frequentemente designado por parâmetro de escala, é estimado paralelamente com α e β . A estimação do valor de k , dá indicação ao analista da variância dos fatores não observados (não mensuráveis ou erros) em B relativamente a A . Por exemplo, um $k'=1.2$ (estimativa do k real) implica que a variância dos erros na cidade B é 20% superior à variância dos erros na cidade A . O exemplo dado representa diferenças na variância dos erros devido a alterações no espaço geográfico, mas podem também dever-se à utilização de diferentes conjuntos de dados (por exemplo, preferências declaradas *versus* preferências reveladas), a diferenças temporais (por exemplo, dados de diferentes censos anuais), entre outros (Train, 2003).

2.4.5.3. Normalização da Escala da Utilidade assumindo Erros Correlacionados

Nos pontos anteriores assumia-se que os erros não eram correlacionados, ou seja, eram independentes entre alternativas. No entanto, quando os mesmos estão correlacionados entre alternativas, ou seja, não são independentes, a normalização da escala da utilidade é mais complexa. Uma vez que apenas a diferença entre utilidades interessa, é mais correto designar-se por normalização da escala das diferenças entre utilidades, mas se os erros entre alternativas estão correlacionados, na sua determinação não é suficiente normalizar a variância dos erros para uma alternativa. Train (2003) apresenta um exemplo para quatro alternativas, considerando que a função utilidade é semelhante à expressão II.7:

$$U_{nj} = V_{nj} + \varepsilon_{nj}, \text{ com } j = 1, \dots, 4$$
Exp. III- 23

e que o vetor com erros é:

$$\varepsilon_n = \langle \varepsilon_{n1} \dots \varepsilon_{n4} \rangle$$
Exp. III- 24

com média igual a zero e matriz de covariância simétrica³:

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \cdot & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \cdot & \cdot & \sigma_{33} & \sigma_{34} \\ \cdot & \cdot & \cdot & \sigma_{44} \end{pmatrix}$$

Como apenas a diferença entre utilidades interessa, o modelo equivalente, considerando a diferença em relação à primeira alternativa, por exemplo o automóvel particular, será:

$$\tilde{U}_{nj1} = \tilde{V}_{nj1} + \tilde{\varepsilon}_{nj1}, \text{ com } j = 2, 3 \text{ e } 4 \text{ e} \quad \text{Exp. III- 25}$$

$$\tilde{U}_{nj1} = U_{nj} - U_{n1}; \quad \text{Exp. III- 26}$$

$$\tilde{V}_{nj1} = V_{nj} - V_{n1}; \quad \text{Exp. III- 27}$$

$$\tilde{\varepsilon}_{nj1} = \langle (\varepsilon_{n2} - \varepsilon_{n1}), (\varepsilon_{n3} - \varepsilon_{n1}), (\varepsilon_{n4} - \varepsilon_{n1}) \rangle \quad \text{Exp. III- 28}$$

A variância da diferença entre cada erro depende da variância e covariância dos erros iniciais, e podem ser expressos da seguinte forma (ver determinação das expressões no Anexo 1 do Capítulo III):

- Variância entre os 1º e 2º erros:

$$\text{Var}(\tilde{\varepsilon}_{n21}) = \sigma_{11} + \sigma_{22} - 2 \cdot \sigma_{12} \quad \text{Exp. III- 29}$$

- Variância entre os 1º e 3º erros:

$$\text{Var}(\tilde{\varepsilon}_{n31}) = \sigma_{11} + \sigma_{33} - 2 \cdot \sigma_{13} \quad \text{Exp. III- 30}$$

- Variância entre os 1º e 4º erros:

$$\text{Var}(\tilde{\varepsilon}_{n41}) = \sigma_{11} + \sigma_{44} - 2 \cdot \sigma_{14} \quad \text{Exp. III- 31}$$

- Covariância dos erros entre os 2ºs e os 3ºs erros:

$$\text{Cov}(\tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n31}) = \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} \quad \text{Exp. III- 32}$$

- Covariância dos erros entre os 2ºs e os 4ºs erros:

$$\text{Cov}(\tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n41}) = \sigma_{24} - \sigma_{12} - \sigma_{14} + \sigma_{11} \quad \text{Exp. III- 33}$$

3 - *Covariância* - É uma medida da associação entre duas variáveis aleatórias determinada como o valor esperado do produto das duas variáveis aleatórias em torno das suas médias: $\text{Cov}(X, Y) = E[(X - \mu_x) \cdot (Y - \mu_y)]$, com μ_x e μ_y as médias respetivamente da variável X e Y.

Matriz de Covariância - Matriz quadrada que contém as variâncias e covariâncias de um conjunto de variáveis aleatórias. A diagonal principal é constituída pelas variâncias de cada variável aleatória, e os restantes elementos são as covariâncias entre as variáveis (Montgomery e Runger, 2007).

- Covariância dos erros entre os 3^{os} e os 4^{os} erros:

$$\text{Cov}(\tilde{\varepsilon}_{n31}, \tilde{\varepsilon}_{n41}) = \sigma_{34} - \sigma_{13} - \sigma_{14} + \sigma_{11} \quad \text{Exp. III- 34}$$

Resultando na seguinte matriz de covariância do vetor das diferenças entre erros:

$$\tilde{\Omega}_1 = \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2 \cdot \sigma_{12} & \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} & \sigma_{24} - \sigma_{12} - \sigma_{14} + \sigma_{11} \\ \cdot & \sigma_{11} + \sigma_{33} - 2 \cdot \sigma_{13} & \sigma_{34} - \sigma_{13} - \sigma_{14} + \sigma_{11} \\ \cdot & \cdot & \sigma_{11} + \sigma_{44} - 2 \cdot \sigma_{14} \end{pmatrix}$$

Nestes casos para definir a escala da utilidade, não é suficiente a normalização de uma das variâncias dos erros originais, é necessário normalizar uma das variâncias entre erros, uma vez que definindo a variância de uma das diferenças entre erros, define-se a escala da utilidade e assim a utilidade (Train, 2003). A forma mais simples é normalizar uma das variâncias entre erros igualando-a à unidade, que resulta na seguinte matriz de covariância de diferenças entre erros em termos de covariâncias dos erros originais:

$$\tilde{\Omega}_1^* = \begin{pmatrix} 1 & \frac{(\sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11})}{m} & \frac{(\sigma_{24} - \sigma_{12} - \sigma_{14} + \sigma_{11})}{m} \\ \cdot & \frac{(\sigma_{11} + \sigma_{33} - 2 \cdot \sigma_{13})}{m} & \frac{(\sigma_{34} - \sigma_{13} - \sigma_{14} + \sigma_{11})}{m} \\ \cdot & \cdot & \frac{(\sigma_{11} + \sigma_{44} - 2 \cdot \sigma_{14})}{m} \end{pmatrix}$$

Onde $m = \sigma_{11} + \sigma_{22} - 2 \cdot \sigma_{12}$ e as utilidades são divididas por $\sqrt{\sigma_{11} + \sigma_{22} - 2 \cdot \sigma_{12}}$ para se obter a escala definida.

A interpretação dos resultados desta matriz, no entanto, é bastante mais complexa, uma vez que, por exemplo, o valor estimado para $\tilde{\Omega}_{1,2}^*$, será a covariância da diferença entre os erros da 1^a e 3^a alternativas, relativamente à variância da diferença entre os erros da 1^a e 2^a alternativa (m). Esta interpretação é ainda mais difícil uma vez que a variância da diferença de erros entre duas alternativas reflete a variância de ambas e a sua covariância (Train, 2003).

2.4.6. Estimação dos Coeficientes dos Modelos

Assumindo que a escolha de cada indivíduo é independente das escolhas de todos os outros (Train, 2003), para a estimação dos valores dos pesos dos atributos/coeficientes (θ) é usual a utilização de *modelos de máxima verosimilhança* considerando, depois de definida uma função matemática específica para cada tipo de modelo $P(i)$, a seguinte expressão:

$$L(\theta) = \prod_{n=1}^N \prod_{i=1}^I P(i)^{\delta_{in}} \quad \text{Exp. III- 35}$$

Onde:

$L(\theta)$: Estimador da máxima verosimilhança, ou seja, os valores do vetor de θ que maximizam a função $P(i)$;

δ_{in} : é definido com o valor de 1 se a n -ésima escolha discreta observada for i e 0 caso contrário;

$P(i)$: Função de distribuição cumulativa para o modelo selecionado, considerando uma função utilidade com parâmetros lineares;

N : número total de observações;

I : número total de alternativas;

Uma vez que através das propriedades dos logaritmos a expressão III.35 pode ser transformada num somatório sem alteração dos seus resultados (Washington *et al.*, 2003; Train, 2003), a estimação dos parâmetros pode ser realizada através da seguinte expressão mais simples, designada *máxima verosimilhança logarítmica* - MVL):

$$LL(\theta) = \ln(L) = \ln \left[\prod_{n=1}^N \prod_{i=1}^I P(i)^{\delta_{in}} \right] = \sum_{n=1}^N \sum_{i=1}^I \delta_{in} \cdot \ln[P(i)] \quad \text{Exp. III- 36}$$

Train (2003) sugere o cálculo da função recorrendo ao seu valor médio, ou seja, dividindo o valor da expressão anterior por N (tamanho da amostra). Esta alteração não altera o local do máximo da expressão e facilita a interpretação de alguns dos procedimentos.

A determinação do vetor dos θ que maximiza a função $LL(\theta)$, que por ser uma função logarítmica de uma probabilidade que toma o valor entre 0 e 1 é sempre negativa, é relativamente simples para apenas um parâmetro com formulação linear ou quadrática (primeira derivada da função $LL(\theta)$ em relação a θ igualada a zero). No entanto, quando existe um vetor de parâmetros (θ) a determinar, ou a sua formulação não é quadrática, torna-se bastante mais complexa. Numericamente a solução passa por partir de um valor inicial θ_0 , e percorrer a função de verosimilhança até que não se consiga encontrar um valor superior. O problema fundamental será saber qual o incremento a utilizar, e Train (2003) sugere a seguinte metodologia:

- O gradiente da função de verosimilhança logarítmica numa iteração t (θ_t), ou seja, a sua primeira derivada (ou inclinação), será:

$$g_t = \left(\frac{\partial LL(\theta)}{\partial \theta} \right)_{\theta_t} \quad \text{Exp. III- 37}$$

Este dá indicação sobre que sentido a seguir de modo a atingir um valor maior na interação seguinte. Terá dimensão $K \times 1$, sendo K o número de parâmetros a estimar.

- A matriz hessiana da função é a matriz das segundas derivadas da função de verosimilhança logarítmica e será:

$$H_t = \left(\frac{\partial g_t}{\partial \theta'} \right)_{\theta_t} = \left(\frac{\partial^2 LL(\theta)}{\partial \theta \partial \theta'} \right)_{\theta_t} \quad \text{Exp. III- 38}$$

Que será uma matriz $K \times K$, e permitirá avaliar o tamanho da interação a realizar de modo a aproximar-se do máximo.

Dos algoritmos existentes para maximização de funções, Train (2003) recomenda o de *Newton-Raphson* (NR), que aproxima o melhor valor para a interação seguinte utilizando uma série de Taylor de segundo grau:

$$\begin{aligned} LL(\theta_{t+1}) &= LL(\theta_t) + (\theta_{t+1} - \theta_t)' \cdot g_t + \frac{1}{2} \cdot (\theta_{t+1} - \theta_t)' \cdot H_t \cdot (\theta_{t+1} - \theta_t) = \\ &= LL(\theta_t) + (\theta_{t+1} - \theta_t)' \cdot g_t + \frac{1}{2} \cdot (\theta_{t+1} - \theta_t)' \cdot H_t \cdot (\theta_{t+1} - \theta_t) \end{aligned} \quad \text{Exp. III- 39}$$

Se se pretende encontrar os valores θ_{t+1} que maximizam a aproximação a $LL(\theta_{t+1})$, então a sua derivada tem que ser igualada a zero:

$$\begin{aligned} \frac{\partial LL(\theta_{t+1})}{\partial \theta_{t+1}} &= 0 \Leftrightarrow 0 + (1-0) \cdot g_t + \frac{2}{2} \cdot (\theta_{t+1} - \theta_t)' \cdot H_t \cdot (1-0) = 0 \Leftrightarrow \\ &\Leftrightarrow g_t + (\theta_{t+1} - \theta_t)' \cdot H_t = 0 \Leftrightarrow -g_t = (\theta_{t+1} - \theta_t)' \cdot H_t \Leftrightarrow -\frac{g_t}{H_t} = (\theta_{t+1} - \theta_t) \Leftrightarrow \quad \text{Exp. III- 40} \\ &\Leftrightarrow \theta_{t+1} = \theta_t - \frac{g_t}{H_t} \end{aligned}$$

Implicando que o incremento para obter o novo valor de θ_{t+1} seja $(-g_t/H_t)$, ou seja, o valor negativo do produto do gradiente pela inversa da matriz hessiana. Como a hessiana representa o grau de curvatura, quanto maior for a curvatura (e, portanto, mais próximo do máximo), menor será o incremento devido à sua proporcionalidade inversa, e quanto menor a curvatura (mais afastado do máximo), maior o incremento. O sinal negativo da expressão deve-se ao facto de os valores da função serem todos negativos pelo que a sua hessiana é negativa e o gráfico é invertido (o máximo na realidade é equivalente ao mínimo de uma função concava). O gradiente positivo indica que o máximo está à frente do ponto em análise e, portanto, o incremento deve ser positivo (para a direita), e um gradiente negativo indica que o máximo está atrás do ponto em análise e o incremento deve ser negativo (movimentar para a esquerda). A figura III.1 permite uma melhor visualização destes conceitos.

No procedimento de NR há três aspetos relevantes a considerar: o caso de funções quadráticas; o tamanho do incremento; e a concavidade. Esta problemática encontra-se desenvolvida no Anexo 2 do Capítulo III.

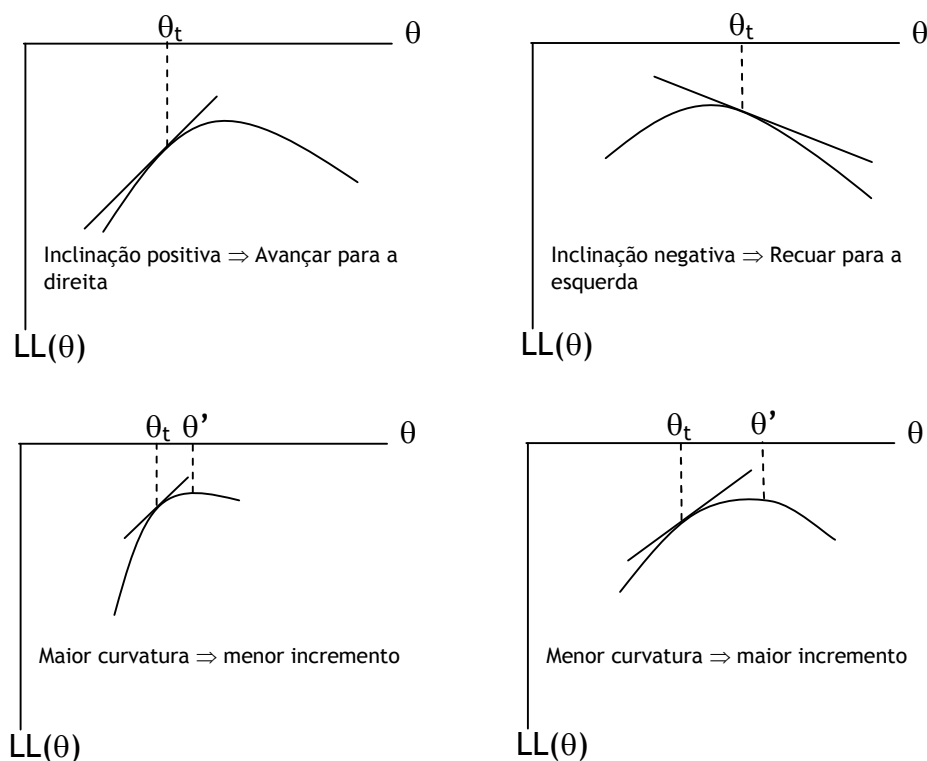


Figura III. 1 - Ilustração da influência da inclinação e da curvatura no processo incremental (traduzido de Train (2003))

2.4.7. Agregação de Dados e Transmissibilidade

Os modelos de escolha discreta são aplicados ao nível das decisões individuais, apesar de ser comum o analista necessitar de alguma agregação em determinadas características, tais como probabilidade média numa população ou a resposta média a alterações de um determinado atributo. Nos modelos de regressão linear estimativas de valores agregados da variável dependente são obtidos através da inserção no modelo de valores agregados das variáveis independentes ou explicativas. Os modelos de escolha discreta não são lineares em termos de variáveis explicativas, uma vez que a sua função cumulativa de probabilidades é uma curva sigmoide em S, pelo que a introdução de valores agregados das mesmas na modelação apresenta estimativas enviesadas das probabilidades médias ou das respostas médias (a função tem valores iniciais e finais com tendências muito diferentes dos valores intermédios comparativamente a uma função linear - figura III.2). A mesma situação aplica-se no cálculo das suas derivadas e elasticidades. Train (2003) refere estudos em que as elasticidades de utilidades médias apresentavam valores duas a três vezes superiores ou inferiores à média das elasticidades das utilidades individuais.

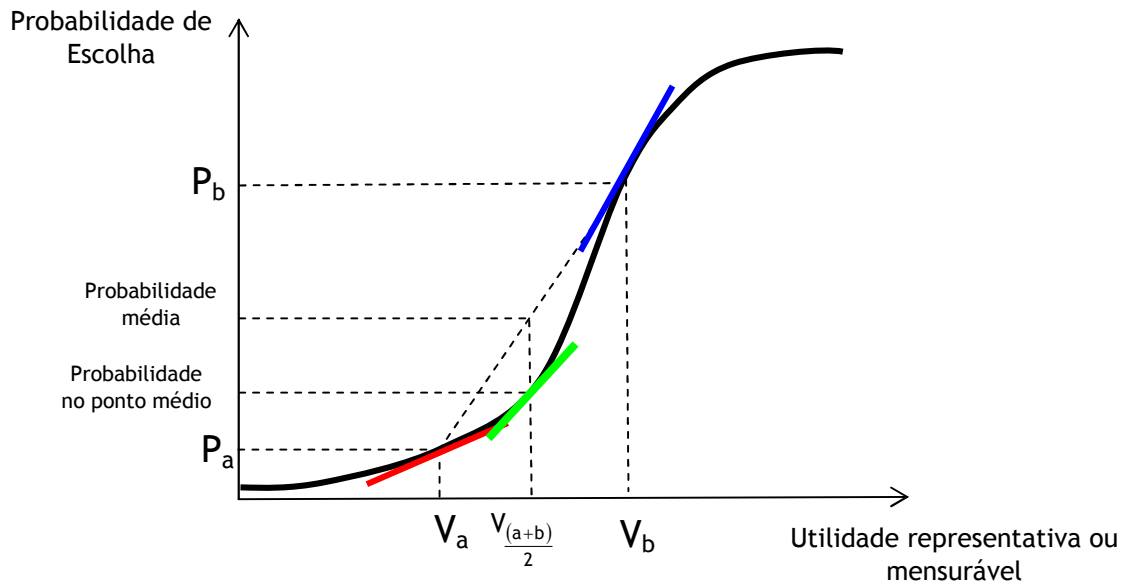


Figura III. 2 - Diferença entre a probabilidade média e a probabilidade calculada para o ponto médio da utilidade representativa ou mensurável (adaptado de Train (2003) e Washington *et al.* (2003))

Existe, portanto, a necessidade de tomar precauções na obtenção de resultados agregados através de modelos de escolha discreta, sugerindo Train (2003) duas metodologias: a enumeração da amostra; e a segmentação ou classificação segundo as variáveis independentes.

Numa primeira metodologia, a mais popular e direta, as probabilidades de escolha de cada indivíduo são somadas ou ponderadas considerando toda a população cujos resultados agregados estejam em análise. Considerando uma amostra de N indivíduos selecionada de uma população (pode ser a amostra utilizada para estimar o modelo ou outra de uma área ou data diferente), cada indivíduo da amostra tem um peso ω_n que representa o número de indivíduos com características semelhantes (por exemplo o rendimento) na população. Em amostras baseadas em fatores exógenos, variáveis independentes, estes pesos correspondem à probabilidade de cada indivíduo ser selecionado para a amostra. Se a amostra for aleatória, estes pesos são iguais para todos os n indivíduos selecionados, e se a amostra for segmentada e depois aleatória, o peso será igual para todos os n indivíduos desse segmento. A estimação do total de indivíduos da população que escolhem a alternativa i será a soma ponderada das probabilidades individuais:

$$\hat{N}_i = \sum_n \omega_n \cdot P_{ni}$$

Exp. III- 41

e a probabilidade média, ou a estimação da cota de mercado, será:

$$\hat{N}_i / N = W(i)$$

Exp. III- 42

Washington *et al.* (2003), no entanto, referem que este procedimento tem restrições na transferência de resultados para outras populações.

Na segunda metodologia, se o número de variáveis independentes ou explicatórias for reduzido, é possível estimar os resultados agregados utilizando uma amostra da população. Seleccionam-se as variáveis independentes que interessam e será a sua combinação que definirá os segmentos sobre os quais se determinará as probabilidades de escolha, e não as características dos indivíduos de cada segmento. Se o analista dispuser de dados sobre o número de indivíduos em cada segmento, o resultado agregado será estimado calculando a probabilidade de escolha dentro de cada segmento s , e realizando a soma ponderada dessas probabilidades:

$$\hat{N}_i = \sum_s \omega_s \cdot P_{si}$$

Exp. III- 43

onde P_{si} é a probabilidade de um indivíduo do segmento s escolher a alternativa i e ω_s é o número de indivíduos no segmento s . Apesar da sua facilidade de aplicação mediante a disponibilidade dos dados, Washington *et al.* (2003) referem que a assunção de homogeneidade da população no segmento pode ser dúbia introduzindo erros consideráveis. Ben-Akiva e Lerman (1985) referem que o ponto fundamental da classificação é a partição do universo de possíveis atributos, variáveis independentes, em subgrupos, uma vez que à medida que o número de subgrupos cresce, aumenta a precisão da estimação, mas também aumenta a necessidade de dados e de capacidade de processamento de informação. Este ponto levanta algumas questões que devem ser consideradas:

- Na classificação é mais importante realizar a divisão atendendo a grupos com pequenas variações na utilidade mensurável do que com pequenas variações nas próprias variáveis independentes;
- O número de variáveis a considerar para a definição da classificação deve ser reduzido para evitar um número inviável de classificações (por exemplo, uma utilidade com 16 variáveis independentes subdivididas em apenas 2 categorias daria $2^{16} = 65\,536$ classes diferentes), sugerindo-se o uso de uma ou duas variáveis consideradas importantes para realizar a classificação (que tenham um impacto significativo na utilidade de pelo menos um modo e com uma vasta distribuição pela população);
- Uma vez que o esforço de processamento é independente do tamanho da classe, sugere-se que estas sejam definidas evitando grupos desproporcionalmente pequenos.

Washington *et al.* (2003) e Ben-Akiva e Lerman (1985) referem ainda outras duas possíveis abordagens para evitar o enviesamento na agregação dos dados:

- A utilização de funções de densidade em integrais explícitos;
- E a definição de momentos de distribuição para diferenciação estatística (geralmente apenas os dois primeiros são considerados, sendo o 1º momento central a média e o 2º a variância).

A utilização de funções de densidade em integrais explícitos pretende representar a distribuição dos atributos na população através de uma distribuição analítica contínua conveniente. A repartição pode ser aproximada através de uma expressão do tipo:

$$W(i) \cong \int_x P_{ni} \cdot \tilde{p}(x) \cdot d(x) \quad \text{Exp. III- 44}$$

onde $\tilde{p}(x)$ é a distribuição aproximada dos atributos x na população. Como os valores de x podem ser contínuos e discretos há que ter em atenção que a resolução do integral pode necessitar de somas e integrais onde apropriado. Esta abordagem explora o facto de a expressão III.44 ser resolúvel para determinadas combinações de P_{ni} e $p(x)$ permitindo a obtenção da cota de mercado através de integração direta, se a aproximação a $p(x)$ for realizada através de uma distribuição conhecida. Ben-Akiva e Lerman (1985) referem duas combinações possíveis a utilizar, os modelos Probit e Logit, considerando os atributos normalmente distribuídos. Os problemas de aplicação desta metodologia advêm da consideração da distribuição normal dos atributos pela população, principalmente para os qualitativos e inteiros (como as variáveis binárias ou o número de automóveis) que podem ser irrealistas. Outro problema referido deve-se ao facto de a soma de produtos de variáveis aleatórias, essenciais para a resolução do integral (ver Ben-Akiva e Lerman (1985) nas pág. 142-145), não ter distribuição normal mesmo que os seus componentes a tenham. Washington *et al.* (2003) refere ainda que a utilização de funções de densidade apresenta a vantagem de permitir uma grande flexibilidade do modelo para aplicação a outras populações, apesar da sua difícil definição tanto teórica como empírica, e de a quantificação da covariância entre as várias variáveis independentes ser complicada.

A diferenciação estatística pretende aproximar o valor esperado de uma função de variáveis aleatórias a partir de informações sobre os momentos da sua distribuição conjunta. Os momentos e momentos cruzados são utilizados para representar a dispersão e a forma da distribuição das variáveis e a sua interação na população. As limitações desta abordagem incluem:

- A dificuldade de reunir informação teórica e empírica para justificar a representação dos momentos e momentos cruzados, podendo ser necessário ignorar alguns ou estimar coerentemente outros;
- Na integração a inclusão de termos corretivos não garante uma maior precisão deste método em relação a segmentação/classificação, uma vez que na metodologia se expande a função de probabilidade de escolha P_{ni} numa série de Taylor em que os termos superiores aos de segunda ordem (momentos superiores aos de segunda ordem) são ignorados. Mesmo a sua inclusão, no sentido de aumentar a sua precisão, não garante uma convergência mais célere do método (Ben-Akiva e Lerman, 1985).

Ben-Akiva e Lerman (1985) sintetizam o problema da agregação de escolhas individuais da seguinte forma:

“... A conclusão mais significativa deste capítulo é que as evidências empíricas sugerem que os erros devido a agregação entre indivíduos podem facilmente ser relativamente reduzidos...”

Refere ainda que as aproximações utilizadas nas várias abordagens apresentadas são fiáveis, resultando em pequenos erros, mesmo com condições iniciais bastante diferentes das condições assumidas.

Outra análise a realizar de modo a avaliar a utilidade do modelo estimado refere-se à sua transmissibilidade ou possibilidade de transferência dos parâmetros estimados espacial e temporalmente. Da perspetiva espacial, tem interesse que o modelo possa ser utilizado noutras cidades, regiões ou culturas, poupando esforços de novas aquisições de dados e de estimação. Da perspetiva temporal, assegura que os parâmetros estimados são estáveis no tempo e valida as previsões realizadas recorrendo aos mesmos. Pode mesmo assumir-se que a estabilidade de parâmetros estimados fornece indicação direta sobre a validade do modelo (Ortúzar e Willumsen, 2001). Os mesmos autores, no entanto, referem que é irrealista esperar que um modelo operacional fundamentado em ciências sociais seja perfeitamente especificado, uma vez que qualquer modelo estimado será em princípio dependente do contexto de estimação. Assim, sugere que não se procure um modelo perfeitamente estável e que não se avalie a transmissibilidade do modelo em termos de valores dos parâmetros estimados em cada contexto. Os autores sugerem antes que a abordagem de transferência de modelos seja encarada como estimações de um modelo inicial que providenciem informações úteis para a estimação do mesmo modelo noutro contexto, necessitando-se ainda assim de procedimentos de atualização dos parâmetros para que o modelo seja representativo.

Neste sentido, Ortúzar e Willumsen (2001), propõem que se faça uma análise comparativa dos parâmetros estimados independentemente nos contextos em análise (temporal ou espacial) e a medição dos erros relacionados com a utilização de um modelo no contexto do outro.

Train (2003) sugere o ajustamento dos parâmetros e/ou das variáveis independentes para refletirem as alterações previstas no futuro. Se o ajustamento for para o modelo agregado por enumeração, a amostra é ajustada de modo a que “pareça” que foi escolhida no futuro. Por exemplo, para prever o número de cidadãos que escolherá determinada alternativa dentro de cinco anos, a amostra atual é ajustada de modo a refletir as alterações socioeconómicas e de outros fatores espectáveis no período em estudo. Este ajustamento pode ser realizado alterando o valor das variáveis independentes (por exemplo o rendimento familiar) associadas a cada decisor selecionado, ou alterando o peso de cada decisor, de modo a refletir as alterações no tempo do número de indivíduos na população com características semelhantes. Se a agregação for por segmentação as alterações no futuro têm de ser refletidas no número de indivíduos em cada segmento e não nas variáveis independentes que as definem. A alteração das variáveis independentes de um indivíduo selecionado apenas lhe modificaria o segmento a que pertence.

Washington *et al.* (2003) sugerem a aplicação de teste de *razões de verosimilhança* para avaliar a qualidade da transmissibilidade temporal e espacial dos parâmetros estimados. Se se considerarem duas regiões ou períodos temporais *a* e *b*, a transmissibilidade dos parâmetros de um modelo poder ser avaliada através da seguinte expressão:

$$\chi^2 = -2 \cdot [LL(\theta_T) - LL(\theta_a) - LL(\theta_b)] \quad \text{Exp. III- 45}$$

onde:

- χ^2 : probabilidade dos modelos terem parâmetros diferentes. Esta variável tem distribuição chi-quadrada⁴ (χ^2) com graus de liberdade iguais à soma do número total de parâmetros estimados em todos os modelos (*a* e *b* ou inclusive outros) subtraído do número de parâmetros estimados no modelo total.
- $LL(\theta_T)$: é a máxima verosimilhança logarítmica estimada com dados de ambas as regiões ou períodos;
- $LL(\theta_a)$: é a máxima verosimilhança logarítmica estimada com os dados de *a*;
- $LL(\theta_b)$ é a máxima verosimilhança logarítmica estimada com os dados de *b*;

4 - Distribuição Chi-Quadrada (χ^2) (Montgomery e Runger, 2007): Uma variável aleatória contínua tem uma distribuição (χ^2) quando a sua função de densidade de probabilidades é igual a:

$$f(x) = \frac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{(k/2)-1} \cdot e^{-x/2}, \text{ com } x > 0$$

Onde:

k - número de graus de liberdade;

Γ - função gama $\Gamma(x) = (x-1)!$, com $x > 0$ e inteiro.

A distribuição (χ^2) tem média igual a *k* e variância igual a $2 \cdot k$, e a sua função de probabilidades é enviesada para a direita. Com o aumento de *k*, a distribuição torna-se mais simétrica e no caso extremo de $k \rightarrow \infty$ a função é igual a distribuição normal.

Neste teste os três modelos (total, região *a*, e região *b*) utilizam todos as mesmas variáveis.

Outra abordagem sugerida também por Washington *et al.* (2003) e Ortúzar e Willumsen (2001) utiliza o designado *teste estatístico de transmissibilidade* (TET) utilizando a expressão:

$$\chi^2 = -2 \cdot [LL_b(\theta_a) - LL(\theta_a)] \quad \text{Exp. III- 46}$$

onde:

- $LL_b(\theta_a)$: máxima verossimilhança logarítmica do modelo utilizando os parâmetros estimados com os dados de *b* nos dados da região *a*;
- $LL(\theta_a)$: máxima verossimilhança logarítmica do modelo utilizando os parâmetros estimados com os dados da região *a*.

O TET pode ser realizado de forma inversa utilizando $LL_a(\theta_b)$ e $LL(\theta_b)$. A função estatística tem distribuição chi-quadrada com graus de liberdade iguais ao número de parâmetros estimados em $LL_b(\theta_a)$ e permite avaliar a probabilidade dos modelos terem parâmetros diferentes. Ortúzar e Willumsen (2001) referem que o TET não é simétrico pelo que é aceitável e razoável que a transmissibilidade numa direção entre contextos não se verifique na direção inversa.

Segundo Washington *et al.* (2003) a combinação destes dois testes assegura uma boa avaliação da transmissibilidade dos parâmetros no tempo ou no espaço, mas Ortúzar e Willumsen (2001) referem ainda um *índice de transferência* (IT) que descreve o grau em que a máxima verossimilhança logarítmica do modelo transferido excede o modelo nulo ou de referência (utilizando, por exemplo, as cotas de mercado) em relação à melhoria conseguida com o modelo aplicado ao novo contexto. Este índice é definido como:

$$IT_b(\theta_a) = \frac{LL_b(\theta_a) - LL_b(\theta_R)}{LL(\theta_b) - LL_b(\theta_R)} \quad \text{Exp. III- 47}$$

Onde:

- $IT_b(\theta_a)$ - índice de transferência do modelo utilizando os parâmetros estimados para a região *b* com os dados da região *a*.
- $LL_b(\theta_a)$ - máxima verossimilhança logarítmica do modelo utilizando os parâmetros estimados com os dados de *b* nos dados da região *a*;
- $LL_b(\theta_R)$ - máxima verossimilhança logarítmica do modelo utilizando os parâmetros estimados com os dados de *b* nos dados da situação de referência ou nula;

$LL(\theta_b)$ - máxima verossimilhança logarítmica do modelo utilizando os parâmetros estimados com os dados da região b .

O IT tem limite superior igual à unidade (quando o modelo transferido tem a mesma precisão que o modelo local), mas não tem limite inferior, pelo que valores negativos implicam apenas que o modelo transferido é pior que o local. Assim, o IT permite avaliar a medida relativa de transmissibilidade do modelo, enquanto o TET permite uma medida estatística dessa transmissibilidade (Ortúzar e Willumsen, 2001).

Train (2003) ainda refere outra consideração sobre a transmissibilidade dos resultados dos modelos, referente às constantes específicas de cada alternativa. Estas são inseridas no modelo para capturar o efeito médio dos fatores não observados ou mensuráveis, mas a sua grandeza pode variar de local para local ou no tempo comparativamente à amostra analisada, pelo que há necessidade de as ajustar de modo a refletir essas alterações. Dados das cotas de mercado das zonas para onde se pretende realizar as previsões podem ser utilizadas para recalibrar as constantes, permitindo o uso do modelo na previsão de alterações das percentagens de indivíduos que escolhem uma alternativa devido a alterações das variáveis explicadoras ou independentes. Train (2003) e Ortúzar e Willumsen (2001) recomendam um processo iterativo utilizando a seguinte expressão:

$$k_j^1 = k_j^0 + \ln \left(\frac{S_j}{\hat{S}_j^0} \right) \quad \text{Exp. III- 48}$$

onde:

- k_j^1 : é a constante específica da alternativa j na iteração 1;
- k_j^0 : é a constante específica da alternativa j na iteração inicial/anterior;
- S_j : percentagem de indivíduos na área de previsão que escolhem a alternativa j no ano base (último com dados disponíveis);
- $\hat{S}_j^0$: percentagem de indivíduos na área de previsão que escolhem a alternativa j estimado pelo modelo de escolha discreta utilizando os valores de k_j^0 ;

Com as novas constantes devem estimar-se de novo as cotas de mercado (percentagem de cidadãos que escolhem determinada alternativa) e comparar com as do ano base, parando as iterações quando as previsões de aproximarem suficientemente das cotas conhecidas na área de previsão (S_j). O modelo com as constantes ajustadas pode ser utilizado para a previsão de alterações nas cotas de mercado em relação ao ano base devido a modificações nas variáveis independentes que afetam as escolhas dos indivíduos.

3. Modelos de Escolha Discreta mais Comuns

3.1. Logit Multinomial (ou Multimodal)

De longe o mais simples e difundido modelo de escolha discreta é o LOGIT simples. A sua popularidade advém da sua formulação para a probabilidade de escolha ser fixa e facilmente interpretável. A sua formulação é a seguinte:

$$P(i | C_n) = P_{ni} = \frac{e^{(\lambda V_{in})}}{\sum_{j \in C_n} e^{(\lambda V_{jn})}} \quad \text{Exp. III- 49}$$

Onde:

- $P(i | C_n) = P_{ni}$: é a probabilidade da opção i do conjunto de alternativas (C_n) ser escolhida pelo indivíduo n ;
- λ : parâmetro de escala relacionado com a variância⁵ σ^2 da distribuição dos erros ($\lambda = \pi / \sqrt{6} \cdot \sigma$), não identificável que pode assumir qualquer valor sem afetar o modelo, uma vez que apenas altera a escala da matriz de covariância e não o resultado. Geralmente é assumido como a unidade (Ben-Akiva e Lerman, 1985; Munizaga *et al.*, 2000; Ortúzar e Willumsen, 2001; Train, 2003);
- V_{in} e V_{jn} : Utilidade mensurável, ou representativa, das alternativas i e j , respetivamente, para o indivíduo n , especificada linearmente em relação aos parâmetros (a probabilidade de escolha, no entanto, não é linear mas uma curva em S).

3.1.1. Pressupostos Subjacentes à Formulação Logit

3.1.1.1. Independência entre Alternativas Irrelevantes - IIA

Os pressupostos subjacentes às características das probabilidades de escolha para este tipo de modelos derivam da *independência de alternativas irrelevantes* (“*independence of irrelevant alternatives*” - IIA). Este pressuposto, algo simplicista, pode aplicar-se a algumas situações, mas muitas situações reais não seguem esta independência. Exemplos destes problemas são os padrões de substituição referidos por Train (2003). Quando os atributos de uma alternativa são melhorados, por exemplo o seu custo diminui, a sua probabilidade de ser escolhida aumenta, levando alguns indivíduos que nas condições iniciais escolheriam outras opções, a

5 - Variância, é uma medida da variabilidade de uma distribuição, definida como o valor esperado do quadrado da variável aleatória em relação à sua média. É calculada através da expressão:

$$\sigma^2 = \sum_x (x - \mu)^2 \cdot f(x) = \sum_x x^2 \cdot f(x) - \mu^2$$

Onde μ é a média ou valor expectável de uma variável aleatória discreta X , calculada pela expressão:

$$\mu = \sum_x x \cdot f(x)$$

O desvio padrão σ , é a raiz quadrada da variância e é a medida de variância mais utilizada (Montgomery e Runger, 2007).

escolher a alternativa melhorada. Considerando que a soma das probabilidades é a unidade, um aumento da probabilidade de escolha de uma leva a uma diminuição da probabilidade de escolha das restantes. Este problema pode ser observado ou através de uma restrição nas razões entre probabilidades e/ou como uma restrição nas elasticidades cruzadas das probabilidades.

Se considerarmos a razão entre probabilidades, segundo o modelo LOGIT para duas alternativas i e j , a mesma é dada por:

$$\frac{P_{ni}}{P_{nj}} = \frac{\frac{e^{V_{ni}}}{\sum_k e^{V_{nk}}}}{\frac{e^{V_{nj}}}{\sum_k e^{V_{nk}}}} = \frac{e^{V_{ni}}}{e^{V_{nj}}} = e^{V_{ni} - V_{nj}} \quad \text{Exp. III- 50}$$

Assim, a razão não depende de outras alternativas além de i e j , ou seja, a probabilidade relativa de escolher i ou j é a mesma independentemente das outras alternativas ou dos seus atributos. Como a razão não depende de outras alternativas diz-se que são independentes de alternativas irrelevantes. Esta assunção é uma representação válida para algumas situações reais, mas, tal como já foi referido, há muitas para as quais não permite previsões corretas.

Para considerarmos o problema em termos de elasticidades cruzadas, é necessário definir elasticidade de uma variável dependente y em relação a uma variável independente x_i . A Elasticidade é dada pela expressão:

$$E(y, x_i) = \frac{\partial y}{\partial x_i} \cdot \frac{x_i}{y} \quad \text{Exp. III- 51}$$

A sua interpretação é a alteração percentual na variável dependente (y) atendendo a uma determinada variação percentual da variável independente (x_i). Na modelação de transportes, as variáveis independentes são os atributos das alternativas modais, e uma elasticidade direta refere-se aos atributos do modo em análise, enquanto uma elasticidade cruzada diz respeito a atributos entre modos alternativos. Como exemplo, uma elasticidade direta dos transportes coletivos em relação ao seu custo de -0,33, significa que um aumento de 1% no custo da viagem levaria a um decréscimo (devido ao sinal negativo) na sua utilização de 0,33%.

Atendendo a esta definição, Train (2003) apresenta as seguintes formulações para as elasticidades direta e cruzada utilizando o modelo LOGIT (com função utilidade linear):

- Elasticidade Direta

$$E(i, X_{ni}) = \frac{\partial P_{ni}}{\partial X_{ni}} \cdot \frac{X_{ni}}{P_{ni}} = \frac{\partial V_{ni}}{\partial X_{ni}} \cdot X_{ni} \cdot (1 - P_{ni}) = \beta_{xi} \cdot X_{ni} \cdot (1 - P_{ni}) \quad \text{Exp. III- 52}$$

- Elasticidade Cruzada:

$$E(i, X_{nj}) = \frac{\partial P_{ni}}{\partial X_{nj}} \cdot \frac{X_{nj}}{P_{ni}} = - \frac{\partial V_{nj}}{\partial X_{nj}} \cdot X_{nj} \cdot P_{nj} = -\beta_{xj} \cdot X_{nj} \cdot P_{nj} \quad \text{Exp. III- 53}$$

Onde:

- $E(i,X)$: Elasticidade da alternativa i em relação ao atributo X ;
- P_{ni} : Probabilidade do indivíduo n escolher a alternativa i ;
- V_{ni} : Parte mensurável, observável ou sistemática da utilidade da alternativa i para o indivíduo n ;
- β_{xj} : Coeficiente do atributo X para a alternativa j , constante para todos os indivíduos do “segmento”, mas que varia segundo o modo.

Assim, se se considerar uma alteração num atributo da alternativa j e se se pretender saber qual o impacto em todas as restantes alternativas (por exemplo i), através da sua elasticidade cruzada (expressão III.53) pode verificar-se que esta é igual para todos os i , pois este não entra na fórmula (Ben-Akiva e Lerman (1985) referem que os modelos Logit têm *elasticidades cruzadas uniformes*). Isto significa que a melhoria no atributo de uma alternativa reduz a probabilidade de escolher qualquer outra na mesma percentagem, isto é, a melhoria de uma alternativa reduz proporcionalmente as restantes. A mesma conclusão pode ser retirada em sentido contrário, a redução da utilidade de uma alternativa aumenta proporcionalmente a probabilidade de escolha das restantes. Este fenómeno, chamado *deslocamento proporcional* (Train, 2003) é outra manifestação da *I/A* (no Anexo 3 do Capítulo III pode verificar-se a definição de elasticidades agregadas segundo Ben-Akiva e Lerman (1985)). Ben-Akiva e Lerman (1985) referem que o pressuposto de *I/A* apenas tem que se verificar em grupos homogêneos da população e não em toda a população, permitindo a aplicação de modelos Logit a uma grande variedade de situações desde que devidamente especificados ao nível das variáveis socioeconómicas independentes, de forma a segmentar a população heterogênea que não verifica a *I/A* em segmentos homogêneos que a verifiquem. Assim, se o pressuposto da *I/A* refletir a realidade ou uma adequada aproximação, o seu emprego poderá trazer vantagens consideráveis. Estas vantagens são evidentes em situações em que o número de alternativas é bastante elevado, uma vez que a condição *I/A* permite estimar os parâmetros do modelo de forma consistente utilizando um subconjunto muito menor de alternativas. Dado que as probabilidades relativas de um subconjunto de alternativas não são afetadas

pelos atributos ou pela existência de outras alternativas não pertencentes ao subconjunto, a consistência dos estimadores também não é afetada pela exclusão de alternativas, permitindo uma poupança considerável nos tempos de processamento. Esta vantagem estende-se a situações em que apenas algumas alternativas interessam do conjunto total de opções, permitindo a estimação credível do modelo apenas para as alternativas mais interessantes (por exemplo, automóvel particular e autocarro).

Train (2003) sugere duas metodologias para testar se as alternativas são efetivamente *IIA*. Uma primeira abordagem sugere a estimação dos parâmetros para um subconjunto de alternativas, comparando-os com os estimados para a totalidade das alternativas. Se estas forem realmente *IIA*, os valores estimados para ambas as situações deverão ser semelhantes. Outra abordagem sugere a estimação dos parâmetros com novos atributos cruzados entre alternativas, isto é, com a utilização das variáveis de uma alternativa na expressão da utilidade de outra. Se a razão entre probabilidades para as alternativas *i* e *j* realmente depender dos atributos e da existência de uma alternativa *k*, e os atributos desta última influenciarem significativamente a utilidade das alternativas *i* e *j* especificadas através da formulação Logit, então existe uma violação da *IIA*. Train (2003) refere que tanto ele próprio como Daniel McFadden desenvolveram métodos para esta verificação, através de regressões ou incluídos na própria modelação Logit.

Assim, assumindo a hipótese de *IIA*, nos anos 60 provou-se (Train, 2003):

- que a mesma implicava que a formulação Logit é consistente com a maximização da utilidade;
- a existência da relação entre a fórmula Logit e a distribuição da utilidade não observada (em oposição às características da probabilidade de escolha);
- e demonstrou-se que a distribuição de erros de valores extremos resultava na fórmula Logit.

Já nos anos 70 completou-se a análise da formulação Logit, demonstrando que a mesma para probabilidades de escolha implicava necessariamente que a utilidade não observada tivesse uma distribuição de valores extremos. Este tipo de distribuição tem uma propriedade bastante útil: *os máximos de uma amostra escolhida aleatoriamente têm a mesma distribuição que os máximos da “população” total*; propriedade que uma distribuição normal não tem, uma vez que os seus máximos não têm uma distribuição normal (Washington et al, 2003). Estas distribuições de valores extremos podem ser do tipo I (*Gumbel*), tipo II (*Fréchet*) ou do tipo III (*Weibull*).

3.1.1.2. Erros Independentes e Igualmente Distribuídos - IID

O modelo Logit pode ser gerado assumindo-se que a utilidade não observada, ou seja, os erros ou resíduos aleatórios (ε) têm uma distribuição de valores extremos do tipo I, de *Gumbel*, ou mutuamente independentes e igualmente distribuídos - IID. Esta assunção, conveniente mas simplista, tem algumas implicações que influenciam a sua aplicabilidade. A distribuição IID dos erros implica que estes tenham distribuições independentes, identicamente distribuídas, um parâmetro de localização (moda) η e um parâmetro de escala positivo $\lambda > 0$.

O problema da diferença de variância de erros pode ser considerado de duas formas diferentes: entre alternativas e entre observações (Munizanga *et al.*, 2000). Um exemplo do primeiro tipo de problema de diferença na variância dos erros verifica-se num estudo de preferências declaradas onde se pretenda um escalonamento de várias opções por ordem de preferência. Não parece credível que para todas as possíveis respostas a distribuição dos erros seja semelhante. Se avaliações com apreciações muito baixas ou muito altas são relativamente precisas, avaliações com apreciação intermédia tendem a ser muito menos precisas esperando-se que as últimas tenham uma maior variância dos erros. O problema dos erros entre observações pode ser exemplificado quando se usam dados recolhidos de formas diferentes, através de preferências declaradas e de preferências reveladas por exemplo, não sendo razoável assumir que a variância dos erros dos dois tipos de observações seja idênticos.

Em relação ao parâmetro de localização η a assunção comum de considerá-lo constante ou nulo, como no Logit binário ou bimodal, não restringe a sua aplicabilidade desde que cada utilidade sistemática ou mensurável tenha um termo constante (Ben-Akiva e Lerman, 1985). A consideração de que os erros têm uma distribuição IID ou *Gumbel* pode ser defendida como uma aproximação a uma densidade normal, mas analiticamente mais conveniente. Outra restrição que a distribuição IID impõe advém de obrigar que todos os erros tenham o mesmo parâmetro de escala λ , o que implica que todos os erros tenham a mesma variância, o que em muitos casos pode não ser real ou ser difícil de provar.

Ben-Akiva e Lerman (1985) apresentam as características básicas das distribuições de valores extremos do tipo I ou *Gumbel*, se as funções de distribuição cumulativa $F(\varepsilon)$ e de densidade de probabilidades $f(\varepsilon)$ forem as seguintes:

$$F(\varepsilon) = e^{-e^{-\lambda(\varepsilon-\eta)}}; \lambda > 0 \quad \text{Exp. III- 54}$$

$$f(\varepsilon) = \lambda \cdot e^{-\lambda(\varepsilon-\eta)} \cdot e^{-e^{-\lambda(\varepsilon-\eta)}} \quad \text{Exp. III- 55}$$

Onde η é o fator de localização (abscissa onde se localiza o máximo) e λ o parâmetro de escala positivo.

As características básicas das distribuições de *Gumbel* são:

- A moda é η ;
- A média é $\eta + \gamma/\lambda$, onde γ é a constante de Euler $\approx 0,5772$;
- A variância $\sigma^2 = \pi^2/(6\lambda^2)$;
- Se ε tem distribuição *Gumbel* com parâmetros (η, λ) e, V e $a > 0$ são qualquer constante escalar, então $a \cdot \varepsilon + V$ tem distribuição *Gumbel* com parâmetros $(a\eta + V, \lambda/a)$;
- Se ε_1 e ε_2 são variáveis independentes com distribuição de *Gumbel* com parâmetros (η_1, λ) e (η_2, λ) , então $\varepsilon^* = \varepsilon_1 - \varepsilon_2$ tem uma distribuição logística:

$$F(\varepsilon^*) = \frac{1}{1 + e^{\lambda(\eta_1 - \eta_2 - \varepsilon^*)}} \quad \text{Exp. III- 56}$$

- Se ε_1 e ε_2 são variáveis independentes com distribuição de *Gumbel* com parâmetros (η_1, λ) e (η_2, λ) , então o $\max(\varepsilon_1, \varepsilon_2)$ tem distribuição de *Gumbel* com parâmetros $\left(\frac{1}{\lambda} \ln(e^{\lambda\eta_1} + e^{\lambda\eta_2}), \lambda \right)$;
- Como consequência da preposição anterior se $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J)$ são J variáveis aleatórias independentes com distribuição de *Gumbel*, com parâmetros respectivamente de (η_1, λ) , (η_2, λ) , ..., (η_J, λ) , então o $\max(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J)$, tem distribuição de *Gumbel* com os parâmetros:

$$\left(\frac{1}{\lambda} \ln \sum_{j=1}^J e^{\lambda\eta_j}, \lambda \right) \quad \text{Exp. III- 57}$$

As primeiras três características especificam a moda, a média e a variância da distribuição. A quarta afirma que a distribuição de *Gumbel* se preserva através de transformações lineares. A quinta descreve a distribuição da diferença entre duas variáveis independentes com distribuição de *Gumbel* com o mesmo parâmetro de escala λ . As duas últimas propriedades especificam que o máximo de variáveis independentes com distribuição de *Gumbel* com parâmetros de escala comuns tem também distribuição de *Gumbel*. O parâmetro de localização λ do máximo $\left(\frac{1}{\lambda} \ln \sum_{j=1}^J e^{\lambda\eta_j} \right)$, é igual ao valor esperado do máximo subtraído de uma constante (γ/λ) .

As figuras III.3 e III.4 seguintes apresentam graficamente os resultados da variação dos parâmetros η e λ das funções de distribuição de probabilidades e da densidade cumulativa, permitindo uma melhor interpretação do seu significado.

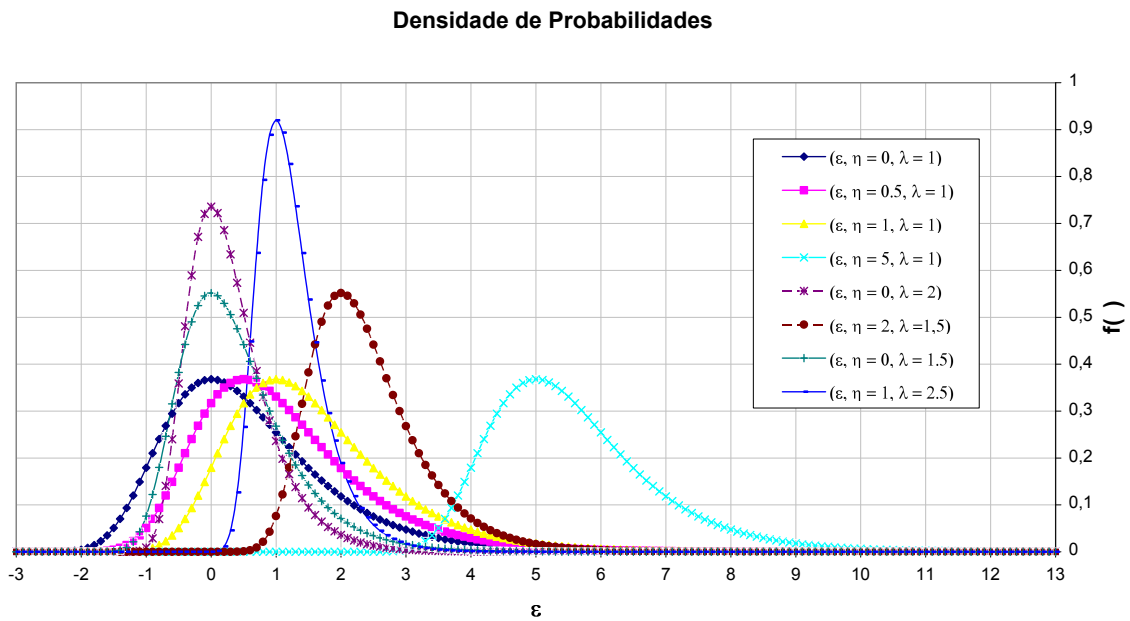


Figura III. 3 - Representação gráfica das densidades de probabilidades da expressão III.55 com distribuição de valores extremos do tipo I ou de *Gumbel* para diferentes valores dos parâmetros de localização e de escala

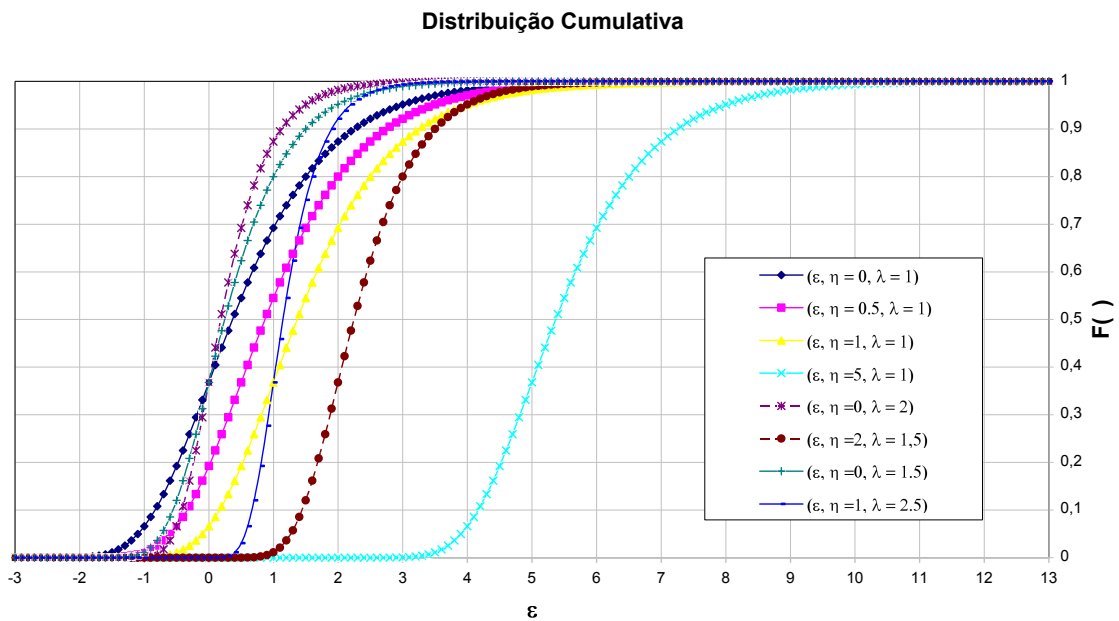


Figura III. 4 - Representação gráfica da distribuição cumulativa da expressão III.54 com distribuição de valores extremos do tipo I ou de *Gumbel* para diferentes valores dos parâmetros de localização e de escala

Ao assumir que a variância é $\pi^2/6$ implicitamente está-se a normalizar a escala da utilidade e numa distribuição de valores extremos, apesar de a sua média não ser nula, a diferença entre os dois termos aleatórios com a mesma média é, esta sim, nula (note-se que apenas a diferença entre utilidades interessa). A utilização de distribuições de valores extremos para

os erros (e portanto uma distribuição logística para a diferença entre erros) é similar a assumir que os erros são normalmente independentes. A distribuição de valores extremos assume caudas ligeiramente maiores que a distribuição normal, indicando que permite comportamentos ligeiramente mais anormais que a distribuição normal, mas em geral pode assumir-se que a diferença entre erros de valores extremos e normalmente independentes é empiricamente insignificante (Train, 2003).

A normalização da escala da utilidade considerando que os fatores não observáveis têm uma distribuição de valores extremos com variância $\pi^2/6$ implica algumas alterações na modelação apresentada. Considere-se que a utilidade pode ser expressa pela seguinte expressão:

$$U_{nj}^* = V_{nj} + \varepsilon_{nj}^* \quad \text{Exp. III- 58}$$

com:

- U_{nj}^* : A real utilidade da alternativa j para o indivíduo n ;
- V_{nj} : Utilidade representativa ou mensurável da alternativa j para o indivíduo n ;
- ε_{nj}^* : Utilidade não mensurável da alternativa j para o indivíduo n .

A utilidade não mensurável tem variância igual a $\sigma^2 \times (\pi^2/6)$, ou seja, é qualquer número expresso como um múltiplo de $\pi^2/6$. Como a escala da utilidade é irrelevante para o comportamento, esta pode ser dividida por σ sem afetar as escolhas, pelo que a utilidade pode ser expressa como:

$$U_{nj} = \frac{V_{nj}}{\sigma} + \varepsilon_{nj}; \text{ onde } \varepsilon_{nj} = \frac{\varepsilon_{nj}^*}{\sigma} \quad \text{Exp. III- 59}$$

Assim, a parte não observada da utilidade passa a ter uma variância $\pi^2/6$:

$$\text{Var}(\varepsilon_{nj}) = \text{Var}\left(\frac{\varepsilon_{nj}^*}{\sigma}\right) = \left(\frac{1}{\sigma^2}\right) \cdot \text{Var}(\varepsilon_{nj}^*) = \left(\frac{1}{\sigma^2}\right) \cdot \sigma^2 \cdot \left(\frac{\pi^2}{6}\right) = \left(\frac{\pi^2}{6}\right) \quad \text{Exp. III- 60}$$

E a probabilidade de escolha será:

$$P_{ni} = \frac{e^{\frac{V_{ni}}{\sigma}}}{\sum_j e^{\frac{V_{nj}}{\sigma}}} \quad \text{Exp. III- 61}$$

Se a utilidade representativa for linear em relação aos parâmetros e tiver coeficientes B^* , a expressão anterior toma a forma:

$$P_{ni} = \frac{e^{\left(\frac{B^*}{\sigma}\right)x_{ni}}}{\sum_j e^{\left(\frac{B^*}{\sigma}\right)x_{nj}}} \quad \text{Exp. III- 62}$$

Pela expressão pode verificar-se que apenas se pode estimar a razão $B^*/\sigma = B'$, não sendo os valores de B^* e σ separadamente identificados, levando à forma generalizada do modelo Logit na sua forma escalonada:

$$P_{ni} = \frac{e^{B' \cdot x_{ni}}}{\sum_j e^{B' \cdot x_{nj}}} \quad \text{Exp. III- 63}$$

Os parâmetros B' são estimados mas na sua interpretação deve atender-se que são uma estimação dos coeficientes B originais divididos pelo parâmetro de escala, ou seja, que os mesmos estimam o efeito de cada variável independente relativamente à variância dos fatores não mensuráveis. Quanto maior a variância dos fatores não mensuráveis/observáveis, menores serão os coeficientes, mesmo que as variáveis independentes tenham o mesmo efeito na utilidade (Train, 2003). Variáveis como a disponibilidade para pagar, o valor do tempo ou outras medidas de taxas de substituição marginais não são afetadas pelo fator de escala, apenas a interpretação das suas magnitudes.

Assumir a independência entre erros, equivale a considerar que a parte não observada da utilidade de uma alternativa é independente da parte não mensurável de outra alternativa, facto que mais uma vez limita a aplicabilidade deste tipo de modelos.

Apesar destas limitações, a assunção de independência entre alternativas pode ser interpretado como um resultado de um modelo bem especificado. Considerando que o termo aleatório referente aos erros é definido como a diferença entre a representação da utilidade medida pelo analista através das variáveis independentes (V_{ni}) e a real utilidade da mesma (U_{ni}), a sua definição e distribuição dependem da especificação da utilidade representativa por parte do analista e não da situação de escolha em si. Assim, se independentes, a avaliação do erro de uma alternativa não fornece ao analista informações sobre os erros das restantes, pelo que se a especificação da utilidade mensurável for suficiente, os erros podem ser considerados como “ruído de fundo”. O objetivo ideal da modelação seria a especificação da utilidade mensurável de forma tão exaustiva que os elementos remanescentes fossem insignificantes, tornando o modelo Logit ideal.

Train (2003) sugere que se na análise se considerar que a parte não observada na especificação da utilidade mensurável é correlacionada entre alternativas, o analista tem três opções:

- Utilizar um outro modelo mais sofisticado que permita a correlação entre erros;
- Reformular a utilidade mensurável de forma a incorporar as fontes de correlação e tornar os erros independentes;
- Ou utilizar o modelo Logit assumindo as condições existentes para a definição da utilidade mensurável ou representativa e considerar os seus resultados como uma aproximação aos valores reais. Esta opção está diretamente relacionada com o objetivo da modelação, assumindo-se que a violação dos pressupostos subjacentes aos modelos Logit (*I/A* e *I/D*) tem maiores consequências na previsão de influências entre alternativas do que na estimação de preferências médias (Train, 2003).

3.1.2. Vantagens e Desvantagens dos Logit Simples

O modelo Logit para a determinação da probabilidade de escolha apresenta diversas propriedades desejáveis:

- A probabilidade P_{ni} está necessariamente entre 0 e 1;
- Quando uma utilidade representativa de uma alternativa aumenta (devido a melhoria de algum atributo mensurável), mantendo as restantes constantes, a probabilidade de escolha também aumenta aproximando-se de 1. Quando a sua utilidade representativa se aproxima de $-\infty$, mantendo as restantes constantes, a sua probabilidade de escolha diminui aproximando-se de 0;
- A probabilidade de escolha nunca é igual a 0, mas se se considerar que uma alternativa não tem hipóteses de ser escolhida, então pode ser retirada do conjunto de opções;
- A probabilidade de escolha igual a 1 só pode ser obtida se só houver uma única alternativa;
- O facto de o denominador ser a soma das utilidades representativas de todas as alternativas, assegura que o resultado da soma de todas as probabilidades seja a unidade, facilitando a interpretação dos seus resultados.

A forma sigmoide da função de distribuição cumulativa de probabilidade de escolha tem ainda as seguintes implicações:

- Se a utilidade representativa for muito reduzida comparativamente à das restantes, um pequeno acréscimo na utilidade dessa alternativa tem pouco impacto na sua probabilidade de escolha. As restantes ainda são suficientemente melhores para não serem afetadas. Do mesmo modo, numa alternativa com uma utilidade muito superior às restantes um acréscimo na sua utilidade também não afeta significativamente a sua probabilidade de escolha sobre as restantes (partes superior e inferior das curvas da figura III.2).

- O ponto onde um acréscimo da utilidade representativa tem maior impacto na probabilidade da sua escolha sobre as restantes será perto do ponto com probabilidade de 0,5, que significa uma probabilidade de 50% da alternativa ser escolhida. Nestes casos, uma pequena alteração da utilidade representativa pode provocar grandes alterações na sua probabilidade de escolha.

Estas características das formas sigmóides são extensíveis aos restantes modelos de escolha discreta que as utilizam para definir as probabilidades de escolha.

Train (2003) apresenta ainda três situações que devem ser consideradas ao ponderar a utilização dos modelos Logit:

- Os Logit conseguem representar variações de gostos sistémicas (relacionadas com as características observáveis dos decisores, como o rendimento, o tamanho do agregado, entre outros), mas não variações aleatórias (que não podem ser atribuídas às variáveis mensuráveis, como os gostos pessoais, antecedentes, entre outros);
- O modelo Logit implica a substituição proporcional entre alternativas (o aumento da probabilidade de uma implica um decréscimo proporcional das restantes) atendendo a especificação da utilidade representativa. Outras formas de substituição necessitam de outro tipo de modelo;
- Se os fatores não observados forem independentes em situações repetidas ao longo do tempo, o modelo Logit consegue apreender a dinâmica de escolhas repetidas, no entanto se estes estiverem correlacionados ao longo do tempo outro modelo terá que ser utilizado.

Ortúzar e Willumsen (2001) referem que o facto de o modelo considerar a independência entre alternativas irrelevantes, foi encarado inicialmente como uma vantagem destes modelos, uma vez que permitia tratar de forma simples a introdução de uma nova alternativa inexistente na altura da calibração do modelo (se os seus atributos fossem conhecidos), sendo atualmente encarado como uma desvantagem que invalida o modelo na presença de alternativas correlacionadas (célebre exemplo da linha de autocarros vermelha e azul).

3.1.3. Erros de Especificação

Existem algumas suposições restritivas que se consideram na formulação dos modelos Logit, que, quando não respeitados, podem resultar em erros de especificação potencialmente graves (Washington *et al.*, 2003). Estes erros podem ser agravados pela inclusão de variáveis irrelevantes ou pela omissão de fatores chave na estimação estatística dos seus coeficientes.

De longe a limitação mais frequente e mais mal entendida é a propriedade de independência entre alternativas irrelevantes. Os problemas surgem principalmente quando apenas algumas

das utilidades partilham os fatores não observados, implicando que na determinação da diferença entre utilidades, os fatores não observados não se anulem.

Outro possível erro de especificação tem a ver com a omissão de variáveis independentes relevantes que pode resultar em estimativas inconsistentes dos parâmetros do modelo e consequentemente das probabilidades de escolha previstas pelo mesmo. Esta omissão é principalmente penalizante se: as variáveis preteridas estão correlacionadas com outras incluídas no modelo; os valores médios das variáveis omitidas variarem nos resultados das alternativas e se as suas constantes específicas não forem incluídas no modelo; ou as variáveis em falta estão correlacionadas com os resultados das alternativas ou têm variâncias diferentes para resultados distintos. Como a probabilidade de alguma destas situações acontecer é elevada, a omissão de variáveis relevantes é um problema bastante grave.

De modo semelhante, a presença de variáveis irrelevantes também constitui um problema, embora de grandeza incomparavelmente inferior. As estimativas dos parâmetros e das probabilidades de escolha mantêm a consistência mesmo na presença de parâmetros irrelevantes, mas os erros da estimação de parâmetros aumentam, diminuindo a eficiência da modelação.

A variação dos parâmetros aleatórios também pode constituir um erro de especificação, uma vez que na formulação Logit se assume que os parâmetros estimados são iguais para todas as observações. Violações a esta assunção podem ocorrer se por algum motivo se acreditar que os parâmetros variam ao longo das observações de um modo não considerado no modelo, como por exemplo a importância do custo da viagem, que será mais importante para os agregados de baixos rendimentos que para os de elevados rendimentos. Uma forma de tentar ultrapassar este tipo de problema pode ser a utilização de uma variável combinada, no exemplo o custo dividido pelo rendimento do agregado, mas parâmetros completamente aleatórios como o efeito da paciência associada ao tempo de viagem, em que não há dados para justificarem a variação, levam a uma estimação inconsistente dos parâmetros e das probabilidades de escolha, e à necessidade de modelos mais robustos.

Outro problema a considerar é a correlação entre as variáveis independentes e variáveis não mensuráveis, que pode tornar a estimação dos parâmetros inconsistente. Washington *et al.* (2003) dão o exemplo da modelação utilizando a distância como variável independente e o conforto como variável não mensurável, que num cenário com transportes coletivos em que o operador, de forma a melhorar o serviço, opta por colocar os veículos mais confortáveis nos percursos mais longos, estão efetivamente correlacionados. De forma similar deve considerar-se a existência de variáveis endógenas, isto é, internas aos sistemas, de forma a não incluir variáveis que aparentemente estão correlacionados mas que na realidade fazem parte do sistema. Um exemplo é a modelação discreta da gravidade de acidentes, incluindo nas variáveis independentes a sinalização de perigo de acidente. Como é nos locais com maior

número de acidentes que esta sinalização está localizada, a sua presença está correlacionada com a gravidade dos acidentes, e portanto é endógena.

Um potencial problema de estimação pode também ocorrer na modelação discreta se a informação de escolhas anteriores for utilizada na estimação da probabilidade de escolha atual. Conceptualmente este procedimento pode ter algum sentido uma vez que se pode considerar que se tenta capturar comportamentos habituais, designados *dependências de estado* (Washington *et al.*, 2003). No entanto, o uso de procedimentos habituais também pode captar heterogeneidades residuais que podem sugerir falsos hábitos. Como exemplo considere-se um modelo onde o termo aleatório inclui características não observadas, e portanto heterogeneidade não observada, geralmente presentes nos condutores ao selecionarem um trajeto. Se uma variável que indique qual a escolha em situações anteriores for incluída no modelo, não há certeza se o parâmetro associado a esta variável revela verdadeiramente os seus hábitos ou se traduz um valor médio no termo aleatório para condutores. Este assunto é de extrema importância uma vez que a presença ou ausência de comportamentos habituais pode levar a conclusões comportamentais bastante diferentes. Isolar dependências de estado efetivas de heterogeneidades não observadas não é fácil, pelo que a interpretação de parâmetros calibrados com base em dados de escolhas anteriores deve ser realizada com extremo cuidado.

3.1.4. Influência da Amostragem na Estimação dos Modelos

Os métodos de estimação dos parâmetros dos modelos Logit necessitam ser adaptados à estratégia de amostragem utilizada para gerar as observações. A metodologia mais utilizada é a da Máxima Verosimilhança e a sua especificação para amostras exógenas e amostras baseadas em escolha é apresentada nos pontos seguintes.

3.1.4.1. Amostras Exógenas

Neste tipo de amostras, quer sejam aleatórias simples quer aleatórias estratificadas atendendo a fatores exógenos à escolha analisada (se forem estratificadas atendendo a sua escolha a estimação é mais complexa), considera-se que todas as variáveis independentes são exógenas à situação de escolha, ou seja, as variáveis que entram na utilidade representativa são independentes da componente não mensurável da utilidade.

Na posse de uma amostra de N decisores a utilizar para a estimação, o procedimento da máxima verosimilhança pode ser aplicado aos modelos Logit devido à sua formulação bem definida considerando que a probabilidade de um indivíduo n escolher a alternativa que de facto se observou que escolheu pode ser representada por:

$$\prod_i (P_{ni})^{\delta_{ni}}$$

Exp. III- 64

Onde $\delta_{ni} = 1$ se a pessoa n escolheu a alternativa i pertencente ao conjunto de alternativas à sua disposição (C_n) e zero caso contrário. Como $\delta_{ni} = 0$ no caso de alternativas não escolhidas, P_{ni} elevado a zero é 1 e portanto a expressão apenas dá a probabilidade de escolha da alternativa i . Se a escolha de cada indivíduo for independente das escolhas dos restantes a probabilidade de cada indivíduo da amostra escolher a alternativa que efetivamente se verificou que escolheu é representada pela expressão III.64, e a sua expressão “logaritmizada” é mais fácil de utilizar.

A função resultante da expressão da máxima verosimilhança logarítmica é em geral côncava para utilidades com parâmetros lineares, pelo que o vetor de parâmetros θ das variáveis independentes que maximizam a função é o que anula a condição de primeira ordem, ou seja, a sua primeira derivada:

$$\frac{\partial LL(\theta)}{\partial \theta} = 0 \quad \text{Exp. III- 65}$$

Considerando a expressão da máxima verosimilhança logarítmica aplicada à fórmula de probabilidade de escolha Logit, a sua primeira derivada é (Ben-Akiva e Lerman, 1985; Train, 2003):

$$\frac{\partial LL(\theta)}{\partial \theta} = \sum_n \sum_i (\delta_{ni} - P_{ni}) \cdot x_{ni} = 0 \quad \text{Exp. III- 66}$$

Ou seja:

$$\sum_n \sum_i \delta_{ni} \cdot x_{ni} = \sum_n \sum_i P_{ni} \cdot x_{ni} \Leftrightarrow \frac{1}{N} \cdot \sum_n \sum_i \delta_{ni} \cdot x_{ni} = \frac{1}{N} \cdot \sum_n \sum_i P_{ni} \cdot x_{ni} \quad \text{Exp. III- 67}$$

Se se considerar que a parte esquerda da expressão representa a média dos parâmetros x nas escolhas observadas dos indivíduos na amostra e a parte direita a média dos parâmetros x das escolhas estimadas para os indivíduos da amostra, então a interpretação da igualdade anterior será que as estimativas dos parâmetros θ pela máxima verosimilhança são os que igualam as médias anteriores. Assim, as estimativas dos parâmetros através da máxima verosimilhança são as que igualam as médias previstas para cada variável independente às médias observadas das mesmas (Train, 2003). Esta propriedade é bastante importante principalmente nas constantes específicas de cada alternativa pois garante (uma vez que na função utilidade assumem o valor 1 quando se refere à alternativa i em análise e 0 para as restantes) através da expressão anterior que a repartição de indivíduos na amostra que escolhem a alternativa i é igual a repartição estimada dos indivíduos que escolhem i . Considerando que a parte esquerda da expressão é a diferença entre a real escolha do indivíduo δ_{ni} e a probabilidade dessa escolha P_{ni} , ou seja, o erro de modelação ou resíduo

(covariância da amostra entre os resíduos e as variáveis independentes), pode verificar-se que a estimativa através da máxima verosimilhança fornece os valores dos parâmetros que anulam essa covariância, ou tornam os resíduos não correlacionados com as variáveis independentes.

Ben-Akiva e Lerman (1985) na página 20 resumem a sua interpretação do seguinte modo:

“... um estimador de máxima verosimilhança é o valor do parâmetro mais provável de ter ocorrido na amostra observada...”

Outro aspeto importante dos θ estimados (máximos da $LL(\theta)$) é que a sua distribuição é similar à distribuição Normal com média θ e matriz de covariância amostral S^2 .

3.1.4.2. Amostras Baseadas em Escolhas

Em algumas situações uma amostra com base em fatores exógenos pode incluir um número bastante reduzido de indivíduos que escolham uma determinada alternativa, por exemplo bicicleta. Se houver um grande interesse em saber as motivações de quem utiliza a bicicleta, uma amostra aleatória teria de ser de grande dimensão para assegurar um número aceitável de ciclistas. Nestas situações o analista pode selecionar a amostra ou parte dela com base na escolha a analisar. Este procedimento pode servir de complemento a uma amostra aleatória simples.

Este tipo de amostragem pode ser puramente baseada em escolhas ou conciliar a amostra baseada em escolhas com amostras atendendo a variáveis exógenas. Nas puramente baseadas em escolhas a população é dividida mediante a alternativa escolhida e os indivíduos são escolhidos aleatoriamente em cada segmento, portanto com razões diferentes. Numa amostragem híbrida combina-se, por exemplo, uma amostragem exógena complementada com uma amostra baseada na escolha do agregado familiar.

A estimação de parâmetros com amostras selecionadas, pelo menos parcialmente, com base nas escolhas dos indivíduos é complexa e segundo Train (2003) varia com a forma exata do procedimento de amostragem. No entanto, se a amostragem for puramente baseada em escolhas e incluir uma constante específica da alternativa na utilidade representativa para cada alternativa, então a estimação do modelo Logit como se a amostra fosse exógena produz estimações consistentes para todos os parâmetros exceto para as constantes específicas. No entanto, estas constantes são enviesadas num valor conhecido que permite o seu ajustamento de forma consistente (Train, 2003). Se se souber qual a repartição na população que escolhe a alternativa j (A_j) e a repartição da amostra baseada em escolhas que também opta pela alternativa j (S_j), então uma estimativa consistente da constante específica ajustada da alternativa j (α'_j) atendendo á constante estimada (α^*_j) será:

$$\alpha'_j = \alpha * _j + \ln\left(\frac{A_j}{S_j}\right)$$

Exp. III- 68

3.1.5. Precisão da Estimação Estatística dos Parâmetros do Modelo Logit

Para uma aferição da precisão da estimação de um modelo os primeiros resultados a avaliar devem ser os sinais e as grandezas relativas dos coeficientes estimados. No entanto, na modelação pretende-se ainda conhecer a contribuição da cada variável na explicação da escolha modal. Assim, uma variável acrescenta informação ao modelo de escolha se o seu parâmetro for estatisticamente significativo, ou seja, se a hipótese nula de que o parâmetro é igual a zero, poder ser rejeitada. Num modelo de regressão linear essa significância estatística é avaliada através dos testes estatísticos de *t* e *F*. No entanto, nos modelos de escolha discreta estes não podem ser utilizados uma vez que apesar de lineares em termos de parâmetros, a sua estimação recorre a probabilidades que incluem logaritmos. Em vez destes, utiliza-se um teste assintótico equivalente designado de *teste de Wald* (Hensher *et al.*, 2005) que relaciona o valor do parâmetro *i* com o seu erro padrão:

$$\text{Wald} = \theta_i / \text{erro padrão}_i$$

Exp. III- 69

Estes valores podem ser calculados e interpretados da mesma forma que os teste estatísticos de *t* nas regressões lineares, através dos seus *p-valores*. Assim, assumindo um nível de confiança de 95% e um erro de 5% (geralmente assumidos como de referência), valores absolutos do teste de *Wald* inferiores a 1,96 sugerem que o parâmetro não é estatisticamente significativo, não se podendo rejeitar a hipótese nula de que θ_i é igual a zero (Hensher *et al.*, 2005).

3.1.6. Ajustamento Global e Teste de Hipóteses

Na estimação da melhor especificação de um modelo convém dispor de uma medida para o grau de ajustamento global do modelo estimado aos dados observados. Para medir o quanto um modelo de escolha discreta se ajusta aos dados é frequente utilizar um parâmetro estatístico designado por *índice de razão de verosimilhança*, também designado de *pseudo- R^2* . Este índice compara o desempenho do modelo com os parâmetros estimados com o modelo com todos os parâmetros nulos (que equivale a não ter modelo, ou seja que todos os modos têm igual repartição modal) ou apenas com as constantes específicas das alternativas (que equivale ao modelo com as repartições de mercado da amostra utilizada). A primeira situação é mais comum, mas Hensher *et al.* (2005) referem que ambas as comparações são comuns na literatura internacional. Esta comparação é feita através da função de máxima verosimilhança logarítmica (FMVL ou LL), podendo o índice ser definido como:

$$\rho^2 = p - R^2 = 1 - \left(\frac{LL(\theta)}{LL(0)} \right)$$

Exp. III- 70

Onde:

- $LL(\theta)$: É o valor da FMVL com os parâmetros estimados;
- $LL(0)$: É o valor da FMVL com os parâmetros todos nulos.

Se os parâmetros estimados em termos de FMVL não resultarem num valor melhor que o modelo com parâmetros nulos, os dois valores são iguais e a sua razão é 1, pelo que o índice toma o valor 0, ou seja, o seu valor mais baixo (se fosse possível $LL(\theta) < LL(0)$, então $LL(\theta)$ não seria o máximo).

No outro extremo, se o modelo fosse tão bom que permitisse prever perfeitamente as escolhas de cada indivíduo da amostra, a FMVL para os parâmetros estimados seria 1, uma vez que a probabilidade de observar as escolhas realmente feitas também seria 1. Como o logaritmo de 1 é 0, a FMVL seria 0 e o índice daria o valor 1, que será o seu maior valor possível. Assim o ρ^2 ($p-R^2$) é caracterizado por ter valores entre 0 e 1. No entanto, a sua interpretação não é tão simples como o coeficiente de correlação (R^2) utilizado nas regressões, que indica a percentagem da variação da variável dependente que é explicada pelo modelo estimado.

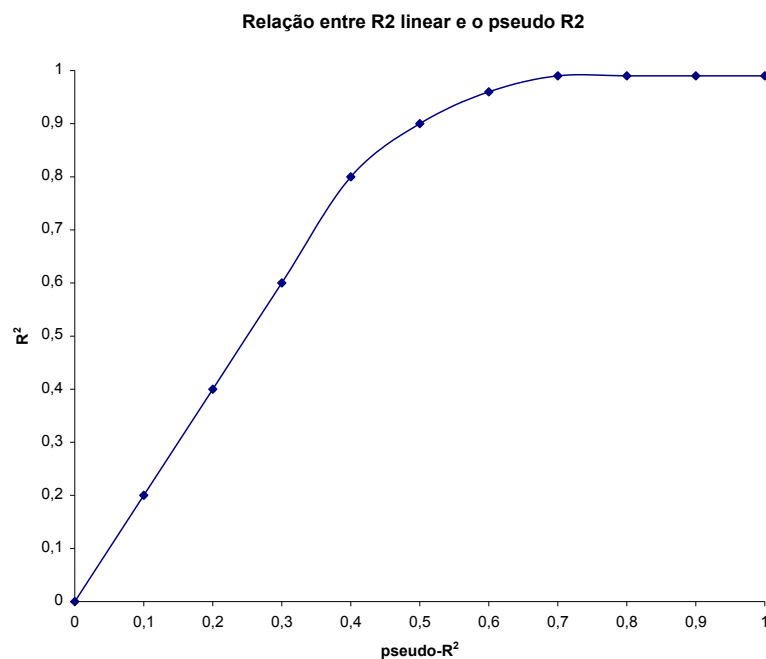


Figura III. 5 - Relação entre R^2 de modelos lineares e R^2 não lineares (pseudo- R^2) definida por Domencich e MacFadden em 1975 (referida em Hensher *et al.*, 2005)

De facto, este parâmetro apesar de similar ao coeficiente de ajustamento das regressões lineares não pode ser interpretado da mesma forma. A figura III.5 representa a relação empírica entre ambos os parâmetros, podendo verificar-se que o R^2 de modelos não lineares (pseudo- R^2) mesmo apresentando valores mais baixos que os de modelos lineares podem

traduzir ajustes aceitáveis aos dados reais. Geralmente valores de pseudo- R^2 superiores a 0,4 já representam bons ajustes.

O índice não tem nenhuma interpretação intuitiva para os valores entre os extremos 0 e 1, apenas representa o incremento percentual na FMVL em relação ao seu valor com os todos os parâmetros nulos, pelo que se pode dizer que entre dois modelos com igual $LL(0)$, o que maior valor no índice apresentar melhor se ajusta aos dados (Train, 2003). Ben-Akiva e Lerman (1985) referem que ρ^2 de um modelo para os mesmos dados de estimação aumenta ou pelo menos mantém o seu valor sempre que mais variáveis forem adicionadas à função utilidade, recomendando assim o seu ajuste para:

$$\bar{\rho}^2 = 1 - \left(\frac{LL(\theta) - K}{LL(0)} \right) \quad \text{Exp. III- 71}$$

Onde K corresponde aos graus de liberdade, ou seja, ao número coeficientes estimados.

Modelos estimados com amostras ou com conjuntos de alternativas diferentes para cada indivíduo não podem ser comparados através do valor dos seus índices de razão de verosimilhança.

Considerando uma hipótese nula H que expresse as restrições aos valores dos parâmetros estimados, as duas hipóteses mais comuns a verificar são:

1ª Hipótese Nula H_0 : Vários parâmetros são nulos;

2ª Hipótese Nula H_1 : Dois ou mais parâmetros têm valor igual.

Quando $LL(\theta)$ é maximizado, os parâmetros θ estimados têm uma distribuição Normal $N(\theta, S^2)$, cujo variância segundo Ortúzar e Willumsen (2001) pode ser determinada por:

$$S^2 = \frac{-1}{\text{média} \left(\frac{\partial^2 LL(\theta)}{\partial \theta^2} \right)} \quad \text{Exp. III- 72}$$

Pelo que se se considerar a hipótese nula H_0 , ou seja, os $\theta_k = 0$, implica que o teste de $t = \theta_k / S^2$ tem uma distribuição normal $N(0,1)$. Apesar de θ poder ser enviesado (distribuição não simétrica), para grandes amostras (superiores a 500 a 1000 observações) esse enviesamento é muito reduzido (e quase coincidente com a distribuição normal) e permite testar t através de uma aproximação à distribuição normal $N(0,1)$ (Ortúzar e Willumsen, 2001). Valores suficientemente grandes de t levam à rejeição da Hipótese nula $\theta_k=0$, e à aceitação de que k -ésimo atributo tem um efeito significativo no modelo. Os valores típicos de t para a rejeição da hipótese nula (quer utilizando a distribuição t com infinitos graus de liberdade, quer a

distribuição normal para uma distribuição *quási-t* com amostras grandes) serão superiores a $\pm 1,96$ para o nível de confiança de 95% (nível de significância de 0,05) e a $\pm 1,65$ para um intervalo de confiança de 90% (nível de significância de 0,10) (Ortúzar e Willumsen, 2001).

Para hipóteses mais complexas a *razão de verosimilhança* é definida por:

$$R = \frac{L(\theta^H)}{L(\theta_b)} \quad \text{Exp. III- 73}$$

Onde:

- $L(\theta^H)$: é o máximo valor restringido da função de verosimilhança atendendo a hipótese nula H de que o modelo com mais restrições não é melhor que o modelo de base;
- $L(\theta_b)$: é o máximo valor da função de verosimilhança sem restrições, ou seja, um modelo considerado de base, geralmente com menos variáveis;

Para testar estatisticamente o ajustamento entre dois modelos semelhantes através da sua razão de verosimilhança é comum recorrer à sua expressão logarítmica (Ortúzar e Willumsen, 2001; Train, 2003, Hensher *et al.*, 2005):

$$LR = -2LL = -2 \cdot \ln R = -2 \cdot [LL(\theta_b) - LL(\theta_R)] \quad \text{Exp. III- 74}$$

Que tem uma distribuição chi-quadrada (χ^2) com graus de liberdade iguais ao número de restrições impostas pela hipótese nula (diferença entre o número de parâmetros estimados) que, dado a função de verosimilhança logarítmica ser sempre negativa, é duas vezes a magnitude da diferença entre os máximos das funções de verosimilhança logarítmica restringida (θ_R ou θ_H) e não restringida de base (θ_b). Se este valor exceder o valor crítico de χ^2 com os graus de liberdade apropriados, então a hipótese nula de que o modelo mais restringido não é melhor que o modelo de base é rejeitada. Se o valor for inferior ao de χ^2 com os graus de liberdade apropriados, então não se pode concluir que o modelo restringido é melhor que o modelo de base (Hensher *et al.*, 2005). Valores típicos da distribuição χ^2 *para apenas um grau de liberdade (mais uma variável independente)* serão de 3,84 para um intervalo de confiança de 90%, de 5,02 para um intervalo de confiança de 95% e de 7,88 para um intervalo de confiança de 99%.

As expressões para as duas hipóteses nulas referidas, para a razão das máximas verosimilhanças serão:

$$1^a \text{ Hipótese nula } H_0: LR^0 = -2 \cdot [LL(0) - LL(\theta)] \quad \text{Exp. III- 75}$$

Na expressão III.75 para testar a hipótese nula (designada de *modelo de igualdade mútua* por Ortúzar e Willumsen, 2001), deve estimar-se o modelo duas vezes, uma vez com a estimação incluindo todas as variáveis independentes (ou todos os coeficientes), e uma segunda sem as mesmas, uma vez que retirando as variáveis se está a forçá-las a terem coeficientes nulos. Na prática resume-se à multiplicação dos logaritmos da probabilidade $\ln(e^0 / \sum_n e^0) = \ln(1 / \sum_n (1))$ por n , ou seja, a dimensão da amostra. O dobro da diferença entre os valores destes máximos é o valor estatístico do teste que é comparado ao valor crítico de chi-quadrado com graus de liberdade equivalentes ao número de variáveis independentes retirados na segunda estimação (ou seja, todos) (Train, 2003).

Para a expressão III.76 o teste à hipótese nula consiste também na estimação do modelo duas vezes, uma primeira com cada uma das variáveis independentes consideradas separadamente, incluindo as duas primeiras, e uma segunda estimação substituindo as duas primeiras variáveis por uma equivalente à sua soma (a adição de duas variáveis força os seus coeficientes a serem iguais). Compara-se o dobro da diferença dos valores máximos obtidos nas estimações anteriores com o valor crítico da distribuição chi-quadrada com apenas um grau de liberdade (Train, 2003).

Ben-Akiva e Lerman (1985) sugerem que para a segunda estimação $LL(\theta^H)$ se considere a repartição modal verificada na amostra selecionada e que se utilize esta como a probabilidade de escolha das diferentes alternativas, o que equivale a considerar a função de máxima verosimilhança logarítmica quando apenas uma constante específica é incluída, geralmente designada por $LL(\theta^C)$.

Hensher *et al.* (2005) e Koppelman e Bhat (2006) sugerem o mesmo teste para comparar quaisquer dois modelos com igual amostra e número de alternativas. Sugerem que o modelo base de comparação seja sempre o que tiver menor número de restrições (ou seja, variáveis a estimar), com graus de liberdade equivalentes à diferença de parâmetros estimados entre modelos e que a hipótese nula a testar é que o modelo mais restritivo não melhora estatisticamente a máxima verosimilhança logarítmica em relação ao modelo de base.

3.2. Modelos de Valores Extremos Generalizados - GEV

O modelo Logit padrão considera que as alternativas são IIA, e portanto que os padrões de substituição são proporcionais entre alternativas. No entanto, esta propriedade tanto pode ser vista como uma restrição à sua aplicação, como o resultado de uma boa especificação do mesmo conseguindo avaliar todas as fontes de correlação entre alternativas na utilidade representativa (mensurável, observada, sistémica, entre outros), de maneira a que apenas ruído de fundo, sem implicações, permaneça. Quando esta especificação não é conseguida e

existem fontes de correlação explícitas que imponham correlações às componentes não observadas das utilidades, o pressuposto *IIA* não é verificado e os modelos Logit padrão não são consistentes na estimação.

Os modelos de valores extremos generalizados (GEV) constituem uma família de modelos que permitem uma variedade de padrões de substituição. O atributo comum a todos os modelos GEV é o facto de considerarem os erros das utilidades de todas as alternativas conjuntamente distribuídas como valores extremos generalizados. Este facto possibilita a correlação entre alternativas e é uma generalização da distribuição de valores extremos para uma variável, que quando têm correlações iguais a zero (não têm correlações), se torna no produto de distribuições de valores extremos independentes que corresponde ao modelo Logit padrão. Além do modelo Logit padrão outros modelos foram desenvolvidos mas poucos têm sido implementados (Train, 2003), pelo que apenas se fará referência ao mais utilizado dos modelos GEV - Logit Hierárquico.

O modelo *Logit Hierárquico* ou “*Nested Logit*” tem sido aplicado em estudos desde os anos 70, e apresenta uma forma funcional mais simples comparativamente à maioria dos restantes modelos da família e permite um considerável conjunto de padrões de substituição (correlação entre erros ou erros correlacionados). No entanto, não é aplicável quando o analista não consegue dividir as alternativas existentes em conjuntos (nichos) de forma a:

- Para duas alternativas do mesmo nicho, a razão de probabilidades de escolha ser independente da existência de todas as outras, isto é, a *IIA* mantém-se dentro de cada nicho;
- Para duas alternativas em nichos diferentes, a razão de probabilidades de escolha pode depender de outra alternativa pertencente aos dois nichos, ou seja, a *IIA* não se mantém em geral para duas alternativas de nichos diferentes.

De um modo geral se for possível a constatação de alguns atributos não observados semelhantes entre alternativas que possam permitir ao analista considerá-las similares, este pode agrupá-las em nichos ou conjuntos. Se se verificarem variações proporcionais semelhantes em alternativas devido a remoção/implementação de uma ou conjunto de alternativas, as mesmas são *IIA* e portanto também podem formar um nicho. Se essas variações de proporção não forem semelhantes, então têm de pertencer a nicho diferentes para que o Logit Hierárquico seja apropriado para modelar a situação.

É comum recorrer à representação invertida de uma árvore de alternativas para ilustrar os nichos. As árvores têm ramos que agrupam conjuntos de alternativas em nichos que depois têm galhos para as alternativas modais. Entre os galhos de cada ramo há *IIA*, ou seja, a proporcionalidade dos padrões de substituição mantém-se, mas entre os ramos não.

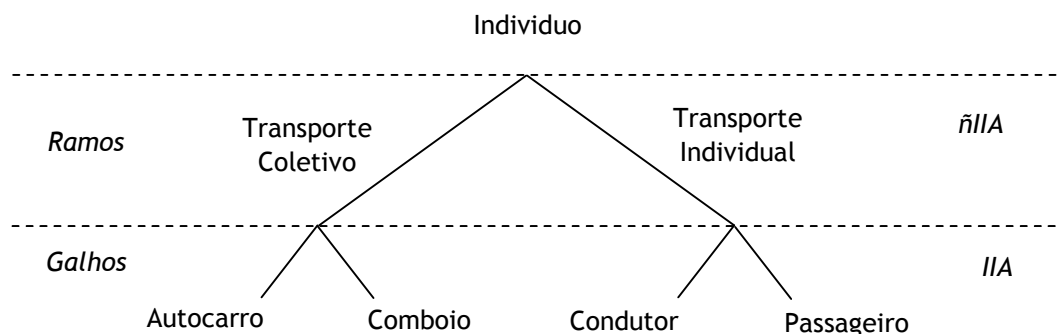


Figura III. 6 - Possível diagrama em árvore para um modelo de escolha modal Logit Hierárquico

Train (2003) refere que vários autores chegaram de diferentes maneiras à conclusão de que o modelo Logit Hierárquico é consistente com a maximização da utilidade. Se um conjunto de J alternativas for dividido em K subconjuntos não sobrepostos designados por B_k (nichos), o modelo Logit Hierárquico é obtido assumindo que o vetor da utilidade não observada tem distribuição cumulativa:

$$F(\epsilon_n) = e^{\left(-\sum_{k=1}^K \left(\sum_{j \in B_k} e^{\frac{\epsilon_{nj}}{\lambda_k}} \right)^{\lambda_k} \right)} \quad \text{Exp. III- 77}$$

Neste modelo GEV a distribuição marginal de cada erro ϵ_{nj} é de valores extremos uni-variável, no entanto, os erros são correlacionados em cada nicho, ou seja, para quaisquer duas alternativas j e m do nicho B_k , ϵ_{nj} está correlacionado com ϵ_{nm} . Duas alternativas de nichos diferentes têm erros não correlacionados, implicando que $\text{Cov}(\epsilon_{nj}, \epsilon_{nm}) = 0$ para qualquer $j \in B_k$ e $m \in B_l$ com $l \neq k$.

O parâmetro λ_k é uma medida do grau de independência dos erros entre alternativas no nicho k , que cresce com o aumento da independência e consecutivamente com a diminuição da correlação. Um valor estatístico para a medição da correlação poderá ser $(1 - \lambda_k)$ uma vez que com o aumento de λ_k , indicação de menos correlação, o seu valor diminui. Apesar de a correlação ser mais complexa que a simples expressão $(1 - \lambda_k)$ Train (2003) sugere-a como uma indicação da correlação. Quando $\lambda_k = 1$ indica completa independência entre alternativas e portanto nenhuma correlação no nicho k . Quando esta situação ocorre em todos os k nichos, significando que todas as alternativas são independentes em todos os nichos, a distribuição GEV torna-se no produto de erros com distribuição de valores extremos independente, tornando-se no modelo Logit padrão.

Neste modelo a probabilidade de uma alternativa $i \in B_k$ ser escolhida será:

$$P_{ni} = \frac{e^{\frac{V_{ni}}{\lambda_k} \cdot \left(\sum_{j \in B_k} e^{\frac{V_{nj}}{\lambda_k}} \right)^{\lambda_k - 1}}}{\sum_{l=1}^K \left(\sum_{j \in B_l} e^{\frac{V_{nj}}{\lambda_l}} \right)^{\lambda_l}}$$

Exp. III- 78

Uma propriedade importante dos Logit Hierárquicos é que apesar de apresentarem *I/A* entre alternativas do mesmo nicho e não entre alternativas de nichos diferentes, uma espécie de *I/A* mantém-se, uma vez que a razão de probabilidades é dependente apenas das alternativas pertencentes aos nichos das alternativas em análise, podendo dizer-se que têm independência de nichos irrelevantes (*I/N*).

O parâmetro λ_k pode ser igual para todos os nichos ou diferente, permitindo correlações entre erros diferentes em cada nicho. Para que o modelo Logit Hierárquico seja consistente com a teoria da maximização da utilidade de todos os possíveis valores das variáveis independentes, o parâmetro $\lambda_k \forall k$ deve estar compreendido entre os valores de 0 e 1 (Train, 2003). A consideração de λ_k como um valor constante implica que todos os indivíduos têm a mesma correlação entre erros, mas tal pode não ser verdade. Se a correlação entre erros variar entre indivíduos atendendo às suas características observadas, λ_k pode ser definido como uma função paramétrica de variáveis socioeconómicas desde que o seu valor seja positivo (um valor negativo contrariaria a teoria da maximização da utilidade).

Train (2003) refere ainda diferentes variações do modelo GEV que se apresentam no Anexo 4 do Capítulo III.

3.3. Probit Multinomial

3.3.1. Formulação Base

O modelo Logit tem três limitações importantes, não consegue representar variações de gosto, apresenta restrições nos padrões de substituição devido à independência de alternativas irrelevantes e não pode ser utilizado com dados obtidos através de inquéritos de preferências declaradas quando os fatores não observados têm correlação no tempo para cada decisor (Train, 2003).

Os modelos Probit conseguem tratar variações de gosto, permitem padrões de substituição e são aplicáveis a dados de preferências declaradas com erros temporalmente correlacionados. As limitações da sua aplicabilidade devem-se fundamentalmente a não terem uma fórmula definida o que complica o seu cálculo e o facto de considerarem que os erros ou componentes não observadas da utilidade são normalmente distribuídos. Apesar de em grandes amostras se poder considerar que a distribuição normal dos erros é uma aproximação adequada para a maioria das situações, existem casos em que essa assunção pode levar a previsões erróneas.

Um exemplo proeminente da desadequação da distribuição normal aos erros está relacionado com os coeficientes da variável independente custo ou preço. Se se assumir que estes erros têm distribuição normal, significa considerar que a sua densidade tem valores para ambos os lados a partir de zero, implicando que haja indivíduos que tenham coeficiente positivo para o custo/preço, significando que a probabilidade de escolha aumenta com o aumento do preço o que contraria claramente a teoria comportamental. Uma distribuição que apenas tenha densidade para um dos lados do zero, como a distribuição lognormal⁶, é claramente mais realista, no entanto, não pode ser incorporada na modelação Probit. À parte destas considerações o modelo Probit tem aplicabilidade quase geral.

A primeira derivação do modelo remota aos anos 20 do séc. XX, mas a generalização da sua especificação para representar vários aspetos do comportamento de escolhas apenas foi realizada nos anos 70 (Ben-Akiva e Lerman, 1985; Train, 2003).

Nos modelos Probit, assume-se que a parte não mensurável, não observável, os resíduos ou erros (ε_n) têm uma distribuição normal com média igual a zero e matriz de covariância Ω , cuja função densidade de probabilidades é:

$$\varphi(\varepsilon_n) = \frac{1}{(2\pi)^{J/2} \cdot |\Omega|^{1/2}} \cdot e^{-\frac{1}{2} \varepsilon_n' \Omega^{-1} \varepsilon_n} \quad \text{Exp. III- 79}$$

Onde J é o número total de alternativas disponíveis e Ω pode depender das variáveis apresentadas a cada decisor.

Assim, a probabilidade de escolha será:

$$\begin{aligned} P(i | C_n) &= \text{Prob}[V_{nj} + \varepsilon_{ni} > V_{ni} + \varepsilon_{nj}, \forall j \in C_n] = \\ &= \int I(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) \cdot \varphi(\varepsilon_n) \cdot d\varepsilon_n \end{aligned} \quad \text{Exp. III- 80}$$

Em que V é a utilidade representativa, $I(\dots)$ é o indicador de uma expressão a definir que verifique a inequação entre parêntesis e o integral é calculado em função de todos valores

6 - Distribuição lognormal - Segundo Montgomery e Runger (2007) uma variável aleatória Lognormal x tem uma distribuição de probabilidades igual à de $x = e^{(w)}$ para uma variável aleatória normal w , um vez que $\ln(x) = w$. Se w tiver uma distribuição normal com média θ e variância ω^2 , então $x = e^w = \exp(w)$ tem distribuição lognormal com uma função de densidade de probabilidade de:

$$f(x) = \frac{1}{x \cdot \omega \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\left[\frac{(\ln x - \theta)^2}{2 \cdot \omega^2} \right]}; \text{ com } 0 < x < \infty$$

Uma função de distribuição cumulativa de:

$$F(x) = \Phi \left[\frac{\ln(x) - \theta}{\omega} \right]; \text{ que tem a particularidade de } F(x) = 0 \text{ para } x \leq 0.$$

E média e variância igual a:

$$\mu(x) = e^{\theta + \omega^2/2}; \quad \sigma^2(x) = e^{2\theta + \omega^2} \cdot (e^{\omega^2} - 1);$$

dos erros ε_n . É possível verificar que, ao contrário do Logit, o integral não tem forma definida, pelo que tem de ser definido numericamente por simulação.

Para facilitar a simulação do integral a probabilidade de escolha pode ser expressa de outra forma, considerando B_{ni} como o conjunto dos erros ε_n que resulta na escolha pelo indivíduo da alternativa i :

$$B_{ni} = \{\varepsilon_n : V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i\} \quad \text{Exp. III- 81}$$

Assim:

$$P_{ni} = \int_{\varepsilon_n \in B_{ni}} \varphi(\varepsilon_n) \cdot d\varepsilon_n \quad \text{Exp. III- 82}$$

Que é um integral sobre apenas alguns dos valores de ε_n , e não de todos os possíveis valores incluídos em B_{ni} (Train, 2003).

As expressões III.80 e III.82 são integrais com dimensão J (número de alternativas) sobre os J erros ε_{nj} , mas como apenas a diferença entre utilidades interessa a probabilidade de escolha pode ser igualmente expressa como integrais de $(J-1)$ dimensões sobre a diferença entre erros. Se a diferença for em relação à alternativa i , para a qual queremos calcular a probabilidade de escolha, então a expressão III.80 toma a forma:

$$\begin{aligned} P_{ni} &= \int I(0 > V_{nj} - V_{ni} + \varepsilon_{nj} - \varepsilon_{ni}, \forall j \neq i) \varphi(\varepsilon_n) \cdot d\varepsilon_n = \\ &= \int I(\tilde{V}_{nji} + \tilde{\varepsilon}_{nji} < 0, \forall j \neq i) \cdot \varphi(\tilde{\varepsilon}_{nji}) \cdot d\tilde{\varepsilon}_{nji} \end{aligned} \quad \text{Exp. III- 83}$$

Onde $\tilde{V}_{nji} = V_{nj} - V_{ni}$ e $\tilde{\varepsilon}_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$.

Esta expressão afirma que a probabilidade de escolha de uma alternativa i é a probabilidade de todas as diferenças de utilidade serem negativas, quando diferenciadas em relação a alternativa i em análise. O vetor $\tilde{\varepsilon}_{nji}$ tem dimensão $J-1$ uma vez que não faz a diferença dos erros da alternativa i com ela mesma (considere-se $\tilde{\varepsilon}_{nji} = \tilde{\varepsilon}_{ni}$), e uma vez que a diferença entre duas variáveis com distribuição normal é também normal, a densidade de probabilidade das diferenças será:

$$\varphi(\tilde{\varepsilon}_{ni}) = \frac{1}{(2 \cdot \pi)^{J/2} \cdot |\tilde{\Omega}_i|^{1/2}} \cdot e^{-\frac{1}{2} \tilde{\varepsilon}_{ni} \cdot \tilde{\Omega}_i^{-1} \cdot \tilde{\varepsilon}_{ni}} \quad \text{Exp. III- 84}$$

Pelo que a probabilidade de escolha da alternativa i passará a ser o integral de dimensão $J-1$:

$$P_{ni} = \int_{\tilde{\varepsilon}_{ni} \in \tilde{B}_{ni}} \varphi(\tilde{\varepsilon}_{ni}) \cdot d\tilde{\varepsilon}_{ni}$$

Exp. III- 85

As expressões anteriores utilizam a matriz de covariância $\tilde{\Omega}_i$ das diferenças entre erros, que é calculada a partir da matriz de covariância entre os mesmos (Ω). Train (2003) apresenta uma metodologia simples que permite definir a matriz de diferença de erros para qualquer situação, designada de matriz M (ver Anexo 5 do Capítulo III).

3.3.2. Normalização

Nos modelos da família Logit, a normalização da escala e do nível da utilidade era realizada automaticamente atendendo aos pressupostos considerados na distribuição dos seus erros (IID com distribuição de valores extremos do tipo I) pelo que a normalização não era realizada de forma explícita. Nos modelos Probit a normalização da escala e do nível não ocorre automaticamente pelo que tem que ser realizada diretamente pelo analista.

A normalização de um modelo está relacionada com a identificabilidade dos parâmetros. Um parâmetro é identificável se puder ser estimado, e será inidentificável se não puder ser estimado, ou seja, se não afetar o comportamento do indivíduo. Num modelo não normalizado podem aparecer parâmetros relacionados com a escala e o nível de utilidade que não afetam o comportamento. A normalização do modelo extrai os parâmetros não identificáveis. Train (2003) refere que a situação é mais complexa devido à dificuldade de saber quais os parâmetros relacionados com a escala e o nível de utilidade, referindo mesmo a publicação de artigos científicos com modelos Probit não normalizados em que nem os autores nem os revisores detetaram quais os parâmetros inidentificáveis. Para a normalização de modelos Probit Train (2003) apresenta um procedimento que pode ser utilizado em todas as situações para assegurar que todos os parâmetros estimados são identificáveis ou para avaliar outros procedimentos de normalização (ver Anexo 6 do Capítulo III).

3.3.3. Variações de Gosto

Tal como já foi referido, os modelos Logit simples não modelam variações de gosto, ou seja, situações em que os coeficientes dos parâmetros variem aleatoriamente entre decisores. Nos modelos Logit consideram-se os coeficientes constantes para todos os decisores, variando apenas de alternativa em alternativa. Os modelos Probit são particularmente indicados para modelar utilidades com coeficientes que podem variar de indivíduo para indivíduo (Ben-Akiva e Lerman, 1985; Ortúzar e Willumsen, 2001; Train, 2003), isto é, que sejam definidos como:

$$U_{nj} = \theta'_n \cdot x_{nj} + \varepsilon_{nj}$$

Exp. III- 86

Onde θ'_n é um vetor de coeficientes para o decisor n que representa a sua variação de gostos pessoais, que se assume que tenha distribuição normal entre a população com média b e

covariância W : $\beta_n \sim N(b, W)$. O objetivo da análise será a estimação destes dois parâmetros. Se se decompuser β_n na sua média e desvio da sua média, teremos a expressão:

$$U_{nj} = b' \cdot x_{nj} + \tilde{\beta}_n \cdot x_{nj} + \varepsilon_{nj}; \text{ onde } \tilde{\beta}_n = \beta_n - b \quad \text{Exp. III- 87}$$

Os dois últimos termos da utilidade são aleatórios e se a sua soma for definida como η_{nj} a utilidade fica com a expressão:

$$U_{nj} = b' \cdot x_{nj} + \eta_{nj} \quad \text{Exp. III- 88}$$

A covariância de η_{nj} depende de W e dos x_{nj} , para que a sua covariância divirja entre indivíduos. Um exemplo para definir a covariância dos η_{nj} de um modelo com duas alternativas teria as utilidades:

$$\begin{aligned} U_{n1} &= \beta_n \cdot x_{n1} + \varepsilon_{n1} \\ U_{n2} &= \beta_n \cdot x_{n2} + \varepsilon_{n2} \end{aligned} \quad \text{Exp. III- 89}$$

Assumindo que os β_n são normalmente distribuídos com média b e variância σ_β , e que os erros são IID (não é um pressuposto necessário) com variância ε_n , permite reescrever as utilidades como:

$$\begin{aligned} U_{n1} &= b \cdot x_{n1} + \eta_{n1} \\ U_{n2} &= b \cdot x_{n2} + \eta_{n2} \end{aligned} \quad \text{Exp. III- 90}$$

Onde η_{n1} e η_{n2} são normalmente distribuídos com média igual a zero (Ben-Akiva e Lerman, 1985; Train, 2003):

$$\mu(\eta_{nj}) = \mu(\tilde{\beta}_n \cdot x_{nj} + \varepsilon_{nj}) = 0 \quad \text{Exp. III- 91}$$

e cada um com variância:

$$\text{Var}(\eta_{nj}) = \text{Var}(\tilde{\beta}_n \cdot x_{nj} + \varepsilon_{nj}) = x_{nj}^2 \cdot \sigma_\beta + \sigma_\varepsilon \quad \text{Exp. III- 92}$$

A sua covariância será:

$$\begin{aligned} \text{Cov}(\eta_{n1}, \eta_{n2}) &= \mu[(\tilde{\beta}_n \cdot x_{n1} + \varepsilon_{n1})(\tilde{\beta}_n \cdot x_{n2} + \varepsilon_{n2})] = \\ &= \mu[(\tilde{\beta}_n^2 \cdot x_{n1} \cdot x_{n2} + \tilde{\beta}_n \cdot x_{n1} \cdot \varepsilon_{n2} + \tilde{\beta}_n \cdot x_{n2} \cdot \varepsilon_{n1} + \varepsilon_{n1} \cdot \varepsilon_{n2})] = \\ &= x_{n1} \cdot x_{n2} \cdot \sigma_\beta \end{aligned} \quad \text{Exp. III- 93}$$

Que define a matriz de covariância da seguinte forma:

$$\begin{aligned}\Omega &= \begin{pmatrix} x_{n1}^2 \cdot \sigma_B + \sigma_\varepsilon & x_{n1} \cdot x_{n2} \cdot \sigma_B \\ x_{n1} \cdot x_{n2} \cdot \sigma_B & x_{n2}^2 \cdot \sigma_B + \sigma_\varepsilon \end{pmatrix} = \\ &= \sigma_B \cdot \begin{pmatrix} x_{n1}^2 & x_{n1} \cdot x_{n2} \\ x_{n1} \cdot x_{n2} & x_{n2}^2 \end{pmatrix} + \sigma_\varepsilon \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\end{aligned}$$

Para a normalização do modelo é conveniente utilizar $\sigma_\varepsilon = 1$ (Train, 2003), resultando na matriz de covariância:

$$\Omega = \sigma_B \cdot \begin{pmatrix} x_{n1}^2 & x_{n1} \cdot x_{n2} \\ x_{n1} \cdot x_{n2} & x_{n2}^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Com este procedimento, e dado que x_{n1} e x_{n2} são observados pelo analista, resta estimar os parâmetros b e σ_B , ficando-se a conhecer a média e a variância do coeficiente aleatório na população.

3.3.4. Padrões de Substituição/Falha no Pressuposto //A.

O modelo Logit, devido ao pressuposto de //A, assume um padrão de substituição proporcional, ou seja, o que uma utilidade aumenta diminui proporcionalmente nas restantes. O modelo Probit permite representar qualquer padrão de substituição, uma vez que diferentes matrizes de covariância Ω permitem diferentes padrões de substituição e ao estimar a matriz de covariância o analista determina o padrão de substituição mais adequado aos dados que possui.

Existem duas abordagens possíveis: a estimação completa da matriz de covariância; ou a imposição de uma estrutura à matriz de covariância que represente fontes particulares de dependência. A última abordagem sugerida geralmente reduz o número de parâmetros a estimar e facilita a sua interpretação. Exemplos destas abordagens sugeridas em Train (2003) são apresentados no Anexo 7 do Capítulo III.

3.3.5. Dados Recolhidos ao Longo do Tempo por Preferências Declaradas

Segundo Train (2003) o modelo Probit para escolhas repetidas é semelhante ao Probit com uma escolha por decisor, diferenciando-se apenas pela dimensão da matriz de covariância dos erros que será maior. No Anexo 8 do Capítulo III apresenta-se uma metodologia para a sua inclusão na modelação sugerida por Train (2003).

3.3.6. Simulação das Probabilidades de Escolha

O modelo Probit para o cálculo das probabilidades de escolha não tem uma expressão definida, pelo que o seu cálculo tem de ser realizado através de uma aproximação. Esta aproximação pode ser realizada numericamente ou por simulação. Train (2003) apresenta dois métodos para uma aproximação não simulada: os métodos de quadratura, e o algoritmo de

Clark (ver pág. 274 de Ortúzar e Willumsen, 2001). No entanto, refere que a sua utilização é limitada e pode produzir aproximações bastante imprecisas e com grau de precisão difícil de avaliar.

Assim, como alternativa, Train (2003) sugere vários métodos de simulação, nomeadamente:

- Simulador de Aceitação-Rejeição;
- Simulador de Aceitação-Rejeição suavizado;
- Simulador GHK;
- Simulador GHK com estimação da máxima verosimilhança.

Destes, Train (2003) e Ortúzar e Willumsen (2001) referem que o mais comum (e também com melhores resultados nas mais diversas situações) é o simulador GHK, assim designado pelas iniciais dos grupos de trabalho que o desenvolveram no início da década de 90. No Anexo 9 do Capítulo III apresenta-se um exemplo de cálculo para o simulador de GHK com 3 alternativas e as especificidades do simulador GHK com estimação da máxima verosimilhança.

3.4. Logit Misto

3.4.1. Formulação Base

O modelo Logit misto, Logit Kernel, Logit de parâmetros aleatórios ou modelo dos componentes dos erros, é considerado o modelo do novo milénio (Ortúzar e Willumsen, 2001), e apesar de ter surgido no final da década de 70 apenas nos finais dos anos 90 é que se reconheceram as suas reais potencialidades.

O modelo Logit misto tem uma grande flexibilidade e consegue aproximar qualquer modelo de utilidade aleatória (Train, 2003). Permite variações de gosto aleatórias, padrões de substituição sem restrições e correlações entre erros ao longo do tempo. Não é restringido a erros com distribuição normal como o Probit e a simulação das suas probabilidades de escolha é computacionalmente simples. Por estes motivos é considerado o modelo do futuro para a modelação de escolhas discretas.

Nos modelos Logit mistos considera-se que os vetores de parâmetros θ associados às variáveis observadas (ou atributos X) variam aleatoriamente dependendo dos gostos individuais de cada decisor (contrariamente aos modelos Logit padrão que os assumem constantes para todos os indivíduos), e que os erros aleatórios têm uma distribuição *IID* de *Gumbel* independentes tanto dos parâmetros como dos atributos (se se considerassem os erros e os atributos com distribuição normal a formulação resultaria no modelo Probit) (Ortúzar e Willumsen, 2001). Estas variações nos parâmetros induzem correlações nas utilidades entre alternativas e entre situações. Os modelos Logit mistos são definidos com base na forma funcional das suas probabilidades de escolha, e são representados por integrais de probabilidades Logit padrão

sobre a densidade dos parâmetros, ou seja, serão todos os modelos cujas probabilidades de escolha que possam ser expressos como:

$$P_{ni} = \int L_{ni}(\theta) \cdot f(\theta) \cdot d\theta \quad \text{Exp. III- 94}$$

Onde L_{ni} é a fórmula geral da probabilidade Logit avaliada para os parâmetros θ , e $f(\theta)$ é a função de densidade de probabilidades. Se a utilidade for linear em relação aos parâmetros a sua fórmula geral será:

$$P_{ni} = \int \left(\frac{e^{\theta \cdot x_{ni}}}{\sum_j e^{\theta \cdot x_{nj}}} \right) \cdot f(\theta) \cdot d\theta \quad \text{Exp. III- 95}$$

As probabilidades Logit mistas são uma média ponderada da fórmula Logit avaliada com diferentes valores de θ , onde os pesos são definidos pela densidade $f(\theta)$. Em estatística uma média ponderada de diversas funções é uma função mista e a densidade que permite a ponderação é designada de distribuição mista. O modelo Logit Misto torna-se num Logit simples com IIA se a distribuição mista dos erros tiver uma variância igual a zero.

A função de distribuição mista pode ser discreta e θ ter um conjunto distinto de valores. Se este tiver M possíveis valores nomeados $b_1, \dots, b_M$, com probabilidade S_m que $\theta = b_m$ então o Logit misto torna-se num modelo de classe latente e a sua probabilidade de escolha será (Train, 2003):

$$P_{ni} = \sum_{m=1}^M S_m \cdot \left(\frac{e^{b_m \cdot x_{ni}}}{\sum_j e^{b_m \cdot x_{nj}}} \right) \quad \text{Exp. III- 96}$$

Esta especificação é útil quando há M segmentos da população com o seu próprio comportamento de escolha e preferências. A percentagem da população do segmento m será S_m e poderá ser estimado em conjunto com os restantes parâmetros b de cada segmento.

Na maioria das suas aplicações, no entanto, o modelo Logit misto considera a função de densidades contínua. Esta densidade pode ser considerada com distribuições normais, lognormais, uniformes, triangulares, gama, entre outros. Assim, uma correta especificação das variáveis independentes e das suas densidades mistas permitem representar qualquer comportamento maximizante da utilidade bem como de alguns comportamentos não maximizantes (Train, 2003).

É de referir que a função densidade caracteriza os parâmetros do modelo, e que os próprios parâmetros dessa função, por exemplo, a média e a variância, são geralmente o objeto da estimação. Assim, se os parâmetros β (média e variância) caracterizarem a função $f(\theta)$, o que caracterizará a probabilidade de escolha Logit misto serão os β e não os θ . A probabilidade de escolha será:

$$P_{ni} = \int L_{ni}(\theta).f(\theta | B).d\theta \quad \text{Exp. III- 97}$$

Que é função de β , o que equivale a dizer que os θ se assemelham aos ε , no sentido em que ambos são termos aleatórios que são integrados externamente para o cálculo das probabilidades.

A probabilidade dos modelos Logit mistos pode ser calculada a partir de comportamentos maximizantes da utilidade de várias maneiras, formalmente equivalentes, mas que permitem interpretações diferentes. As mais conhecidas são baseadas em coeficientes aleatórios (os θ variam entre decisores ao contrário do considerado nos Logit padrão em que eram constantes) ou incluindo componentes de erros que criam correlações entre utilidades de diferentes alternativas.

3.4.2. Coeficientes Aleatórios

A derivação mais simples e mais utilizada em aplicações recentes é baseada em coeficientes aleatórios. A utilidade que um indivíduo n recebe de uma alternativa j , de entre J alternativas disponíveis, será:

$$U_{nj} = \theta_n x_{nj} + \varepsilon_{nj} \quad \text{Exp. III- 98}$$

Onde os x_{nj} são as variáveis observadas relacionadas com a alternativa e com o indivíduo, θ_n é o vetor de coeficientes das variáveis observadas para o indivíduo n representando o seu gosto, e ε_{nj} é um termo aleatório que tem distribuição *IID*. Os coeficientes θ_n variam entre os decisores da população com densidade de probabilidades $f(\theta)$. Esta densidade é uma função dos parâmetros θ e representa a média e a covariância dos θ na população. É semelhante à especificação do modelo Logit só que não é constante entre os indivíduos. O analista consegue observar os valores de x_{nj} mas não os de θ_n nem de ε_{nj} . Se os coeficientes fossem conhecidos o modelo seria um Logit padrão uma vez que os erros são considerados com distribuição *IID*.

Assim, a probabilidade condicional de θ_n seria:

$$L_{ni}(\theta_n) = \frac{e^{\theta_n \cdot x_{ni}}}{\sum_j e^{\theta_n \cdot x_{nj}}} \quad \text{Exp. III- 99}$$

No entanto, uma vez que θ_n não é conhecido, a probabilidade não pode ser condicional, pelo que a probabilidade incondicional é o integral $L_{ni}(\theta_n)$ para todas as possíveis variáveis de θ_n :

$$P_{ni} = \int \left(\frac{e^{\theta \cdot x_{ni}}}{\sum_j e^{\theta \cdot x_{nj}}} \right) \cdot f(\theta) \cdot d\theta \quad \text{Exp. III- 100}$$

Designada por probabilidade Logit mista. Será o analista que definirá a distribuição para os coeficientes e que estimará os parâmetros dessa distribuição.

As distribuições mais comuns são a normal ou a lognormal: $\theta \sim N(b, W)$ ou $\ln \theta \sim N(b, W)$, com b e W estimados. A distribuição lognormal é útil quando se sabe que o parâmetro tem o mesmo sinal para todos os decisores, tal como o coeficiente do preço que se sabe ser negativo para toda a gente (todos consideram uma utilidade menor para uma alternativa com o aumento do seu custo).

Outras distribuições são também utilizadas como: a distribuição uniforme e a triangular (ver Anexo 10 do Capítulo III). Na distribuição uniforme os θ são uniformemente distribuídos entre $b-s$ e $b+s$ (b é a média e s a amplitude “Spread” e são estimados). Na distribuição triangular os θ têm uma densidade positiva entre $b-s$ que sobe linearmente até b e depois decresce linearmente até $b+s$, assumindo a forma de um triângulo (b e s são estimados como para a distribuição uniforme mas a densidade em vez de plana tem um pico); entre outros. Estas distribuições têm a vantagem de serem limitadas de ambos os lados evitando o problema da distribuição normal e lognormal que permitem coeficientes exageradamente grandes para alguma porção dos decisores. Ao restringir $s=b$ o analista pode assegurar que os coeficientes têm o mesmo sinal para todos os decisores (Train, 2003).

As variações de gosto relacionadas com os atributos observados do decisor (variáveis socioeconómicas) podem ser integradas pela especificação da variável independente e/ou através de distribuições mistas. Como exemplo, o custo pode ser integrado na especificação dividido pelos rendimentos de modo a que o seu valor ou importância relativa decresça com o aumento dos rendimentos (situação normal). Este coeficiente aleatório representaria então a variação, entre a população com o mesmo rendimento, da importância que depositam no custo. Segundo Train (2003) a avaliação média do custo decresce com o aumento do rendimento enquanto a variância em torno da média se mantém fixa. Os coeficientes que

definem as distribuições dos parâmetros θ também podem ser especificados em função de características observadas nos decisores.

3.4.3. Componentes dos Erros

Outra maneira de especificar o modelo Logit misto pode ser através da decomposição dos erros de forma a criar correlações entre as utilidades de diferentes alternativas. Neste caso a utilidade será especificada como:

$$U_{ni} = \alpha' \cdot x_{ni} + \mu_n' \cdot z_{nj} + \varepsilon_{nj} \quad \text{Exp. III- 101}$$

Onde x_{nj} e z_{nj} são vetores de variáveis observadas relacionadas com a alternativa j , α' é um vetor de coeficiente fixos, μ' é um vetor de termos aleatórios com média igual a zero, e ε_{nj} tem distribuição *IID*. Os termos z_{nj} são componentes dos erros que em conjunto com os ε_{nj} formam a parte hipotética (estocástica) da utilidade. Assim, a parte não observada da utilidade (a parte aleatória) será:

$$\eta_{nj} = \mu_n' \cdot z_{nj} + \varepsilon_{nj} \quad \text{Exp. III- 102}$$

Esta poderá ser correlacionada entre alternativas dependendo da especificação de z_{nj} . No modelo Logit padrão os z_{nj} são nulos de forma a não haver correlação entre alternativas e a matriz de covariância é diagonal, implicando que haja *I/A* e que os padrões de substituição sejam proporcionais. Quando z_{nj} não é nulo, a utilidade é correlacionada entre alternativas e a covariância será:

$$\text{Cov}(\eta_{ni}, \eta_{nj}) = E(\mu_n' \cdot z_{ni} + \varepsilon_{ni})(\mu_n' \cdot z_{nj} + \varepsilon_{nj}) = z_{ni}' \cdot W \cdot z_{nj} \quad \text{Exp. III- 103}$$

Sendo $E(\dots)$ o valor esperado ou média da expressão (...), e W a matriz de covariância de μ_n . Será a escolha das variáveis apropriadas a incluir na componente dos erros que definirá os possíveis padrões de correlação e assim os padrões de substituição.

Assim, estas duas especificações são formalmente equivalentes, uma vez que os coeficientes aleatórios θ_n podem ser decompostos na sua média α (relacionados com os coeficientes fixos) e nos seus desvios μ_n (efetivamente relacionados com os termos aleatórios) que entram na componente dos erros.

3.4.4. Padrões de Substituição

Os modelos Logit mistos não exibem a *I/A* ou padrões restritivos de substituição dos Logit padrão (proporcionais), uma vez que a razão entre probabilidades Logit mistas depende de todos os dados incluindo atributos de alternativas diferentes de i e j , dado que o denominador da fórmula Logit está inserido no integral e, portanto, não pode ser anulado:

$$\frac{P_{ni}}{P_{nj}} = \frac{\int \left(\frac{e^{\theta_n \cdot x_{ni}}}{\sum_k e^{\theta_n \cdot x_{nk}}} \right) \cdot f(\theta) \cdot d\theta}{\int \left(\frac{e^{\theta_n \cdot x_{nj}}}{\sum_k e^{\theta_n \cdot x_{nk}}} \right) \cdot f(\theta) \cdot d\theta} \quad \text{Exp. III- 104}$$

De igual modo a elasticidade cruzada é diferente para cada i , uma vez que (Train, 2003):

$$E_{nix_{nj}^m} = -\frac{1}{P_{ni}} \cdot \int \theta^m \cdot L_{ni}(\theta) \cdot L_{nj}(\theta) \cdot f(\theta) \cdot d\theta = -\int \theta^m \cdot L_{nj}(\theta) \cdot \left[\frac{L_{ni}(\theta)}{P_{ni}} \right] \cdot f(\theta) \cdot d\theta \quad \text{Exp. III- 105}$$

Onde θ^m é o m -ésimo elemento do vetor θ . Assim, o padrão de substituição depende da especificação realizada pelo analista das variáveis e da função mista, podendo ser determinadas empiricamente.

3.4.5. Inclusão de Dados Recolhidos por Escolhas Repetidas - Dados de Painéis.

A especificação do modelo Logit misto, de modo a incluir diversas escolhas de cada indivíduo da amostra, trata os coeficientes que entram na utilidade como se variassem entre decisores mas fossem constantes entre situações de escolha para cada indivíduo. Assim a utilidade para a alternativa i na situação de escolha t pelo indivíduo n será definida como:

$$U_{nit} = \theta_n \cdot x_{nit} + \varepsilon_{nit} \quad \text{Exp. III- 106}$$

Em que o termo dos erros tem distribuição *IID* de valores extremos sobre o tempo, indivíduos e alternativas (Train, 2003). Se se considerar uma sequência de T alternativas num período temporal $i = \{i_1, \dots, i_T\}$, a probabilidade condicional de um indivíduo realizar uma sequência de escolhas é o produto de fórmulas Logit padrão (Train, 2003), ou seja:

$$L_{ni}(\theta) = \prod_{t=1}^T \left[\frac{e^{\theta_n \cdot x_{nit}}}{\sum_j e^{\theta_n \cdot x_{njt}}} \right] \quad \text{Exp. III- 107}$$

Uma vez que os erros são independentes no tempo. A probabilidade incondicional será o integral do produto anterior sobre todos os valores de θ :

$$P_{ni} = \int L_{ni}(\theta) \cdot f(\theta) \cdot d\theta \quad \text{Exp. III- 108}$$

Assim, a diferença do modelo Logit misto para várias situações de escolha comparativamente a apenas uma situação por indivíduo é a integração envolver o produto de fórmulas Logit padrão para cada período temporal em vez de apenas uma. A sua simulação é realizada de forma similar à situação de apenas uma escolha (Train, 2003).

Assumindo que os gostos pessoais não são constantes no tempo, e portanto que podem variar durante o período a que se referem as várias observações, implica que os coeficientes a estimar associados a cada decisor variem também, podendo esta variação no tempo ser especificadas de diferentes formas. Se se considerar que os gostos pessoais possam estar correlacionados nas várias situações de escolha, a função de utilidade pode ser expressa (Train, 2003):

$$\begin{aligned} U_{nit} &= \theta_{nt} \cdot x_{nit} + \varepsilon_{nit} \\ \theta_{nt} &= b + \tilde{\theta}_{nt} \\ \tilde{\theta}_{nt} &= \rho \cdot \tilde{\theta}_{nt-1} + \mu_{nt} \end{aligned} \quad \text{Exp. III- 109}$$

Em que b é constante e μ_{nt} é *IID* em relação aos indivíduos e ao tempo e ρ é a auto-correlação⁷ estimada através do processo de máxima verosimilhança.

3.4.6. Especificidades de Identificação do Modelo Logit Misto

O conceito do modelo Logit misto pode ser sintetizado como um modelo de escolha discreta em que as perturbações ou erros das utilidades são formados por uma parte semelhante aos modelos Probit (podendo, no entanto, ter qualquer distribuição) adicionada a outra parte com distribuição *IID* de valores extremos semelhante aos modelos Logit padrão. Segundo Walker (2002), existem vários exemplos de aplicações deste tipo de modelo onde os problemas de identificação têm sido ignorados. Vários artigos científicos foram publicados com erros de especificação devido a estes problemas de identificação, que levam à estimação de parâmetros de atributos e das distribuições de erros aparentemente credíveis mas que não correspondem à otimização do modelo. Tal como existem procedimentos de identificação dos parâmetros para os modelos Logit padrão, como a restrição da escala à unidade ou a anulação da constante específica da alternativa a considerar como referência, existe a necessidade de desenvolver um conjunto de orientações para a identificação de modelos Logit mistos. Walker (2002) e Ben-Akiva *et al.* (2001) apresentam algumas sugestões encontradas empiricamente de modo a conseguir minimizar os problemas referidos. Estas sugestões são exemplos de

7 - Enquanto a correlação é definida entre diferentes variáveis, a auto-correlação de um processo aleatório, em estatística, descreve a correlação do processo em diferentes pontos no tempo. É definida como a auto-covariância normalizada através da divisão pela variância. Se um processo tiver um valor no tempo t de X_t e no tempo s de X_s , se tiver ainda média μ e variância σ^2 , então:

$$\rho(t,s) = \frac{\text{Cov}(X_t, X_s)}{\text{Var}(X_t, X_s)} = \frac{E[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \cdot \sigma_s} = \frac{Y_s}{Y_0}$$

Onde $E[.]$ é o valor esperado, e se o processo estiver bem definido tem valores entre 1 (correlação perfeita) e -1 (anti-correlação perfeita) (Washington *et al.*, 2003; Hensher *et al.*, 2005).

estimação de diversas estruturas de erros tais como: estrutura de covariância sem restrições, similar à estimação para modelos Probit puros; modelos heteroscedásticos; modelos hierárquicos; e modelos Logit de parâmetros aleatórios, apesar de muitas outras também poderem ser especificadas.

Além das especificações gerais para os modelos de escolha discreta para a normalização da escala e da utilidade, existem ainda outras que necessitam de ser avaliadas para que o modelo seja identificável e assim estimável. Estudos empíricos (Walker, 2002) têm revelado que a consideração da ordem, da característica e da definição positiva da matriz de covariâncias entre diferenças de utilidade, se apresentam como verificações adicionais que permitem avaliar o número de parâmetros estimáveis (ver Anexo 11 do Capítulo III). O motivo de necessidade de normalizações adicionais provém da infinidade de possíveis soluções que resultam da hipotética estrutura de covariância assumida, de modo a estabelecer a existência de uma solução única que não altere a estrutura do modelo. A condição de definição positiva permite verificar se a normalização escolhida é válida e se os restantes parâmetros estimados (portanto, não normalizados) conseguem replicar a estrutura do modelo subjacente. As normalizações comuns para outros modelos de escolha discreta, e aparentemente óbvias (por exemplo, $\sigma = 1$ ou $\mu = 1$), podem não ser válidas para os modelos Logit mistos pois a estrutura do modelo impede a recuperação da matriz de covariância das diferenças entre utilidades subjacente. Se as normalizações forem válidas os resultados do modelo mediante a estimação com essas mesmas normalizações serão idênticos. Assim, todas as normalizações válidas para uma determinada especificação resultam em matrizes de covariâncias idênticas, ou seja, os parâmetros estimados com uma determinada normalização (por exemplo, $\sigma_{11} = 0$) podem ser convertidos nos parâmetros estimados impostos por uma normalização diferente (por exemplo, $\sigma_{11} = 1$), desde que ambas sejam válidas (ver Anexo 11 do Capítulo III).

A maioria dos programas informáticos para tratamento estatístico de modelos de escolha discreta permite a estimação de resultados mesmo que o modelo não seja identificável, necessitando que a sua aplicação seja antecedida de uma correta validação suplementar do mesmo por parte do analista. Para avaliar a identificabilidade do modelo, Walker (2002) sugere a análise da matriz *hessiana* e das estatísticas de t , cuja indeterminação é sinal claro de erros de identificação do modelo. Outro sinal de erros de identificação do modelo é a existência de erros padrão muito grandes ou a instabilidade dos parâmetros para valores iniciais diferentes ou quando o número de extração para simulação aumenta (sinal de a matriz hessiana ser próxima de singular). Ainda assim, existem evidências empíricas que os modelos Logit mistos não necessitam de tantas restrições para ser identificáveis como, por exemplo, os Probit, sugerindo Walker (2002) mesmo assim a sua normalização pois a diferença entre os resultados não apresenta valores que justifiquem a alteração da metodologia de identificação.

Nas restrições a aplicar a modelos Logit mistos deve atender-se que os mesmos estudos empíricos apontam que a escolha do elemento das perturbações da parte Probit a anular não seja arbitrária. Sugere-se que se escolha a alternativa, ou uma delas, com variância mínima para alternativa de base, e portanto, a anular.

Para evitar problemas de estimação devido a amostras muito reduzidas, a fracas rotinas de otimização, a valores iniciais errados, a problemas relacionados com a escala, ou com um número insuficiente de extrações nas simulações, sugere-se que se verifique a estabilidade da estimação dos parâmetros para sucessivos incrementos nas extrações e com a alteração dos valores iniciais de estimação.

Assim, para que a integração seja verosímil, existem ainda outros aspetos a definir *a priori*, nomeadamente (Hensher *et al.* 2003, 2005 - ver também Anexo 11 do Capítulo III): a seleção dos parâmetros aleatórios considerados mais importantes; as suas funções de distribuição; o número de extrações a realizar na simulação; a possibilidade de integrar heterogeneidade de preferências nos parâmetros aleatórios; a possibilidade de correlações entre situações de escolha e/ou entre parâmetros; os problemas inerentes à consideração da disponibilidade para pagarem; e a existência de atributos endógenos (Train, 2009; Louviere *et al.*, 2005; Guevara e Ben-Akiva, 2009). Destes problemas, a seleção dos parâmetros aleatórios, as suas funções de distribuição e o número de extração a realizar, são comuns a todas as modelações, podendo as restantes existir ou não.

3.4.7. Simulação

Os modelos Logit mistos são adequados para a estimação através de métodos de simulação. Atendendo à função utilidade que tem coeficientes aleatórios distribuídos com densidade $f(\theta)$ definida pelo analista, a estimação pretenderá conhecer os parâmetros que definem $f(\theta)$, por exemplo a média e a variância dos θ . O procedimento é análogo à simulação para os Probit:

1. Escolha aleatória de um θ da distribuição $f(\theta|\beta)$, designando-o de θ^r para a r -ésima escolha;
2. Cálculo da sua probabilidade Logit $L_{ni}(\theta^r)$;
3. Repetição dos passos anteriores R vezes;
4. A média ponderada das R probabilidades calculadas será a probabilidade de escolha estimada:

$$\bar{P}_{ni} = \frac{1}{R} \sum_{r=1}^R L_{ni}(\theta^r)$$

Exp. III- 110

Esta média ponderada é uma estimativa não enviesada de P_{ni} , tem uma variância que diminui à medida que R aumenta, é estritamente positiva (propriedade útil para a estimação dos parâmetros pela maximização da verosimilhança), é suave (permite a sua segunda derivação) em relação aos parâmetros θ e às variáveis x , e a soma das probabilidades de todas as alternativas é igual à unidade, útil para realizar previsões (Train, 2003).

Estas probabilidades são depois aplicadas na função de verosimilhança logarítmica para resultar na verosimilhança logarítmica simulada (SLL):

$$SLL = \sum_{n=1}^N \sum_{j=1}^J d_{nj} \cdot \ln(\tilde{P}_{nj}) \quad \text{Exp. III- 111}$$

Onde d_{nj} tem o mesmo significado do δ_{in} da MVL para os modelos Logit. A maximização desta função permite determinar os θ que possibilitam o cálculo das probabilidades Logit mistas. É comum realizar várias extrações para cada observação de modo a manter a independência das probabilidades simuladas em relação aos decisores.

Ortúzar e Willumsen (2001) referem que se se utilizarem valores aleatórios padrão, ou pseudoaleatórios, para o cálculo de probabilidades, que o número de repetições a efetuar seja no mínimo de 1 000. Ortúzar e Willumsen (2001) e Hensher *et al.* (2003) referem ainda estudos que usando sequências de Halton conseguiram apenas com 125 repetições uma precisão semelhante a 2 000 repetições com sequências aleatórias padrão (ver Anexo 12 do Capítulo III).

4. Integração de Monte Carlo e Sequências de Extração

Modelos sem forma definida, como o Probit ou o Logit misto, necessitam de uma avaliação numérica que na prática tem sido realizada através de simulação de extrações aleatórias de determinadas distribuições. Os métodos numéricos mais comuns para a estimação de integrais múltiplos podem ser divididos em duas categorias: métodos de quadratura polinomial (“*polynomial cubature*”) ou métodos de simulação de Monte Carlo (Silva e Garrido, 2003).

O primeiro tipo de modelos tem aplicação limitada a especificações de modelos Logit mistos de pequena dimensão (poucos atributos).

4.1. Integração de Monte Carlo

Os métodos de simulação de Monte Carlo padrão simulam integrais multidimensionais substituindo a média contínua (o valor real do integral) por uma discreta utilizando uma série de números (pseudo) aleatórios selecionados de modo a cobrir um hipercubo definido (que contém o espaço de integração) cuja distribuição nesse espaço é posteriormente utilizada para determinar a média.

Se se considerar o seguinte exemplo bidimensional:

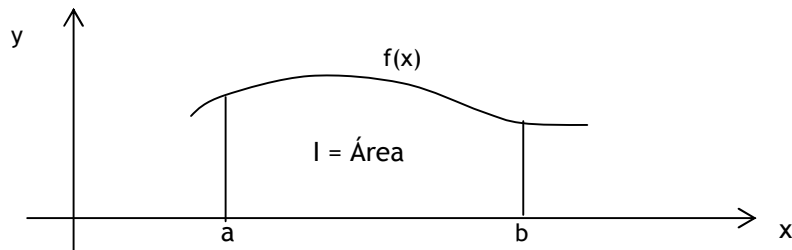


Figura III. 7 - Exemplo de cálculo do integral bidimensional (Área) de uma função

O método numérico (analítico) de determinar a área I , isto é de calcular o integral, pode ser calculado aproximadamente como uma soma (Krykova, 2003):

$$I = \int_a^b f(x)dx \approx \frac{(b-a)}{N} \sum_{i=1}^N f(x_i) \quad \text{Exp. III- 112}$$

A estimação de Monte Carlo aproxima I através de muitas amostras aleatórias x_i e pondera a média das suas contribuições $f(x_i)$. Pode assim generalizar-se que o método de Monte Carlo resolve integrais multidimensionais através da aproximação:

$$\theta = \int_B \phi(x)f(x)dx \approx \frac{1}{N} \sum_{\substack{i=1 \\ x_i \in B}}^N \phi(x_i) \quad \text{Exp. III- 113}$$

Onde x_i , com $i = 1, \dots, N$; são amostras aleatórias independentes da distribuição $f(x)$ que são obtidas de N amostras aleatórias independentes com distribuição uniforme $[0,1]^S$, em que S é a dimensão do cubo unitário, $B \in \mathcal{H}^S$ (Krykova, 2003). Em termos computacionais os valores aleatórios necessários ao método de Monte Carlo são gerados por rotinas determinísticas, pelo que na realidade são valores pseudoaleatórios.

As vantagens do método de Monte Carlo, com base na sua simplicidade de implementação e sentido intuitivo, podem ser definidas como (Krykova, 2003):

- Facilidade de implementação e aplicabilidade a problemas de integração complexos. Existe apenas a necessidade de obter pontos amostrais aleatórios x e avaliar a função $f(x)$ nesses pontos;
- Permite a integração de problemas com muitas dimensões;

- Os erros padrão não dependem da dimensão do integral ao contrário da maioria dos métodos numéricos;
- Permitem replicar fenómenos extremamente complicados utilizando geradores de valores aleatórios para simular flutuações normais na natureza.

No entanto, o método de Monte Carlo não é a solução para todos os problemas, apresentando algumas limitações das quais se salientam (Krykova, 2003):

- Existe a necessidade de decidir previamente quantos pontos escolher e quão pormenorizados devem ser;
- A sua convergência é lenta, com uma taxa (ou rapidez com que o erro diminui com o aumento do tamanho da amostra) proporcional a $1/\sqrt{N}$, que significa que para diminuir o erro para metade se tem que aumentar quatro vezes a amostra;
- Os resultados são probabilísticos podendo a estimativa do integral estar errada, havendo apenas limites probabilísticos para esses erros;
- O limite de erro (ou intervalo de confiança) não reflete qualquer regularidade adicional do integrando;
- O método depende do valor inicial da estimação.

A precisão dos métodos de simulação de Monte Carlo depende da cobertura e do número de pontos considerados na região de integração. O aspeto mais importante neste tipo de simulação é a localização dos pontos selecionados de forma a cobrir eficientemente o espaço de integração. Para que se alcance um erro de simulação na estimação de parâmetros razoavelmente baixo, que corresponde a cobrir o espaço de integração de forma equitativa e completa, há que realizar um grande número de extrações que podem ser morosos até para modelos de tamanho médio (Train, 2000; Silva e Garrido, 2003).

4.2. Sequências Não Aleatórias

Pesquisas noutras áreas de investigação, fora da área dos transportes, têm experimentado séries não aleatórias na simulação de Monte Carlo - designadas de simulações Quasi-Monte Carlo -, cuja definição determinística pretende cobrir de forma mais uniforme a região de integração. Estas séries são geralmente designadas de *baixa discrepância*, *sequências quasi-aleatórias* ou *extrações inteligentes*, escolhidas deterministicamente com base numa distribuição equitativa de forma a minimizar os erros. Em geral esta alteração permite que a estimação do integral convirja para a solução em $(\ln N)^S / N$, onde S é o número de dimensões do integral, em vez dos usais $1/\sqrt{N}$. Na realidade o termo “quasi-aleatório” é enganador uma vez que não têm nada de aleatório, sendo estes determinados através algoritmos que

pretendem evitar ao máximo a aproximação dos diferentes valores, podendo dizer-se que os métodos Quasi-Monte Carlo (QMC) são a versão determinística do Monte Carlo (MC) tradicional (Krykova, 2003).

As sequências de baixa discrepância são um conjunto de pontos s -dimensionais, que preenchem a área da amostra de forma mais eficiente e com menor discrepância que um conjunto de valores pseudoaleatórios comuns. Estas sequências permitem que os métodos QMC tenham vantagens em relação ao modelo tradicional (Krykova, 2003):

- O integral é aproximado utilizando uma sequência de pontos bem definida;
- Em geral, os QMC permitem melhores estimativas dos pontos para um esforço computacional similar comparativamente aos MC tradicionais;
- Os valores quasi-aleatórios permitem uma convergência mais rápida, o que se traduz na necessidade de menor número de extrações para obter o mesmo nível de precisão.

As sequências de baixa discrepância são baseadas nas sequências de *Van der Corput*, apresentando reduzidos erros de integração, sendo as mais comuns as sequências de Hammersley, de Niederreiter, de Halton, de Sobol, e de Faure (Silva e Garrido, 2003; Krykova, 2003). Destas, as que maior aplicação prática têm tido, principalmente na área de Transportes, são as três últimas, quer na sua forma original ou com modificações.

A utilização de extrações inteligentes em vez das aleatórias comuns tem apresentado um elevado potencial na diminuição do tempo de processamento, através da diminuição do número de extrações necessárias para obter resultados aceitáveis. Este procedimento é tanto mais rápido quanto menor for o ciclo das extrações, pelo que sequências com base em números primos menores são mais rápidas que as baseadas em valores de dimensão elevada.

As sequências de *Van der Corput* são as mais simples das sequências de baixa discrepância unidimensionais. Para obter o n -ésimo ponto x_n com base no primo p , tem que se decompor o inteiro n na base p (Krykova, 2003):

$$n = \sum_{i=0}^l a_i(n) \cdot p^i \quad \text{Exp. III- 114}$$

O número quasi-aleatório é obtido pela transposição dos dígitos de $a_i(n)$ para valores decimais:

$$x_n = \Phi_p(n) = \sum_{i=0}^l \frac{a_i(n)}{p^{i+1}} \quad \text{Exp. III- 115}$$

Como exemplo considere-se o primo $p = 3$ e o número $n = 19$. O número 19 na base 3 será $19 = 2.3^2 + 0.3^1 + 1.3^0 = 201$, que refletido em valores decimais será:

$$x_{19} = \Phi_3(19) = \frac{2}{27} + \frac{0}{9} + \frac{1}{3} = \frac{11}{27} = 0,407407\dots$$

Que está no intervalo $[0,1]$.

A sequência com base no número primo 2 será:

$$\begin{aligned} n = 1; 1 = 1.2^0 = 1; \quad x_1 = \Phi_2(1) &= \frac{1}{2} = 0,5 \\ n = 2; 1 = 1.2^1 + 0.2^0 = 2; \quad x_2 = \Phi_2(2) &= \frac{0}{2} + \frac{1}{4} = 0,25 \\ n = 3; 1 = 1.2^1 + 1.2^0 = 21; \quad x_3 = \Phi_2(3) &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4} = 0,75 \\ n = 4; 1 = 1.2^2 + 0.2^1 + 0.2^0 = 100; \quad x_4 = \Phi_2(4) &= \frac{0}{2} + \frac{0}{4} + \frac{1}{8} = 0,125 \\ &\vdots \end{aligned}$$

Neste caso os primeiros dois ciclos de base 2 serão $(0, 1/2)$ e $(1/4, 3/4)$.

Um tipo destas extrações inteligentes são as sequências de *Halton*, que permitem variâncias de simulação para parâmetros estimados através de 100 valores de *Halton* razoavelmente menores que com 1000 extrações aleatórias (Train, 2000). No entanto, Chiou e Walker (2006) apresentam estudos em que se verificou que a aplicação prática de modelos de simulação com 1000 extrações tanto aleatórias, como de Halton, e mesmo de Halton baralhadas (“*shuffled*”), podia gerar resultados enganadores, revelando que o número de 200 extrações, tipicamente aplicadas nos trabalhos empíricos com modelos Logit mistos, pode ser insuficiente para uma correta estimação dos parâmetros. A caracterização das sequências de *Halton* é apresentada no Anexo 12 do Capítulo III.

5. Aplicação de Modelos de Escolha Discreta em Estudos Reais de Transporte

A utilização de modelos de escolha discreta numa grande variedade de situações reais é vasta a nível internacional, demonstração clara da sua versatilidade, fiabilidade e coerência de resultados. A sua aplicação a problemas de transporte também é referida por vários autores de renome internacional, mas a maioria das suas aplicações utiliza modelos Logit multinomiais. Apesar das suas potencialidades, a utilização de modelos Logit Mistos tem-se centrado principalmente em estudos teóricos, verificando-se na última década o surgimento de estudos com a sua aplicação a casos práticos. No quadro III.1 apresentam-se algumas dessas aplicações práticas a problemas de escolhas modais, encontradas a nível internacional,

cujos resultados servirão de valores de referência para comparação com os obtidos nos modelos sugeridos no presente estudo.

É possível constatar que o número de variáveis analisadas varia entre duas e doze, e que o número de alternativas também varia entre duas e dez (neste último caso cinco modos e as respectivas combinações entre estes). A maioria das especificações consideradas para as variáveis é a contínua, e as variáveis relacionadas com custos são especificadas de diversas e distintas formas, relevando ser extremamente particularizadas para cada situação. A maioria das variáveis referidas são para a caracterização socioeconómica dos indivíduos, verificando-se a utilização de algumas potenciais variáveis caracterizadoras de políticas de ordenamento do território ou de mobilidade urbana, nomeadamente:

- Origem/destino da viagem em grandes cidades/Destino no centro histórico;
- Frequência média horária/Regularidade dos Bus;
- Viagem pendular;
- Existência de ligação direta por TC/Número de transbordos/Tempo de espera para transbordos;
- Densidade de empregos na zona de destino;
- Mora no CBD;
- Distância da viagem;
- Custo de estacionamento.

A sua utilização não é, no entanto, sistemática nem extensiva, e a sua combinação, nomeadamente as relacionadas com restrições ao livre uso do automóvel, praticamente inexistentes.

Quadro III. 1 - Estudos internacionais sobre escolhas modais com modelos de escolha discreta

Tipo de estudo	Modelo utilizado	Modos considerados	Variáveis analisadas	Tipologia	Unidades	Amostra	Referência
Modo de transporte	MNL	- Train - Air	Rendimento	Continua	/\$CA1000	2000	Bhat, 1998a
			Viagem sozinho	Binária	-		
			Sexo Feminino		-		
			OD em Grandes cidades		-		
			Custo da viagem	Continua / Genérica	\$CA		
			Tempo de viagem no veículo		minutos		
			Tempo de viagem fora do veículo		minutos		
			Frequência		Partidas/dia		
Modo de Transporte para lazer	MNL	- Shared Auto - TC	Automóveis/Adultos do agregado	Continua	Veic./hab.	3000	Bhat, 1998b
			Presença de crianças no agregado	Binária	-		
			Sexo feminino		-		
			Custo da viagem	Continua / Genérica	\$US cent		
			Tempo total de viagem		minutos		
			Tempo de viagem fora do veículo/distância		min./milha		
Modo de transporte para o emprego	MNL	- Auto - TC	Rendimento	Continua	/\$US1000	1617	Bhat, 2000
			Automóveis/trabalhador		Veic./hab		
			Custo da viagem		\$US		
			Tempo total de viagem		Minutos		
			Tempo de viagem fora do veículo/distância		Min./milha		
			Ln(tempo de viagem equivalente dentro do veículo)		Minutos		
Zona de recreio escolhida	MNL (parâmetros não lineares)	Não discriminado	Tempo de viagem equivalente dentro do veículo	Continua	Veic./hab	7770	Pozsgay e Bhat, 2002
			Número de veículos per capita no agregado		anos		
			Idade		-		
			Jovens (5-21 anos) no agregado	Binária	-		
			Viagem sozinho		-		
			Agregado com baixo rendimentos anuais (<€45 000/ano)		-		
			Trabalhador		-		
Modo de transporte	MNL	- Auto - TC	Tempo de viagem	Continua	Horas	1067	Caschetta e Papola, 2003
			Custo da viagem		€		
			Frequência média diária		TC/h		
			Destino ao centro histórico	Binária	-		
			Viagem pendular		-		
			Ligação directa por TC		-		

Quadro III. 2 - Estudos internacionais sobre escolhas modais com modelos de escolha discreta
(continuação)

Tipo de estudo	Modelo utilizado	Modos considerados	Variáveis analisadas	Tipologia	Unidades	Amostra	Referência
Modo de transporte para o emprego	MNL	- Auto - Bus+Auto - Carpool - Bus+Walk	Custo da viagem /rendimento líquido	Continua	\$/cent/min.	771	Train, 2003
			Tempo dentro do veículo		Minutos		
			Tempo a andar				
			Tempo de espera para transbordo				
			Regularidade dos BUS				
			Rendimento Familiar		\$US		
			Número de transbordos		-		
			Número de condutores no agregado		condutores		
			Número de automóveis/conductor		Veic./cond.		
			Densidade de empregos na área de trabalho		?		
Modo de transporte para o emprego	MNL	- Shared Auto - TC - Bike - Walk	Cabeça de casal é trabalhador	Binária	-	5029	Koppelman e Bhat, 2006
			Mora no CBD				
			Tempo de viagem		minutos		
			Custo da viagem		\$US cent		
Modo de transporte para compras/Outro	MNL	- Shared Auto - TC - Bike - Walk	Rendimento	Continua / Genérica	\$/US 1000	3157	Koppelman e Bhat, 2006
			Custo da viagem				
			Tempo de viagem		\$US cent		
			Tempo de viagem fora do veículo/distância		minutos		
			Rendimento do agregado		Min./milhas		
			Pessoas do agregado		\$/US 1000		
			Veículos no agregado per capita		habitantes		
			Agregado sem veículo		Veic./hab.		
			Destino ao CBD		-		
			Idade				
Modo de transporte para o emprego	MNL	- Auto - TC	Rendimento anual individual para TC	Discreta (média da gama)	Anos	1840	Hensher e Rose, 2007
			Sexo		\$/SAU 1000		
			Tamanho do agregado		1-M; 0-F		
			Número de crianças no agregado		Habitantes		
			Custo da viagem		SAU		
			Tempo de viagem dentro do veículo				
			Tempo de acesso ao TC		minutos		
			Tempo de espera para TC				
			Tempo de saída				

Quadro III. 3 - Estudos internacionais sobre escolhas modais com modelos de escolha discreta
(continuação)

Tipo de estudo	Modelo utilizado	Modos considerados	Variáveis analisadas	Tipologia	Unidades	Amostra	Referência
Modo de transporte para outras viagens	MNL	- Auto - TC	Idade Rendimento anual individual para TC Sexo Tamanho do agregado Número de crianças no agregado Custo da viagem Custo do estacionamento Tempo de viagem dentro do veículo Tempo de acesso ao TC Tempo de espera para TC Tempo de saída	Discreta (média da gama) Discreta (máximo da gama) Binária Contínua	Anos \$/SAU 1000 - habitantes \$AU minutos	1941	Hensher e Rose, 2007
		- Auto - Metro - Bus - Táxi - Comb	Tempo de viagem Custo da viagem	Contínua / Genérica	minutos \$US		
		- TC - Bike - Walk - Auto	Distância Disponibilidade diária de AUTO Mau tempo	Contínua Binária	km -		
		- Walk - Bus - Bike	- Motivo da viagem - Sexo - Grau de Conforto - Profissão - Distância - Custo da viagem	Binária / Genérica Discreta / Genérica Contínua / Genérica	- 3 níveis km ¥RMB €		
		10 modos: - Condutor - Passageiro - Táxi - Metro - Bus - Combinações	- Custo da viagem/salário horário - Tempo de viagem - Tempo de espera - Tempo a andar a pé - Número de transbordos - Número de automóveis - Sexo - Rendimento anual - Licença de condução - Tempo de viagem - Custo da viagem - Distância	Contínua / Genérica	min -		
		- Auto - Bus		Binária	-		
				Contínua	minutos Yuan/km km		
Modo de transporte a escola	MNL	- TC - Bike - Walk - Auto	Distância Disponibilidade diária de AUTO Mau tempo	Contínua Binária	km -	9300	Müller et al., 2008
Modo de transporte para viagens pendulares	ML	- Walk - Bus - Bike	- Motivo da viagem - Sexo - Grau de Conforto - Profissão - Distância - Custo da viagem	Binária / Genérica Discreta / Genérica Contínua / Genérica	- 3 níveis km ¥RMB €	664	Li et al. (2010)
Modo de transporte	ML	10 modos: - Condutor - Passageiro - Táxi - Metro - Bus - Combinações	- Custo da viagem/salário horário - Tempo de viagem - Tempo de espera - Tempo a andar a pé - Número de transbordos - Número de automóveis - Sexo - Rendimento anual - Licença de condução - Tempo de viagem - Custo da viagem - Distância	Contínua / Genérica	min -	1290	Yáñez et al. (2010)
Modo de transporte	MNL	- Auto - Bus		Binária	-	986	Bao et al. (2010)
				Contínua	minutos Yuan/km km		

6. Síntese

Se a regressão linear múltipla se apresenta como uma ferramenta com enorme potencial em muitos ramos da engenharia, a sua limitação a variáveis contínuas inviabiliza a sua utilização em situações em que o conjunto de possíveis respostas é bastante conciso. Assim nos finais da década de 70 surgiram os modelos de escolha discreta que, nomeadamente através dos trabalhos de Daniel MacFadden (Train, 2003) com uma base teórica consecutivamente consolidada, se apresentaram como ferramentas potentes para a modelação das escolhas de indivíduos pelo facto de permitirem modelar escolhas com base em variáveis contínuas mas também discretas, isto é, seleccionadas de um conjunto finito de opções.

Nos modelos de escolha discreta assume-se que cada alternativa tem associada uma utilidade bruta para um determinado indivíduo, e que o analista não possui a informação completa sobre os elementos considerados para a escolha de uma alternativa. Assim, o analista assume que a utilidade é representada por dois componentes: uma mensurável, determinística, função de atributos mensuráveis e outra parte aleatória, probabilística, que reflete o gosto pessoal de cada indivíduo, agregado a eventuais erros de observação ou medição. A forma como o analista considera esta última parte define o tipo de modelo a utilizar.

A probabilidade de um indivíduo escolher a opção será o valor esperado de uma função indicadora considerando todos os possíveis valores dos fatores não observáveis. Para calcular esta probabilidade, o respectivo integral pode ser avaliado de três maneiras: através de expressões de forma completamente fixa (por exemplo os Logit multinomiais), simulação completa (por exemplo os Probit) e através de simulação e forma fixa parciais (por exemplo os Logit mistos).

Para a especificação e estimação dos modelos de escolha discreta há que atender que apenas as diferenças entre utilidades importam e que a escala da utilidade é arbitrária. Assim, existem algumas restrições gerais que se devem impor a todos os modelos de escolha discreta, nomeadamente a normalização da utilidade e da escala. Esta normalização caracteriza-se por considerar apenas a diferença entre utilidades de uma alternativa em relação às restantes, ou melhor, a diferença entre os erros de uma utilidade em relação aos erros das restantes, bem como à imposição de restrições a determinados elementos, geralmente as variâncias e covariâncias, de modo a tornar o modelo identificável e assim estimável.

Uma restrição usual é a definição do valor da constante específica de determinada alternativa entre as disponíveis (geralmente igualando-a a zero). Uma constante específica de uma alternativa apreende o efeito médio de todos os fatores não incluídos no modelo, semelhante à constante dos modelos de regressão. Se esta constante for incluída, a parte não mensurável da utilidade (ε) terá média igual a zero por definição. Para a normalização da escala das

diferenças entre utilidades deve atender-se que normalizar a variância dos erros é equivalente a normalizar a escala da utilidade e que depende da distribuição de probabilidades assumida para os erros. Como procedimento geral, recomenda-se a normalização da matriz de variâncias de diferenças de utilidade, geralmente através da definição do seu valor no canto superior esquerdo com valor igual à unidade e à divisão das restantes variâncias e covariâncias pela expressão do elemento da variância normalizado.

Assumindo que a escolha de cada indivíduo é independente das escolhas de todos os outros, depois de definida uma função matemática específica para a probabilidade de escolha segundo o modelo escolhido, a estimação dos valores dos pesos/coeficientes dos atributos é geralmente realizada através da utilização do método *da Máxima Verosimilhança Logarítmica*. Este método iterativo necessita de um algoritmo que permita determinar o vetor de coeficientes que o maximize, recomendando-se o algoritmo *Newton-Raphson*.

Dos diversos modelos de escolha discreta desenvolvidos desde os anos 70, o mais simples e difundido é o Logit Multinomial. Este modelo assume o pressuposto de *I/A*, que os seus erros são *IID* com uma distribuição de valores extremos do tipo I ou de *Gumbel* e que as variáveis são igualmente valorizadas por toda a população em estudo. Estas simplificações tornam o seu uso adequado a estudos para avaliação de valores de base para ajustamentos dos modelos de escolha discreta e para os valores dos parâmetros para as diversas variáveis em análise. A facilidade de normalização de utilidades e de escala, a sua forma fixa e a fácil interpretação dos seus resultados são características que fomentaram a sua utilização nas últimas décadas. No entanto, estas mesmas simplificações, tornam-nos inapropriados para situações reais mais complexas, nomeadamente quando os seus erros não são *IID*, quando há correlações entre variáveis e/ou entre alternativas ou quando há variações de gosto por parte de cada indivíduo e/ou ao longo do tempo.

Para situações mais complexas o modelo Logit Misto apresenta-se o mais promissor uma vez que tem uma grande flexibilidade e consegue aproximar qualquer modelo de utilidade aleatória. Permite variações de gosto aleatórias (coeficientes dos atributos com valores diferentes entre indivíduos), padrões de substituição sem restrições (não impõem as restrições resultantes da *I/A*) e correlações entre erros ao longo do tempo. Não é restringido a erros com distribuição normal como o Probit e a simulação das suas probabilidades de escolha é computacionalmente relativamente simples.

Os modelos Logit mistos são definidos com base na forma funcional das suas probabilidades de escolha, e são representados por integrais de probabilidades Logit padrão sobre a densidade dos parâmetros. Os erros aleatórios têm uma distribuição *IID* de *Gumbel* independentes tanto dos parâmetros como dos atributos, e os coeficientes dos atributos têm distribuição definida pelo analista.

A forma mais expedita de determinar as probabilidades de escolha segundo um modelo Logit Misto, a utilizar na maximização da verosimilhança logarítmica, é através da simulação do integral através do método de Monte Carlo (ou Quasi-Monte Carlo).

Apesar de ainda existir necessidade de uma melhor fundamentação teórica, nomeadamente para a normalização e identificabilidade dos modelos Logit Mistos, os mesmos apresentam-se como uma ferramenta poderosa que permite a inclusão de diferentes fatores na probabilidade de escolha, através de abordagens computacionalmente pouco complicadas. De facto, estudos empíricos já permitem definir um conjunto de regras para que a estimação através de Logit mistos seja coerente e mais fidedigna.

No entanto, atendendo ao esforço necessário para uma correta definição de um modelo de escolha modal, nomeadamente na seleção das variáveis mais relevantes no processo, a utilização de modelos Logit Mistos apenas é aconselhada após extensas análises exploratórias com modelos mais céleres (Hensher *et al.*, 2005). Atendendo à facilidade de aplicação dos modelos Logit Multinomiais, tanto ao nível do tempo de estimação, quer ao nível da interpretação dos seus resultados, é comum recorrer aos mesmos para essas análises exploratórias fundamentais. A sua utilização permite uma extensa exploração de quais as variáveis com maior potencial e qual a sua melhor especificação, permitindo avaliar a coerência e relevância das mesmas e fornecendo dados de base para comparação com os resultados de modelos mais complexos. Esta comparação é possível quer ao nível dos valores dos parâmetros de cada variável, quer em relação ao ajustamento global do modelo aos dados utilizados.

Capítulo IV

Informação de Base

1. Introdução

Tendo como objetivo avaliar o impacto de alterações socioeconómicas da população, sistémicas e externas aos sistemas de transporte, bem como a influência de diferentes medidas das políticas urbanas para uma mobilidade sustentável, procedeu-se a um caso de estudo na Área Metropolitana do Porto (AMP) no ano de 2000. Este estudo serviu para analisar o potencial destas medidas na influência das escolhas modais dos cidadãos nas suas deslocações diárias.

Optou-se pela área referida devido à existência de uma extensa base de dados sobre as características dos cidadãos e das suas viagens provenientes do Inquérito à Mobilidade da AMP realizado pelo INE (INE, 2002). Na caracterização do sistema de TC e da oferta de estacionamento no concelho do Porto foi utilizado um Sistema de Informação Geográfica (SIG) com dados recolhidos pela Faculdade de Engenharia da Universidade do Porto (FEUP) para a Autoridade Metropolitana de Transportes do Porto (AMTP) e para a Câmara Municipal do Porto (CMP) (Rodrigues *et al.*, 2005). Atendendo ao período temporal do inquérito, todos os dados recolhidos foram devidamente ajustamento de modo a caracterizar a realidade no ano 2000.

No presente capítulo faz-se uma caracterização da área de estudo, em termos demográficos, espaciais, principais acessos rodoviários existentes, redes de TC existentes e oferta de estacionamento nas áreas urbanas, nomeadamente no concelho do Porto. Nesta caracterização apresentam-se igualmente as simplificações assumidas devido às limitações encontradas nesses levantamentos. Atendendo aos dados recolhidos das diferentes fontes citadas apresentam-se as variáveis de base consideradas no estudo, caracterizando-as e apresentando o processamento realizado para viabilizar a sua utilização nos modelos de escolha modal. Identificam-se os problemas/fragilidades/simplificações das bases de dados existentes/utilizadas permitindo ilações para futuros estudos de transportes, de forma a garantir a melhor qualidade/utilidade dos dados recolhidos/tratados. No final do capítulo apresentam-se as diferentes amostras disponíveis na base de dados, considerando a existência de dados resultantes da combinação de diferentes variáveis.

2. Caracterização da AMP

2.1. População e Território Estudado

A AMP está dividida em 16 municípios (quadro IV.1), mas para aproveitar os dados estatísticos e as divisões administrativas utilizadas no INE e, atendendo a dificuldades que ocorreram na obtenção de dados sobre a oferta de TC, optou-se por apenas avaliar os municípios pertencentes à NUT III - Grande Porto (GP) (ver Anexo 1 do Capítulo IV).

Quadro IV. 1 - Municípios inseridos na AMP (INE, 2004)

Municípios da AMP	Nº de Freguesias
Arouca	20
Espinho	5
Gondomar	12
Maia	17
Matosinhos	10
Oliveira de Azeméis	19
Porto	15
Póvoa do Varzim	12
Santa Maria da Feira	31
Santo Tirso	24
São João da Madeira	1
Trofa	8
Vale de Cambra	9
Valongo	5
Vila do Conde	30
Vila Nova de Gaia	24

Quadro IV. 2 - Caracterização dos Municípios da NUT III - Grande Porto com dados de 2000 (INE, 2004)

Município da NUT III - Grande Porto	Nº de Freguesias	Area 2001 km2	População Residente Hab.	Densidade populacional Hab/km2
Espinho	5	21,1	33 701	1595,7
Gondomar	12	131,4	164 096	1248,5
Maia	17	83,2	120 111	1443,3
Matosinhos	10	61,9	167 026	2697,4
Porto	15	41,6	263 131	6332,9
Póvoa do Varzim	12	82,1	63 470	773,2
Valongo	5	75,8	86 005	1134,3
Vila do Conde	30	149,0	74 391	499,2
Vila Nova de Gaia	24	168,7	288 749	1711,8
Total/Média	130	814,8	1 260 680	1937,4

É possível verificar que em 2000 a população total residente na NUT III - Grande Porto era de 1 260 680 habitantes numa área de 815 km² resultando numa densidade populacional média de 1937 hab./km² (ver quadro IV.2). Esta densidade é muito superior à densidade populacional média nacional em 2000 de 110,8 hab./km² (INE, 2004).

2.2. Rede de Transportes

2.2.1. Rede Viária

Em relação à rede viária a AMP, em 2000, era servida por várias vias que faziam a ligação entre os seus municípios na direção Norte-Sul são (ver figura IV.1):

- IP1 (atual IP1/A20)
- A1 (atual A1/IC1)
- IC1/N107 (atual A28)
- A3 (atual A3/IP1)
- N14 (atual Via Norte)

E na direção este-oeste:

- IC24 (atual IC24/A41)
- N12
- A3 (atual IC23/VCI)
- A4 (atual A4/IP4)

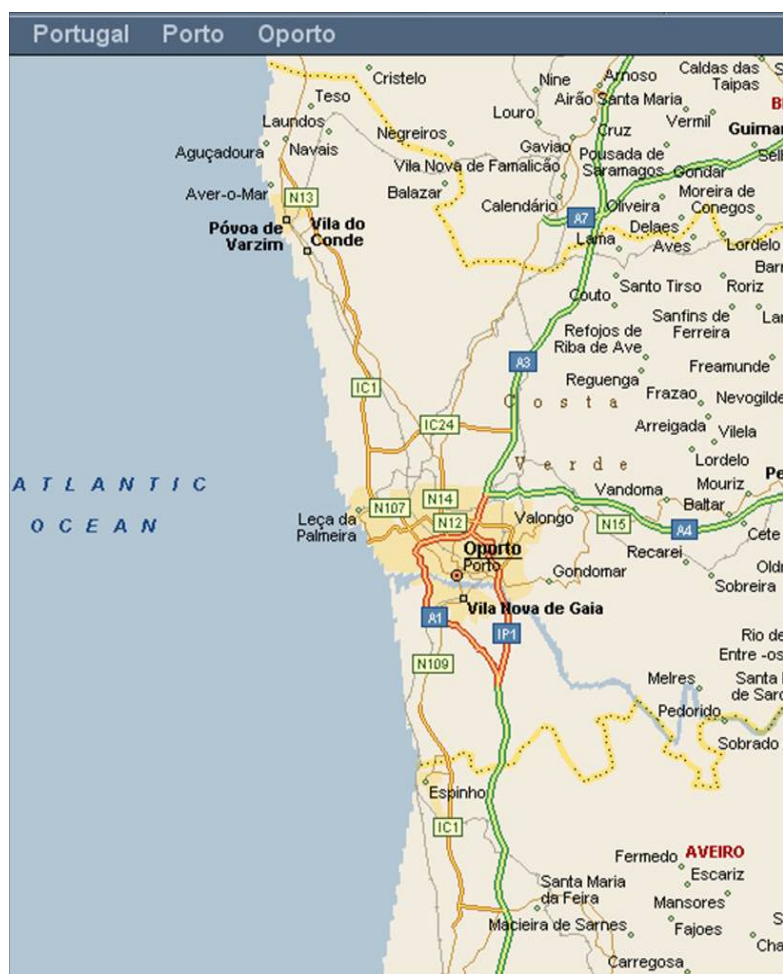


Figura IV. 1 - Rede viária principal na AMP em 2000 (Microsoft Corporation, 2001)

2.2.2. Rede de Transportes Coletivos

A rede de TC que servia a AMP em 2000 era composta por operadores de autocarros privados e públicos e pelo transporte ferroviário. Em termos de transporte ferroviário a oferta limitava-se à linha do Norte que servia os municípios de Espinho, Vila Nova de Gaia, Porto, Maia e Valongo, e à Linha da Póvoa (entretanto reestruturada em 2002 para linha de Metro) que ligavam os municípios de Vila do Conde e Póvoa do Varzim à cidade do Porto. Da lista de cerca de 40 operadores de TC fornecida pelo Instituto da Mobilidade e dos Transportes Terrestres (IMTT)⁸ e pela CMP, que operavam na AMP em 2008, apenas se conseguiu informação, quer de carreiras georreferenciadas quer de horários, de 8 operadores, que se consideraram semelhantes aos serviços em 2000:

- A. Nogueira e Costa, Lda.
- Arriva Portugal Transportes, Lda.
- Autoviação Castelo da Maia, Lda.
- Empresa de Transportes Gondomarense, Lda.
- Maia Transportes, Lda.
- Resende - Atividades Turísticas, Lda.
- Valpi Bus
- Sociedade de Transportes Coletivos do Porto (STCP)

Os operadores referidos representam a totalidade dos operadores para determinados pares de geração-atração na parte norte da AMP, e na modelação da sua influência apenas esses pares foram considerados.

Os trajetos das diferentes carreiras foram obtidos através de um levantamento realizado pela FEUP em 2005 para a AMTP, com recurso a um SIG. Atendendo à escassez de tempo e de recursos humanos e financeiros não foi possível obter mais dados, pelo que se considerou que os trajetos das carreiras e horários não se tinham alterado substancialmente desde 2000. É de lamentar a renitência encontrada por parte de alguns operadores privados em ceder informação (pública) sobre os trajetos e horários adjudicados.

Em relação ao operador público STCP, o mesmo forneceu a rede em SIG e os horários de 2000. A rede total utilizada para o estudo é apresentada no Anexo 2 do Capítulo IV. Pode verificar-se que a informação recolhida cobre quase toda a área a norte da AMP.

2.2.3. Oferta de Estacionamento

Para a avaliação da oferta de estacionamento na AMP utilizou-se o levantamento realizado pela FEUP em 2005 para a AMTP (Rodrigues *et al.*, 2005), mas que apenas avaliou a oferta de estacionamento no concelho do Porto. O facto de o levantamento ter sido realizado em 2005,

⁸ - Dados fornecidos através de contacto estabelecido com a Eng.^a Maria Teresa Malheiro do Núcleo de Atividades e Transportes em Março de 2008

impôs a necessidade de avaliar quais as alterações existentes na oferta entre 2000 e 2005. Informação cedida pela CMP permitiu considerar que a oferta de estacionamento adjacente à via, embora podendo ter sofrido algumas alterações pontuais de localização, não sofreu grandes alterações em quantidade, considerando-se assim os valores de 2005. Não foi, no entanto, possível avaliar o estacionamento ilegal e informal, pelo que se optou por não incluir este tipo de estacionamento na oferta existente em 2000.

Para corrigir a oferta de estacionamento em parque para o ano 2000, fez-se uma análise do tamanho dos estacionamentos em parque em 2005, e concluiu-se que dos 366 parques avaliados, 55 (representando 59,5% da oferta levantada) tinham mais de 200 lugares. De modo a não tornar a análise demasiado demorada e dispendiosa, optou-se por apenas avaliar estes parques. Um levantamento local nos parques seleccionados permitiu avaliar as alterações relativamente ao ano 2000 (apenas em 4 não foi possível determinar quais as alterações). Para os restantes parques de menor dimensão assumiu-se que a oferta em 2000 era semelhante à existente em 2005.

No levantamento da oferta de estacionamento existente, foram diferenciadas várias tipologias de estacionamento que são descriminadas no quadro IV. 3.

Quadro IV. 3 - Tipologias de oferta de estacionamento consideradas (baseado em Rodrigues *et al.*, 2005)

OEVPUPP	Oferta de estacionamento na via pública de uso publico pago
OEVPUPI	Oferta de estacionamento na via pública de uso publico isento
OEVPUPrE	Oferta de estacionamento na via pública de uso privado para entidades
OEVPUPrA	Oferta de estacionamento na via pública de uso privado alugados
TOVP	Total de Oferta na via publica
OEPCond	Oferta de estacionamento em parque de acesso condicionado
OEPConcess	Oferta de estacionamento em parque concessionados pagos
OEPMUPI	Oferta de estacionamento em parque municipais de uso publico isentos
OEPMUPP	Oferta de estacionamento em parque municipais de uso publico pagos
OEPPUPI	Oferta de estacionamento em parque privados de uso publicos isentos
OEPPUPP	Oferta de estacionamento em parque privados de uso publicos pagos
TOPq	Total de Oferta em parques
OEOfertaTotal	Total de Oferta

Os resultados do levantamento da oferta de estacionamento e do seu ajustamento para o ano 2000 podem ser consultados no quadro IV. 4.

Apesar de o levantamento da oferta de estacionamento pela FEUP ter utilizado zonas elementares mais pequenas que as zonas utilizadas no Inquérito à Mobilidade do INE em 2000, o facto de não se conseguir relacionar as zonas de atração das viagens dos inquéritos com essas zonas elementares implicou a sua agregação em zonas maiores coincidentes com o zonamento do INE. Assim, no concelho do Porto o zonamento considerado utilizou as

Quadro IV. 4 - Oferta de estacionamento no município do Porto ajustados para o ano 2000

152

Assim, o concelho do porto foi dividido em 26 zonas de tráfego e para a análise da influência das variáveis relacionadas com o estacionamento apenas se consideraram as viagens com atração para as mesmas. Este facto deve ser considerado na análise dos resultados das variáveis sobre estacionamento, uma vez que a limitação de possíveis dados resultantes de combinações de apenas 26 valores originais resulta em parâmetros obtidos a partir de dados com variedade limitada e, por conseguinte, pouca riqueza para serem estatisticamente relevantes.

3. O Inquérito à Mobilidade da População Residente em 2000

Segundo INE (2002), o inquérito à mobilidade da população residente no ano 2000 foi realizado entre Abril e Junho de 2000 por entrevista direta e com inquéritos por amostragem, com unidade igual a subsecção estatística, área homogénea que costuma corresponder ao quarteirão em áreas urbanas e a lugares ou parte de lugares nas áreas rurais.

A amostra foi dimensionada para ser representativa ao nível da freguesia para os concelhos do Grande Porto e ao nível do concelho para os restantes concelhos, atendendo aos Censos de 1991, admitindo-se um erro de amostragem de aproximadamente 10% (por concelho ou freguesia) para um intervalo de confiança para o total de 95%. Para o Grande Porto o número de indivíduos da amostra foi de 137 804 (64,5%) num total de 213 727 indivíduos entrevistados em toda a AMP (INE, 2002).

Segundo o INE (2002), a amostra foi criada a partir de subsecções selecionadas aleatoriamente entre as subsecções pertencentes à área em estudo. O controlo foi realizado através da reinquirição de 5% da amostra com taxas de consistência na ordem dos 98%.

3.1. Caracterização da Informação Original Recolhida na Base de Dados

A utilização da base de dados do Inquérito à mobilidade de 2000 deveu-se à quantidade de informação disponível através de um organismo público credível, o INE. O inquérito envolveu um total de 213 727 inquéritos, correspondendo a 506 715 viagens realizadas em dias úteis na AMP. Os dados recolhidos, extrapolados para o universo em análise são equivalentes a 2 736 614 indivíduos e 6 801 867 viagens realizadas pelos mesmos.

Além do tamanho da base de dados, outro ponto que justificou a sua utilização foi a variedade de informação recolhida e a área de intervenção do inquérito. A recolha e tratamento estatístico de um estudo da dimensão do realizado pelo INE comportam custos financeiros e recursos humanos que inviabilizam a sua elaboração para estudos pontuais ou académicos.

Quadro IV. 5 - Variáveis recolhidas para caracterizar os Agregados Domésticos Privado (ADP)s
(INE, 2002)

APURADP	
TABELA DOS ADP'S	
CAMPO	DESCRIÇÃO DAS VARIÁVEIS
FDT	Código da Divisão Administrativa 1º nível, corresponde ao Distrito onde reside o ADP
FCC	Código da Divisão Administrativa 2º nível, corresponde ao Concelho de residência do ADP
FFR	Código da Divisão Administrativa 3º nível, corresponde à Freguesia de residência do ADP
Fsítio	Código do sítio de residência (tabela Tab00sítios do Inquérito à Mobilidade 2000) do ADP entrevistado
FUA	Código da Unidade de alojamento onde reside o ADP entrevistado
FADP	Nº de ordem do ADP residente na unidade de alojamento
Ponderador	Peso do ADP da amostra em relação ao total dos ADP's do estrato
FNumInd	Nº de indivíduos no ADP
FNVEÍCULOS	Nº de Automóveis ligeiros à disposição do ADP
FNMOTOCICLOS	Nº de motociclos/ciclomotres à disposição do ADP
FNAUTODiario	Nº de automóveis ligeiros à disposição do ADP com utilização diária
FNAUTOFS	Nº de automóveis ligeiros à disposição do ADP com utilização apenas aos fins de semana
FNAUTOPontual	Nº de automóveis ligeiros à disposição do ADP com utilização pontual
FNMOTODiario	Nº de motociclos/ciclomotres à disposição do ADP com utilização diária
FNMOTOFS	Nº de motociclos/ciclomotres à disposição do ADP com utilização apenas aos fins de semana
FNMOTOPontual	Nº de motociclos/ciclomotres à disposição do ADP com utilização pontual
FT10COD	Nível de rendimento líquido mensal do ADP, de acordo com a tabela Tab10Rend
FDMTTP	Despesa mensal do ADP com transportes públicos (escudos)
FDMPasseSimples	Despesa mensal do ADP com Passe simples
FDMPasseCombinado	Despesa mensal do ADP com Passe combinado
FDMCOMB	Despesa mensal do ADP com combustíveis
FDMESTAC	Despesa mensal do ADP com estacionamento
FDMOUTRAS	Outras despesas mensais do ADP relacionadas com o uso da viatura
FDAMV	Despesas anuais do ADP com a Manutenção da(s) viatura(s)
FDASV	Despesas anuais do ADP com seguros da(s) viatura(s)
FDAOUTRAS	Outras despesas anuais do ADP relacionadas com a utilização da viatura

A variedade de informação recolhida pelo inquérito à mobilidade incluiu três grupos distintos:

- Identificação e caracterização dos agregados domésticos privados (ADP) e de cada indivíduo;
- Caracterização das viagens e deslocações efetuadas por cada indivíduo residente na unidade de alojamento;
- Opinião sobre o sistema de transporte para indivíduos com 15 ou mais anos de idade;

Nos quadros IV.5, IV.6, IV.7 e IV.8 é apresentada a informação original recolhida nesses três grupos de informação.

Quadro IV. 6 - Variáveis recolhidas para caracterizar cada indivíduo (INE, 2002)

APURINDIV	
TABELA DE INDIVÍDUOS	
CAMPO	DESCRIÇÃO DAS VARIÁVEIS
IDT	Código da Divisão Administrativa 1º nível, corresponde ao Distrito onde reside o indivíduo entrevistado
ICC	Código da Divisão Administrativa 2º nível, corresponde ao Concelho de residência do indivíduo entrevistado
IFR	Código da Divisão Administrativa 3º nível, corresponde à Freguesia de residência do indivíduo entrevistado
Isitio	Código do sítio de residência (tabela Tab00sítios do Inquérito à Mobilidade 2000) do indivíduo entrevistado
IUA	Código da Unidade de alojamento onde reside o Indivíduo entrevistado
IADP	Nº de ordem do ADP residente na unidade de alojamento
IIND	Nº de ordem do indivíduo dentro do ADP
Ponderador	Peso do indivíduo da amostra em relação à população do estrato
ISEXO	Sexo do indivíduo (a variável assume os valores M ou F)
ICodIdade	Código correspondente ao escalão etário do indivíduo (tabela tabcodidades)
IT2COD	Código correspondente à relação de parentesco do indivíduo com o representante do ADP (tabela Tab02RelRepresent)
IT3COD	Código correspondente à Condição perante o Trabalho (tabela Tab03CPT)
IT4COD	Código correspondente ao sector de actividade da empresa onde o indivíduo trabalha (tabela Tab04sectoractiv)
IESTUDA	Resposta à pergunta "Estuda?", a variável assume os valores S ou N
IREFORMA	Resposta à pergunta "Reformado?", a variável assume os valores S ou N
IT5COD	Código correspondente ao nível de instrução do indivíduo (tabela Tab05Educ)
IsitioTrab	Código do sítio onde o indivíduo trabalha (de acordo com a tabela Tab00sítios)
IsitioEst	Código do sítio onde o indivíduo estuda (de acordo com a tabela Tab00sítios)
ICartaCond	Se o indivíduo tem Carta de Condução = S, Se o indivíduo não tem Carta de Condução = N
IT7FREQ	Resposta à pergunta E1, de acordo com a tabela Tab07ModoTranspFreq
IT11COD1	Código correspondente à prioridade 1 na resposta à pergunta E2 (motivo pelo qual utiliza principalmente o automóvel)
IT11COD2	Código correspondente à prioridade 2 na resposta à pergunta E2 (motivo pelo qual utiliza principalmente o automóvel)
IT11COD3	Código correspondente à prioridade 3 na resposta à pergunta E2 (motivo pelo qual utiliza principalmente o automóvel)
IE301Class	Opinião do indivíduo quanto ao tempo de percurso, de acordo com o quadro E3 do questionário
IE301Escala	Importância do atributo (tempo de percurso) para o indivíduo
IE302Class	Opinião do indivíduo quanto às condições de transbordo, de acordo com o quadro E3 do questionário
IE302Escala	Importância do atributo (condições de transbordo) para o indivíduo
IE303Class	Opinião do indivíduo quanto à frequência das circulações, de acordo com o quadro E3 do questionário
IE303Escala	Importância do atributo (frequência das circulações) para o indivíduo
IE304Class	Opinião do indivíduo quanto à segurança na circulação, de acordo com o quadro E3 do questionário
IE304Escala	Importância do atributo (segurança na circulação) para o indivíduo
IE305Class	Opinião do indivíduo quanto à segurança pessoas e bens, de acordo com o quadro E3 do questionário
IE305Escala	Importância do atributo (segurança pessoas e bens) para o indivíduo
IE306Class	Opinião do indivíduo quanto ao conforto, de acordo com o quadro E3 do questionário
IE306Escala	Importância do atributo (conforto) para o indivíduo
IE307Class	Opinião do indivíduo quanto ao preço, de acordo com o quadro E3 do questionário
IE307Escala	Importância do atributo (preço) para o indivíduo
IE308Class	Opinião do indivíduo quanto à facilidade de aceder ao transporte, de acordo com o quadro E3 do questionário
IE308Escala	Importância do atributo (facilidade de aceder ao transporte) para o indivíduo
IE309Class	Opinião do indivíduo quanto à ligação entre diferentes modos de transporte, de acordo com o quadro E3 do questionário
IE309Escala	Importância do atributo (ligação entre diferentes modos de transporte) para o indivíduo
IE310Class	Opinião do indivíduo quanto ao atendimento, de acordo com o quadro E3 do questionário
IE310Escala	Importância do atributo (atendimento) para o indivíduo
IE311Class	Opinião do indivíduo quanto à pontualidade, de acordo com o quadro E3 do questionário
IE311Escala	Importância do atributo (pontualidade) para o indivíduo
IE312Class	Opinião do indivíduo quanto à adequação de percurso, de acordo com o quadro E3 do questionário
IE312Escala	Importância do atributo (adequação de percursos) para o indivíduo
IE313Class	Apreciação global do serviço de transportes públicos, de acordo com o quadro E3 do questionário
IE313Escala	Importância dos transportes públicos para o indivíduo
IE41Class	Concorda (S) ou não concorda (N) com a 1ª medida apresentada no quadro E4
IE41Escala	Importância que o indivíduo atribui à 1ª medida apresentada no quadro E4
IE42Class	Concorda (S) ou não concorda (N) com a 2ª medida apresentada no quadro E4
IE42Escala	Importância que o indivíduo atribui à 2ª medida apresentada no quadro E4
IE43Class	Concorda (S) ou não concorda (N) com a 3ª medida apresentada no quadro E4
IE43Escala	Importância que o indivíduo atribui à 3ª medida apresentada no quadro E4
IE44Class	Concorda (S) ou não concorda (N) com a 4ª medida apresentada no quadro E4
IE44Escala	Importância que o indivíduo atribui à 4ª medida apresentada no quadro E4
IE45Class	Concorda (S) ou não concorda (N) com a 5ª medida apresentada no quadro E4
IE45Escala	Importância que o indivíduo atribui à 5ª medida apresentada no quadro E4
IE46Class	Concorda (S) ou não concorda (N) com a 6ª medida apresentada no quadro E4

Quadro IV. 7 - Variáveis recolhidas para caracterizar as viagens (INE, 2002)

APURVIAGESEM	
TABELA DAS VIAGENS NUM DIA UTIL	
CAMPO	DESCRIÇÃO DAS VARIÁVEIS
VSEMDT	Código da Divisão Administrativa 1º nível, corresponde ao Distrito onde reside o indivíduo entrevistado
VSEMCC	Código da Divisão Administrativa 2º nível, corresponde ao Concelho de residência do indivíduo entrevistado
VSEMFRR	Código da Divisão Administrativa 3º nível, corresponde à Freguesia de residência do indivíduo entrevistado*
VSEMSítio	Código do sítio de residência (tabela Tab00sítios do Inquérito à Mobilidade 2000) do indivíduo entrevistado
VSEMUUA	Código da Unidade de alojamento onde reside o indivíduo entrevistado
VSEMAPDP	Nº de ordem do ADP residente na unidade de alojamento
VSEMINDD	Nº de ordem do indivíduo dentro do ADP
VSEMNVIag	Numero de ordem da viagem do indivíduo
PonderVIagSem	Ponderador da viagem do indivíduo
VSEMnumTraj	Numero de trajectos (deslocações) da viagem
VSEMDURTotal	Duração total da viagem em minutos (diferença entre a hora inicial do 1º trajecto e a hora final do último trajecto)
VSEMDURLiquida	Soma das durações das deslocações que fazem parte da viagem, em minutos
Mod1	Nº de deslocações (pertencentes à viagem) realizadas no modo 1 da tabela tab07ModoTransp
Mod2	Nº de deslocações (pertencentes à viagem) realizadas no modo 2 da tabela tab07ModoTransp
Mod3	Nº de deslocações (pertencentes à viagem) realizadas no modo 3 da tabela tab07ModoTransp
Mod4	Nº de deslocações (pertencentes à viagem) realizadas no modo 4 da tabela tab07ModoTransp
Mod5	Nº de deslocações (pertencentes à viagem) realizadas no modo 5 da tabela tab07ModoTransp
Mod6	Nº de deslocações (pertencentes à viagem) realizadas no modo 6 da tabela tab07ModoTransp
Mod7	Nº de deslocações (pertencentes à viagem) realizadas no modo 7 da tabela tab07ModoTransp
Mod8	Nº de deslocações (pertencentes à viagem) realizadas no modo 8 da tabela tab07ModoTransp
Mod9	Nº de deslocações (pertencentes à viagem) realizadas no modo 9 (91, 92, 93, 94 e 95) da tabela tab07ModoTransp
Mod10	Nº de deslocações (pertencentes à viagem) realizadas no modo 10 da tabela tab07ModoTransp
Mod11	Nº de deslocações (pertencentes à viagem) realizadas no modo 11 da tabela tab07ModoTransp
Mod12	Nº de deslocações (pertencentes à viagem) realizadas no modo 12 da tabela tab07ModoTransp
Mod13	Nº de deslocações (pertencentes à viagem) realizadas no modo 13 da tabela tab07ModoTransp
Mod14	Nº de deslocações (pertencentes à viagem) realizadas no modo 14 da tabela tab07ModoTransp
Mod15	Nº de deslocações (pertencentes à viagem) realizadas no modo 15 da tabela tab07ModoTransp
VSEMSITINI	Sítio de Origem da viagem, de acordo com a tabela Tab00sítios
VSEMSITFIM	Sítio de Destino da viagem, de acordo com a tabela Tab00sítios
VSEMDataINI	Data de início da viagem
VSEMHOrINI	Hora de início da viagem
VSEMDataFIM	Data de fim da viagem
VSEMHOrFIM	Hora de fim da viagem
VSEMT6INI	Motivo da viagem imediatamente anterior à observada, de acordo com a tabela Tab06MotivosDestino
VSEMT6COD	Motivo da viagem em observação de acordo com a tabela Tab06MotivosDestino

Quadro IV. 8 - Variáveis recolhidas para caracterizar cada deslocação (INE, 2002)

APURDELOCASEM	
TABELA DAS DESLOCAÇÕES NUM DIA ÚTIL	
CAMPO	DESCRIÇÃO DAS VARIÁVEIS
VSEMDT	Código da Divisão Administrativa 1º nível, corresponde ao Distrito onde reside o indivíduo entrevistado
VSEMCC	Código da Divisão Administrativa 2º nível, corresponde ao Concelho de residência do indivíduo entrevistado
VSEMFR	Código da Divisão Administrativa 3º nível, corresponde à Freguesia de residência do indivíduo entrevistado
VSEMSitio	Código do sítio de residência (tabela Tab00sítios do Inquérito à Mobilidade 2000) do indivíduo entrevistado
VSEMUA	Código da Unidade de alojamento onde reside o indivíduo entrevistado
VSEMAP	Nº de ordem do ADP residente na unidade de alojamento
VSEMIND	Nº de ordem do indivíduo dentro do ADP
VSEMNTraj	Nº de ordem da deslocação
VSEMNViag	Nº de ordem da viagem
PonderViagSem	Ponderador da deslocação do indivíduo (igual ao ponderador da viagem)
VSEMSITINI	Sítio de origem da deslocação (tabela Tab00sítios)
VSEMSITFIM	Sítio de destino da deslocação (tabela Tab00sítios)
VSEMDatINI	Data de início da deslocação
VSEMDatFIM	Data de fim da deslocação
VSEMDuracao	Duração da deslocação (Hora de início - Hora de fim) em minutos
VSEMCodViag	Código de viagem: 1 início de viagem, 2 continuação de viagem
VSEMT6INI	Motivo da deslocação anterior
VSEMT7INI	Modo de transporte da deslocação anterior
VSEMT6COD	Motivo da deslocação
VSEMT7COD	Modo de transporte da deslocação
VSEMT8COD	Forma de estacionamento do veículo (quando o modo é automóvel)
VSEMT9COD	Título de transporte utilizado na deslocação quando o modo é 1, 2, 4, 6 ou 7

3.2. Condicionantes na Utilização da Base dos Dados do INE

Na análise aos dados levantados verificou-se que parte dos inquéritos não apresentava características necessárias à sua utilização na modelação. Os motivos para a inviabilidade da sua utilização devem-se a diferentes fatores:

- A limitação da análise a viagens com origem e destino dentro da AMP levou a exclusão das viagens com origem ou final fora da mesma. Esta limitação levou à exclusão de 197 782 viagens das 506 715 em análise;
- Atendendo à dimensão espacial da AMP considerou-se que tempos de viagem apreendidos superiores a 120 minutos seriam erros de apreensão dos inquiridos ou erros de codificação, inviabilizando a sua utilização. Esta condicionante excluiu 2 802 viagens das 308 933 em análise;
- A impossibilidade de definir em concreto o local de origem e de destino de viagens intrazonais (dentro da mesma zona de tráfego) levou à eliminação das mesmas da análise. Esta condicionante levou à exclusão de 115 817 viagens das 306 131 em análise;
- Verificou-se que do total de possíveis modos de transporte, isolados e combinados, que apenas 6 eram utilizados por mais de 1,5% dos inquiridos e apresentavam uma amostra suficiente para eventuais análises de escolhas modais. Esta condicionante levou à exclusão de 8 388 viagens das 190 314 em análise;

- Os treze motivos de viagem originais, foram reestruturados e apenas se consideraram os cinco mais significativos. Esta condicionante excluiu 19 294 viagens das 181 926 em análise;
- Atendendo às condicionantes anteriores, a base de dados foi reduzida a 162 632 viagens correspondendo a 8 448 pares de zonas de Geração-Atracção (GA)⁹. De forma a assegurar uma quantidade mínima de dados para cada par de zonas GA garantindo alguma representatividade das condições reais para viagens entre os pares GA, analisou-se o número total de viagens entre os mesmos (quadro IV.9).

Quadro IV. 9 - Variáveis recolhidas para caracterizar cada deslocação (INE, 2002)

	>=100 viagens GA	>=50 viagens GA	>=20 viagens GA	Total
nº de pares GA	327	694	1730	8448
nº de viagens	70180	95426	126849	162632

Considerou-se que a seleção dos pares GA com um número total de viagens entre si igual ou superior a 50 garantia um número suficiente de viagens (95 426) e pares GA (694) para uma análise de dados com diversidade suficiente para resultados credíveis com recursos moderados.

Assim, a aplicação dos critérios referidos reduziu a base de dados para 95 426 viagens correspondentes a 694 pares GA.

3.3. Outras Condicionantes à Utilização da Base dos Dados Seleccionada

Além das limitações referidas, durante o estudo ainda se verificaram mais alguns problemas relacionados com a forma como foi feita a especificação de diferente informação na BD, que condicionou e limitou a utilização dos dados levantados:

- A informação recolhida sobre o motivo da viagem não distinguiu a realização de compras da obtenção de serviços financeiros ou públicos. Esta distinção poderia ser útil para o planeamento de sistemas urbanos uma vez que são dois tipos de viagens bastante comuns mas com características/necessidades operacionais diferentes;
- A informação da idade do inquirido foi codificada em escalões etários limitando a utilização da mesma, pela incoerência nos limites dos escalões principalmente até aos 20 anos. Os escalões originais definiram as idades dos 0 aos 4 anos, dos 5 aos 9 anos, dos 10 aos 14 anos e dos 15 aos 19 anos, cuja segmentação aparentemente não tem uma justificação totalmente lógica, na ótica de estudos de transportes. Atendendo às

⁹ - Atendendo ao objetivo do estudo de incluir na análise variáveis relacionadas com estacionamento optou-se por utilizar pares de Geração-Atracção (GA) em vez de pares origem destino (OD). Os pares GA distinguem-se dos OD nas viagens de regresso a casa, onde a zona de destino (casa) não corresponde à atracção mas antes a zona de geração, passando a zona de atracção a ser a origem da viagem, local onde efetivamente as necessidades de estacionamento são mais condicionantes para as escolhas modais.

necessidades de mobilidade e de autonomia, uma segmentação mais lógica seria dos 0 aos 5 e dos 6 aos 10 anos (pré-escolar e 1º Ciclo do Ensino Básico) para crianças totalmente dependentes dos pais; dos 11 aos 18 anos (final do ensino secundário) para adolescentes com potencial independência dos pais e possíveis utilizadores de TC; mantendo as restantes categorias, atendendo à possibilidade de poderem possuir licença de condução e a escolha modal depender de outros fatores que não a impossibilidade de utilizar o automóvel;

- Um problema similar foi detetado na codificação do rendimento líquido mensal do agregado definido por classes, cuja agregação por classes limita a sua utilização como variável contínua;
- Apesar de no inquérito original supostamente se diferenciarem as condições de acesso ao automóvel na família, distinguindo condutores e o número de passageiros, na base de dados utilizada esta distinção não estava disponível;
- Na definição do modo de TC não foi possível especificar qual a carreira escolhida, inviabilizando a avaliação concreta das características de cada viagem;
- Na informação recolhida sobre o custo do estacionamento e da viagem em TC, a sua agregação em custos mensais impossibilita a avaliação do impacto dessas medidas entre ou nas zonas de geração e de atração, ou para as diferentes viagens do dia;
- A informação sobre a origem e o destino das viagens levantou problemas na localização exata das mesmas devido ao tamanho das zonas de tráfego utilizadas, coincidindo com as unidades estatísticas do INE, dificultando significativamente a definição da distância percorrida em cada viagem;
- Estudos de mobilidade que pretendam servir de base para modelos de escolha discreta necessitam de informação relevante mesmo das alternativas não escolhidas, limitação que inquéritos à mobilidade tradicionais não conseguem colmatar. No presente estudo, variáveis como o tempo de viagem ou a distância percorrida podem ter valores diferentes para modos distintos, mesmo entre os mesmos pares GA.
- Outro aspeto muito importante a considerar na análise dos resultados é o desfasamento temporal na recolha de informação de diferentes fontes. Volta-se a salientar que o inquérito à mobilidade foi realizado em 2000, e que o levantamento da oferta de TC e do estacionamento em 2005. Apesar deste desfasamento procurou-se atenuá-lo, através de uma análise das alterações verificadas entre os dois períodos, que possivelmente poderão ter atenuado as diferenças que eventualmente se verificaram nesse período.

4. Opções Assumidas no Tratamento da Informação

4.1. Distância Percorrida na Viagem

Na definição da distância entre as zonas de GA, a impossibilidade de definir com exatidão o ponto de origem e de destino da viagem devido ao tamanho das zonas de tráfego, levou à assunção dos seus centróides. As coordenadas dos centróides geométricos, retiradas do SIG foram ajustados no *Google Earth*© (fotos de 2007) atendendo às densidades de ocupação do solo e de modo a que estivessem sobre uma via relevante. Apenas nas zonas de tráfego homogêneas em termos de ocupação do solo, se considerou que o centróide ajustado coincidia com o centróide geométrico (por exemplo nas áreas mais centrais do concelho do Porto). Considerou-se que as densidades de uso do solo em 2007 seriam relativamente semelhantes às existentes em 2000.

Devido à dificuldade de identificar inequivocamente o trajeto das viagens, para a definição da distância percorrida durante a viagem utilizou-se uma aproximação designada *distância mais curta entre os centróides espaciais da cada zona de tráfego de geração e de atração*, em km. Esta distância mais curta foi, à partida, calculada através do *software AutoRoute2001*®, transformando as coordenadas obtidas do SIG no *Datum 73* para o *Datum WGS84*, de modo a ter em consideração a rede de estradas o mais aproximada possível da rede existente no ano de 2000.

A constatação de alguns valores exagerados para o uso do modo pedonal, levou à correção das distâncias pedonais para valores possíveis de alcançar a velocidades pedonais espectáveis. Apesar de normalmente velocidades de 1,25 a 1,5 m/s serem valores comuns para a maioria da população, optou-se por considerar uma velocidade pedonal de 1,0 m/s (ou seja, 3,6km/h) para os cálculos, de forma a atender aos grupos etários extremos, crianças e idosos (cerca de 13% da amostra). As distâncias percorridas para as viagens a pé foram assim determinadas através da expressão seguinte, em que *tp* é o tempo de viagem apreendido quando o modo escolhido é o pedonal:

$$D2(\text{Walk}) [\text{km}] = t_p [\text{min.}] / 60 * 3,6 [\text{km/h}]$$

Exp. IV - 1

4.2. Frequências Horárias de BUS

A frequência de passagem de unidades de transporte é geralmente considerada na análise de desempenho de sistemas de TC como a frequência horária média de uma determinada carreira. No entanto, no presente estudo, e uma vez que não foi possível determinar qual a carreira utilizada, definiu-se a oferta de TC para determinada viagem como a oferta potencial total, determinada através da soma de todas as carreiras que ligavam os centróides de GA em determinados períodos do dia. Assumiu-se, ainda, que os inquiridos tinham conhecimento dessa oferta quer fossem utilizadores de TC ou não.

Na aquisição de dados sobre os horários dos autocarros houve alguma dificuldade em obter a colaboração de alguns dos operadores privados, nomeadamente a sul do rio Douro. Assim, admitiu-se que a informação recolhida era representativa apenas dos operadores relevantes a norte do rio Douro existente em 2000. A incerteza da informação e falta de cobertura espacial para os concelhos de Espinho, Pova do Varzim, Vila do Conde e Vila Nova de Gaia, levou à sua exclusão no referente às informações sobre os serviços de TC e por conseguinte da análise de escolhas modais nos modelos envolvendo variáveis sobre o serviço de TC.

Na definição dos períodos a ponderar na análise da oferta de serviços de TC, consideraram-se as horas de ponta ao longo do dia tradicionalmente utilizadas em estudos de tráfego para diferentes usos do solo, bem como os resultados da hora de início e de fim para os diferentes motivos de viagem na base de dados original (Gonçalves *et al.*, 2011). Atendendo aos resultados dessa análise, optou-se por avaliar a oferta para cada trajeto durante nos seguintes períodos:

- Manhã - 7:30 - 9:30
- Meio da manhã - 10:30-12:30
- Tarde - 17:00-19:30
- Horário laboral - 7:30-19:30

Atendendo à oferta existente em cada período avaliado utilizou-se um SIG e cada carreira foi dividida nos seus trajetos de ida e de volta. Todas as carreiras que apenas utilizavam parte do trajeto de outra mais abrangente, foram duplicadas, renomeadas e delimitadas no seu início e final. Na contabilização do número de carreiras em cada trajeto, fez-se a distinção dos trajetos completos e dos trajetos parciais de modo a que apenas quando houvesse uma sobreposição espacial ambas fossem contabilizadas. Assim, para cada trajeto de ida e volta das carreiras de autocarros com horários disponíveis, contabilizaram-se o número total das que iniciavam e terminavam o seu trajeto dentro da cada gama de horários estabelecidos. Esta simplificação permitiu assegurar que todos os trajetos considerados para cada carreira se verificavam dentro dos períodos em análise. Estes valores foram acrescentados à tabela de atributos de cada elemento de linha associada a cada carreira no SIG. É de salientar, no entanto, que a simplificação assumida pode efetivamente excluir algumas carreiras mais longas, cuja início ou fim esteja fora do período de análise, mesmo que constituam uma alternativa considerada por parte dos cidadãos.

Considerou-se que todas as carreiras cujos trajetos passassem dentro de uma área de influência equivalente a uma circunferência de raio 1000m com origem no centróide da zona, tinham potencial para atrair passageiros para trajetos superiores ao dobro dessa distância (entre os centróides de GA da viagem). Nas viagens com trajetos inferiores anularam-se as frequências verificadas, considerando-se pouco provável percursos pedonais quase tão longos quanto o trajeto total a vencer. Estas análises foram realizadas para todos os trajetos entre

os pares GA considerados na base de dados e em ambos os sentidos caso existissem. Para avaliar a correção desta assunção, foram ainda definidos mais dois conjuntos de variáveis sobre as frequências nos períodos selecionados. Num primeiro conjunto considerou-se que os pares GA muito próximos, em vez de ter uma frequência nula caracterizadora de um serviço existente mas não utilizado, não tinham valores e portanto não pertenciam à amostra com dados sobre a frequência dos TC. Num segundo conjunto para essas mesmas zonas foram quantificados os trajetos de todas as carreiras que potencialmente as serviam, de forma similar ao procedimento para as zonas menos próximas.

O resultado desta análise foi uma tabela para os diferentes pares GA com o total de trajetos existentes em cada período horário assumido. A sua divisão pelo número de horas em cada período definiu uma frequência média horária entre pares GA que cada inquirido tinha potencialmente à sua disposição.

4.3. Velocidade Comercial dos BUS

Apesar de ser um indicador valorizado pelos utilizadores de TC, a velocidade comercial pressupõe a duração total real da viagem de TC e a distância real percorrida. Atendendo à dificuldade na distinção da carreira efetivamente utilizada pelos inquiridos, optou-se por utilizar uma variável aproximada. Esta variável designada de *velocidade comercial equivalente* foi definida como a duração média da viagem para cada par GA sempre que o modo escolhido não era o Bus e como o tempo de viagem realmente apreendido quando a viagem era realizada de Bus (T2a). Estes tempos foram relacionados não com a distância efetivamente percorrida pelo inquirido mas pela distância mais curta entre centróides de GA que se considerou ser o trajeto teoricamente utilizado em viagem de automóvel ou pelo melhor serviço possível de BUS (D2). Assim, a variável não reflete a velocidade efetivamente alcançada pelo modo Bus na viagem em análise mas uma velocidade média relativa à situação de viagem mais favorável (quer de automóvel quer pela potencial melhor carreira Bus) ponderando todas as viagens de Bus entre determinado par GA.

No estudo apenas se avaliou a velocidade comercial equivalente para o modo BUS (V2). A variável é calculada através da fórmula:

$$V2 \text{ [km/h]} = D2 \text{ [km]} / T2a \text{ [min.]} \times 60$$

Exp. IV - 2

4.4. Tempos de Viagem Apreendidos

O facto de as zonas de tráfego fora do concelho do Porto coincidirem com as respetivas Freguesias, conferiu-lhes uma dimensão que dificultou a correta definição da OD de cada viagem, e consequentemente não permite distinguir pares GA com características muito diferentes. Ainda assim, para atenuar respostas pouco realistas e ter tempos de viagem mesmo quando um determinado modo não era escolhido, Hensher *et al.* (2005) sugerem o cálculo dos tempos médios de viagem entre cada par GA por modo. Apesar de mitigar os

problemas referidos, a existência de pares GA sem viagens num determinado modo, impede a obtenção de tempos médios de viagem para todas as viagens e para todos os modos, diminuindo a amostra disponível para a análise desta variável.

Salienta-se o facto de a duração da viagem poder dificultar a definição do período de oferta a considerar nos TC, tal como referido no ponto 4.2, podendo certas viagens ter início ou fim no período selecionado e, devido à duração da mesma, ter fim ou início fora dele.

5. Definição e Justificação das Variáveis Utilizadas na Análise

5.1. Variáveis Socioeconómicas

Este tipo de variável pretende introduzir no modelo de escolha modal eventuais influências de fatores sociais quer próprias de cada indivíduo quer do agregado familiar em que se insere, e das condições económicas que vivencia. Este tipo de variáveis costumam ser incluídas na parte sistemática da função de utilidade quer linearmente, mais comum, quer associadas a variáveis caracterizadoras da qualidade dos serviços de transporte.

Na modelação das escolhas modais a utilização de combinação de variáveis binárias leva a que, por redundância de informação, uma das categorias seja normalizada a zero. A definição de valores pertencentes à categoria normalizada será caracterizada pela combinação de zeros nas restantes categorias. Assim, para definição de uma metodologia geral e permitir uma interpretação mais fácil dos resultados optou-se por normalizar sempre a variável binária representativa da categoria mais favorável à utilização do automóvel particular. A verificação de qual a categoria mais favorável ao automóvel deve basear-se na ponderação da lógica comportamental subjacente bem como os resultados da base de dados.

5.1.1. Rendimento Líquido Mensal do Agregado

Esta variável apesar de ter características para ser definida como contínua foi codificada na base de dados original como discreta ordinal (Rb). A sua codificação foi realizada atendendo a gamas de valores de rendimento aproximados convertidos a partir de escudos (1€=200,482\$). A codificação original foi a seguinte:

- Rb = 1: Rendimentos mensais de 0-500€;
- Rb = 2: Rendimentos mensais de 501-1000€;
- Rb = 3: Rendimentos mensais de 1001-1500€;
- Rb = 4: Rendimentos mensais de 1501-2000€
- Rb = 5: Rendimentos mensais >2000€;

Esta variável caracteriza o potencial nível económico do agregado familiar que pode influenciar as opções modais dos seus elementos. Condições económicas elevadas permitem, teoricamente, um maior leque de alternativas e tornam medidas restritivas baseadas em

custos para o utilizador menos eficazes. Um aumento do rendimento líquido mensal do agregado poderá incentivar o uso do automóvel em detrimento dos restantes modos, não só pela probabilidade de adquirir e utilizar o mesmo, como pela diluição das diferenças na apreensão dos custos da sua utilização em relação às alternativas.

É uma variável discreta sem uma distribuição definida (ver figura IV.2).

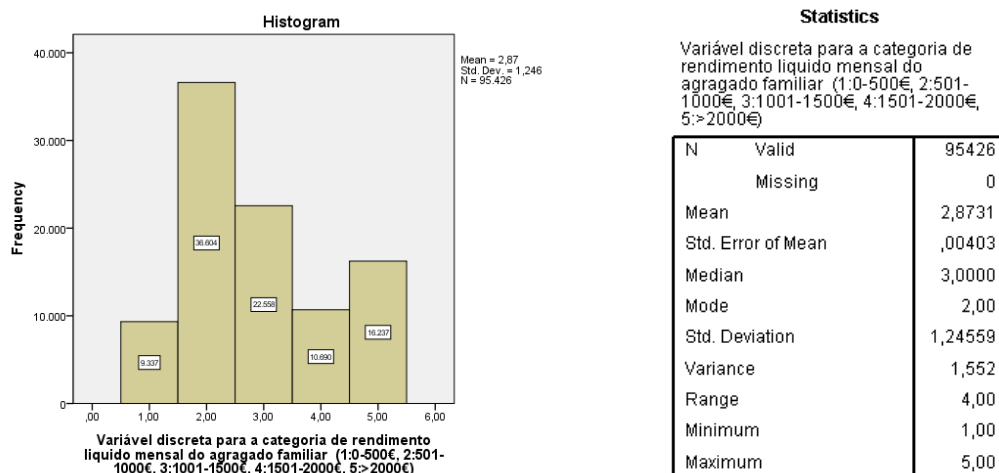


Figura IV. 2 - Dados estatísticos e distribuição da variável *Rb* através do *software* SPSS©

Para a análise de escolhas modais foram ainda definidas variáveis binárias para testar a eventual não-linearidade da sua influência. Foram definidas cinco variáveis binárias, uma para cada categoria:

- R1: Rendimentos mensais de 0-500€;
- R2: Rendimentos mensais de 501-1000€;
- R3: Rendimentos mensais de 1001-1500€;
- R4: Rendimentos mensais de 1501-2000€
- R5: Rendimentos mensais >2000€;

No quadro IV.10 é possível constatar que o nível de rendimento mais elevado é o que apresenta maior repartição modal para o automóvel. Atendendo a este resultado, a categoria normalizada para o rendimento do agregado familiar foi a R5> 2000€.

Quadro IV. 10 - Repartição modal verificada para cada categoria binária do rendimento líquido mensal do agregado

		Rendimento líquido mensal										Total
		R1	%	R2	%	R3	%	R4	%	R5	%	
Modo de transporte da viagem em estudo	Bus+Walk	364	3,9%	1013	2,8%	549	2,4%	160	1,5%	178	1,1%	2264
	BUS	2972	31,8%	7971	21,8%	4123	18,3%	1604	15,0%	1653	10,2%	18323
	Outro Bus	423	4,5%	1668	4,6%	822	3,6%	284	2,7%	421	2,6%	3618
	Moto	787	8,4%	3132	8,6%	1402	6,2%	509	4,8%	545	3,4%	6375
	Walk	2322	24,9%	4929	13,5%	2381	10,6%	936	8,8%	1152	7,1%	11720
	Auto	2469	26,4%	17891	48,9%	13281	58,9%	7197	67,3%	12288	75,7%	53126
Total		9337		36604		22558		10690		16237		95426

Para testar uma especificação discreta alternativa definiu-se ainda uma variável discreta (R) com valores iguais aos máximos de cada intervalo:

- R = 500€: Rendimentos mensais de 0-500€;
- R = 1000€: Rendimentos mensais de 501-1000€;
- R = 1500€: Rendimentos mensais de 1001-1500€;
- R = 2000€: Rendimentos mensais de 1501-2000€
- R = 2500€: Rendimentos mensais >2000€;

Esta variável apresenta as mesmas distribuições que a variável discreta original.

5.1.2. Rendimento Líquido Mensal do Agregado *per Capita*

Além do tipo de variáveis anteriores ponderou-se ainda a definição de outro tipo que permitisse incorporar o impacto da dimensão do agregado familiar no nível económico do mesmo (Rc). Rendimentos similares para estruturas familiares muito diferentes tendem a ter influências distintas devido à parte disponível para cada um. Para a sua definição utilizou-se a última especificação referida para o rendimento líquido do agregado (R) e o número de indivíduos do agregado (FNumInd):

$$Rc \text{ [euros/mês/residente]} = R \text{ [euros/mês]} / FNumInd \quad \text{Exp. IV - 3}$$

A sua formulação implicou que a mesma tivesse características para ser definida como contínua com uma distribuição aproximadamente lognormal (figura IV.3).

No sentido de avaliar a sua eventual não-linearidade foram também definidas variáveis binárias para diferentes categorias de rendimentos *per capita*, atendendo aos valores que limitavam amostras relativamente semelhantes:

- Rc1: Rendimentos mensais *per capita* ≤ 250€/mês/residente;
- Rc2: Rendimentos mensais *per capita* > 250 & < 500€/mês/residente ;
- Rc3a: Rendimentos mensais *per capita* > 500€/mês/residente ;

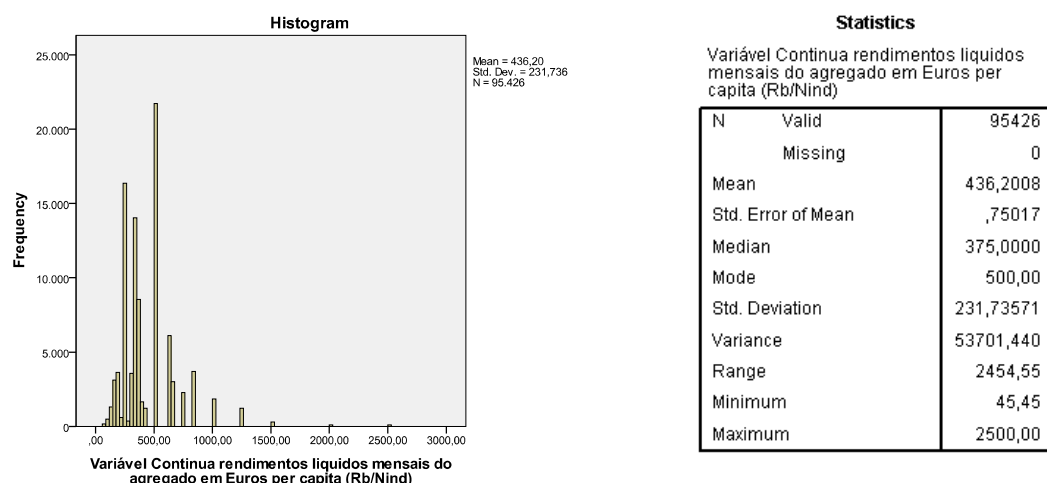


Figura IV. 3 - Dados estatísticos e distribuição da variável *Rc* através do *software* SPSS©

Quadro IV. 11 - Repartição modal verificada para cada categoria binária do rendimento líquido mensal do agregado *per capita*

		Rendimento líquido mensal per capita						Total
		Rc1	%	Rc2	%	Rc3a	%	
Modo de transporte da viagem em estudo	Bus+Walk	844	3,3%	1227	2,4%	193	1,0%	2264
	BUS	6688	26,0%	9592	18,8%	2043	11,0%	18323
	Outro Bus	1615	6,3%	1682	3,3%	321	1,7%	3618
	Moto	2220	8,6%	3668	7,2%	487	2,6%	6375
	Walk	4192	16,3%	6025	11,8%	1503	8,1%	11720
	Auto	10131	39,4%	28906	56,6%	14089	75,6%	53126
Total		25690		51100		18636		95426

A análise das repartições modais para cada uma das categorias binárias definidas (quadro IV.11), permite verificar que a categoria de rendimentos mais elevados aparentemente potencia uma maior utilização do automóvel. Assim, também para esta variável, a categoria selecionada para a normalização foi a mais elevada (Rc3a).

5.1.3. Sexo do Inquirido

Esta variável (*Sx*) binária foi definida segundo os dados do inquérito, correspondendo ao valor zero quando o inquirido era masculino e ao valor um quando era feminino ($M=0$ ou $F=1$). Esta variável permite avaliar diferenças de comportamento entre homens e mulheres, nomeadamente distinguir diferenças de mentalidades, hábitos ou comportamentos diferenciados geralmente associados ao género.

É possível constatar que a amostra apresenta um equilíbrio entre inquiridos do sexo masculino e do sexo feminino. O senso comum sugere que o sexo feminino tem maior propensão para a utilização das alternativas ao automóvel particular. No quadro IV.12 é possível verificar que

na base de dados efetivamente as alternativas ao automóvel com exceção do modo Moto e Outros Bus, são mais utilizadas pelas mulheres que pelos homens.

Quadro IV. 12 - Repartição da variável Sexo do inquirido

Sx	n	%
M =0	47452	49,7%
F =1	47974	50,3%
Total	95426	100,0%

Quadro IV. 13 - Repartição modal da variável Sexo do inquirido

		Sexo				Total
		M=0	%	F=1	%	
Modo de transporte da viagem em estudo	Bus+Walk	750	1,6%	1514	3,2%	2264
	BUS	6661	14,0%	11662	24,3%	18323
	Outro Bus	1969	4,1%	1649	3,4%	3618
	Moto	4027	8,5%	2348	4,9%	6375
	Walk	4773	10,1%	6947	14,5%	11720
	Auto	29272	61,7%	23854	49,7%	53126
Total		47452		47974		95426

Esta propensão pode dever-se a motivos sociais, tanto por habitualmente as mulheres não gostarem tanto de conduzir, como por tradicionalmente ser o cônjuge masculino a conduzir e a levar o automóvel consigo. Este facto deve ser levado em conta na interpretação dos resultados, uma vez que os mesmos podem traduzir uma realidade algo imposta em vez de uma escolha real das mulheres.

5.1.4. Idade do Inquirido

A variável idade do inquirido, apesar de ter características para ser especificada como contínua, foi codificada na base de dados original como discreta ordinal. Esta variável deverá permitir inferir em que medida o estágio da vida do inquirido pode influenciar as suas opções modais. A codificação original, no entanto, era extensa e não correspondia a escalões que em termos de mobilidade justificassem a segmentação assumida (ver quadro IV.14). Assim, para o estudo fez-se uma nova recodificação discreta ordinal (ldb), atendendo a pressupostos que poderiam influenciar as escolhas modais:

- ldb=1: 0 a 9 anos (escalões 1 e 2 originais) - Crianças que viajam essencialmente com os pais
- ldb=2: 10 a 19 anos (escalões 3 e 4 originais) - Estudantes que podem viajar de TC e genericamente ainda não têm idade para ter carta de carro

- ldb=3: 20 a 34 anos (escalões 5, 6 e 7 originais) - Solteiros ativos ou casais jovens com crianças com dependência em termos de mobilidade;
- ldb=4: 35 a 64 anos (escalões 8, 9, 10, 11, 12 e 13 originais) - Pessoas maduras em idade ativa, geralmente sem crianças dependentes em termos de mobilidade;
- ldb=5: 65+ (escalão 14 original) - Reformados/Avós que podem eventualmente cuidar dos netos

Quadro IV. 14 - Codificação do escalão etário original na base de dados

Codigo	Gama de Idades
1	0-4
2	5-9
3	10-14
4	15-19
5	20-24
6	25-29
7	30-34
8	35-39
9	40-44
10	45-49
11	50-54
12	55-59
13	60-64
14	65+

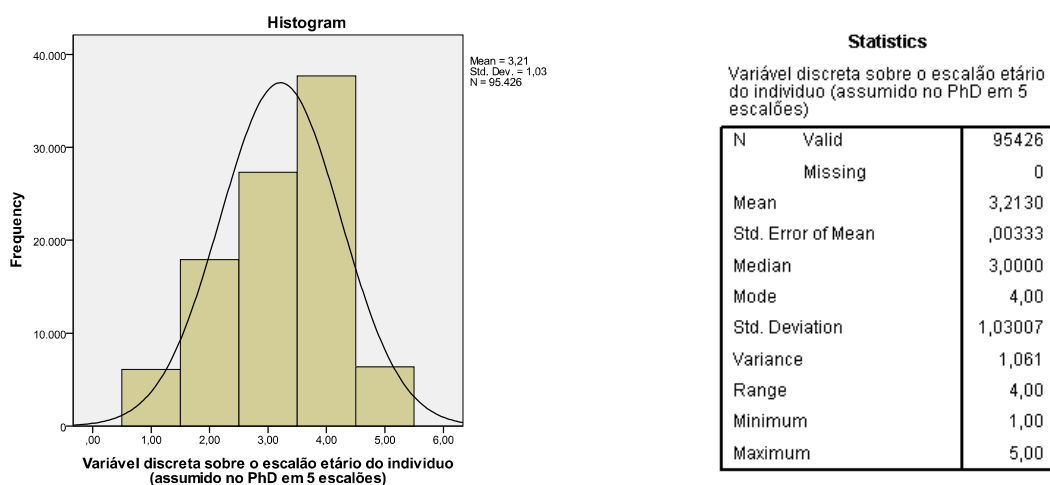


Figura IV. 4 - Dados estatísticos e distribuição da variável *ldb* através do software SPSS©

Salienta-se que mesmo com a nova codificação, não é possível segregar os menores de 18 anos dos restantes inquiridos, que atendendo à idade mínima para obtenção da licença de condução automóvel seria um limite relevante a analisar. De igual forma não foi possível diferenciar as idades pré-escolares das escolares, que também poderão influenciar os padrões de mobilidade familiar. Atendendo ainda assim à segmentação assumida, a figura IV.4 apresenta a distribuição e os dados estatísticos da mesma na base de dados utilizada.

Quadro IV. 15 - Repartição das idades dos inquiridos na amostra

	ldb	n	%
Escala etário do indivíduo (5 escalões)	0-9 anos	6107	6,4
	10-19 anos	17924	18,8
	20-34 anos	27312	28,6
	35-65 anos	37701	39,5
	+65 anos	6382	6,7
Total		95426	100,0

O senso comum sugere que atendendo aos escalões etários definidos, os escalões 1 (0-9 anos), 2 (10-19 anos) e 5 (+65 anos) serão os que menores propensões terão para o uso do automóvel particular. Esta tendência pode dever-se quer à falta de habilitações para conduzir (por ser novo ou já menos propenso a conduzir), quer por existirem alternativas com condições mais favoráveis para estes escalões. Estas condições tornam expectável que os mesmos escalões tenham maior propensão para o uso dos autocarros e do modo pedonal. É de salientar que o escalão etário mais baixo (0-9 anos) poderá apresentar resultados inesperados pelo facto de não possuírem opção de escolha, estando diretamente dependentes das escolhas dos pais.

Para testar a eventual não-linearidade dos impactos, foram definidas variáveis binárias para os diferentes escalões etários. A repartição modal de cada escalão na base de dados é apresentada no quadro IV.16.

Quadro IV. 16 - Repartição modal segundo os escalões etários

		Escala etário								Total	
		0-9 anos	%	10-19 anos	%	20-34 anos	%	35-65 anos	%	+65 anos	%
Modo de transporte da viagem em estudo	Bus+Walk	42	0,7%	900	5,0%	388	1,4%	715	1,9%	219	3,4%
	BUS	523	8,6%	6026	33,6%	3716	13,6%	6159	16,3%	1899	29,8%
	Outro Bus	301	4,9%	1814	10,1%	592	2,2%	814	2,2%	97	1,5%
	Moto	127	2,1%	938	5,2%	2372	8,7%	2635	7,0%	303	4,7%
	Walk	830	13,6%	2989	16,7%	2121	7,8%	4293	11,4%	1487	23,3%
	Auto	4284	70,1%	5257	29,3%	18123	66,4%	23085	61,2%	2377	37,2%
Total		6107		17924		27312		37701		6382	

Atendendo aos resultados das repartições modais verifica-se que o escalão com maior uso do automóvel particular é o escalão dos 0 aos 9 anos. Apesar de aparentemente não ser coerente, o facto de o modo automóvel corresponder a condutores e passageiros, torna o resultado plausível é até expectável, uma vez que crianças desta idade não têm opção de escolha e são levadas pelos pais. Tendo em conta todas estas questões optou-se por normalizar as variáveis binárias do escalão etário através do grupo dos 0 aos 9 anos.

5.1.5. Nível de Instrução

O nível de instrução do inquirido foi codificado na base de dados original como discreta com 10 níveis.

Quadro IV. 17 - Codificação do nível de instrução original na base de dados

Código	Nível de Instrução
1	Não sabe ler nem escrever
2	Sabe ler e escrever
3	Ensino básico - 1º ciclo (4º ano/antiga 4ª classe)
4	Ensino básico - 2º ciclo (6º ano/antigo 2º ano)
5	Ensino básico - 3º ciclo (9º ano/antigo 5º ano)
6	Ensino secundário - 11º ano (antigo 7º ano)
7	Ensino secundário - 12º ano
8	Curso superior não universitário
9	Curso superior universitário
10	Pós Universitário

Para o estudo, foi realizada uma agregação considerada mais adequada às características do estudo, em 4 escalões (In) com os níveis de ensino existentes em Portugal:

In = 1: Analfabeto (original 1, incluindo crianças em idade pré-escolar)

In = 2: Básico (original 2, 3, 4 e 5)

In = 3: Secundário (original 6 e 7)

In = 4: Superior (original 8, 9 e 10)

Com esta variável pretende-se caracterizar a potencial influência do nível de instrução nas escolhas modais. O senso comum sugere que níveis de instrução superiores tenderão a exigir maiores níveis de conforto, mas que também podem desenvolver consciências mais ambientalistas. Atendendo aos dados da figura IV.5 e do quadro IV.18, é possível constatar que 70% da amostra tem o nível básico de instrução.

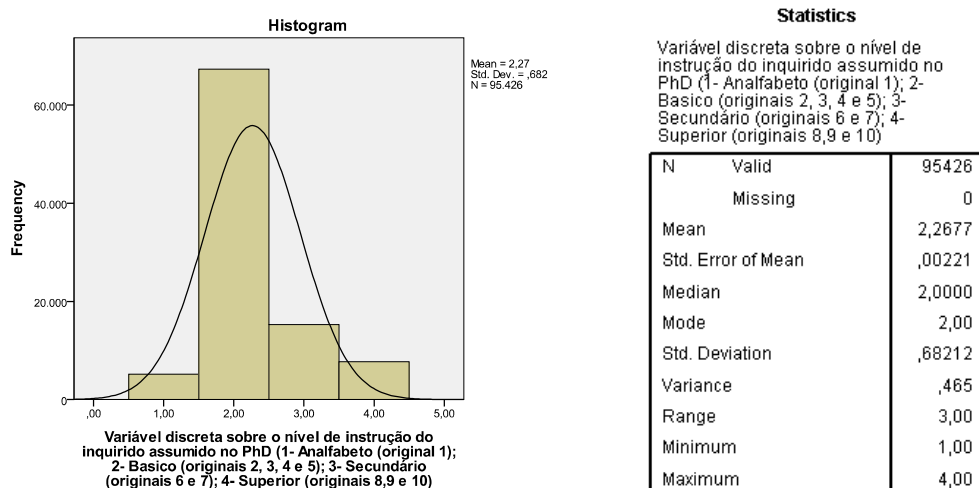


Figura IV. 5 - Dados estatísticos e distribuição da variável *In* através do *software* SPSS©

Quadro IV. 18 - Repartição do nível de instrução dos inquiridos na amostra

	In	n	%
Nível de Instrução	Analfabeto	5147	5,4%
	Básico	67289	70,5%
	Secundário	15290	16,0%
	Superior	7700	8,1%
Total		95426	100,0%

Atendendo ao pressuposto de que maior instrução é equivalente a maiores oportunidades e melhor nível de vida será expectável que o aumento do nível de instrução leve a uma maior utilização do automóvel particular em contraste com as alternativas. A análise à repartição modal de cada nível é apresentada no quadro IV.19.

Quadro IV. 19 - Repartição modal segundo os níveis de instrução

		Nível de Instrução								Total
		Analfabeto	%	Básico	%	Secundário	%	Superior	%	
Modo de transporte da viagem em estudo	Bus+Walk	99	1,9%	1831	2,7%	301	2,0%	33	0,4%	2264
	BUS	786	15,3%	13903	20,7%	2969	19,4%	665	8,6%	18324
	Outro Bus	237	4,6%	3122	4,6%	238	1,6%	21	0,3%	3618
	Moto	248	4,8%	5701	8,5%	413	2,7%	13	0,2%	6375
	Walk	716	13,9%	8909	13,2%	1540	10,1%	555	7,2%	11720
	Auto	3061	59,5%	33823	50,3%	9829	64,3%	6413	83,3%	53128
Total		5147		67289		15290		7700		95426

É possível constatar que aparentemente o nível de instrução superior potencia uma maior utilização do automóvel, tendo-se optado por este nível para a normalização.

Para analisar a sua eventual não-linearidade também foram definidas variáveis binárias para os diferentes níveis de instrução.

5.1.6. Número de Automóveis Disponíveis Diariamente *per Capita* no Agregado

Esta variável contínua (Nc), também designada de *taxa de motorização familiar*, foi definida através da combinação de duas variáveis contínuas da base de dados original:

$$NC [\text{automóveis/residente}] = FNAUTODiario / FNumInd$$

Exp. IV - 4

Caracteriza o potencial grau de acesso ao automóvel particular diariamente por cada elemento do agregado familiar e pretende-se que permita dar uma estimativa do potencial de mobilidade em automóvel particular dos elementos do agregado familiar. Em simultâneo apresenta-se como uma forma de integrar indiretamente no modelo o potencial nível económico da família (número de automóveis) e a dimensão do agregado familiar (número de elementos do agregado).

É uma variável contínua conjugada que relaciona duas variáveis socioeconómicas com uma distribuição que poderá ser aproximadamente lognormal (ver figura IV.6).

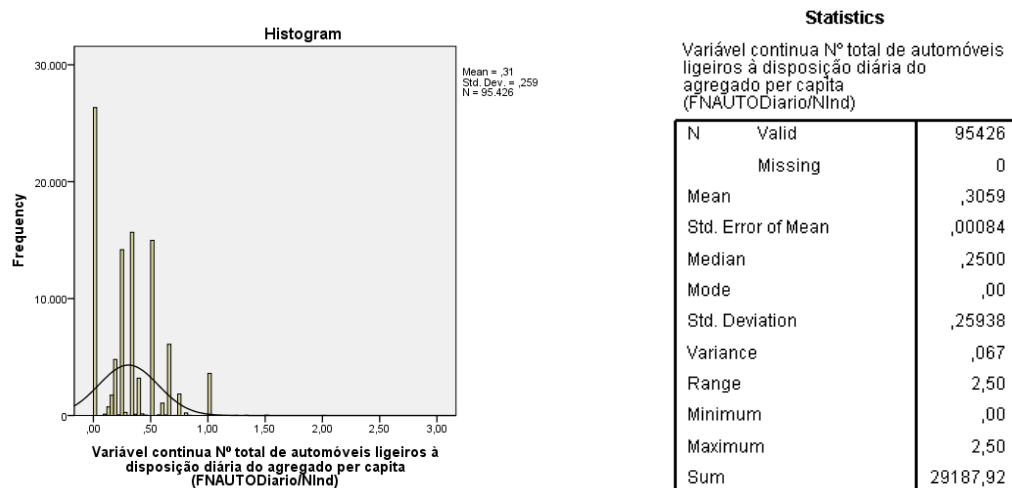


Figura IV. 6 - Dados estatísticos e distribuição da variável NC através do *software* SPSS©

Um aumento do número de automóveis ligeiros no agregado tenderá a incentivar o seu uso em detrimento dos restantes modos. Apesar de aparentemente ser uma variável específica para os condutores, optou-se por considerá-la para todos os modos uma vez que não foi possível distinguir os condutores dos passageiros, assumindo-se que o número de carros disponíveis no agregado (e não os disponíveis para o inquirido) era um fator comum a todos os inquiridos independentemente do modo escolhido.

Para a análise de escolhas modais foram ainda definidas variáveis binárias para testar a eventual não-linearidade da sua influência. Para a definição dos intervalos de cada gama de valores foram consideradas as que apresentavam amostras com tamanho consideradas suficientes e valores que descrevessem facilmente a proporção de veículos por morador do agregado. Foram consideradas 5 categorias:

- NC0 = 0 veíc./resid.: Corresponde aos agregados sem automóvel disponível;
- NC1 >0 e <=0,25 veíc./resid.: Corresponde aos agregados familiares com automóvel disponível, mas que o divide pelo menos entre 4 pessoas;
- NC2 >0,25 e <=0,5 veíc./resid.: Corresponde aos agregados familiares com automóvel mas que o divide entre 4 e 2 pessoas;
- NC3 >0,5 e <=0,75 veíc./resid.: Corresponde aos agregados familiares com automóvel mas que o divide entre 2 a 1,33 pessoas;
- NC4a >0,75 veíc./resid.: Corresponde aos agregados familiares com automóveis para menos de 1,33 pessoas;

Na base de dados foi possível verificar a repartição das variáveis binárias apresentada no Quadro IV - 20.

Quadro IV. 20 - Repartição das variáveis binárias consideradas sobre a taxa de motorização familiar da base de dados

Variável binária	Amostra	Percentagem
NC0 (=0 veíc./resid.)	26344	27,6%
NC1 (>0 & <=0,25 veíc./resid.)	21637	22,7%
NC2 (>0,25 & <=0,5 veíc./resid.)	34403	36,1%
NC3 (>0,5 & <=0,75 veíc./resid.)	9105	9,5%
NC4a (>0,75 veíc./resid.)	3937	4,1%
	95426	100,0%

Analisando a repartição modal verificada para cada variável binária constata-se que a partir dos 0,25 veíc./resid. o uso do automóvel é predominante (Quadro IV.21).

No caso da taxa de motorização familiar verifica-se que a categoria mais favorável ao automóvel é a NC4a, com uma taxa superior a 0,75 veíc./resid., coerente com o senso comum de que a maior disponibilidade do mesmo potencia o seu uso. Assim, será esta a categoria a normalizar.

Quadro IV. 21 - Repartição modal verificada para cada categoria da taxa de motorização familiar na base de dados

		Taxa de motorização Familiar										Total
		NC0	%	NC1	%	NC2	%	NC3	%	NC4a	%	
Modo de transporte da viagem em estudo	Bus+Walk	1197	4,5%	513	2,4%	508	1,5%	43	0,5%	3	0,1%	2264
	BUS	9302	35,3%	4097	18,9%	4248	12,3%	580	6,4%	96	2,4%	18323
	Outro Bus	1420	5,4%	1172	5,4%	856	2,5%	133	1,5%	37	0,9%	3618
	Moto	3170	12,0%	1456	6,7%	1505	4,4%	190	2,1%	54	1,4%	6375
	Walk	6320	24,0%	2339	10,8%	2573	7,5%	316	3,5%	172	4,4%	11720
Auto		4935	18,7%	12060	55,7%	24713	71,8%	7843	86,1%	3575	90,8%	53126
Total		26344		21637		34403		9105		3937		95426

Foi ainda ponderada a hipótese de utilizar a taxa de motorização familiar por adulto do agregado. Atendendo aos limites definidos no inquérito do INE para os grupos etários consideraram-se adultos todos os maiores de 19 anos, no entanto, nos agregados familiares em que se verificou que nem todos os seus elementos realizaram viagens ou responderam ao inquérito não foi possível clarificar o número exato de adultos. Esta constatação impossibilitou uma avaliação precisa do número total de adultos de todos os agregados, levando a uma diminuição para cerca de dois terços da amostra considerada (63 735 inquéritos) que definiram a base de dados designada de *NCA_b*. Análises posteriores revelaram um impacto no ajustamento dos modelos exploratórios muito similar à taxa de motorização *per capita*, considerando-se a mesma nas análises para estudos futuros.

5.1.7. Disponibilidade Diária de Moto no Agregado

A variável sobre a *disponibilidade diária de Moto no agregado familiar* (Mot) foi especificada como binária. Apesar de a base de dados original permitir a sua definição como contínua, assumiu-se como relevante apenas a sua disponibilidade ou não para justificar o eventual uso do modo Moto. A sua disponibilidade poderá potenciar o seu uso relativamente ao automóvel, uma vez que permite uma mobilidade/conveniência semelhante, quando não há necessidade de levar outros passageiros ou quando as condições climáticas não forem muito adversas.

Quadro IV. 22 - Repartição da variável disponibilidade diária de Moto (Mot) no agregado na base de dados

Mot	n	%
N =0	79797	83,6%
S =1	15629	16,4%
Total	95426	100,0%

É possível constatar que a maioria dos inquiridos selecionados não tem Moto disponível. Uma análise à repartição modal para esta variável confirma um maior uso deste modo por parte dos inquiridos que o têm disponível (quadro IV.23).

Quadro IV. 23 - Repartição modal verificada para a disponibilidade diária de Moto na base de dados

		Disponibilidade diária de Moto no agregado familiar				Total
		N=0	%	S=1	%	
Modo de transporte da viagem em estudo	Bus+Walk	1855	2,3%	409	2,6%	2264
	BUS	16392	20,5%	1931	12,4%	18323
	Outro Bus	2653	3,3%	965	6,2%	3618
	Moto	805	1,0%	5570	35,6%	6375
	Walk	10772	13,5%	948	6,1%	11720
	Auto	47320	59,3%	5806	37,1%	53126
Total		79797		15629		95426

5.1.8. Licença de Condução de Automóvel

A variável sobre a disponibilidade de licença de condução de automóvel do inquirido foi especificada na base de dados original como discreta com dois possíveis valores, sim ou não. Para o estudo foi definida a variável binária (Lc) sobre a possibilidade de utilização do TI através da disponibilidade de licença de condução (S=1; N=0). Esta variável deverá permitir definir se o inquirido é passageiro cativo de TC ou voluntário, permitindo avaliar o seu impacto em diferentes segmentos da população.

Quadro IV. 24 - Repartição da variável Lc na base de dados

Lc	n	%
S=1	49245	51,6%
N=0	46181	48,4%
Total	95426	100,0%

É possível constatar que a base de dados tem uma repartição muito semelhante de inquiridos com licença de condução automóvel e sem a mesma. O senso comum sugere que a existência da mesma será um fator potenciador do uso do automóvel.

5.2. Variáveis Caracterizadoras da Viagem e da Organização das Atividades Diárias

Este tipo de variável pretende introduzir no modelo de escolha modal os factos relacionados com as características das viagens, da posição relativa entre as zonas de GA e da organização das atividades diárias dos inquiridos. Serão estes fatores os que permitirão associar as características socioeconómicas de cada viajante com as características de cada viagem.

5.2.1. Modo de Transporte Utilizado

O modo de transporte utilizado na viagem em análise é uma variável discreta nominal. Na modelação das escolhas modais será a variável dependente a modelar. A base de dados original discriminou o modo de transporte em 19 tipos distintos (quadro IV.25), cuja possível combinação foi aferida através do número de deslocações por modo em cada viagem com um único propósito.

Quadro IV. 25 - Modos de transportes considerados na base de dados original (INE, 2002)

Código	Modo de transporte
1	Autocarro STCP
2	Autocarro urbano de outros operadores
3	Autocarro de excursão
4	Outro tipo de transporte em autocarro
5	Veículo rodoviário (ligeiro ou pesado) de transporte escolar ou de trabalhadores de empresa
6	Eléctrico
7	Comboio
8	Taxi/Letra A
91	Automóvel só condutor
92	Automóvel condutor com 1 passageiro
93	Automóvel condutor com 2 passageiros
94	Automóvel condutor com 3 passageiros
95	Automóvel condutor com 4 ou mais passageiros
10	Motociclo/Ciclomotor
11	Bicicleta
12	A pé
13	Avião
14	Barco
15	Outro modo de transporte

Quadro IV. 26 - Modos de transportes agregados

Codigo	Modo de transporte
1	So Auto (Codigo original 9 - Inclui condutores e passageiros);
2	So Bus (Codigos originais 1,2 - Inclui STCP e Operadores privados);
3	So Outros Bus (Codigos originais 3,4,5 - Inclui Transporte Escolar, Excursoes e outras viagens de BUS);
4	So Comb Bus (Codigos originais 1,2,3,4,5 - Inclui qualquer combinaçao dos diferentes modos BUS);
5	so train (Codigo original 7);
6	so moto (Codigo original 10);
7	so bike (Codigo original 11);
8	so outros (Codigos originais 13,14,15 - Inclui Aviao, Barco ou outro modo);
9	so Walk (Codigo original 12);
10	so taxi (Codigo original 8)
11	auto+bus;
12	outros bus+auto;
13	bus+train;
14	outros bus+train;
15	auto+train;
16	bus+Walk;
17	outros bus+Walk;
18	comb outros.

Apesar de na descrição dos dados do inquérito se apresentarem várias tipologias de utilização do automóvel, constatou-se que a base de dados utilizada apenas distinguia uma única classe para a utilização do mesmo, sem diferenciação entre condutor e passageiro(s).

Para o estudo agruparam-se os modos em 10 classes para viagens feitas integralmente num único modo e mais 8 classes para as realizadas com combinações de modos (quadro IV.26). Esta codificação permitiu distinguir todas as viagens efetuadas com distinção unívoca do modo de transporte.

Tal como foi já referido, a constatação de que a maioria das viagens foi realizada apenas por seis modos de transportes, levou à purga da amostra utilizada de todas as viagens realizadas pelos restantes modos. Assim definiu-se uma variável discreta nominal (Mtrp) com os seguintes valores:

- Mtrp = 1: Combinação BUS e WALK;
- Mtrp = 2: Só BUS;
- Mtrp = 3: Só Outros BUS;
- Mtrp = 4: Só MOTO;
- Mtrp = 5: Só WALK;
- Mtrp = 6: Só Auto;

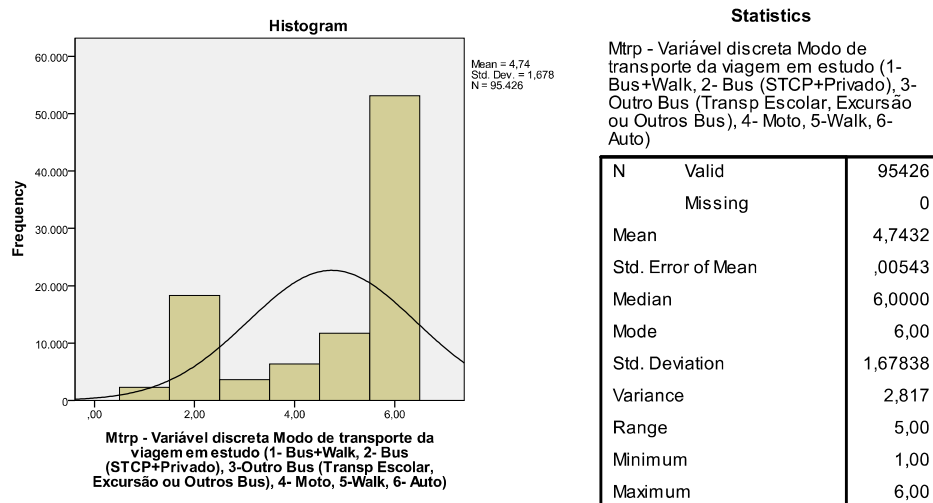


Figura IV. 7 - Dados estatísticos e distribuição da variável MTrp através do *software* SPSS©

Na codificação utilizada, a escolha do modo Auto com código mais elevado deveu-se à forma como o programa NLogit 4.0 faz a modelação considerando o modo a normalizar e, portanto, como referência para as diferenças de utilidade, o último da lista de modos.

Quadro IV. 27 - Repartição do modo de transporte utilizado na viagem em estudo na amostra

MTrp		n	%
Modo de transporte da viagem em estudo	Bus+Walk	2264	2,4%
	BUS	18323	19,2%
	Outro Bus	3618	3,8%
	Moto	6375	6,7%
	Walk	11720	12,3%
	Auto	53126	55,7%
Total		95426	100,0%

5.2.2. Distância entre as Zonas de Geração-Atração

A distância mais curta entre os centróides espaciais da cada zona de tráfego de GA, em km, permite avaliar a proximidade entre zonas e a sua influência no uso de modos estudados. Será também esta distância mais curta (D2), teoricamente, a utilizada pelos utilizadores de TI ou pelos TC com o melhor serviço possível.

A sua definição confere à mesma características contínuas com uma distribuição aproximadamente lognormal. Uma vez que não é a distância real da viagem em análise, e portanto a distância percorrida especificamente por cada modo, mas um valor estimado entre a origem e o destino, será uma variável comum para todos as alternativas com as devidas adaptações para o modo Walk.

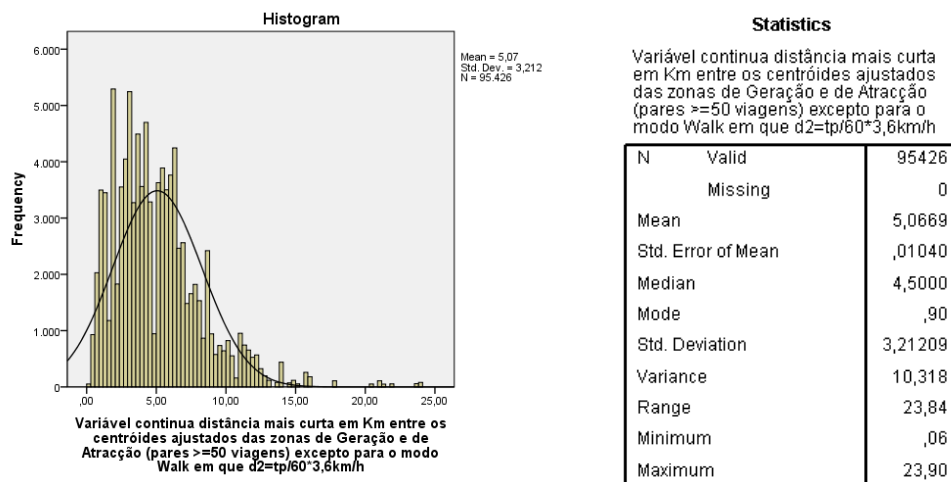


Figura IV. 8 - Dados estatísticos e distribuição da variável D2 através do *software* SPSS©

O senso comum sugere que o aumento da distância diminui a atratividade do uso das alternativas relativamente ao automóvel.

Para avaliar a sua linearidade definiram-se cinco variáveis binárias para diferentes distâncias (Dxi) que serviram para testar diferentes especificações. Os intervalos de valores para as diferentes variáveis foram definidos atendendo à gama de valores encontrados e atendendo às especificidades das alternativas, nomeadamente o modo pedonal.

Quadro IV. 28 - Variáveis binárias para a distância mais curta entre centróides

Variável binária	Amostra	Percentagem
Dx1 ($D2 \leq 1,0 \text{ km}$)	6325	6,6%
Dx2 ($D2 > 1,0$ e $\leq 2,0 \text{ km}$)	10105	10,6%
Dx3 ($D2 > 2,0$ e $\leq 5,0 \text{ km}$)	36202	37,9%
Dx4 ($D2 > 5,0$ e $\leq 10,0 \text{ km}$)	35881	37,6%
Dx5 ($D2 > 10,0 \text{ km}$)	6913	7,2%
Total	95426	100,00%

Atendendo à repartição modal verificada para as diferentes variáveis binárias optou-se pela normalização da variável referente à maior distância. Esta variável normalizada variou consoante as gamas de valores utilizados. No quadro IV.29 é possível constatar que a partir de distâncias superiores a 2,0km a utilização do automóvel é predominante.

Quadro IV. 29 - Repartição modal para as diferentes distâncias entre centróides

		Distância mais curta entre centróides									Total
		Dx1	%	Dx2	%	Dx3	%	Dx4	%	Dx5	
Modo de transporte da viagem em estudo	Bus+Walk	1	0,0%	197	1,9%	874	2,4%	954	2,7%	238	2264
	BUS	24	0,4%	1167	11,5%	7824	21,6%	7822	21,8%	1486	18324
	Outro Bus	0	0,0%	123	1,2%	1532	4,2%	1598	4,5%	365	3618
	Moto	0	0,0%	350	3,5%	3095	8,5%	2608	7,3%	322	6375
	Walk	6234	98,6%	4649	46,0%	770	2,1%	67	0,2%	0	11721
	Auto	66	1,0%	3619	35,8%	22107	61,1%	22832	63,6%	4502	53128
Total		6325		10105		36202		35881		6913	95426

5.2.3. Duração Total da Viagem Apreendida

A variável duração da viagem, ou tempo de viagem, foi aferida no inquérito à mobilidade de várias formas. Para a integração desta informação na modelação das escolhas modais optou-se por considerar a duração total da viagem incluindo paragens por intrinsecamente incluir a percepção do tempo de viagem em andamento, dos congestionamentos e de eventuais atrasos/problemas operacionais. Esta variável tem características contínuas e a unidade é o minuto. É de salientar que esta variável não corresponde ao tempo total de viagem real, mas sim ao apreendido pelo inquirido. Apesar de poder apresentar valores enviesados em relação à real duração da viagem, a sua utilização foi considerada mais apropriada por traduzir a sua percepção parte dos seus utilizadores, cuja motivação se pretende modelar. A constatação de distâncias entre pares GA que não ultrapassavam 25km, levou à purga da amostra dos inquéritos com tempos de viagem exageradas ($T_p > 120\text{min.}$).

A variável tempo total de viagem (T_p) é uma variável que na modelação será genérica, isto é, que para a mesma viagem e para o mesmo indivíduo terá diferentes valores para diferentes modos. Esta característica leva a que na modelação haja necessidade de incluir valores para os modos embora disponíveis não utilizados nessa viagem. A forma tradicional de realizar os inquéritos à mobilidade apenas recolhe informação sobre os aspetos operacionais do modo utilizado, não permitindo a comparação com os modos não utilizados.

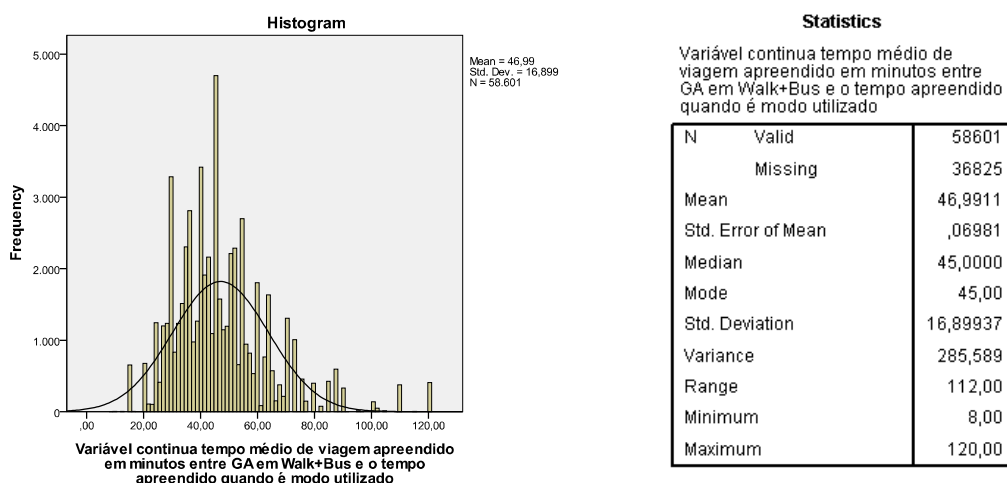
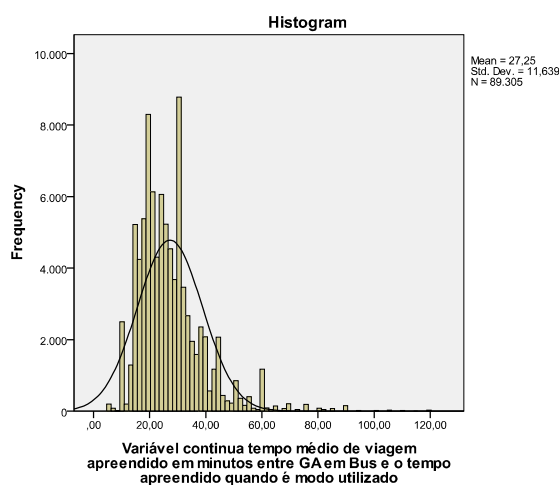


Figura IV. 9 - Dados estatísticos e distribuição da variável T1a através do *software* SPSS©

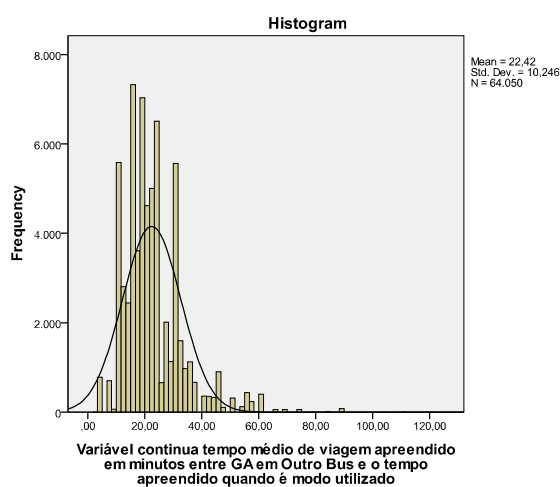


Statistics

Variável continua tempo médio de viagem apreendido em minutos entre GA em Bus e o tempo apreendido quando é modo utilizado

N	Valid	89305
	Missing	6121
Mean		27,2485
Std. Error of Mean		,03895
Median		25,0000
Mode		30,00
Std. Deviation		11,63856
Variance		135,456
Range		115,00
Minimum		5,00
Maximum		120,00

Figura IV. 10 - Dados estatísticos e distribuição da variável T2a através do *software* SPSS©

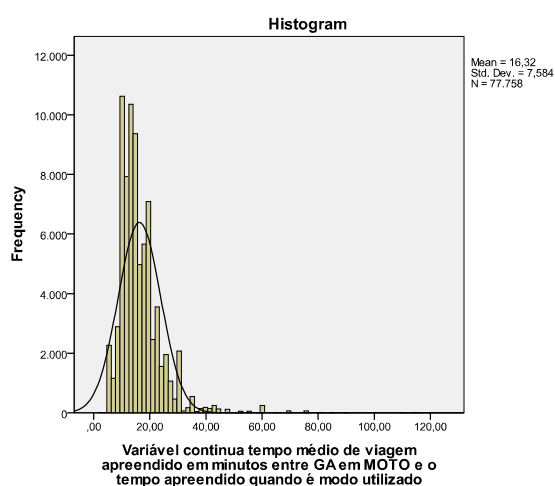


Statistics

Variável continua tempo médio de viagem apreendido em minutos entre GA em Outro Bus e o tempo apreendido quando é modo utilizado

N	Valid	64050
	Missing	31376
Mean		22,4170
Std. Error of Mean		,04049
Median		20,8333
Mode		30,00
Std. Deviation		10,24601
Variance		104,981
Range		117,00
Minimum		3,00
Maximum		120,00

Figura IV. 11 - Dados estatísticos e distribuição da variável T3a através do *software* SPSS©



Statistics

Variável continua tempo médio de viagem apreendido em minutos entre GA em MOTO e o tempo apreendido quando é modo utilizado

N	Valid	77758
	Missing	17668
Mean		16,3206
Std. Error of Mean		,02720
Median		15,0000
Mode		10,00
Std. Deviation		7,58386
Variance		57,515
Range		118,00
Minimum		2,00
Maximum		120,00

Figura IV. 12 - Dados estatísticos e distribuição da variável T4a através do *software* SPSS©

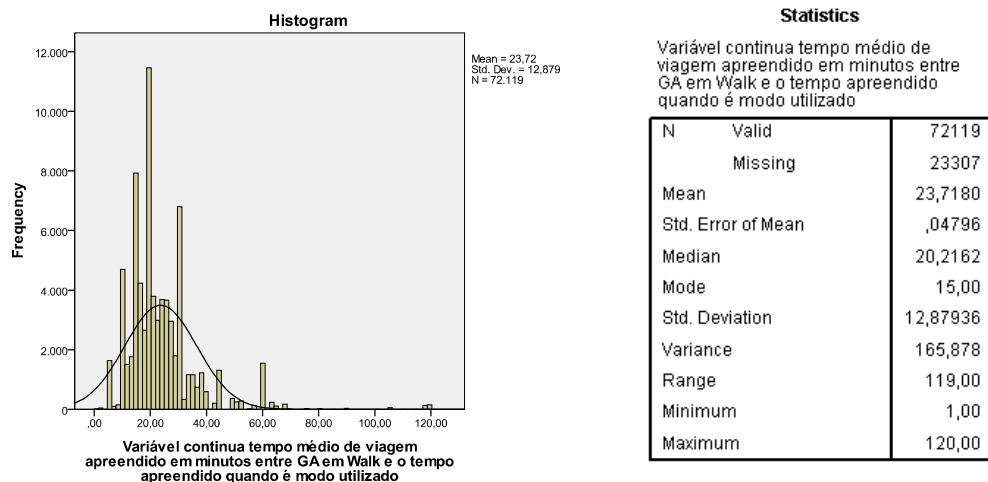


Figura IV. 13 - Dados estatísticos e distribuição da variável T5a através do *software* SPSS©

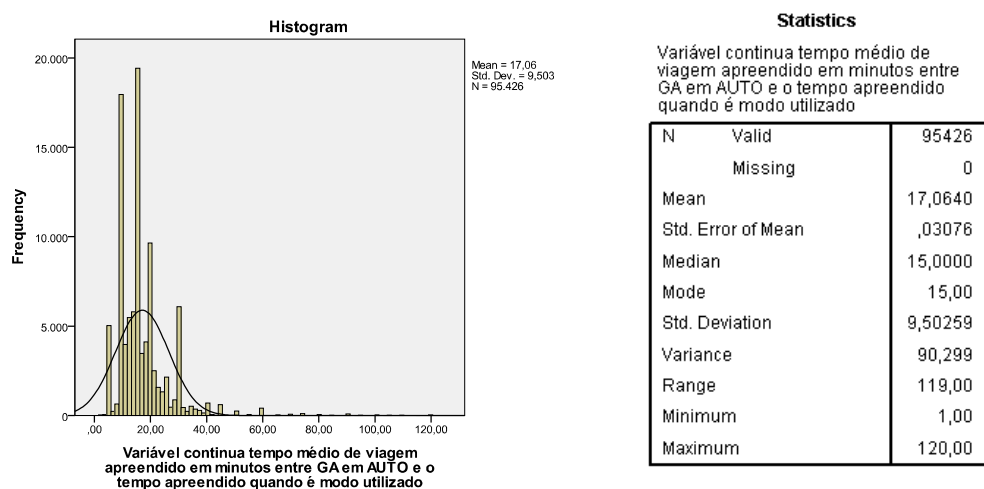


Figura IV. 14 - Dados estatísticos e distribuição da variável T6a através do *software* SPSS©

Para colmatar esta lacuna Hensher *et al.* (2005), sugere a análise do valor médio para cada modo alternativo entre cada par GA, como valor a considerar para as alternativas não escolhidas. Esta metodologia poderá colmatar apreensões erradas das variáveis excluindo valores demasiado extremos, mas retira a variabilidade de valores entre pares GA, com eventuais pontos de origem e destino bastante distintos dentro de cada zona, que poderiam justificar certas escolhas modais. Atendendo a estas limitações, foram definidas variáveis contínuas genéricas para cada modo considerado, com o tempo total de viagem apreendido para o modo utilizado e com a média do tempo total de viagem entre cada par GA por modo para as alternativas não escolhidas (Tia). O senso comum sugere que aumentos do tempo de viagem geralmente diminuem a utilidade do modo.

Koppelman e Bhat (2006), sugerem que o logaritmo do tempo de viagem (Til) apresenta melhores resultados na modelação de escolhas modais. Assim, foram também definidas novas

variáveis correspondentes aos logaritmos das variáveis genéricas anteriores para posterior teste nos modelos.

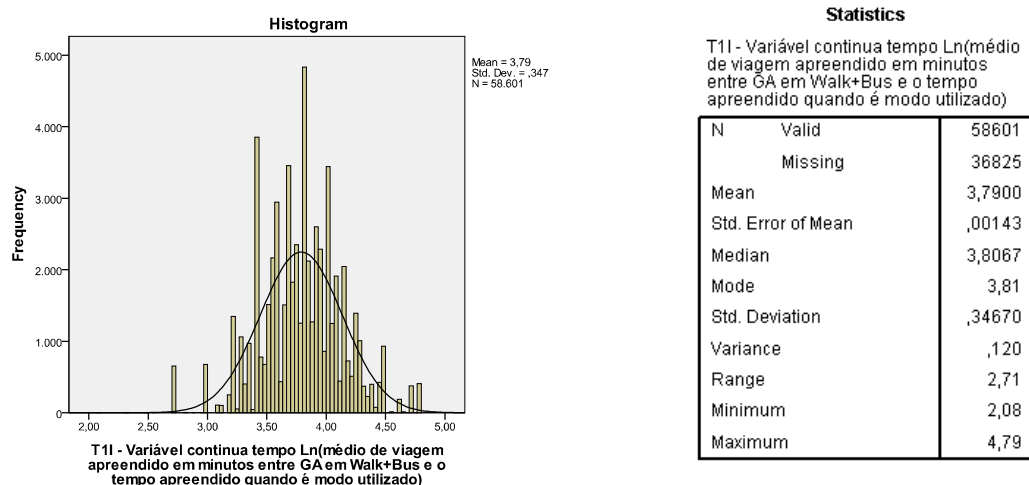


Figura IV. 15 - Dados estatísticos e distribuição da variável T1l através do *software* SPSS©

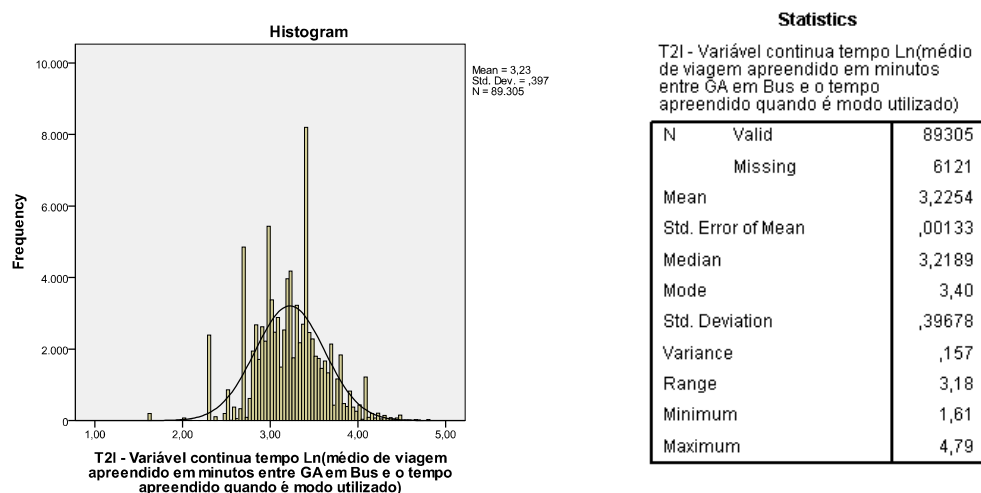


Figura IV. 16 - Dados estatísticos e distribuição da variável T2l através do *software* SPSS©

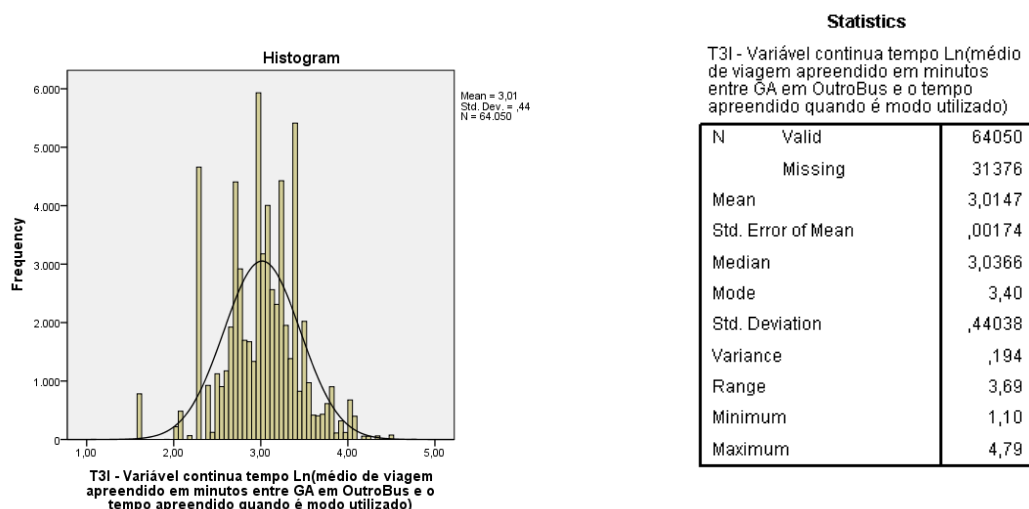
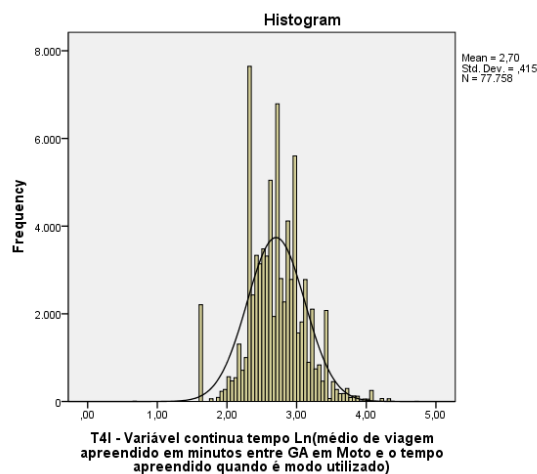


Figura IV. 17 - Dados estatísticos e distribuição da variável T3l através do *software* SPSS©

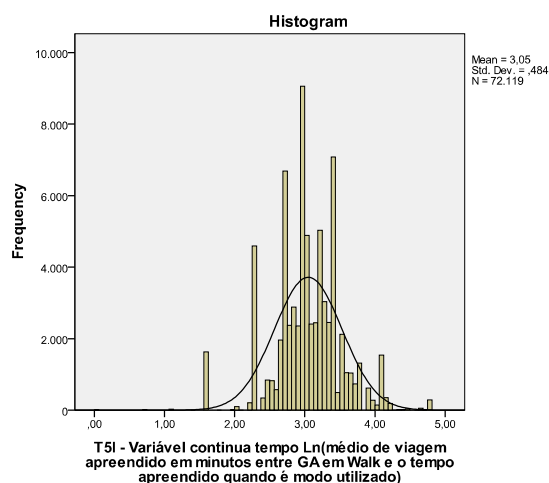


Statistics

T4l - Variável continua tempo Ln(médio de viagem apreendido em minutos entre GA em Moto e o tempo apreendido quando é modo utilizado)

N	Valid	77758
	Missing	17668
Mean		2,7042
Std. Error of Mean		,00149
Median		2,7081
Mode		2,30
Std. Deviation		,41467
Variance		,172
Range		4,09
Minimum		,69
Maximum		4,79

Figura IV. 18 - Dados estatísticos e distribuição da variável T4l através do *software* SPSS©

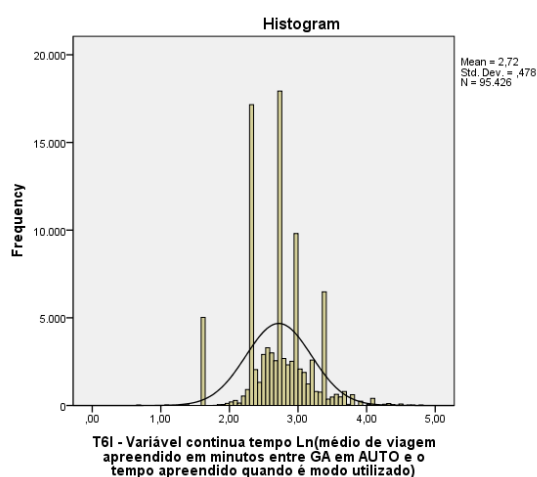


Statistics

T5l - Variável continua tempo Ln(médio de viagem apreendido em minutos entre GA em Walk e o tempo apreendido quando é modo utilizado)

N	Valid	72119
	Missing	23307
Mean		3,0492
Std. Error of Mean		,00180
Median		3,0065
Mode		2,71
Std. Deviation		,48356
Variance		,234
Range		4,79
Minimum		,00
Maximum		4,79

Figura IV. 19 - Dados estatísticos e distribuição da variável T5l através do *software* SPSS©



Statistics

T6l - Variável continua tempo Ln(médio de viagem apreendido em minutos entre GA em AUTO e o tempo apreendido quando é modo utilizado)

N	Valid	95426
	Missing	0
Mean		2,7190
Std. Error of Mean		,00155
Median		2,7081
Mode		2,71
Std. Deviation		,47849
Variance		,229
Range		4,79
Minimum		,00
Maximum		4,79

Figura IV. 20 - Dados estatísticos e distribuição da variável T6l através do *software* SPSS©

É possível constatar que a logaritmação da variável tempo de viagem torna as distribuições das mesmas muito próximas da distribuição normal.

5.2.4. Motivo da Viagem

O motivo da viagem foi codificado na base de dados original como uma variável discreta com 13 escalões apresentados no quadro IV.30.

Atendendo às amostras de cada motivo e à sua potencial influência para as escolhas modais diárias, tal como já foi referido, agruparam-se alguns motivos e purgaram-se os considerados menos relevantes da amostra. Assim, no estudo, foi realizada a seguinte agregação considerada mais adequada, em cinco motivos (Mt) atendendo aos originais da base de dados:

Mt = 1: Trabalho (motivo original 1)

Mt = 2: Regresso ao domicílio (motivo original 2)

Mt = 3: Escola (motivo original 3)

Mt = 4: Lazer (motivo original 8, 9, 10 e 11)

Mt = 5: Compras/serviços (motivo original 12)

Com esta variável pretende-se caracterizar a potencial influência do motivo da viagem nas escolhas modais. O senso comum sugere que motivos mais regulares e ou com estadias tendencialmente mais longas, como a ida para o emprego ou para a escola, possam ter maior propensão para ser realizadas por modos alternativos ao automóvel, atendendo à menor necessidade de flexibilidade/disponibilidade/rapidez do modo de transporte. O motivo escola terá ainda o catalisador de serem jovens geralmente sem carta de condução, com alguma autonomia e com condições de aquisição de título de transporte de TC mais baratos que podem potenciar o seu uso. Este último fator é comum também aos cidadãos mais idosos.

Quadro IV. 30 - Codificação dos motivos de viagem na base de dados original

Código	Motivo da Viagem
1	Trabalho
2	Regresso ao domicílio
3	Escola
4	Transporte de familiares
5	Deslocação em Serviço
6	Vendas e serviços
7	Saúde
8	Religião
9	Férias/fim-de-semana
10	Lazer
11	Visitas a familiares
12	Compras/serviços
13	Outros

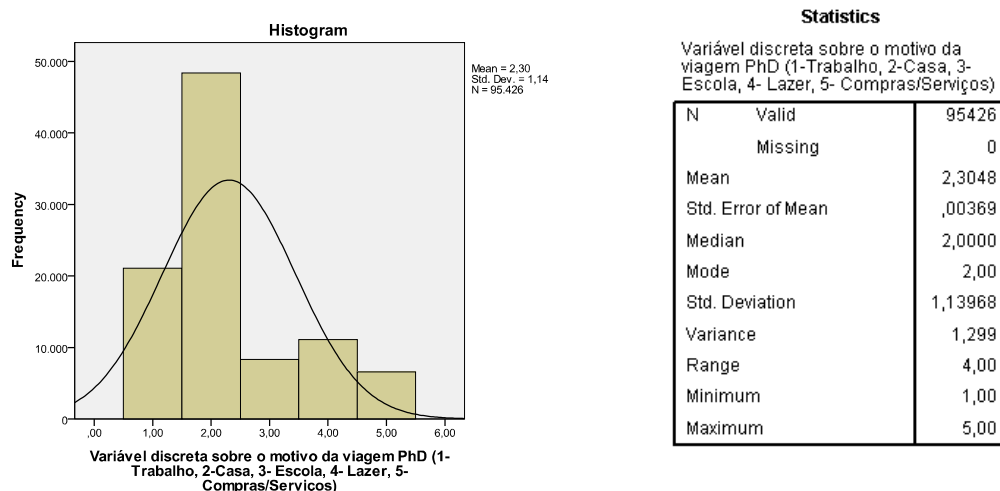


Figura IV. 21 - Dados estatísticos e distribuição da variável *Mt* através do *software* SPSS©

Para analisar a sua linearidade foram definidas variáveis binárias para os diferentes motivos de viagem.

Quadro IV. 31 - Repartição do motivo da viagem na amostra

Mt		n	%
Motivo da Viagem	Tr: Trabalho	21088	22,1%
	Rg: Regresso a casa	48380	50,7%
	Es: Escola	8305	8,7%
	Lz: Lazer	11092	11,6%
	CS: Compras/serviços	6561	6,9%
Total		95426	100,0%

A análise à repartição modal de cada motivo é apresentada no quadro IV.32:

Quadro IV. 32 - Repartição modal segundo o motivo da viagem

		Motivo da Viagem										Total
		Tr	%	Rg	%	Es	%	Lz	%	CS	%	
Modo de transporte da viagem em estudo	Bus+Walk	398	1,9%	1192	2,5%	407	4,9%	149	1,3%	118	1,8%	2264
	BUS	3386	16,1%	9278	19,2%	2781	33,5%	1684	15,2%	1194	18,2%	18324
	Outro Bus	653	3,1%	1810	3,7%	935	11,3%	137	1,2%	83	1,3%	3618
	Moto	2224	10,5%	3201	6,6%	129	1,6%	548	4,9%	273	4,2%	6375
	Walk	1819	8,6%	5910	12,2%	1377	16,6%	1594	14,4%	1020	15,5%	11721
Auto		12608	59,8%	26989	55,8%	2676	32,2%	6980	62,9%	3873	59,0%	53128
Total		21088		48380		8305		11092		6561		95426

É possível constatar que o motivo que maior utilização faz do automóvel é o Lazer. Atendendo à regra estabelecida será este o motivo a normalizar no estudo da linearidade desta variável.

5.2.5. Duração das Atividades

Da base de dados original foi possível avaliar a duração da estadia em cada atividade em minutos (De), através da diferença entre a hora de chegada de uma viagem e a hora de saída de viagem seguinte de um mesmo indivíduo. A fórmula de cálculo da duração da estadia limitou a amostra para esta variável às viagens que não eram a última do dia (61 358), uma vez que essas não tinham valor para a hora de saída da viagem seguinte. Postulou-se que atividades mais longas poderiam potenciar o uso de modos alternativos ao automóvel uma vez que não necessitavam de tanta flexibilidade e disponibilidade (note-se que este efeito pode também ser representado de forma indireta pela variável Motivo de Viagem).

Esta variável foi definida como contínua e a sua distribuição apresenta três picos distintos (ver figura IV.22). O primeiro pico engloba as atividades até 180m de duração, o segundo entre aproximadamente os 180 e 360 minutos, o terceiro verifica-se aproximadamente entre os 500 e os 700 minutos.

Atendendo a gama de valores verificada e às durações de estadias geralmente consideradas em estudos de tráfego, para avaliar a linearidade da variável definiram-se 5 variáveis binárias (quadro IV.33).

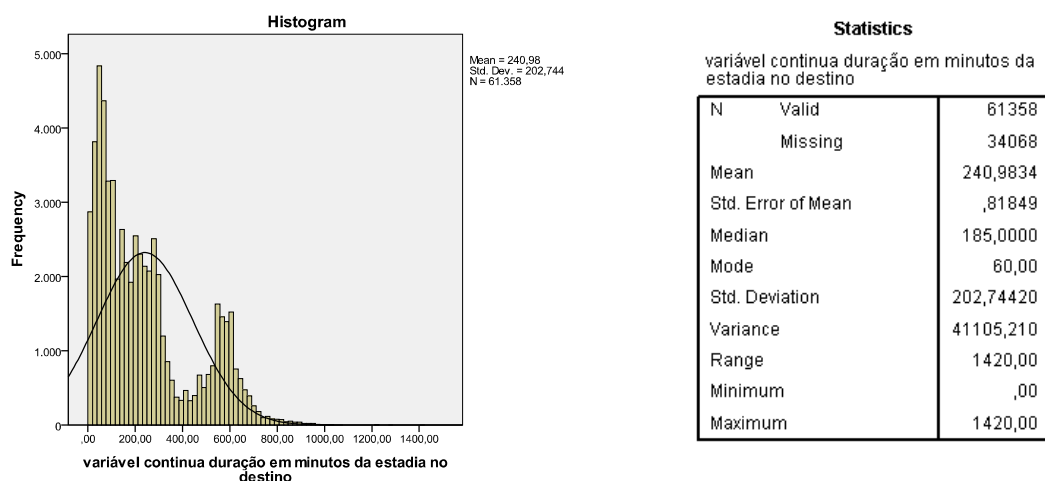


Figura IV. 22 - Dados estatísticos e distribuição da variável *duração da estadia (De)* através do software SPSS©

Quadro IV. 33 - Variáveis binárias para a duração das atividades

Variável binária	Amostra	Porcentagem
De1 (Curta duração: De ≤ 60 min.)	13281	21,6%
De2 (Média duração: De > 60 e ≤ 120 min.)	10358	16,9%
De3 (Média duração: De > 120 e ≤ 240 min.)	14097	23,0%
De4 (Longa Duração: De > 240 e ≤ 480 min.)	12252	20,0%
De5 (Muito longa Duração: De > 480 min.)	11370	18,5%
Total	61358	100,00%

A análise da repartição modal para cada variável binária (quadro IV.34) revelou que o aumento da duração da estadia apresenta um maior uso do TC. Assim, a normalização das variáveis binárias foi realizada através da sua categoria com durações menores (De1).

Quadro IV. 34 - Repartição modal para as diferentes categorias duração da atividade

		Número total de viagens diárias								Total	
		De1	%	De2	%	De3	%	De4	%	De5	%
Modo de transporte da viagem em estudo	Bus+Walk	190	1,4%	169	1,6%	266	1,9%	297	2,4%	408	3,6%
	BUS	1645	12,4%	1728	16,7%	2701	19,2%	2441	19,9%	2750	24,2%
	Outro Bus	266	2,0%	199	1,9%	332	2,4%	544	4,4%	908	8,0%
	Moto	1049	7,9%	536	5,2%	948	6,7%	1069	8,7%	872	7,7%
	Walk	1905	14,3%	1434	13,8%	2010	14,3%	1541	12,6%	742	6,5%
Auto		8226	61,9%	6292	60,7%	7840	55,6%	6360	51,9%	5690	50,0%
Total		13281		10358		14097		12252		11370	
											61358

5.2.6. Número Total de Viagens Diárias

Da base de dados original foi possível avaliar o número total de viagens diárias de um indivíduo (nV). Postulou-se que um maior número de viagens diárias poderia ser um motivo para a escolha do automóvel, devido à maior flexibilidade e disponibilidade que normalmente oferece. Esta variável foi definida como contínua (apenas com valores inteiros) e a gama de possíveis valores variou entre 1 e 18 viagens diárias.

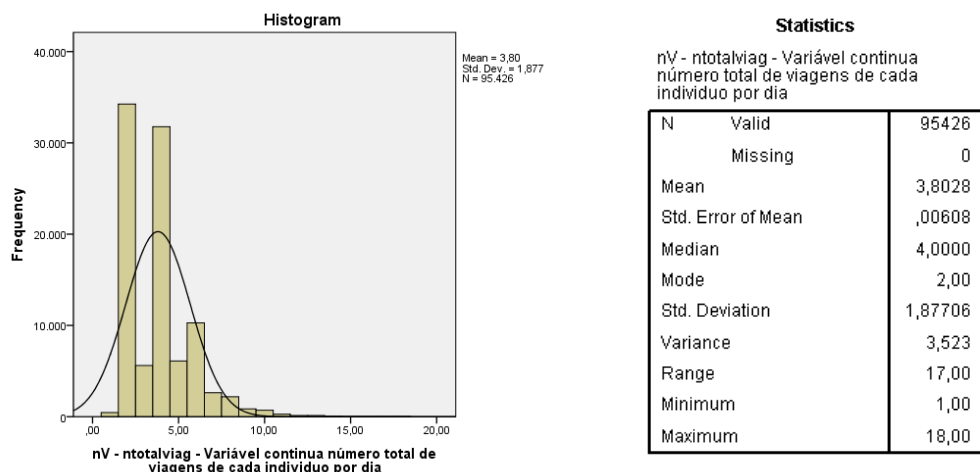


Figura IV. 23 - Dados estatísticos e distribuição da variável nV através do *software* SPSS©

É possível verificar que a maioria das viagens diárias se centra em 2 e 4 viagens diárias e que a variável apresenta uma distribuição aproximadamente exponencial. Atendendo à gama de valores verificada e para avaliar a linearidade da variável definiram-se 3 variáveis binárias, de forma a ter amostras sensivelmente iguais e que representassem grupos de situações com alguma similaridade (quadro IV.35).

Quadro IV. 35 - Variáveis binárias para o número total de viagens diárias

Variável binária	Amostra	Porcentagem
nV1 (nV ≤ 2 viag./dia)	34/03	36,4%
nV2 (nV > 2 e ≤ 4 viag./dia)	37378	39,2%
nV3a (nV > 4 viag./dia)	23345	24,5%
Total	95426	100,00%

Quadro IV. 36 - Repartição modal para as diferentes categorias do total de viagens diárias

		Número total de viagens diárias						Total
		nV1	%	nV2	%	nV3a	%	
Modo de transporte da viagem em estudo	Bus+Walk	1362	3,9%	682	1,8%	220	0,9%	2264
	BUS	9818	28,3%	6675	17,9%	1830	7,8%	18323
	Outro Bus	2199	6,3%	1051	2,8%	368	1,6%	3618
	Moto	1868	5,4%	2771	7,4%	1736	7,4%	6375
	Walk	4347	12,5%	5273	14,1%	2100	9,0%	11720
	Auto	15109	43,5%	20926	56,0%	17091	73,2%	53127
Total		34703		37378		23345		95426

A análise da repartição modal para cada variável binária (quadro IV.36) revelou que o aumento do número de viagens diárias apresenta um maior uso do automóvel. Assim, a normalização das variáveis binárias foi realizada através da sua categoria mais elevado (nV3a).

5.3. Variáveis Caracterizadoras das Políticas de Mobilidade

5.3.1. Variáveis Caracterizadoras das Políticas de Estacionamento

5.3.1.1. Densidade de Oferta face à Procura de Estacionamento

Atendendo à informação recolhida sobre o estacionamento na cidade do Porto e à sua correção para o ano de 2000, procurou-se definir variáveis que teoricamente pudessem influenciar as escolhas modais. A partir dos dados obtidos sobre a oferta de estacionamento nas zonas de atração do concelho do Porto foi definida a densidade de oferta de estacionamento face à procura de estacionamento (OV). Esta variável foi calculada através da razão entre a oferta total de estacionamentos existentes na zona de atração (pagos ou gratuitos, em parque e na via), pelo número total de viagens para a mesma zona de atração. Estas viagens totais para a zona de atração foram determinadas para a base de dados completa, independentemente do modo utilizado, e com valores extrapolados (utilizando os ponderadores de cada viagem da base de dados original) de modo a simular toda a procura existente. A sua formulação torna a variável contínua e o senso comum sugere que um aumento da densidade de oferta de estacionamento face à procura tenderá a potenciar o uso do automóvel. Mais uma vez a variável está limitada a um numerador apenas com 26 possíveis resultados.

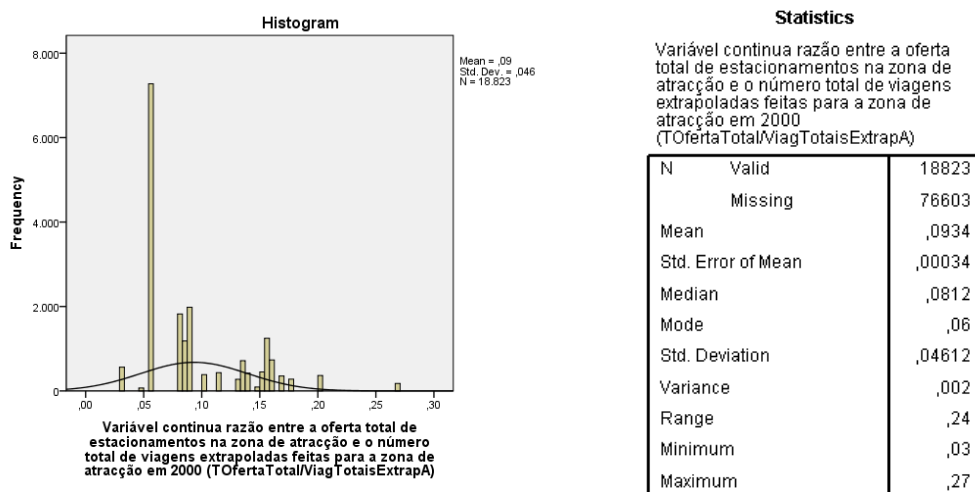


Figura IV. 24 - Dados estatísticos e distribuição da variável OV através do *software* SPSS©

Na figura IV.24 é possível constatar que a variável não apresenta uma distribuição definida e que os seus valores variam entre 0,03 e 0,27.

Atendendo aos valores verificados e aos tamanhos das amostras definiram-se quatro variáveis binárias de modo a permitir avaliar a linearidade do seu impacto (quadro IV.37).

Quadro IV. 37 - Variáveis binárias para a densidade de oferta de estacionamento face à procura de estacionamento

Variável binária	Amostra	Percentagem
OV1 (OV ≤ 0,075)	7908	42,0%
OV2 (OV > 0,075 e ≤ 0,1)	4987	26,5%
OV3 (OV > 0,1 e ≤ 0,15)	2322	12,3%
OV4 (OV > 0,15)	3606	19,2%
Total	18823	100,00%

Quadro IV. 38 - Repartição modal para as diferentes categorias da densidade de oferta de estacionamento face à procura de estacionamento vago

		Probabilidade de encontrar um lugar vago								Total
		OV1	%	OV2	%	OV3	%	OV4	%	
Modo de transporte da viagem em estudo	BUS+walk	307	3,9%	190	3,8%	75	3,2%	86	2,4%	658
	BUS	3265	41,3%	1743	35,0%	448	19,3%	789	21,9%	6246
	Outro BUS	32	0,4%	13	0,3%	11	0,5%	45	1,2%	101
	Moto	58	0,7%	55	1,1%	11	0,5%	27	0,7%	151
	Walk	2067	26,1%	1225	24,6%	1111	47,8%	672	18,6%	5076
	Auto	2179	27,6%	1761	35,3%	666	28,7%	1987	55,1%	6594
Total		7908		4987		2322		3606		18823

Do quadro IV.38 não é possível dizer que a influência da variável é inequívoca mas verifica-se uma tendência para uma maior utilização do automóvel nas zonas com maior densidade de

oferta de estacionamento face à procura. Esta tendência justifica a normalização das variáveis binárias em função da categoria mais elevada (OV4).

5.3.1.2. Proporção de Estacionamento Pago

A outra variável definida foi a proporção de estacionamentos pagos, calculados a partir da razão entre o total de estacionamentos pagos (na via e em parque) e o total de oferta de estacionamento na zona de atração de viagens (PO). Relembra-se que a distinção entre zonas de destino e de atração é crucial para a análise de oferta de estacionamento, uma vez que nas viagens de regresso a casa a origem é a zona de atração e não de geração e, portanto, a oferta de estacionamento que eventualmente possa influenciar a escolha do modo é a da zona de origem/atração e não em casa. O senso comum sugere que o aumento da proporção de estacionamentos pagos deverá diminuir a atratividade de uso do automóvel e aumentar o uso das alternativas.

A variável foi definida como contínua, mas está limitada a dados de apenas 26 zonas de atração, reduzindo a amostra com dados para 18 823 inquéritos. A variável não apresenta uma distribuição distinta e varia entre o valor 0 e 0,88 (figura IV.25).

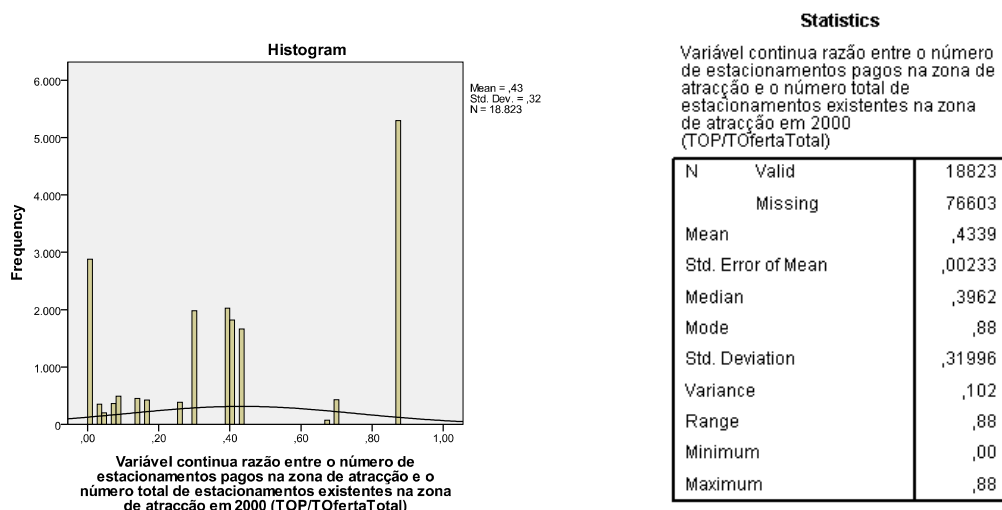


Figura IV. 25 - Dados estatísticos e distribuição da variável PO através do *software* SPSS©

O estudo da sua linearidade levou à definição de três variáveis binárias, atendendo aos possíveis valores e à definição de amostras relativamente semelhantes (quadro IV.39).

Quadro IV. 39 - Variáveis binárias para a proporção de estacionamento pago

Variável binária	Amostra	Percentagem
PO1 (PO ≤ 0,25)	5156	27,4%
PO2 (PO > 0,25 e ≤ 0,5)	7870	41,8%
PO3 (PO > 0,5)	5797	30,8%
Total	18823	100,00%

O estudo da repartição modal para estas variáveis binárias confirmou um menor uso do automóvel com o aumento de lugares de estacionamento pago (quadro IV.40). Este resultado levou à normalização da categoria mais baixa, tendencialmente mais favorável à utilização do automóvel (PO1).

Quadro IV. 40 - Repartição modal para as diferentes categorias da proporção de estacionamentos pagos

		Proporção de estacionamentos pagos						Total
		PO1	%	PO2	%	PO3	%	
Modo de transporte da viagem em estudo	Bus+Walk	137	2,7%	295	3,7%	226	3,9%	658
	BUS	1220	23,7%	2601	33,0%	2424	41,8%	6246
	Outro Bus	52	1,0%	37	0,5%	12	0,2%	101
	Moto	41	0,8%	69	0,9%	41	0,7%	151
	Walk	1110	21,5%	2415	30,7%	1550	26,7%	5076
	Auto	2596	50,3%	2453	31,2%	1544	26,6%	6594
Total		5156		7870		5797		18823

5.3.1.3. Combinação das Variáveis Binárias sobre Políticas de Estacionamento

Atendendo aos resultados anteriores colocou-se a hipótese de definir variáveis binárias para a combinação das variáveis anteriores. Para estas variáveis binárias analisou-se as amostras e repartições modais para as diferentes combinações (quadros IV.41, IV.42 e IV.43).

Quadro IV. 41 - Amostras para a possível combinação de variáveis binárias sobre políticas de estacionamento

Amostras		Pior Auto			Total
		PO1 $\leq 0,25$	PO2 $> 0,25$ & $\leq 0,5$	PO3 $> 0,5$	
Pior Auto ↑	OV1 $\leq 0,075$	0	2342	5366	7908
	OV2 $> 0,075$ & $\leq 0,1$	320	4667	0	4987
	OV3 $> 0,1$ & $\leq 0,15$	1230	661	431	2322
	OV4 $> 0,15$	3606	0	0	3606
	Total	5156	7870	5797	18823

Quadro IV. 42 - Amostras de pares GA para a possível combinação de variáveis binárias sobre políticas de estacionamento

Amostras		Pior Auto		
		PO1 $\leq 0,25$	PO2 $> 0,25$ & $\leq 0,5$	PO3 $> 0,5$
Pior Auto ↑	OV1 $\leq 0,075$	0	27 Pares; 21 Gera; 2 Atrac	39 Pares; 38 Gera; 2 Atrac
	OV2 $> 0,075$ & $\leq 0,1$	14 Pares; 12 Gera; 4 Atrac	48 Pares; 31 Gera; 3 Atrac	0
	OV3 $> 0,1$ & $\leq 0,15$	5 Pares; 4 Gera; 2 Atrac	8 Pares; 8 Gera; 2 Atrac	6 Pares; 6 Gera; 1 Atrac
	OV4 $> 0,15$	36 Pares; 19 Gera; 7 Atrac	0	0

Apesar de as repartições modais das diferentes combinações não apresentarem resultados muito evidentes, verificam-se tendências coerentes de uma diminuição do uso do automóvel quando há um aumento nas restrições ao estacionamento. A indefinição de uma tendência clara poderá derivar da reduzida amostra de pares GA com dados sobre o estacionamento de algumas combinações. Ainda assim, definiram-se variáveis binárias de combinação de

variáveis sobre estacionamentos cuja repartição e definição são apresentados no quadro IV.44.

Quadro IV. 43 - Repartições modais para possível combinação de variáveis binárias sobre políticas de estacionamento

Rep. Modal	Pior Auto					
	PO1 <=0,25		PO2 >0,25 & <=0,5		PO3 >0,5	
OV1 <=0,075	BUS	0	BUS	34,7%	BUS	44,4%
	Walk		Walk	32,4%	Walk	23,2%
	Auto		Auto	27,7%	Auto	27,5%
OV2 >0,075 & <=0,1	BUS	36,3%	BUS	34,9%	BUS	0
	Walk	10,9%	Walk	25,5%	Walk	
	Auto	49,1%	Auto	34,4%	Auto	
OV3 >0,1 & <=0,15	BUS	25,6%	BUS	14,1%	BUS	9,3%
	Walk	32,8%	Walk	60,7%	Walk	71,2%
	Auto	36,7%	Auto	21,8%	Auto	16,2%
OV4 >0,15	BUS	21,9%	BUS	0	BUS	0
	Walk	18,6%	Walk		Walk	
	Auto	55,1%	Auto		Auto	

Apesar de as repartições modais das diferentes combinações não apresentarem resultados muito evidentes, verificam-se tendências coerentes de uma diminuição do uso do automóvel quando há um aumento nas restrições ao estacionamento. A indefinição de uma tendência clara poderá derivar da reduzida amostra de pares GA com dados sobre o estacionamento de algumas combinações. Ainda assim, definiram-se variáveis binárias de combinação de variáveis sobre estacionamentos cuja repartição e definição são apresentados no quadro IV.44.

Quadro IV. 44 - Amostras das variáveis binárias de combinação de variáveis de políticas de estacionamento

Variável binária	Amostra	Porcentagem
P13 (OV1=1 & PO3=1)	5366	39,9%
P12 (OV1=1 & PO2=1)	2542	18,9%
P22 (OV2=1 & PO2=1)	4667	34,7%
P21 (OV2=1 & PO1=1)	320	2,4%
P33 (OV3=1 & PO3=1)	431	3,2%
P32 (OV3=1 & PO2=1)	661	4,9%
P31 (OV3=1 & PO1=1)	1230	9,1%
P41 (OV4=1 & PO1=1)	3606	26,8%
Total	13457	100,0%

A ordem pela qual foram apresentadas as variáveis no quadro IV.44 reflete uma ordenação decrescente em termos de restrições ao uso do automóvel.

Do quadro IV.43 é possível verificar que efetivamente a variável P41 apresenta a maior repartição modal para o automóvel, tendo sido a variável escolhida para normalização.

5.3.2. Variáveis Caracterizadoras das Políticas de Transportes Coletivos

Este tipo de variável pretende quantificar a sua potencial influência nas opções dos inquiridos, de forma a avaliar quais as políticas que maior impacto poderão ter nas suas escolhas modais diárias. Apenas se avaliaram os trajetos de TC diretos, isto é sem transbordos, por se considerar que estes são claramente os que têm maior potencial para servir de alternativa ao TI e pela dificuldade na obtenção de dados sobre esses transbordos. Da variedade de indicadores de desempenho referidos em muitos estudos internacionais sobre a qualidade dos serviços de TC, a velocidade comercial, o tempo de viagem e a frequência horária são dos mais frequentes e valorizados pelos utilizadores (Seco e Gonçalves, 2007).

5.3.2.1. Razão Tempo de Viagem por Bus/Tempo de Viagem por Auto

Uma variável geralmente valorizada pelos utilizadores de TC é o tempo de viagem. No presente estudo postulou-se que mais relevante que o tempo de viagem em si, já considerado na variável genérica (Tia), tenderia a ser a diferença ou o rácio dos tempos de viagem entre TC (T2a) e automóvel (T6a). O senso comum sugere que quanto menor for esta relação, e portanto menor for o tempo de viagem por TC relativamente ao automóvel, maior seria a sua utilidade. Assim definiu-se uma variável razão entre o tempo total de viagem de BUS e o tempo total de viagem por automóvel (TBA). A sua formulação foi definida como:

$$TBA [\%] = T2a [\text{min.}] / T6a [\text{min.}]$$

Exp. IV - 5

A sua especificação é contínua e na figura IV.26 é possível constatar que apresenta uma distribuição aproximadamente lognormal. Tal como para a variável velocidade comercial equivalente, a utilização de tempos médios de viagens entre as zonas de GA quando o modo Bus não é a escolha para a viagem levou à exclusão dos inquéritos de viagens entre zonas sem viagens de Bus.

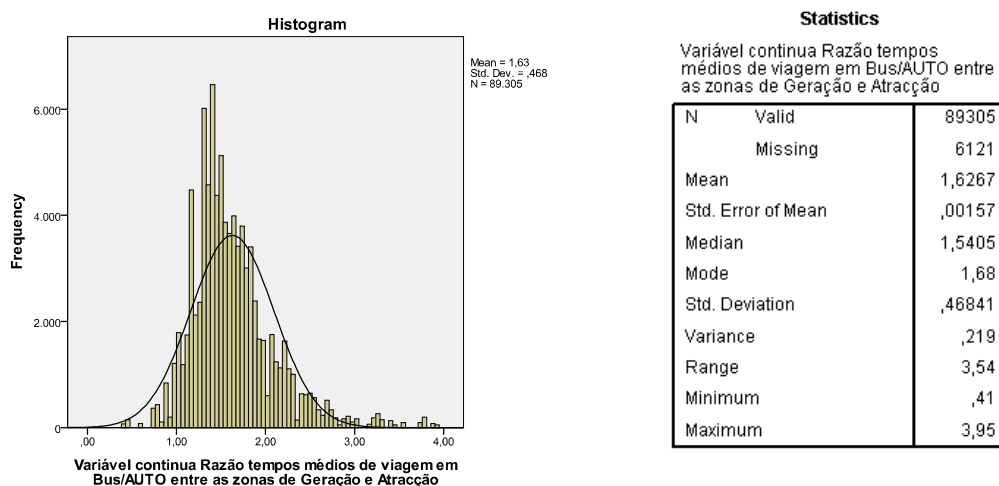


Figura IV. 26 - Dados estatísticos e distribuição da variável TBA através do *software* SPSS©

Para avaliar a sua linearidade definiram-se quatro variáveis binárias atendendo às gamas de possíveis valores e à obtenção de amostras com tamanho razoável (quadro IV.45). A análise das suas repartições modais permitiu verificar um aumento da utilização do automóvel com o aumento da relação dos tempos de viagem e uma diminuição da utilização dos BUS (quadro IV.46). Estes resultados sugerem que, seguindo o critério para a seleção da variável a normalizar, seja a gama de valores mais elevados a variável a excluir do modelo.

Quadro IV. 45 - Amostras das variáveis binárias da razão tempo de viagem BUS/AUTO

Variável binária	Amostra	Percentagem
TB1 (TBA≤1)	3503	3,9%
TB2 (TBA>1 e ≤1,5)	36419	40,8%
TB3 (TBA>1,5 e ≤2)	34695	38,9%
TB4 (TBA>2)	14688	16,4%
Total	89305	100,00%

Quadro IV. 46 - Repartição modal para as diferentes gamas de relações tempos de viagem Bus/AUTO

		Razão tempo médio de viagem BUS/AUTO								Total
		TB1	%	TB2	%	TB3	%	TB4	%	
Modo de transporte da viagem em estudo	Bus+Walk	120	3,4%	1023	2,8%	794	2,3%	308	2,1%	2245
	BUS	820	23,4%	8159	22,4%	7511	21,6%	1833	12,5%	18324
	Outro Bus	54	1,5%	1170	3,2%	1360	3,9%	677	4,6%	3261
	Moto	197	5,6%	1928	5,3%	2080	6,0%	1124	7,7%	5329
	Walk	917	26,2%	5006	13,7%	3495	10,1%	1698	11,6%	11116
Auto		1395	39,8%	19133	52,5%	19455	56,1%	9048	61,6%	49032
Total		3503		36419		34695		14688		89305

5.3.2.2. Frequência Média Horária de BUS

Outro indicador de desempenho dos TC geralmente valorizado pelos seus utilizadores é a frequência horária do serviço. Este indicador traduz a disponibilidade do serviço em determinados períodos do dia e deverá permitir uma organização prévia das atividades diárias a quem pretende utilizar os serviços de TC. É importante não só para os utilizadores diários mas também para os esporádicos, uma vez que frequências muito baixas tornam a utilização do sistema de TC muito menos atrativa em eventuais atrasos nas chegadas às paragens ou interrupções do serviço. Atendendo aos diferentes motivos de viagem considerados e aos períodos do dia em que geralmente acontecem as respetivas viagens (Gonçalves *et al.*, 2011), foram definidas quatro variáveis contínuas para diferentes períodos diários:

- Fm: Frequência média horária no período da manhã entre as 7:30 e as 9:30;
- Fa: Frequência média horária no período do Meio da manhã entre as 10:30 e as 12h30;
- Ft: Frequência média horária no período da tarde entre as 17:00 e as 19:30;

- Ftl: Frequência média horária no período laboral entre as 7:30 e as 19:30.

Considerou-se que os TC teoricamente estariam à disposição de qualquer pessoa que viajasse entre as zonas de geração e de atração, sendo uma variável comum para qualquer indivíduo independentemente do modo de transporte utilizado. O senso comum sugere que o seu aumento promove o uso dos BUS mas não garante a diminuição do uso dos TI. Durante o estudo assumiu-se que para pares GA muito próximos (até 2 km) os TC não eram alternativa ao modo pedonal, anulando-se as suas frequências. No entanto, ponderaram-se outras formas de tratar essas zonas, definindo-se ainda mais dois conjuntos das mesmas variáveis com ligeiras alterações na sua definição:

- Fi: Frequências médias horárias no período *i* com valor zero para os pares GA muito próximos;
- Fi0: Frequências médias horárias no período *i* onde os pares de GA muito próximos foram excluídos na análise, com diminuição da amostra;
- Fi1: Frequências médias horárias no período *i* onde nos pares de GA muito próximos foi considerada toda a potencial oferta verificada;

Nas figuras IV.27 à IV.38 apresentam-se as distribuições das diferentes variáveis consideradas.

É possível constatar que as variáveis apresentam características contínuas e distribuições aproximadamente exponenciais. A falta de dados entre alguns pares de GA, quer por não existir informação sobre TC para essas zonas (excluídas para estas análises), quer por não existir serviço dos mesmos, levou a uma redução considerável da amostra.

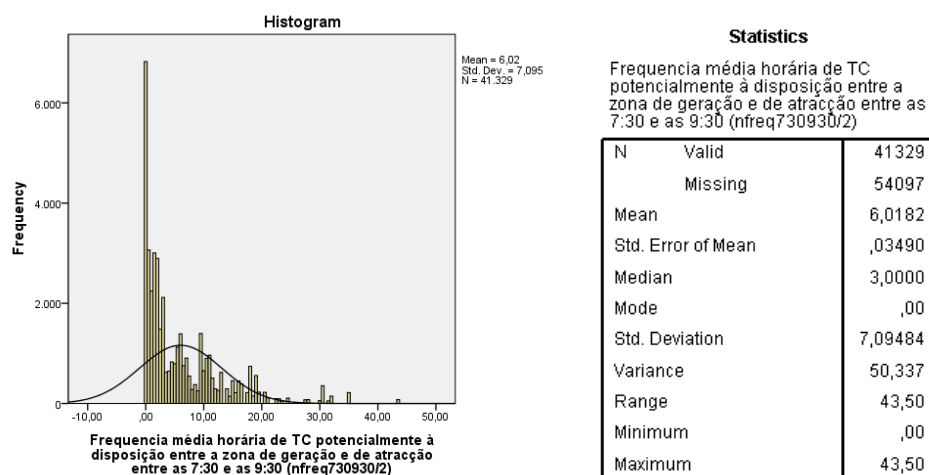
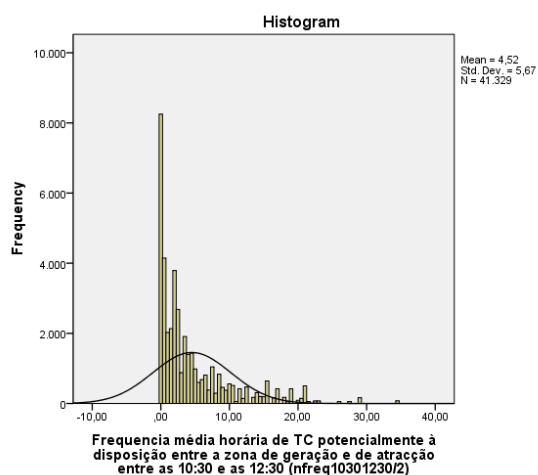


Figura IV. 27 - Dados estatísticos e distribuição da variável Fm através do *software* SPSS©

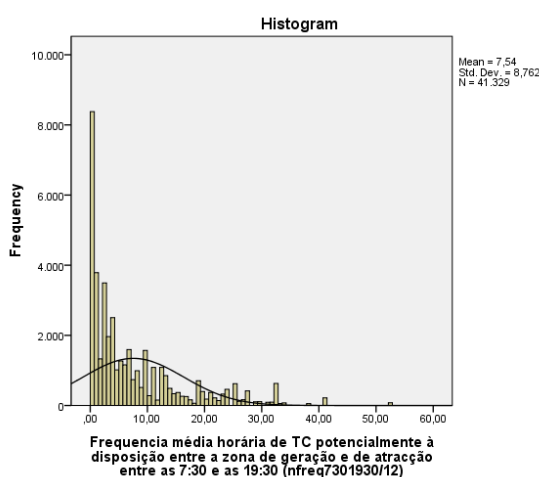


Statistics

Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 10:30 e as 12:30 (nfreq10301230/2)

N	Valid	41329
	Missing	54097
Mean		4,5221
Std. Error of Mean		,02789
Median		2,5000
Mode		,00
Std. Deviation		5,67031
Variance		32,152
Range		34,50
Minimum		,00
Maximum		34,50

Figura IV. 28 - Dados estatísticos e distribuição da variável Fa através do *software* SPSS©

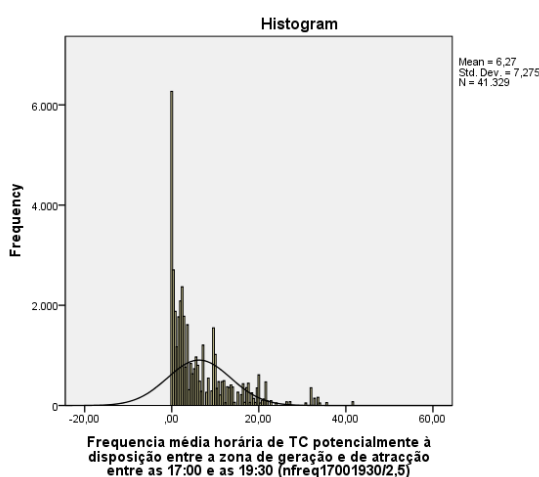


Statistics

Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 7:30 e as 19:30 (nfreq7301930/12)

N	Valid	41329
	Missing	54097
Mean		7,5450
Std. Error of Mean		,04310
Median		3,9167
Mode		,00
Std. Deviation		8,76217
Variance		76,776
Range		52,75
Minimum		,00
Maximum		52,75

Figura IV. 29 - Dados estatísticos e distribuição da variável Ft através do *software* SPSS©



Statistics

Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 17:00 e as 19:30 (nfreq17001930/2,5)

N	Valid	41329
	Missing	54097
Mean		6,2712
Std. Error of Mean		,03579
Median		3,2000
Mode		,00
Std. Deviation		7,27519
Variance		52,928
Range		41,60
Minimum		,00
Maximum		41,60

Figura IV. 30 - Dados estatísticos e distribuição da variável Ftl através do *software* SPSS©

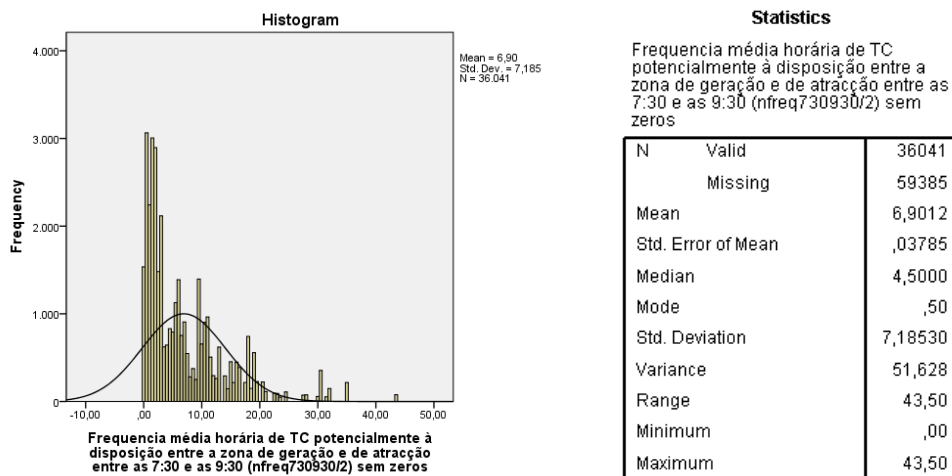


Figura IV. 31 - Dados estatísticos e distribuição da variável Fm0 através do *software* SPSS©

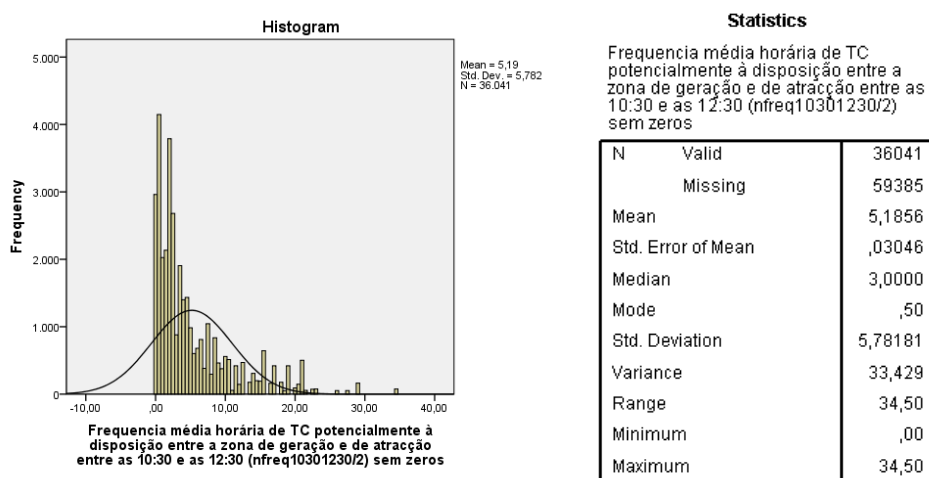


Figura IV. 32 - Dados estatísticos e distribuição da variável Fa0 através do *software* SPSS©

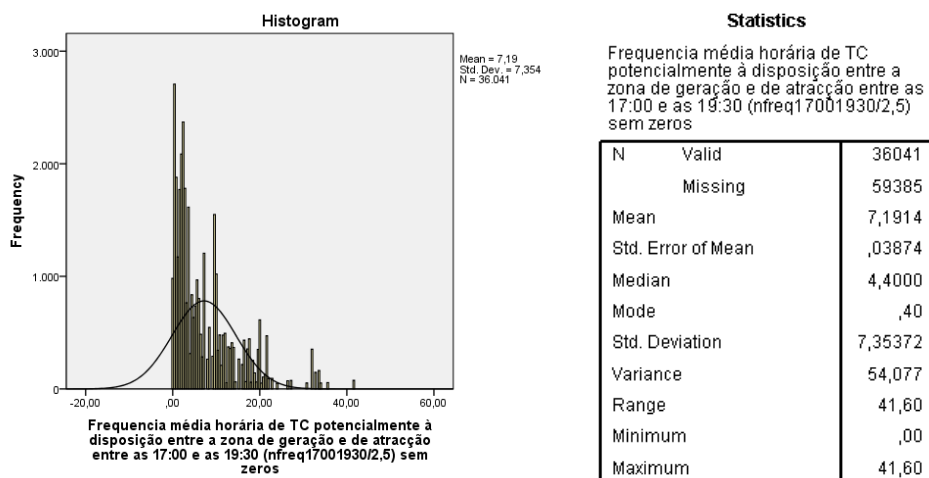
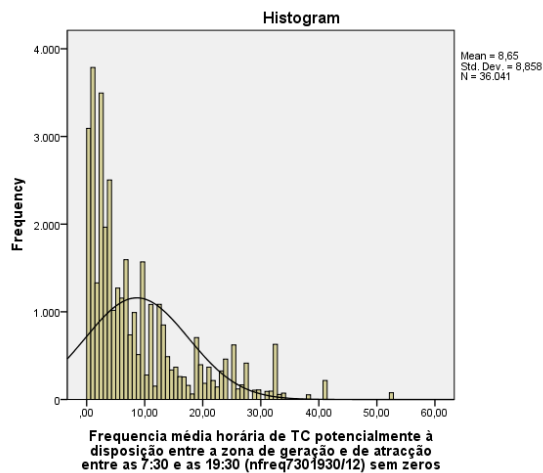
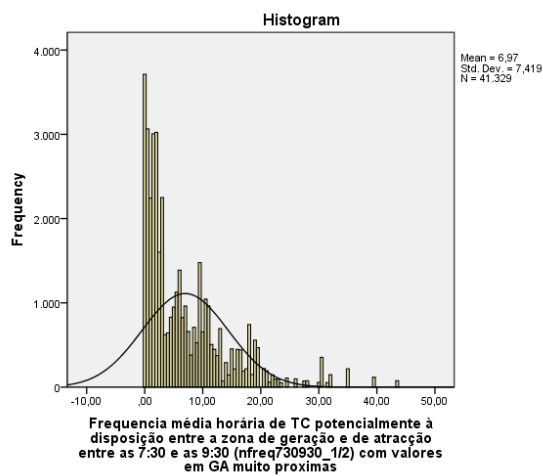


Figura IV. 33 - Dados estatísticos e distribuição da variável Ft0 através do *software* SPSS©



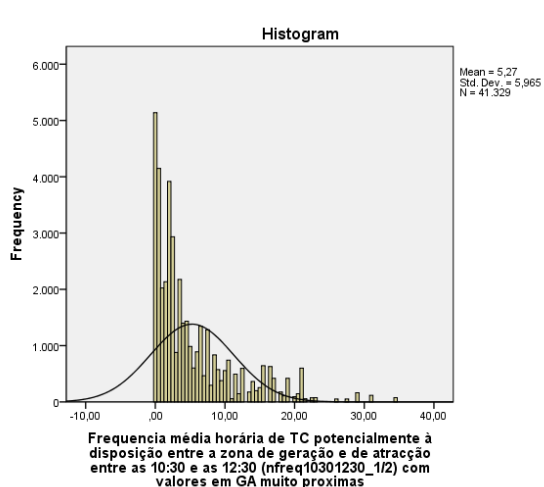
Statistics		
Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 7:30 e as 19:30 (nfreq7301930/12) sem zeros		
N	Valid	36041
	Missing	59385
Mean		8,6520
Std. Error of Mean		,04666
Median		5,2500
Mode		,92
Std. Deviation		8,85790
Variance		78,462
Range		52,67
Minimum		,08
Maximum		52,75

Figura IV. 34 - Dados estatísticos e distribuição da variável Ftl0 através do *software* SPSS©



Statistics		
Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 7:30 e as 9:30 (nfreq730930_1/2) com valores em GA muito próximas		
N	Valid	41329
	Missing	54097
Mean		6,9669
Std. Error of Mean		,03649
Median		4,5000
Mode		,00
Std. Deviation		7,41921
Variance		55,045
Range		43,50
Minimum		,00
Maximum		43,50

Figura IV. 35 - Dados estatísticos e distribuição da variável Fm1 através do *software* SPSS©



Statistics		
Frequecia média horária de TC potencialmente à disposição entre a zona de geração e de atracção entre as 10:30 e as 12:30 (nfreq10301230_1/2) com valores em GA muito próximas		
N	Valid	41329
	Missing	54097
Mean		5,2707
Std. Error of Mean		,02934
Median		3,0000
Mode		,00
Std. Deviation		5,96476
Variance		35,578
Range		34,50
Minimum		,00
Maximum		34,50

Figura IV. 36 - Dados estatísticos e distribuição da variável Fa1 através do *software* SPSS©

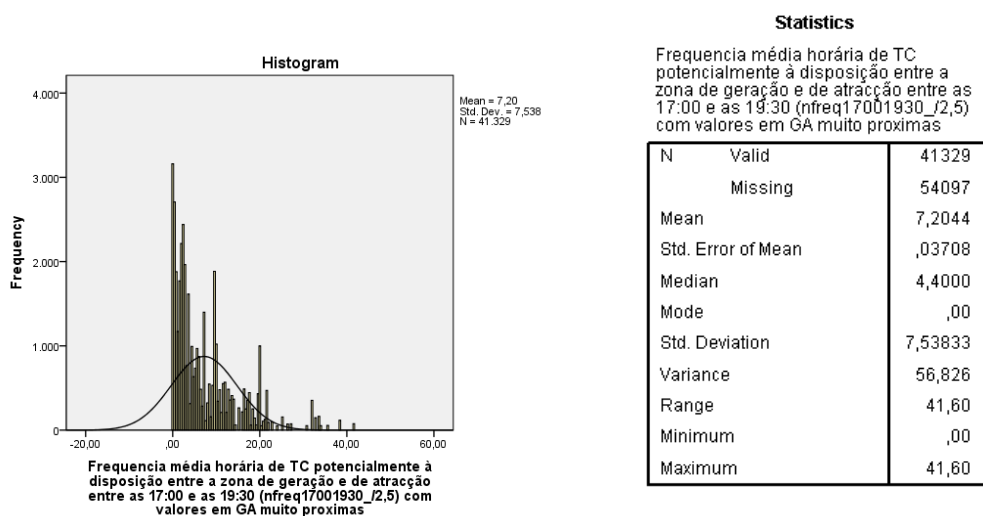


Figura IV. 37 - Dados estatísticos e distribuição da variável Ft1 através do *software* SPSS©

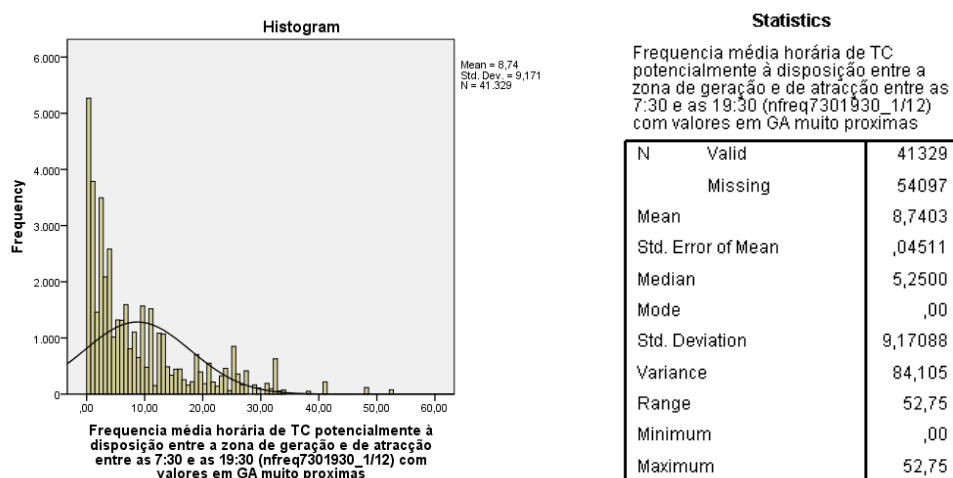


Figura IV. 38 - Dados estatísticos e distribuição da variável Ftl1 através do *software* SPSS©

Quadro IV. 47 - Amostras das variáveis binárias sobre a frequência média horária diária sem exclusão das áreas muito próximas

Variável binária	Amostra	Percentagem
Fl1 (Ftl1<=2TC/h)	10514	25,4%
Fl2 (Ftl1>2 e <=4TC/h)	7418	17,9%
Fl3(Ftl1>4 e <=12TC/h)	12274	29,7%
Fl4 (Ftl1>12TC/h)	11123	26,9%
Total	41329	100,00%

Para a avaliação da sua linearidade definiram-se variáveis binárias atendendo à gama de valores verificados e aos geralmente associadas a diferentes qualidades de serviço (Seco e Gonçalves, 2006). Nos exercícios de modelação das escolhas modais optou-se por utilizar a frequência média horária diária para caracterizar a oferta de TC sem exclusão das áreas mais

próximas (Ftl1), apresentando-se as variáveis binárias apenas para esta variável (quadro IV.47 e quadro IV.48).

Quadro IV. 48 - Repartição modal para as diferentes gamas de frequências médias horárias diárias

		Frequência média horária diária								Total
		F1	%	F2	%	F3	%	F4	%	
Modo de transporte da viagem em estudo	Bus+Walk	165	1,6%	199	2,7%	339	2,8%	299	2,7%	1002
	BUS	1464	13,9%	1825	24,6%	3136	25,5%	3454	31,1%	9880
	Outro Bus	426	4,1%	163	2,2%	124	1,0%	74	0,7%	787
	Moto	571	5,4%	219	3,0%	324	2,6%	116	1,0%	1230
	Walk	1722	16,4%	1150	15,5%	1685	13,7%	2765	24,9%	7322
	Auto	6166	58,6%	3862	52,1%	6666	54,3%	4415	39,7%	21111
Total		10514		7418		12274		11123		41329

Dos resultados do quadro IV.48 verifica-se que aparentemente em zonas com maior frequência horária existe uma maior utilização do Bus e uma menor utilização do automóvel. Este resultado levou a que a normalização das variáveis binárias fosse feita através da variável correspondente às frequências mais baixas (F1).

5.3.2.3. Velocidade comercial equivalente dos BUS

A velocidade comercial, calculada através do tempo que o veículo demora desde a paragem inicial até à paragem final, incluindo atrasos resultantes das paragens e das condições de tráfego, é um dos indicadores de serviço mais valorizados pelos utilizadores de TC (Seco e Gonçalves, 2007). No presente estudo realizou-se uma adaptação utilizando uma *velocidade comercial equivalente* em km/h, restringindo a análise apenas ao modo BUS (V2). A sua análise deverá permitir comparar a velocidade da viagem entre diferentes trajetos e uma viagem equivalente utilizando trajeto mais direto, assumindo-se que seria o utilizado com o TI ou pela melhor carreira de TC.

Para os inquéritos em que o modo escolhido não era o Bus, a velocidade comercial equivalente foi determinada com recurso à média dos tempos de viagem apreendidos de todas as viagens de Bus entre cada par GA. A ausência de pares GA com viagens feitas de Bus, levou a uma ligeira redução da amostra para 89 305 inquéritos com informação completa. A forma de cálculo desta variável dá-lhe características contínuas e os dados estatísticos da mesma são apresentados na figura IV.39.

Esta variável tem uma distribuição lognormal e o senso comum sugere que o aumento da velocidade comercial equivalente de Bus tenderia a aumentar a sua utilidade.

Para avaliar a sua linearidade, definiram-se 3 variáveis binárias atendendo às gamas de valores geralmente observadas em sistemas reais (Seco e Gonçalves, 2006).

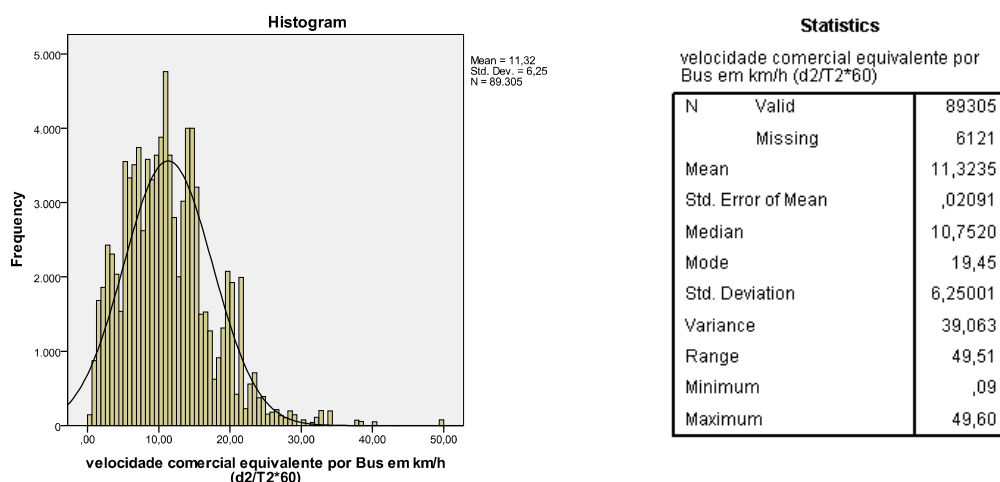


Figura IV. 39 - Dados estatísticos e distribuição da variável V2 através do *software* SPSS©

Quadro IV. 49 - Amostras das variáveis binárias da velocidade comercial equivalente de BUS

Variável binária	Amostra	Porcentagem
V21 (V2<=10km/h)	40410	45,2%
V22 (V2>10 e <=20km/h)	40284	45,1%
V23 (V2>20km/h)	8611	9,6%
Total	89305	100,00%

Quadro IV. 50 - Repartição modal para as diferentes gamas de velocidades comerciais equivalentes de BUS

		Velocidade comercial equivalente de BUS						Total
		V21	%	V22	%	V23	%	
Modo de transporte da viagem em estudo	Bus+Walk	995	2,5%	918	2,3%	332	3,9%	2245
	BUS	7888	19,5%	8953	22,2%	1482	17,2%	18323
	Outro Bus	1011	2,5%	1809	4,5%	441	5,1%	3261
	Moto	1788	4,4%	2883	7,2%	658	7,6%	5329
	Walk	10883	26,9%	216	0,5%	17	0,2%	11116
Auto		17845	44,2%	25505	63,3%	5681	66,0%	49032
Total		40410		40284		8611		89305

Da análise das repartições modais (quadro IV.50), verifica-se que o aumento da velocidade comercial equivalente não apresenta resultados inequívocos de um aumento do uso do modo Bus, verificando-se paralelamente um aumento relativo do uso do automóvel. Estes resultados podem dever-se a diferentes fatores:

- Erros na definição da variável;
- A formulação utilizada não ser a mais correta;
- A influência da velocidade de Bus não ser adequadamente retratada pela especificação utilizada;

Das possíveis causas para os resultados verificados a última parece ser a mais razoável. Geralmente as velocidades comerciais mais elevadas de Bus acontecem fora das áreas urbanas consolidadas, onde o seu serviço e competitividade face ao automóvel são drasticamente reduzidos nomeadamente devido ao efeito resultante de uma redução significativa nas frequências de serviço, e, eventualmente, também do aumento da velocidade equivalente do automóvel particular. Atendendo a estes pressuposto, os resultados apresentam-se efetivamente coerentes, mas a variável velocidade comercial equivalente aparentemente perde relevância para justificar escolhas modais perante outros fatores externos mais preponderantes. Atendendo aos resultados, uma eventual utilização destas variáveis carece de normalização, que se assumiu que seja feita pela gama de velocidades mais elevadas (V23).

6. Amostras Disponíveis

As dificuldades verificadas na obtenção de dados para todas as variáveis consideradas limitaram as amostras disponíveis para a modelação das escolhas modais. Dos resultados apresentados anteriormente constatou-se que as variáveis que maiores limitações impõem à amostra passível de ser utilizada são:

- Tempo total de viagem e variáveis que o consideram na sua formulação como a velocidade comercial dos Bus ou a razão de tempos de viagem BUS/Auto;
- Duração da estadia;
- Variáveis relacionadas com o estacionamento;
- Frequência média horária diária do BUS.

Quadro IV. 51 - Amostras disponíveis atendendo às diferentes combinações de variáveis

Socioeconómicas									Caracterizadoras da viagem e da organização das actividades diárias								Políticas de mobilidade						Amostra
Variável	NL	R	RL	Sx	Id	In	Mot	Lc	Mtrp	Mt	I1a	I1i	D2	nV	De	Estacionamentos			IC				
																PO	OV	P1j	V2	TBA	Ft11		
6 modos	x	x	x	x	x	x	x	x	-	x			x	x						x			95426
									-														89305
									-						x								61358
									-													x	41329
									-											x	x		40099
									-			x											30559
									-							x						x	26576
									-			x				x							19555
									-									x					18823
									-										x		x		18773
									-										x		x	x	17005
									-							x			x				12214
									-			x								x	x		10264
									-			x							x				962
BWA									-											x			78470
									-													x	38310
									-											x	x		37226
									-										x				17913
									-										x		x		17863
									-										x		x	x	16197
									-			x							x				13399
									-			x							x		x		12793

No quadro IV.51 é possível verificar as amostras disponíveis atendendo às possíveis combinações de variáveis. A redução da amostra devido ao tempo total de viagem deve-se

fundamentalmente à inexistência de dados para todos os modos entre todos os pares GA. A redução da amostra devido às variáveis sobre políticas de estacionamento e do serviço de TC deve-se à dificuldade de obtenção de dados de todos os pares GA. A redução da amostra devido à duração da estadia deve-se à falta de dados para as viagens diárias únicas ou para a última viagem do dia.

A redução da amostra devido a estes fatores levou à ponderação da redução dos modos de transporte a considerar, nomeadamente para as análises envolvendo o tempo total de viagem, de seis para apenas os três mais relevantes - autocarro, pedonal e automóvel (base de dados BWA). A constatação de que para a maioria das combinações as amostras não apresentavam grandes diferenças sugere que efetivamente as restantes alternativas não dispõem de amostras significativas. Verificou-se ainda que nas combinações das variáveis sobre o tempo de viagem e políticas de estacionamento a amostra para apenas 3 modos de transporte aumentava relativamente à situação com 6 modos.

Estes resultados sugerem que para modelos mais complexos com a combinação de variáveis sobre o tempo de viagem, políticas de estacionamento e sobre os serviços de TC se poderá ter que reduzir as alternativas modais avaliadas apenas às três mais relevantes, de modo a obter amostras com tamanho e informação suficiente para a modelação de escolhas modais.

7. Síntese

Atendendo aos elevados custos temporais, financeiros e de meios humanos para realizar um levantamento de dados de base exaustivo para a avaliação da influência de diferentes fatores nas opções modais diárias dos cidadãos, procurou-se utilizar dados pré-existentes levantados por instituições idóneas e credíveis. Assim, e atendendo à quantidade de informação recolhida, optou-se por utilizar os dados obtidos pelo Inquérito à Mobilidade na AMP em 2000, realizado pelo INE. Esta base de dados recolheu informação sobre as viagens diárias em dias de semana (apesar de não terem sido utilizadas, também estavam disponíveis informações sobre viagens ao fim de semana), sobre cada indivíduo que as realizou e sobre as características do seu agregado familiar. A base de dados incluía ainda a avaliação da qualidade do sistema de transportes por cada um dos inquiridos. A esta informação juntou-se a informação recolhida pela FEUP para a AMTP sobre a oferta de TC, nomeadamente autocarros, e sobre a oferta de estacionamento no concelho do Porto.

A utilização de levantamentos pré-existentes, não especificados para determinado objetivo mas para outros mais abrangentes, leva à existência de algumas inconsistências que por vezes limitam o aproveitamento pleno de todo o potencial de informação disponível. Atendendo ao objetivo do levantamento do INE, a análise da mobilidade na AMP, houve alguns aspetos metodológicos na definição das variáveis recolhidas que deveriam ser revistos em trabalhos

futuros, já que não implicam custos financeiros significativos e permitiriam potenciar significativamente a utilidade da base de dados resultante:

- A impossibilidade de definir em concreto o local de origem e de destino de viagens intrazonais (dentro da mesma zona de tráfego) quando as zonas são muito extensas deverá levar a uma subdivisão estatística menor ou a uma codificação mais inequívoca dos pontos de origem e destino. Se dentro do concelho do Porto essa subdivisão foi realizada, os critérios para a sua definição deveriam ser conhecidos e atender a aspetos relacionados com a mobilidade, nomeadamente distâncias pedonais, homogeneidade de usos do solo, tipologias e quantidade de oferta de estacionamento, entre outros. Em relação às zonas fora do concelho do Porto, a divisão administrativa de Freguesia utilizada, em determinadas zonas mais rurais, leva a adoção de zonas muito extensas que em termos de mobilidade podem ter realidades muito distintas, sugerindo-se a distinção dos diferentes aglomerados com população relevante;
- Sugere-se, assim, que na definição das zonas de GA das viagens o zonamento consiga diferenciar zonas não muito extensas, bem caracterizadas em termos de homogeneidade de uso do solo, características dos sistemas de transportes ou características socioeconómicas da população residente. No caso de áreas muito extensas ou menos homogéneas em termos de uso do solo ou de aspetos relacionados com a mobilidade, sugere-se a codificação inequívoca dos locais de origem e de destino das viagens;
- Em relação ao modo de transporte utilizado sugere-se que os inquéritos avaliem também quais as alternativas disponíveis e quais as suas características operacionais apreendidas pelo inquirido. A especificação da carreira de Bus/Train/Tram utilizada, ou possível de utilizar, apresenta-se como um dado que poderá facilitar a apreensão das características operacionais das viagens cuja caracterização seja mais dúbia;
- Nos motivos da viagem as características operacionais de viagens para compras ou obtenção de serviços financeiros ou institucionais justificam a sua clara diferenciação;
- Na definição da tipologia de variáveis a utilizar, para uma análise mais coerente, sugere-se, sempre que possível e socialmente exequível, a escolha de uma especificação contínua para o rendimento mensal do agregado, uma vez que as gamas de valores utilizadas foram escassas, nomeadamente para os rendimentos superiores. A especificação contínua elimina igualmente a linearidade imposta às categorias que podem não traduzir diferenças comportamentais reais entre os inquiridos pertencentes a cada um dos níveis. Em alternativa poderá ser inquirido o salário horário (ou dados que o possibilitem determinar), permitindo simultaneamente a avaliação do valor do tempo. Na mesma lógica da tipologia a utilizar, a definição das gamas dos escalões etários

deveria refletir divisões mais ajustadas às diferentes características de mobilidade ou autonomia perante as escolhas modais, nomeadamente nos escalões mais jovens;

- É também importante que no modo de transporte utilizado se consiga distinguir entre condutores e passageiros de automóveis particulares, cujas características intrínsecas e impacto sobre o sistema de transportes são bastante distintos. Apesar de esta distinção existir nos inquéritos originais, a base de dados analisada não apresentava essa diferenciação;

- Na aquisição de informação sobre o custo do estacionamento e da viagem em TC, sugere-se a inquirição desses custos (ou estimativa) efetivamente verificados para cada uma das viagens/atividades em análise, e não um valor total mensal/diário. Esta distinção poderá revelar as gamas de valores a considerar para alcançar determinadas metas de utilização destes sistemas;

Além das dificuldades encontradas no tratamento e utilização dos dados obtidos no Inquérito à Mobilidade, salientam-se as dificuldades para a obtenção de dados sobre as redes de TC, nomeadamente interurbanas e privadas, bem com na caracterização da oferta de estacionamento fora do concelho do Porto. Os problemas mais relevantes relacionam-se com:

- Falta de organização e abertura de muitos operadores privados a explorarem carreiras suburbanas ou em meios urbanos mais pequenos, recusando a divulgação de dados como horários, trajetos e paragens, e sem a sua definição em documentos físicos ou digitais;
- O problema referido anteriormente reflete as falhas existentes à data da recolha de dados na entidade reguladora da área dos transportes de passageiros, o Instituto da Mobilidade e dos Transportes Terrestres (IMTT), cuja monitorização da qualidade e do cumprimento das adjudicações efetuadas se revelou muito desatualizado. Uma entidade que regula um sector cujos serviços são dispendiosos e essenciais a parte da população, deveria ter informação atualizada sobre operadores, horários e trajetos, em suporte digital e disponíveis ao público;
- A lacuna existente na maioria dos municípios de pequena e média dimensão, não só sobre a localização, quantidade e modos de exploração dos locais de estacionamento na sua área de influência, assim como de num registo da sua evolução temporal, dificulta a avaliação das operações de planeamento de transporte no passado e de servir como base para futuras decisões;

Apesar das dificuldades enunciadas foi possível definir um conjunto de variáveis com potencial para poderem introduzir informação relevante para a modelação das escolhas modais. Parte das variáveis consideradas foram baseadas em estudos de escolhas modais

aplicados aos transportes apresentados na bibliografia internacional, sendo outras comuns a modelos genéricos de escolha discreta. A limitação da informação completa para todos os modos ou todos os pares GA, levou à redução da amostra disponível para os modelos mais complexos com combinação de um número razoável de variáveis. Esta restrição reduziu um pouco a riqueza/variedade de informação disponível, limitando um pouco o desenvolvimento de modelos mais complexos, com maior número e tipologia de variáveis.

Capítulo V

Modelações Preliminares

1. Introdução

É internacionalmente aceite que os modelos Logit Multinomiais (MNL) apresentam na sua fundamentação teórica limitações que tornam a sua aplicação em problemas mais complexos desapropriada. No entanto, vários autores (Hensher *et al.* (2005), Train (2003)), referem os possíveis ganhos de eficiência resultantes de um estudo aprofundado sobre que variáveis introduzir nos modelos, bem como quais as tipologias a utilizar, antes de, eventualmente, passar a modelações mais complexas e específicas. A facilidade de cálculo e de interpretação de resultados dos modelos MNL levam à sua recomendação para estudos exploratórios aprofundados, pois permitem combinar análises mais céleres sobre o potencial explicativo das diferentes variáveis a utilizar, com a obtenção de resultados que servem para base de comparação com as análises mais complexas.

No presente capítulo, faz-se a apresentação de um conjunto de modelações preliminares, com cada uma das variáveis selecionadas para a modelação das escolhas modais, isolada das restantes, de modo a avaliar a sua importância relativa, coerência de sinal, especificação com maior potencial de adicionar informação relevante, e a seleção da melhor variável em situações com múltiplas hipóteses disponíveis. Assim, neste capítulo são comparados os resultados de modelos MNL apenas com cada uma das variáveis selecionadas, nas suas diferentes especificações, contínuas e discretas; é avaliada a eventual não linearidade da especificação contínua através do uso de variáveis binárias; e para as variáveis com resultados menos coerentes ou nas que se suspeitou de serem passíveis de correlação, são testadas eventuais relações de simbiose/complementaridade ou de correlação através do coeficiente de Pearson e de modelos com combinação das mesmas.

Apresenta-se ainda uma análise descritiva da estabilidade modal de cada inquirido ao longo das suas viagens diárias. Nesta análise, além de avaliar a relação entre o modo de transportes de cada viagem ao longo do dia atendendo ao modo utilizado na primeira viagem do dia, avaliou-se a constância de uso deste último ao longo do dia.

O capítulo finaliza com uma síntese das variáveis reveladas como mais importantes, bem como das suas melhores especificações, para utilização nos modelos exploratórios com combinação de variáveis.

2. Enquadramento Base

A avaliação do modelo que melhor se adequa para determinada situação não se apresenta como uma tarefa direta e simples. Koppelman e Bhat (2006) referem que, para tomadas de decisão devidamente fundamentadas, o processo deve combinar:

- O uso da intuição, ponderada eventualmente por referência face a problemas similares tratados na bibliografia de referência;
- Uma análise qualitativa;
- E testes estatísticos.

Para avaliar a adequabilidade de diferentes especificações e modelos sugerem tanto testes estatísticos como avaliações informais de sinais, magnitudes, ou magnitudes relativas entre parâmetros, baseadas no conhecimento das relações comportamentais subjacentes que influenciam as escolhas modais. Vários modelos de diferentes especificações podem, na prática, ajustar-se igualmente aos dados existentes mas implicar relações e previsões muito diferentes. O uso da intuição e da experiência na tomada de decisões relacionadas com a modelação são essenciais, uma vez que é quase impossível determinar a “melhor” especificação para um modelo apenas com testes estatísticos.

Uma das abordagens mais comuns para o processo de desenvolvimento do modelo é iniciá-lo com uma especificação mínima, incluindo as variáveis consideradas essenciais a qualquer modelo razoável. A partir desta especificação mínima, devem incrementar-se alterações e testar os seus resultados num esforço de melhorar o seu realismo comportamental e o seu ajustamento empírico aos dados. Estes incrementos devem, ainda assim, evitar tornar o modelo demasiado complexo, mas permitir que a sua estimação clarifique e elucide sobre a adequação das mesmas na melhoria do modelo final.

Na análise dos resultados estatísticos, nomeadamente dos testes de significância estatística dos parâmetros, há que atender que um valor elevado indiciando que determinado parâmetro não pode ser considerado significativamente diferente de zero, e portanto sem impacto, não implica a remoção da variável do modelo. Koppelman e Bhat (2006) referem que se o analista tiver fortes motivos para considerar que uma variável é importante, e se o seu parâmetro tiver um sinal correto, é razoável a sua manutenção no modelo. No caso de uma especificação com codificação através de várias variáveis binárias, Hensher *et al.* (2005) referem que a não significância estatística de apenas uma ou algumas destas variáveis não justifica em teoria a exclusão do conjunto da especificação. No caso de modelos mais complexos em que a parcimónia é essencial, se se verificar que a eliminação da variável não significativa não apresenta grande impacto na constante específica da utilidade e no ajustamento global do modelo, os mesmos autores sugerem a sua eliminação.

2.1. Tipologias de Especificações Mais Comuns

As variáveis a introduzir nos modelos podem ser especificadas de diferentes formas atendendo às suas particularidades e ao tipo de relação que as mesmas apresentam com a variável dependente. Com exceção da tipologia de combinação de binárias, os modelos MNL necessitam de determinar um único parâmetro que representa a influência média que os diferentes valores da variável apresentam na variável dependente.

No caso de a variável independente apresentar valores numéricos cuja grandeza apresenta conjuntos de valores dispersos entre si, a tipologia discreta apresenta-se como a forma mais correta de introduzir informação no modelo. No entanto, estas tipologias podem apresentar problemas de representatividade quando a relação variável dependente versus variável independente não é linear (crescente ou decrescente), para cada modo de transporte. No caso da utilização de variáveis discretas pode suceder que apesar de apresentarem um efeito linear parcial, o mesmo tenha “declives” diferentes entre diferentes categorias. No caso de se verificar que a relação não apresenta características lineares ou exiba diferenças significativas de declividade entre categorias, a opção deverá ser a utilização de variáveis binárias combinadas para os diferentes escalões. A seleção dos diferentes escalões deverá, no entanto, ser alvo de uma análise cuidada uma vez que as mesmas devem representar um número restrito de conjuntos com uma lógica teórica fundamentada. A definição de um grande número de combinações pode dificultar a modelação, quer devido ao aumento da complexidade do modelo, quer pela especificação excessiva das categorias com aplicação apenas a um determinado problema.

Hensher *et al.* (2005) referem que na avaliação da melhor tipologia a utilizar se deve atender à semelhança de resultados obtidos entre as especificações lineares e não lineares, mas também à melhoria de ajustamento global do modelo e à correção dos sinais dos parâmetros. A melhoria do ajustamento global do modelo através de especificação não linear pode não justificar a sua utilização se o impacto não for bastante significativo. Esta melhoria do ajustamento global pode ser aferida através do teste estatístico da razão das MVL (-2LL) entre modelos, desde que tenham amostra e número de alternativas disponíveis iguais. Os mesmos autores sugerem um teste de *Wald* com restrições lineares para avaliar a hipótese de a diferença entre declives de dois parâmetros ser significativamente diferente de zero. O teste, no entanto, envolve restrições lineares entre cada par de parâmetros a avaliar. No caso de muitas variáveis a analisar, para vários modos de transporte, este teste torna-se demasiado moroso, tendo-se optado no presente conjunto de análises preliminares por uma abordagem empírica qualitativa, fazendo o teste apenas às variáveis binárias que apresentassem tendências claramente lineares.

Se algumas das variáveis apresentam características teóricas que as enquadram inequivocamente numa única tipologia (como o sexo ou a disponibilidade de licença de

condução, inequivocamente binárias), outras há cuja especificação pode levantar mais dúvidas. Nos pontos seguintes apresentar-se-ão as análises empíricas realizadas às variáveis cuja especificação não se considerou evidente. Os resultados apresentados resumem-se aos das alternativas mais relevantes, podendo os resultados das restantes ser consultados nos ficheiros com os *outputs* do programa *NLogit*® anexados em formato digital ao presente documento.

2.2. Especificação Discreta Alternativa

Na definição dos valores da cada categoria nas especificações discretas, apesar de subjacente a uma lógica de crescimento ou decrescimento consoante a variável a representar, a grandeza relativa entre os seus valores não costuma ser alvo de grande ponderação. A forma tradicional de implementar estas categorias traduz-se na definição/imposição de valores inteiros geralmente com comportamento linear entre si, como por exemplo:

- 1 - Rendimentos 0-500€;
- 2 - Rendimentos 501-1000€;
- 3 - Rendimentos 1001-1500€;
- 4 - Rendimentos 1501-2000€;
- 5 - Rendimentos >2000€.

Esta definição, no entanto, impõe uma linearidade entre valores que pode não traduzir a real relação entre categorias podendo levar a resultados pouco realistas. Por exemplo, é expectável que o impacto nas escolhas modais de acréscimos em rendimento mais baixos seja maior que para rendimentos já de si grandes.

Para avaliar a possibilidade de uma metodologia alternativa analisou-se a hipótese de substituir os valores que definem cada categoria, geralmente inteiros e lineares entre si, por valores que traduzissem a não linearidade verificada nos resultados. Esta metodologia foi aplicada às variáveis cujo valor do parâmetro para cada uma das categorias não apresentasse grandes discrepâncias entre os diferentes modos.

Os valores utilizados nesta especificação linear discreta reajustada foram as médias aritméticas dos valores estimados para os parâmetros de cada categoria para os diferentes modos. Estes valores médios são resultantes de modelações MNL recorrendo a apenas uma categoria isoladamente. As variáveis que apresentaram maior estabilidade de parâmetros foram o número de automóveis ligeiros disponíveis diariamente no agregado *per capita* e o rendimento líquido mensal do agregado. Os resultados dos modelos utilizando estas variáveis discretas reajustadas (Rj e NCj) para as alternativas mais relevantes comparativamente aos

modelos com as mesmas na sua especificação discreta linear são apresentadas no quadro V.1 e 2.

Quadro V. 1 - Resultados dos modelos preliminares para o reajustamento dos valores discretos lineares pelos valores médios dos parâmetros de cada categoria, para o rendimento mensal do agregado

Modelo	MNL1.79		MNL1.7a	
Modo	B	W	B	W
Rb	-0,4556	-0,4964		
Rj			0,8210	0,9087
Amostra	95426		95426	
p-R ²	0,0285		0,0298	
MVL (LL)	-119964,7		-119803,8	
* - p-valor >0.05 & <=0.1 ** - p-valor >0.1 & <=0.25 *** - p-valor >0.25				

Quadro V. 2 - Resultados dos modelos preliminares para o reajustamento dos valores discretos lineares pelos valores médios dos parâmetros de cada categoria, para o número de veículos disponíveis diariamente *per capita* do agregado

Modelo	MNL1.26		MNL1.13a	
Modo	B	W	B	W
NC	-5,0011	-5,1170		
NCj			0,7559	0,7858
Amostra	95426		95426	
p-R ²	0,1003		0,1023	
MVL (LL)	-111107,3		-110854,97	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

Os resultados destas análises, no entanto, não revelaram ganhos significativos nos ajustamentos globais dos modelos reajustados (nomeadamente comparados com a utilização da especificação não linear) e a sua determinação apresenta-se exageradamente específica para cada situação, inviabilizando a sua utilização de forma generalizada. A inversão do sinal dos parâmetros deve-se ao comportamento das novas categorias que deixaram de ser crescentes para ser decrescentes, mas que facilmente poderiam ser corrigidas, entrando com sinal inverso nos modelos.

3. Análise Isolada de Variáveis Socioeconómicas

3.1. Rendimento Líquido Mensal do Agregado

3.1.1. Análise Comportamental de Diferentes Especificações

A variável *rendimento líquido mensal do agregado familiar* apresentava na base de dados original uma tipologia de variável discreta ordinal. A análise realizada visou avaliar a possibilidade de a utilizar como variável discreta ou como variável binária. A variável original estava dividida em 5 categorias originais com valores de 1 a 5 para cada gama de rendimentos (Rb). Uma especificação discreta alternativa foi considerada, utilizando as 5 categorias com valores máximos do rendimento para cada categoria (R). Na análise como combinação de

variáveis binárias, foram consideradas 5 variáveis distintas para cada categoria (Ri). Para a análise da possível influência do rendimento líquido mensal do agregado na escolha modal de um indivíduo realizaram-se modelos MNL utilizando isoladamente cada uma das especificações referidas. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.3.

Quadro V. 3 - Resultados dos modelos preliminares da influência do rendimento líquido mensal do agregado

Modelo	MNL1.28		MNL1.79		MNL3.11	
Modo	B		B	W	B	W
R	-0,0009	-0,001				
Rb			-0,456	-0,496		
R1					2,191	2,306
R2					1,198	1,078
R3					0,836	0,648
R4					0,505	0,327
R5					-	-
Amostra	95426		95426		95426	
p-R ²	0,029		0,029		0,032	
MVL (LL)	-119964,7		-119964,7		-119596,3	
* - p-valor >0.05 & <=0,1 ** - p-valor >0,1 & <=0.25 *** - p-valor >0.25						

Todos os parâmetros têm o sinal esperado. Nos modelos com especificação linear discreta o sinal negativo traduz o comportamento expectável da perda de utilidade das alternativas ao automóvel particular com o aumento do nível de rendimento. Na especificação com combinação de binárias normalizou-se o nível de rendimento mais elevado normalmente associado a uma maior utilização do automóvel. Esta normalização torna o sinal positivo e valor decrescente dos parâmetros com o aumento do nível de rendimento coerente, uma vez que comparativamente ao nível de rendimento mais elevado, as alternativas são tanto mais atrativas quanto menor for o rendimento.

Verifica-se que as especificações discretas lineares, apesar de apresentarem grandezas diferentes, têm o mesmo ajustamento global, evidenciando que estas formas de definir as categorias nas especificações discretas são equivalentes na informação introduzida. Dos resultados é ainda possível constatar que a especificação com combinação de binárias apresenta um ajustamento global do modelo um pouco melhor ($p\text{-}R^2 = 0,032$ relativamente a 0,029 das especificações discretas lineares). Para comparação estatística dos resultados entre modelos assumiu-se que o modelo de base seria a especificação com menos restrições, neste caso as especificações discretas (MNL1.79 ou 1.28). Utilizando o teste estatístico da razão das MVL (-2LL), com graus de liberdade iguais à diferença de parâmetros estimados entre os modelos a comparar, obtemos:

$$\begin{aligned}
& -2 \times (LL_{\text{modelo base(MNL1.79)}} - LL_{\text{modelo estimado(MNL3.11)}}) \sim \chi^2_{(GL)} \Rightarrow \\
& \Rightarrow -2 \times (-119964,7 - (-119596,3)) \sim \chi^2_{(gl=25-10)} \Rightarrow \\
& \Rightarrow 736,8 \sim \chi^2_{15} > \chi^2_{15}(\alpha = 0,05) = 25,0
\end{aligned}$$

Exp. V - 6

O valor crítico de Chi-quadrado para 15 graus de liberdade e um intervalo de confiança de 95% ($\alpha=0,05$), é de 25,0, cerca de 29 vezes inferior ao teste -2LL, pelo que se pode recusar a hipótese nula de que o “novo” modelo (com especificação de combinação de binárias) não melhora estatisticamente a MVL em relação ao modelo de comparação (com especificação discreta linear). O mesmo resultado pode ser verificado com o modelo MNL1.28, cujo número de variáveis a determinar e o valor de MVL são exatamente iguais aos do MNL1.79. Assim, pode-se concluir que o modelo com especificação como combinação de binárias é estatisticamente mais próximo de zero do que os modelos com especificação discreta, e portanto melhor.

3.1.2. Avaliação de Comportamentos Não-lineares

Apesar de a especificação com combinação de binárias apresentar resultados globais ligeiramente melhores que as especificações discretas lineares, a facilidade de definição das segundas poderá justificar a sua utilização se a variável em estudo apresentar características lineares. Neste sentido realizaram-se análises de comportamento das utilidades parciais de cada modo para as diferentes especificações. Os resultados destas análises de comportamento, obtidas da sua aplicação a modelos MNL, são apresentados nas figuras V.1, V.2 e V.3.

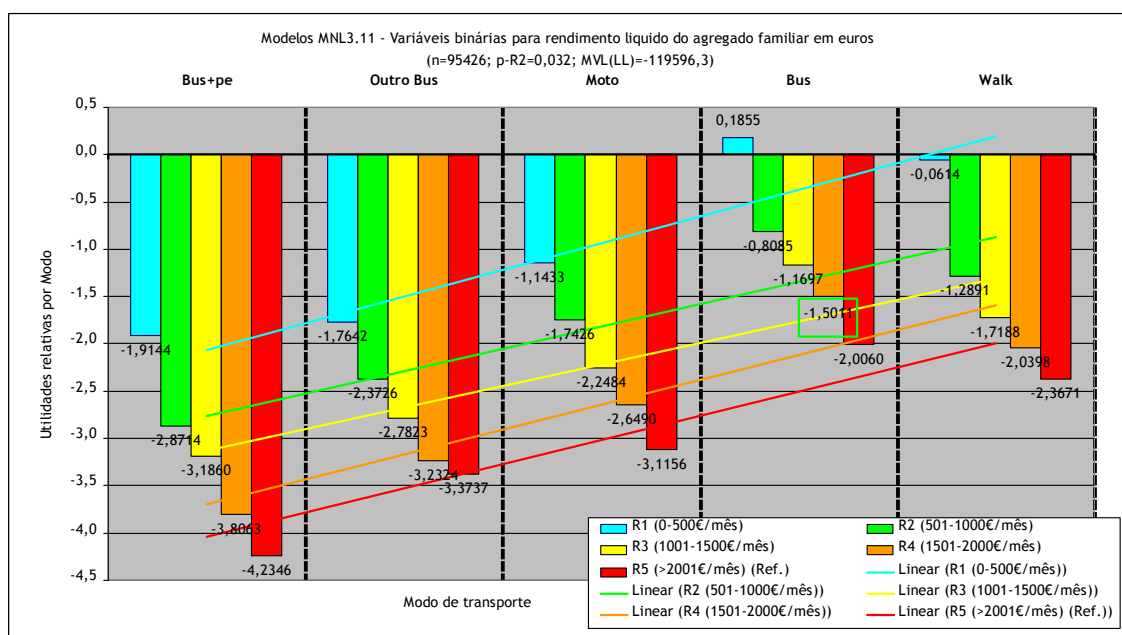


Figura V. 1 - Utilidades parciais para os diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável binária para cada categoria

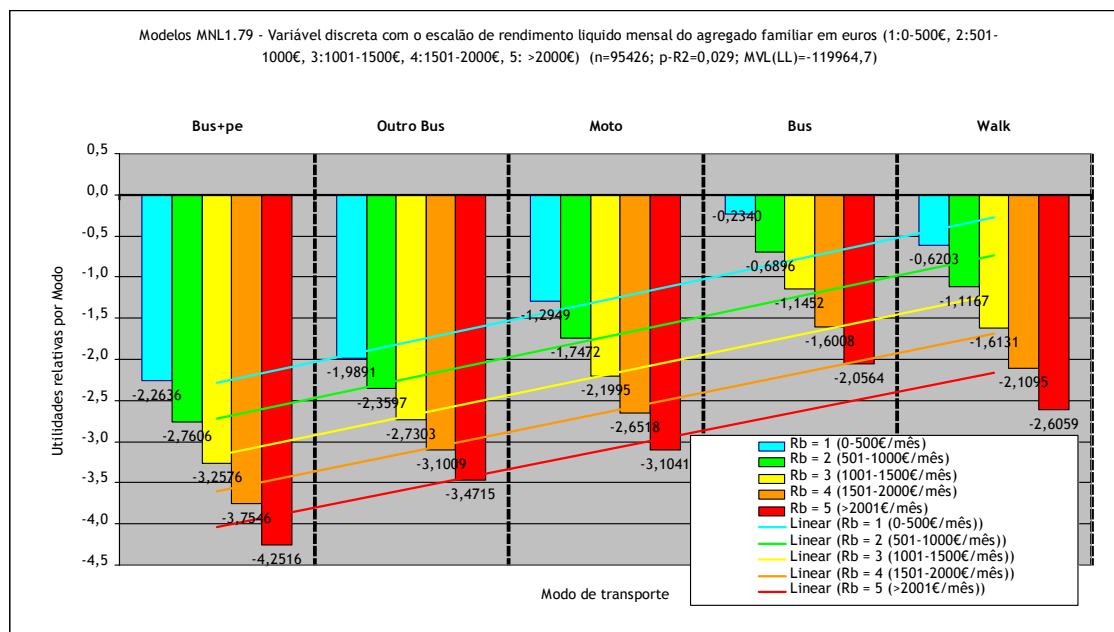


Figura V. 2 - Utilidades parciais dos diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável discreta por categorias

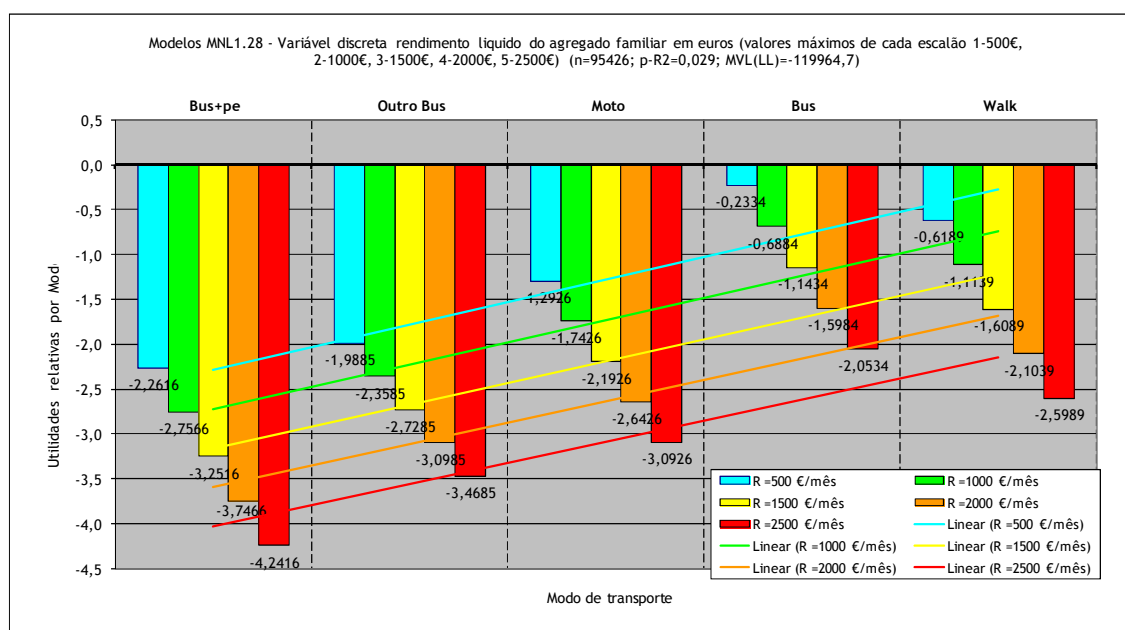


Figura V. 3 - Utilidades parciais dos diferentes modos de transporte considerando o rendimento líquido mensal do agregado familiar como variável discreta com o valor máximo de cada categoria

Da análise qualitativa das utilidades relativas, a especificação com combinação de binárias não aparenta ter um comportamento linear para os modos BUS e Walk nem para o nível de rendimentos mais baixos.

Para avaliar a linearidade dos parâmetros, Hensher *et al.* (2005) recomendam o teste de *Wald* com restrições lineares. Considerando a figura V.4, se os parâmetros para diferentes variáveis binárias da mesma variável discreta apresentarem uma relação linear, a diferença entre parâmetros mantém-se (expressão V-2).

$$b(2) - b(1) = b(3) - b(2) = b(4) - b(3) \quad \text{Exp. V - 7}$$

Para que esta restrição se verifique há a necessidade de a decompor em duas equações:

$$\begin{cases} b(2) - b(1) = b(3) - b(2) \Rightarrow b(2) - b(1) - b(3) + b(2) = 0 \\ b(3) - b(2) = b(4) - b(3) \Rightarrow b(3) - b(2) - b(4) + b(3) = 0 \end{cases} \Rightarrow \begin{cases} 2 \times b(2) - b(1) - b(3) = 0 \\ 2 \times b(3) - b(2) - b(4) = 0 \end{cases} \quad \text{Exp. V - 8}$$

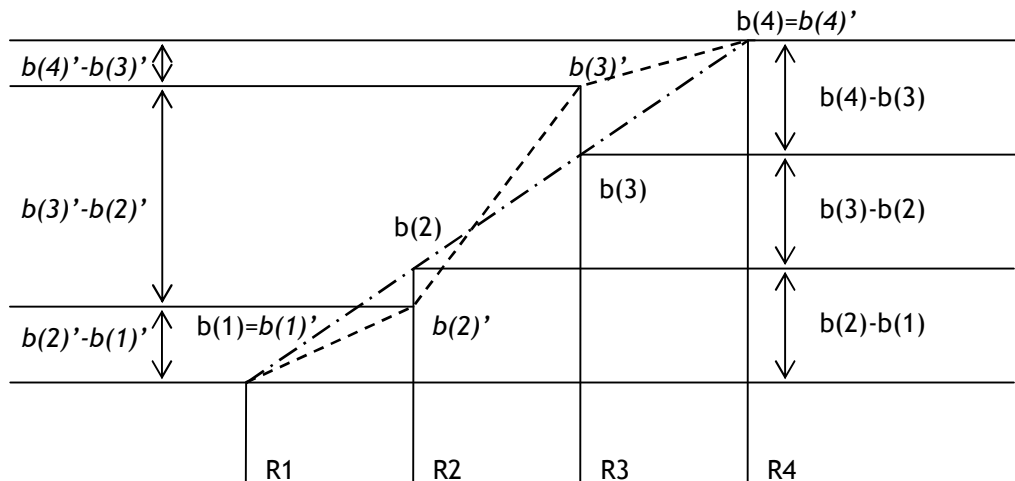


Figura V. 4 - Esquema representativo de relações lineares entre parâmetros de variáveis com combinação de discretas

Para testar esta hipótese realizou-se o teste de *Wald* ao modelo com especificação não linear para as diferentes alternativas (MNL3.11wald_BP, MNL3.11wald_OB, MNL3.11wald_M, MNL3.11wald_B e MNL3.11wald_W) com as duas restrições lineares (2 graus de liberdade) para que as diferenças entre parâmetros fossem iguais entre si (e portanto lineares).

No quadro V.4 é possível verificar que a análise qualitativa aparentemente é razoável, com uma clara rejeição da linearidade dos parâmetros para os modos Bus e Walk. Em relação aos restantes modos, apesar de estatisticamente a hipótese nula ser rejeitada, os valores do teste revelam-se muito próximos dos valores críticos sugerindo que a sua relação é aproximadamente linear.

Quadro V. 4 - Resultados do teste de *Wald* para o modelo MNL3.11 para avaliar a linearidade dos parâmetros da combinação de binárias para o rendimento mensal do agregado

Modo	Teste de <i>Wald</i> Chi-quadrado	Valor crítico para $\alpha = 0,05$ e 2 GL	Comentário
BW	45,71	5,99	Hipótese nula dos parâmetros serem lineares rejeitada
OB	6,70		Hipótese nula dos parâmetros serem lineares rejeitada, mas por muito pouco
M	8,21		Hipótese nula dos parâmetros serem lineares rejeitada, mas por muito pouco
B	313,15		Hipótese nula dos parâmetros serem lineares rejeitada
W	422,55		Hipótese nula dos parâmetros serem lineares rejeitada

Comparando o comportamento das variáveis para cada um dos modos de transporte, este apresenta-se coerente com uma diminuição da utilidade de todas as alternativas relativamente a um aumento dos rendimentos do agregado, comparativamente ao modo de referência - o automóvel particular.

Na análise com combinação de variáveis binárias verifica-se que a declividade para as diferentes categorias tem pequenas diferenças. A diferença mais acentuada está associada à categoria de rendimentos mais baixa e deve-se aos resultados dos modos BUS e Walk cujas utilidades são mais afetadas (favoravelmente) nesta categoria que para os restantes modos.

De uma forma geral, estatisticamente o modelo com combinação de binárias é melhor, mas com exceção dos modos BUS e Walk para rendimentos mais baixos, as especificações lineares apresentam resultados suficientemente aproximados para poderem ser consideradas como uma opção viável.

3.1.3. Rendimento Líquido Mensal *versus* Rendimento Líquido Mensal *per capita*

Atendendo à potencial influência que o rendimento apresenta na teoria comportamental, colocou-se a hipótese de avaliar se o rendimento disponível para cada elemento do agregado teria maior capacidade explicativa que o rendimento global. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.5.

Atendendo aos resultados anteriores, apenas se fez a análise comparativa com o rendimento global do agregado com a especificação de combinação de binárias, uma vez que a mesma apresentava resultados melhores que as especificações discretas lineares.

Dos resultados é possível verificar que o modelo com o rendimento líquido mensal do agregado apresenta ajustamentos ligeiramente melhores que o rendimento *per capita*. Em todos os modelos os parâmetros têm sinal e comportamento correto. Dos testes de estatística -2LL pode verificar-se que os modelos com mais restrições (combinação de binárias) são estatisticamente melhores que o modelo com especificação discreta linear (-2LL é entre 53 e 88 vezes superior ao valor crítico de qui-quadrado). É também possível verificar que entre o modelo com o rendimento global e o *per capita*, o modelo com a variável global melhora estatisticamente a MVL e está mais perto de zero. Assim, aparentemente a especificação com o rendimento líquido mensal do agregado tem maior capacidade explicativa.

Quadro V. 5 - Resultados dos modelos preliminares comparativos entre o rendimento líquido mensal do agregado e o rendimento líquido mensal do agregado *per capita*

Modelo	MNL1.27		MNL3.07		MNL3.11	
Modo	B	W	B	W	B	W
RC	-0,0023	-0,0018				
RC1			1,516	1,356		
RC2			0,828	0,670		
RC3a			-	-		
R						
R1					2,191	2,306
R2					1,198	1,078
R3					0,836	0,648
R4					0,505	0,327
R5					-	-
Amostra	95426		95426		95426	
p-R ²	0,023		0,025		0,032	
MVL (LL)	-120684,6		-120391,4		-119596,3	
Restrições	10		15		25	
GL			5		15	
-2LL	Base		586,4		2176,6	
			Base		1590,2	
Chi ² ($\alpha=0,05$)	Base		11,0		25,0	
			Base		18,3	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

3.2. Sexo do Inquirido

Para a análise da possível influência do sexo na escolha modal de um indivíduo utilizou-se um modelo MNL com especificação binária. Atendendo à natureza da mesma, não foi testada nenhuma outra especificação. Os resultados desse modelo para as alternativas são apresentados no quadro V.6.

Dos resultados é possível verificar que o modelo apresenta um ajustamento global reduzido mas que tem parâmetros com sinais coerentes, mostrando, como se esperava, uma valorização média positiva das alternativas ao automóvel por parte das mulheres, exceto do modo MOTO. É de salientar que esta valorização pode não significar que as mulheres têm preferência pelas alternativas ao automóvel em relação aos homens, mas que efetivamente as

utilizam mais, quer por não gostarem ou não poderem conduzir o automóvel, nomeadamente por tradicionalmente ser o homem quem conduz e leva o automóvel consigo, particularmente quando só existe um veículo no agregado familiar.

Quadro V. 6 - Resultados dos modelos preliminares da influência do Sexo do inquirido

Modelo	MNL1.29					
Modo	B+W	B	OB	M	P	A
Sx (0=M,1=F)	0,907	0,765	0,027***	-0,335	0,580	-
Amostra	95426					
p-R ²	0,012					
MVL (LL)	-121950,6					

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

3.3. Escalão Etário

3.3.1. Análise Comportamental de Diferentes Especificações

A variável *escalão etário* do indivíduo, apesar de ter características para uma especificação contínua, foi codificada como discreta ordinal na base de dados original. Na recodificação realizada para a sua utilização na modelação foram definidas cinco classes. Também esta variável não apresentava à partida características que indiciassem um comportamento linear, principalmente para alguns modos alternativos, cuja utilização costuma ser mais acentuada em escalões etários intermédios (por exemplo o BUS costuma ser mais utilizado pelos adolescentes do que por jovens adultos ativos). Para a análise da possível influência do escalão etário na escolha modal de um indivíduo realizaram-se modelos MNL utilizando as especificações discreta e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.7.

Nos modelos com especificação linear discreta o sinal negativo indicia uma perda de utilidade das alternativas ao automóvel particular com o aumento da idade.

Na especificação com combinação de binárias normalizou-se a idade dos 0 aos 9 anos, escalão que na base de dados estava associado a uma maior utilização do automóvel. Esta constatação poderá dever-se ao seu transporte por parte dos pais. No modelo com combinação de variáveis binárias não se verifica um comportamento linear crescente o que pode justificar o pior ajustamento do modelo linear face ao não linear. Atendendo à normalização realizada na especificação com combinação de binárias os resultados apresentam sinais coerentes com o senso comum, mostrando que comparativamente aos mais jovens as alternativas terão maior utilidade para os jovens adolescentes e para os reformados, que expectavelmente têm menos acesso ao automóvel, do que para a população ativa.

Quadro V. 7 - Resultados dos modelos preliminares da influência do escalão etário

Modelo	MNL1.65		MNL3.16	
Modo	B	W	B	W
Idb	-0,101	-0,039		
Id1			-	-
Id2			2,240	1,077
Id3			0,519	-0,504
Id4			0,782	-0,041***
Id5			1,879	1,172
Amostra	95426		95426	
p-R ²	0,006		0,046	
MVL (LL)	-122753,5		-117695,1	
	* - p-valor >0,05 & ≤0,1		** - p-valor >0,1 & ≤0,25	
			*** - p-valor >0,25	

É, ainda, possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento bem superior que o modelo com especificação discreta. O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.8.

Quadro V. 8 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.65	MNL3.16
Amostra	95426	95426
Id	10 Parâmetros	
Idi	25 Parâmetros	
LL _{base}	-122753,5	-122753,5
LL	-122753,5	-117695,1
-2LL	-	10116,8
Graus de liberdade	-	15
Chi ² (α=0,05)	-	25,0

O valor crítico de Chi-quadrado para 15 graus de liberdade e um intervalo de confiança de 95% ($\alpha=0,05$), é de 25,0, cerca de 404 vezes inferior ao teste estatístico -2LL, pelo que se pode recusar a hipótese nula de que o “novo” modelo (com especificação de combinação de binárias) não melhora estatisticamente a MVL em relação ao modelo base (com especificação discreta).

3.3.2. Avaliação de Comportamentos Não Lineares

Atendendo aos resultados anteriores, realizou-se uma análise das utilidades parciais com recurso a variáveis binárias para cada escalão para avaliar a linearidade dos parâmetros. Os resultados desta análise apresentam-se na figura V.5.

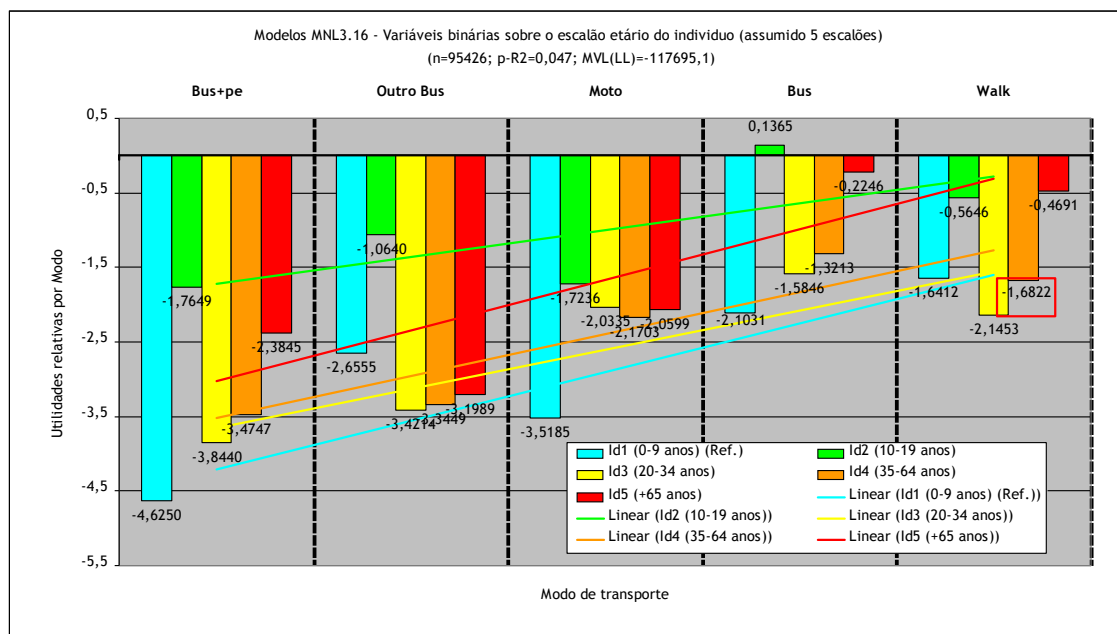


Figura V. 5 - Utilidades parciais para os diferentes modos de transporte considerando o escalão etário como variável binária para cada categoria

Da análise dos resultados das utilidades relativas foi possível constatar que o comportamento destas variáveis não só não era linear como era bastante diferenciado entre os modos alternativos. Esta diferenciação entre modos pode ser aferida pelos diferentes declives de cada escalão. Apesar destas diferenças, o comportamento dos diferentes escalões é coerente com o senso comum para os diferentes modos.

Pelos resultados parece claro que a única especificação possível será a não linear.

3.4. Nível de Instrução

3.4.1. Análise Comportamental de Diferentes Especificações

A variável *nível de instrução* apresentava na base de dados original uma tipologia de variável discreta nominal. Na sua especificação para a modelação foram definidas quatro categorias de instrução de ordem crescente. Apesar da sua definição como discreta, considerou-se que a sua influência nas escolhas modais não apresentava características para ser especificada com comportamento linear, definindo-se para o efeito uma variável binária para cada nível. Para a análise da possível influência do nível de instrução na escolha modal de um indivíduo, realizaram-se modelos MNL utilizando as especificações discreta e a combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.9.

Quadro V. 9 - Resultados dos modelos preliminares da influência do nível de instrução

Modelo	MNL1.81		MNL3.108	
Modo	B	W	B	W
ln	-0,390	-0,460		
ln1			0,907	0,994
ln2			1,377	1,113
ln3			1,069	0,594
ln4			-	-
Amostra	95426		95426	
p-R ²	0,013		0,021	
MVL (LL)	-121836,3		-120869,4	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

Nos modelos com especificação linear discreta o espectável sinal negativo indicia uma perda de utilidade das alternativas ao automóvel particularmente com o aumento do nível de instrução. Na especificação com combinação de binárias normalizou-se o nível de instrução mais elevado normalmente associado a uma maior utilização do automóvel, devido às prováveis melhores condições de vida. Esta normalização torna o sinal positivo coerente, uma vez que comparativamente ao nível de instrução mais elevado, as alternativas são mais atrativas quando o nível de instrução é menor. No entanto, não se verifica um comportamento linear crescente. Se seria expectável um comportamento decrescente dos parâmetros, revelando menor atratividade das alternativas com o aumento do nível de instrução, o nível analfabeto apresenta uma atratividade inferior às restantes categorias. Este resultado aparentemente contraria a teoria comportamental de uma maior probabilidade de uso das alternativas por não disporem das mesmas condições de acesso ao automóvel. Esta aparente contradição pode dever-se aos analfabetos não serem condutores mas passageiros de automóvel, nomeadamente crianças que apesar de não conduzirem pertencem a agregados que o utilizam, ou de, apesar de analfabetos, poderem ter condições económicas para ter acesso ao automóvel.

Quadro V. 10 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.81	MNL3.108
Amostra	95426	95426
In	10 Parâmetros	
Ini	20 Parâmetros	
LL _{base}	-121836,3	-121836,3
LL	-121836,3	-120869,4
-2LL	-	1933,8
Graus de liberdade	-	10
Chi ² (α=0,05)	-	18,3

Dos resultados é ainda possível verificar que o modelo com combinação de variáveis binárias apresenta um melhor ajustamento que o modelo com especificação discreta. O teste estatístico a esse ajustamento através da razão das MVL (-2LL) é apresentado no quadro V.10.

O valor crítico de Chi-quadrado para 10 graus de liberdade e um intervalo de confiança de 95% ($\alpha=0,05$), é de 18,3, cerca de 107 vezes inferior ao teste estatístico -2LL, pelo que se pode recusar a hipótese nula de que o “novo” modelo (com especificação de combinação de binárias) não melhora estatisticamente a MVL em relação ao modelo base (com especificação discreta).

Estes resultados sugerem que a especificação com combinação de variáveis binárias é melhor que a especificação discreta mas justificam outras análises com combinação com outras variáveis de forma a tentar perceber este comportamento menos expectável do nível de instrução analfabeto devido provavelmente a correlações.

3.4.2. Avaliação de Comportamentos Não Lineares

A assunção de que esta variável não apresentava características para uma especificação discreta linear foi verificada através de uma análise das utilidades parciais para cada modo de transporte, obtidas num modelo MNL com especificação de combinação de binárias. Os resultados do mesmo são apresentados na figura V.6.

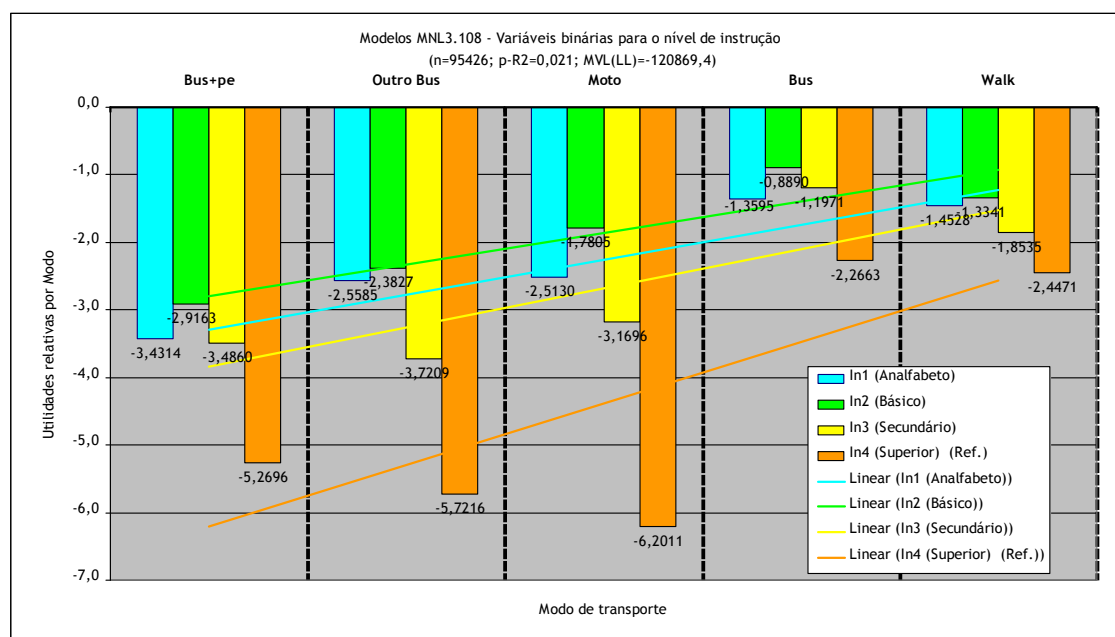


Figura V. 6 - Utilidades parciais para os diferentes modos de transporte considerando o nível de instrução como variável binária para cada categoria

É possível constatar que, apesar de se verificarem tendências similares para cada modo de transporte, as utilidades não seguem um comportamento linear nem decrescente devido aos

resultados do nível analfabeto. A declividade entre modos é relativamente semelhante, mas o impacto do nível de instrução mais elevado é significativamente maior do que para as outras categorias. É possível verificar a aparente contradição de as utilidades das alternativas serem menores para os analfabetos do que para a categoria com ensino Básico. Estas características, aliadas ao resultado do teste estatístico da razão das MVL, sugerem claramente que a especificação com combinação de binárias será a mais apropriada.

3.5. Número de Automóveis Disponíveis no Agregado *per capita*

3.5.1. Avaliação Comportamental de Diferentes Especificações

Outra variável socioeconómica cuja especificação necessita de avaliação é o *número total de automóveis ligeiros diariamente à disposição do agregado per capita, ou taxa de motorização familiar*. A sua tipologia original é contínua mas realizou-se um estudo com a sua desagregação em cinco variáveis binárias para avaliar as características dessa continuidade/linearidade. Para a análise da possível influência da taxa de motorização familiar realizaram-se modelos MNL utilizando as especificações discreta e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.11.

Quadro V. 11 - Resultados dos modelos preliminares da influência da taxa de motorização familiar

Modelo	MNL1.26		MNL3.29	
Modo	B	W	B	W
NC	-5,001	-5,117		
NC0			4,251	3,282
NC1			2,538	1,394
NC2			1,856	0,772
NC3			1,013	-0,177*
NC4a			-	-
Amostra	95426		95426	
p-R ²	0,100		0,104	
MVL (LL)	-111107,3		-110636,3	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

Dos resultados é possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento ligeiramente superior que o modelo com especificação contínua. Ambos os modelos apresentam ajustamentos globais elevados relativamente às variáveis anteriores. Nos modelos com especificação linear discreta o sinal negativo indicia uma perda de utilidade das alternativas ao automóvel particular com a taxa de motorização familiar. Este resultado era expectável atendendo a que um maior número de veículos disponível geralmente está associado a uma maior utilização dos mesmos.

Na especificação com combinação de binárias normalizou-se a categoria de mais de 0,75 automóveis por residente no agregado (ou seja, 1,3 indivíduos por automóvel). Verifica-se que

os parâmetros têm sinal positivo e um comportamento decrescente que, atendendo à normalização realizada, são coerentes. A utilidade das alternativas, face ao automóvel e à taxa de motorização familiar mais elevada, vai diminuindo à medida que a taxa vai aumentando, uma vez que a probabilidade de poder utilizar o automóvel também aumenta. É de salientar o impacto significativo da categoria de agregado sem automóvel ($NC0 = 0$ veíc./resid.) comparativamente à categoria seguinte, cujo parâmetro é aproximadamente o dobro, sugerindo uma influência não linear desta variável. Este resultado revela ainda que a não existência de automóvel disponível no agregado tem um impacto significativo na utilização das alternativas e que, portanto, qualquer política de mobilidade sustentável tenderá a beneficiar da existência de instrumentos que potenciem a existência de agregados familiares onde haja uma opção pela dispensa total de carro próprio.

O teste estatístico aos ajustamentos através da razão das MVL (-2LL) é apresentado no quadro V.12.

Quadro V. 12 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.26	MNL3.29
Amostra	95426	95426
NC	10 Parâmetros	
NCi	25 Parâmetros	
LL _{base}	-111107,3	-111107,3
LL	-111107,3	-110636,3
-2LL		942
Graus de liberdade		15
Chi ² ($\alpha=0,05$)		25,0

O valor crítico de Chi-quadrado é cerca de 37 vezes inferior ao teste estatístico -2LL, pelo que o modelo com especificação como combinação de binárias é estatisticamente mais próximo de zero que o modelo com especificação contínua, e portanto melhor.

3.5.2. Avaliação de Comportamentos Não Lineares

Atendendo aos resultados anteriores analisou-se a linearidade dos parâmetros estimados, através da comparação das utilidades relativas de cada alternativa. Nas figuras V-7 e V-8 apresentam-se os resultados dos modelos quer com combinação de variáveis binárias quer com uma especificação contínua.

Da análise dos resultados do modelo com especificação binária é possível constatar uma tendência decrescente da utilidade das alternativas com o aumento da variável, coerente, mas com declividades entre modos com algumas diferenças, nomeadamente para as categorias extremas. Essas diferenças verificam-se fundamentalmente nos modos BUS e Walk. Para avaliar a linearidade dos parâmetros foram realizados testes de *Wald* (MNL3.29wald_BP,

MNL3.29wald_OB, MNL3.29wald_M, MNL3.29wald_B e MNL3.29wald_W), cujos resultados são apresentados no quadro V.13.

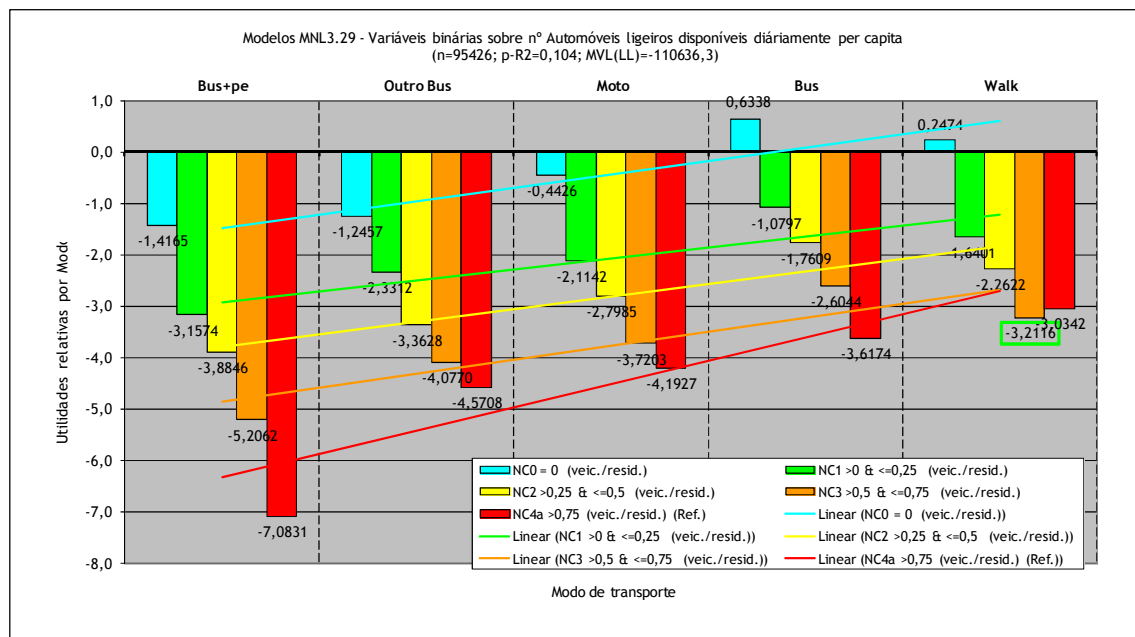


Figura V. 7 - Utilidades parciais para os diferentes modos de transporte considerando a taxa de motorização familiar especificada como variável binária para cada categoria

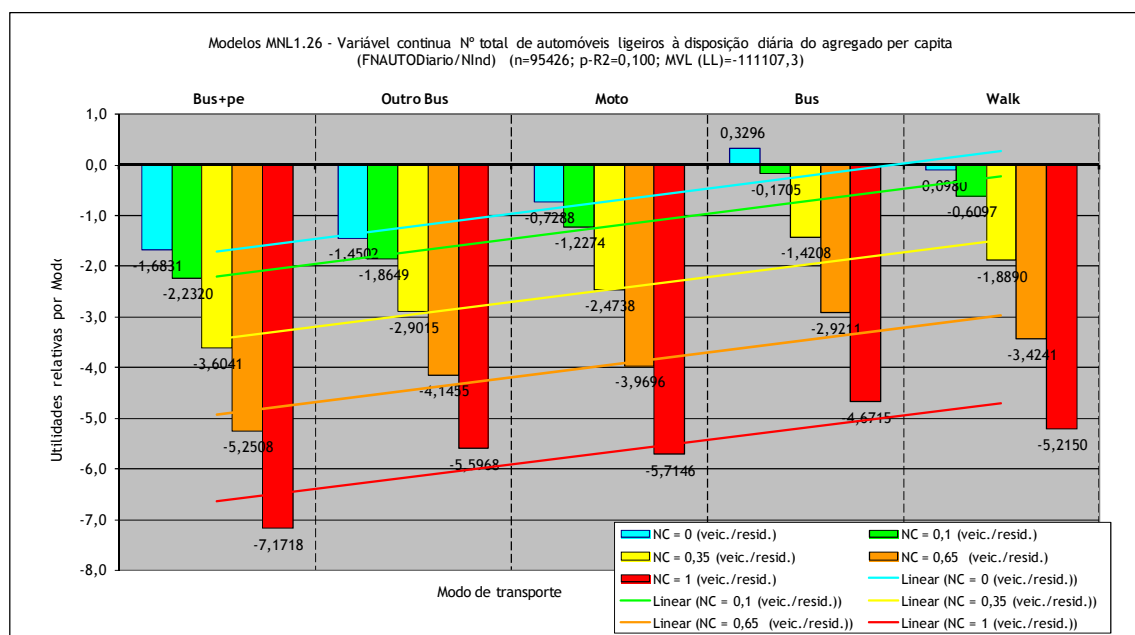


Figura V. 8 - Utilidades parciais para os diferentes modos de transporte considerando a taxa de motorização familiar especificada como variável contínua

Quadro V. 13 - Resultados do teste de *Wald* para o modelo MNL3.29 para avaliar a linearidade dos parâmetros da combinação de binárias para a taxa de motorização familiar

Modo	Teste de <i>Wald</i> Chi-quadrado	Valor crítico para $\alpha = 0,05$ e 2 GL	Comentário
BW	92,63	5,99	<i>Hipótese nula dos parâmetros serem lineares rejeitada</i>
OB	12,89		Hipótese nula dos parâmetros serem lineares rejeitada, mas por muito pouco
M	254,30		<i>Hipótese nula dos parâmetros serem lineares rejeitada</i>
B	660,30		<i>Hipótese nula dos parâmetros serem lineares rejeitada</i>
W	626,97		<i>Hipótese nula dos parâmetros serem lineares rejeitada</i>

Os resultados sugerem, com exceção do modo Outros Bus, claramente que os parâmetros das restantes alternativas não apresentam características lineares.

Comparativamente à especificação com combinação de binárias, a especificação contínua perde alguma informação ($p\text{-}R^2 = 0,100$ relativamente a 0,104 da especificação não linear), que aparentemente se deve à pior representação para as categorias extremas, diminuindo a declividade entre a categoria mais baixa e as restantes e aumentando a que se verifica entre as remanescentes. Mais uma vez, são os modos BUS e Walk que maiores discrepâncias apresentam relativamente aos resultados da especificação não linear.

Esta variável perde alguma informação com a especificação linear e o seu comportamento revela alguns resultados extremos que necessitam de uma avaliação mais cuidada. A comparação dos resultados das especificações sugere que a linear não é a representação mais correta, só devendo ser utilizada em situações em que a especificação com combinação de variáveis binárias por motivos de modelação estiver condicionada.

3.6. Disponibilidade de Moto no Agregado

Para a análise da possível influência da disponibilidade de Moto no agregado familiar na escolha modal de um indivíduo utilizou-se um modelo MNL com especificação binária. Atendendo à natureza da mesma não foi testada nenhuma outra especificação. O modelo MNL1.88 não convergiu não permitindo avaliar a coerência dos sinais e grandezas. Esta falta de convergência poderá advir desta variável ter influência provavelmente apenas para o modo MOTO, não apresentando relações significativas com as restantes alternativas. Apesar deste resultado, a mesma deverá ser avaliada em modelos com combinação com outras variáveis de forma a avaliar eventuais correlações não perceptíveis.

3.7. Licença de Condução

Para a análise da possível influência da existência de licença de condução na escolha modal de um indivíduo, independentemente de haver carro(s) no agregado familiar, utilizou-se um modelo MNL apenas com especificação binária, dada a natureza da mesma. Os resultados desse modelo para as alternativas são apresentados no quadro V.14.

Quadro V. 14 - Resultados dos modelos preliminares da influência da Licença de Condução

Modelo	MNL1.33					
Modo	B+W	B	OB	M	P	A
Lc (0=N,1=S)	-2,601	-2,299	-2,306	-1,593	-2,120	-
Amostra	95426					
p-R ²	0,098					
MVL (LL)	-111414,3					
* - p-valor >0,05 & <=0,1			** - p-valor >0,1 & <=0,25		*** - p-valor >0,25	

Dos resultados é possível verificar que o modelo tem um valor de ajustamento global significativo $p-R^2 = 0,098$ relativamente às variáveis anteriores (semelhante a relevância da taxa de motorização familiar), sugerindo que a mesma tem uma influência importante nas escolhas modais. Os seus parâmetros têm sinais negativos coerentes, evidenciando uma desvalorização significativa das alternativas quando há possibilidade de utilização do automóvel. A grandeza dos parâmetros sugere uma influência semelhante para todas as alternativas, com exceção do modo MOTO, onde se verifica um impacto significativamente menor.

4. Análise Isolada de Variáveis Relacionadas com a Posição Relativa entre a Zona de Geração e a de Atração

4.1. Afastamento entre a Geração e a Atração

4.1.1. Análise Comportamental de Diferentes Especificações

Para a análise da possível influência do afastamento entre a origem e o destino da viagem considerou-se a variável *distância mais curta entre os centróides das zonas de geração e de atração em km*. Esta variável apresenta características para uma especificação contínua. No sentido de avaliar este pressuposto, realizaram-se modelos MNL com esta variável utilizando as especificações contínua e combinação de variáveis binárias. Na sua especificação com combinação de binárias realizou-se a sua codificação atendendo às gamas de valores com amostras com tamanho representativo e cujo comportamento para algum dos modos de transporte fosse distinto. Assim, a especificação não linear foi definida com quatro categorias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.15.

Quadro V. 15 - Resultados dos modelos preliminares da influência da distância mais curta em km entre centróides das zonas de geração e de atracção das viagens

Modelo	MNL1.53		MNL3.83	
Modo	B	W	B	W
D2	-0,005*	-2,656		
Dx1			0,066***	10,559
Dx2			-0,055**	6,262
Dx3			0,038	2,654
Dx4a			-	-
Amostra	95426		95426	
p-R ²	0,194		0,194	
MVL (LL)	-99549,0		-99498,8	

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

É possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento ligeiramente superior que o modelo com especificação continua. Ambos os modelos apresentam ajustamentos globais muito elevados relativamente às variáveis anteriores. Nos modelos com especificação linear discreta o sinal negativo indicia uma perda de utilidade das alternativas Bus, Moto e Walk relativamente ao automóvel particular com o aumento da distância entre centróides. Pelos parâmetros verifica-se, como seria expectável, um impacto muito mais significativo no modo pedonal, e uma influência muito reduzida nos restantes modos (ver modelos completos nos anexos digitais). O parâmetro para o modo Bus só apresenta significância estatística para um intervalo de confiança de 90%, podendo sugerir que o impacto da distância no uso do Bus é relativamente semelhante ao impacto no Automóvel. Este resultado pode dever-se à forma como foi definida a variável distância, igual para cada par GA e para todos os modos motorizados.

Na especificação com combinação de binárias normalizou-se a categoria de mais de 5,0km de distância. Verifica-se que os parâmetros para todas as alternativas exceto Walk, para a distância ≤1,0km, não apresentam significância estatística, podendo justificar-se pela quase exclusividade do uso do modo pedonal para estas distâncias. Para as restantes categorias os sinais dos parâmetros variam mas sugerem que a Moto e o Bus têm maior utilidade para distâncias entre os 2,0 e os 5,0km, e que modo Walk deixa totalmente de ser alternativa ao automóvel a partir de 2,0km. Este resultado revela que a distância entre centróides tem maior impacto nas curtas distâncias, permitindo justificar a preponderância muito grande do modo pedonal, particularmente nas viagens até 1km, e que em relação a distâncias superiores o seu impacto na utilidade dos Bus é relativamente equivalente ao automóvel.

Atendendo aos resultados para o modo Walk, valorizado sobretudo para distâncias até 1km, ponderou-se a hipótese de conjugar as duas especificações. O modelo MNL3.80 conjuga a variável contínua com a variável binária para distâncias até 1 km (Dx1), enquanto o modelo MNL3.84 conjuga a especificação contínua com a combinação das quatro variáveis binárias para diferentes distâncias (Dxi). Os resultados do teste estatístico da razão de MVL (quadro

V.16) para estes modelos permitem concluir que a combinação da especificação contínua com a combinação de binárias é significativamente melhor que a das especificações em separado (MNL3.83). Verifica-se também que a especificação com combinação de binárias (MNL3.84) também é estatisticamente superior à especificação contínua (MNL1.53).

Quadro V. 16 - Resultados do teste estatístico da razão das MVL com o modelo com especificação contínua como base

	MNL1.53	MNL3.83	MNL3.84	MNL3.80
D2	10 Parâmetros		X	X
Dxi		20 Parâmetros	25 Parâmetros	
Dx1				15 Parâmetros
p-R ²	0,194	0,194	0,202	0,198
LL _{base}	-99549,0	-99549,0	-99549,0	-99549,0
LL	-99549,0	-99498,8	-98602,9	-99003,9
-2LL		100,4	1892,2	1090,2
Graus de liberdade		10	15	5
Chi ² ($\alpha=0,05$)		18,3	25,0	11,1

Do teste estatístico da razão das MVL em relação ao modelo apenas com combinação de binárias (quadro V.17), também se constata que as combinações são melhores que o modelo base MNL3.83 (entre 5 e 100 vezes maiores que o valor crítico do teste de qui-quadrado). Comparando os modelos de combinações contínua/binárias, o modelo com o conjunto de binárias (MNL 3.84) apresenta um melhor resultado que apenas com a variável Dx1 (MNL3.80).

Quadro V. 17 - Resultados do teste estatístico da razão das MVL com modelo com especificação com combinação de binárias como base

	MNL3.83	MNL3.84	MNL3.80
D2		X	X
Dxi	20 Parâmetros	25 Parâmetros	
Dx1			15 Parâmetros
p-R ²	0,194	0,202	0,198
LL _{base}	-99498,8	-99498,8	-99498,8
LL	-99498,8	-98602,9	-99003,9
-2LL		1791,8	989,8
Graus de liberdade		5	10
Chi ² ($\alpha=0,05$)		11,1	18,3

Se os resultados sugerem que o modelo combinando especificações contínuas com binárias é aparentemente mais correta para a variável distância mais curta, a constatação de que essa influência não é relevante para determinadas alternativas a partir de determinadas distâncias leva à necessidade de ponderar não aumentar a complexidade do modelo pela introdução de maior número de variáveis para um aumento de ajustamento global apenas ligeiro, apesar de estatisticamente significativo.

4.1.2. Avaliação de Comportamentos Não Lineares

Para auxiliar na ponderação referida anteriormente, realizou-se um estudo comparativo das utilidades relativas das alternativas para os diferentes modelos de combinação. Nas figuras V-9 e 10 apresentam-se os resultados dos modelos quer com combinação de variáveis binárias quer com a especificação contínua.

Dos resultados obtidos para as utilidades parciais, constata-se que, apesar de o ajustamento global do modelo para ambas as especificações ser praticamente igual, os declives da especificação não linear com combinação de binárias não são similares para todos os modos. Estes resultados podem devem-se ao facto de para distâncias curtas (≤ 1 km) a maior parte das viagens ser realizada a pé e não existir amostra para os restantes modos. A especificação linear contínua aparentemente não consegue identificar os problemas para as distâncias ≤ 1 km para os modos alternativos ao pedonal. Esta constatação levou à consideração de uma eventual combinação da especificação linear e da especificação de combinação de binárias para avaliar a possibilidade de colmatar as discrepâncias verificadas. As figuras V-11 e 12 apresentam os resultados dessas combinações.

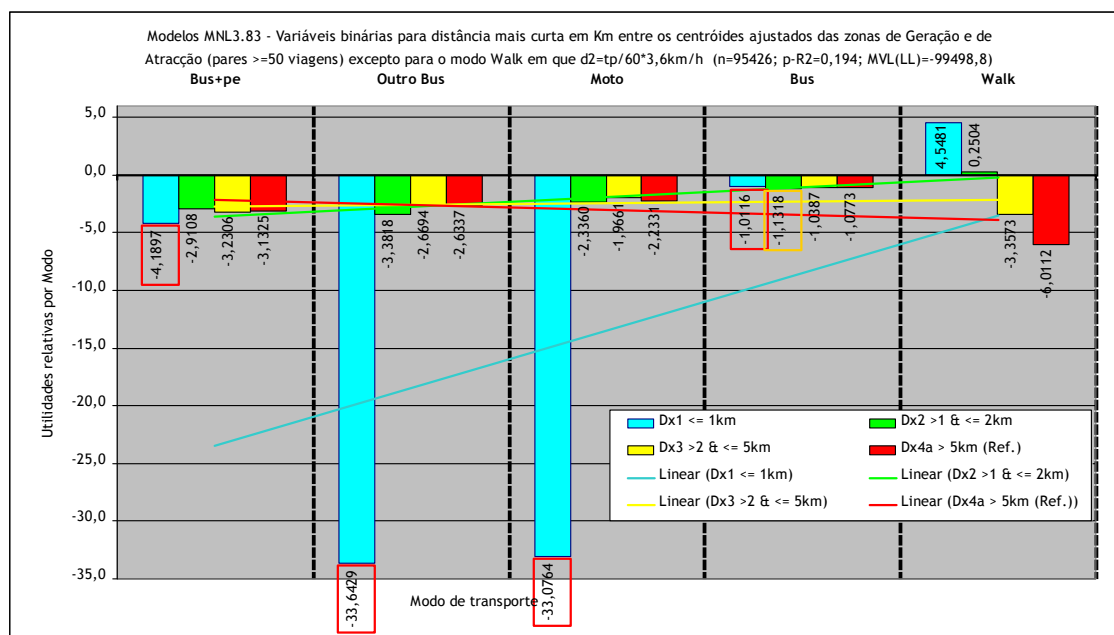


Figura V. 9 - Utilidades parciais para os diferentes modos de transporte considerando a distância mais curta entre centróides em km como variável não continua

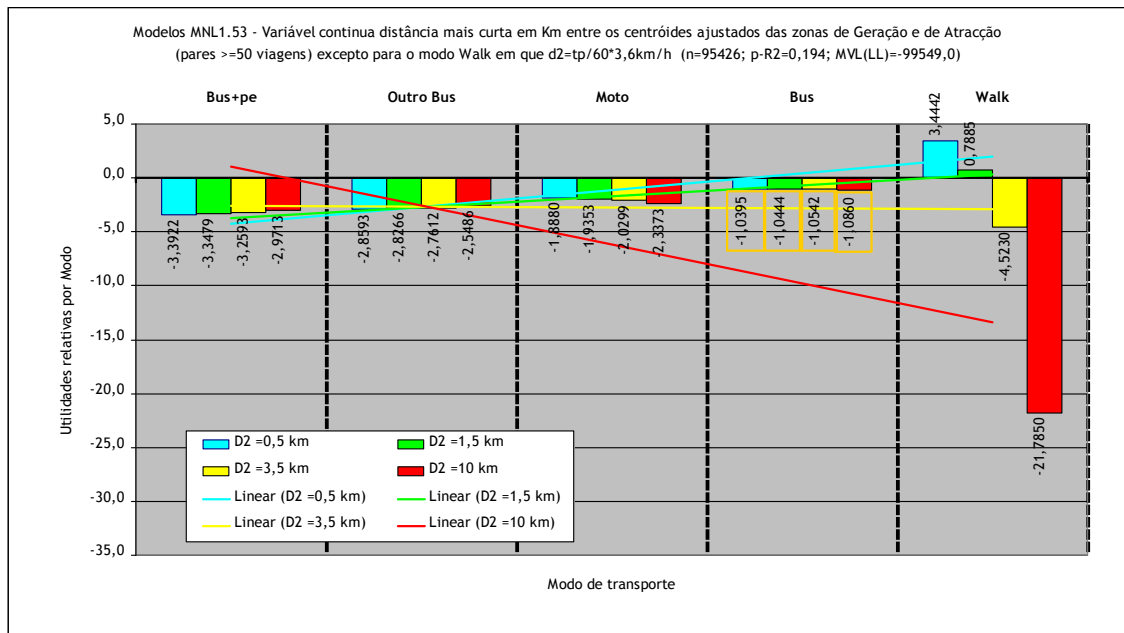


Figura V. 10 - Utilidades parciais para os diferentes modos de transporte considerando a distância mais curta entre centróides em km como variável contínua

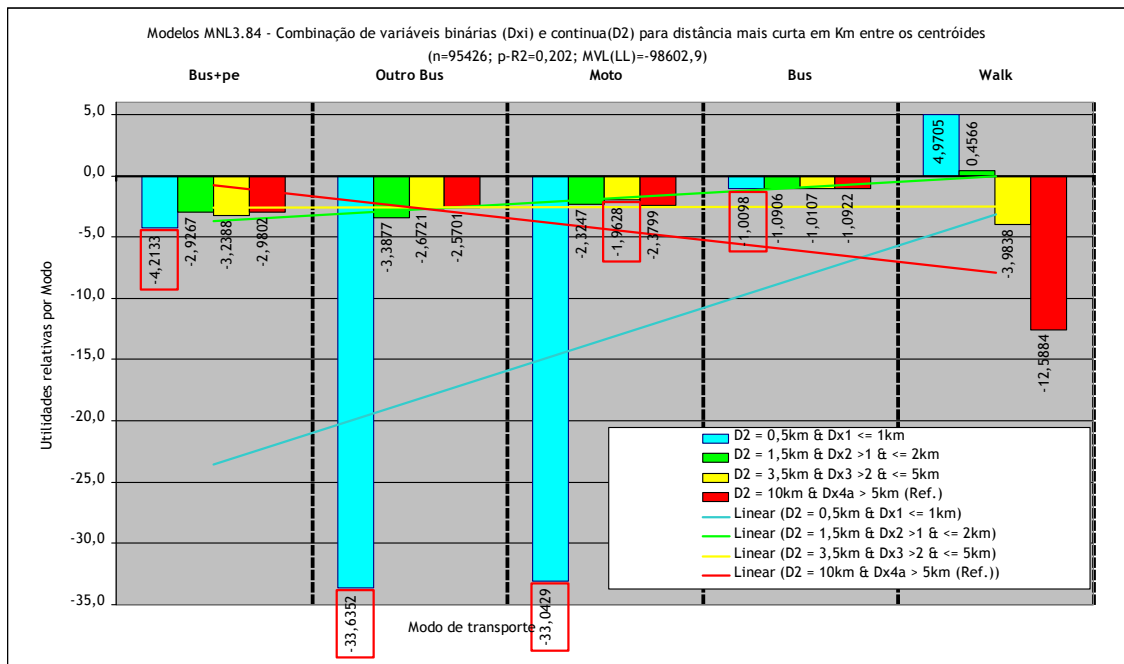


Figura V. 11 - Utilidades parciais para os diferentes modos de transporte considerando a combinação das especificações contínua e binárias da distância mais curta entre centróides em km

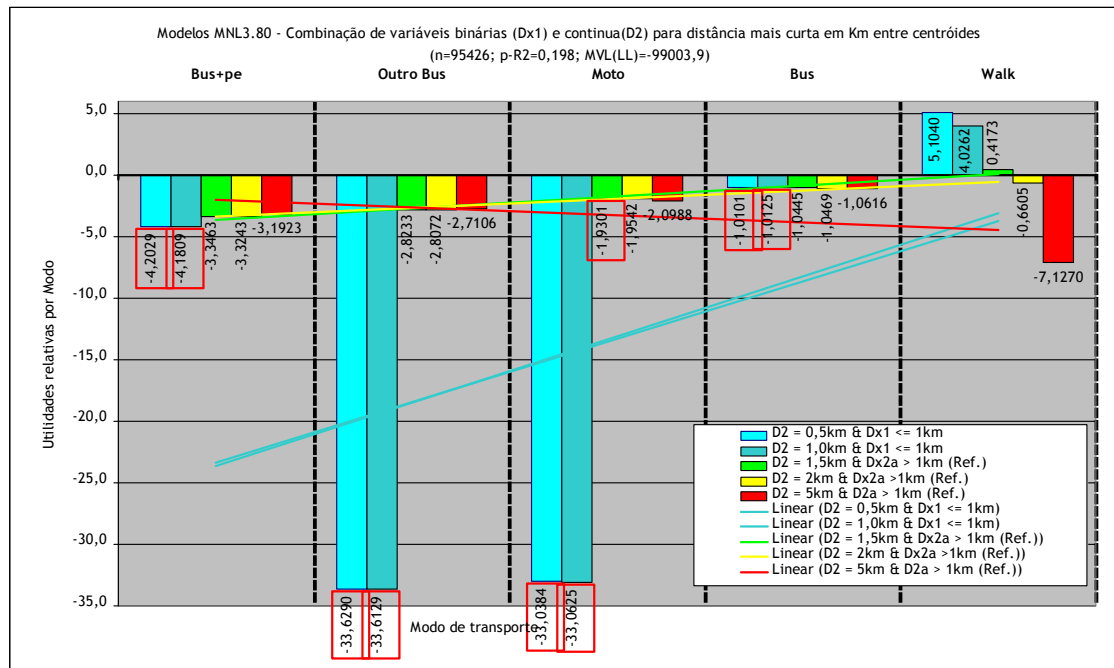


Figura V. 12 - Utilidades parciais para os diferentes modos de transporte considerando a combinação das especificações contínua e binárias (apenas Dx1 <=1km) da distância mais curta entre centróides em km

Atendendo ao ajustamento global e aos sinais, os resultados revelaram que a combinação das duas especificações melhora ligeiramente o ajustamento global do modelo, mas que a mesma se deve principalmente à combinação com a variável binária distância mais curta dx1. É possível verificar que os resultados do modelo MNL3.80 são muito semelhantes aos resultados da especificação com combinação de binárias. Destes resultados, e atendendo à maior complexidade introduzida com a combinação da variável contínua com todas as binárias, aparentemente a combinação variável contínua com uma binária para as distâncias mais curtas (dx1) apresenta-se com resultados suficientemente coerentes para ser a melhor especificação da variável distância mais curta entre centróides. A sua utilização combinada deve, no entanto, ainda ser validada através dos resultados verificados nos modelos mais complexos com combinação de outras variáveis.

4.2. Duração Total da Viagem por Modo

4.2.1. Análise Comportamental de Diferentes Especificações

Para a análise da possível influência da duração total de viagem entre a geração e a atracção da viagem, utilizou-se o tempo declarado (e portanto apreendido) por cada indivíduo para o modo escolhido. Esta variável tem características para, tal como por exemplo o custo da viagem, ser diferenciada entre alternativas para um mesmo indivíduo, sendo portanto genérica. Dada a falta de dados para as alternativas não escolhidas, utilizou-se a Duração Média da Viagem apreendida para todas as viagens entre cada par GA nessas alternativas.

Esta metodologia apesar de permitir definir a variável como genérica (TiA), reduziu a amostra disponível (30 559 inquéritos completos) pelo facto de nem todos os pares GA apresentarem viagens em todos os modos considerados. Os resultados do modelo MNL com as mesmas são apresentados no quadro V.18.

Quadro V. 18 - Resultados dos modelos preliminares da influência do tempo total de viagem em minutos entre a zona de geração e de atracção das viagens

Modelo	MNL1.70
Modo	Todos
TiA	-0,002
Amostra	30559
p-R ²	0,00006
MVL (LL)	-39684,6

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Dos resultados é possível verificar que o modelo apresenta um ajustamento global muito reduzido. O sinal negativo do parâmetro genérico é coerente com outros estudos internacionais, uma vez que tempos de viagem maiores são geralmente avaliados negativamente pela maioria dos indivíduos, ou seja, quanto maior o tempo de viagem menor a sua atratividade.

É de salientar que o facto de ser um parâmetro genérico implica valores da variável diferentes para cada alternativa para um mesmo indivíduo, não necessitando de normalização em relação a um modo para ser estimado pelo modelo, sendo constante para todos os modos. Este facto implica que também afete a utilidade do modo automóvel. Para a definição da utilidade relativa que lhe está associada haverá a necessidade de se conhecer o tempo da viagem para os diferentes modos disponíveis, e será essa diferença de tempos de viagem que definirá a utilidade relativa de cada modo.

4.2.2. Avaliação de uma Especificação Alternativa

Koppelman e Bhat (2006) sugerem que o uso do logaritmo do valor do tempo tende a apresentar resultados melhores que a especificação da variável como contínua linear. Tal foi testado apresentando-se os resultados desse modelo no quadro V.19.

Dos resultados é possível verificar que o modelo com especificação logarítmica apresenta um ajustamento global modesto mas claramente superior ao modelo com especificação contínua linear. Em ambas a especificações o sinal negativo do parâmetro genérico é coerente pelo que se pode concluir que a especificação logarítmica aparenta ser mais correta que a linear contínua simples.

Quadro V. 19 - Resultados dos modelos preliminares da influência do tempo total de viagem em minutos entre a zona de geração e de atracção das viagens e do logaritmo do mesmo

Modelo	MNL1.70	MNL1.80
Modo	Todos	Todos
Tia	-0,002	
Til = ln(Tia)		-0,752
Amostra	30559	30559
p-R ²	0,00006	0,017
MVL (LL)	-39684,6	-39013,4

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

5. Análise Isolada de Variáveis Relacionadas com a Organização das Atividades Diárias

5.1. Motivo da Viagem

5.1.1. Análise Comportamental de Diferentes Especificações

A primeira variável relacionada com a organização das atividades diárias considerada foi o *motivo da viagem*. Esta variável é de natureza discreta nominal, não apresentando à partida características para poder ser especificada de forma linear. Para a análise da possível influência do motivo da viagem nas escolhas modais, realizaram-se modelos MNL utilizando a especificações discreta e combinação de binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.20.

Quadro V. 20 - Resultados dos modelos preliminares da influência do motivo da viagem

Modelo	MNL3.89		MNL1.87	
Modo	B	W	B	W
Mt			0,040	0,145
Tr	0,1072	-0,4592		
Rg	0,3541	-0,0420**		
Es	1,4604	0,8124		
Lz	-	-		
CS	0,2451	0,1426		
Amostra	95426		95426	
p-R ²	0,0204		0,004	
MVL (LL)	-120967,6		-123003,6	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Dos resultados é possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento relativamente baixo, em linha com os valores de outras variáveis menos relevantes, mas superior que a especificação discreta linear. Na análise com especificação de combinação de binárias a categoria normalizada representa as viagens para Lazer. Tal como esperado, a influência destas variáveis nos diferentes modos é bastante diferente e não apresenta tendências lineares. Apesar destas diferenças é possível constatar algumas tendências coerentes. Todas as alternativas exceto Walk são avaliadas positivamente

para viagens para o trabalho e de regresso a casa relativamente ao automóvel para Lazer. Todos os modos alternativos exceto a Moto apresentam uma influência positiva no caso de viagens para a escola e para compras ou serviços relativamente ao automóvel para Lazer. Para viagens para a escola e para compras ou serviços o modo pedonal apresenta uma influência positiva relativamente ao automóvel, resultado que deverá ser cuidadosamente avaliado noutras análises, uma vez que este resultado pode ser devido ao efeito de curtas distâncias entre pontos de GA.

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.21.

Quadro V. 21 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.87	MNL3.89
Amostra	95426	95426
Mt	10 Parâmetros	
Mti	20 Parâmetros	
LL _{base}	-123003,6	-123003,6
LL	-123003,6	-120967,6
-2LL		4072
Graus de liberdade		10
Chi ² ($\alpha=0,05$)		18,3

O valor crítico de Chi-quadrado é cerca de 223 vezes inferior ao teste estatístico -2LL concluindo-se que o modelo com especificação como combinação de variáveis binárias é estatisticamente mais próximo de zero que o modelo com especificação discreta, e portanto melhor.

5.1.2. Avaliação de Comportamentos Não Lineares

Para uma melhor visualização dos resultados do ponto anterior compararam-se as utilidades parciais das alternativas do modelo com especificação de combinação de binárias. Os resultados dessa análise apresentam-se na figura V.13.

Constata-se que, de facto, a variável não apresenta características lineares nem declives similares entre modos. O comportamento, apesar de não linear, apresenta-se ainda assim coerente com o esperado, com uma maior valorização das alternativas ao automóvel (com exceção do modo MOTO) para viagens para a escola. O modo pedonal também se revela o menos atrativo para viagens para o trabalho e de regresso a casa. Atendendo aos resultados, esta variável não apresenta efetivamente características que possam ser representadas por uma variável discreta linear, apresentando-se a especificação não linear com a única forma de captar as espectáveis diferentes influências do motivo da viagem nos diferentes modos de transporte.

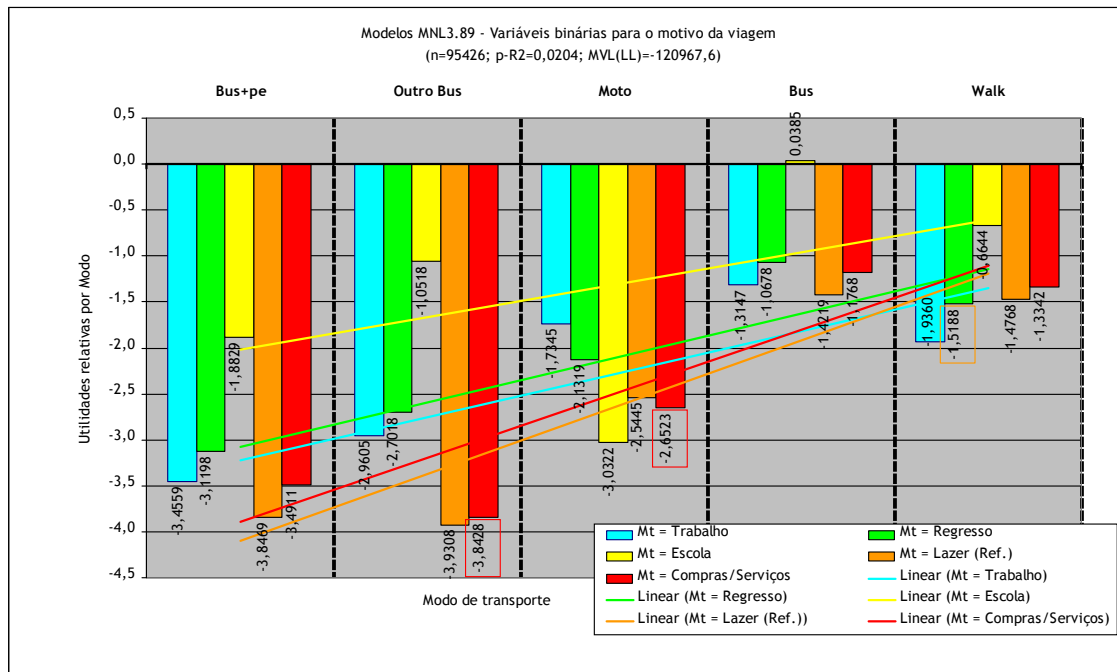


Figura V. 13 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para o motivo da viagem

5.2. Duração da Atividade no Destino da Viagem

5.2.1. Análise Comportamental de Diferentes Especificações

A variável *duração da estadia no destino em minutos* tem características para a classificação como contínua. Para avaliar a viabilidade desta especificação foram definidos cinco intervalos atendendo às durações de estadia mais comuns para diferentes atividades, especificando cinco variáveis binárias. Para a análise da sua possível influência nas escolhas modais realizaram-se modelos MNL utilizando as especificações contínua e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.22. Na análise com especificação de combinação de binárias, a categoria normalizada representa as estadias de curta duração (≤ 60 minutos) que na base de dados correspondiam à categoria com maior utilização do automóvel.

No modelo com especificação linear contínua o sinal positivo para todos os modos exceto o pedonal indicia uma maior utilidade deste modo para as atividades de curta duração.

Da análise dos resultados da especificação com combinação de binárias, é possível constatar que os modos que envolvem o BUS não só apresentavam parâmetros positivos como crescentes com o aumento da duração da atividade. Estes resultados sugerem que, como era expectável, estes modos ganham utilidade à medida que a duração da atividade também aumenta, comparativamente à estadia normalizada. Para os modos moto e pedonal a influência da duração das atividades é menos estável, apresentando um comportamento

variável mas que sugere uma tendência para menor utilidade com o aumento da duração da atividade.

Quadro V. 22 - Resultados dos modelos preliminares da influência da duração em minutos da atividade no destino da viagem

Modelo	MNL1.71		MNL3.120a	
Modo	B	W	B	W
De	0,0012	-0,0009		
De1			-	-
De2			0,3172	-0,0160***
De3			0,5439	0,1017
De4			0,6519	0,0452**
De5			0,8825	-0,5743
Amostra	61358		61358	
p-R ²	0,011		0,014	
MVL (LL)	-78298,4		-78027,4	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

Dos resultados é ainda possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento global ligeiramente superior que o modelo com especificação contínua. Ambos os modelos apresentam ajustamentos globais reduzidos. O teste estatístico da razão das MVL (-2LL), com modelo base MNL1.71 com especificação contínua é apresentado no quadro V.23.

Quadro V. 23 - Resultados do teste estatístico da razão das MVL com modelo com especificação contínua como base

	MNL1.71	MNL3.120a
Amostra	61358	61358
De	10 Parâmetros	
Dei	20 Parâmetros	
LL _{base}	-78298,4	-78298,4
LL	-78298,4	-78027,4
-2LL		542
Graus de liberdade		10
Chi ² (α=0,05)		18,3

O valor crítico de Chi-quadrado é cerca de 29 vezes inferior ao teste estatístico -2LL, pelo que o modelo com especificação como combinação de variáveis binárias é estatisticamente mais próximo de zero que o modelo com especificação contínua, e portanto melhor.

5.2.2. Avaliação de Comportamentos Não Lineares

Para avaliar a linearidade dos valores dos parâmetros compararam-se as utilidades parciais para cada modo de transporte obtidas da aplicação a modelos MNL, apresentando-se os resultados nas figuras V.14 e V.15.

Da análise do comportamento e dos sinais é possível constatar que a especificação não linear apresenta declividades diferentes para as diferentes categorias, mas uma tendência semelhante para o aumento da utilidade das alternativas face ao automóvel (com exceção do modo Walk) com o aumento da duração da estadia. O modo pedonal é o único cuja utilidade decresce com o aumento da duração da estadia. As utilidades relativas numa análise qualitativa apresentam um comportamento quase linear, com exceção para os modos Moto e Walk. Para avaliar a linearidade entre parâmetros para as alternativas, realizou-se o teste de *Wald* ao modelo com especificação não linear (MNL3.120aWald_BP, MNL3.120aWald_OB, MNL3.120aWald_M, MNL3.120aWald_B e MNL3.120aWald_W) com duas restrições lineares para que a diferença entre parâmetros fossem iguais entre si (e portanto lineares). Os resultados do teste são apresentados no quadro V.24.

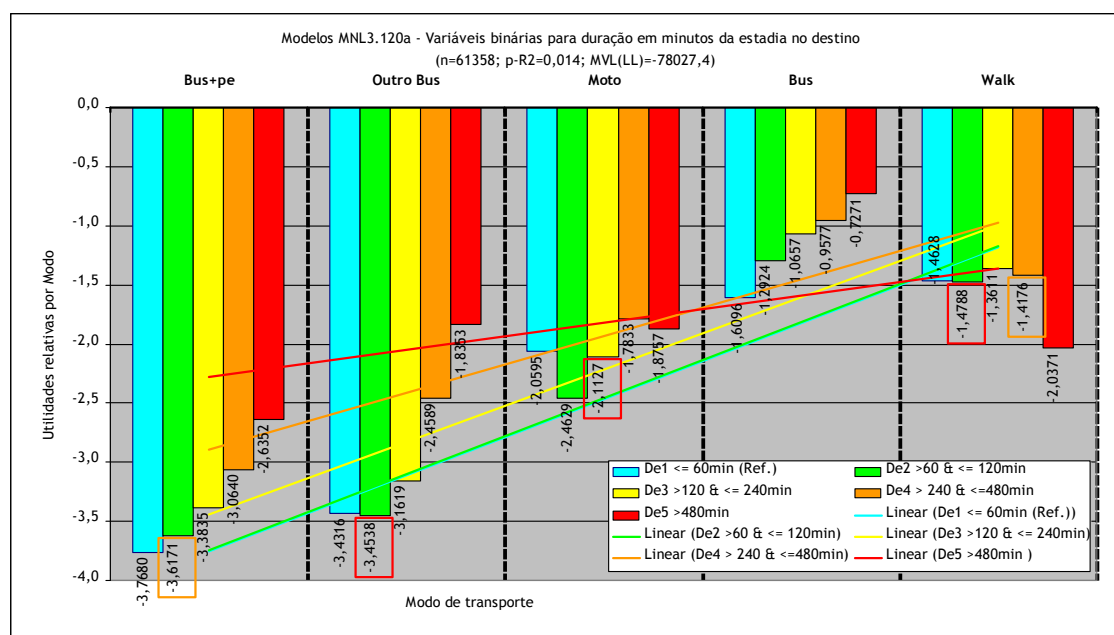


Figura V. 14 - Utilidades parciais para os diferentes modos de transporte considerando a especificação binária da duração da estadia no destino em minutos

A análise dos resultados do teste de *Wald* corrobora a análise qualitativa, revelando que estatisticamente os parâmetros para os modos BW e B apresentam um comportamento linear, e que, apesar de estatisticamente não ser significativo, para o modo OB também apresentam valores próximos de um comportamento linear. Apenas para os modos Moto e Walk o teste rejeita a hipótese de a diferença entre parâmetros ser igual.

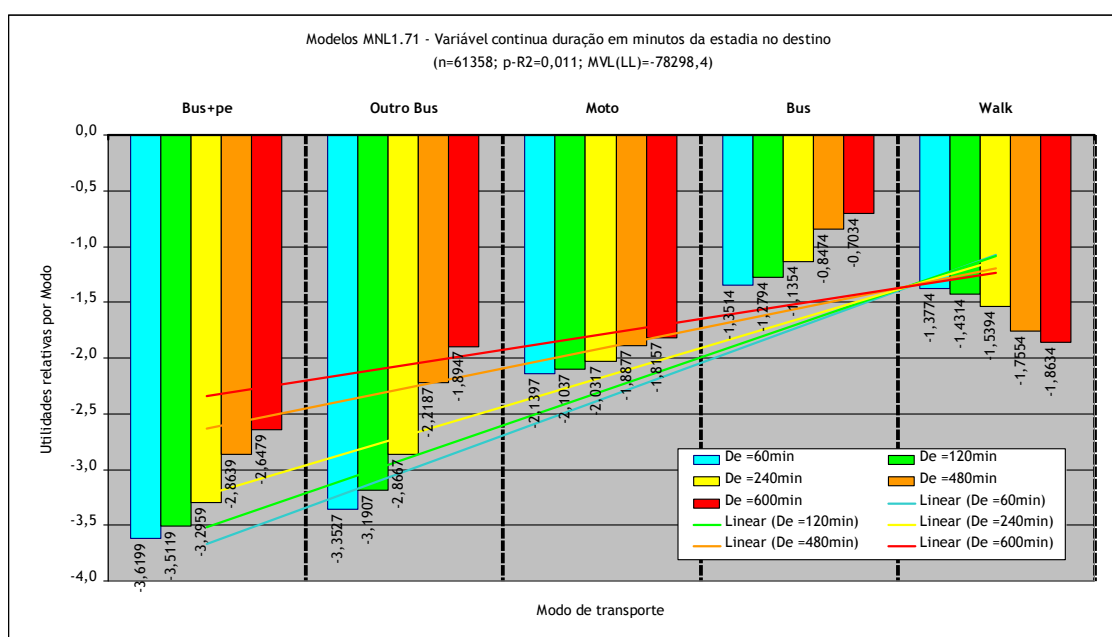


Figura V. 15 - Utilidades parciais para os diferentes modos de transporte considerando a especificação contínua da duração da estadia no destino em minutos

Quadro V. 24 - Resultados do teste de *Wald* para o modelo MNL3.120a para avaliar a linearidade dos parâmetros da combinação de binárias para a duração da atividade

Modo	Teste de <i>Wald</i> Chi-quadrado	Valor crítico para $\alpha = 0,05$ e 2 GL	Comentário
BW	2,45	5,99	Hipótese nula dos parâmetros serem lineares não pode ser rejeitada
OB	11,13		Hipótese nula dos parâmetros serem lineares rejeitada, mas por muito pouco
M	44,48		Hipótese nula dos parâmetros serem lineares rejeitada
B	5,36		Hipótese nula dos parâmetros serem lineares não pode ser rejeitada
W	144,22		Hipótese nula dos parâmetros serem lineares rejeitada

Assim, os resultados sugerem que a especificação não linear é melhor que a especificação contínua. Apesar disso, a especificação contínua poderá ser utilizada no caso de a modelação necessitar de uma especificação mais simples, em que as discrepâncias verificadas para os modos M e W se considerem ignoráveis perante outros aspetos da modelação, como a rapidez de cálculo ou uma menor complexidade da mesma.

5.2.3. Avaliação de uma Especificação Alternativa

Atendendo aos resultados ponderou-se também a possibilidade de especificar a variável duração da estadia no local de destino em minutos através do seu logaritmo como forma de melhorar a sua capacidade explicativa (quadro V.25).

Quadro V. 25 - Resultados dos modelos preliminares com diferentes especificações para a variável duração da estadia

Modelo	MNL1.71		MNL1.82		MNL1.83		MNL1.85	
Modo	B	W	B	W	B	W	B	W
De	0,0012	-0,0009						
DeL = ln(De)			0,2718	-0,0664				
De1m					0,0012	-0,0009		
D1L = ln(De1m)							0,2571	-0,0640
Amostra	61358		60954		61358		61358	
p-R ²	0,0108		0,0098		0,0108		0,0096	
MVL (LL)	-78298,4		-77911,0		-78298,4		-78393,2	
* - p-valor >0,05 & <=0,1			** - p-valor >0,1 & <=0,25			*** - p-valor >0,25		

O modelo com a especificação logarítmica não revelou ajustamentos globais melhores. A existência do valor Zero minutos de estadia em algumas viagens levou a uma ligeira redução da amostra para 60 954. Para colmatar esta inconsistência optou-se por substituir os valores de zero por 1 minuto de estadia, considerado o tempo mínimo de uma paragem. A comparação dos modelos com esta variável corrigida e com a sua especificação logarítmica revelou ajustamentos semelhantes à variável sem correção e um ajustamento pior para a especificação logarítmica, enquanto os sinais dos parâmetros se mantiveram coerentes. Assim, os resultados sugerem não se justificar a utilização da especificação logarítmica para esta variável.

5.3. Número de Viagens no Dia

5.3.1. Análise Comportamental de Diferentes Especificações

A variável *número de viagens diárias de um indivíduo* apresenta características de variável contínua inteira. A análise com especificação não linear foi realizada com três variáveis binárias atendendo à gama de possíveis valores constatados na base de dados. Para a análise da sua possível influência nas escolhas modais, realizaram-se modelos MNL utilizando as especificações discreta e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.26 (restantes valores em anexo digital).

Dos resultados é possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento ligeiramente inferior que o modelo com especificação contínua. Ambos os modelos apresentam ajustamentos globais reduzidos.

Quadro V. 26 - Resultados dos modelos preliminares da influência do número total de viagens num dia

Modelo	MNL1.66		MNL3.136	
Modo	B	W	B	W
nV	-0,4131	-0,1821		
nV1			1,8032	0,8508
nV2			1,0916	0,7182
nV3a			-	-
Amostra	95426		95426	
p-R ²	0,031		0,031	
MVL (LL)	-119661,3		-119695,5	
	* - p-valor >0,05 & <=0,1		** - p-valor >0,1 & <=0,25	
			*** - p-valor >0,25	

No modelo com especificação linear o sinal negativo para todos os modos alternativos é coerente com a assunção que a flexibilidade e disponibilidade do automóvel o tornam no modo mais atrativo quando se pretende realizar varias viagens, durante o dia. Os modos para os quais o número total de viagens diárias apresenta menor impacto são o modo pedonal e Moto, mais uma vez de acordo com a flexibilidade e disponibilidade que estes modos também normalmente apresentam.

Na análise com especificação de combinação de binárias a categoria normalizada representa movimentações diárias com mais de quatro viagens. Em relação a esta categoria todos os modos apresentaram valores positivos, sugerindo que efetivamente para a categoria mais elevada o automóvel é realmente o modo mais valorizado. Com exceção do modo moto, todas as outras alternativas apresentaram tendência de utilidades decrescentes com o aumento do número de viagens diárias. O decréscimo da utilidade do modo pedonal deve ser devidamente ponderado noutras modelações, pois a redução a utilidade deste modo pode dever-se não exclusivamente ao aumento do número de viagens diárias mas também a um aumento da distância entre atividades.

Quadro V. 27 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.66	MNL3.136
Amostra	95426	95426
nV	10 Parâmetros	
nVi	15 Parâmetros	
LL _{base}	-119661,3	-119661,3
LL	-119661,3	-119695,5
-2LL		-68,4
Graus de liberdade		5
Chi ² (α=0,05)		11,1

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.27 e verifica-se que o valor crítico de Chi-quadrado é superior ao teste estatístico -2LL pelo que não se pode recusar

a hipótese nula de que o “novo” modelo (com especificação de combinação de binárias) não melhora estatisticamente a MVL em relação ao modelo base (com especificação contínua). Assim, pode-se concluir que o modelo com especificação contínua é estatisticamente mais próximo de zero que o modelo com especificação como combinação de variáveis binárias, e portanto melhor.

É possível verificar que, apesar de muito semelhantes, o ajustamento global com a especificação linear é efetivamente ligeiramente melhor que para a não linear. A justificação para estes resultados pode dever-se ou a uma desadequada especificação considerada nas variáveis binárias ou a um efetivo impacto linear desta variável para os diferentes modos.

5.3.2. Avaliação de Comportamentos Não Lineares

Apesar dos resultados anteriores, para avaliar a linearidade dos parâmetros compararam-se as utilidades parciais das diferentes alternativas para ambas as especificações. Os resultados dessa comparação apresentam-se nas figuras V.16 e V.17.

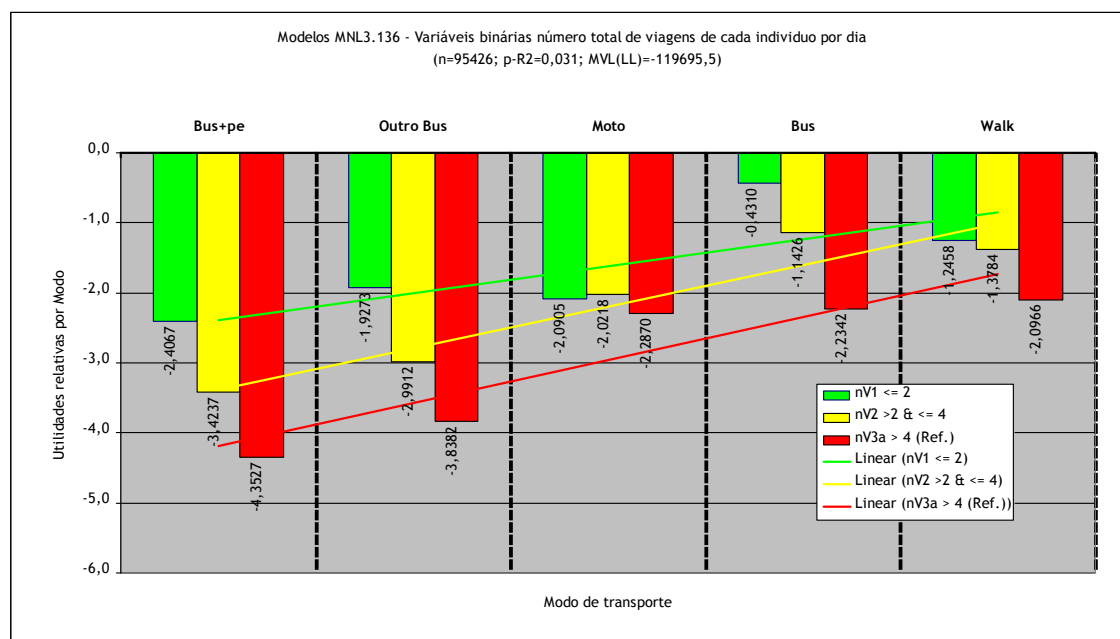


Figura V. 16 - Utilidades parciais para os diferentes modos de transporte considerando a especificação não linear para o número de viagens diárias de um indivíduo

O comportamento de ambos os modelos é coerente com a diminuição da utilidade de todas as alternativas com o aumento no número de viagens diárias comparativamente ao automóvel particular. A especificação linear apresenta declividades entre categorias semelhantes à não linear e o comportamento para os diferentes modos em ambas as especificações também é similar. Parece, assim, possível assumir que entre as especificações estudadas a linear é a mais apropriada para representar esta variável.

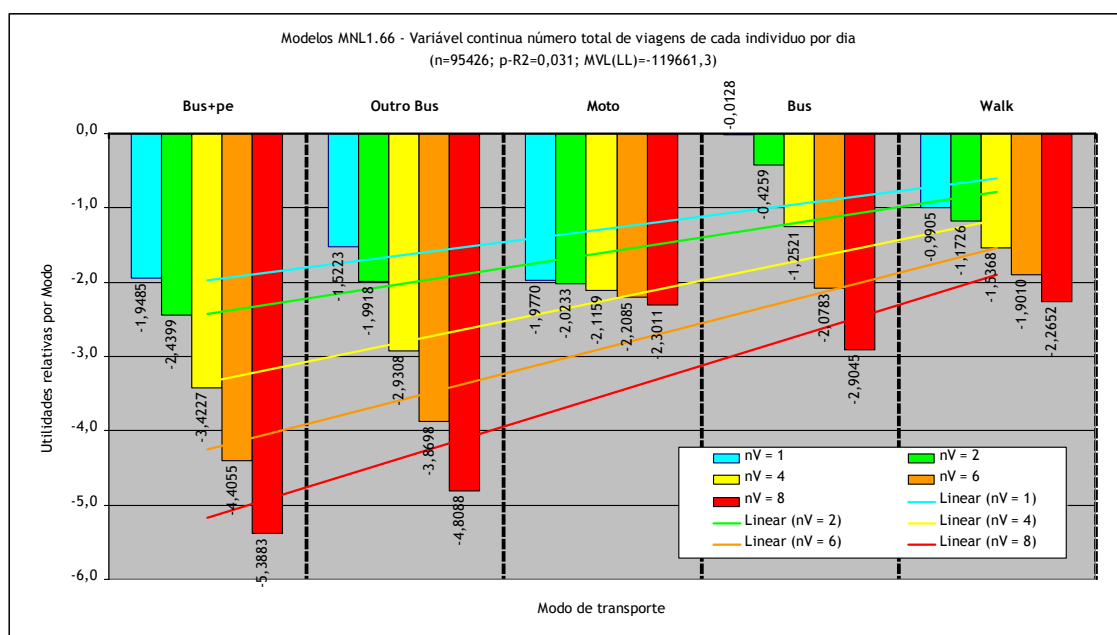


Figura V. 17 - Utilidades parciais para os diferentes modos de transporte considerando a especificação linear para o número de viagens diárias de um indivíduo

6. Variáveis Relativas a Efeitos Isolados Relacionados com Políticas de Estacionamento

6.1. Rácio “Oferta de Estacionamento / Total de Viagens Atraídas”

6.1.1. Análise Comportamental de Diferentes Especificações

A variável relacionada com estacionamento que traduz a *razão entre a oferta total de estacionamento na zona de atracção com o total de viagens extrapoladas para a zona de atracção* (independentemente do modo utilizado) devido à sua forma de cálculo apresenta características de uma variável contínua. Para avaliar a linearidade da especificação contínua, realizou-se uma análise paralela utilizando quatro variáveis binárias. Na sua definição assumiram-se os valores que permitiam amostras semelhantes para cada categoria. É de salientar a severa redução na amostra em relação às análises anteriores devido à falta de informação nas zonas externas à cidade do Porto. Para a análise da possível influência da proporção de lugares de estacionamento relativamente ao total de viagens para uma zona de atracção nas escolhas modais, ou seja, da densidade de oferta de estacionamento face à procura, realizaram-se modelos MNL utilizando as especificações contínua e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.28.

Na especificação com combinação de binárias normalizou-se a categoria mais elevada pois corresponde a uma proporção de oferta de estacionamento superior o que geralmente deverá potenciar o uso do automóvel. Nos resultados é ainda possível verificar que ambos os modelos

apresentam ajustamentos globais reduzidos, apresentando a especificação contínua um ajustamento próximo de metade do da especificação com combinação de binárias. No modelo com especificação contínua o sinal negativo para todos os modos exceto Outros Bus (autocarros não regulares) confirma uma menor utilidade destes modos quando a oferta de estacionamento aumenta.

Quadro V. 28 - Resultados dos modelos preliminares da influência da proporção de estacionamentos relativamente ao total de viagens para uma zona de atracção

Modelo	MNL1.55		MNL3.71	
Modo	B	W	B	W
OV	-11,3784	-6,6556		
OV1			1,3280	1,0314
OV2			0,9133	0,7212
OV3			0,5271	1,5958
OV4			-	-
Amostra	18823		18823	
p-R ²	0,018		0,031	
MVL (LL)	-23479,9		-23186,5	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

Quadro V. 29 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.55	MNL3.71
Amostra	18823	18823
OV	10 Parâmetros	
OVi	20 Parâmetros	
LL _{base}	-23479,9	-23479,9
LL	-23479,9	-23186,5
-2LL		586,8
Graus de liberdade		10
Chi ² (α=0,05)		18,3

Da análise dos resultados da especificação com combinação de binárias é possível constatar que o modo BUS não só apresenta parâmetros positivos como decrescentes com o aumento da oferta de estacionamento. Este comportamento é coerente com a teoria comportamental de que o aumento da oferta de estacionamento potencia o uso do automóvel e diminui a utilidade das alternativas. Para os restantes modos esse comportamento não é tão homogêneo. Os parâmetros para os modos Bus+Walk, Moto e pedonal, apesar de positivos não têm tendência linear, e para o modo Outros Bus o sinal é contrário ao expectável ou sem significância estatística. Estes resultados, no entanto, devem ser analisados com cautela, pois apenas são representativos de viagens com atracção no Porto, reduzindo significativamente as amostras dos modos menos utilizados o que poderá, pelo menos em parte, justificar os resultados menos coerentes, particularmente numa análise onde apenas se aplica uma

variável explicativa, não se podendo portanto avaliar eventuais influências cruzadas entre diferentes tipos de variáveis explicativas.

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.29.

O valor crítico de Chi-quadrado é de 18,3, 32 vezes inferior ao teste estatístico -2LL, concluindo-se que o modelo com especificação de combinação de binárias é estatisticamente mais próximo de zero que o modelo com especificação como contínua, e portanto melhor.

6.1.2. Avaliação de Comportamentos Não Lineares

No sentido de avaliar a linearidade dos resultados, compararam-se as utilidades relativas das alternativas para as especificações referidas. Os resultados destas análises são apresentados nas figuras V.18 e V.19.

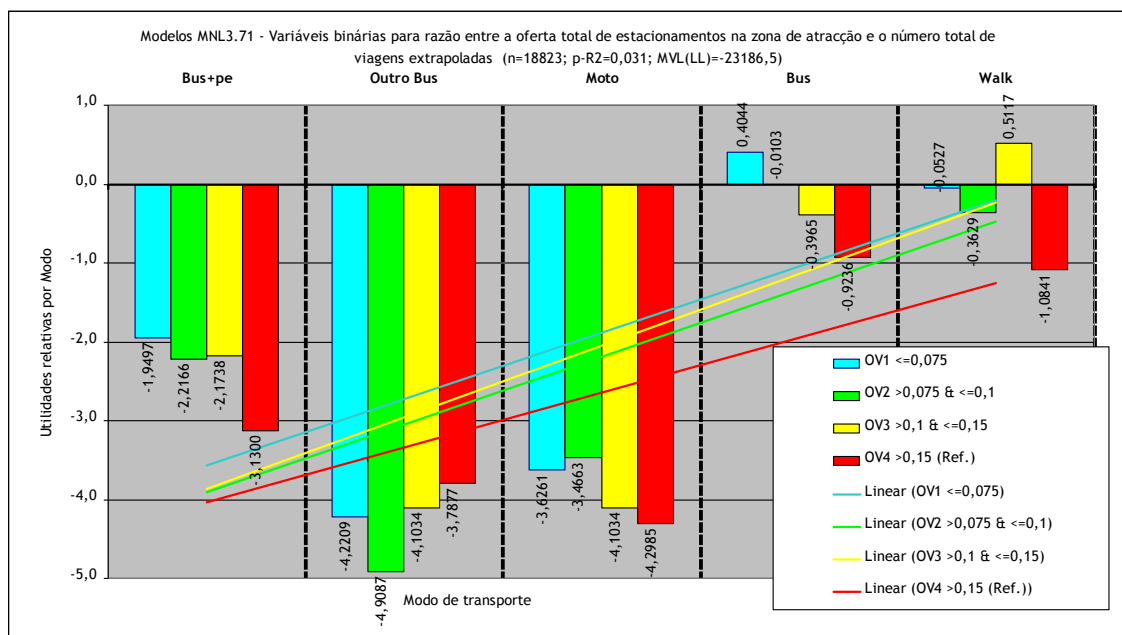


Figura V. 18 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão oferta de lugares versus viagens totais extrapoladas

Os resultados das utilidades relativas revelam que a especificação contínua diminui consideravelmente o ajustamento global do modelo. Esta perda parece dever-se à não linearidade da influência desta variável nos diferentes modos, mais acentuada nos modos pedonal e outros BUS. Note-se, no entanto, que algum deste comportamento diferenciado associado a alguns dos modos poderá ser influência cruzada de outras variáveis explicativas, que uma análise onde apenas se aplica uma variável explicativa, não pode avaliar.

É, ainda assim, possível verificar que além da não linearidade das utilidades das diferentes categorias para alguns modos, que a influência das mesmas varia de modo para modo, como se pode constatar pelas diferenças nos gráficos de ambas as especificações.

Assim, os resultados sugerem que a especificação linear para esta variável não é a melhor forma de introduzir esta informação no modelo, para permitir simular o impacto diferenciado de cada categoria nos diferentes modos alternativos ao automóvel. A utilização da especificação contínua só deverá ser ponderada caso se verifique que a sua combinação com outras variáveis tem resultados semelhantes à combinação com as especificações binárias, ou os seus parâmetros se revelem mais coerentes com a teoria comportamental.

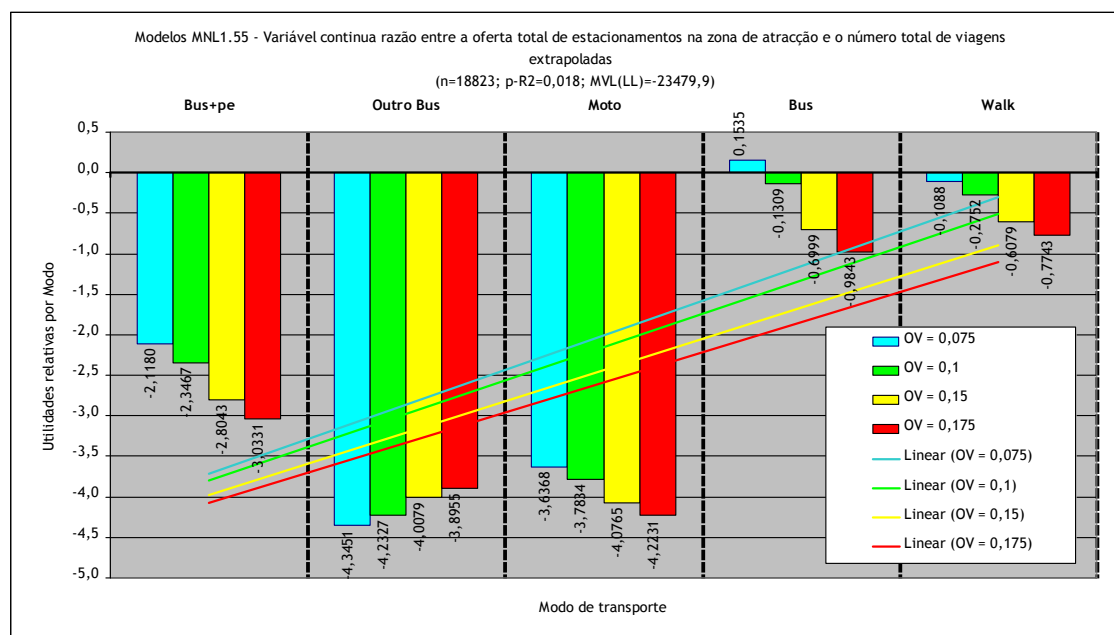


Figura V. 19 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a razão oferta de lugares versus viagens totais extrapoladas

6.2. Rácio “Oferta de Estacionamento Paga / Total”

6.2.1. Análise Comportamental de Diferentes Especificações

Outra variável sobre políticas de estacionamento analisada foi a *razão entre lugares pagos e o total de lugares disponíveis na zona de atracção*, ou seja, a percentagem de estacionamentos pagos. Esta variável também apresenta características contínuas, tendo-se também estudado uma especificação não linear com 3 variáveis binárias considerando valores que definissem amostras semelhantes. Para a análise da possível influência da proporção de estacionamentos pagos na zona de atracção nas escolhas modais realizaram-se modelos MNL utilizando as especificações contínua e combinação de variáveis binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.30.

Na especificação com combinação de variáveis binárias normalizou-se a categoria mais baixa pois corresponde a uma proporção de estacionamentos pagos menor, o que deverá potenciar o uso do automóvel.

Quadro V. 30 - Resultados dos modelos preliminares da influência da proporção de estacionamentos pagos na zona de atracção

Modelo	MNL1.54		MNL3.70	
Modo	B	W	B	W
PO	1.4860	0,8094		
PO1			-	-
PO2			0,8137	0,8340
PO3			1,2062	0,8534
Amostra	18823		18823	
p-R ²	0,016		0,019	
MVL (LL)	-23546,6		-23476,1	
* - p-valor >0,05 & <=0,1		** - p-valor >0,1 & <=0,25		*** - p-valor >0,25

Dos resultados é possível verificar que o modelo com combinação de variáveis binárias apresenta um ajustamento ligeiramente superior ao modelo com especificação contínua. Ambos os modelos apresentam ajustamentos globais reduzidos. No modelo com especificação contínua o sinal positivo para todos os modos alternativos ao automóvel, confirma a tendência expectável para uma maior utilidade dos mesmos quando a proporção de estacionamentos pagos aumenta.

Da análise dos resultados da especificação com combinação de binárias é possível constatar que, com exceção do modo Outros Bus, os parâmetros das diferentes alternativas apresentavam o sinal positivo esperado e um comportamento crescente (exceto o modo MOTO e Outro Bus) com o aumento da proporção de lugares pagos. Este comportamento é coerente com a teoria comportamental de que o aumento do estacionamento pago potencia o uso das alternativas ao automóvel.

Quadro V. 31 - Resultados do teste estatístico da razão das MVL com modelo com especificação contínua como base

	MNL1.50	MNL3.70
Amostra	18823	18823
PO	10 Parâmetros	
POi	15 Parâmetros	
LL _{base}	-23546,6	-23546,6
LL	-23546,6	-23476,1
-2LL		141
Graus de liberdade		5
Chi ² (α=0,05)		11,070

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.31. O valor crítico de Chi-quadrado é cerca de 13 vezes inferior ao teste estatístico -2LL, concluindo-se que o modelo com especificação como combinação de variáveis binárias aparentemente é estatisticamente mais próximo de zero que o modelo com especificação contínua, e portanto melhor.

6.2.2. Avaliação de Comportamentos Não Lineares

Para avaliar a validade da especificação contínua realizou-se um estudo do comportamento das utilidades relativas das alternativas para ambas as especificações. Os resultados da sua aplicação a modelos MNL são apresentados nas figuras V.20 e V.21.

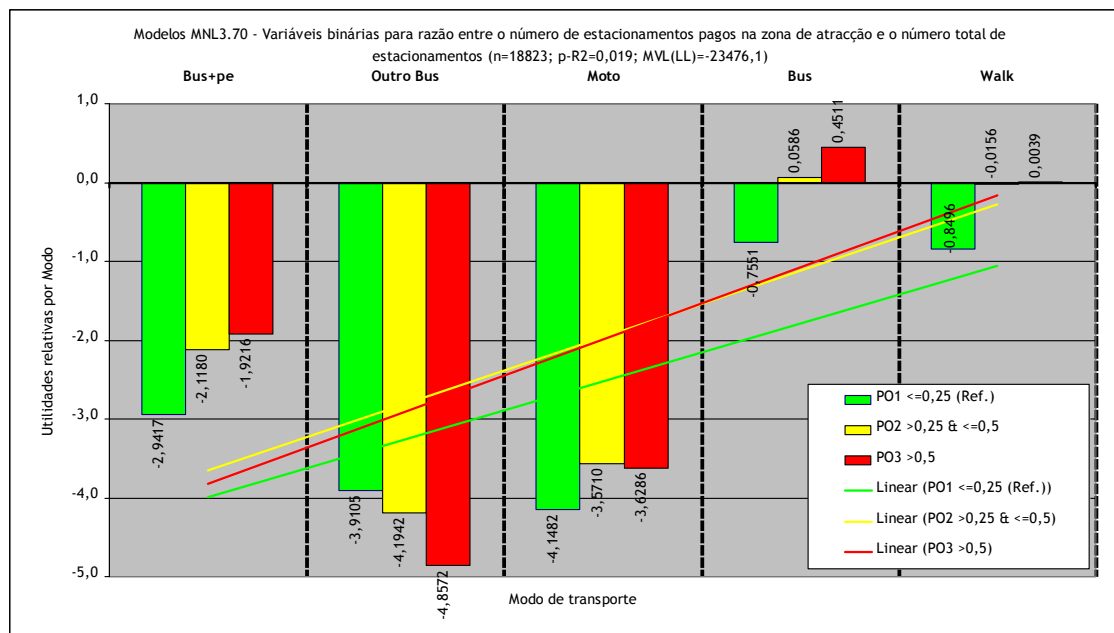


Figura V. 20 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão oferta de lugares pagos versus oferta total de estacionamentos na zona de atracção

A especificação com combinação de binárias permite verificar que apesar de declividades muito diferentes principalmente para a categoria mais baixa $PO \leq 0,25$, as diferentes categorias apresentam comportamento quase linear com alguma coerência nos diferentes modos. Com exceção do modo Outros Bus, todos os outros modos aumentam a sua utilidade relativa com o aumento da percentagem de lugares pagos e portanto menos atrativo para o uso do automóvel. É possível constatar nas utilidades relativas que a especificação contínua perde alguma capacidade explicativa. A especificação contínua apresenta resultados similares aos da especificação não linear com exceção para os modos BUS e Walk, onde o crescimento da proporção de lugares pagos tem um impacto mais significativo.

Estes resultados podem dever-se ao facto de a proporção de estacionamentos pagos aumentar nas áreas mais centrais, que coincide com as zonas onde os modos Bus e Walk têm maior utilização, quer pela proximidade de zonas de geração e atracção, quer devido a uma rede de serviços mais densa. A especificação contínua apresenta assim, algum potencial para poder representar a influência desta variável, mas a especificação não linear parece revelar o maior impacto da mesma nas alternativas mais relevantes (BUS e WALK). Uma opção definitiva por uma das especificações deverá avaliar a sua relação ajustamento *versus* complexidade

adicional introduzida do modelo, bem como a coerência dos seus sinais e o comportamento dos seus parâmetros devido a eventuais possíveis efeitos cruzados de outras variáveis.

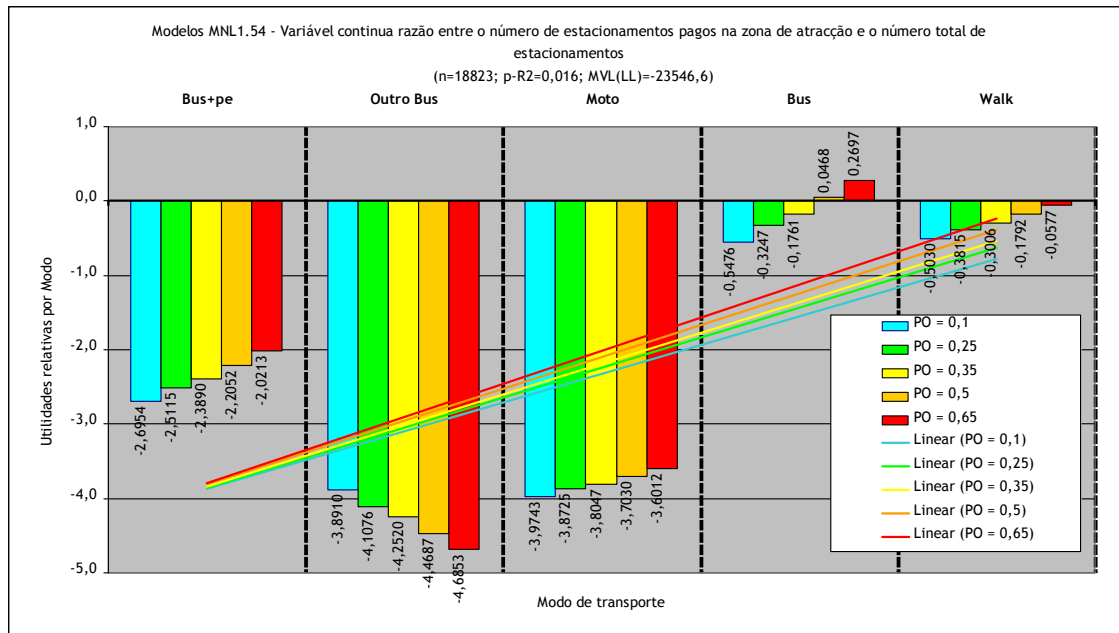


Figura V. 21 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a razão oferta de lugares pagos versus oferta total de estacionamentos na zona de atracção

7. Análise Isolada de Variáveis Relacionadas com o Serviço de Transportes Coletivos

7.1. Rácio “Tempo Viagem BUS / Auto”

7.1.1. Análise Comportamental de Diferentes Especificações

A variável *razão entre a duração média da viagem por Bus e por automóvel particular* apresenta características para ser especificada como uma variável contínua. Para avaliar o seu eventual comportamento linear, definiram-se quatro variáveis binárias atendendo a proporções que se consideraram relevantes e que apresentassem amostras com dimensão razoável. Para a análise da sua possível influência nas escolhas modais realizaram-se modelos MNL utilizando as especificações contínua e combinação de binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.32.

Na especificação com combinação de variáveis binárias normalizou-se a categoria mais elevada pois corresponde a uma situação em que a duração média da viagem de Bus é superior ao dobro do tempo médio da mesma viagem de automóvel, o que deverá potenciar o uso do automóvel. A redução verificada na amostra deve-se ao facto de existirem pares GA em que não houve viagens de BUS, impossibilitando o cálculo da duração média da viagem por BUS.

Quadro V. 32 - Resultados dos modelos preliminares da influência dos tempos médios de viagem de Bus e de Automóvel

Modelo	MNL1.64		MNL3.91	
Modo	B	W	B	W
TBA	-0,6216	-0,5235		
TB1			1,0652	1,2536
TB2			0,7443	0,3323
TB3			0,6449	-0,0437**
TB4			-	-
Amostra	89305		89305	
p-R ²	0,007		0,008	
MVL (LL)	-114898,5		-114751,4	

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Dos resultados é possível verificar que os modelos apresentam ajustamentos globais muito semelhantes e reduzidos. No modelo com especificação contínua o sinal negativo para as alternativas, exceto Outro Bus e Moto, parece confirmar uma diminuição da utilidade das mesmas quando a razão dos tempos médio de viagem aumenta. Os resultados para os modos Outro Bus e Moto apresentam parâmetros positivos provavelmente pelo facto de as razões maiores se verificarem em zonas periféricas ou rurais, zonas onde a utilização destes modos também costuma ser mais necessária atendendo à falta de alternativas. O parâmetro negativo para o modo pedonal também poderá estar relacionado com estes fatores externos, nomeadamente com o seu maior uso nas zonas urbanas centrais onde esta razão de tempos de viagem costuma apresentar valores menores. De facto, as maiores discrepâncias entre os tempos de viagens de automóvel e de Bus acontecem geralmente em zonas interurbanas ou periurbanas, locais onde o modo pedonal poderá não ter grande expressividade pelas distâncias a percorrer, e onde os modos Moto e Outros Bus são uma alternativa viável devido à natural redução do serviço de Bus regular. Esta possibilidade sugere a necessidade de uma análise mais aprofundada de possíveis efeitos cruzados através modelos com combinação de outras variáveis.

Da análise dos resultados da especificação com combinação de binárias é possível constatar que seguem as tendências referidas anteriormente. Com exceção do modo Outros Bus e Moto, os parâmetros das diferentes alternativas apresentam o sinal positivo esperado e comportamento decrescente com o aumento da razão entre tempos médio de viagem. Este comportamento é coerente com a teoria comportamental de que o aumento do tempo de viagem de Bus relativamente ao automóvel potencia o uso deste. Os restantes modos apresentam sinal positivo e comportamento decrescente com o aumento da razão, indiciando que à medida que o automóvel é mais rápido que o Bus, que os mesmos vão perdendo utilidade relativamente ao automóvel.

O teste estatístico da razão das MVL (-2LL) comparando ambas as especificações é apresentado no quadro V.33.

Quadro V. 33 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.64	MNL3.91
Amostra	89305	89305
TBA	10 Parâmetros	
TBi	20 Parâmetros	
LL _{base}	-114898,5	-114898,5
LL	-114898,5	-114751,4
-2LL		294,2
Graus de liberdade		10
Chi ² ($\alpha=0,05$)		18,3

O valor crítico de Chi-quadrado é cerca de 16 vezes inferior ao teste estatístico -2LL, concluindo-se que o modelo com especificação como combinação de variáveis binárias é estatisticamente mais próximo de zero que o modelo com especificação contínua, e portanto melhor.

7.1.2. Avaliação de Comportamentos Não Lineares

A análise do possível comportamento linear desta variável foi realizada através da comparação qualitativa das utilidades relativas das alternativas para ambas as especificações. Os seus resultados são apresentados nas figuras V.22 e V.23.

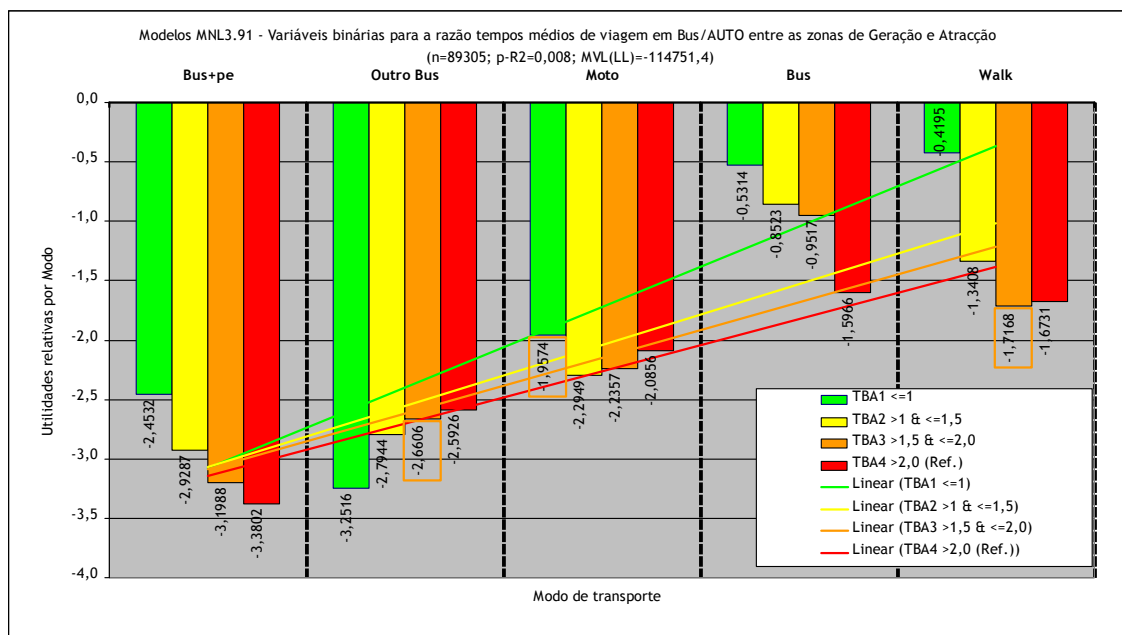


Figura V. 22 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a razão entre a duração média da viagem por BUS e AUTO entre as zonas de GA

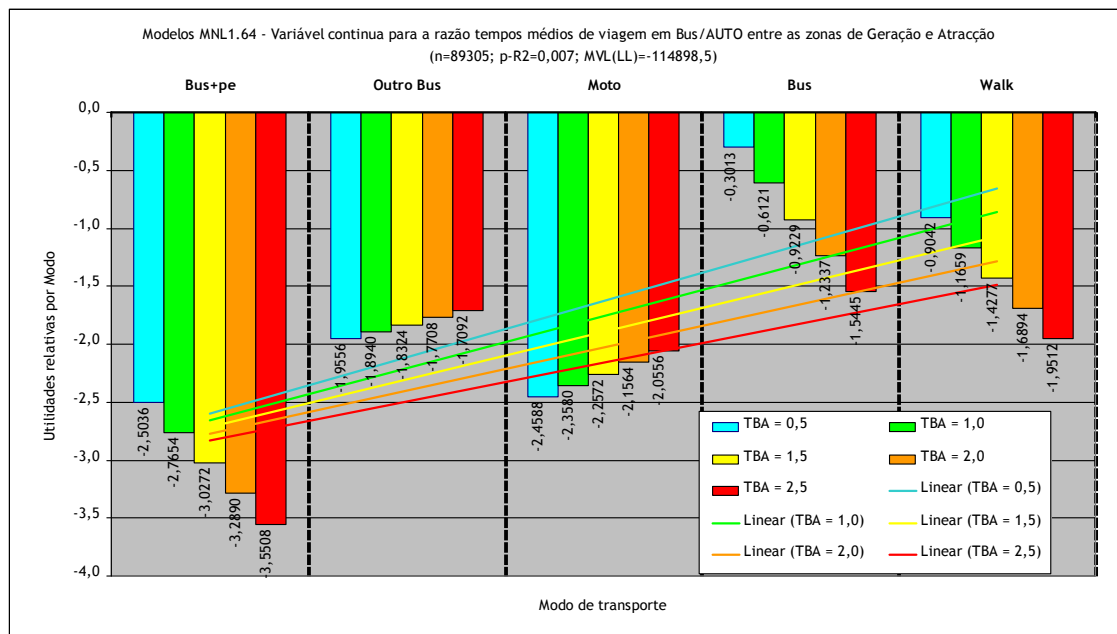


Figura V. 23 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a razão entre a duração média da viagem por BUS e AUTO entre as zonas de GA

Os resultados das utilidades relativas revelam que o modelo com especificação contínua perde alguma informação, possivelmente devida à pior representação da sua influência nos modos pedonal e Moto. Com exceção da categoria inferior ($TBA < 1$), os declives para os diferentes modos são semelhantes e o comportamento das variáveis tem tendências lineares para a maioria dos modos.

Quadro V. 34 - Resultados do teste de *Wald* para o modelo MNL3.91 para avaliar a linearidade dos parâmetros da combinação de binárias para razão de tempos de viagem BUS/AUTO

Modo	Teste de <i>Wald</i> Chi-quadrado	Valor crítico para $\alpha = 0,05$ e 2 GL	Comentário
BW	2,91	5,99	Hipótese nula dos parâmetros serem lineares não pode ser rejeitada
OB	4,42		Hipótese nula dos parâmetros serem lineares não pode ser rejeitada
M	18,28		Hipótese nula dos parâmetros serem lineares rejeitada, mas por pouco
B	17,40		Hipótese nula dos parâmetros serem lineares rejeitada, mas por pouco
W	94,29		Hipótese nula dos parâmetros serem lineares rejeitada

Para avaliar estatisticamente a linearidade entre parâmetros para as alternativas realizou-se o teste de *Wald* no modelo com especificação não linear (MNL3.91Wald_BP,

MNL3.91Wald_OB, MNL3.91Wald_M, MNL3.91Wald_B e MNL3.91Wald_W) com uma restrição linear para que a diferença entre parâmetros fossem iguais entre sim (e portanto lineares). Os resultados do teste são apresentados no quadro V.34.

A análise dos resultados do teste de *Wald* corrobora a análise qualitativa, revelando que estatisticamente os parâmetros para os modos BW e OB apresentam um comportamento linear, e que, apesar de não ter significância estatística, os modos Moto e Bus também apresentam valores próximos de um comportamento linear. Apenas para o modo Walk o teste rejeita a hipótese de a diferença entre parâmetros seja igual.

Assim, apesar das pequenas discrepâncias descritas, a especificação contínua apresenta potencial como alternativa aparentemente viável à combinação de variáveis binárias.

7.2. Frequências Horárias Médias dos Serviços BUS

7.2.1. Análise Comportamental de Diferentes Especificações

Na especificação contínua desta variável foram testadas várias possibilidades relacionadas com frequências médias para diferentes períodos do dia (representadas pelas frequências equivalentes horárias). As variáveis que melhores resultados apresentaram foram respetivamente a frequência média relativa ao período de ponta do tráfego da manhã e a frequência média diária (período laboral). Apesar de apresentar resultados aparentemente ligeiramente melhores, optou-se por não utilizar a variável relacionada com o período de ponta de tráfego da manhã, por se considerar que a frequência média diária seria um indicador mais fidedigno do serviço disponibilizado aos cidadãos. Esta assunção baseou-se no facto de se considerar que os cidadãos valorizam mais serviços distribuídos ao longo do período laboral que podem utilizar em caso de necessidade, relativamente a serviços concentrados apenas em algum(ns) período(s) específico(s) do dia que podem condicionar a viagem de retorno.

A variável *frequência média horária diária de TC potencialmente à disposição entre as 7h30 e as 19h30*, devida à sua forma de cálculo apresenta características contínuas. Atendendo às gamas de frequências geralmente existentes em meios urbanos, foram definidas quatro variáveis binárias de modo a avaliar o comportamento dos parâmetros desta variável.

Para a análise da sua possível influência realizaram-se modelos MNL utilizando as especificações contínua e combinação de binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.35.

Na especificação com combinação de binárias normalizou-se a categoria mais baixa pois corresponde a uma situação em que as frequências médias diárias são muito baixas (tempos médios de espera superiores a 30 minutos) sendo pior alternativa ao uso do automóvel.

Dos resultados é possível verificar que ambos os modelos apresentam ajustamentos globais reduzidos e muito similares.

Quadro V. 35 - Resultados dos modelos preliminares da influência da frequência média horária diária dos BUS

Modelo	MNL1.78		MNL3.137	
Modo	B	W	B	W
Ftl1	0,0458	0,0436		
Fl1			-	-
Fl2			0,6883	0,0641**
Fl3			0,6838	-0,0997
Fl4			1,1924	0,8076
Amostra	41329		41329	
p-R ²	0,021		0,023	
MVL (LL)	-51066,9		-50952,4	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25				

No modelo com especificação contínua o sinal positivo para todos os modos exceto Outro Bus e Moto, confirma um aumento da utilidade dos mesmos quando as frequências médias dos Bus aumentam. Os resultados para os modos Outro Bus e Moto, com parâmetros negativos que sugerem que esta variável tem impacto negativo nos mesmos, bem como o parâmetro positivo para o modo pedonal, poderão estar relacionados, mais uma vez, com os locais onde as frequências potenciais apresentam valores maiores. As mesmas localizam-se nas zonas centrais do Porto, onde a rede de transportes coletivos é mais densa e múltiplos trajetos se entrecruzam e onde a distância entre as zonas GA eventualmente não permite que os outros modos sejam competitivos. Esta possibilidade sugere mais uma vez a necessidade de análises mais aprofundadas destes resultados com modelos com combinação de outras variáveis.

Da análise dos resultados da especificação com combinação de binárias, é possível constatar que seguem as tendências referidas anteriormente. Com exceção do modo Outros Bus e Moto (com comportamento inverso), os parâmetros das alternativas que incluem BUS apresentam o sinal positivo esperado e comportamento crescente com o aumento da frequência média horária diária. Este comportamento é coerente com o senso comum de que o aumento da frequência dos serviços de transportes coletivos diminui a atratividade das restantes alternativas que não incluem o BUS. É também possível verificar que a influência de ter frequências médias horárias de 2 a 4 TC/h e de 4 a 12 TC/h é muito similar, que eventualmente não justifica a sua subdivisão. Os resultados para o modo pedonal sugerem um comportamento similar aos parâmetros para as alternativas com BUS, mas mais acentuado, com uma valorização apenas das frequências médias muito elevadas. Estes resultados podem, mais uma vez, dever-se mais à proximidade entre zonas de GA nas áreas urbanas do que às frequências potenciais muito elevadas (efetivamente existentes nessas zonas).

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.36.

Quadro V. 36 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.78	MNL3.137
Amostra	41329	41329
Ftl	10 Parâmetros	
Fli	20 Parâmetros	
LL _{base}	-51066,9	-51066,9
LL	-51066,9	-50952,4
-2LL		229
Graus de liberdade		10
Chi ² ($\alpha=0,05$)		18,3

O valor crítico de Chi-quadrado de 18,3, é 12 vezes inferior ao teste estatístico -2LL, concluindo-se que o modelo com especificação como combinação de binárias é estatisticamente mais próximo de zero que o modelo com especificação contínua, e portanto melhor.

7.2.2. Avaliação de Comportamentos Não Lineares

A avaliação da linearidade de comportamento dos parâmetros da variável em estudo foi realizada através da análise qualitativa das utilidades relativas para ambas as especificações. Os resultados destas análises apresentam-se nas figuras V.24 e V.25.

Os resultados das utilidades parciais sugerem que as categorias têm uma tendência aproximadamente linear coerente, revelando um aumento da utilidade para os modos Bus e combinação Bus+Walk com o aumento da frequência média horária diária de Bus. Apesar da ligeira perda de ajustamento, a especificação contínua representa o comportamento das diferentes categorias de forma similar à combinação de variáveis binárias, com maiores discrepâncias verificadas no modo Walk. As divergências nos declives entre modos nas duas especificações não sugerem grandes diferenças na influência das categorias, verificando-se as maiores discrepâncias para a categoria (Ftl1>4 & ≤12 TC/h).

Para avaliar estatisticamente a linearidade entre parâmetros para as alternativas realizou-se o teste de *Wald* ao modelo com especificação não linear (MNL3.137Wald_BP, MNL3.137Wald_OB, MNL3.137Wald_M, MNL3.137Wald_B e MNL3.137Wald_W) com uma restrição linear para que a diferença entre parâmetros fossem iguais entre sim (e portanto lineares). Os resultados do teste são apresentados no quadro V.37.

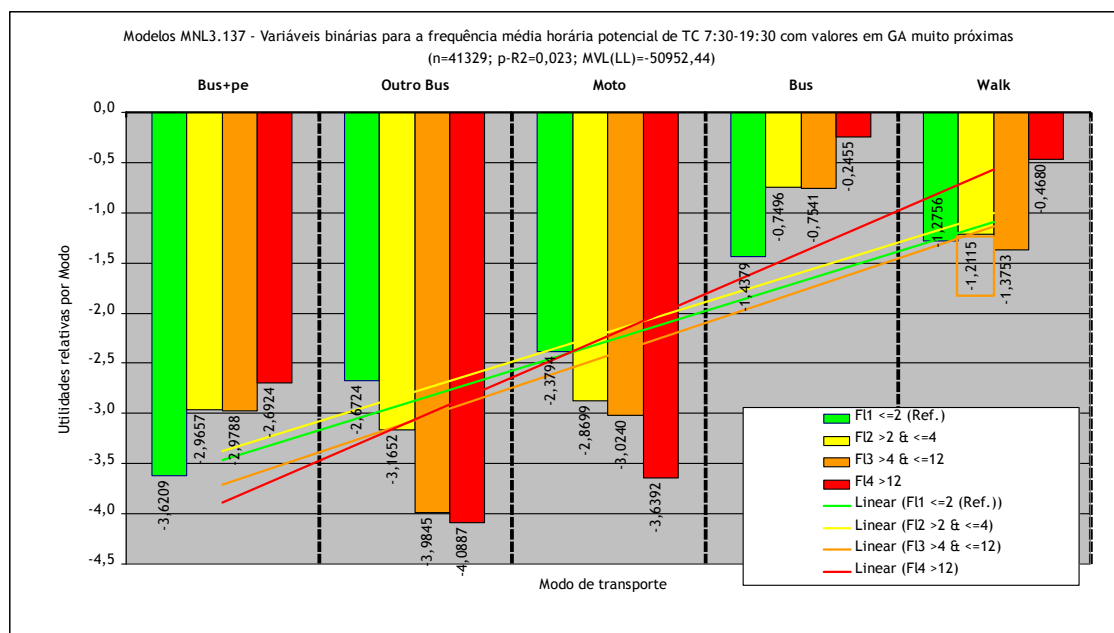


Figura V. 24 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a frequência média horária de TC potencialmente à disposição entre as 7h30 e as 19h30

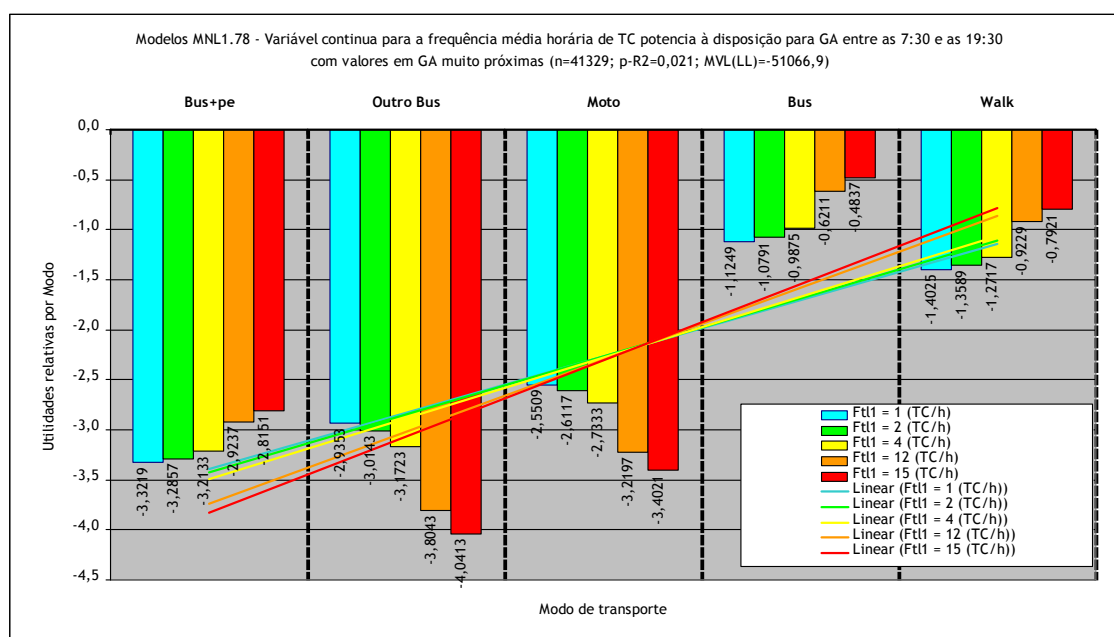


Figura V. 25 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a frequência média horária de TC potencialmente à disposição entre as 7h30 e as 19h30

Quadro V. 37 - Resultados do teste de *Wald* para o modelo MNL3.137 para avaliar a linearidade dos parâmetros da combinação de binárias para a frequência média diária

Modo	Teste de <i>Wald</i> Chi-quadrado	Valor crítico para $\alpha = 0,05$ e 2 GL	Comentário
BW	4,22	5,99	Hipótese nula dos parâmetros serem lineares não pode ser rejeitada
OB	9,65		Hipótese nula dos parâmetros serem lineares rejeitada, mas por pouco
M	7,98		Hipótese nula dos parâmetros serem lineares rejeitada, mas por pouco
B	82,3		Hipótese nula dos parâmetros serem lineares rejeitada
W	244,6		Hipótese nula dos parâmetros serem lineares rejeitada

A análise dos resultados do teste de *Wald* corrobora a análise qualitativa, revelando que estatisticamente os parâmetros para o modo Bus+Walk apresentam um comportamento linear, e que, apesar de estatisticamente não ser significativo, para os modos Moto e Outros Bus também apresentam valores próximos de um comportamento linear. Apenas para os modos Walk e Bus o teste rejeita a hipótese de a diferença entre parâmetros ser igual.

Atendendo aos resultados verificados, a especificação com combinação de variáveis binárias é aparentemente melhor, podendo ainda assim assumir-se que a especificação linear tem potencial para poder representar a influência desta variável nos modelos de forma viável.

7.2.3. Outras Especificações da Frequência Média Horária de BUS

Na avaliação da influência da qualidade dos serviços de TC, mais concretamente dos BUS, foram utilizadas quatro variáveis correspondentes à oferta potencial em quatro períodos de tempo distintos: período de ponta de tráfego da manhã (Fm), Meio da manhã (Fa), da tarde (Ft) e a oferta potencial ao longo do período laboral (Ftl). Para analisar qual a melhor especificação realizaram-se modelos com as diferentes variáveis contínuas cujos resultados forma comparados. Os resultados destes modelos são apresentados nos quadros IV.38, V.39, V.40 e V.41.

Nos mesmos quadros é possível verificar a distinção de três tipos de variáveis para cada um dos quatro períodos analisados. Esta distinção baseou-se na forma como se considerou o valor das frequências nos pares GA muito próximos (<2km). Para estas zonas adotou-se para as primeiras variáveis o valor de zero (Fx), assumindo que, apesar de na realidade poderem existir serviços de TC, a proximidade entre zonas os torna desprezáveis pois não são uma alternativa ao modo pedonal. Nas segundas variáveis para essas mesmas zonas (Fx0) considerou-se que não havia um serviço e as células ficaram vazias (diminuindo a amostra

para o modelo). Nas terceiras variáveis (Fx1), para as mesmas zonas consideraram-se todos os potenciais trajetos que as ligavam.

Quadro V. 38 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego da manhã

Modelo	MNL1.56		MNL1.72		MNL1.75	
Modo	B	W	B	W	B	W
Fm	0,0445	0,0225				
Fm0			0,0614	0,0631		
Fm1					0,0614	0,0556
Amostra	41329		36041		41329	
p-R ²	0,0116		0,0248		0,0227	
MVL (LL)	-51556,7		-43123,3		-50975,0	

* - p-valor >0,05 & ≤0,1 ** - p-valor >0,1 & ≤0,25 *** - p-valor >0,25

Quadro V. 39 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego do Meio da manhã

Modelo	MNL1.57		MNL1.73		MNL1.76	
Modo	B	W	B	W	B	W
Fa	0,0509	0,0298				
Fa0			0,0690	0,0759		
Fa1					0,0707	0,0694
Amostra	41329		36041		41329	
p-R ²	0,0105		0,0226		0,0213	
MVL (LL)	-51613,8		-43222,8		-51049,5	

* - p-valor >0,05 & ≤0,1 ** - p-valor >0,1 & ≤0,25 *** - p-valor >0,25

Quadro V. 40 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período de ponta de tráfego da tarde

Modelo	MNL1.58		MNL1.74		MNL1.77	
Modo	B	W	B	W	B	W
Ft	0,0404	0,0178				
Ft0			0,0561	0,0572		
Ft1					0,0565	0,0503
Amostra	41329		36401		41329	
p-R ²	0,0100		0,0220		0,0201	
MVL (LL)	-51636,9		-43248,9		-51109,5	

* - p-valor >0,05 & ≤0,1 ** - p-valor >0,1 & ≤0,25 *** - p-valor >0,25

Quadro V. 41 - Resultados dos modelos preliminares comparativos entre as diferentes frequências médias horárias de BUS no período laboral

Modelo	MNL1.59		MNL1.74a		MNL1.78	
Modo	B	W	B	W	B	W
Ftl	0,0326	0,0161				
Ftl0			0,0453	0,0487		
Ftl1					0,0458	0,0436
Amostra	41329		36041		41329	
p-R ²	0,0101		0,0225		0,0210	
MVL (LL)	-51633,1		-43225,6		-51066,9	

* - p-valor >0,05 & ≤0,1 ** - p-valor >0,1 & ≤0,25 *** - p-valor >0,25

Dos resultados pode verificar-se que a terceira especificação (Fx1) com valores para todas as zonas independentemente da sua proximidade, permitiram melhores ajustamentos que as primeiras e parâmetros muito semelhantes às segundas mas com amostras maiores. Atendendo a estes resultados considerou-se que a terceira especificação era a mais promissora.

Comparando os resultados dos diferentes períodos de análise, as frequências da manhã e do período do Meio da manhã apresentam um ajustamento ligeiramente melhor que as do período laboral. No entanto, e dada a pequena diferença verificada, considerou-se a frequência média diária com maior potencial para ser utilizada, uma vez que permite atenuar algumas deficiências resultantes dos problemas de levantamento e tratamento de dados referidos no Capítulo IV. Os parâmetros, apesar de alterações de grandeza, mantêm as tendências para todas as tipologias testadas.

7.3. Velocidades Comerciais Equivalentes dos Serviços BUS

7.3.1. Análise Comportamental de Diferentes Especificações

A última variável analisada foi a *velocidade comercial equivalente por BUS*, cuja formulação a caracteriza como uma variável contínua. Para testar a sua linearidade foram definidas três variáveis binárias com valores correspondentes a velocidades operacionais normalmente verificadas em áreas urbanas (Seco e Gonçalves, 2006). Para a análise da sua possível influência nas escolhas modais, desenvolveram-se modelos MNL utilizando as especificações contínua e combinação de binárias. Os resultados desses modelos para as duas alternativas mais importantes (Bus e Walk) são apresentados no quadro V.42.

Quadro V. 42 - Resultados dos modelos Preliminares da influência da velocidade comercial equivalente do BUS

Modelo	MNL1.63		MNL3.90a	
Modo	B	W	B	W
V2	-0,03223	-0,8927		
V21			0,5274	5,3171
V22			0,2968	1,0403
V23			-	-
Amostra	89305		89305	
p-R ²	0,175		0,076	
MVL (LL)	-95406,9		-106839,0	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

O ajustamento do modelo com especificação contínua é significativo e equivalente em grandeza ao do modelo com distâncias mais curtas. Verifica-se que os modos Moto e Outros Bus têm um sinal positivo e o modo Bus+Walk não apresenta significância estatística para a influência desta variável. Nesta especificação seria também, à partida, expectável que o parâmetro para o modo Bus apresentasse um sinal positivo, revelando que o aumento da sua velocidade comercial (ainda que equivalente) valorizava a sua utilidade. No entanto,

constatou-se o inverso, sugerindo um decréscimo de utilidade com o efetivo aumento da velocidade comercial do BUS.

Do mesmo modo, na especificação com combinação de variáveis binárias, numa primeira abordagem tender-se-ia a esperar que a categoria com velocidades comerciais equivalentes mais baixas corresponderia à situação mais favorável para o uso do automóvel. No entanto, constatou-se que na base de dados a maior utilização do automóvel corresponde efetivamente à categoria de maior velocidade comercial do BUS (V23: $V2 > 20\text{km/h}$), que normalmente deve corresponder a zonas interurbanas.

Esta aparente discrepância, tornada mais relevante pela acentuada diferença entre os valores do $p\text{-}R^2$, muito significativo para a especificação contínua e pouco relevante para a combinação de binárias, sugere que deverá haver fatores explicativos alternativos cujo efeito se sobrepõe decisivamente ao efeito positivo do aumento da velocidade comercial. Estes fatores poderão dever-se à forma como a variável foi especificada, através de conjugação de outras duas variáveis já de si relevantes, ou devido a uma correlação da velocidade em relação à localização relativamente à área urbana. A variável pode estar a representar tipos de serviços Bus com características operacionais muito diferentes. Geralmente, em zonas urbanas consolidadas com boas redes de transportes coletivos há uma grande repartição modal para os mesmos mas as velocidades de operação máximas atingidas não ultrapassam os 20km/h , sendo usuais valores menores que 10km/h (Seco e Gonçalves, 2006). Nas zonas periféricas e interurbanas, com menos restrições operacionais devido às distâncias entre paragens ou aos congestionamentos de tráfego, as velocidades operacionais aumentam substancialmente para valores acima dos 20km/h , mas a repartição modal dos BUS tende a diminuir significativamente uma vez que o seu nível de competitividade face ao automóvel também diminui devido a múltiplos outros fatores.

Atendendo à normalização realizada, os valores positivos e decrescentes com o aumento da velocidade comercial para o Bus e Walk, e os resultados com pouca significância estatística para os restantes modos, são passíveis de ser explicados por esta realidade.

O teste estatístico da razão das MVL (-2LL) é apresentado no quadro V.43.

O valor crítico de Chi-quadrado de 11,1, é 2080 vezes superior ao teste estatístico -2LL, podendo concluir-se que o modelo com especificação contínua é estatisticamente mais próximo de zero que o modelo com especificação como combinação de variáveis binárias, e portanto potencialmente melhor.

Quadro V. 43 - Resultados do teste estatístico da razão das MVL com modelo com especificação discreta como base

	MNL1.63	MNL3.90a
Amostra	89305	89305
V2	10 Parâmetros	
V2i		15 Parâmetros
LL _{base}	-95406,9	-95406,9
LL	-95406,9	-106839,0
-2LL		-22864,2
Graus de liberdade		5
Chi ² ($\alpha=0,05$)		11,1

Atendendo a estes resultados, à diferença de ajustamentos entre as especificações lineares e não lineares e à forma de cálculo da variável velocidade comercial equivalente media, analisar-se-á mais adiante se os mesmos se podem dever a outros fatores, nomeadamente à proximidade entre zonas de GA nas áreas urbanas, e portanto à variável distância mais curta entre centróides. Esta constatação poderá sugerir que a velocidade comercial, tal como foi determinada, não é a melhor forma de introduzir mais informação no modelo.

7.3.2. Avaliação de Comportamentos Não Lineares

A análise da linearidade dos parâmetros desta variável foi novamente realizada através da comparação das utilidades relativas para ambas as especificações. Os resultados destas análises apresentam-se nas figuras V.26 e V.27.

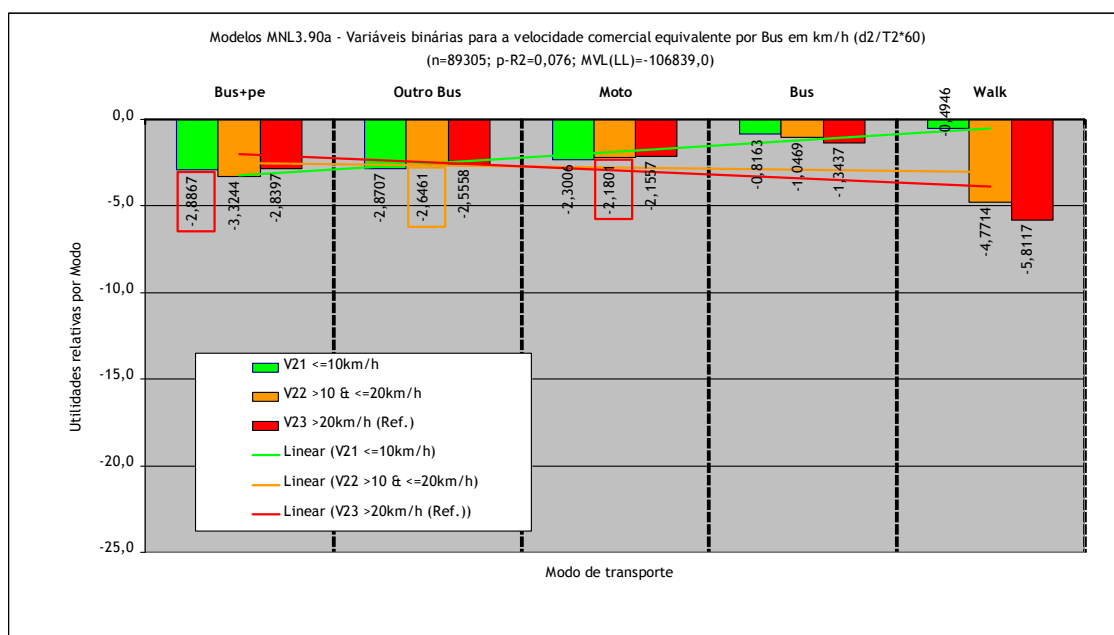


Figura V. 26 - Utilidades parciais para os diferentes modos de transporte considerando variáveis binárias para a velocidade comercial equivalente para BUS em km/h

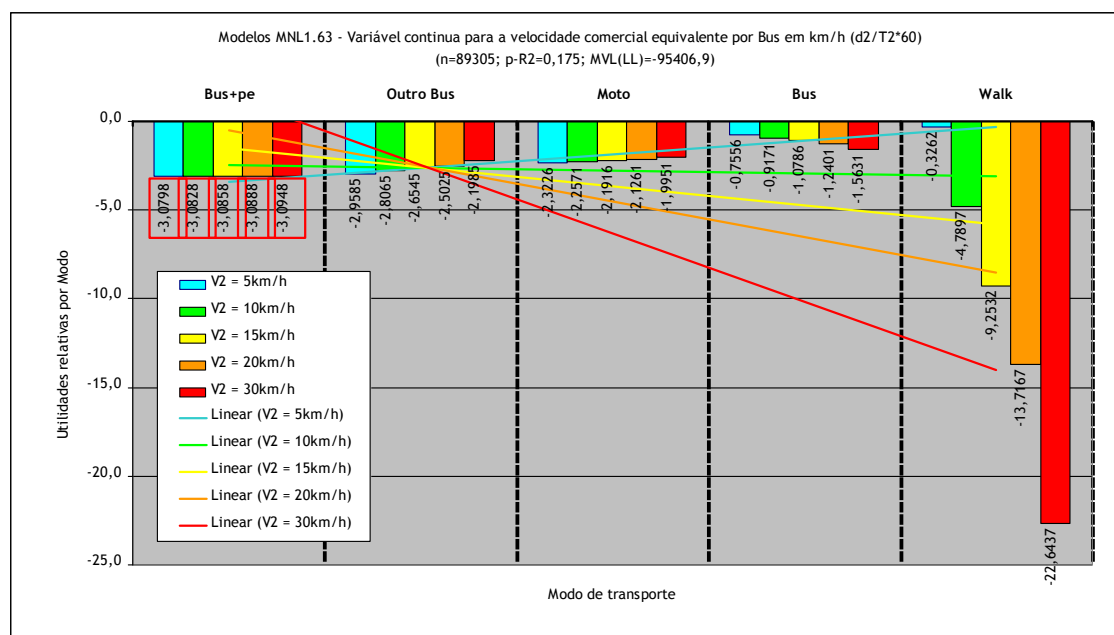


Figura V. 27 - Utilidades parciais para os diferentes modos de transporte considerando uma variável contínua para a velocidade comercial equivalente para BUS em km/h

Os resultados das utilidades relativas corroboraram as tendências contrárias ao esperado, verificados anteriormente. O modo Walk apresenta resultados de uma maior influência da variável na especificação contínua que na não linear. Para os restantes modos a especificação não linear apresenta resultados muito semelhantes à linear, que sugerem que a variável tem uma influência predominantemente linear. Da análise qualitativa dos resultados sugere-se a especificação contínua para representar esta variável, com uma nota de atenção para avaliar a sua possível colinearidade com outras variáveis que possam justificar o aumento anormal do ajustamento do modelo face à especificação não linear.

8. Síntese das Análises Isoladas

No quadro V.44 apresenta-se um resumo dos resultados das análises anteriores onde é possível avaliar quais as variáveis mais relevantes, justificando a sua inclusão no modelo de base. É também possível comparar os ajustamentos obtidos com diferentes especificações e entre diferentes variáveis.

Dessa comparação é possível verificar que as variáveis com grande potencial explicativo, a considerar num modelo base, serão a existência de licença de condução (Lc), o número de automóveis disponíveis diariamente no agregado *per capita* (NCi) e a distância entre centróides das zonas de GA de viagens (D2 e/ou Dxi). A variável velocidade comercial equivalente por Bus (V2), apesar de apresentar um ajustamento elevado, revela uma discrepância de ajustamento muito acentuada com a sua especificação com combinação de binárias. Este resultado indicia algum problema que justifica ser alvo de mais análises antes

de ser considerada como variável base. Sugere-se uma verificação se a sua formulação (combinação de outras duas variáveis relevantes) não estará a replicar informação já introduzida pelas suas variáveis primárias.

Quadro V. 44 - Quadro síntese dos resultados dos modelos preliminares com as diferentes especificações de cada variável analisada

Modelo	Esp. Discreta	Esp. Continua	Comb. Binárias	Amostra	p-R ²	LL
1.28	R			95426	0,029	-119964,7
3.11			R1	95426	0,032	-119596,3
1.29			Sx	95426	0,012	-121950,6
1.65	ldb			95426	0,006	-122753,5
3.16			ldi	95426	0,046	-117695,1
1.81	ln			95426	0,013	-121836,3
3.108			lni	95426	0,021	-120869,4
1.26		NC		95426	0,100	-111107,3
3.29			NCi	95426	0,104	-110636,3
1.88			Mot	95426	-	-
1.33			Lc	95426	0,098	-111414,3
1.53		D2		95426	0,194	-99549,0
3.83			Dxi	95426	0,194	-99498,8
1.70		Tia (gen)		30559	0,00006	-39684,6
1.80		Til (gen)		30559	0,017	-39013,4
3.89			Tr,Rg,Es,Cs	95426	0,018	-121275,0
1.71		De		61358	0,011	-78298,4
3.120a			Dei	61358	0,014	-78027,4
1.66		nV		95426	0,031	-119661,3
3.136			nVi	95426	0,031	-119695,5
1.54		PO		18823	0,016	-23546,6
3.70			POi	18823	0,019	-23476,1
1.55		OV		18823	0,018	-23479,9
3.71			OVi	18823	0,031	-23186,5
1.64		TBA		89305	0,007	-114898,5
3.91			TBi	89305	0,008	-114751,4
1.78		Ftl1		41329	0,021	-51066,9
3.137			Fli	41329	0,023	-50952,4
1.63		V2		89305	0,175	-95406,9
3.90a			V2i	89305	0,076	-106839,0

9. Avaliação de Questões de Correlação/Complementaridade entre Variáveis

Atendendo aos resultados anteriores e às variáveis mais relevantes de cada vetor de intervenção, procurou-se avaliar se essa relevância se mantinha com a sua combinação potenciando a sua utilização num modelo base a ser incluído nos modelos mais complexos, ou se pelo contrário a sua utilização simultânea perdia relevância ou coerência.

Atendendo ainda aos resultados verificados com os modelos com apenas uma variável, independentemente da sua especificação, estudou-se também a possibilidade de algumas aparentes incoerências com o senso comum poderem ser explicadas e ultrapassadas através da conjugação dessa variável em simultâneo com outra cujo modelo coordenado/cruzado resultasse em valores mais coerentes.

9.1. Correlação entre a taxa de motorização familiar e a existência de licença de condução

As duas variáveis socioeconómicas com aparente maior impacto nas escolhas modais foram o número de automóveis disponíveis no agregado *per capita* (NCi) e a existência de licença de condução. No entanto, a sua utilização em separado não permite o teste de eventuais efeitos de potenciação ou atenuação resultantes da sua utilização combinada. As combinações mais relevantes atendendo às possíveis repercussões na atitude perante o automóvel estão descritas no quadro V.45:

Quadro V. 45 - Possíveis combinações da disponibilidade de automóvel e de licença de condução e possíveis repercussões na atitude perante o automóvel

a) NC0=0	a1) Lc=1 - Existem carros disponíveis no agregado e o indivíduo tem carta de condução: Indivíduo tem possibilidade de fazer a viagem como condutor ou como passageiro voluntariamente;
	a2) Lc =0 - Existem carros disponíveis no agregado mas o indivíduo não tem carta de condução: Indivíduo pode realizar a viagem apenas como passageiro (voluntário ou involuntário);
b) NC0=1	b1) Lc =1 - Não existem carros disponíveis no agregado mas o indivíduo tem carta de condução: Indivíduo tem que utilizar as alternativas mas é um potencial condutor se no futuro vier a ter acesso ao automóvel;
	b2) Lc =0 - Não existem carros disponíveis no agregado e indivíduo não tem carta de condução: Indivíduo é utilizador cativo das alternativas (não tem condições de utilizar o automóvel) ou voluntário (tem condições económicas mas por motivos ideológicos não quer utilizar o automóvel).

Assim, apenas a conjugação de ambas as variáveis poderá permitir esta distinção. Para avaliar esta possibilidade, realizou-se um modelo com a combinação de ambas as variáveis cujos resultados foram comparados com os modelos isolados. Estes resultados são apresentados no quadro V.46.

Quadro V. 46 - Resultados do modelo MNL exploratório combinando o número de automóveis diariamente à disposição do agregado *per capita* e a existência de licença de condução

Modelo	MNL1.33		MNL3.29		MNL3.72	
Modo	B	W	B	W	B	W
Lc	-2,299	-2,120			-1,970	-1,799
NC0			4,251	3,282	3,235	2,367
NC1			2,538	1,394	1,585	0,539
NC2			1,857	0,772	1,121	0,119**
NC3			1,013	-0,177*	0,533	-0,598
NC4a			-	-	-	-
Amostra	95426		95426		95426	
p-R ²	0,0978		0,1041		0,1592 ($\Sigma=0,2019$)	
MVL (LL)	-111414,3		-110636,3		-103828,9	
Restrições	10		25		30	
GL	Base		15		20	
			Base		5	
-2LL	Base		1556,6		15171,4	
			Base		13614,8	
Chi ²	Base		25,0		31,4	
($\alpha=0,05$)			Base		11,1	

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Dos resultados é possível constatar que, apesar da diminuição em grandeza, os sinais e comportamento dos parâmetros do modelo combinado se mantiveram coerentes. O modelo combinado melhora o ajustamento em relação a qualquer um dos modelos isolados, e o teste estatístico -2LL permite concluir que estatisticamente o modelo combinado é melhor que cada um dos modelos isolados. O ajustamento global do modelo melhorou consideravelmente, mas verifica-se que não corresponde à soma dos ajustamentos dos modelos isolados (0,2019). Este resultado sugere que os ajustamentos não são aditivos, ou seja, que parece haver alguma informação sobreposta em cada uma das variáveis incluídas. Ainda assim, em conjunto têm uma capacidade explicativa muito significativa e coerente, considerando-se fazer sentido a sua inclusão combinada como modelo base a usar como referência para avaliação de outros modelos mais complexos.

Para avaliar estatisticamente a correlação entre estas variáveis calculou-se o coeficiente de Pearson entre as especificações binárias de ambas as variáveis. Os resultados são apresentados no quadro V.47.

É possível constatar que os resultados sugerem pouca correlação entre a maioria das variáveis binárias da taxa de motorização familiar e da licença de condução. Apesar de baixa, os coeficientes têm sinal e comportamento coerente, apresentando sinal negativo para agregados sem automóveis (aparentemente não se justifica o investimento em obter licença de condução), e sinal positivo para as restantes categorias. Os resultados sugerem igualmente que para taxas de motorização familiar superiores a 0,5 veíc./hab. a correlação com a licença de condução é baixa mas muito semelhante.

Quadro V. 47 - Resultados da análise de correlação entre as variáveis automóveis diariamente à disposição do agregado *per capita* e licença de condução através do *software* SPSS®

Correlações		Lc (S=1, N=0)
NC = 0	Pearson Correlation	-0,298 ²
	Sig. (2-tailed)	0,000
	N	95426
NC >0 & <=0,25	Pearson Correlation	-0,068 ¹
	Sig. (2-tailed)	0,000
	N	95426
NC >0,25 & <=0,5	Pearson Correlation	0,157 ²
	Sig. (2-tailed)	0,000
	N	95426
NC >0,5 & <=0,75	Pearson Correlation	0,173 ²
	Sig. (2-tailed)	0,000
	N	95426
NC >0,75	Pearson Correlation	0,178 ²
	Sig. (2-tailed)	0,000
	N	95426

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

9.2. Correlação entre o Escalão Etário e o Nível de Instrução

Das análises anteriores foi possível constatar que existiam algumas discrepâncias nos resultados do modelo com incorporação isolada do nível de instrução, nomeadamente para o nível analfabeto. Postulou-se a hipótese de que essas discrepâncias podiam resultar de haver uma grande parte de crianças em idade pré-escolar (e portanto analfabeta ainda) que utilizava o automóvel por serem levadas pelos pais, ficando assim com as suas escolhas modais inteiramente dependentes das escolhas dos mesmos e, portanto, associadas ao correspondente nível de instrução dos mesmos. Esta hipótese, se confirmada, leva a uma distorção dos resultados num modelo que incorpore apenas a variável nível de instrução. Para avaliar esta hipótese realizou-se uma análise de correlações entre as especificações binárias e combinaram-se as variáveis grupo etário e nível de instrução comparando os resultados à luz do senso comum. Esses resultados são apresentados nos quadros V.48 e V.49 respetivamente.

Quadro V. 48 - Resultados da análise de correlação entre as variáveis grupo etário e nível de instrução através do *software* SPSS®

Correlações		In 1 - Analfabeto	In 2 - Básico	In 3 - Secundário	In 4 - Superior
Id 1:0-9 anos	Pearson Correlation	0,613 ⁴	-0,162 ²	-0,114 ²	-0,077 ¹
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	95426	95426	95426	95426
Id 2:10-19 anos	Pearson Correlation	-0,110 ²	0,114 ²	0,029 ¹	-0,138 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	95426	95426	95426	95426
Id 3:20-34 anos	Pearson Correlation	-0,141 ²	-0,154 ²	0,204 ²	0,101 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	95426	95426	95426	95426
Id 4:35-64 anos	Pearson Correlation	-0,137 ²	0,118 ²	0,116 ²	0,072 ¹
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	95426	95426	95426	95426
Id 5:+65 anos	Pearson Correlation	0,093 ¹	0,034 ¹	-0,076 ¹	-0,032 ¹
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	95426	95426	95426	95426

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

Quadro V. 49 - Resultados do modelo MNL exploratório combinando o grupo etário com o nível de instrução

Modelo	MNL3.108		MNL3.16		MNL3.51	
Modo	B	W	B	W	B	W
In1	0,907	0,994			1,368	0,801
In2	1,377	1,113			1,054	0,819
In3	1,069	0,594			0,777	0,420
In4	-	-			-	-
Id1			-	-	-	-
Id2			2,240	1,077	2,493	1,151
Id3			0,519	-0,504	0,907	-0,302
Id4			0,782	-0,041***	1,098	0,076**
Id5			1,879	1,172	2,100	1,234
Amostra	95426		95426		95426	
p-R ²	0,0212		0,0469		0,0643 ($\Sigma=0,0681$)	
MVL (LL)	-120869,4		-117695,1		-115546,4	
Restrições	20		25		40	
GL	Base		5		20	
			Base		15	
-2LL	Base		6348,6		10646	
			Base		4297,4	
Chi ² ($\alpha=0,05$)	Base		11,0		31,4	
			Base		25,0	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

É possível constatar no quadro V.48 que efetivamente existe uma correlação forte entre as crianças entre os 0 e 9 anos e o nível de instrução analfabeto, sendo as restantes correlações fracas ou inexistentes, embora coerentes. O escalão etário +65 anos não apresenta correlação com nenhum dos níveis de instrução. Para o escalão etário dos 10 aos 19 anos aparentemente a amostra inclui mais adolescentes no nível básico que secundário. Os inquiridos da amostra com idades entre 20 e 34 anos apresentam correlações positivas para os níveis de instrução secundário e superior. Os inquiridos mais velhos também apresentam correlações positivas com o nível básico.

Os resultados do quadro V.49 demonstram que o modelo combinado melhorou o ajustamento global, sendo o $p-R^2$ combinado muito próximo do somatório dos $p-R^2$ dos modelos parcelares. Este resultado sugere que a capacidade explicativa que ambas as variáveis introduzem no modelo é essencialmente aditiva. Além disso, os sinais dos parâmetros passaram a ser todos coerentes, com a discrepância para o nível de instrução analfabeto atenuada, passando os parâmetros a ter um comportamento decrescente com o aumento do nível de instrução. O teste estatístico -2LL rejeita a hipótese nula de que o modelo combinado não seja melhor que ambos os isolados.

9.3. Correlação entre as Variáveis sobre Estacionamento

Os resultados dos modelos com as variáveis sobre estacionamento também foram testados para verificar alguma correlação entre as mesmas ou a existência de algum efeito potenciador ou não linear do conjunto. Na análise isolada verificou-se um comportamento geral mais coerente para a variável relacionada com a percentagem de estacionamento pagos, mas uma maior capacidade explicativa da variável que avaliava a densidade de oferta de estacionamento face à procura de estacionamento.

Quadro V. 50 - Resultados da análise de correlação entre as variáveis sobre estacionamento através do software SPSS®

Correlações		OV <=0,075	OV >0,075 & <=0,1	OV >0,1 & <=0,15	OV >0,15
PO <=0,25	Pearson Correlation	-0,553 ⁴	-0,282 ²	0,215 ²	0,793 ⁴
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	18823	18823	18823	18823
PO >0,25 & <=0,5	Pearson Correlation	-0,167 ²	-0,101 ²	-0,101 ³	-0,413 ³
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	18823	18823	18823	18823
PO >0,5	Pearson Correlation	0,683 ⁴	-0,401 ³	-0,099 ¹	-0,325 ³
	Sig. (2-tailed)	0,000	0,000	0,000	0,000
	N	18823	18823	18823	18823

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

A análise de correlações é apresentada no quadro V.50. É possível verificar uma grande correlação positiva entre as zonas com pouca oferta de estacionamento e uma maior percentagem de estacionamentos pagos e entre as zonas com maior oferta de estacionamento e menor percentagem de estacionamento pago. Também se verifica uma grande correlação negativa entre as zonas com poucos estacionamentos e grande percentagem destes pagos. As restantes situações apresentam correlações médias a fracas. Os resultados corroboram a assunção habitual de haver mais estacionamentos pagos onde estes são mais escassos.

Para avaliar os efeitos da sua conjugação, apresentam-se os resultados do modelo com combinação de ambas no quadro V. 51.

Quadro V. 51 - Resultados do modelo MNL exploratório combinando as variáveis sobre estacionamentos

Modelo	MNL3.70		MNL3.71		MNL3.114	
Modo	B	W	B	W	B	W
OV1			1,3280	1,0314	1,1002	-0,2808
OV2			0,9133	0,7212	0,8558	-0,6546
OV3			0,5271	1,5958	0,4731	0,9654
OV4			-	-	-	-
PO1	-	-			-	-
PO2	0,8137	0,8340			0,063***	1,440
PO3	1,2062	0,8535			0,300	1,242
Amostra	18823		18823		18823	
p-R ²	0,0186		0,0308		0,0384 (Σ=0,0494)	
MVL (LL)	-23476,1		-23186,5		-23002,5	
Restrições	15		20		30	
GL	Base		5		15	
			Base		10	
-2LL	Base		579,2		947,2	
			Base		368	
Chi ²	Base		11,1		25,0	
(α=0,05)			Base		18,3	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25						

É possível constatar que o modelo combinado é estatisticamente melhor que qualquer um dos modelos isolados. Em relação aos parâmetros, existem alterações significativas não só em termos de valor absoluto como de sinal.

Os parâmetros das variáveis sobre os estacionamentos pagos alteraram substancialmente de grandeza mas mantiveram a tendência crescente, exceto para o modo pedonal. Os parâmetros dos modos Outro Bus e Moto deixaram de ter significância estatística sugerindo a possibilidade de os mesmos não serem influenciados pelo pagamento de estacionamento.

Os parâmetros das variáveis sobre a densidade de oferta de estacionamento face à procura de estacionamento apresentam uma redução do seu valor absoluto. Em relação ao sinal, com exceção do modo pedonal, as alternativas mantiveram o sinal. Em relação ao

comportamento, o Bus manteve a tendência decrescente, e as restantes alternativas mantiveram o comportamento não linear. Os modos Outro Bus e Moto, tal como para o estacionamento pago, não apresentam significância estatística para a maior parte das categorias.

Em relação ao ajustamento global do modelo combinado verifica-se que o mesmo não indicia uma aditividade completa das variáveis (0,0494), sugerindo, como esperado pelos resultados da análise de correlação, a existência de alguma informação similar nas duas variáveis.

Quadro V. 52 - Resultados dos modelos exploratórios combinando variáveis sobre estacionamentos

Modelo	MNL3.74		MNL3.114		MNL3.76	
Modo	B	W	B	W	B	W
OV	-8,018	-6,191				
PO	0,626	0,098***				
OV1			1,100	-0,281		
OV2			0,856	-0,655		
OV3			0,473	0,965		
OV4			-	-		
PO1			-	-		
PO2			0,063***	1,440		
PO3			0,300	1,242		
P13					1,404	0,914
P12					1,147	1,240
P22					0,938	0,786
P21					0,621	-0,417
P33					0,364*	2,563
P32					0,486	2,108
P31					0,563	0,969
P41					-	-
Amostra	18823		18823		18823	
p-R ²	0,0201		0,0384		0,0401	
MVL (LL)	-23440,6		-23002,5		-22962,5	
Restrições	15		30		40	
GL	Base		15		25	
			Base		10	
-2LL	Base		876,2		956,2	
			Base		80	
Chi ² (α=0,05)	Base		25,0		37,6	
			Base		18,3	
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25						

A partir destes resultados não foi possível retirar uma conclusão clara sobre a melhor forma de utilizar esta informação. Para avaliar outras possíveis formas de introduzir esta informação, realizou-se um modelo com combinação de variáveis binárias relacionando pares de combinações das variáveis PO e OV, e outro com a combinação das suas variáveis contínuas. Os resultados destes modelos e a sua comparação com o modelo anterior para as alternativas mais relevantes são apresentados no quadro V.52. Os parâmetros estão dispostos

por ordem crescente em relação ao que o senso comum sugeriria como mais favorável à utilização do automóvel.

O modelo com a combinação das variáveis contínuas (MNL3.74) apresenta um ajustamento global inferior aos outros modelos, e o teste -2LL mostra que estatisticamente o modelo não é melhor que os modelos não lineares. Em relação aos parâmetros, estes apresentam sinais coerentes, com a atratividade das alternativas a diminuir com o aumento da oferta de estacionamento e com um aumento de atratividade com o incremento da percentagem de estacionamento pagos. Verifica-se que a taxa de estacionamento pago não tem significância estatística para os modos Moto e pedonal e que o modo Outros Bus também não revela significância estatística para a densidade de oferta de estacionamento face à procura de estacionamento.

O modelo com variáveis binárias de combinações apresenta uma ligeira melhoria de ajustamento global, que o teste estatístico -2LL valida para um intervalo de confiança de 95%. Em relação aos parâmetros observa-se uma tendência decrescente para o modo Bus, sem significância para os modos Outro Bus e Moto, e sem uma tendência clara para o modo Walk e Bus+Walk. Estas tendências não aparentam, no entanto, ser lineares. Em relação aos parâmetros sem significância estatística, tal como nos modelos anteriores e provavelmente devido à redução da amostra, os modos Outros Bus e Moto aparentemente não são influenciados por estas variáveis. Em relação aos sinais, 22 dos 25 parâmetros com significância estatística apresentam valores positivos, de acordo com o esperado atendendo à normalização da variável caracterizadora da situação mais favorável ao uso do automóvel.

Observando os resultados não é possível definir com certeza absoluta a tipologia de variável mais adequada para inserção desta informação no modelo. As variáveis como especificadas não apresentam uma tendência completamente definida, sugerindo que a sua influência não é totalmente clara. Admitindo que a possibilidade de estacionar influencia de facto as escolhas modais, principalmente o uso do automóvel, os resultados sugerem que as especificações utilizadas poderão não ser a melhor forma de avaliar essa influência ou que a mesma possa estar correlacionada com outros fatores ainda desconhecidos. Para estudos posteriores, no sentido de avaliar eventuais correlações com outras variáveis a utilizar, sugere-se, ainda assim, a especificação com combinação de binárias para cada variável em separado (OVi e POi). O modelo com estas variáveis apresenta um ajustamento muito semelhante à especificação com variáveis binárias de combinação (Pij) mas a interpretação dos seus parâmetros é mais simples e direta.

9.4. Correlação entre a Velocidade Comercial Equivalente de BUS e a Distância Mais Curta entre Centróides ou o Tempo Médio de Viagem

Dos resultados apresentados anteriormente foi possível constatar que o modelo com a variável contínua para a velocidade comercial equivalente para BUS (V2) apresentava um valor de ajustamento muito superior ao verificado com a especificação com combinação de binárias. O aumento substancial do ajustamento combinado com os resultados para os modos Bus e pedonal e a forma de cálculo da velocidade comercial equivalente, sugerem que possa haver uma grande correlação da mesma com a variável distância mais curta entre centróides e eventualmente com a duração média da viagem por Bus. Para avaliar esta possível correlação realizaram-se análise cujos resultados são apresentados nos quadros V.53 e V.54.

Quadro V. 53 - Resultados da análise de correlação entre as variáveis contínuas velocidade comercial equivalente de BUS (V2) e distância mais curta entre centróides (D2) através do *software SPSS®*

Correlações		D2
V2	Pearson Correlation	0,744 ⁴
	Sig. (2-tailed)	0,000
	N	89305

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

Quadro V. 54 - Resultados da análise de correlação entre a variável contínua velocidade comercial equivalente de BUS (V2) e as binárias da duração média da viagem (Til) através do *software SPSS®*

Correlações		T1I	T2I	T3I	T4I	T5I	T6I
V2	Pearson Correlation	0,054 [*]	-0,103 ²	0,076 ¹	0,228 ²	0,102 ²	0,021 ¹
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000
	N	57603	89305	60265	72006	67825	89305

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

Das análises é possível verificar uma grande correlação positiva entre a velocidade e a distância mais curta, mas uma correlação muito fraca e negativa com a duração média da viagem (na sua versão logarítmica). Ambas as correlações apresentam sinais coerentes, sugerindo que as velocidades maiores acontecem quando as distâncias entre GA também são maiores, e que tempos médios de viagem menores acontecem quando há velocidades maiores.

A avaliação do efeito conjugado destas variáveis foi realizada através de modelos exploratórios, cujos resultados são apresentados nos quadros V.55 e V.56. A combinação da velocidade comercial equivalente com a distância mais curta revelou um ajustamento muito aquém da soma dos ajustamentos dos modelos isolados. Apesar de haver uma redução na amostra com a inclusão da variável velocidade, verifica-se que o modelo combinado tem um ajustamento pouco superior ao modelo apenas com a distância mais curta. Tal resultado sugere que a informação adicionada ao modelo comparativamente à capacidade explicativa da variável distância mais curta é reduzida, e que provavelmente a informação existente na variável velocidade provém da sua formulação, que já inclui a distância mais curta.

Quadro V. 55 - Resultados dos modelos exploratórios com combinação das variáveis velocidade comercial equivalente do BUS e distância mais curta entre centróides

Modelo	MNL1.53		MNL1.63		MNL8a.14	
Modo	B	W	B	W	B	W
D2	-0,0049**	-2,6557			0,0713	-2,2113
V2			-0,0323	-0,8927	-0,0607	-0,1902
Amostra	95426		89305		89305	
p-R ²	0,194		0,175		0,201 ($\Sigma=0,369$)	
MVL (LL)	-99549,0		-95406,9		-92403,6	
Restrições	10		10		15	
GL			Base		5	
-2LL			Base		6006,6	
Chi ² ($\alpha=0,05$)			Base		11,070	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Quadro V. 56 - Resultados dos modelos exploratórios com combinação das variáveis velocidade comercial equivalente do BUS, distância mais curta entre centróides e tempo médio de viagem

Modelo	MNL1.80		MNL3.138		MNL8a.15	
Modo	B	W	B	W	B	W
D2			0,0644	-2,6434	0,2484	-1,6508
V2					-0,1085	-0,4559
Til = Ln(Tia)	-0,7525		-0,6252		-0,6923	
Amostra	30559		30559		30559	
p-R ²	0,017		0,187		0,199	
MVL (LL)	-39013,4		-32274,1		-31790,3	
Restrições	6		11		16	
GL	Base		5		10	
			Base		5	
-2LL	Base		13478,6		14446,2	
			Base		967,6	
Chi ² ($\alpha=0,05$)	Base		11,070		18,307	
			Base		11,070	

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Relativamente aos parâmetros para a distância, os sinais do modelo combinado corrigem a não significância estatística da distância para os BUS, mantêm o sinal para quase todos os

restantes modos mas alteram a sua grandeza. Em relação à velocidade, os sinais dos parâmetros mantêm-se no modelo combinado havendo um ajuste significativo das suas grandezas. Atendendo aos resultados anteriores, a variável velocidade comercial equivalente de BUS não parece acrescentar capacidade explicativa relevante ao modelo comparativamente à variável distância.

Como a velocidade na sua formulação também incluiu o tempo de viagem de BUS, compararam-se os resultados do modelo com combinação das três variáveis (MNL8a.15) com os do modelo apenas com a distância e a variável genérica logaritmo da duração média da viagem entre centróides (MNL3.138).

É possível constatar que para os modos mais importantes (e portanto menos sensíveis à significativa redução da amostra) os sinais e grandezas dos parâmetros do modelo apenas com a distância e com a velocidade (MNL8a.14) são muito semelhantes aos do modelo com a distância e o tempo de viagem (MNL3.138). A introdução da velocidade altera a grandeza dos parâmetros e mantém os sinais dos modos mais importantes para a distância, mantendo-se os parâmetros para a duração média da viagem estáveis. O ajustamento global do modelo não apresenta grandes ganhos com a introdução da variável velocidade comercial equivalente.

Estes resultados sugerem que aparentemente o elevado ajustamento do modelo só com a variável velocidade comercial equivalente com especificação contínua advém da sua formulação e da informação contida nas variáveis primárias distância e tempo de viagem, aparentemente não justificando a sua introdução no modelo.

10. Estabilidade Modal ao Longo das Viagens Diárias

10.1. Modo Escolhido na 1ª Viagem do Dia *versus* Modo da Viagem Atual

Apesar de as escolhas modais ao longo do dia não poderem ser incluídas no modelo MNL uma vez que matematicamente este não permite relacionar escolhas anteriores ou posteriores à viagem em análise, o senso comum sugere que o modo utilizado em determinada viagem pode estar diretamente relacionada com a liberdade/utilidade para poder mudar de modo ao longo do dia. Geralmente essa liberdade/utilidade é limitada e está associada ao modo utilizado na primeira viagem do dia, especialmente quando os modos das primeiras viagens não são o modo pedonal ou combinação de modos.

		Modo de transporte utilizado na viagem actual										Total	
		AUTO	%	BUS	%	Outro Bus	%	Moto	%	Walk	%	Walk+Bus	%
modo de transporte utilizado na 1ª viagem do dia (se não for a 1ª viagem)	AUTO	34445	94,1%	941	2,6%	156	0,4%	170	0,5%	796	2,2%	91	0,2%
	BUS	1142	10,1%	9333	82,9%	29	0,3%	63	0,6%	546	4,9%	141	1,3%
	Outro Bus	322	14,3%	35	1,6%	1790	79,7%	35	1,6%	82	2,8%	3	0,1%
	Comb Bus	3		4		1		0		1		1	
	Train	24		24		2		3		17		2	
	Moto	391	8,8%	17	0,4%	12	0,3%	3953	89,1%	60	1,4%	5	0,1%
	Bike	38		4		0		2		9		0	
	Outro	59		5		1		3		1		0	
	Walk	1592	17,9%	681	7,6%	83	0,9%	93	1,0%	6338	71,1%	131	1,5%
	Taxi	20		28		0		2		17		2	
	Bus+Auto	68		52		3		3		9		14	
	Outro Bus+Auto	32		0		5		0		2		1	
	Bus+Train	15		11		0		2		3		1	
	Auto+Train	28		0		0		0		4		0	
	Walk+Bus	148	10,4%	230	16,2%	3	0,2%	19	1,3%	97	6,8%	924	65,0%
	Outro Bus+Walk	17		1		8		0		6		2	
	Outra Comb	211		24		7		29		69		21	
	Total	38555		11388		2100		4377		8037		1339	

Figura V. 28 - Análise do modo de cada viagem em relação ao modo da primeira viagem do dia

Esta liberdade associada aos modos pedonal, e provavelmente aos modos combinados, deve-se à facilidade de alteração de modo por não necessitarem de trazer o veículo de volta ou por haver maior propensão para alterações modais. Estas relações espectáveis foram verificadas na base de dados completa e são apresentadas na figura IV.28. Os seus resultados revelam que o modo utilizado numa determinada viagem ao longo do dia é na grande maioria das vezes o utilizado também na primeira viagem do dia. Os modos que apresentam maior estabilidade/fidelidade são o Auto (84%), o Bus (83%), Outro Bus (80%) e Moto (89%), descendo o modo Walk para 71% e Bus+Walk para 65%. Verifica-se que mesmo com viagens iniciais de Walk, 18% das restantes viagens do dia são feitas de Auto, e se o primeiro modo do dia for a combinação Bus+Walk, 16% das restantes são apenas de Bus, 10% passam para o Auto e apenas 7% se mantém apenas de Walk.

Realizou-se igualmente uma avaliação da correlação entre o modo de transporte de determinada viagem e o modo da viagem anterior, cujos resultados foram similares aos obtidos na análise anterior.

10.2. Estabilidade do Modo Escolhido ao Longo das Atividades do Dia

Considerou-se também útil analisar a estabilidade da escolha modal ao longo de toda a cadeia de viagens realizada ao longo do dia, verificando a eventual existência de padrões comportamentais característicos, possivelmente diferenciados para distintos grupos socioeconómicos.

Numa primeira fase procurou-se avaliar a estabilidade do modo utilizado ao longo do dia para diferentes condições sociais. Para esta análise foi considerada a amostra com dados sobre a taxa de motorização familiar por adulto e em que todos os elementos tenham sido entrevistados ou tenham realizado viagens no dia em estudo (NCa_b). A necessidade de amostras com tamanho razoável limitou a análise, pelo que foram seleccionados três tipos de indivíduos com diferentes níveis de acesso ao automóvel:

- Tipo I: Homens, com licença de condução, idade entre 35 e 65 anos e com uma taxa de motorização familiar por adulto entre 0,5 e 1 veíc./hab., representativo do grupo populacional com maior propensão para escolher o automóvel;
- Tipo II: Mulher com licença de condução com idade entre 35 e 65 anos e com uma taxa de motorização familiar por adulto entre 0,5 e 1 veíc./hab., para comparação com os resultados para os homens;
- Tipo III: Mulher sem licença de condução com idade entre 35 e 65 anos e sem automóvel disponível, representativa do grupo populacional menos propenso a andar de automóvel;

Das figuras V.29, V.30 e V.31 é possível verificar que para as viagens realizadas por automóvel:

- A maioria dos indivíduos do Tipo I utilizam-no para todas as viagens ao longo de todo o dia (81%), resultado coerente com a sua propensão para a essa utilização;

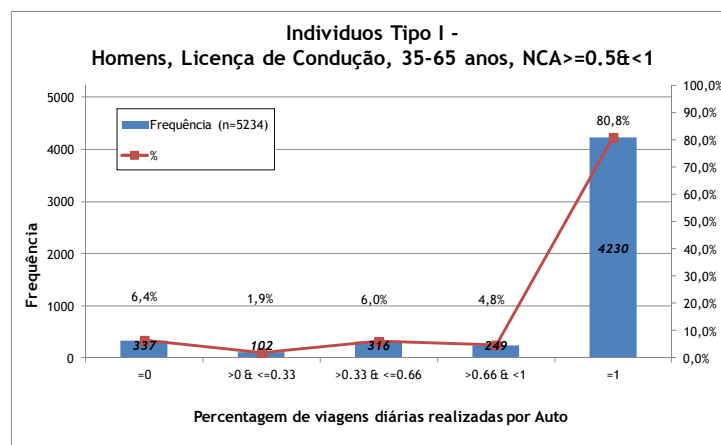


Figura V. 29 - Estabilidade do uso do modo Auto ao longo do dia de indivíduos do Tipo I

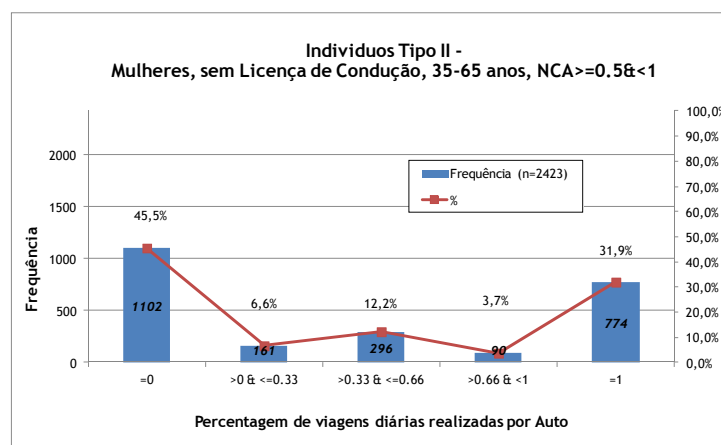


Figura V. 30 - Estabilidade do uso do modo Auto ao longo do dia para indivíduos do Tipo II

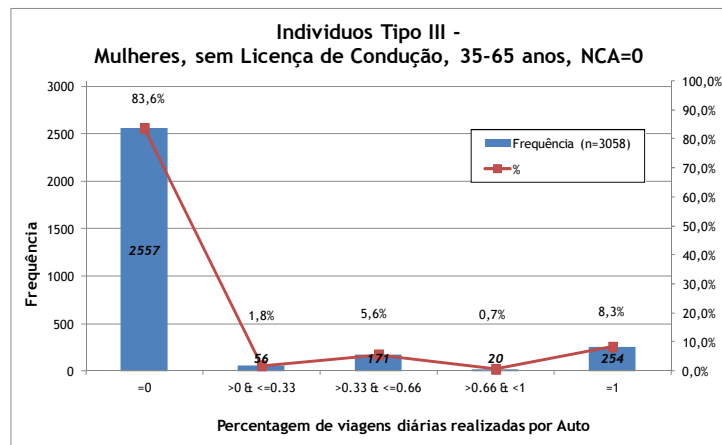


Figura V. 31 - Estabilidade do uso do modo Auto ao longo do dia para indivíduos do Tipo III

- Para os indivíduos do Tipo II, mulheres com as mesmas condições dos homens do tipo I, verifica-se que os resultados se modificam drasticamente. Apenas 32% utilizam sempre o automóvel, enquanto 45,5% nunca o utilizam. Este resultado sugere que para as mesmas condições sociais, nomeadamente sem um automóvel por adulto no agregado, que a mulher não tem o mesmo nível de acesso ao uso do mesmo. No entanto, é relevante mencionar que a grande maioria dos casos (77,4%), revela que grande parte das escolhas ao longo do dia continua a ser uniforme (ou se anda sempre de automóvel ou não se anda nunca);
- Para os indivíduos do Tipo III verifica-se exatamente o oposto do tipo I, com 84% destas mulheres a não realizar nenhuma viagem de automóvel ao longo do dia. Este resultado corrobora o pressuposto de ser o grupo com menor propensão para o uso do automóvel;

Estes resultados sugeriram uma análise mais aprofundada para os restantes modos mais relevantes e as suas eventuais relações com o modo escolhido para a primeira viagem do dia.

Assim, foram definidas 3 variáveis atendendo à sua sequência nas viagens diárias de cada indivíduo. Na base de dados original foram avaliadas todas as viagens realizadas por cada um dos inquiridos ao longo do dia e o modo de transporte utilizado em cada uma delas. Verificou-se o modo de transporte utilizado em cada viagem para as seguintes situações:

- 1ª Viagem do dia
- Viagens intermédias ao longo do dia
- Viagem final do dia

Os modos de transporte considerados foram os 18 modos agregados definidos no capítulo IV. Para todas as viagens foi verificado o modo de transporte utilizado na 1ª viagem do dia, existindo 18 possíveis modos, e para as restantes apenas se verificou o modo utilizado na viagem em estudo. Atendendo à simplificação realizada para a base de dados utilizada, no

estudo para as viagens intermédias e de final de dia apenas se dispunham dos 6 modos mais relevantes. As viagens únicas ao longo do dia foram consideradas apenas como 1ª viagem e não nas últimas viagens diárias. Atendendo a estas condicionantes as amostras para cada tipo de viagem são apresentadas no quadro V.57.

Quadro V. 57 - Amostras para os diferentes tipos de viagens analisadas para avaliar a influência do modo escolhido na 1ª viagem diária nas restantes viagens diárias

Situação	Amostra	%
1ª Viagem do dia	29 630	31,1%
Viagens Intermédias	32 185	33,7%
Últimas viagens	33 611	35,2%
Total	95 426	100,0%

A repartição pelos modos de transporte considerados mais relevantes foi avaliada para distâncias até 1km, entre 1 e 2Km e para viagens superiores a 2km.

Na análise da continuidade do modo utilizado na 1ª viagem diária teve-se em consideração os modos que permitem maior mobilidade e conveniência, e a sua importância em termos sociais. Atendendo à probabilidade de viagens realizadas através de combinação de modos apresentarem comportamentos distintos das monomodais, considerou-se também as viagens feitas por Bus+Walk. Das 1ª viagens analisadas foram excluídas as realizadas por outros modos que não os seis considerados no estudo. Dada a reduzida amostra para o modo OB como modo utilizado na primeira viagem do dia optou-se também por excluir este modo da análise. Atendendo a estas opções a amostra limitou-se a 62 630 viagens. Os resultados da análise são apresentados no quadro V.58.

É possível constatar que para as viagens para menos de 1km de distância, a maioria das viagens realizadas é independente do modo utilizado na 1ª viagem do dia. Para estas viagens a predominância do modo pedonal é muito forte, sempre acima de 90%. Apenas nas últimas viagens de quem realizou a primeira de Bus+Walk não se verifica este resultado. No entanto a amostra desta situação é extremamente reduzida pelo que não é possível tirar nenhuma conclusão definitiva.

Quadro V. 58 - Análise dos modos utilizados nas viagens intermédias e finais do dia comparativamente ao modo utilizado na primeira viagem do dia para diferentes distâncias

Modo da 1ª Viagem do dia	Modo das viagens intermédias									Modo da última viagem do dia								
	Modo	D2<=1km	%	D2>1 & <=2km	%	D2>2km	%	Total	%	Modo	D2<=1km	%	D2>1 & <=2km	%	D2>2km	%	Total	%
Auto	Auto	24	8,2%	1334	88,2%	16865	96,3%	18223	94,3%	Auto	19	10,2%	1053	85,5%	15150	95,6%	16222	93,9%
	BUS	2	0,7%	26	1,7%	414	2,4%	442	2,3%	BUS	0	0,0%	16	1,3%	483	3,0%	499	2,9%
	OutroBus	0	0,0%	4	0,3%	80	0,5%	84	0,4%	Outro Bus	0	0,0%	4	0,3%	68	0,4%	72	0,4%
	Moto	0	0,0%	5	0,3%	94	0,5%	99	0,5%	Moto	0	0,0%	2	0,2%	69	0,4%	71	0,4%
	Walk	267	91,1%	139	9,2%	30	0,2%	436	2,3%	Walk	168	89,8%	154	12,5%	38	0,2%	360	2,1%
	Walk+BUS	0	0,0%	5	0,3%	38	0,2%	43	0,2%	Walk+BUS	0	0,0%	3	0,2%	45	0,3%	48	0,3%
	Total	293	1,5%	1513	7,8%	17521	90,7%	19327		Total	187	1,1%	1232	7,1%	15853	91,8%	17272	
BUS	Auto	0	0,0%	28	6,7%	418	11,6%	446	10,7%	Auto	0	0,0%	42	8,5%	654	10,1%	696	9,8%
	BUS	4	2,5%	290	69,0%	3062	85,2%	3356	80,4%	BUS	7	6,2%	320	64,9%	5650	87,3%	5977	84,4%
	OutroBus	0	0,0%	0	0,0%	17	0,5%	17	0,4%	Outro Bus	0	0,0%	2	0,4%	10	0,2%	12	0,2%
	Moto	0	0,0%	1	0,2%	29	0,8%	30	0,7%	Moto	0	0,0%	1	0,2%	32	0,5%	33	0,5%
	Walk	156	97,5%	97	23,1%	28	0,8%	281	6,7%	Walk	106	93,8%	119	24,1%	40	0,6%	265	3,7%
	Walk+BUS	0	0,0%	4	1,0%	40	1,1%	44	1,1%	Walk+BUS	0	0,0%	9	1,8%	88	1,4%	97	1,4%
	Total	160	3,8%	420	10,1%	3594	86,1%	4174		Total	113	1,6%	493	7,0%	6474	91,4%	7080	
Moto	Auto	0	0,0%	7	4,1%	196	8,4%	203	8,1%	Auto	0	0,0%	6	6,3%	182	10,1%	188	9,8%
	BUS	0	0,0%	0	0,0%	10	0,4%	10	0,4%	BUS	0	0,0%	0	0,0%	7	0,4%	7	0,4%
	OutroBus	0	0,0%	0	0,0%	6	0,3%	6	0,2%	Outro Bus	0	0,0%	0	0,0%	6	0,3%	6	0,3%
	Moto	0	0,0%	156	91,2%	2115	90,8%	2271	90,2%	Moto	0	0,0%	80	84,2%	1602	88,8%	1682	87,6%
	Walk	18	100,0%	8	4,7%	2	0,1%	28	1,1%	Walk	20	100,0%	8	8,4%	4	0,2%	32	1,7%
	Walk+BUS	0	0,0%	0	0,0%	1	0,0%	1	0,0%	Walk+BUS	0	0,0%	1	1,1%	3	0,2%	4	0,2%
	Total	18	0,7%	171	6,8%	2330	92,5%	2519		Total	20	1,0%	95	5,0%	1804	94,0%	1919	
Walk	Auto	1	0,1%	101	8,7%	712	52,4%	814	18,2%	Auto	5	0,3%	97	6,0%	676	52,6%	778	17,5%
	BUS	0	0,0%	55	4,7%	350	25,8%	405	9,1%	BUS	1	0,1%	37	2,3%	238	18,5%	276	6,2%
	OutroBus	0	0,0%	4	0,3%	52	3,8%	56	1,3%	Outro Bus	0	0,0%	2	0,1%	25	1,9%	27	0,6%
	Moto	0	0,0%	2	0,2%	51	3,8%	53	1,2%	Moto	0	0,0%	1	0,1%	39	3,0%	40	0,9%
	Walk	1937	99,9%	999	85,6%	142	10,4%	3078	69,0%	Walk	1544	99,6%	1475	91,1%	241	18,8%	3260	73,2%
	Walk+BUS	0	0,0%	6	0,5%	52	3,8%	58	1,3%	Walk+BUS	0	0,0%	7	0,4%	66	5,1%	73	1,6%
	Total	1938	43,4%	1167	26,1%	1359	30,4%	4464		Total	1550	34,8%	1619	36,3%	1285	28,9%	4454	
Walk+BUS	Auto	0	0,0%	7	11,3%	58	15,0%	65	13,4%	Auto	10	100,0%	9	10,3%	74	8,8%	93	9,9%
	BUS	0	0,0%	11	17,7%	81	21,0%	92	18,9%	BUS	0	0,0%	12	13,8%	126	15,0%	138	14,8%
	OutroBus	0	0,0%	0	0,0%	1	0,3%	1	0,2%	Outro Bus	0	0,0%	0	0,0%	2	0,2%	2	0,2%
	Moto	0	0,0%	2	3,2%	8	2,1%	10	2,1%	Moto	0	0,0%	2	2,3%	7	0,8%	9	1,0%
	Walk	38	100,0%	14	22,6%	13	3,4%	65	13,4%	Walk	0	0,0%	11	12,6%	11	1,3%	22	2,4%
	Walk+BUS	0	0,0%	28	45,2%	225	58,3%	253	52,1%	Walk+BUS	0	0,0%	53	60,9%	618	73,7%	671	71,8%
	Total	38	7,8%	62	12,8%	386	79,4%	486		Total	10	1,1%	87	9,3%	838	89,6%	935	62630

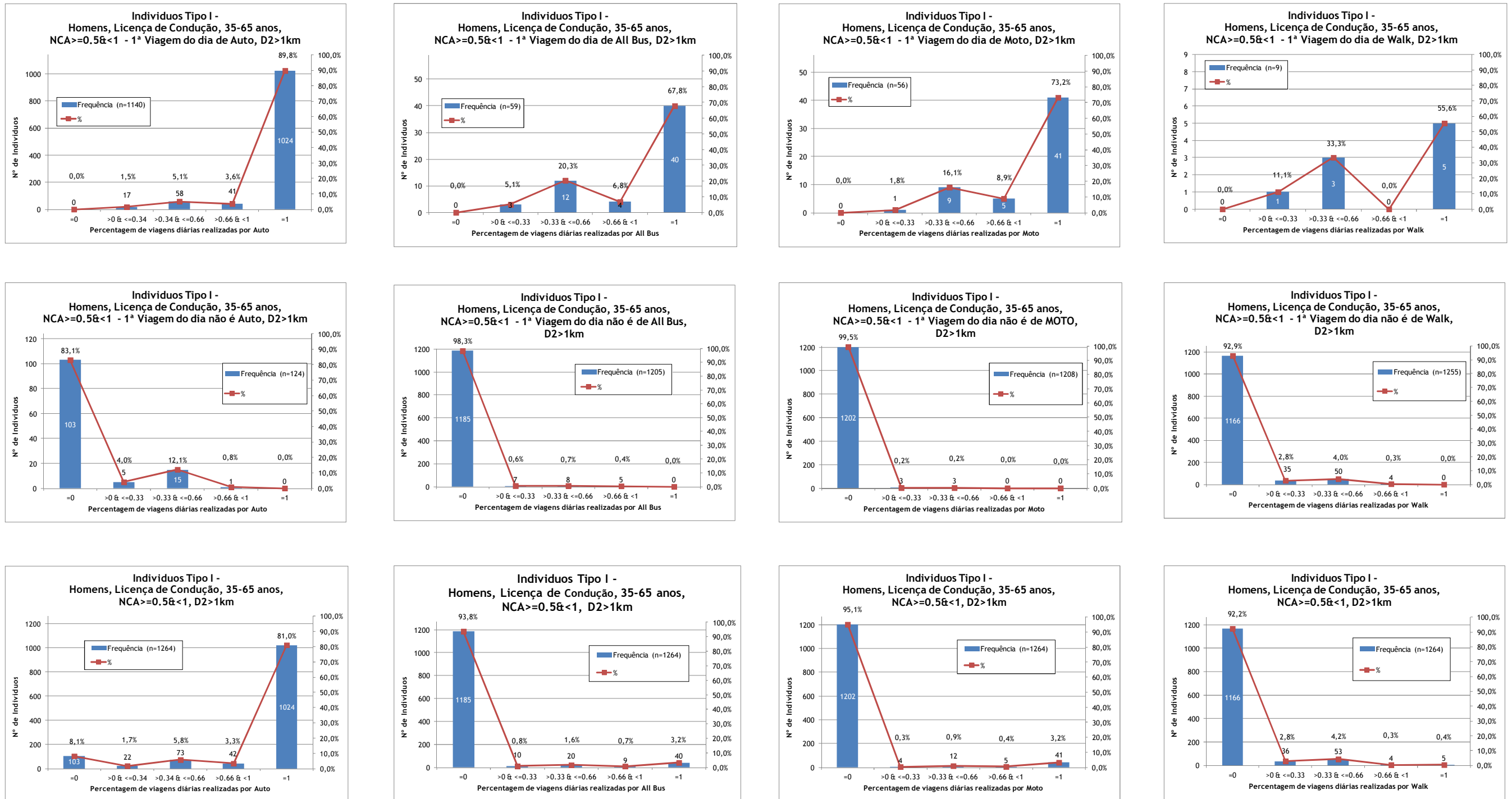


Figura V. 32 - Análise da estabilidade do modo ao longo do dia para indivíduos Tipo I, atendendo ao modo utilizado na primeira viagem do dia

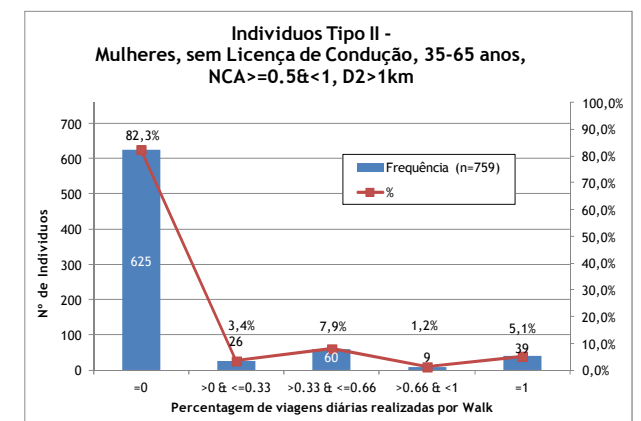
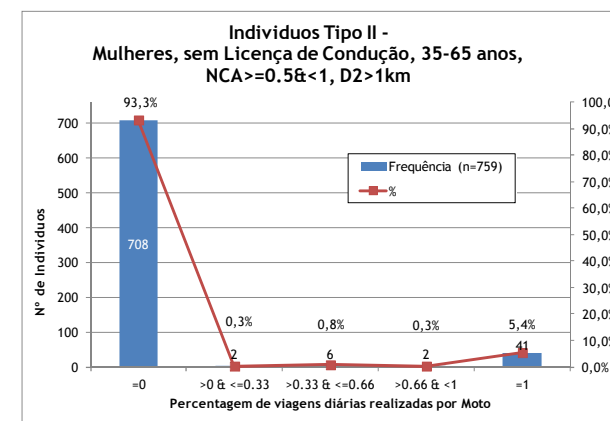
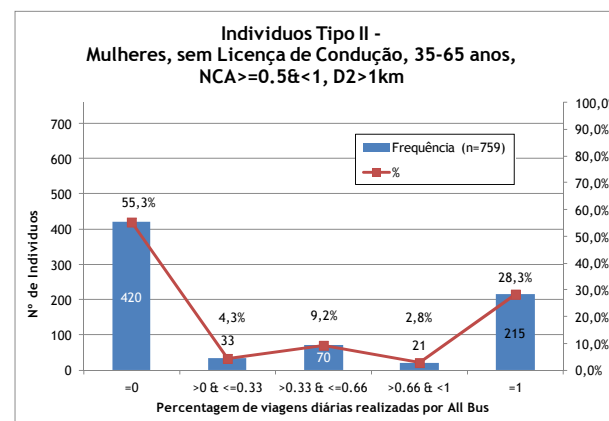
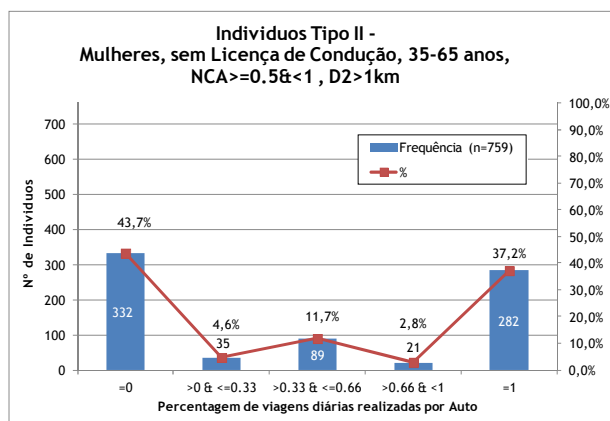
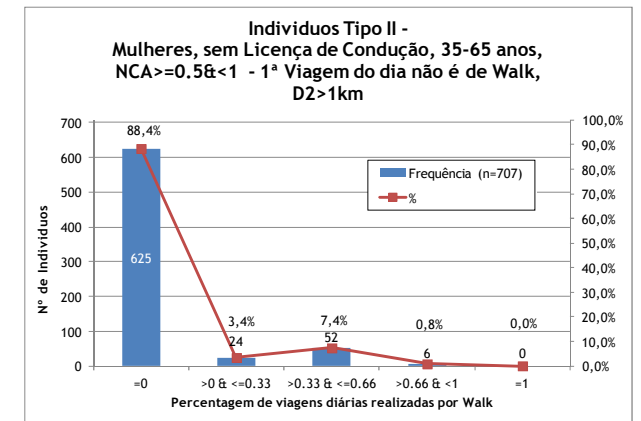
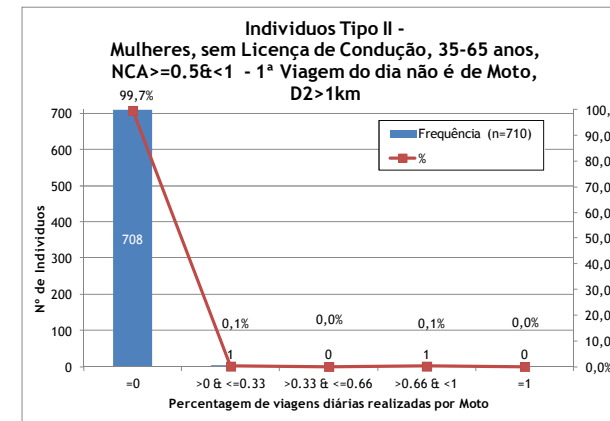
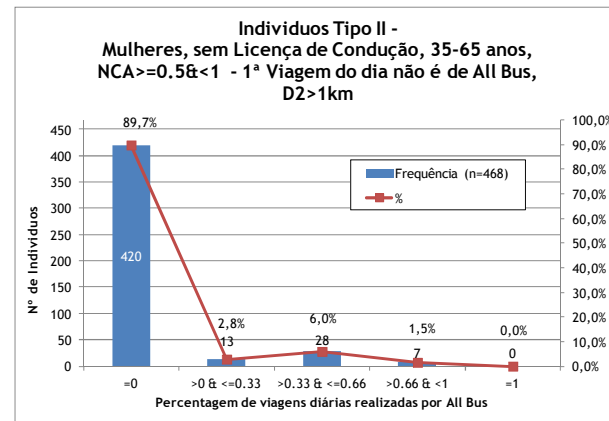
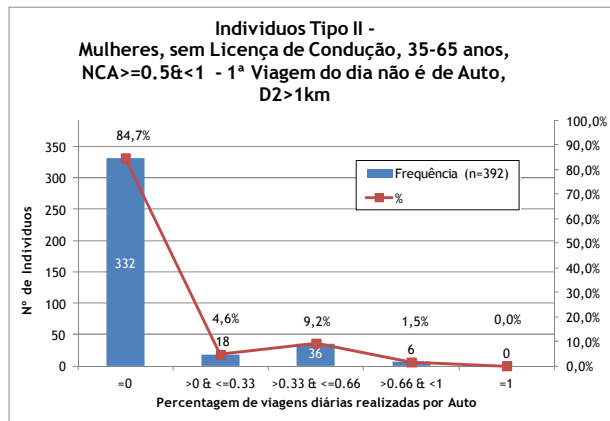
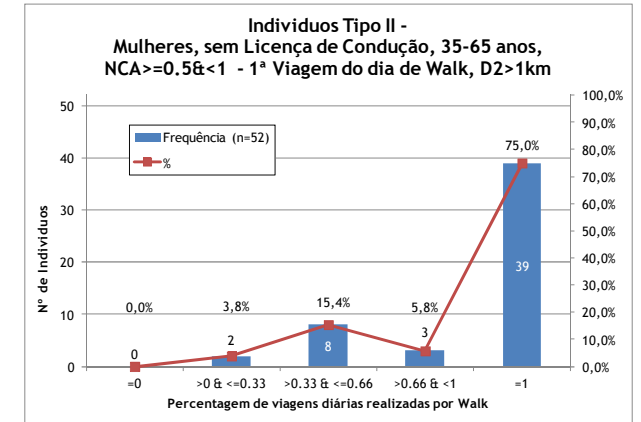
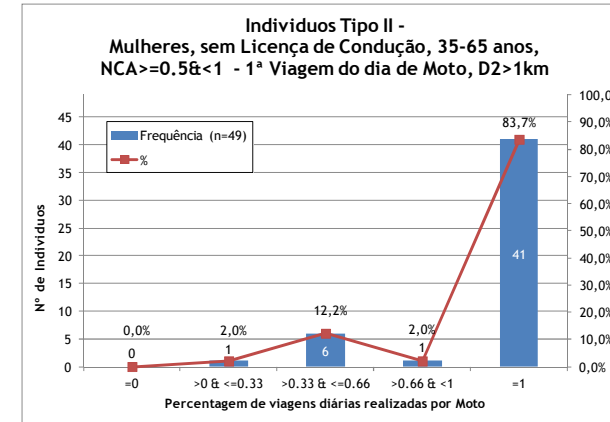
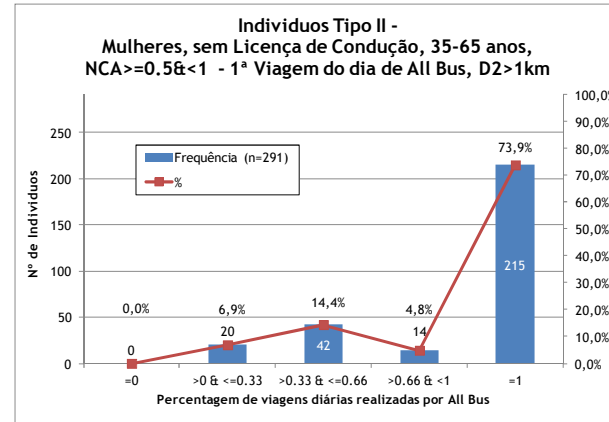
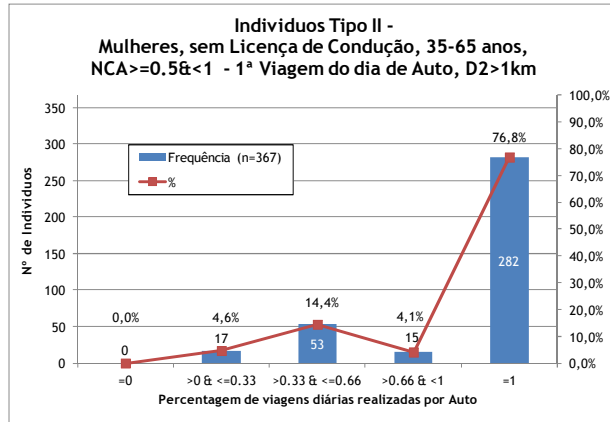


Figura V. 33 - Análise da estabilidade do modo ao longo do dia para indivíduos Tipo II, atendendo ao modo utilizado na primeira viagem do dia

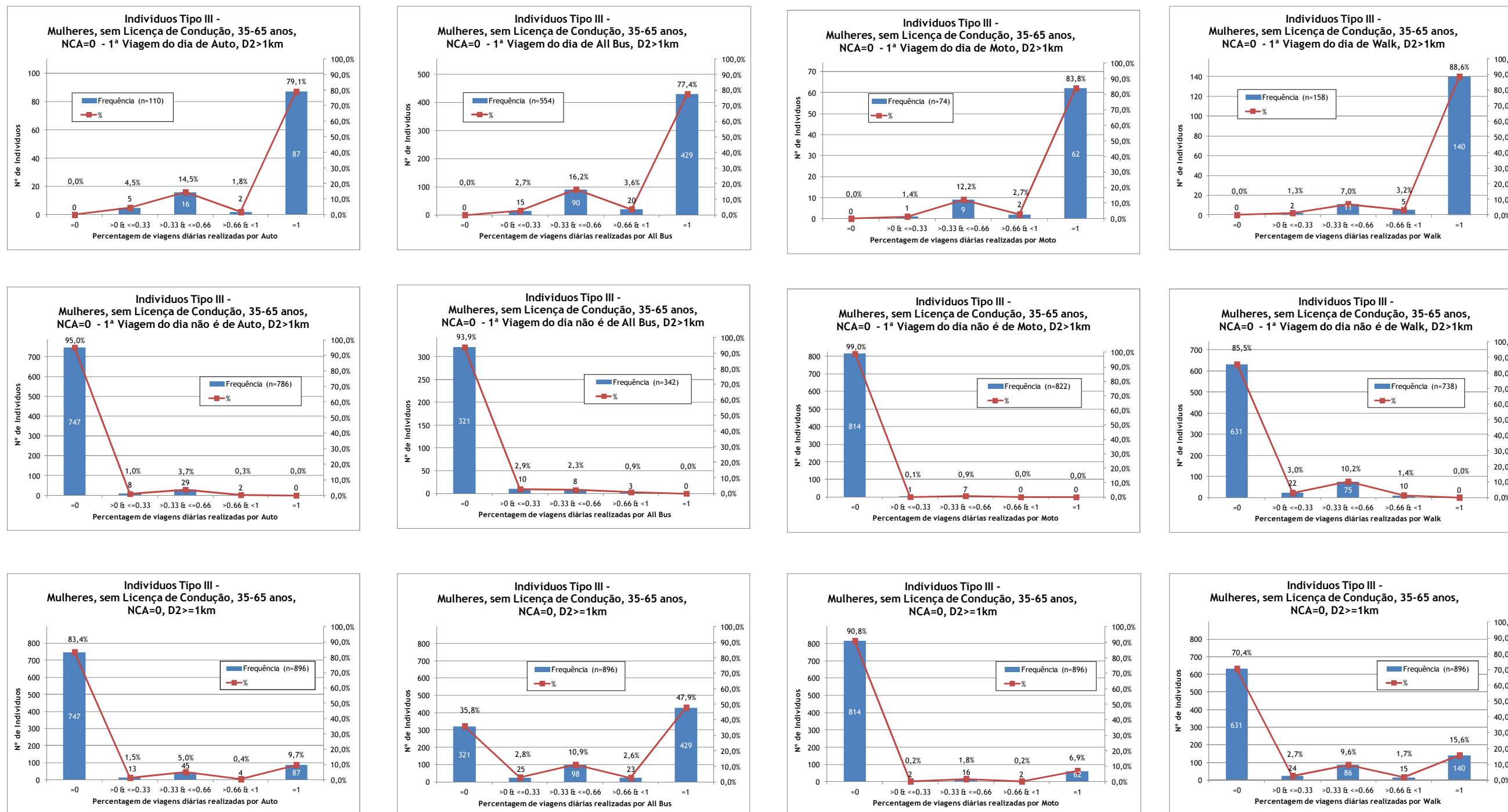


Figura V. 34 - Análise da estabilidade do modo ao longo do dia para indivíduos Tipo III, atendendo ao modo utilizado na primeira viagem do dia

Nos restantes casos pode verificar-se que:

- Com exceção do modo combinado Bus+Walk, o comportamento para as viagens intermédias e para as últimas viagens do dia é muito semelhante;
- Para os modos mais relevantes e para distâncias superiores a 2km, a grande maioria das viagens intermédias e finais são realizadas no mesmo modo da primeira viagem do dia;
- Quando a 1ª viagem do dia é realizada de determinado modo motorizado não combinado, as restantes acima de 1km também são realizadas pelo mesmo modo (sempre acima de 65%);
- Quando a 1ª viagem do dia é realizada de determinado modo motorizado não combinado, com o aumento da distância há tendência para um acréscimo de viagens realizadas por automóvel;
- Como seria de esperar, para as viagens cujo primeiro modo foi pedonal, para distâncias superiores a 2km há tendência para acentuado aumento no uso do Auto e Bus nas restantes viagens do dia;
- Para o modo combinado Bus+Walk, para distâncias superiores a 2km, apesar da predominância do mesmo modo, nas restantes viagens verifica-se um aumento do uso do modo Auto e Bus, com predominância para este último;
- O modo pedonal ainda apresenta alguma relevância para distâncias até 2km, mesmo quando na 1ª viagem do dia não se utilizou o modo pedonal;

A conjugação das duas análises anteriores levou ao estudo da estabilidade do modo de transporte para os três tipos de indivíduos sugeridos atendendo à distância e ao modo utilizado na primeira viagem do dia. Nesta análise, para facilitar a interpretação dos resultados, e tendo em atenção as maiores semelhanças entre estas alternativas, optou-se por juntar todos os modos com Bus num modo designado *All Bus*. Tendo em consideração a hegemonia do modo pedonal para distâncias até 1km, a análise apenas focou distâncias superiores. Os gráficos com essas análises podem ser consultados nas figuras V.32, V.33 e V.34.

Dos gráficos é possível constatar que as tipologias de indivíduos assumidos mantêm o modo de transporte ao longo do dia para a maioria das viagens. Para todos os modos e para todas as tipologias, a percentagem de indivíduos que fazem todas as viagens no modo da primeira viagem é superior a 55% e, na maioria dos casos, acima dos 85%. A tendência inversa verifica-se na avaliação do uso de alternativas que não sejam as utilizadas na primeira viagem do dia, em que sempre mais de 75% dos indivíduos não fazem nenhuma viagem durante o resto do dia nas mesmas. Este resultado demonstra que qualquer política de mobilidade sustentável deverá incidir na opção de fomentar o uso do modo de transportes sustentáveis para todas as

viagens do dia e logo na viagem inicial a partir de casa. Aparentemente, existe uma relutância ou dificuldade em mudar de modo ao longo do dia e parece credível que estas conclusões possam ser extrapoladas para outras tipologias de indivíduo.

Entender como é feita esta primeira escolha, ou como a sucessão de escolhas ao longo do dia se interliga, poderá ajudar a influenciá-la e a promover políticas que tenham maior probabilidade de sucesso.

11. Síntese

Neste capítulo utilizaram-se modelos MNL para realizar análises preliminares sobre as variáveis, e suas diferentes especificações, que apresentavam maior potencial para simular o processo de decisão nas escolhas modais. Atendendo às tipologias de especificações mais comuns, compararam-se a coerência de sinais, os ajustamentos globais e a eventual relação linear entre as grandezas dos seus parâmetros.

A suspeita de eventuais correlações ou complementaridade entre algumas variáveis justificou um estudo mais pormenorizado das mesmas. Estas análises foram realizadas através de coeficientes de Pearson, e da comparação dos resultados de modelos MNL exploratórios com a sua combinação.

Assim, dos resultados das análises preliminares iniciais foi possível identificar quais as variáveis e especificações que potencialmente poderão fornecer melhores resultados. O resumo das mesmas, bem como a sua classificação como variáveis de base, que se revelaram mais significativas e devendo fazer parte de qualquer modelo base inicial, ou adicionais, que ainda necessitam de ser testadas combinadas com as de base, é apresentado no quadro V.59.

De forma resumida pôde verificar-se que as variáveis que se identificaram como mais relevantes nas escolhas modais foram a taxa de motorização familiar, a existência de licença de condução e a distância mais curta entre centróides. Esta constatação levou à decisão de as incluir numa modelação base a considerar sempre nos modelos onde se pretenda estudar combinações das restantes variáveis.

Para as variáveis remanescentes menos proeminentes mas cujos resultados sugerem que podem introduzir alguma informação relevante ao modelo sugere-se que, na sua maioria, são melhor representadas pelas suas especificações com combinação de binárias. Apenas a duração média da viagem (genérica e versão logarítmica), o número de viagens diárias, a razão entre tempos de viagem Bus/Auto e a frequência média diária de Bus são sugeridas com especificação contínua, quer pela sua natureza quer pelos resultados verificados na análise da sua linearidade.

Quadro V. 59 - Síntese das variáveis e especificações sugeridas para os modelos combinados

Designação	Variável	Base	Adicional	Especificação
Rendimento líquido mensal do agregado (€/mês)	Ri		X	Binária
Sexo (0=M, 1=F)	Sx		X	Binária
Nível de Instrução	Ini & Idi		X	Binárias combinadas
Escalão etário				
Número de automóveis disponíveis diariamente per capita no agregado (veic./hab)	NCi	X		Binária
Disponibilidade diária de Moto no agregado	Mot		x	Binária
Licença de Condução (0=N, 1=S)	Lc	X		Binária
Distância mais curta entre centróides (km)	D2	X		Contínua
	Dxi		X	Binária
Logaritmo do tempo médio de viagem por modo (genérica) (minutos)	Til		X	Ln de contínua
Motivo da viagem	Tr, Rg, Es, Cs		X	Binárias
Duração da estadia no destino (minutos)	Dei		X	Binária
Número de viagens diárias	nV		X	Contínua
Probabilidade de encontrar estacionamento	OVi & POi		X	Binárias Combinadas
Taxa de estacionamento pago				
Razão entre tempo de viagem Bus/Auto	TBA		X	Contínua
Potencial frequência média horária de BUS no período laboral (BUS/h)	Ftl(1)		X	Contínua

Salienta-se ainda a opção por combinar as variáveis que aparentemente se revelaram simbióticas, realçando-se particularmente o escalão etário e o nível de instrução, cuja combinação revelou a correção de incoerências no comportamento dos seus parâmetros para determinadas gamas. Como exemplo de complementaridade refere-se ainda que apenas a combinação das variáveis de base taxa de motorização familiar e disponibilidade de licença de condução permite a distinção de todas as diferentes situações relativamente à oportunidade de utilização do automóvel, razão pela qual se sugere em cada caso a avaliação da eventual relevância da sua utilização em simultâneo.

Em relação às variáveis relacionadas com as políticas de mobilidade, a velocidade comercial equivalente revelou que apesar de um elevado ajustamento global do modelo com a sua especificação contínua quando usada isoladamente, a informação que introduz no modelo é muito semelhante à introduzida pela distância mais curta entre centróides, variável primária utilizada para a definição da velocidade. Este facto é comprovado pela elevada correlação verificada entre as mesmas. Esta constatação sugere a sua eliminação como variável explicativa.

Ainda em relação às variáveis relacionadas com a oferta de estacionamento, a dificuldade em definir a sua melhor especificação levou tendencialmente à opção de combinar as suas especificações como combinação de binárias de cada variável em separado, uma vez que apresenta um dos melhores ajustamentos entre as diferentes especificações, com a vantagem de uma fácil interpretação dos seus parâmetros.

Em relação a todas as variáveis e especificações selecionadas, sugere-se a sua análise cuidada em relação à coerência de sinais, de comportamento e de relevância para o ajustamento global do modelo, na sua utilização em modelos combinados com melhor potencial explicativo.

Relativamente à estabilidade do modo de transporte ao longo do dia esta análise revelou que na maior parte das viagens se utiliza o modo escolhido para a primeira viagem do dia. Numa primeira avaliação da eventual influência de diferentes condições sociais na estabilidade do modo utilizado ao longo do dia, selecionaram-se grupos com amostras de tamanho suficiente para obter resultados com significância estatística. Foi possível constatar que as condições sociais influenciam o modo escolhido para realizar as viagens, mas que a estabilidade de escolha ao longo do dia é comum a todas as situações. A maioria dos inquiridos realiza quase todas as suas viagens no mesmo modo, geralmente coincidente com o modo da primeira viagem do dia.

Verifica-se ainda que, um dos poucos fatores que se mostra verdadeiramente capaz de “quebrar” esta fidelidade é a distância de viagem, já que para pequenas distâncias a atratividade do modo Pedonal se revela determinante.

Estes resultados sugerem uma fidelidade, voluntária ou imposta, ao modo escolhido à saída de casa que indicia a importância de poder influenciar as escolhas logo a partir da primeira viagem.

Capítulo VI

Modelações Combinadas

1. Enquadramento das Análises

A avaliação isolada das diferentes especificações de cada variável permitiu aferir o seu potencial explicativo nas escolhas modais da população em estudo e qual a especificação com melhores resultados, embora com a limitação de não se poder testar o efeito combinado ou sobreposto de diferentes variáveis. Neste capítulo será avaliada a possibilidade da sua combinação integrada em modelações mais completas. Este estudo foi feito de forma progressiva, com a formulação de um modelo exploratório de base, ao qual foram acrescentadas sucessivamente mais variáveis. Em cada iteração avaliou-se a estabilidade dos parâmetros, a coerência dos sinais e dos seus comportamentos e a melhoria do ajustamento global, comparativamente ao modelo até aí identificado como melhor. Este procedimento permitiu explorar quais as variáveis a integrar no modelo final e servir de referência tanto para os parâmetros como para o ajustamento global a considerar em modelos mais complexos.

Assim, no capítulo é apresentada a evolução do ajustamento com a introdução de diferentes variáveis, bem como a redução que implicam na amostra possível de utilizar. Atendendo à acentuada redução das amostras implícita à utilização das variáveis sobre políticas de mobilidade, definiram-se formulações complementares com integração de conjuntos diferentes de variáveis explicativas, incluindo apenas as que se considerou serem mais significativas, para situações em que se pretenda utilizar essas variáveis.

Após a definição das variáveis a utilizar procurou-se validar os resultados comparando as previsões dos modelos exploratórios com as repartições modais efetivamente verificadas nas respetivas bases de dados.

A possibilidade de existirem tendências não detetadas devido à falta de capacidade explicativa dos modelos exploratórios, justificou ainda a realização de análises adicionais com vista a avaliar situações de interdependência e não linearidade das variáveis, nomeadamente as socioeconómicas mais valorizadas.

O capítulo conclui-se com uma comparação dos parâmetros obtidos nos modelos exploratórios com o tipo de influência, linear ou não linear, positiva ou negativa, e gama de valores sugeridos para os mesmos em estudos internacionais.

2. Modelos Multinomiais Logit com Combinação de Variáveis

2.1. Modelo Base

Os resultados obtidos no capítulo anterior, nomeadamente atendendo às variáveis que maior capacidade explicativa demonstraram, levou à consideração do *modelo de base* MNL7.40, que incluiu as variáveis Lc, NCi e D2, tendo-se ainda testado a inclusão da variável Dx1:

- Lc: Variável binária Licença de condução (0-Não; 1-Sim);
- NCi: Combinação de variáveis binárias sobre o número de automóveis disponíveis diariamente no agregado *per capita*, ou seja, a taxa de motorização familiar (veíc./hab);
- D2: Variável contínua distância mais curta entre os centróides de GA da viagem (km);
- Dx1: Variável binária que toma o valor 1 para distâncias mais curtas entre os centróides de GA até 1,0 km.

Dos resultados apresentados no quadro VI.1 destaca-se o ajustamento global do modelo de base ($p\text{-}R^2=0,335$), bastante significativo, sugerindo que estas variáveis têm efetivamente grande capacidade explicativa das escolhas modais. Lembra-se que um $p\text{-}R^2$ de 0,335 neste tipo de modelos equivale a um R^2 de cerca de 0,6 dos modelos de regressão.

Quadro VI. 1 - Resultados do modelo exploratório de base MNL7.40

Variável	MNL7.40					Obs.
	BW	B	OB	M	W	
ASC	-5,457	-1,929	-3,037	-2,870	4,604	
Lc	-2,248	-1,963	-2,027	-1,234	-1,811	- Parâmetros com sinal e grandeza coerentes
NC0	4,461	3,221	2,253	3,157	2,552	
NC1	2,804	1,586	1,246	1,543	0,724	
NC2	2,309	1,115	0,430	0,997	0,393	- Parâmetros com sinal e comportamento coerentes
NC3	1,277	0,528	-0,021***	0,233**	-0,288	
NC4a	-	-	-	-	-	
D2	0,060	0,011	0,047	-0,034	-2,665	- Parâmetros com sinal e grandeza coerentes
Amostra	95426					
p-R ²	0,3349					- Ajustamento considerável
LL	-82136,3					

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Em relação aos parâmetros das variáveis para os diferentes modos, é possível constatar:

- Lc: Parâmetros apresentam sinal coerente, diminuindo a utilidade de todas as alternativas com a existência de licença de condução. Os modos mais penalizados são os modos relacionados com Bus, seguido do modo pedonal,

enquanto o modo moto é o menos penalizado com um parâmetro com quase metade da grandeza do parâmetro mais elevado.

- NCi: Parâmetros com comportamento e sinal coerente. Verifica-se que a não existência de um veículo no agregado tem um impacto significativo na atratividade das alternativas, com grandezas quase duas vezes superiores que os valores da categoria imediatamente seguinte. O comportamento ao longo das categorias é coerente com o senso comum de que o aumento da taxa de motorização do agregado torna as alternativas ao automóvel cada vez menos atrativas.
- D2: Parâmetros com sinais expectáveis, com uma valorização modesta dos modos com Bus e uma desvalorização também modesta da Moto relativamente ao automóvel. Apenas o modo pedonal apresenta um impacto negativo relativo mais acentuado. Estes resultados sugerem que a distância mais curta entre centróides (potencialmente semelhantes para os diferentes modos motorizados) para percursos maiores não tem grande influência na escolha das alternativas ao automóvel com exceção do modo pedonal. Estes resultados corroboram o senso comum que para distâncias curtas o modo pedonal é mais adequado, diminuindo drasticamente a utilidade do mesmo quando a distância aumenta significativamente. Para estas distâncias maiores, o modelo sugere que esta variável não permite distinguir grandes diferenças em relação à utilidade do automóvel.

Testou-se ainda a hipótese de combinar no modelo de base uma variável binária para avaliar a possível importância de incluir uma variável que identifique viagens muito curtas ($Dx1 \leq 1,0\text{km}$) (ver quadro VI.2), de modo a representar o impacto muito particular que a característica “Distância” tende a ter ao nível da escolha do modo pedonal (MNL7.42).

A comparação deste modelo com o modelo de base (MNL7.40) permitiu verificar a estabilidade dos resultados dos parâmetros do modelo de base e que, tal como esperado, a variável Dx1 apenas tem significância, e bem elevada, para o modo pedonal. Os ganhos de ajustamento global são ligeiros mas suficientes para que o teste -2LL considere estatisticamente o modelo com Dx1 melhor que o modelo de base com um intervalo de confiança superior a 95%. Atendendo aos reduzidos ganhos no ajustamento global, à estabilidade dos parâmetros para as alternativas motorizadas, e à aparente falta de relação da variável Dx1 com esses mesmos modos, sugere-se que a introdução no modelo final desta variável apenas seja ponderada na função da utilidade do modo pedonal. A dificuldade de diferenciar as funções de utilidade das diferentes alternativas devido à forma como foi recolhida a informação (apenas para o modo escolhido) e o acentuado aumento de informação

a modelar para a sua transformação em dados para todas as alternativas (6 x 95 426 = 572 556 registros), levou à assunção de que o modelo de referência seria o modelo MNL7.40.

Quadro VI. 2 - Resultados do modelo exploratório base e comparação com os resultados acrescentando a variável Dx1

Variável	MNL7.40					MNL7.42					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-5,457	-1,929	-3,037	-2,870	4,604	-5,461	-1,937	-3,040	-2,867	3,185	- Parâmetros estáveis exceto para W
Lc	-2,248	-1,963	-2,027	-1,234	-1,811	-2,252	-1,967	-2,030	-1,237	-1,884	- Parâmetros com sinal e grandeza estáveis
NC0	4,461	3,221	2,253	3,157	2,552	4,468	3,227	2,259	3,158	2,793	- Parâmetros com sinal e comportamento estáveis
NC1	2,804	1,586	1,246	1,543	0,724	2,808	1,590	1,250	1,542	0,932	
NC2	2,309	1,115	0,430	0,997	0,393	2,312	1,118	0,432	0,996	0,557	
NC3	1,277	0,528	-0,021***	0,233**	-0,288	1,279	0,530	-0,020***	0,232**	-0,237	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,060	0,011	0,047	-0,034	-2,665	0,060	0,012	0,048	-0,034	-2,123	- Parâmetros com sinal e grandeza estáveis exceto W
Dx1						-0,542***	0,288**	-29,8***	-30,2***	2,843	- Parâmetros sem significância exceto para W. Parâmetros para W coerente
Amostra			95426					95426			- Mesma amostra
p-R ²			0,3349					0,3397			- Ajustamento muito semelhante
LL			-82136,3					-81546,5			
Param. GL			35					40			
-2LL			-					5			
Chi ² (α=0,05)			-					1179,6			- Hipótese nula dos modelos serem iguais pode ser recusada
			-					11,1			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

2.2. Modelo de Referência com Inclusão de mais Variáveis Socioeconômicas

Após a definição do modelo de referência começou a avaliar-se a possibilidade de melhorar o modelo através da inclusão de variáveis que nas análises exploratórias isoladas apresentaram resultados coerentes e significativos embora com menor capacidade explicativa do que as de base.

Numa primeira etapa foram incluídas as formulações que se verificou terem maior capacidade explicativa de entre as restantes variáveis socioeconômicas. A comparação dos resultados do

modelo de referência com o modelo adicionado de todas as variáveis socioeconômicas é apresentada no quadro VI.3.

Quadro VI. 3 - Comparação do modelo de referência MNL7.40 com o modelo com a adição das restantes variáveis socioeconômicas

Variável	MNL7.40					MNL7.37					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-5,457	-1,929	-3,037	-2,870	4,604	-8,660	-3,835	-5,220	-7,453	3,121	- Sinais estáveis, grandezas alteraram substancialmente
Lc	-2,248	-1,963	-2,027	-1,234	-1,811	-2,076	-2,054	-1,471	-1,745	-1,925	- Sinais estáveis, grandezas de OB e M alteraram
NC0	4,461	3,221	2,253	3,157	2,552	4,388	3,333	1,887	1,465	2,389	- Sinais estáveis, comportamento coerente e grandezas dos modos OB e M com alterações
NC1	2,804	1,586	1,246	1,543	0,723	2,637	1,662	0,668	0,078***	0,702	
NC2	2,309	1,115	0,430	0,997	0,393	2,246	1,207	0,082***	-0,003***	0,396	
NC3	1,277	0,528	-0,022***	0,233**	-0,288	1,266	0,625	-0,288**	-0,265**	-0,211**	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,060	0,011	0,047	-0,034	-2,665	0,060	0,014	0,050	-0,090	-2,648	- Sinais e grandezza estáveis
Dx1											
R1						0,320	0,276	0,357	0,699	0,308	- Comportamento deixa de ser claramente decrescente
R2						0,325	0,283	0,169	0,377	0,329	
R3						0,419	0,293	0,063***	0,164	0,201	
R4						0,193*	0,298	-0,068***	0,140*	0,101**	
R5						-	-	-	-	-	
ln1						0,213***	-0,692	1,570	2,359	-0,292	- Comportamento dos parâmetros inverte sentido e passa a ser crescente com o aumento da instrução para os modos mais relevantes. Grande variação nas grandezas
ln2						0,451	-0,274	1,660	2,433	-0,067***	
ln3						0,736	0,244	0,946	2,219	0,005***	
ln4						-	-	-	-	-	
ld1						-	-	-	-	-	
ld2						2,846	2,117	1,696	1,553	1,640	- Mantém sinais e comportamentos coerentes. Alteração das grandezas
ld3						1,926	1,507	0,501	2,714	1,182	
ld4						2,107	1,650	0,306	2,433	1,439	
ld5						2,305	1,858	-0,393	2,118	1,600	
Sx						0,581	0,411	-0,176	-0,745	0,166	
Mot						0,062***	-0,478	0,540	3,763	-0,089**	- Sinais e significâncias coerentes
Amostra			95426					95426			- Ajustamento global melhora significativamente
p-R ²			0,3349					0,431			
LL			-82236,3					-70267,2			
Param.			35					100			
GL			-					65			
-2LL			-					23938			- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² $\alpha=0,05$)			-					84,82			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

É possível constatar que o modelo combinado com todas as variáveis socioeconômicas apresenta um ajustamento significativamente melhor e que essa melhoria tem significância estatística (-2LL é cerca de 280 vezes superior ao valor crítico da estatística de chi-quadrado). As constantes específicas apresentam alterações significativas nas suas grandezas e em relação aos parâmetros das diferentes variáveis verificou-se que:

- Variáveis base:

NCi: Os sinais dos parâmetros para estas variáveis mantiveram-se estáveis, e as suas grandezas, com exceção dos modos Outros Bus e Moto, mantiveram-se similares;

LC: Manteve os sinais e a ordem de grandeza não modificou significativamente exceto para os modos Outro Bus e Moto (modos que menos “valorizam” a existência de licença de condução);

D2: Parâmetros com sinal e grandeza similar ao modelo de base, exceto para o modo Moto.

- Outras variáveis socioeconómicas:

Ri: Os sinais dos parâmetros da maioria destas variáveis são coerentes com a normalização utilizada, mas apresentam uma redução significativa relativamente ao modelo isolado. Esta redução na grandeza pode dever-se ao facto de parte da sua capacidade explicativa, demonstrada quando usadas individualmente, poder ser representada através das variáveis de base, nomeadamente o NCi. Uma análise de correlações entre as suas variáveis binárias revelou uma correlação média e positiva (0,370) especialmente entre os rendimentos inferiores (0-500€) e a inexistência de automóveis no agregado ($NC0=1$). Os parâmetros deixam de ter um comportamento claramente decrescente, e para o modo Bus a similaridade de grandezas sugere que aparentemente não há distinção da influência das diferentes categorias. Para os modos Outros Bus e Moto esta variável não apresenta significância estatística, sugerindo que a mesma ou não têm influência para a escolha destes modos, que a amostra não permite avaliar essa influência, ou que relativamente à categoria normalizada não se verificam diferenças significativas.

Ini e Idi: Considerando os parâmetros de ambas as variáveis isoladamente a grandeza dos parâmetros das variáveis instrução apresentou uma acentuada redução, e houve alteração de comportamento e de sinal dos mesmos. O modo Bus+Walk manteve os parâmetros positivos mas passou de um comportamento decrescente para crescente com o aumento do nível de instrução, indiciando que o automóvel terá maior atratividade para os analfabetos que para os indivíduos com algum nível de instrução. Também para o modo Bus se verificou uma alteração de comportamento, passando de uma tendência decrescente para uma tendência crescente, com sinal negativo para as duas primeiras categorias. Este resultado, analisado isoladamente, não é totalmente coerente com o senso comum, uma vez que sugere que o Bus é uma alternativa menos útil para analfabetos e indivíduos

com o ensino básico que o automóvel para indivíduos com ensino superior. Os modos Outros Bus e Moto, apesar da redução na grandeza mantêm sinais positivos e comportamentos relativamente decrescentes com o aumento do nível de instrução. O modo pedonal, além da redução da grandeza, mudou para sinal negativo ou para parâmetros sem significância estatística, sugerindo que o nível de instrução não tem impacto na utilidade deste modo relativamente ao uso do automóvel por indivíduos com formação superior, e que para os analfabetos o modo pedonal é menos atrativo que o automóvel. Lembra-se, no entanto, que a avaliação destes resultados tem que ser enquadrada pelo facto de que a maior parte dos indivíduos analfabetos são crianças até aos 5 anos sem opção de escolha por viajarem com os pais.

Em relação aos parâmetros das variáveis idade, comparativamente ao modelo combinado com a instrução, verifica-se uma alteração moderada da grandeza dos valores, mantendo para a maioria das situações o sinal. O comportamento das mesmas, apesar de não linear, continua coerente com o senso comum, melhorando a coerência principalmente para o modo Pedonal, Moto e Outros Bus.

Quadro VI. 4 - Combinação dos parâmetros das variáveis *Idi* e *Ini* do modelo MNL7.37

		Id 1	Id 2	Id 3	Id 4	Id 5
BW	ln 1	-	3,059	2,139	2,320	2,518
	ln 2	-	3,297	2,377	2,558	2,756
	ln 3	-	3,582	2,662	2,843	3,041
	ln 4	-	-	-	-	-
B	ln 1	-	1,425	0,815	0,958	1,166
	ln 2	-	1,843	1,233	1,376	1,584
	ln 3	-	2,361	1,751	1,894	2,102
	ln 4	-	-	-	-	-
OB	ln 1	-	3,266	2,071	1,876	1,177
	ln 2	-	3,356	2,161	1,966	1,267
	ln 3	-	2,642	1,447	1,252	0,553
	ln 4	-	-	-	-	-
M	ln 1	-	3,912	5,073	4,792	4,477
	ln 2	-	3,986	5,147	4,866	4,551
	ln 3	-	3,772	4,933	4,652	4,337
	ln 4	-	-	-	-	-
W	ln 1	-	1,348	0,890	1,147	1,308
	ln 2	-	1,573	1,115	1,372	1,533
	ln 3	-	1,645	1,187	1,444	1,605
	ln 4	-	-	-	-	-

Para a análise da conjugação dos parâmetros, tal como sugerido no capítulo anterior, realizou-se a conjugação dos parâmetros de ambas as variáveis. A sua influência relativa conjunta, dado serem parâmetros de variáveis binárias, traduz-se pela adição dos seus parâmetros (quadro VI.4). Verifica-se que todas as combinações apresentam parâmetros positivos, coerentes com o esperado atendendo à normalização efetuada. No que diz respeito ao comportamento dos

parâmetros, a tendência crescente dos parâmetros com o aumento do nível de instrução para os modos Bus+Walk, Bus e Walk mantém-se. Esta aparente contradição pode dever-se a uma maior consciência ambiental com o aumento do nível de instrução. Em relação à idade, o comportamento mantém-se com uma maior atratividade das alternativas Bus+Walk, Bus e Walk pelos escalões Id2 e Id5, enquanto a Moto atrai mais os escalões Id3 e Id4.

Sx: Relativamente à variável Sexo do indivíduo, é possível constatar que apesar de haver um reajustamento com diminuição da grandeza dos parâmetros, os mesmos mantiveram o sinal coerente, sugerindo que apenas os modos Outros Bus e, particularmente, Moto são menos valorizados (utilizados) pelas mulheres.

Mot: Relativamente à variável disponibilidade diária de Moto no agregado, esta passa a ter parâmetros com significância estatística, demonstrando que a sua conjugação com outras variáveis introduz informação relevante para o modelo. Essa relevância aparentemente não se verifica para os modos com componente pedonal. Dos resultados verifica-se que a existência de Moto, como espectável, é mais valorizada pelo modo Moto, mas que os utilizadores de Outros Bus também avaliam positivamente esta variável, verificando-se uma diminuição da atratividade do modo Bus.

Tendo em consideração os resultados anteriores, mais concretamente os da variável rendimento líquido mensal do agregado familiar, verifica-se que aparentemente perdeu poder explicativo da utilidade das diferentes alternativas ao automóvel. Também a variável instrução apresentou diferenças significativas que em algumas situações contrariam o senso comum. Em função dos possíveis efeitos de correlação entre estas variáveis e as variáveis de base, considerou-se importante avaliar a possibilidade de os mesmos resultarem especificamente de correlações entre o rendimento do agregado e o nível de instrução (geralmente níveis de instrução superiores estão associados a rendimentos globais superiores) ou destas com as variáveis de base. Assim, testaram-se modelos retirando tanto a variável rendimento como a variável instrução, avaliando-se os resultados obtidos comparativamente ao modelo de referência com todas as variáveis socioeconómicas (quadros VI.5 e VI.6).

É possível verificar que a exclusão das variáveis binárias relativas ao rendimento líquido mensal do agregado não afeta significativamente a capacidade explicativa do modelo, uma vez que o aumento do ajustamento do modelo com a inclusão da mesma, apesar de estatisticamente significativo, é muito limitado. É igualmente possível constatar que os parâmetros das constantes específicas e das restantes variáveis mantêm grandezas e sinais estáveis. Estes resultados sugerem que a informação incluída no rendimento líquido mensal do agregado aparenta já estar incluída noutra variável, muito provavelmente no número de

automóveis *per capita* disponíveis no agregado familiar. A existência de maior número de veículos está normalmente associada a uma maior capacidade financeira do agregado.

Quadro VI. 5 - Comparação do modelo MNL7.48 (sem rendimento) com o modelo de referência com a adição de todas as variáveis socioeconómicas

Variável	MNL7.48					MNL7.37					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,545	-3,723	-5,214	-7,369	3,193	-8,660	-3,835	-5,220	-7,453	3,121	- Parâmetros estáveis
Lc	-2,078	-2,056	-1,489	-1,768	-1,935	-2,076	-2,054	-1,471	-1,745	-1,925	- Parâmetros estáveis
NC0	4,506	3,419	1,979	1,648	2,476	4,388	3,333	1,887	1,465	2,389	- Parâmetros estáveis. Ligeiras alterações de grandeza para M
NC1	2,745	1,740	0,686	0,132***	0,735	2,637	1,662	0,668	0,078***	0,702	
NC2	2,314	1,254	0,086***	0,047***	0,403	2,246	1,207	0,082***	-0,003***	0,396	
NC3	1,228	0,620	-0,339*	-0,330*	-0,267*	1,266	0,625	-0,288**	-0,265**	-0,211**	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,060	0,014	0,050	-0,092	-2,649	0,060	0,014	0,050	-0,090	-2,648	
Dx1											
R1						0,320	0,276	0,357	0,699	0,308	
R2						0,325	0,283	0,169	0,377	0,329	
R3						0,419	0,293	0,063***	0,164	0,201	
R4						0,193*	0,298	-0,068***	0,140*	0,101**	
R5						-	-	-	-	-	
In1	0,291**	-0,625	1,660	2,534	-0,170**	0,213***	-0,692	1,570	2,359	-0,292	- Ligeiras alterações de grandeza.
In2	0,534	-0,208	1,727	2,557	0,046***	0,451	-0,274	1,660	2,433	-0,067***	
In3	0,800	0,291	0,988	2,296	0,080***	0,736	0,244	0,946	2,219	0,005***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
Id2	2,839	2,113	1,702	1,561	1,637	2,846	2,117	1,696	1,553	1,640	
Id3	1,954	1,534	0,535	2,752	1,217	1,926	1,507	0,501	2,714	1,182	
Id4	2,123	1,666	0,337	2,475	1,453	2,107	1,650	0,306	2,433	1,439	
Id5	2,312	1,868	-0,296	2,264	1,619	2,305	1,858	-0,393	2,118	1,600	
Sx	0,581	0,412	-0,173	-0,747	0,169	0,581	0,411	-0,176	-0,745	0,166	- Parâmetros estáveis
Mot	0,064***	-0,478	0,520	3,723	-0,090**	0,062***	-0,478	0,540	3,763	-0,089**	- Parâmetros estáveis
Amostra			95426					95426			- Ajustamentos muito próximos
p-R ²			0,430					0,431			
LL			-70377,5					-70267,2			
Param.			80					100			- Hipótese nula dos modelos serem iguais pode ser recusada
GL								20			
-2LL								220,6			
Chi ² ($\alpha=0,05$)								31,4			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

No estudo comparativo entre o modelo sem o nível de instrução e o modelo com todas as variáveis socioeconómicas, verifica-se que a exclusão da variável instrução leva a um menor ajustamento do modelo, relativamente ao mais completo bem como do modelo sem a variável rendimento. Esta diferença, apesar de reduzida, é estatisticamente significativa tornando ambos os modelos melhores que o modelo sem instrução. Em relação às constantes específicas, verificam-se alterações acentuadas de grandeza para os modos Bus, Outros Bus e Moto. É ainda possível constatar que as variáveis binárias sobre o rendimento líquido do agregado continuam com uma tendência pouco coerente, principalmente para os modos Bus e Walk, mas que mantêm sinal e grandezas semelhantes.

Quadro VI. 6 - Comparação do modelo de referência MNL7.49 (sem instrução) com o modelo com a adição de todas as variáveis socioeconômicas

Variável	MNL7.49					MNL7.37					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,405	-4,253	-3,897	-5,376	2,953	-8,660	-3,835	-5,220	-7,453	3,121	- Grande alteração de grandeza para B, M e OB
Lc	-2,053	-1,935	-1,621	-1,799	-1,895	-2,076	-2,054	-1,471	-1,745	-1,925	- Parâmetros estáveis
NC0	4,425	3,268	2,117	1,650	2,372	4,388	3,333	1,887	1,465	2,389	- Parâmetros estáveis exceto para OB e M.
NC1	2,681	1,601	0,907	0,272*	0,691	2,637	1,662	0,668	0,078***	0,702	
NC2	2,292	1,159	0,299*	0,185**	0,388	2,246	1,207	0,082***	-0,003***	0,396	
NC3	1,309	0,593	-0,133***	-0,156***	-0,220**	1,266	0,625	-0,288**	-0,265**	-0,211**	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,060	0,013	0,050	-0,090	-2,647	0,060	0,014	0,050	-0,090	-2,648	- Parâmetros estáveis
Dx1											- Parâmetros relativamente estáveis
R1	0,325	0,204	0,440	0,798	0,285	0,320	0,276	0,357	0,699	0,308	
R2	0,334	0,225	0,255	0,473	0,314	0,325	0,283	0,169	0,377	0,329	
R3	0,435	0,266	0,122*	0,240	0,194	0,419	0,293	0,063***	0,164	0,201	
R4	0,204*	0,289	-0,034***	0,206	0,097**	0,193*	0,298	-0,068***	0,140*	0,101**	
R5	-	-	-	-	-	-	-	-	-	-	
In1						0,213***	-0,692	1,570	2,359	-0,292	
In2						0,451	-0,274	1,660	2,433	-0,067***	
In3						0,736	0,244	0,946	2,219	0,005***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Ligeiras alterações de grandeza
Id2	3,032	2,441	1,682	1,580	1,771	2,846	2,117	1,696	1,553	1,640	
Id3	2,098	1,843	0,454	2,727	1,303	1,926	1,507	0,501	2,714	1,182	
Id4	2,238	1,890	0,351	2,474	1,550	2,107	1,650	0,306	2,433	1,439	
Id5	2,413	2,051	-0,346	2,159	1,690	2,305	1,858	-0,393	2,118	1,600	
Sx	0,585	0,434	-0,211	-0,767	0,171	0,581	0,411	-0,176	-0,745	0,166	- Ligeira alteração de grandeza para OB
Mot	0,058***	-0,507	0,579	3,816	-0,089**	0,062***	-0,478	0,540	3,763	-0,089**	- Parâmetros estáveis
Amostra			95426					95426			- Ligeira melhoria do ajustamento global com introdução de Ini
p-R ²			0,428					0,431			
LL			-70679,5					-70267,2			
Param.			85					100			
GL			-					15			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL			-					824,6			
Chi ² (α=0,05)			-					25,0			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Para as restantes variáveis regista-se uma estabilidade de sinais e, para a maioria das variáveis, uma estabilidade de grandeza.

Dos resultados anteriores verifica-se que a inclusão das restantes variáveis socioeconômicas no modelo de referência melhora o ajustamento global mas revela alterações significativas e retira coerência ao comportamento das variáveis rendimento e nível de instrução.

Aparentemente a informação contida na variável rendimento líquido mensal do agregado familiar pode já estar considerada através de outra variável, provavelmente NCi, e a sua exclusão, que foi assumida, não retira grande capacidade explicativa ao modelo, permitindo pelo contrário limitar a sua complexidade e o número global de variáveis. Uma possível explicação para este facto pode resultar da forma como foi definido o rendimento mensal,

resultante da relutância da maioria das pessoas em o explicitar, implicando a sua distinção por gamas de valores e, portanto, oferecendo menos precisão na informação. O número de automóveis no agregado, pelo contrário, além de potencialmente correlacionado com o nível económico, é de explicitação relativamente fácil e fidedigna.

Quanto à variável nível de instrução, a sua inclusão melhora ligeiramente a capacidade explicativa do modelo em relação ao modelo com o rendimento familiar, mas os seus parâmetros isoladamente apresentam uma tendência crescente com o aumento da categoria para as principais alternativas (Bus e Walk). A sua análise combinada mantém esse comportamento mas com sinal positivo, podendo dever-se a uma maior consciência ambiental dos cidadãos com mais instrução.

Assim, os parâmetros das variáveis sobre o rendimento familiar mensal e do nível de instrução perdem coerência em termos de comportamento. A sua introdução em separado no modelo revelou que o rendimento aparentemente não introduz informação adicional relevante. A estabilidade de resultados com a variável nível de instrução comparativamente aos do modelo com o rendimento, não justificam a inclusão desta última. O comportamento dos parâmetros do nível de instrução sugere, ainda assim, uma análise cuidada nos modelos mais complexos.

Assim, o modelo de referência proposto para o conjunto das variáveis socioeconómicas é o modelo MNL7.48 (modelo socioeconómico base).

2.3. Modelo de Referência com Inclusão da Variável Genérica Tempo Total de Viagem

Numa segunda etapa analisou-se a possibilidade de acrescentar mais informação ao modelo através da combinação do modelo de base socioeconómico com o tempo total de viagem. Das análises preliminares constatou-se que a versão logarítmica da variável apresentava maior capacidade explicativa que a sua versão simples. A formulação analítica desta variável, genérica e portanto com valores distintos para os diferentes modos, levou a uma redução da amostra para cerca de um terço e à impossibilidade de comparação dos seus ajustamentos globais com o modelo de referência por resultarem de amostras distintas. Para minorar este problema aplicou-se o modelo de referência à base de dados com tempos de viagens para todos os modos e comparou-se a estabilidade dos parâmetros com o modelo de referência com a base de dados completa. Os resultados desses modelos e a comparação dos parâmetros são apresentados no quadro VI.7.

É possível constatar que os parâmetros apresentaram algumas alterações de grandeza e de comportamento. Estas alterações são, no entanto, razoáveis atendendo à diminuição da amostra e à provável purga de dados relevantes para determinadas categorias das variáveis analisadas. Admitindo que as alterações se deveram a essa redução da amostra, analisou-se a

introdução da variável genérica tempo de viagem, cujos resultados são apresentados no quadro VI.8.

Quadro VI. 7 - Comparação dos resultados do modelo de referência MNL7.48 com o mesmo modelo com a base de dados reduzida MNL7.48a

Variável	MNL7.48					MNL7.48a					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,545	-3,723	-5,214	-7,369	3,193	-7,847	-4,296	-4,806	-6,327	3,407	- Alteração da grandeza mas estabilidade de sinais
Lc	-2,078	-2,056	-1,489	-1,768	-1,935	-2,323	-2,264	-1,608	-1,721	-2,247	- Grandezas alteraram ligeiramente, sinais estáveis
NC0	4,506	3,419	1,979	1,648	2,476	3,392	2,870	1,912	1,268	2,429	- Alterações de grandeza significativas. Comportamento coerente
NC1	2,745	1,740	0,686	0,132***	0,735	1,977	1,434	0,861	-0,102***	0,580	
NC2	2,314	1,254	0,086***	0,047***	0,403	1,556	0,963	0,163***	-0,331**	0,290**	
NC3	1,228	0,620	-0,339*	-0,330*	-0,267*	0,364***	0,342**	-0,599*	-0,566	-0,582	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis exceto para B e BW
D2	0,060	0,014	0,050	-0,092	-2,649	0,045	0,068	0,051	-0,106	-2,877	
Dx1											
R1											
R2											
R3											
R4											
R5											
ln1	0,291**	-0,625	1,660	2,534	-0,170**	0,765*	-0,079***	1,257	2,416	-0,189***	- Grandezas com alterações para a maioria das variáveis. Comportamentos estáveis exceto M e W.
ln2	0,534	-0,208	1,727	2,557	0,046***	0,833	0,066***	1,304	2,050	0,085***	
ln3	0,800	0,291	0,988	2,296	0,080***	0,878	0,380	0,694*	1,771	0,074***	
ln4	-	-	-	-	-	-	-	-	-	-	
ld1	-	-	-	-	-	-	-	-	-	-	- Alterações de grandeza nos modos mais relevantes BW, B e W. Comportamento estável exceto para a ld5 para BW e B
ld2	2,839	2,113	1,702	1,561	1,637	3,008	2,431	1,706	1,582	1,857	
ld3	1,954	1,534	0,535	2,752	1,217	1,882	1,741	0,716	2,696	1,579	
ld4	2,123	1,666	0,337	2,475	1,453	2,231	1,875	0,559	2,591	1,737	
ld5	2,312	1,868	-0,296	2,264	1,619	2,164	1,572	-0,373*	1,880	1,769	
Sx	0,581	0,412	-0,173	-0,747	0,169	0,631	0,364	-0,300	-0,968	0,073***	- Alterações de grandeza. Sinais estáveis. Para W perde significância
Mot	0,064***	-0,478	0,520	3,723	-0,090**	0,686	-0,080**	0,622	3,578	-0,111***	- Alterações significativas para BW e B. Sinais estáveis.
Amostra	95426					30559					- Amostra reduzida a 1/3
p-R ²	0,430					0,416					- Não é possível comparar ajustamentos
LL	-70377,5					-23181,3					

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

É possível constatar que os parâmetros e as constantes específicas apresentam valores estáveis, com poucas diferenças entre modelos. O ajustamento global do modelo com a inclusão do tempo de viagem é apenas ligeiramente melhor, mas estatisticamente significativo. As variáveis base apresentam comportamento e sinais coerentes. Os resultados para as restantes variáveis socioeconómicas revelam que a combinação nível de instrução *versus* escalão etário (quadro VI.9) mantém o comportamento, pouco expectável, crescente da utilidade das alternativas Bus+Walk e Bus com o aumento da instrução, e que a variável grupo etário apresenta um comportamento diferente para os cidadãos mais velhos (ld5) que avaliam as alternativas (exceto Walk) de forma menos atrativa que no modelo de referência

anterior. Esta alteração de comportamento para as variáveis do grupo etário pode dever-se à purga da base de dados de parte deste escalão mais idoso.

Atendendo aos resultados anteriores, e apesar da severa redução da amostra imposta pelo uso da variável genérica tempo de viagem, optou-se por considerar o modelo de referência com a inclusão da mesma MNL7.51.

Quadro VI. 8 - Comparação dos resultados do modelo de referência MNL7.48a com a inclusão do tempo de viagem MNL7.51

Variável	MNL7.48 ^a					MNL7.51					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,847	-4,296	-4,806	-6,327	3,407	-7,187	-3,935	-4,455	-6,236	3,529	- Alterações na grandeza, estabilidade do sinal
Lc	-2,323	-2,264	-1,608	-1,721	-2,247	-2,315	-2,244	-1,600	-1,732	-2,284	- Parâmetros estáveis
NC0	3,392	2,870	1,912	1,268	2,429	3,363	2,852	1,879	1,228	2,393	- Parâmetros estáveis
NC1	1,977	1,434	0,861	-0,102***	0,580	1,955	1,402	0,814	-0,146***	0,519	
NC2	1,556	0,963	0,163***	-0,331**	0,290**	1,529	0,940	0,113***	-0,364**	0,219***	
NC3	0,364***	0,342**	-0,599*	-0,566	-0,582	0,362***	0,319**	-0,635*	-0,584	-0,642	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,045	0,068	0,051	-0,106	-2,877	0,042	0,065	0,038	-0,103	-2,761	
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,765*	-0,079***	1,257	2,416	-0,189***	0,820*	-0,080***	1,238	2,346	-0,181***	- Parâmetros estáveis
In2	0,833	0,066***	1,304	2,050	0,085***	0,856	0,057***	1,282	1,990	0,098***	
In3	0,878	0,380	0,694*	1,771	0,074***	0,905	0,379	0,673*	1,743	0,071***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
Id2	3,008	2,431	1,706	1,582	1,857	3,009	2,418	1,698	1,562	1,933	
Id3	1,882	1,741	0,716	2,696	1,579	1,879	1,699	0,677	2,698	1,609	
Id4	2,231	1,875	0,559	2,591	1,737	2,223	1,824	0,506	2,598	1,753	
Id5	2,164	1,572	-0,373*	1,880	1,769	2,120	1,489	-0,433*	1,871	1,788	
Sx	0,631	0,364	-0,300	-0,968	0,073***	0,632	0,365	-0,302	-0,969	0,064***	- Parâmetros estáveis
MOT	0,686	-0,080**	0,622	3,578	-0,111***	0,677	-0,073**	0,618	3,581	-0,152**	- Parâmetros estáveis
Til								-0,600			- Sinal coerente
Amostra			30559					30559			- Ajustamentos muito semelhantes
p-R ²			0,416					0,421			
LL			-23181,3					-22969,0			
Param.			80					81			
GL			-					1			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL			-					424,6			
Chi ² ($\alpha=0,05$)			-					3,8			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Quadro VI. 9 - Combinação dos parâmetros das variáveis *Idi* e *Ini* do modelo MNL7.51

		Id 1	Id 2	Id 3	Id 4	Id 5
BW	In 1	-	3,829	2,699	3,043	2,940
	In 2	-	3,865	2,735	3,079	2,976
	In 3	-	3,914	2,784	3,128	3,025
	In 4	-	-	-	-	-
B	In 1	-	2,338	1,619	1,744	1,409
	In 2	-	2,475	1,756	1,881	1,546
	In 3	-	2,797	2,078	2,203	1,868
	In 4	-	-	-	-	-
OB	In 1	-	2,936	1,915	1,744	0,805
	In 2	-	2,980	1,959	1,788	0,849
	In 3	-	2,371	1,350	1,179	0,240
	In 4	-	-	-	-	-
M	In 1	-	3,908	5,044	4,944	4,217
	In 2	-	3,552	4,688	4,588	3,861
	In 3	-	3,305	4,441	4,341	3,614
	In 4	-	-	-	-	-
W	In 1	-	1,752	1,428	1,572	1,607
	In 2	-	2,031	1,707	1,851	1,886
	In 3	-	2,004	1,680	1,824	1,859
	In 4	-	-	-	-	-

2.4. Modelo de Referência com Inclusão das Variáveis Relacionadas com a Organização das Atividades ao Longo do Dia

2.4.1. Duração da Estadia no Destino

A conjugação do modelo de base com as variáveis socioeconómicas mais relevantes, com o tempo de viagem e com a duração da estadia limitou a amostra a apenas 19 555 inquéritos com dados para todas as variáveis. Assim o ajustamento global do modelo não pode ser comparado aos modelos anteriores.

Para colmatar esta limitação realizou-se o modelo de referência MNL7.51 para a base de dados reduzida aos inquéritos com dados sobre todas as variáveis em análise (MNL7.51a). No quadro VI.10, podem verificar-se os resultados e comparar os parâmetros das diferentes variáveis.

É possível constatar que a maioria dos parâmetros apresenta valores estáveis. As alterações de grandeza dos parâmetros de algumas variáveis parecem dever-se à redução da amostra, com provável purga de dados relativamente a algumas situações.

O modelo de referência resultante da aplicação à base de dados reduzida foi posteriormente comparado com o modelo incluindo a duração da estadia no destino. Os resultados destes modelos são apresentados no quadro VI.11.

Quadro VI. 10 - Comparação dos resultados do modelo de referência com inclusão do tempo de viagem com o modelo de referência incluindo a duração da estadia no destino

Variável	MNL7.51					MNL7.51a					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,187	-3,935	-4,455	-6,236	3,529	-7,193	-3,701	-4,425	-6,563	3,548	- Maioria dos parâmetros estáveis
Lc	-2,315	-2,244	-1,600	-1,732	-2,284	-2,255	-2,289	-1,684	-1,776	-2,385	- Parâmetros estáveis
NC0	3,363	2,852	1,879	1,228	2,393	3,819	2,851	2,367	1,122	2,604	- Alteração da grandeza dos parâmetros das alternativas menos relevantes e da categoria NC3. Comportamento coerente.
NC1	1,955	1,402	0,814	-0,146***	0,519	2,357	1,330	1,199	-0,325***	0,534*	
NC2	1,529	0,940	0,113***	-0,364**	0,219***	1,940*	0,834	0,454***	-0,552*	0,230***	
NC3	0,362***	0,319**	-0,635*	-0,584	-0,642	0,181***	0,055***	-0,350***	-0,840	-0,717	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,042	0,065	0,038	-0,103	-2,761	0,033*	0,071	0,042	-0,090	-2,949	- Parâmetros relativamente estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,820*	-0,080***	1,238	2,346	-0,181***	1,008*	-0,069***	0,944	2,802	-0,141***	- Maioria dos parâmetros estável. Comportamento para as alternativas menos relevantes é coerente e para as mais relevantes quase não têm significância
In2	0,856	0,057***	1,282	1,990	0,098***	0,857*	0,085***	0,982	1,864	0,169***	
In3	0,905	0,379	0,673*	1,743	0,071***	1,070	0,443	0,399***	1,647	0,222***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros para BW, B e OB relativamente estáveis. Para M e W alterações na grandeza. Comportamento coerente, exceto para Id5 para BW e B
Id2	3,009	2,418	1,698	1,562	1,933	2,935	2,365	1,674	2,203	2,223	
Id3	1,879	1,699	0,677	2,698	1,609	1,866	1,654	0,838	3,367	1,907	
Id4	2,223	1,824	0,506	2,598	1,753	2,134	1,709	0,514	3,376	2,046	
Id5	2,120	1,489	-0,433*	1,871	1,788	1,916	1,248	-0,399**	2,354	2,087	
Sx	0,632	0,365	-0,302	-0,969	0,064***	0,655	0,410	-0,335	-0,950	0,129**	- Parâmetros estáveis e coerentes
MOT	0,677	-0,073**	0,618	3,581	-0,152**	0,707	-0,066***	0,595	3,622	-0,340	- Maioria dos parâmetros estáveis e coerentes
Til			-0,600					-0,900			- Alteração da grandeza e sinal coerente
De1											
De2											
De3											
De4											
De5											
nV											
Tr											
Rg											
Es											
Lz											
Cs											
Amostra			30559					19555			- Redução da amostra para 2/3
p-R ²			0,421					0,445			- Não é possível comparar
LL			-22969,0					-14021,9			
Param.			81					81			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

É possível constatar que, apesar de algumas alterações de grandeza das constantes específicas, os parâmetros para as diferentes variáveis permaneceram razoavelmente

estáveis, indiciando que a introdução da variável duração da estadia aparentemente não tem correlação com as restantes. O ajustamento global melhorou ligeiramente, e o teste estatístico -2LL permite concluir que essa melhoria é significativa (cerca de 10 vezes superior ao valor crítico de Chi-quadrado) podendo considerar-se que o modelo mais complexo é melhor que o de referência.

Quadro VI. 11 - Comparação dos resultados do modelo de referência com inclusão do tempo de viagem com o modelo de referência incluindo a duração da estadia no destino

Variável	MNL7.51 ^a					MNL7.53					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,193	-3,701	-4,425	-6,563	3,548	-7,589	-4,165	-4,780	-6,646	3,623	- Algumas alterações de grandeza. Sinais estáveis
Lc	-2,255	-2,289	-1,684	-1,776	-2,385	-2,229	-2,259	-1,636	-1,789	-2,379	- Parâmetros estáveis
NC0	3,819	2,851	2,367	1,122	2,604	3,782	2,829	2,295	1,126	2,607	- Parâmetros estáveis
NC1	2,357	1,330	1,199	-0,325***	0,534*	2,302	1,296	1,127	-0,339**	0,509**	
NC2	1,940*	0,834	0,454***	-0,552*	0,230***	1,906*	0,806	0,392***	-0,566	0,214***	
NC3	0,181***	0,055***	-0,350***	-0,840	-0,717	0,168***	0,046***	-0,381***	-0,830	-0,729	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,033*	0,071	0,042	-0,090	-2,949	0,037*	0,073	0,047	-0,086	-2,950	- Parâmetros estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
ln1	1,008*	-0,069***	0,944	2,802	-0,141***	0,966*	-0,075***	0,876*	2,856	-0,137***	- Parâmetros estáveis
ln2	0,857*	0,085***	0,982	1,864	0,169***	0,762**	0,051***	0,856	1,872	0,159***	
ln3	1,070	0,443	0,399***	1,647	0,222***	1,068	0,453	0,400***	1,650	0,226***	
ln4	-	-	-	-	-	-	-	-	-	-	
ld1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis, exceto para alguns de ld5
ld2	2,935	2,365	1,674	2,203	2,223	2,912	2,349	1,672	2,228	2,225	
ld3	1,866	1,654	0,838	3,367	1,907	1,893	1,670	0,884	3,398	1,928	
ld4	2,134	1,709	0,514	3,376	2,046	2,215	1,753	0,611	3,315	2,063	
ld5	1,916	1,248	-0,399**	2,354	2,087	2,217	1,408	0,014***	2,427	2,123	
Sx	0,655	0,410	-0,335	-0,950	0,129**	0,665	0,416	-0,325	-0,940	0,128**	- Parâmetros estáveis
MOT	0,707	-0,066***	0,595	3,622	-0,340	0,708	-0,053***	0,600	3,630	-0,343	- Parâmetros estáveis
Til			-0,900					-0,893			- Parâmetros estável
De1						-	-	-	-	-	- Parâmetros com comportamento crescente, exceto M. Não parece ter significância para o modo pedonal.
De2						-0,101***	0,218	-0,300*	-0,380	-0,127***	
De3						0,255**	0,500	0,093***	0,126**	-0,212**	
De4						0,747	0,666	0,646	0,319	-0,007***	
De5						0,956	0,819	1,122	-0,069***	0,019***	
nV											
Tr											
Rg											
Es											
Lz											
CS											
Amostra			19555					19555			- Ligeira melhoria do ajustamento global
p-R ²			0,445					0,452			
LL			-14021,9					-13857,0			
Param.			81					101			- Hipótese nula dos modelos serem iguais pode ser recusada
GL			-					20			
-2LL			-					329,8			
Chi ² (α=0,05)			-					31,4			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

A variável duração da estadia apresenta valores com sinal e comportamento coerentes. Os modos com Bus ganham utilidade com o aumento da duração da estadia, enquanto o modo

moto para estadias muito longas perde utilidade. O modo pedonal não tem significância estatística para estadias superiores a 60 minutos.

Estes resultados sugerem que as estadias de longa duração, geralmente para emprego ou para a escola, em que a autonomia permitida pelo modo de transporte não é tão relevante, apresentam potencial para serem mais servidas por TC. Aparentemente o modo pedonal serve para realizar viagens de curta duração, associadas provavelmente a viagens a curta distância.

Atendendo aos resultados considerou-se que a introdução da variável duração da estadia permite a introdução de informação adicional ao modelo de referência, pelo que o modelo MNL7.53 passou a ser o modelo de referência.

2.4.2. Número Total de Viagens Diárias

Outra variável relacionada com a organização das atividades ao longo do dia, considerada relevante, foi o número total de viagens realizadas durante o dia. O estudo da sua combinação com o modelo anterior é apresentado no quadro VI.12.

Analisando os resultados verifica-se que o ajustamento do modelo com a inclusão do número total de viagens durante o dia melhora ligeiramente o ajustamento global do modelo de referência, e que essa melhoria é estatisticamente significativa segundo o teste estatístico - 2LL (cerca de 20 vezes superior ao valor crítico).

Em relação às constantes específicas verificam-se alterações de grandeza mas com estabilidade de sinais. Quanto aos parâmetros das diferentes variáveis, constata-se que os mesmos são relativamente estáveis e coerentes para as variáveis:

- Licença de condução (Lc);
- Número de automóveis disponíveis diariamente no agregado *per capita* (NCi);
- Distância mais curta entre centróides (D2);
- Instrução (Ini)
- Sexo (Sx)
- Disponibilidade diária de Moto no agregado (Mot)
- Logaritmo do tempo total de viagem (Til);

A variável Idade (Idi), apresenta ligeiras modificações na grandeza dos parâmetros, nomeadamente para os modos Bus e Outros Bus no escalão Id5 (+65 anos).

A variável duração da estadia no destino (Dei) apresenta ligeiras alterações de grandeza principalmente para as estadias mais prolongadas. Salienta-se a redução da utilidade do modo Bus para estadias muito longas comparativamente a estadias de média e longa duração.

Quadro VI. 12 - Comparação dos resultados do modelo de referência com duração da estadia com o mesmo incluindo a duração da estadia no destino e o número de viagens diárias

Variável	MNL7.53					MNL7.54					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,589	-4,165	-4,780	-6,646	3,623	-6,949	-3,162	-4,079	-6,987	4,418	- Alteração da grandeza. Sinais estáveis
Lc	-2,229	-2,259	-1,636	-1,789	-2,379	-2,158	-2,154	-1,561	-1,801	-2,279	- Parâmetros estáveis
NC0	3,782	2,829	2,295	1,126	2,607	3,766	2,804	2,273	1,140	2,595	- Parâmetros estáveis.
NC1	2,302	1,296	1,127	-0,339**	0,509**	2,300	1,289	1,122	0,327***	0,521*	
NC2	1,906*	0,806	0,392***	-0,566	0,214***	1,913*	0,817	0,399***	-0,554*	0,230***	
NC3	0,168***	0,046***	-0,381***	-0,830	-0,729	0,190***	0,083***	-0,337***	-0,858	-0,662*	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,037*	0,073	0,047	-0,086	-2,950	0,039*	0,076	0,048	-0,084	-2,962	
Dx1											
R1											
R2											- Parâmetros estáveis.
R3											
R4											
R5											
In1	0,966*	-0,075***	0,876*	2,856	-0,137***	0,953*	-0,097***	0,881*	2,829	-0,171***	
In2	0,762**	0,051***	0,856	1,872	0,159***	0,745**	0,035***	0,856	1,835	0,122***	- Parâmetros estáveis.
In3	1,068	0,453	0,400***	1,650	0,226***	1,063	0,456	0,410***	1,654	0,201***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	
Id2	2,912	2,349	1,672	2,228	2,225	2,938	2,400	1,714	2,249	2,269	- Maioria dos parâmetros estáveis. Algumas alterações de grandeza para o modo B e para o escalão Id5, que revelam comportamento diferente.
Id3	1,893	1,670	0,884	3,398	1,928	1,952	1,778	0,968	3,349	2,014	
Id4	2,215	1,753	0,611	3,315	2,063	2,234	1,788	0,651	3,296	2,090	
Id5	2,217	1,408	0,014***	2,427	2,123	2,105	1,261	-0,093***	2,511	1,988	
Sx	0,665	0,416	-0,325	-0,940	0,128**	0,643	0,381	-0,351	-0,930	0,104***	
MOT	0,708	-0,053***	0,600	3,630	-0,343	0,717	-0,044***	0,609	3,620	-0,312	- Parâmetros estáveis
Til			-0,893					-0,883			- Parâmetro estável
De1	-	-	-	-	-	-	-	-	-	-	- Alteração de grandezas e de comportamento.
De2	-0,101***	0,218	-0,300*	-0,380	-0,127***	-0,155***	0,130**	-0,358	-0,357	-0,186**	
De3	0,255**	0,500	0,093***	0,126**	-0,212**	0,193**	0,403	0,025***	0,162**	-0,282	
De4	0,747	0,666	0,646	0,319	-0,007***	0,619	0,453	0,501	0,378	-0,159***	
De5	0,956	0,819	1,122	-0,069***	0,019***	0,638	0,324	0,764	0,104***	-0,365	
nV						-0,147	-0,236	-0,165	0,075	-0,178	- Comportamento coerente exceto para M.
Tr	-	-	-	-	-	-	-	-	-	-	
Rg											
Es											
Lz											
CS											
Amostra			19555					19555			- Ligeira melhoria do ajustamento global
p-R ²			0,452					0,456			
LL			-13857,0					-13744,1			
Param. GL			101					106			
-2LL								5			- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² (α=0,05)								225,8			
								11,1			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

A variável número total de viagens diárias (nV) apresenta parâmetros com sinal negativo coerente, com exceção do modo Moto. Aparentemente o aumento do número de viagens diárias tem um impacto na utilidade da Moto similar ao automóvel, provavelmente devido à autonomia que ambos os modos oferecem.

Os resultados sugerem que a introdução da variável número total de viagens diárias acrescenta alguma informação adicional ao modelo, que apresenta pouca correlação com a maioria das restantes variáveis (apenas na duração da estadia se verificaram alterações relevantes), apontando, assim, que o modelo de referência deverá passar a ser o MNL7.54.

2.4.3. Motivo da Viagem

2.4.3.1. Impacto no Modelo dos Motivos de Viagem

A última variável introduzida sobre a organização das atividades ao longo do dia foi o motivo da viagem. Os resultados deste modelo são comparados com o modelo anterior no quadro VI.13.

Apesar da variação da grandeza das constantes específicas, mais uma vez é possível constatar a estabilidade dos parâmetros para a maioria das variáveis socioeconómicas, apresentando apenas ligeiras alterações nas suas grandezas mas mantendo a coerência de sinal e de comportamento. A variável que maiores alterações apresenta é a variável escalão etário, voltando a apresentar uma maior valorização das alternativas Bus+Walk e Walk pelos mais idosos.

Esta estabilidade também se verificou nas variáveis relacionadas com o afastamento entre as atividades, distância e tempo de viagem.

Os resultados das variáveis relacionadas com a organização das atividades ao longo do dia, no entanto e como se poderia esperar tendo em atenção a sua expectável correlação, não apresentaram a mesma estabilidade. Os parâmetros da variável duração da estadia sofreram alterações da sua grandeza e significância, aparentando já não influenciar as diferentes alternativas, nomeadamente os modos Outro Bus e Moto. Os parâmetros para a influência do número total de viagens diárias apresentam algumas diferenças de grandeza mas mantêm a coerência do sinal. A introdução das variáveis sobre o motivo da viagem alterou os parâmetros das variáveis sobre a duração da estadia, nomeadamente para as estadias mais longas (De ≥ 240 min.).

As variáveis relacionadas com o motivo da viagem, apresentam sinal positivo para quase todas as alternativas, expectável atendendo à normalização realizada, apenas revelando uma diminuição da utilidade dos modos Moto e Walk para viagens para a escola e para Compras/Serviços respetivamente. Os comportamentos podem ser considerados coerentes com o senso comum de que as alternativas ao automóvel são mais atrativas para as atividades regulares do que para atividades de lazer. Os modos com Bus são os mais valorizados para as viagens para a escola.

Quadro VI. 13 - Comparação dos resultados do modelo de referência MNL7.54 com o modelo de referência incluindo o motivo da viagem

Variável	MNL7.54					MNL7.55					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-6,949	-3,162	-4,079	-6,987	4,418	-7,750	-3,629	-4,997	-6,822	4,384	- Alteração das grandezas para BW, B e OB. Sinais estáveis.
Lc	-2,158	-2,154	-1,561	-1,801	-2,279	-2,193	-2,184	-1,584	-1,892	-2,288	- Parâmetros estáveis
NC0	3,766	2,804	2,273	1,140	2,595	3,808	2,821	2,318	1,138	2,656	- Parâmetros estáveis
NC1	2,300	1,289	1,122	0,327***	0,521*	2,306	1,275	1,133	-0,377**	0,545*	
NC2	1,913*	0,817	0,399***	-0,554*	0,230***	1,919*	0,802	0,401***	-0,588	0,246***	
NC3	0,190***	0,083***	-0,337***	-0,858	-0,662*	0,192***	0,072***	-0,323***	-0,952	-0,642*	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,039*	0,076	0,048	-0,084	-2,962	0,042	0,080	0,054	-0,080	-2,971	
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,953*	-0,097***	0,881*	2,829	-0,171***	1,042*	0,015***	1,100	2,813	0,108***	- Maioria dos parâmetros estáveis.
In2	0,745**	0,035***	0,856	1,835	0,122***	0,734**	0,049***	0,877	1,831	0,160***	
In3	1,063	0,456	0,410***	1,654	0,201***	1,045	0,460	0,403***	1,718	0,215***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Alteração de grandezas e de comportamento para BW e W
Id2	2,938	2,400	1,714	2,249	2,269	2,721	2,232	1,524	2,341	2,183	
Id3	1,952	1,778	0,968	3,349	2,014	2,038	1,808	1,158	2,809	2,086	
Id4	2,234	1,788	0,651	3,296	2,090	2,333	1,837	0,862	2,736	2,193	
Id5	2,105	1,261	-0,093***	2,511	1,988	2,359	1,469	0,315***	2,259	2,220	
Sx	0,643	0,381	-0,351	-0,930	0,104***	0,669	0,408	-0,319	-0,946	0,153**	- Parâmetros estáveis
Mot	0,717	-0,044***	0,609	3,620	-0,312	0,716	-0,037***	0,615	3,654	-0,305	- Parâmetros estáveis
Til			-0,883					-0,886			- Parâmetro estável
De1	-	-	-	-	-	-	-	-	-	-	- Alterações de grandeza principalmente para De4 e De5.
De2	-0,155***	0,130**	-0,358	-0,357	-0,186**	-0,091***	0,178	-0,289*	-0,317	-0,152***	
De3	0,193**	0,403	0,025***	0,162**	-0,282	0,187**	0,345	-0,079***	0,047***	-0,386	
De4	0,619	0,453	0,501	0,378	-0,159***	0,436	0,223	0,142***	0,181**	-0,355	
De5	0,638	0,324	0,764	0,104***	-0,365	0,402	0,030***	0,351	-0,214**	-0,601	
nV	-0,147	-0,236	-0,165	0,075	-0,178	-0,200	-0,280	-0,204	0,041	-0,204	- Ligeiras alterações de grandeza mas sinais estáveis
Tr						1,027	0,809	1,128	1,041	0,169***	- As alternativas com Bus mais valorizadas para viagens para a escola. Comportamento coerente
Rg						1,440	0,981	1,322	0,691	0,246**	
Es						1,437	1,056	1,632	-0,854	-0,370*	
Lz						-	-	-	-	-	
CS						0,716	0,354	0,566	0,241**	-0,511	
Amostra			19555					19555			- Melhoria do ajustamento global
p-R ²			0,452					0,465			
LL			-13857,0					-13526,3			
Param.			106					126			
GL								20			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL								661,4			
Chi ² ($\alpha=0,05$)								31,4			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

O facto de a variável duração da estadia apresentar variações acentuadas na grandeza dos seus parâmetros com a introdução da variável motivo da viagem, sugere assim, de facto, que a sua capacidade explicativa poderá ter informação correlacionada com a introduzida com o motivo da viagem. Esta hipótese foi estudada através da análise das correlações entre estas variáveis e da comparação de resultados do modelo sem a variável duração da estadia.

2.4.3.2. Correlação entre Motivo da Viagem e Duração da Estadia

No sentido de avaliar se o pressuposto enunciado no ponto anterior era real, calcularam-se as correlações entre as variáveis binárias para os diferentes motivos de viagem e para as diferentes durações de estadia consideradas. Os resultados desse cálculo apresentam-se no quadro VI.14.

Quadro VI. 14 - Resultados de correlações entre as variáveis binárias sobre o motivo da viagem e sobre a duração das estadias através o *software* SPSS®

Correlações		Variável binária para viagens para trabalho	Variável binária para viagens de regresso a casa	Variável binária para viagens para a escola	Variável binária para viagens em lazer	Variável binária para viagens para compras/serviços
variável binária duração da estadia ≤60min (curta duração)	Pearson Correlation	-0,301 ²	0,276 ²	-0,182 ²	0,070 ¹	0,186 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000
	N	61358	61358	61358	61358	61358
variável binária duração da estadia >60min & ≤120min (média duração)	Pearson Correlation	-0,264 ²	0,138 ²	-0,122 ²	0,150 ²	0,163 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000
	N	61358	61358	61358	61358	61358
variável binária duração da estadia >120min & ≤240min (média duração)	Pearson Correlation	0,005 ¹	-0,003 ¹	-0,019 ¹	0,045 ¹	-0,035 ¹
	Sig. (2-tailed)	0,258	0,445	0,000	0,000	0,000
	N	61358	61358	61358	61358	61358
variável binária duração da estadia >240min & ≤480min (longa duração)	Pearson Correlation	0,230 ²	-0,168 ²	0,166 ²	-0,120 ²	-0,152 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000
	N	61358	61358	61358	61358	61358
variável binária duração da estadia >480min (muito longa duração)	Pearson Correlation	0,332 ³	-0,249 ²	0,160 ²	-0,143 ²	-0,159 ²
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000
	N	61358	61358	61358	61358	61358

* - Sem significância estatística; 1- Sem correlação; 2- Pouca correlação; 3- Média correlação; 4- Elevada correlação

É possível constatar que para durações entre 120 e 240 minutos, as correlações com os motivos da viagem ou não têm significância estatística ou não revelam correlação. Quanto as restantes situações a maioria apresenta significância estatística, mas as correlações são fracas a médias. Os parâmetros apresentam, no entanto, grandezas e sinais com alguma coerência:

- Para as viagens para o trabalho, verificam-se correlações positivas para estadias com duração superior a 240 minutos, e correlações negativas para durações inferiores. Estes resultados sugerem que as viagens para o trabalho estão relacionadas com as estadias mais longas;
- As viagens para a escola apresentam resultados similares ao motivo anterior, mas aparentemente com correlações mais fracas;

- As viagens para casa, revelam correlações positivas para as estadias de menor duração, e correlações negativas para estadias superiores a 240 minutos. Estes resultados sugerem que estas viagens para casa são de curta a média duração, provavelmente relacionadas com retornos a casa à hora do almoço. Recordar-se que as durações das estadias não puderam ser determinadas para as últimas viagens do dia por falta de informação.
- As viagens para lazer e compras ou serviços apresentam resultados semelhantes, revelando correlações fracas mas que sugerem uma correlação positiva para estadias menores que 120 minutos e correlação negativa para estadias superiores a 240 minutos.

Além deste estudo, procurou-se também analisar as diferenças verificadas nos modelos MNL devido à utilização de ambas as variáveis, isoladamente e em conjunto. Assim, realizou-se um modelo sem duração da estadia mas com o motivo da viagem, apresentando-se os seus resultados para comparação no quadro VI.15.

É possível constatar que apesar de algumas alterações na grandeza dos valores, nomeadamente nas constantes específicas e nos parâmetros das variáveis sobre a organização das atividades, que tanto os sinais como as tendências dos valores entre categorias se mantêm. Estes resultados, nomeadamente a perda de significância de parte dos parâmetros para a duração da estadia e a estabilidade de sinais e de tendências das restantes variáveis no modelo apenas com os motivos da viagem, sugerem que a informação da duração da estadia não melhora significativamente a capacidade explicativa do modelo, e corrobora que provavelmente parte dela já estará incluída na informação introduzida pelo motivo da viagem. O motivo da viagem, teoricamente, incorpora indiretamente uma gama de valores característicos para a duração de determinada atividade. É possível ainda relacionar o motivo da viagem com outras características importantes de diferentes tipos de viagens, nomeadamente a frequência de ocorrência das mesmas, a sua obrigatoriedade, a sua estabilidade em termos de horários e de distribuição geográfica, características que dificilmente poderão ser representadas apenas pela variável duração da estadia, considerando-se que a variável motivo poderá ser bastante mais relevante em termos de capacidade explicativa. Como exemplo desta questão referem-se as viagens pendulares para o trabalho e para a escola que se apresentam com características de ocorrência diária, de obrigatoriedade, e com horários e localização geográfica geralmente estáveis.

Quadro VI. 15 - Comparação dos resultados do modelo de referência MNL7.55 com o modelo sem duração da estadia no destino

Variável	MNL7.56a					MNL7.55					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,539	-3,487	-4,900	-6,935	4,075	-7,750	-3,629	-4,997	-6,822	4,384	- Grandezas relativamente estáveis. Sinais estáveis
Lc	-2,190	-2,189	-1,584	-1,883	-2,276	-2,193	-2,184	-1,584	-1,892	-2,288	- Parâmetros estáveis
NC0	3,815	2,800	2,349	1,137	2,625	3,808	2,821	2,318	1,138	2,656	- Parâmetros estáveis
NC1	2,311	1,254	1,155	-0,375**	0,525*	2,306	1,275	1,133	-0,377**	0,545*	
NC2	1,919*	0,784	0,422***	-0,591	0,230***	1,919*	0,802	0,401***	-0,588	0,246***	
NC3	0,194***	0,052***	-0,297***	-0,974	-0,647*	0,192***	0,072***	-0,323***	-0,952	-0,642*	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,042	0,079	0,053	-0,082	-2,971	0,042	0,080	0,054	-0,080	-2,971	- Parâmetros estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
ln1	1,062*	-0,004***	1,158	2,778	-0,146***	1,042*	0,015***	1,100	2,813	0,108***	- Parâmetros estáveis
ln2	0,762**	0,027***	0,924	1,816	0,135***	0,734**	0,049***	0,877	1,831	0,160***	
ln3	1,045	0,457	0,409***	1,712	0,193***	1,045	0,460	0,403***	1,718	0,215***	
ln4	-	-	-	-	-	-	-	-	-	-	
ld1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
ld2	2,698	2,249	1,511	2,324	2,200	2,721	2,232	1,524	2,341	2,183	
ld3	2,016	1,822	1,145	2,784	2,095	2,038	1,808	1,158	2,809	2,086	
ld4	2,308	1,869	0,836	2,723	2,209	2,333	1,837	0,862	2,736	2,193	
ld5	2,351	1,529	0,256***	2,273	2,233	2,359	1,469	0,315***	2,259	2,220	
Sx	0,669	0,409	-0,322	-0,945	0,149**	0,669	0,408	-0,319	-0,946	0,153**	- Parâmetros estáveis
MOT	0,720	-0,045***	0,627	3,651	-0,305	0,716	-0,037***	0,615	3,654	-0,305	- Parâmetros estáveis
Til			-0,890					-0,891			- Parâmetro estável
De1						-	-	-	-	-	
De2						-0,091***	0,178	-0,289*	-0,317	-0,152***	
De3						0,187**	0,345	-0,079***	0,047***	-0,386	
De4						0,436	0,223	0,142***	0,181**	-0,355	
De5						0,402	0,030***	0,351	-0,214**	-0,601	
nV	-0,230	-0,261	-0,260	0,059	-0,175	-0,200	-0,280	-0,204	0,041	-0,204	- Ligeiras alterações de grandeza. Sinais coerentes
Tr	1,283	0,775	1,397	1,116	-0,031***	1,027	0,809	1,128	1,041	0,169***	- Alterações na grandeza dos parâmetros. Sinais mantêm-se nos parâmetros com significância
Rg	1,440	0,927	1,367	0,702	0,267*	1,440	0,981	1,322	0,691	0,246**	
Es	1,674	1,048	1,844	-0,765	0,212**	1,437	1,056	1,632	-0,854	-0,370*	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,613	0,327	0,491	0,222**	-0,422	0,716	0,354	0,566	0,241**	-0,511	
Amostra			19555					19555			- Ligeira melhoria do ajustamento global
p-R ²			0,463					0,465			
LL			-13572,1					-13526,3			
Param.			106					126			
GL								20			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL								91,6			
Chi ² (α=0,05)								31,4			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Quadro VI. 16 - Comparação dos resultados dos modelos sem duração da estadia no destino (MNL7.56a) e sem motivo da viagem (MNL7.54)

Variável	MNL7.54					MNL7.56a					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-6,949	-3,162	-4,079	-6,987	4,418	-7,539	-3,487	-4,900	-6,935	4,075	- Alterações de grandeza
Lc	-2,158	-2,154	-1,561	-1,801	-2,279	-2,190	-2,189	-1,584	-1,883	-2,276	- Parâmetros estáveis
NC0	3,766	2,804	2,273	1,140	2,595	3,815	2,800	2,349	1,137	2,625	- Parâmetros estáveis
NC1	2,300	1,289	1,122	-0,327***	0,521*	2,311	1,254	1,155	-0,375**	0,525*	
NC2	1,913*	0,817	0,399***	-0,554*	0,230***	1,919*	0,784	0,422***	-0,591	0,230***	
NC3	0,190***	0,083***	-0,337***	-0,858	-0,662*	0,194***	0,052***	-0,297***	-0,974	-0,647*	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,039*	0,076	0,048	-0,084	-2,962	0,042	0,079	0,053	-0,082	-2,971	- Parâmetros estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,953*	-0,097***	0,881*	2,829	-0,171***	1,062*	-0,004***	1,158	2,778	-0,146***	- Parâmetros estáveis. Apenas In1 têm ligeiras alterações para OB
In2	0,745**	0,035***	0,856	1,835	0,122***	0,762**	0,027***	0,924	1,816	0,135***	
In3	1,063	0,456	0,410***	1,654	0,201***	1,045	0,457	0,409***	1,712	0,193***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Grandezas alteraram. Principalmente para Id2 e Id3 e para os modos OB e M.
Id2	2,938	2,400	1,714	2,249	2,269	2,698	2,249	1,511	2,324	2,200	
Id3	1,952	1,778	0,968	3,349	2,014	2,016	1,822	1,145	2,784	2,095	
Id4	2,234	1,788	0,651	3,296	2,090	2,308	1,869	0,836	2,723	2,209	
Id5	2,105	1,261	-0,093***	2,511	1,988	2,351	1,529	0,256***	2,273	2,233	
Sx	0,643	0,381	-0,351	-0,930	0,104***	0,669	0,409	-0,322	-0,945	0,149**	- Parâmetros estáveis
Mot	0,717	-0,044***	0,609	3,620	-0,312	0,720	-0,045***	0,627	3,651	-0,305	- Parâmetros estáveis
Til			-0,886					-0,890			- Parâmetros estável
De1	-	-	-	-	-						
De2	-0,155***	0,130**	-0,358	-0,357	-0,186**						
De3	0,193**	0,403	0,025***	0,162**	-0,282						
De4	0,619	0,453	0,501	0,378	-0,159***						
De5	0,638	0,324	0,764	0,104***	-0,365						
nV	-0,147	-0,236	-0,165	0,075	-0,178	-0,230	-0,261	-0,260	0,059	-0,175	- Ligeira alteração na grandeza para BW e OB. Sinais estáveis
Tr	-	-	-	-	-	1,283	0,775	1,397	1,116	-0,031***	
Rg						1,440	0,927	1,367	0,702	0,267*	
Es						1,674	1,048	1,844	-0,765	0,212**	
Lz						-	-	-	-	-	
CS						0,613	0,327	0,491	0,222**	-0,422	
Amostra			19555					19555			- Aumento do ajustamento global
p-R ²			0,452					0,463			
LL			-13857,0					-13572,1			
Param. GL			106					106			
-2LL											
Chi ² (α=0,05)											

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Atendendo aos resultados considerou-se não se justificar a utilização de ambos os grupos de variáveis no modelo. A opção pelo conjunto a integrar o modelo de referência foi tomada comparando ambos os modelos (ver quadro VI.16), onde foi possível constatar uma ligeira melhoria do ajustamento global com o modelo sem duração da estadia.

A estabilidade de comportamentos e de sinais, apesar de ligeiras alterações de grandeza, parecem permitir assumir que o modelo sem duração da estadia é mais viável para introduzir a informação que aparentemente ambas, pelo menos parcialmente, introduzem no modelo.

Assim, o novo modelo de referência assumido será o MNL7.56a.

2.5. Modelo de Referência Acrescido de Variáveis Representativas da Oferta de Estacionamento na Zona de Atração

2.5.1. Eliminação dos Efeitos Resultantes de Diferenças de Dimensão nas Bases de Dados

As limitações verificadas no levantamento de dados referentes à oferta de estacionamento na AMP, onde apenas se obteve informação do concelho do Porto, combinado com a falta de dados sobre os tempos médios de viagem entre as zonas de GA para todos os modos analisados, levaram a uma acentuada redução da amostra para a sua combinação. Os 962 inquéritos com dados para todas estas variáveis revelaram não ter informação suficiente, ou de qualidade suficiente, para uma modelação tão complexa.

Este problema levou à ponderação de nos modelos seguintes, onde se analisam os efeitos potenciais decorrentes da aplicação de instrumentos de gestão de estacionamento, se retirar a variável tempo total de viagem pelos diferentes modos. A sua exclusão do modelo de referência não revelou variações acentuadas de comportamento dos parâmetros das diferentes variáveis, verificando-se as maiores alterações nas constantes específicas (quadro VI.17) e uma limitada redução do ajustamento global.

É possível verificar que os sinais e comportamento da maioria dos parâmetros são estáveis, havendo essencialmente algumas diferenças na grandeza dos valores. As alterações mais relevantes verificam-se nas constantes específicas e para alguns parâmetros do modo Walk. Estas diferenças, no entanto, não são muito acentuadas sugerindo que a exclusão da variável tempo total de viagem não deverá reduzir nem alterar significativamente a capacidade explicativa dos modelos seguintes não justificando a redução na base de dados que implica. Assim, o modelo de referência assumido para este conjunto de análises foi o MNL7.57a.

Quadro VI. 17 - Modelo de referência MNL7.56a comparado com o modelo sem a variável tempo total médio de viagem por modo

Variável	MNL7.57a					MNL7.56a					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,375	-3,988	-5,337	-7,022	3,989	-7,539	-3,487	-4,900	-6,935	4,075	- Alterações de grandeza para BW, B e OB. Sinais estáveis
Lc	-2,176	-2,200	-1,574	-1,853	-2,216	-2,190	-2,189	-1,584	-1,883	-2,276	- Parâmetros estáveis
NC0	3,776	2,812	2,368	1,184	2,647	3,815	2,800	2,349	1,137	2,625	- Parâmetros estáveis. Algumas alterações de grandeza para W
NC1	2,263	1,283	1,192	-0,321***	0,589*	2,311	1,254	1,155	-0,375**	0,525*	
NC2	1,884*	0,804	0,469***	-0,552*	0,315***	1,919*	0,784	0,422***	-0,591	0,230***	
NC3	0,135***	0,089***	-0,268***	-0,971	-0,573*	0,194***	0,052***	-0,297***	-0,974	-0,647*	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,048	0,084	0,073	-0,088	-3,161	0,042	0,079	0,053	-0,082	-2,971	- Alteração de grandeza ligeira para OB e W. Sinais estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,959*	-0,020***	1,191	2,845	-0,214***	1,062*	-0,004***	1,158	2,778	-0,146***	- Parâmetros estáveis. Apenas para W apresenta alterações de grandeza. Não significância continua
In2	0,719**	0,021***	0,936	1,913	0,086***	0,762**	0,027***	0,924	1,816	0,135***	
In3	1,013*	0,439	0,441***	1,784	0,164***	1,045	0,457	0,409***	1,712	0,193***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Maioria dos parâmetros estáveis.
Id2	2,680	2,303	1,546	2,332	2,097	2,698	2,249	1,511	2,324	2,200	
Id3	1,986	1,902	1,178	2,748	2,016	2,016	1,822	1,145	2,784	2,095	
Id4	2,299	1,963	0,891	2,679	2,161	2,308	1,869	0,836	2,723	2,209	
Id5	2,424	1,661	0,317***	2,303	2,206	2,351	1,529	0,256***	2,273	2,233	
Sx	0,666	0,409	-0,319	-0,939	0,159**	0,669	0,409	-0,322	-0,945	0,149**	- Parâmetros estável
Mot	0,741	-0,050***	0,630	3,646	-0,234**	0,720	-0,045***	0,627	3,651	-0,305	- Parâmetros estáveis
Til								-0,891			
De1											
De2											
De3											
De4											
De5											
nV	-0,245	-0,267	-0,268	0,051	-0,181	-0,230	-0,261	-0,260	0,059	-0,175	- Parâmetros estáveis
Tr	1,320	0,781	1,434	1,142	0,013***	1,283	0,775	1,397	1,116	-0,031***	- Parâmetros estáveis
Rg	1,462	0,933*	1,374	0,741	0,289*	1,440	0,927	1,367	0,702	0,267*	
Es	1,679	1,037	1,799	-0,761	0,189***	1,674	1,048	1,844	-0,765	0,212**	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,639	0,397	0,546	0,227**	-0,371	0,613	0,327	0,491	0,222**	-0,422	
Amostra			19555					19555			- Ligeira melhoria do ajustamento global
p-R ²			0,452					0,463			
LL			-13851,6					-13572,1			
Param. GL			105					106			
-2LL								1			
Chi ² (α=0,05)								559			- Hipótese nula dos modelos serem iguais pode ser recusada
								3,8			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

2.5.2. Modelo de Referência Combinado com Variáveis Binárias sobre a Oferta de Estacionamento

O modelo de referência sem a variável tempo total de viagem foi assim comparado com o mesmo modelo combinado com as variáveis sobre os estacionamentos. Devido à redução da amostra, no entanto, o número de inquéritos com disponibilidade diária de Moto no agregado

foi reduzido a apenas 3% (566 em 18823 inquéritos), impossibilitando o modelo de convergir para uma solução.

Assim, optou-se por retirar também esta variável para os modelos seguintes, apresentando-se o modelo de referência sem esta variável para comparação no quadro VI.18.

Quadro VI. 18 - Comparação dos parâmetros do modelo de referência MNL7.57a com o mesmo modelo sem a variável disponibilidade diária de Moto no agregado

Variável	MNL7.57a					MNL7.58					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,375	-3,988	-5,337	-7,022	3,989	-8,450	-3,968	-5,402	-6,999	4,008	- Parâmetros estáveis
Lc	-2,176	-2,200	-1,574	-1,853	-2,216	-2,165	-2,217	-1,570	-1,859	-2,175	- Parâmetros estáveis
NC0	3,776	2,812	2,368	1,184	2,647	3,961	2,830	2,528	2,496	2,624	- Alterações da grandeza de alguns parâmetros, principalmente para M. Comportamento continua coerente. NC3 sem significância
NC1	2,263	1,283	1,192	-0,321***	0,589*	2,398	1,284	1,308	0,991	0,631	
NC2	1,884*	0,804	0,469***	-0,552*	0,315***	1,962*	0,809	0,538**	0,352**	0,332***	
NC3	0,135***	0,089***	-0,268***	-0,971	-0,573*	0,165***	0,080***	-0,241***	-0,365**	-0,567*	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,048	0,084	0,073	-0,088	-3,161	0,052	0,083	0,076	-0,054	-3,182	- Parâmetros estáveis exceto para M
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,959*	-0,020***	1,191	2,845	-0,214***	1,012*	-0,004***	1,242	3,135	-0,143***	- Parâmetros estáveis exceto para M. Parâmetros com tendência crescente
In2	0,719**	0,021***	0,936	1,913	0,086***	0,791**	0,024***	0,995	2,608*	0,091***	
In3	1,013*	0,439	0,441***	1,784	0,164***	1,033	0,412	0,445***	2,045	0,157***	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis exceto para Id5 para M
Id2	2,680	2,303	1,546	2,332	2,097	2,766	2,340	1,625	2,362	2,123	
Id3	1,986	1,902	1,178	2,748	2,016	2,058	1,952	1,252	2,758	2,083	
Id4	2,299	1,963	0,891	2,679	2,161	2,318	1,980	0,912	2,573	2,183	
Id5	2,424	1,661	0,317***	2,303	2,206	2,356	1,682	0,266***	1,435	2,267	
Sx	0,666	0,409	-0,319	-0,939	0,159**	0,677	0,415	-0,312	-0,898	0,192*	- Parâmetros estáveis
Mot	0,741	-0,050***	0,630	3,646	-0,234**						
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,245	-0,267	-0,268	0,051	-0,181	-0,252	-0,272	-0,274	0,053	-0,205	- Parâmetros estáveis exceto para W
Tr	1,320	0,781	1,434	1,142	0,013***	1,301	0,741	1,412	0,969	0,022***	- Parâmetros estáveis exceto para M
Rg	1,462	0,933*	1,374	0,741	0,289*	1,454	0,904	1,361	0,647	0,309	
Es	1,679	1,037	1,799	-0,761	0,189***	1,654	0,982	1,768	-0,833	0,204**	
Lz	-	-	-	-	-	-	-	-	-	-	
Cs	0,639	0,397	0,546	0,227**	-0,371	0,609	0,360	0,517	0,127***	-0,364	
Amostra			19555					19555			
p-R ²			0,452					0,394			- Ajustamento global piorou
LL			-13851,6					-15299,1			
Param.			105					100			
GL								5			
-2LL								-2895			- Hipótese nula dos modelos serem iguais pode ser recusada, mas modelo de base é melhor
Chi ² ($\alpha=0,05$)								11,1			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

É possível constatar que o modelo perde ajustamento com a exclusão da variável sobre a disponibilidade diária de Moto no agregado. Essa perda de informação é evidente para o modo Moto, cuja maioria dos parâmetros sofre alterações de grandeza. Apesar desta perda de informação, a limitação referida anteriormente impôs a adoção do modelo MNL7.58 como modelo de referência.

No entanto, a diferença nas amostras com dados sobre o estacionamento não permite a comparação dos ajustamentos dos modelos com o de referência, optando-se mais uma vez por recalcular o mesmo para a amostra com dados sobre estacionamento (MNL7.58a). Os resultados desse modelo podem ser comparados com o de referência no quadro VI.19.

É possível verificar uma acentuada alteração do ajustamento global do modelo e uma grande modificação das grandezas de quase todos os parâmetros.

É ainda possível constatar que o comportamento e sinais dos parâmetros da maioria das variáveis socioeconômicas se mantêm estável, com exceção das variáveis escalão etário e nível de instrução. Verifica-se que o nível de instrução analfabeto deixa de ter significância estatística para as alternativas, e os restantes níveis apresentam valorizações muito semelhantes. Estes resultados parecem indiciar que o nível de instrução, para este modelo e esta amostra, não permite diferenciar preferências por nenhuma alternativa.

As variáveis relacionadas com o motivo da viagem também apresentam grandes alterações de comportamento e de sinal, perdendo significância para as alternativas menos relevantes.

Apesar da redução da amostra, a sua grandeza não parece justificar a grande diferença de valores entre parâmetros nos dois modelos. Uma possível justificação será a purga da base de dados dos modos menos relevantes e de grande parte dos inquiridos analfabetos (eventualmente devida a uma distribuição etária muito diferente para este universo, com muito menos jovens até aos 5 anos), que tenderá a obrigar o modelo a recalibrar todos os parâmetros de forma diferente.

Estes resultados podem também dever-se à limitação da amostra a inquéritos com zona de atração no concelho do Porto (onde foi possível realizar o levantamento da oferta de estacionamento), que pode ter contribuído na purga referida. Salienta-se, ainda assim, que apesar da potencial perda de diversidade de informação, o modelo apresenta um melhor ajustamento aos dados sobranes.

O modelo de referência com base de dados reduzida MNL7.58a foi posteriormente comparado com o modelo incluindo a combinação de variáveis binárias sobre estacionamentos. Os resultados desses modelos podem ser comparados no quadro VI.20.

Quadro VI. 19 - Comparação dos parâmetros do modelo de referência MNL7.58 com o mesmo modelo com base de dados reduzida devido às variáveis sobre estacionamento

Variável	MNL7.58					MNL7.58a					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,450	-3,968	-5,402	-6,999	4,008	-8,063	-4,089	-5,157	-7,332	2,660	- Alterações significativas de grandeza exceto para B
Lc	-2,165	-2,217	-1,570	-1,859	-2,175	-1,778	-1,671	-0,962	-1,432	-1,435	- Alterações significativas de grandeza. Sinais coerentes
NC0	3,961	2,830	2,528	2,496	2,624	5,578	4,311	2,825	3,514	3,068	- Alterações significativas da grandeza. Comportamento contínuo coerente. NC3 sem significância exceto para B.
NC1	2,398	1,284	1,308	0,991	0,631	3,442	2,187	2,154	1,427	0,875	
NC2	1,962*	0,809	0,538**	0,352**	0,332***	2,914	1,835	1,108***	0,840**	0,556	
NC3	0,165***	0,080***	-0,241***	-0,365**	-0,567*	1,725**	1,008	1,776*	0,700***	0,178***	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,052	0,083	0,076	-0,054	-3,182	0,080	0,018	0,027***	-0,013***	-2,451	- Alteração significativa na grandeza. Sinais estáveis.
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	1,012*	-0,004***	1,242	3,135	-0,143***	0,489**	0,209**	0,723**	-0,229***	0,298**	- Grandes alterações na grandeza e de comportamento dos parâmetros. In1 perde significância.
In2	0,791**	0,024***	0,995	2,608*	0,091***	1,181	0,500	0,820*	0,556*	0,448	
In3	1,033	0,412	0,445***	2,045	0,157***	1,105	0,482	0,068***	0,598*	0,344	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Grandes alterações na grandeza. Comportamento de Id2 altera e perde relevância para as alternativas com B
Id2	2,766	2,340	1,625	2,362	2,123	1,654	1,914	-0,470**	2,054	1,608	
Id3	2,058	1,952	1,252	2,758	2,083	1,888	2,135	-0,924	2,549	1,587	
Id4	2,318	1,980	0,912	2,573	2,183	1,702	1,996	-1,027	1,148	1,640	
Id5	2,356	1,682	0,266***	1,435	2,267	2,085	2,461	-0,705**	0,159***	1,865	
Sx	0,677	0,415	-0,312	-0,898	0,192*	0,531	0,502	0,464	-1,794	0,397	- Grande alteração da grandeza. OB muda de sinal e W ganha significância
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,252	-0,272	-0,274	0,053	-0,205	-0,210	-0,133	-0,163	0,243	-0,165	- Grande alteração de grandeza exceto para BW. Sinais estáveis
Tr	1,301	0,741	1,412	0,969	0,022***	0,592	0,383	0,377***	0,533*	0,150**	- Alterações de grandeza e de sinal
Rg	1,454	0,904	1,361	0,647	0,309	0,362	0,243	-0,227***	0,076***	0,302	
Es	1,654	0,982	1,768	-0,833	0,204**	0,231***	0,197*	0,036***	-0,515**	0,027	
Lz	-	-	-	-	-	-	-	-	-	-	
Cs	0,609	0,360	0,517	0,127***	-0,364	0,400*	0,460	-1,463*	-0,930**	0,547	
Amostra			19555					18823			- Amostras diferentes
p-R ²			0,394					0,444			
LL			-15299,1					-13299,4			
Param.			100					100			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Quadro VI. 20 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58a com a introdução das variáveis sobre estacionamentos

Variável	MNL7.58a					MNL9.09					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,063	-4,089	-5,157	-7,332	2,660	-8,492	-4,496	-5,016	-7,415	3,233	- Alterações de grandeza. Sinais estáveis
Lc	-1,778	-1,671	-0,962	-1,432	-1,435	-1,817	-1,719	-0,849	-1,472	-1,486	- Ligeiras alterações de grandeza. Sinais estáveis.
NC0	5,578	4,311	2,825	3,514	3,068	5,489	4,212	3,054	3,448	3,112	- Parâmetros estáveis.
NC1	3,442	2,187	2,154	1,427	0,875	3,336	2,069	2,338	1,353	0,871	
NC2	2,914	1,835	1,108**	0,840**	0,556	2,865	1,774	1,197**	0,818**	0,503	
NC3	1,725**	1,008	1,776*	0,700***	0,178***	1,698**	0,941	1,896*	0,698***	0,109***	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,080	0,018	0,027***	-0,013***	-2,451	0,068	-0,005***	0,045**	-0,032***	-2,520	- Alteração de grandeza ligeiras. Sinais estáveis exceto para B.
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,489**	0,209**	0,723**	-0,229***	0,298**	0,499**	0,170***	0,749**	-0,330***	0,266**	- Parâmetros estáveis. Não significância de In1 continua
In2	1,181	0,500	0,820*	0,556*	0,448	1,169	0,430	0,941	0,496**	0,417	
In3	1,105	0,482	0,068***	0,598*	0,344	1,123	0,446	0,170***	0,556*	0,352	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis exceto para Id5.
Id2	1,654	1,914	-0,470**	2,054	1,608	1,648	1,950	-0,526**	2,048	1,657	
Id3	1,888	2,135	-0,924	2,549	1,587	1,844	2,065	-1,044	2,486	1,570	
Id4	1,702	1,996	-1,027	1,148	1,640	1,657	1,928	-1,122	1,101**	1,605	
Id5	2,085	2,461	-0,705**	0,159***	1,865	1,997	2,326	-0,799**	0,081***	1,694	
Sx	0,531	0,502	0,464	-1,794	0,397	0,525	0,515	0,493	-1,767	0,403	- Parâmetros estável
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,210	-0,133	-0,163	0,243	-0,165	-0,203	-0,123	-0,170	0,249	-0,149	- Parâmetros estáveis
Tr	0,592	0,383	0,377***	0,533*	0,150**	0,544	0,344	0,470**	0,476*	0,134**	- Parâmetros estáveis. Pequenas alterações para CS
Rg	0,362	0,243	-0,227***	0,076***	0,302	0,338	0,223	0,328***	0,037***	0,315	
Es	0,231***	0,197*	0,036***	-0,515**	0,027	0,251**	0,276	0,114***	-0,493**	0,142***	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,400*	0,460	-1,463*	-0,930**	0,547	0,297**	0,303	-1,163***	-1,053**	0,433	
OV1						0,116***	0,746	-0,769**	0,190***	-1,077	- Parâmetros para B com sinal e comportamento coerente. Para W sinal não coerente e comportamento não definido. Não significativos para as restantes alternativas
OV2						-0,231***	0,392	-2,360	0,356***	-1,260	
OV3						0,595	0,207	-0,765*	-0,099***	-0,890	
OV4						-	-	-	-	-	
PO1						-	-	-	-	-	Sinais e comportamento coerente. Não significativo para OB e M.
PO2						0,792	0,225*	0,934**	0,066***	0,457	
PO3						0,855	0,540	-0,392***	0,521***	1,419	
Amostra			18823					18823			
p-R ²			0,444					0,459			- Melhoria do ajustamento global
LL			-13299,4					-12954,1			
Param.			100					125			
GL								25			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL								690,6			
Chi ² (α=0,05)								37,7			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

É possível constatar que a maioria das variáveis apresenta sinais e comportamentos estáveis em relação ao modelo de referência (MNL7.58a). O ajustamento global do modelo é melhorado com a introdução das variáveis sobre estacionamento e é estatisticamente melhor (o valor do teste -2LL é cerca de 18 vezes superior ao valor crítico). Verifica-se a perda de significância da variável distância para quase todos os modos exceto Walk, compreensível face à limitação da base de dados às viagens com atração no Porto.

As variáveis sobre a razão da oferta de estacionamento total e o número total de viagens para a zona de atração não apresentam significância estatística para as alternativas menos relevantes (Bus+Walk, Outro Bus e Moto). Apresentam ainda assim um comportamento coerente para o modo Bus e aparentemente incoerente para o modo Walk. Esta incoerência para o modo pedonal, com sinal negativo relativamente à categoria normalizada (com maior oferta de estacionamento), parece sugerir que o modo pedonal é menos valorizado em zonas com menor número de estacionamento disponível nas quais o automóvel, teoricamente, não é tão fácil de utilizar. Esta aparente contradição pode dever-se ao facto de o modo pedonal ser mais utilizado nas zonas urbanas centrais onde as distâncias a percorrer entre atividades são geralmente menores mas onde tradicionalmente a oferta total de estacionamento também costuma ter maior densidade. Esta constatação pode implicar que o sinal negativo da utilidade relativa da densidade de oferta de estacionamento verificada para as viagens pedonais reflita o efeito da sua localização na área urbana e não a efetiva influência da oferta de estacionamento (no quadro VI.21 constata-se que 1/3 das viagens pedonais se efetuam entre zonas centrais do Porto). Este pressuposto sugere a sua remoção da função utilidade do modo pedonal.

As variáveis sobre a percentagem de estacionamentos pagos na zona de atração apresentam sinais e comportamentos coerentes, com um aumento da atratividade das alternativas com o aumento do número de estacionamentos pagos. Salienta-se novamente que os resultados para o modo pedonal podem advir da sua localização na zona urbana (mais estacionamentos pagos nas áreas centrais) do que à real influência do estacionamento pago.

Quadro VI. 21 - Avaliação das zonas de origem e de destino das viagens pedonais

		Zona de Destino da viagem										Total
		Zonas centrais do Porto		Zonas periféricas do Porto		Zonas centrais fora do Porto		Zonas periféricas fora do Porto		Zonas ultraperiféricas fora do Porto		
			%		%		%		%		%	
Zona de Origem da viagem	Zonas centrais do Porto	3878	33,1%	11	0,1%	147	1,3%	1	0,0%	1	0,0%	4038
	Zonas periféricas do Porto	11	0,1%	415	3,5%	33	0,3%	355	3,0%	1026	8,8%	1840
	Zonas centrais fora do Porto	157	1,3%	32	0,3%	653	5,6%	102	0,9%	1	0,0%	945
	Zonas periféricas fora do Porto	0	0,0%	341	2,9%	109	0,9%	114	1,0%	330	2,8%	894
	Zonas ultraperiféricas fora do Porto	2	0,0%	992	8,5%	1	0,0%	278	2,4%	2730	23,3%	4003
Total		4048		1791		943		850		4088		11720

2.5.3. Modelo de Referência Combinado com Variáveis Contínuas sobre a Oferta de Estacionamento

No sentido de avaliar a possibilidade de introduzir a informação relacionada com a oferta de estacionamento de forma mais simples sem grande perda de representatividade, ponderou-se a substituição das variáveis binárias sobre estacionamentos pela sua especificação contínua. Os resultados desses modelos podem ser comparados no quadro VI.22.

É possível constatar que a especificação com combinação de binárias é estatisticamente melhor que a especificação contínua. A maioria dos parâmetros das restantes variáveis apresenta-se estável, com algumas alterações de grandeza nomeadamente nas constantes específicas. O modelo com a especificação contínua perde significância estatística em alguns parâmetros, mas o seu ajustamento global é muito semelhante ao modelo com especificação com combinação de binárias. As variáveis contínuas sobre os estacionamentos apresentam comportamentos e sinais coerentes para a maioria dos parâmetros, tornando o modelo mais simples, mas mantêm, ainda assim, o sinal positivo não coerente para a variável razão de oferta de estacionamento com o número total de viagens atraídas para o modo pedonal.

Esta semelhança ao nível da capacidade explicativa, associada à simplificação do modelo com especificação contínua, e à sua coerência em termos de sinal aliada à estabilidade da maioria dos parâmetros, sugere o modelo MNL9.21 como uma possibilidade credível para modelo de referência.

2.6. Modelo de Referência com Inclusão das Variáveis Relacionadas com Transportes Coletivos

2.6.1. Eliminação dos Efeitos Resultantes de Diferenças de Dimensão nas Bases de Dados

A dificuldade de levantamento de dados relativamente às frequências horárias dos operadores privados, principalmente da zona sul da AMP, limitou o número de inquéritos com dados sobre os mesmos. Esta constatação aliada à redução da amostra devido à duração média da viagem levou à manutenção da assunção anterior de não as combinar na análise com as variáveis sobre os TC. Numa primeira fase optou-se pela análise da sua influência em separado, relativamente às variáveis sobre estacionamento.

Para poder avaliar a estabilidade dos parâmetros aplicou-se o modelo de referência MNL7.58a, sem a inclusão das variáveis sobre estacionamento, à base de dados reduzida devido à falta de dados sobre os TC (MNL7.58b). Os resultados desse modelo podem ser comparados com os resultados anteriores no quadro VI.23.

Quadro VI. 22 - Comparação dos resultados dos modelos de referência com a introdução das variáveis sobre estacionamentos contínuas e binárias

Variável	MNL9.21					MNL9.09					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,737	-3,701	-5,063	-7,230	1,332	-8,492	-4,496	-5,016	-7,415	3,233	- Alterações de grandeza nas alternativas mais relevantes. Sinais estáveis
Lc	-1,802	-1,712	-0,876	-1,450	-1,479	-1,817	-1,719	-0,849	-1,472	-1,486	- Parâmetros estáveis
NC0	5,488	4,186	2,949	3,443	3,086	5,489	4,212	3,054	3,448	3,112	- Parâmetros estáveis
NC1	3,350	2,051	2,278	1,348	0,849	3,336	2,069	2,338	1,353	0,871	
NC2	2,872	1,762	1,140***	0,809**	0,541	2,865	1,774	1,197**	0,818**	0,503	
NC3	1,676**	0,933	1,826*	0,673***	0,145***	1,698**	0,941	1,896*	0,698***	0,109***	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,062	-0,006***	0,044**	-0,031***	-2,532	0,068	-0,005***	0,045**	-0,032***	-2,520	
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,478**	0,197***	0,715**	-0,252***	0,370*	0,499**	0,170***	0,749**	-0,330***	0,266**	- Parâmetros estáveis
In2	1,134	0,437	0,852*	0,514**	0,470	1,169	0,430	0,941	0,496**	0,417	
In3	1,081	0,445	0,117***	0,576*	0,360	1,123	0,446	0,170***	0,556*	0,352	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
Id2	1,677	1,963	-0,520**	2,083	1,697	1,648	1,950	-0,526**	2,048	1,657	
Id3	1,823	2,064	-0,950	2,494	1,624	1,844	2,065	-1,044	2,486	1,570	
Id4	1,648	1,935	-1,031	1,117**	1,662	1,657	1,928	-1,122	1,101**	1,605	
Id5	1,982	2,325	-0,682**	0,081***	1,768	1,997	2,326	-0,799**	0,081***	1,694	- Parâmetros estáveis
Sx	0,533	0,510	0,471	-1,777	0,397	0,525	0,515	0,493	-1,767	0,403	
Mot											
Til											
De1											
De2											
De3											
De4											- Parâmetros estáveis
De5											
nV	-0,200	-0,120	-0,169	0,246	-0,156	-0,203	-0,123	-0,170	0,249	-0,149	
Tr	0,554	0,343	0,408***	0,487*	0,135**	0,544	0,344	0,470**	0,476*	0,134**	- Parâmetros estáveis
Rg	0,345	0,221	0,274***	0,049	0,313	0,338	0,223	0,328***	0,037***	0,315	
Es	0,272**	0,265	0,038***	-0,489**	0,112***	0,251**	0,276	0,114***	-0,493**	0,142***	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,290**	0,304	-1,230**	-1,050**	0,457	0,297**	0,303	-1,163***	-1,053**	0,433	- Sinal coerente exceto para W e OB
OV	-3,545	-5,021	0,553***	-1,547***	7,259						
PO	0,570	0,864	-1,304	0,655**	1,947						
OV1						0,116***	0,746	-0,769**	0,190***	-1,077	- Sinal e comportamento para B coerente. W com tendência coerente
OV2						-0,231***	0,392	-2,360	0,356***	-1,260	
OV3						0,595	0,207	-0,765*	-0,099***	-0,890	
OV4						-	-	-	-	-	
PO1						-	-	-	-	-	- Sinais e comportamento das alternativas mais relevantes coerente
PO2						0,792	0,225*	0,934***	0,066***	0,457	
PO3						0,855	0,540	-0,392***	0,521***	1,419	
Amostra			18823					18823			
p-R ²			0,456					0,459			- Hipótese nula dos modelos serem iguais pode ser recusada
LL			-13022,0					-12954,1			
Param.			110					125			
GL								15			
-2LL								135,8			
Chi ² (α=0,05)								25,0			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Quadro VI. 23 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58a com a redução da base de dados aos inquéritos com dados sobre os TC e estacionamento

Variável	MNL7.58a					MNL7.58b					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-8,063	-4,089	-5,157	-7,332	2,660	-7,693	-4,057	-5,237	-7,930	2,687	- Parâmetros estáveis exceto para BW e M
Lc	-1,778	-1,671	-0,962	-1,432	-1,435	-1,795	-1,711	-1,024	-1,345	-1,458	- Parâmetros estáveis
NC0	5,578	4,311	2,825	3,514	3,068	5,413	4,191	2,693	4,458	2,989	- Parâmetros com alterações de grandeza exceto para OB e W. Comportamento com tendência correta
NC1	3,442	2,187	2,154	1,427	0,875	3,232	2,059	2,124	2,450	0,818	
NC2	2,914	1,835	1,108***	0,840**	0,556	2,678	1,761	1,016***	1,826*	0,542	
NC3	1,725**	1,008	1,776*	0,700***	0,178***	1,201**	0,853	1,692**	1,823*	0,143***	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,080	0,018	0,027***	-0,013***	-2,451	0,106	0,023	0,071	-0,026***	-2,397	- Parâmetros estáveis exceto para OB e BW. Sinais corretos. Ambas valorizaram
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,489**	0,209**	0,723**	-0,229***	0,298**	0,241***	0,160***	0,651***	-0,391***	0,241***	- Parâmetros com significância estáveis
In2	1,181	0,500	0,820*	0,556*	0,448	1,192	0,513	0,796*	0,474**	0,442	
In3	1,105	0,482	0,068***	0,598*	0,344	1,114	0,533	-0,142***	0,502**	0,363	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros das alternativas mais relevantes estáveis. Restantes apresentam alterações de grandeza mas mantêm comportamento
Id2	1,654	1,914	-0,470**	2,054	1,608	1,446	1,984	-0,440**	1,749	1,560	
Id3	1,888	2,135	-0,924	2,549	1,587	1,642	2,210	-0,880*	2,153	1,548	
Id4	1,702	1,996	-1,027	1,148	1,640	1,444	2,047	-0,875	0,565***	1,602	
Id5	2,085	2,461	-0,705**	0,159***	1,865	1,819	2,499	-0,598***	-1,162***	1,776	
Sx	0,531	0,502	0,464	-1,794	0,397	0,492	0,459	0,456	-1,633	0,353	- Ligeira alteração da grandeza. Alteração maior para M. Sinais estáveis
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,210	-0,133	-0,163	0,243	-0,165	-0,222	-0,150	-0,141*	0,251	-0,168	- Parâmetros estáveis
Tr	0,592	0,383	0,377***	0,533*	0,150**	0,596	0,406	0,305***	0,557*	0,164**	- Parâmetros com significância estáveis.
Rg	0,362	0,243	-0,227***	0,076***	0,302	0,372	0,243	0,160***	0,041***	0,302	
Es	0,231***	0,197*	0,036***	-0,515**	0,027	0,218***	0,253	0,089***	-0,749*	0,048***	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,400*	0,460	-1,463*	-0,930**	0,547	0,190***	0,425	-1,542**	-0,747***	0,516	
Amostra			18823					17005			- Amostras diferentes
p-R ²			0,444					0,446			
LL			-13299,4					-12021,0			
Param.			100					100			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

É possível verificar que a redução na amostra não foi muito acentuada (cerca de 10%), e que a maioria dos parâmetros se mostrou estável.

As alterações mais relevantes verificaram-se na valorização do parâmetro sobre a taxa de motorização familiar para o modo Moto. Esta valorização pode dever-se a uma maior relevância deste modo na base de dados reduzida.

A variável distância valoriza-se ligeiramente para as alternativas com o modo Bus.

O nível de instrução apresenta algumas alterações mas apenas nos parâmetros que não têm significância estatística, mantendo ainda assim sinais e comportamento.

A variável escalão etário apresenta estabilidade dos parâmetros das alternativas mais relevantes, e algumas alterações de grandeza nos restantes, mas mantém comportamentos.

A variável sexo apresenta ligeiras reduções de grandeza mas mantém sinais coerentes.

Os motivos de viagem mantêm comportamentos apesar de alterações de grandeza nos parâmetros sem significância.

Atendendo aos resultados considerou-se que o modelo de referência apresentava valores com pequenas alterações de grandeza, que se podem justificar devido a redução da amostra, mas que mesmo assim segue em grande parte tendências coerentes com o senso comum.

2.6.2. Modelos de Referência Combinando apenas Variáveis Representando Oferta de Transportes Coletivos

O modelo de referência com base de dados reduzida MNL7.58b foi comparado com o modelo incluindo as variáveis sobre TC (MNL11.7). Os resultados desses modelos podem ser comparados no quadro VI.24.

É possível constatar a estabilidade de valores e de comportamentos dos parâmetros de todas as variáveis. Apenas as constantes específicas apresentam um ajustamento de grandeza espectável dada a introdução de mais 10 variáveis (2 x 5 alternativas). Estes resultados sugerem que a informação introduzida pelas variáveis sobre os TC não está aparentemente significativamente correlacionada com a das restantes variáveis. A melhoria do ajustamento global do modelo, apesar de ligeira é estatisticamente significativa.

Os resultados das variáveis relacionadas com os TC apresentam sinais coerentes com o senso comum, fundamentalmente para os modos com Bus. O modo Outros Bus apresenta sinais contrários aos das alternativas com Bus, e o modo pedonal sugere que é mais valorizado quando a qualidade do serviço dos TC também é melhor.

Na interpretação dos resultados deve atender-se, novamente, ao facto das frequências médias horárias diárias altas e razões de velocidade em Bus relativamente a viagens em Auto mais baixas, acontecem essencialmente nos núcleos urbanos consolidadas onde efetivamente as alternativas com Bus e o Walk (devido a distâncias mais reduzida) são mais utilizadas. Fora dos núcleos urbanos esses serviços pioram e a utilização do automóvel é preponderante.

Quadro VI. 24 - Análise dos resultados do modelo de referência MNL7.58b com a introdução das variáveis sobre TC

Variável	MNL7.58b					MNL11.7					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,693	-4,057	-5,237	-7,930	2,687	-7,102	-3,373	-6,010	-7,295	2,703	- Alterações de grandeza exceto para W. Sinais estáveis
Lc	-1,795	-1,711	-1,024	-1,345	-1,458	-1,812	-1,730	-1,025	-1,374	-1,476	- Parâmetros estáveis
NC0	5,413	4,191	2,693	4,458	2,989	5,373	4,149	2,676	4,459	2,985	- Parâmetros estáveis
NC1	3,232	2,059	2,124	2,450	0,818	3,196	2,020	2,152	2,435	0,820	
NC2	2,678	1,761	1,016***	1,826*	0,542	2,661	1,742	1,007***	1,818*	0,538	
NC3	1,201**	0,853	1,692**	1,823*	0,143***	1,690**	0,843	1,670**	1,836*	0,123***	
NC4a	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
D2	0,106	0,023	0,071	-0,026***	-2,397	0,111	0,027	0,078	-0,025***	-2,418	
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,241***	0,160***	0,651***	-0,391***	0,241***	0,259***	0,173***	0,605***	-0,390***	0,252***	- Parâmetros estáveis
In2	1,192	0,513	0,796*	0,474**	0,442	1,169	0,489	0,779*	0,449**	0,443	
In3	1,114	0,533	-0,142***	0,502**	0,363	1,115	0,534	-0,173***	0,485**	0,375	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
Id2	1,446	1,984	-0,440**	1,749	1,560	1,492	2,026	-0,452**	1,778	1,569	
Id3	1,642	2,210	-0,880*	2,153	1,548	1,653	2,213	-0,882*	2,152	1,542	
Id4	1,444	2,047	-0,875	0,565***	1,602	1,460	2,056	-0,863	0,572***	1,600	
Id5	1,819	2,499	-0,598***	-1,162***	1,776	1,807	2,480	-0,547***	-1,163***	1,744	
Sx	0,492	0,459	0,456	-1,633	0,353	0,488	0,455	0,448	-1,642	0,359	- Parâmetros estáveis
Mot											
Til											
De1											
De2											
De3											
De4											
De5											Parâmetros estáveis
nV	-0,222	-0,150	-0,141*	0,251	-0,168	-0,215	-0,142	-0,157	0,261	-0,161	
Tr	0,596	0,406	0,305***	0,557*	0,164**	0,583	0,394	0,345***	0,551*	0,154**	
Rg	0,372	0,243	0,160***	0,041***	0,302	0,375	0,247	0,191***	0,032***	0,313	
Es	0,218***	0,253	0,089***	-0,749*	0,048***	0,242***	0,278	0,100***	-0,754*	0,069***	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,190***	0,425	-1,542**	-0,747***	0,516	0,146***	0,376	-1,497**	0,796***	0,498	
Ftl1						0,012	0,012	-0,015**	-0,003***	0,016	- Parâmetros com sinal esperado e sem significância para OB e M
TBA						-0,538	-0,590	0,635	-0,403**	-0,155*	- Parâmetros com sinal esperado exceto OB.
Amostra			17005					17005			- Ligeira melhoria do ajustamento global
p-R ²			0,446					0,449			
LL			-12021,0					-11934,7			
Param.			100					110			
GL								10			- Hipótese nula dos modelos serem iguais pode ser recusada
-2LL								172,6			
Chi ² ($\alpha=0,05$)								18,3			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

2.7. Modelo de Referência com Inclusão das Variáveis Relacionadas com Transportes Coletivos e Estacionamentos

2.7.1. Modelo Base com Combinação de Variáveis Binárias

Após a análise da combinação do modelo de referência com as variáveis sobre TC e sobre estacionamento em separado, avaliou-se a sua estabilidade com a combinação de ambas no mesmo modelo. Os resultados deste modelo apresentam-se no quadro VI.25.

Da análise dos resultados pode verificar-se que:

- As constantes específicas para os modos Outro Bus, Moto e Walk apresentam alterações de grandeza;
- Os parâmetros da licença de condução mantêm sinais e grandezas do modelo de referência;
- Os parâmetros da variável taxa de motorização do agregado familiar mantêm sinais, a maioria das grandezas e comportamento coerente;
- Os parâmetros da variável distância mais curta mantêm os sinais e a grandeza. Apenas o modo Bus perde significância estatística;
- Os parâmetros da variável nível de instrução mantêm os sinais e a grandeza, continuando a não influenciar os modos Outro Bus e Moto. Verifica-se também que para os analfabetos continua sem significância estatística;
- As variáveis sobre a idade apresentaram parâmetros com sinal, comportamento e grandezas similares ao modelo de referência, havendo algumas alterações para o modo Outro Bus;
- A variável sexo apresenta parâmetros com sinais coerentes e grandezas semelhantes ao modelo de referência;
- A variável número total de viagens diárias mantêm os sinais e as grandezas dos parâmetros;
- Em relação ao motivo da viagem, os modos Outro Bus e Moto mantêm a perda de significância para qualquer motivo. Vários parâmetros com significância mantêm grandezas muito semelhantes;

Quadro VI. 25 - Avaliação da estabilidade dos resultados do modelo de referência MNL7.58b com a introdução simultânea das variáveis sobre TC e estacionamento (combinação de binárias)

Variável	MNL7.58b					MNL12.16					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,693	-4,057	-5,237	-7,930	2,687	-7,628	-4,134	-5,843	-7,522	3,115	- Alterações de grandeza para OB, M e W. Sinais coerentes
Lc	-1,795	-1,711	-1,024	-1,345	-1,458	-1,833	-1,760	-0,926	-1,379	-1,524	- Parâmetros estáveis
NC0	5,413	4,191	2,693	4,458	2,989	5,328	4,087	2,911	4,444	3,012	- Parâmetros quase todos estáveis
NC1	3,232	2,059	2,124	2,450	0,818	3,136	1,948	2,283	2,427	0,812	
NC2	2,678	1,761	1,016***	1,826*	0,542	2,646	1,702	1,079***	1,820*	0,476	
NC3	1,201**	0,853	1,692**	1,823*	0,143***	1,702**	0,802	1,786*	1,848*	0,064***	
NC4a	-	-	-	-	-	-	-	-	-	-	-
D2	0,106	0,023	0,071	-0,026***	-2,397	0,103	0,006***	0,103	-0,046***	-2,469	- Parâmetros estáveis exceto para B que perde significância
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,241***	0,160***	0,651***	-0,391***	0,241***	0,281***	0,154***	0,684**	-0,460***	0,251***	- Parâmetros estáveis
In2	1,192	0,513	0,796*	0,474**	0,442	1,193	0,455	0,918	0,424**	0,425	
In3	1,114	0,533	-0,142***	0,502**	0,363	1,153	0,512	-0,082***	0,465**	0,383	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	- Parâmetros com significância estáveis
Id2	1,446	1,984	-0,440**	1,749	1,560	1,445	2,030	-0,510**	1,767	1,605	
Id3	1,642	2,210	-0,880*	2,153	1,548	1,631	2,183	-1,021	2,114	1,551	
Id4	1,444	2,047	-0,875	0,565***	1,602	1,434	2,017	-1,003	0,543***	1,584	
Id5	1,819	2,499	-0,598***	-1,162***	1,776	1,760	2,404	-0,689**	-1,212***	1,609	
Sx	0,492	0,459	0,456	-1,633	0,353	0,477	0,465	0,494	-1,620	0,358	- Parâmetros estáveis
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,222	-0,150	-0,141*	0,251	-0,168	-0,214	-0,140	-0,157	0,256	-0,151	- Parâmetros estáveis
Tr	0,596	0,406	0,305***	0,557*	0,164**	0,542	0,369	0,439***	0,519*	0,151**	- Parâmetros estáveis exceto para OB. Para OB e M não têm significância
Rg	0,372	0,243	0,160***	0,041***	0,302	0,351	0,229	0,306***	0,013***	0,327	
Es	0,218***	0,253	0,089***	-0,749*	0,048***	0,228**	0,315	0,239***	-0,739*	0,169	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,190***	0,425	-1,542**	-0,747***	0,516	0,084***	0,273	-1,278**	-0,865***	0,408	
OV1						0,131***	0,561	-0,479***	0,351***	-1,157	- Perdem significância para BW, OB e M. Parâmetros para B com comportamento coerente
OV2						-0,338***	0,313	-2,477	0,216***	-1,308	
OV3						0,628**	0,182*	-0,767*	-0,090***	-0,964	
OV4						-	-	-	-	-	
PO1						-	-	-	-	-	- Parâmetros para BW e OB com comportamento não coerente. Sem significância para OB e M
PO2						0,724***	0,174**	1,126*	-0,115***	0,466	
PO3						0,590	0,458	-0,214***	0,109***	1,456	
Ftl1						0,006**	0,006	-0,017**	-0,006***	0,009	- Só tem significância para B e W, com sinal correto mas com maior importância para W
TBA						-0,372	-0,215	0,475**	-0,204***	0,049***	- Só tem significância para BW e B. Sinal coerente para BW, B e M
Amostra			17005					17005			
p-R ²			0,446					0,460			- Melhoria do ajustamento global
LL			-12021,0					-11696,9			
Param.			100					135			
GL								35			
-2LL								648,2			- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² (α=0,05)								49,8			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

- A razão entra a oferta de estacionamento e o total de viagens para a zona de atração apresenta comportamento coerente para o modo Bus. Os restantes modos apresentam um comportamento variável ou não apresentam significância estatística para estas variáveis. Comparando os resultados com os do modelo de referência com variáveis sobre estacionamentos, verifica-se que a maioria dos parâmetros é estável com algumas alterações de grandeza, que se podem justificar devido à redução da amostra;
- A percentagem de estacionamentos pagos apresenta comportamentos coerentes para os modos Bus e Walk, mas para a maior parte dos parâmetros verifica-se que não apresentam significância estatística. O modo Moto também apresenta comportamento coerente. Comparativamente ao modelo de referência combinado apenas com as variáveis de estacionamento, apenas os modos Bus e Walk apresentam valores estáveis, e o modo Bus+Walk modifica o seu comportamento;
- A variável frequência média horária diária de TC apenas apresenta significância estatística para os modos Bus e Walk, o seu sinal positivo é coerente mas aparentemente têm maior relevância para o modo Walk que para o Bus. Comparando os resultados com o modelo de referência com as variáveis sobre os TC, verificou-se uma redução da grandeza dos parâmetros das alternativas mais relevantes e a perda de significância para o modo Bus+Walk;
- Os parâmetros da variável razão de tempos médios de viagem em Bus *versus* Auto apenas apresentam significância estatística para os modos Bus e Bus+Walk, com sinal negativo coerente uma vez que tempos relativos de viagem por Bus maiores diminuem a sua atratividade. A sua influência parece ser maior para o Bus+Walk do que para o Bus. Comparando os resultados aos do modelo de referência com variáveis sobre os TC, verifica-se a perda de significância para os modos Outros Bus e a manutenção para a Moto e Walk. Verifica-se uma redução da grandeza para todos os parâmetros mas o sinal para o Bus mantém a coerência.

2.7.2. Teste à utilização de Variáveis de Estacionamento Contínuas

Atendendo aos resultados preliminares sobre a melhor tipologia para as variáveis sobre estacionamento, avaliou-se a possibilidade de as substituir no modelo de referência com a combinação de variáveis binárias pela sua especificação contínua de modo a simplificar o modelo. Os resultados desses modelos são apresentados no quadro VI.26.

É possível verificar que o modelo com especificação contínua apresenta um ajustamento global inferior ao modelo com combinações binárias. Apesar de ter significância estatística (cerca de 7 vezes superior ao valor crítico), o acréscimo de ajustamento do modelo com combinação de binárias não é elevado.

Quadro VI. 26 - Comparação dos resultados do modelo de referência com a introdução das variáveis sobre TC e estacionamento contínuas e combinação de binárias

Variável	MNL12.20					MNL12.16					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-7,091	-3,691	-5,497	-7,352	1,120	-7,628	-4,134	-5,843	-7,522	3,115	- Alterações de grandeza principalmente para W. Sinais estáveis
Lc	-1,819	-1,750	-0,967	-1,370	-1,514	-1,833	-1,760	-0,926	-1,379	-1,524	- Parâmetros estáveis
NC0	5,314	4,064	2,751	4,424	2,980	5,328	4,087	2,911	4,444	3,012	- Parâmetros estáveis
NC1	3,143	1,934	2,201	2,408	0,787	3,136	1,948	2,283	2,427	0,812	
NC2	2,638	1,691	1,031***	1,815*	0,513	2,646	1,702	1,079***	1,820*	0,476	
NC3	1,663**	0,794	1,730**	1,826*	0,101***	1,702**	0,802	1,786*	1,848*	0,064***	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,093	0,004***	0,079	-0,038***	-2,484	0,103	0,006***	0,103	-0,046***	-2,469	- Parâmetros estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
In1	0,257***	0,183***	0,612***	-0,423***	0,356**	0,281***	0,154***	0,684**	-0,460***	0,251***	- Maioria dos parâmetros estáveis.
In2	1,149	0,467	0,799*	0,427**	0,480	1,193	0,455	0,918	0,424**	0,425	
In3	1,100	0,513	-0,132***	0,464**	0,394	1,153	0,512	-0,082***	0,465**	0,383	
In4	-	-	-	-	-	-	-	-	-	-	
Id1	-	-	-	-	-	-	-	-	-	-	
Id2	1,485	2,042	-0,489**	1,770	1,652	1,445	2,030	-0,510**	1,767	1,605	- Parâmetros estáveis
Id3	1,616	2,177	-0,902*	2,114	1,610	1,631	2,183	-1,021	2,114	1,551	
Id4	1,429	2,019	-0,872	0,550***	1,642	1,434	2,017	-1,003	0,543***	1,584	
Id5	1,754	2,396	-0,551***	-1,200***	1,691	1,760	2,404	-0,689**	-1,212***	1,609	
Sx	0,490	0,463	0,446	-1,624	0,356	0,477	0,465	0,494	-1,620	0,358	- Parâmetros estáveis
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,213	-0,138	-0,162	0,258	-0,160	-0,214	-0,140	-0,157	0,256	-0,151	- Parâmetros estáveis
Tr	0,557	0,371	0,351***	0,513*	0,155**	0,542	0,369	0,439***	0,519*	0,151**	- Parâmetros estáveis.
Rg	0,362	0,230	0,222***	0,009***	0,329	0,351	0,229	0,306***	0,013***	0,327	
Es	0,252***	0,310	0,118***	-0,754*	0,142***	0,228**	0,315	0,239***	-0,739*	0,169	
Lz	-	-	-	-	-	-	-	-	-	-	
CS	0,082***	0,276*	-1,340**	-0,859***	0,437	0,084***	0,273	-1,278**	-0,865***	0,408	
OV1						0,131***	0,561	-0,479***	0,351***	-1,157	- B com comportamento coerente
OV2						-0,338***	0,313	-2,477	0,216***	-1,308	
OV3						0,628**	0,182*	-0,767*	-0,090***	-0,964	
OV4						-	-	-	-	-	
PO1						-	-	-	-	-	
PO2						0,724***	0,174**	1,126*	-0,115***	0,466	- B e W com comportamento coerente
PO3						0,590	0,458	-0,214***	0,109***	1,456	
OV	-3,017*	-3,048	-0,807***	-1,511***	7,974						- Sinal coerente exceto para W. Grande valorização para o modo W
PO	0,391*	0,801	-0,908**	0,337***	1,992						- Sinal coerente exceto para OB
Ftl1	0,007**	0,006	-0,014***	-0,007***	0,009	0,006**	0,006	-0,017**	-0,006***	0,009	Parâmetros estáveis
TBA	-0,275*	-0,179	0,463**	-0,210***	0,023***	-0,372	-0,215	0,475**	-0,204***	0,049***	- Algumas alterações de grandeza nos modos com B
Amostra			17005					17005			
p-R ²			0,456					0,460			- Ligeira melhoria do ajustamento global
LL			-11779,4					-11696,9			
Param.			120					135			
GL								15			
-2LL								165			- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² (α=0,05)								25,0			

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

É possível constatar alterações de grandeza nas constantes específicas, principalmente para o modo Walk. Em relação às restantes variáveis, os seus parâmetros apresentam-se muito semelhantes para ambos os modelos, sugerindo que a introdução das diferentes especificações das variáveis sobre estacionamento não tem aparentemente influência relevante nas restantes variáveis.

As variáveis contínuas relacionadas com o estacionamento apresentam sinal e comportamento coerentes para a maioria dos parâmetros. Em relação à densidade de oferta de estacionamento face à procura, os sinais negativos dos parâmetros são coerentes, excetuando o resultado para o modo Walk. O valor do parâmetro para o modo Walk apresenta uma grandeza positiva invulgar, cuja justificação pode dever-se a um acentuado uso do modo pedonal nas zonas centrais, provavelmente nas que apresentam maior oferta de estacionamento. A percentagem de estacionamentos pagos apresenta significância e o sinal expectável para as alternativas mais relevantes.

O parâmetro da variável taxa de estacionamentos pagos apresenta valor positivo coerente para as alternativas Bus e Walk, não sendo significativo para as restantes alternativas. Mais uma vez se verifica uma influência mais acentuada no modo pedonal que no Bus, provavelmente devido ao motivo referido anteriormente.

As variáveis relacionadas com os serviços de TC apresentaram valores para a frequência média horária diária estáveis e com significância para as alternativas mais relevantes (Bus e Walk), e a razão de tempos de viagem apresenta o sinal expectável e significância para o modo Bus.

Estes resultados sugerem que apesar da melhoria do ajustamento global do modelo com variáveis de estacionamento com especificação binária, que a especificação contínua apresenta resultados muito semelhantes, demonstrando potencial para ser uma alternativa viável na simplificação do modelo.

2.7.3. Teste de Modelo sem Variáveis Socioeconómicas menos Relevantes

Atendendo aos resultados para as variáveis grupo etário e nível de instrução cujo comportamento se alterou com a adição das restantes variáveis comparativamente aos resultados dos modelos isolados, analisou-se um modelo mais simples com menos variáveis socioeconómicas, no sentido de avaliar a possibilidade de utilizar um modelo menos complexo. Optou-se pela manutenção das variáveis sexo e número de viagens diárias pela estabilidade de valores que apresentaram ao longo dos modelos. A comparação desses resultados é apresentada no quadro VI.27.

Quadro VI. 27 - Comparação do modelo de referência MNL12.16 com o modelo sem as variáveis socioeconômicas consideradas menos relevantes

Variável	MNL12.17					MNL12.16					Obs.
	BW	B	OB	M	W	BW	B	OB	M	W	
ASC	-5,619	-2,066	-5,190	-6,041	4,662	-7,628	-4,134	-5,843	-7,522	3,115	- Alterações significativas de grandeza. Sinais estáveis
Lc	-1,751	-1,589	-1,667	-1,474	-1,411	-1,833	-1,760	-0,926	-1,379	-1,524	Sinais estáveis. Grandezas estáveis exceto para OB e B.
NC0	5,604	4,230	3,030	4,210	3,153	5,328	4,087	2,911	4,444	3,012	- Ligeiras alterações de grandeza principalmente para OB e M. Tendências coerentes
NC1	3,277	1,931	2,531	2,288	0,864	3,136	1,948	2,283	2,427	0,812	
NC2	2,740	1,675	1,320**	1,672**	0,511	2,646	1,702	1,079***	1,820*	0,476	
NC3	1,732*	0,765	1,928*	1,855*	0,115***	1,702**	0,802	1,786*	1,848*	0,064***	
NC4a	-	-	-	-	-	-	-	-	-	-	
D2	0,111	0,013**	0,090	-0,044***	-2,470	0,103	0,006***	0,103	-0,046***	-2,469	- Parâmetros estáveis
Dx1											
R1											
R2											
R3											
R4											
R5											
In1						0,281***	0,154***	0,684**	-0,460***	0,251***	
In2						1,193	0,455	0,918	0,424**	0,425	
In3						1,153	0,512	-0,082***	0,465**	0,383	
In4						-	-	-	-	-	
Id1						-	-	-	-	-	
Id2						1,445	2,030	-0,510**	1,767	1,605	
Id3						1,631	2,183	-1,021	2,114	1,551	
Id4						1,434	2,017	-1,003	0,543***	1,584	
Id5						1,760	2,404	-0,689**	-1,212***	1,609	
Sx	0,456	0,465	0,264**	-1,658	0,356	0,477	0,465	0,494	-1,620	0,358	- Parâmetros estáveis exceto para OB
Mot											
Til											
De1											
De2											
De3											
De4											
De5											
nV	-0,183	-0,110	-0,240	0,314	-0,128	-0,214	-0,140	-0,157	0,256	-0,151	- Parâmetros estáveis
Tr	0,618	0,411	0,081***	0,691	0,241	0,542	0,369	0,439***	0,519*	0,151**	- Parâmetros estáveis exceto para OB e M
Rg	0,360	0,214	0,191***	0,081***	0,347	0,351	0,229	0,306***	0,013***	0,327	
Es	0,219***	0,250	0,236***	-0,303***	0,191**	0,228**	0,315	0,239***	-0,739*	0,169***	
Lz	-	-	-	-	-	-	-	-	-	-	
Cs	0,120***	0,310	-1,458**	-1,118**	0,439	0,084***	0,273	-1,278**	-0,865***	0,408	- Parâmetros estáveis exceto para OB
OV1	0,214***	0,587	-0,355***	0,454***	-1,145	0,131***	0,561	-0,479***	0,351***	-1,157	
OV2	-0,286***	0,299	-2,306	0,245***	-1,312	-0,338***	0,313	-2,477	0,216***	-1,308	
OV3	0,619	0,162**	-0,603**	-0,185***	-0,961	0,628**	0,182*	-0,767*	-0,090***	-0,964	
OV4	-	-	-	-	-	-	-	-	-	-	- Parâmetros estáveis
PO1	-	-	-	-	-	-	-	-	-	-	
PO2	0,706	0,200*	1,013*	-0,087***	0,469	0,724***	0,174**	1,126*	-0,115***	0,466	
PO3	0,577	0,491	-0,263***	-0,086***	1,454	0,590	0,458	-0,214***	0,109***	1,456	
Ftl1	0,006**	0,005	-0,018**	-0,006***	0,008	0,006**	0,006	-0,017**	-0,006***	0,009	- Parâmetros estáveis
TBA	-0,342	-0,191	0,447**	-0,166***	0,053***	-0,372	-0,215	0,475**	-0,204***	0,049***	- Parâmetros estáveis
Amostra			17005					17005			
p-R ²			0,445					0,460			- Melhoria do ajustamento global
LL			-12003,1					-11696,9			
Param.			100					135			
GL								35			
-2LL								612,4			
Chi ² (α=0,05)								49,8			- Hipótese nula dos modelos serem iguais pode ser recusada

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

Dos resultados é possível verificar que o modelo mais completo MNL12.16 apresenta um ajustamento global superior, que o teste -2LL considera estatisticamente significativa (cerca

de 12 vezes superior ao valor crítico). Apesar disso, os parâmetros para a maioria das variáveis mantêm sinais, comportamentos e grandezas em ambos os modelos:

- As constantes específicas apresentaram alterações de grandeza relevantes para a maioria das alternativas;
- A licença de condução mantém parâmetros com sinal coerente e com valores semelhantes, com exceção do modo Outros Bus e Bus;
- A taxa de motorização familiar apresenta parâmetros com sinal, comportamento e grandezas semelhantes. As maiores discrepâncias verificam-se para a categoria NC0 e para os modos Outro Bus e Moto;
- A distância mais curta entre centróides revela estabilidade dos parâmetros tanto em sinal como em grandeza;
- O sexo do inquirido revela parâmetros estáveis com exceção do modo Outros Bus cuja grandeza diminui e perdeu significância estatística comparativamente ao modelo mais completo MNL12.16;
- O número total de viagens diárias manteve parâmetros com sinal e grandeza estáveis.
- Da análise dos parâmetros sobre o motivo da viagem, verifica-se alguma estabilidade de valores e comportamento, mantendo-se a não significância estatística destas variáveis para os modos Outro Bus e Moto face ao modelo MNL12.16;
- As variáveis sobre a razão da oferta de estacionamento em relação ao número total de viagens, mantêm a significância estatística apenas para as alternativas mais relevantes, bem como sinais, grandeza e comportamento coerente para o Bus, relativamente ao modelo MN12.16.
- Os parâmetros para a percentagem de estacionamento pagos mantêm sinal e grandeza;
- A variável da frequência média diária de Bus apresenta parâmetros estáveis;
- A variável da razão entre a duração média da viagem de Bus e a duração média da viagem por AUTO apresenta valores estáveis em sinal e grandeza.

Estes resultados sugerem que o modelo com apenas as variáveis socioeconómicas consideradas mais relevantes apresenta um potencial para modelar as escolhas modais de forma bastante semelhante ao modelo com as variáveis nível de instrução e grupo etário, com ajustamentos próximos e parâmetros similares, preservando, portanto, o essencial da capacidade explicativa. Estes resultados sugerem também, no entanto, que a informação destas variáveis acrescenta alguma capacidade explicativa e que aparentemente não estão correlacionadas com as restantes variáveis combinadas. Na sua utilização deve ponderar-se a relação complexidade acrescida *versus* o aumento do ajustamento global. Atendendo aos resultados

verificados e à melhoria do ajustamento global, parece justificar-se a manutenção das variáveis.

2.7.4. Redução da Base de Dados aos Modos Alternativos mais Relevantes

Os problemas de redução da amostra devida à combinação de variáveis, mais concretamente com as variáveis sobre estacionamento e sobre os TC, revelou também que os modos alternativos menos relevantes, Combinação Bus+Walk, Outros Bus e Moto, deixavam de ter significância estatística para muitas variáveis. Este problema deve-se à purga da amostra destes modos, não apresentando dados suficientes para que o modelo os possa considerar devidamente.

Neste sentido, apenas para a análise incluindo a influência das variáveis sobre o estacionamento e sobre os TC reduziu-se a base de dados aos inquiridos cujo modo de transporte fosse apenas um dos três modos de transporte mais relevantes, Bus, Walk e Auto. Considerando esta base de dados, realizaram-se modelos similares aos modelos mais completos com a base de dados integral e avaliou-se a estabilidade dos parâmetros. Os resultados dessas análises são apresentados no quadro VI.28.

É possível constatar que a purga das alternativas menos relevantes não diminuiu consideravelmente a amostra, e que efetivamente a amostra dos modos agora excluídos era muito pequena. Comparando os resultados dos modelos MNL12.16 e 12.17 com as suas versões reduzidas a três modos, constata-se que os modelos com redução das alternativas apresentam parâmetros muito estáveis, generalizadamente consistentes e estatisticamente têm ajustamentos globais muito significativos, sugerindo que efetivamente os dados retirados não introduziam informação em quantidade e com riqueza suficiente que justifiquem a sua manutenção no modelo.

Comparando os resultados com e sem as variáveis idade e nível de instrução, constata-se uma ligeira perda de ajustamento global devido à sua exclusão. O teste estatístico -2LL comprova que o modelo mais completo é melhor que o modelo de base (cerca de 19 vezes superior ao valor crítico). Relativamente aos parâmetros, é possível verificar que apesar de algumas ligeiras diferenças de grandeza que os sinais, a coerência e o comportamento dos mesmos é muito similar. Esta constatação parece comprovar que as variáveis retiradas não apresentam correlações significativas com as variáveis mantidas no modelo. As alterações mais significativas em termos de grandeza verificaram-se nas constantes específicas e nos parâmetros da variável sobre a disponibilidade de licença de condução.

Quadro VI. 28 - Comparação do modelo de referência com o modelo apenas com as variáveis socioeconômicas consideradas mais relevantes utilizando a base de dados reduzida às alternativas mais relevantes

Variável	MNL12.17BWA		MNL12.16BWA		Obs.
	B	W	B	W	
ASC	-2,146	4,494	-4,156	2,919	- Similares aos modelos com todas as alternativas
Lc	-1,578	-1,407	-1,748	-1,528	- Similares aos modelos com todas as alternativas
NC0	4,237	3,139	4,101	2,998	- Similares aos modelos com todas as alternativas
NC1	1,931	0,851	1,949	0,801	
NC2	1,682	0,494	1,712	0,464	
NC3	0,790	0,091***	0,820	0,039***	
NC4a	-	-	-	-	
D2	0,011***	-2,444	0,005***	-2,444	- Similares aos modelos com todas as alternativas
Dx1					
R1					
R2					
R3					
R4					
R5					
In1			0,093***	0,265***	- Similares ao modelo com todas as alternativas
In2			0,436	0,420	
In3			0,490	0,396	
In4			-	-	
Id1			-	-	- Similares ao modelo com todas as alternativas
Id2			2,019	1,628	
Id3			2,158	1,600	
Id4			1,942	1,592	
Id5			2,358	1,632	
Sx	0,466	0,377	0,474	0,378	- Similares aos modelos com todas as alternativas
Mot					
Til					
De1					
De2					
De3					
De4					
De5					- Similares aos modelos com todas as alternativas
nV	-0,108	-0,130	-0,138	-0,155	
Tr	0,471	0,282	0,426	0,192**	- Similares aos modelos com todas as alternativas
Rg	0,256	0,371	0,266	0,352	
Es	0,318	0,249*	0,342	0,212**	
Lz	-	-	-	-	
Cs	0,356	0,489	0,323	0,461	- Similares aos modelos com todas as alternativas
OV1	0,639	-1,101	0,651	-1,106	
OV2	0,316	-1,232	0,342	-1,215	
OV3	0,165**	-0,946	0,192*	-0,945	
OV4	-	-	-	-	
PO1	-	-	-	-	- Similares aos modelos com todas as alternativas
PO2	0,171**	0,441	0,140**	0,431	
PO3	0,445	1,475	0,418	1,479	
Ftl1	0,006	0,010	0,007	0,010	- Similares aos modelos com todas as alternativas
TBA	-0,183	0,085***	-0,209	0,087***	- Similares aos modelos com todas as alternativas
Amostra	16197		16197		- Melhoria do ajustamento global
p-R ²	0,509		0,522		
LL	-8718,4		-8493,6		
Param.	40		54		- Hipótese nula dos modelos serem iguais pode ser recusada
GL			14		
-2LL			449,6		
Chi ² (α=0,05)			23,7		

* - p-valor >0,05 & ≤0,1

** - p-valor >0,1 & ≤0,25

*** - p-valor >0,25

2.7.5. Exclusão do Motivo da Viagem

Nos modelos anteriores foi possível verificar que os parâmetros das variáveis Motivo de Viagem apresentavam grandezas muito semelhantes, sugerindo que as mesmas não induziam uma preferência declarada por determinada alternativa. Esta hipótese levou à comparação dos modelos sem estas variáveis, cujos resultados são apresentados no quadro VI.29.

Quadro VI. 29 - Avaliação do impacto da variável motivo da viagem no modelo de referência com as variáveis socioeconômicas essenciais e com a base de dados reduzida às alternativas mais relevantes

Variável	MNL12.18BWA		MNL12.17BWA		Obs.
	B	W	B	W	
ASC	-1,892	4,800	-2,146	4,494	-Alteração de grandezas. Sinais estáveis
Lc	-1,544	-1,399	-1,578	-1,407	- Parâmetros estáveis
NC0	4,222	3,139	4,237	3,139	- Parâmetros estáveis
NC1	1,941	0,866	1,931	0,851	
NC2	1,690	0,511	1,682	0,494	
NC3	0,806	0,111***	0,790	0,091***	
NC4a	-	-	-	-	- Parâmetros estáveis
D2	0,011***	-2,442	0,011***	-2,444	
Dx1					
R1					
R2					
R3					
R4					
R5					
ln1					
ln2					
ln3					
ln4					
ld1					
ld2					
ld3					
ld4					
ld5					
Sx	0,480	0,394	0,466	0,377	- Parâmetros estáveis
Mot					
Til					
De1					
De2					
De3					
De4					
De5					
nV	-0,109	-0,135	-0,108	-0,130	- Parâmetros estáveis
Tr			0,471	0,282	
Rg			0,256	0,371	
Es			0,318	0,249*	
Lz			-	-	
Cs			0,356	0,489	
OV1	0,665	-1,085	0,639	-1,101	- Parâmetros estáveis
OV2	0,342	-1,226	0,316	-1,232	
OV3	0,187*	-0,950	0,165**	-0,946	
OV4	-	-	-	-	
PO1	-	-	-	-	- Parâmetros estáveis
PO2	0,170**	0,440	0,171**	0,441	
PO3	0,436	1,466	0,445	1,475	
Ft1	0,006	0,009	0,006	0,010	- Parâmetros estáveis
TBA	-0,189	0,078***	-0,183	0,085***	- Parâmetros estáveis
Amostra	16197		16197		
p-R ²	0,508		0,509		- Ligeira melhoria do ajustamento global
LL	-8740,1		-8718,4		
Param.	32		40		
GL			8		
-2LL			43,4		- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² (α=0,05)			15,5		

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Dos resultados é possível constatar que o ajustamento global é praticamente igual, apesar de o teste estatístico -2LL comprovar que estatisticamente o modelo com o motivo da viagem é melhor que o modelo de referência (cerca de três vezes superior ao valor crítico). Em relação aos restantes parâmetros é possível constatar a sua estabilidade, verificando-se apenas alterações de grandeza nas constantes específicas. Esta constatação sugere que efetivamente o motivo da viagem acrescenta alguma informação relevante ao modelo que aparentemente não é introduzida pelas restantes variáveis e que essa informação não aparenta ter correlações fortes com as mesmas. Assim, a alteração da grandeza das constantes específicas sugere que o motivo da viagem não seja retirado do modelo, apesar do reduzido impacto no ajustamento global, salvaguardado os casos onde, eventualmente, se verifique manifesta necessidade de simplificar o modelo.

2.7.6. Inclusão da Duração Média da Viagem por Modo

Com a purga das alternativas menos relevantes, colocou-se a hipótese de que a variável Duração Média da Viagem pudesse ser novamente introduzida no modelo, uma vez que a sua inclusão poderia agora não impor uma redução tão acentuada da amostra. Esta hipótese foi testada e os resultados comparados com o modelo anterior. A diferença no tamanho da amostra implicou a aplicação do modelo MNL12.17BWA à base de dados reajustada (MNL12.17BWA_a), cuja estabilidade é comparada no quadro VI.30.

É possível constatar que a maioria dos parâmetros revela estabilidade de valores, sinais e comportamentos. Os parâmetros da taxa de motorização familiar apresentam algumas alterações de grandeza para o modo Bus, que se podem dever à redução da amostra (redução de cerca de 22%). Os parâmetros da percentagem de estacionamento pagos perdem significância estatística para o modo Bus. Os resultados sugerem que apesar da estabilidade da maioria dos parâmetros se deve ponderar devidamente a redução da amostra, uma vez que há perda de significância dos parâmetros da percentagem de estacionamento pago.

Ainda assim, no sentido de avaliar a pertinência desta redução, verificou-se o ganho no ajustamento com a introdução da Duração Média da Viagem no modelo de referência MNL12.17BWA_a (quadro VI.31).

Constatou-se que para a maioria das variáveis os parâmetros se mantinham estáveis. A introdução da Duração Média da Viagem retirou significância estatística aos parâmetros da razão de tempos de viagem entre Bus e Auto. Este resultado sugere uma possível correlação entre as duas variáveis, expectável atendendo à definição analítica da variável razão de tempos de viagem. O ligeiro ganho no ajustamento global, embora estatisticamente relevante (cerca de 3 vezes superior ao valor crítico), aparentemente não justifica a redução da amostra decorrente da introdução do tempo médio de viagem, uma vez que aparentemente a informação já é fornecida pela razão de tempos de viagem.

Quadro VI. 30 - Avaliação do impacto no modelo de referência MNL12.17BWA da redução da base de dados

Variável	MNL12.17BWA_a		MNL12.17BWA		Obs.
	B	W	B	W	
ASC	-2,016	4,648	-2,146	4,494	- Alteração de grandezas
Lc	-1,623	-1,437	-1,578	-1,407	- Parâmetros estáveis
NC0	4,064	3,147	4,237	3,139	- Alterações de grandeza para o modo B. Estáveis para W
NC1	1,742	0,834	1,931	0,851	
NC2	1,410	0,434	1,682	0,494	
NC3	0,449*	0,026***	0,790	0,091***	
NC4a	-	-	-	-	- Parâmetro para B passa a ter significância estatística
D2	0,059	-2,376	0,011***	-2,444	
Dx1					
R1					
R2					
R3					
R4					
R5					
ln1					
ln2					
ln3					
ln4					
ld1					
ld2					
ld3					
ld4					
ld5					
Sx	0,442	0,363	0,466	0,377	- Parâmetros estáveis
Mot					
Til					
De1					
De2					
De3					
De4					
De5					
nV	-0,091	-0,119	-0,108	-0,130	- Parâmetros estáveis
Tr	0,487	0,285	0,471	0,282	- Parâmetros estáveis
Rg	0,262	0,375	0,256	0,371	
Es	0,259	0,207**	0,318	0,249*	
Lz	-	-	-	-	
Cs	0,366	0,514	0,356	0,489	
OV1	0,795	-1,185	0,639	-1,101	- Ligeiras alterações de grandeza
OV2	0,444	-1,344	0,316	-1,232	
OV3	0,187**	-1,105	0,165**	-0,946	
OV4	-	-	-	-	
PO1	-	-	-	-	- Parâmetros para B deixam de ter significância estatística mas mantém comportamento
PO2	0,111***	0,402	0,171**	0,441	
PO3	0,277*	1,357	0,445	1,475	
Ftl1	0,007	0,009	0,006	0,010	- Parâmetros estáveis
TBA	-0,222	0,042	-0,183	0,085***	- Parâmetros estáveis. Parâmetro ganha significância estatística para W
Amostra	12793		16197		- Amostras diferentes
p-R ²	0,488		0,509		
LL	-7144,5		-8718,4		
Param.	40		40		

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Quadro VI. 31 - Comparação dos resultados do modelo de referência MNL 12.17BWA_a com a introdução do tempo médio de viagem

Variável	MNL12.17BWA_a		MNL12.19BWA		Obs.
	B	W	B	W	
ASC	-2,016	4,648	-2,104	4,541	- Parâmetros relativamente estáveis
Lc	-1,623	-1,437	-1,618	-1,436	- Parâmetros estáveis
NC0	4,064	3,147	4,046	3,144	- Parâmetros estáveis
NC1	1,742	0,834	1,732	0,832	
NC2	1,410	0,434	1,404	0,433	
NC3	0,449*	0,026***	0,444*	0,023***	
NC4a	-	-	-	-	
D2	0,059	-2,376	0,058	-2,333	- Parâmetros estáveis
Dx1					
R1					
R2					
R3					
R4					
R5					
In1					
In2					
In3					
In4					
Id1					
Id2					
Id3					
Id4					
Id5					
Sx	0,442	0,363	0,441	0,361	- Parâmetros estáveis
Mot					
Til			-0,159		- Sinal coerente
De1					
De2					
De3					
De4					
De5					
nV	-0,091	-0,119	-0,089	-0,118	- Parâmetros estáveis
Tr	0,487	0,285	0,487	0,279	
Rg	0,262	0,375	0,268	0,370	
Es	0,259	0,207**	0,267	0,203	- Parâmetros estáveis
Lz	-	-	-	-	
Cs	0,366	0,514	0,361	0,512	
OV1	0,795	-1,185	0,796	-1,183	
OV2	0,444	-1,344	0,443	-1,356	- Parâmetros estáveis
OV3	0,187**	-1,105	0,183**	-1,102	
OV4	-	-	-	-	
PO1	-	-	-	-	
PO2	0,111***	0,402	0,106***	0,409	- Parâmetros estáveis
PO3	0,277*	1,357	0,266*	1,363	
Ftl1	0,007	0,009	0,007	0,009	- Parâmetros estáveis
TBA	-0,222	0,042	-0,119**	0,085***	- Parâmetros perdem significância
Amostra	12793		12793		
p-R ²	0,488		0,488		- Ligeira melhoria do ajustamento global
LL	-7144,5		-7138,9		
Param.	40		41		
GL			1		
-2LL			11,2		
Chi ² (α=0,05)			3,8		- Hipótese nula dos modelos serem iguais pode ser recusada
* - p-valor >0,05 & <=0,1 ** - p-valor >0,1 & <=0,25 *** - p-valor >0,25					

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Assim, o modelo de referência selecionado mantém-se o MNL 12.17BWA.

3. Resultados Globais Fundamentais

3.1. Grau de (In)dependência dos Efeitos dos Diferentes Fatores

Ao longo das análises anteriores foi possível verificar um aumento do ajustamento global dos modelos. Se alguma dessa melhoria se pode dever a reduções da amostra em análise, outra parte deve-se à introdução de informação relevante nos modelos através da adição de mais variáveis ao modelo de base. Nas figuras VI.1 e VI.2 apresentam-se os gráficos com a evolução do ajustamento global, avaliado pelo coeficiente $p-R^2$, com a introdução sucessiva de mais variáveis ao modelo de base inicial (MNL7.40), para as análises com 6 modos de transporte e para aquelas onde se avaliaram apenas os três modos mais relevantes. Nos gráficos, como termo de referência, é apresentada igualmente a amostra utilizada para cada modelo.

Verifica-se que o modelo MNL3.139 (modelo de base socioeconómico) apenas com as variáveis socioeconómicas consideradas mais relevantes permite obter um $p-R^2$ de 0,262 para uma amostra de 95 426, equivalente a um R^2 de uma regressão linear de cerca de 0,5. Este resultado sugere que as variáveis socioeconómicas apresentam algum poder explicativo, mas as análises subsequentes realizadas sugerem, no entanto, alguma sobreposição de informação entre estas variáveis e entre estas e outras também analisadas.

É possível constatar que para uma amostra de 95 426 o modelo MNL7.40a, apenas com as variáveis socioeconómicas que revelaram maior capacidade explicativa, adicionadas à variável sobre a distância entre GA e à disponibilidade diária de Moto no agregado, introduz informação que justifica mais de metade do ajustamento máximo alcançado, e um acréscimo de 52% no $p-R^2$ comparativamente ao modelo de base. A introdução das restantes variáveis socioeconómicas (MNL7.37) melhora o ajustamento 65% relativamente ao modelo de base, mas apenas 8% em relação ao modelo MNL7.40a. Constata-se igualmente que a exclusão da variável do rendimento mensal do agregado (MNL7.48) apenas reduz 0,2% o ajustamento global em relação ao MNL7.37. O modelo com maior número de variáveis que permitem utilizar a amostra completa (MNL7.57) apresenta um ajustamento cerca de 70% superior ao do modelo de base e 12% superior ao do modelo MNL7.40a. Este último modelo sugere que a introdução das variáveis número de viagens diárias e Motivo da viagem apenas aumenta 3,5% o ajustamento global em relação ao modelo sem as mesmas (MNL7.48).

De forma a diferenciar os ganhos de ajustamento devido à redução da amostra relativamente aos ganhos devido à introdução de informação adicional através de mais variáveis, apresentam-se os modelos de referência de cada etapa para ambas as amostras.

Assim, é possível verificar que a redução da amostra devido à duração média da viagem (em cerca de 68% para 30 559) reduz o ajustamento em 3,3%. A introdução da duração média da viagem ao modelo de referência (MNL7.48a) melhora o ajustamento global apenas 1,2% e a

sua combinação com o número de viagens diárias e o motivo da viagem (MNL7.51a) melhoram o ajustamento 5,0%.

A falta de dados completos devido à duração média da viagem e à duração das atividades levou à redução para apenas 20,5% (19 555) da amostra inicial. Esta redução levou a um aumento do ajustamento global de 1,6% (relativamente ao MNL7.57) a 6% (relativamente ao MNL7.56). É possível verificar que a introdução da duração da estadia apenas aumenta o ajustamento 0,4% (MNL7.55) e que retirar a duração média da viagem e a duração das atividades reduz o ajustamento global 2,8% (MNL7.57a).

A introdução das variáveis relacionadas com as políticas de estacionamento leva a uma redução da amostra inicial para 19,7% (18 823), uma redução de 4% em relação à amostra anterior, e a uma redução do ajustamento global de 1,8%. É possível verificar que a introdução da combinação de variáveis binárias sobre o estacionamento melhora o ajustamento global 3,4% (MNL9.09).

A introdução das variáveis relacionadas com os serviços de TC reduz a amostra a apenas 18% da inicial (17 005), uma redução de cerca de 10% em relação à amostra anterior, e melhora o ajustamento global 0,5%. A adição das variáveis relacionadas com os TC melhora ao ajustamento global apenas 0,7% (MNL11.7).

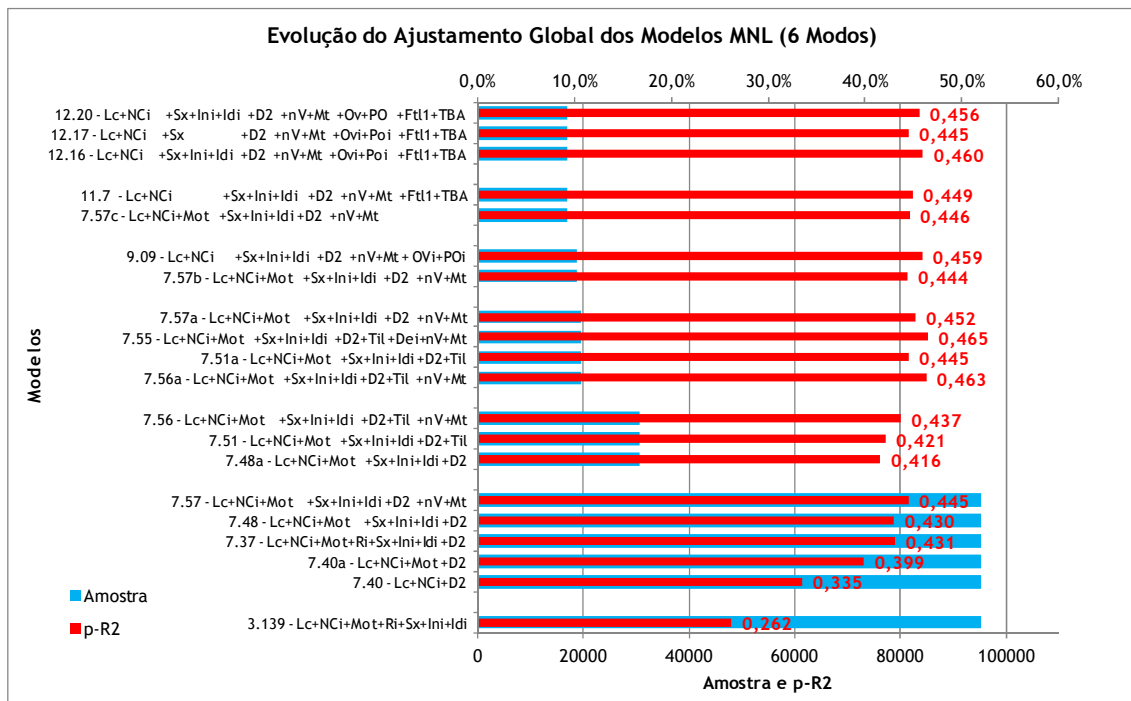


Figura VI. 1 - Evolução do ajustamento global dos modelos MNL com a introdução das variáveis mais relevantes para a análise com 6 modos.

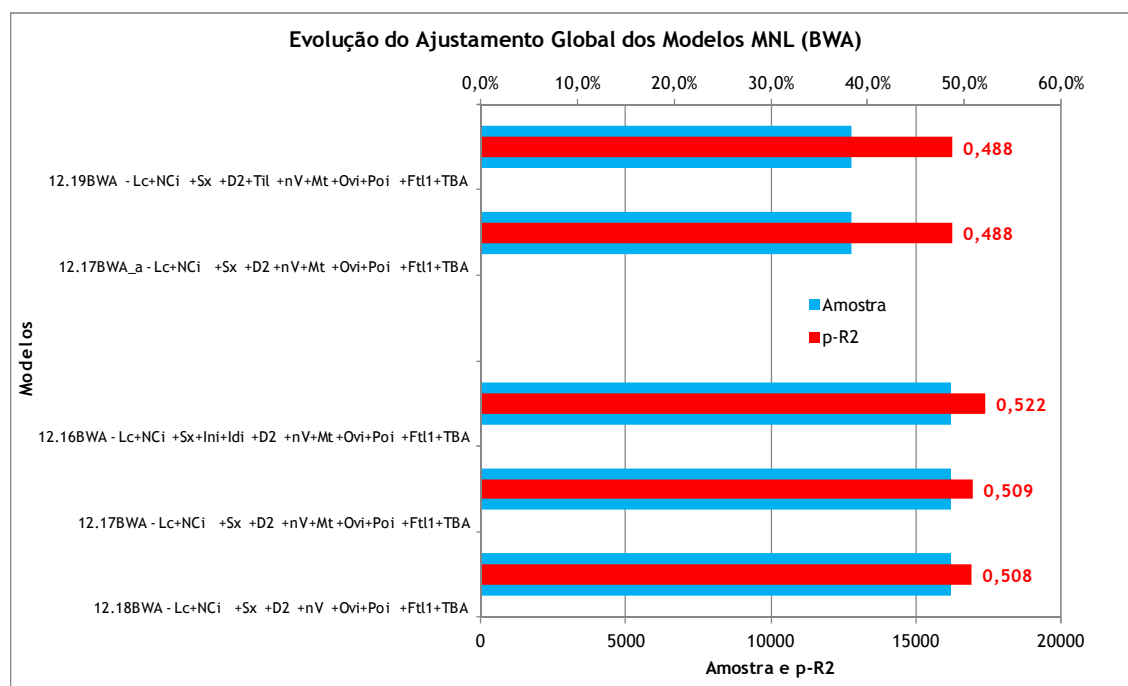


Figura VI. 2 - Evolução do ajustamento global dos modelos MNL com a introdução das variáveis mais relevantes para a análise com os 3 modos mais relevantes.

A adição das variáveis binárias sobre estacionamento ao modelo combinadas com as variáveis sobre os TC mantém a amostra de 17 005, e melhora o ajustamento global em 2,4% (MNL12.16). A exclusão das variáveis socioeconómicas menos relevantes (idade e instrução) do modelo reduz o ajustamento em 3,3% (MNL12.17) e a substituição das variáveis binárias sobre o estacionamento pelas suas homologas contínuas reduz o ajustamento global em 0,9% (MNL12.20) relativamente ao modelo anterior (MNL12.16).

Nestes resultados é possível verificar que as variáveis socioeconómicas e a distância entre GA são as que mais informação introduzem no modelo, e que as variáveis relacionadas com o serviço de TC são as que menor contributo parecem dar. Algumas aparentemente não são muito relevantes uma vez que a sua remoção não tem grande impacto nos ajustamentos globais (como por exemplo a duração das atividades e da duração média da viagem por modo). A remoção das variáveis Idade e instrução leva a uma redução pouco significativa (3,3%) do ajustamento global, pelo que a sua inclusão deve ser devidamente ponderada atendendo à complexidade que adicionam ao modelo com mais 35 parâmetros para calcular ($5 \times 3 + 5 \times 4$). A análise à diferença de ajustamento dos modelos com especificação binária *versus* especificação contínua em modelos combinados, considerando todos os resultados anteriores, foi realizada apenas para as variáveis sobre a oferta de estacionamento, dada a incerteza da sua melhor especificação verificada nos estudos preliminares. Esta análise permitiu verificar que a especificação binária melhorou efetivamente o ajustamento global, mas que esse acréscimo não foi muito acentuado. Estes resultados reforçam a ideia de que

nenhuma das especificações se evidencia, carecendo a decisão sobre a sua escolha de uma ponderação sobre o benefício de um eventual acréscimo de complexidade, inerente ao uso da especificação com combinação de binárias, face ao reduzido aumento do ajustamento global obtido. Para o presente modelo, o reduzido acréscimo de complexidade associado à especificação binária considerada sugere a sua seleção para representar o potencial impacto de políticas de estacionamento.

A redução da amostra devido à diminuição do número de alternativas modais (para apenas 16 197 inquéritos), dada a pouca significância de quase todos os parâmetros para as alternativas Outro Bus e Moto, apresenta uma melhoria do ajustamento global de cerca de 14% (atendendo ao modelo MNL12.17). Para a base de dados com apenas 3 alternativas a introdução das variáveis idade e instrução aumenta o ajustamento global em 2,6% em relação ao modelo MNL12.17BWA, enquanto a remoção do motivo da viagem apenas piora o ajustamento global em 0,2% em relação ao mesmo. É possível verificar que a introdução da variável Duração Média da Viagem por modo, não apresenta melhoramento no ajustamento global do modelo e impõe uma redução da amostra de 21% (12 793) em relação à amostra anterior.

Os melhores ajustamentos globais verificados tanto para a base de dados com as 6 alternativas como para as 3 alternativas, de 0,465 para o modelo MNL7.55 e de 0,522 para o modelo MNL12.16BWA, podem ser considerados muito significativos equivalendo a coeficientes de ajustamentos R^2 de regressões múltiplas lineares de cerca de 0,85 a 0,90.

3.2. Modelos Integrados Fundamentais

Dos resultados anteriores foi possível constatar que a modelação pode ser subdividida em duas situações complementares:

- Modelo sem variáveis sobre políticas de mobilidade;
- Modelo com variáveis sobre políticas de mobilidade.

Esta distinção justifica-se fundamentalmente pela clara preponderância do impacto das variáveis socioeconómicas e das relacionadas com a organização das atividades ao longo do dia, relativamente ao impacto das medidas de políticas de mobilidade. Outra limitação ao uso de todas as variáveis analisadas advém da redução de informação imposta pelas variáveis sobre políticas de mobilidade. Esta redução, devido à falta de toda a informação para parte das alternativas ou para determinados pares GA, retira riqueza/diversidade aos dados de base podendo limitar a apreensão da real influência das diferentes variáveis socioeconómicas analisadas.

Outro fator relevante na definição de dois modelos complementares foi a constatação que, apesar de teoricamente relevantes, os resultados dos modelos revelaram que as medidas de política de mobilidade analisadas, apesar de apresentarem parâmetros com comportamento

coerente atendendo à teoria comportamental, nomeadamente para o modo Bus, não evidenciaram impactos muito significativos nas escolhas modais. Estes resultados podem resultar de diversos fatores:

- Um desconhecimento efetivo dos cidadãos das medidas analisadas conjugadas com uma falha na sua perceção em contextos reais;
- Uma efetiva reduzida influência das mesmas nas escolhas modais diárias dos cidadãos inquiridos, quer consciente quer inconscientemente;
- Pela redução da amostra devido à dificuldade verificada na recolha de informação sobre políticas de mobilidade;
- Devido a deficiente capacidade das variáveis selecionadas de representar o impacto das políticas de mobilidade ou na sua correta quantificação.

A existência destes dados apenas para as viagens com atração no concelho do Porto, levou à limitação da modelação das escolhas modais para essas zonas, e devida à purga na base de dados de informação para as alternativas menos relevantes, apenas para duas alternativas ao automóvel (Bus e Walk). Esta limitação leva a que o modelo com variáveis sobre políticas de mobilidade apenas possa ser utilizado para viagens atraídas para zonas urbanas centrais, onde essas medidas geralmente são mais comuns e incisivas. O modelo sem variáveis sobre políticas de mobilidade aplica-se também a viagens entre zonas suburbanas e interurbanas, onde as políticas de mobilidade analisadas são mais incipientes.

3.2.1. Modelo sem Variáveis sobre Políticas de Mobilidades

Dos modelos sem variáveis de políticas de mobilidade analisados o que melhores resultados apresentou, combinando ajustamento global, tamanho da amostra, quantidade de variáveis e coerência de parâmetros foi o modelo MNL7.57 (quadro VI.32).

É possível constatar que:

- Lc: Apresenta sinal negativo coerente com o senso comum de que a existência de licença de condução torna o automóvel mais atrativo. Maior impacto verifica-se nos modos com Bus regular e o menor impacto é nos Outros Bus. As grandezas semelhantes dos parâmetros sugerem uma influência muito similar para todas as alternativas.
- Nci: Apresenta sinais positivos coerentes para a maioria dos parâmetros. O comportamento decrescente dos parâmetros com o aumento da taxa de motorização familiar é também coerente com o senso comum uma vez que aumenta a probabilidade de utilizar o automóvel. Verifica-se um aumento acentuado da utilidade das alternativas quando não há automóvel disponível

(NC0) e que para os modos Outro Bus e Moto as taxas de motorização elevadas ou não influenciam a utilidade dos mesmos ou a influência é similar à utilidade para a categoria normalizada.

Quadro VI. 32 - Modelo de referência MNL7.57 para a modelação sem políticas de mobilidade

Variável	MNL7.57					Obs.
	BW	B	OB	M	W	
ASC	-8,220	-3,287	-5,355	-7,546	3,727	- Sinais coerentes
Lc	-1,951	-1,935	-1,351	-1,800	-1,810	- Sinais corretos
NC0	4,488	3,401	1,928	1,662	2,459	- Sinais e comportamento coerente
NC1	2,721	1,725	0,624	0,125***	0,737	
NC2	2,303	1,250	0,041***	0,045***	0,416	
NC3	1,265	0,627	-0,356*	-0,365	-0,226**	
NC4a	-	-	-	-	-	
D2	0,045	0,001***	0,037	-0,089	-2,686	- Sinais e grandezas coerentes
Dx1						
R1						
R2						
R3						
R4						
R5						
ln1	0,242***	-0,653	1,722	2,528	-0,179**	- Comportamento coerente para OB e M. Falta de significância para W
ln2	0,456	-0,272	1,668	2,507	0,022***	
ln3	0,781	0,278	0,987	2,307	0,082***	
ln4	-	-	-	-	-	
ld1	-	-	-	-	-	- Parâmetros com comportamento plausível exceto para M e ld5
ld2	2,801	2,111	1,672	1,568	1,698	
ld3	2,119	1,703	0,712	2,407	1,381	
ld4	2,251	1,798	0,476	2,155	1,560	
ld5	2,522	2,032	0,024***	2,201	1,670	
Sx	0,551	0,383	-0,193	-0,714	0,145	- Parâmetros com sinais e grandezas coerentes
Mot	0,087**	-0,449	0,547	3,720	-0,047***	- Sem significância para BW e W e sinal e grandeza coerente para B e M
Til						
De1						
De2						
De3						
De4						
De5						
nV	-0,302	-0,257	-0,298	0,008***	-0,208	- Sinais coerentes. Sem significância para M
Tr	0,907	0,598	1,465	0,931	0,178	- Sinais coerentes e comportamentos plausíveis. Parâmetros para CS quase todos sem significância estatística
Rg	0,762	0,440	1,177	0,399	0,267	
Es	1,242	0,897	1,753	-0,695	0,307	
Lz	-	-	-	-	-	
Cs	0,153**	0,101*	0,213**	-0,232	-0,061***	
Amostra			95426			- Amostra completa
p-R ²			0,445			- Bom ajustamento
LL			-68588,0			
Param.			105			

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

- D2: A distância mais curta entre centróides tem impacto positivo mas pouco relevante para as alternativas com Bus. Tem um impacto negativo pouco relevante também para o Moto mas muito acentuado para o modo Walk. Estes resultados sugerem que o afastamento entre a zona de GA apenas é significativamente relevante para o modo Walk, enquanto para as restantes alternativas não as diferencia de forma acentuada em relação ao automóvel. Na eventual possibilidade de se conseguir diferenciar as funções de utilidade na modelação, sugere-se a inclusão de variáveis binárias que especifique distâncias

inferiores a 2km apenas para o modo pedonal, quer através de uma única variável quer através de duas para distâncias até 1km e entre 1 e 2km.

- Ini: Os parâmetros do nível de instrução não têm significância estatística para o modo Walk, apresentam um comportamento decrescente com o aumento do nível de instrução para os modos Outros Bus e Moto e um comportamento crescente, pouco coerente, para os modos Bus+Walk e Bus. É de salientar a valorização do modo Moto relativamente aos restantes modos para o nível de analfabeto, que poderá dever-se à sua maior utilização pelas idades mais ativas mas com baixos índices de instrução, nomeadamente homens. Este facto poderá justificar também os valores negativos inesperados para os parâmetros do Bus e Walk para os níveis de instrução mais baixos, conjugado com a influência do grupo de crianças até aos 5 anos transportadas pelos pais.
- Idi: Comportamento variável mas coerente. As alternativas ao Auto com exceção da Moto são mais valorizadas pelos adolescentes. A moto é mais valorizada pelas idades ativas, mas o parâmetro para Id5 é algo inesperado.
- Sx: As mulheres valorizam (ou utilizam mais) as alternativas com Bus regular, e valorizam negativamente (ou utilizam menos) os outros Bus e principalmente a Moto. Em relação ao modo pedonal, o modelo revela algo inesperadamente uma valorização de cerca de metade da do modo Bus para as mulheres.
- Mot: A disponibilidade diária de Moto no agregado aparentemente não tem influência para os modos Bus+Walk e Walk. Apresenta uma grande valorização (esperada) para o modo Moto e uma perda de utilidade para o modo Bus. Este resultado aliado ao do nível de instrução parece sugerir que para a população analfabeta a Moto é uma alternativa mais utilizada que o modo Bus, muito provavelmente pela sua disponibilidade.
- nV: O aumento do número de viagens diárias tem um impacto negativo semelhante para quase todas as alternativas ao Auto, com exceção da Moto. Aparentemente esta variável não apresenta impacto na utilidade do modo Moto, comparativamente ao Auto (talvez devido à flexibilidade que ambos permitem).
- Motivo da Viagem: Os parâmetros revelam que todas as alternativas exceto o modo Moto, são mais valorizadas nas viagens para a escola que o automóvel. Para compras e serviços, aparentemente nenhuma alternativa apresenta relevância estatística que sugira uma maior utilidade que o modo Auto, pelo menos comparativamente às viagens para Lazer. Nas viagens para o trabalho a alternativa mais valorizada é o Outro Bus, bem como para a escola e para o regresso a casa.

Assim, o modelo sugerido resulta nas seguintes funções utilidade para cada um dos modos:

$$\begin{aligned}
U_{BW} &= -8,220 + (-1,951 \times Lc) + (4,488 \times NC0 + 2,721 \times NC1 + 2,303 \times NC2 + 1,265 \times NC3) + (0,045 \times D2) \\
&\quad + (0,242 \times \ln1 + 0,456 \times \ln2 + 0,781 \times \ln3) + (2,801 \times Id2 + 2,119 \times Id3 + 2,251 \times Id4 + 2,522 \times Id5) + (0,551 \times Sx) \\
&\quad + (0,087 \times Mot) + (-0,302 \times nV) + (0,907 \times Tr + 0,762 \times Rg + 1,242 \times Es + 0,153 \times CS) \\
U_B &= -3,287 + (-1,935 \times Lc) + (3,401 \times NC0 + 1,725 \times NC1 + 1,250 \times NC2 + 0,627 \times NC3) + (0,001 \times D2) \\
&\quad + (-0,653 \times \ln1 - 0,272 \times \ln2 + 0,278 \times \ln3) + (2,111 \times Id2 + 1,703 \times Id3 + 1,798 \times Id4 + 2,032 \times Id5) + (0,383 \times Sx) \\
&\quad + (-0,449 \times Mot) + (-0,257 \times nV) + (0,598 \times Tr + 0,440 \times Rg + 0,897 \times Es + 0,101 \times CS) \\
U_{OB} &= -5,355 + (-1,351 \times Lc) + (1,928 \times NC0 + 0,624 \times NC1 + 0,041 \times NC2 + 0,356 \times NC3) + (0,037 \times D2) \\
&\quad + (1,722 \times \ln1 + 1,668 \times \ln2 + 0,987 \times \ln3) + (1,672 \times Id2 + 0,712 \times Id3 + 0,476 \times Id4 + 0,024 \times Id5) + (-0,193 \times Sx) \\
&\quad + (0,547 \times Mot) + (-0,298 \times nV) + (1,465 \times Tr + 1,177 \times Rg + 1,753 \times Es + 0,213 \times CS) \\
U_M &= -7,546 + (-1,800 \times Lc) + (1,662 \times NC0 + 0,125 \times NC1 + 0,045 \times NC2 - 0,365 \times NC3) + (-0,089 \times D2) \\
&\quad + (2,528 \times \ln1 + 2,507 \times \ln2 + 2,307 \times \ln3) + (1,568 \times Id2 + 2,407 \times Id3 + 2,155 \times Id4 + 2,201 \times Id5) + (-0,714 \times Sx) \\
&\quad + (3,720 \times Mot) + (0,008 \times nV) + (0,931 \times Tr + 0,399 \times Rg - 0,695 \times Es - 0,232 \times CS) \\
U_W &= 3,727 + (-1,810 \times Lc) + (2,459 \times NC0 + 0,737 \times NC1 + 0,416 \times NC2 - 0,226 \times NC3) + (-2,686 \times D2) \\
&\quad + (-0,179 \times \ln1 + 0,022 \times \ln2 + 0,082 \times \ln3) + (1,698 \times Id2 + 1,381 \times Id3 + 1,560 \times Id4 + 1,670 \times Id5) + (0,145 \times Sx) \\
&\quad + (-0,047 \times Mot) + (-0,208 \times nV) + (0,178 \times Tr + 0,267 \times Rg + 0,307 \times Es - 0,061 \times CS) \\
U_A &= 0
\end{aligned}$$

Os resultados das análises realizadas sugerem que as variáveis fundamentais são a Licença de condução (Lc), a taxa de motorização familiar (NCi), a distância mais curta entre centróides, principalmente para o modo pedonal (D2) e o sexo do inquirido (Sx) devido à sua estabilidade ao longo das modelações.

Como variáveis complementares, que introduzem alguma informação relevante mas com menor impacto nas escolhas, verificou-se o nível de instrução (lni) e o grupo etário (Idi), geralmente em combinação, a disponibilidade diária de Moto (Mot), o número total de viagens diárias (nV) e o motivo da viagem.

O ajustamento global do modelo proposto é de $p-R^2 = 0,445$, correspondente a um coeficiente de regressão linear equivalente de cerca de 0,80, ou seja, muito significativo.

3.2.2. Modelo com Variáveis sobre Políticas de Mobilidades

Dos modelos com variáveis de políticas de mobilidade analisados o que melhores resultados apresentou, combinando ajustamento global, tamanho da amostra e coerência de parâmetros foi o modelo MNL12.16BWA (quadro VI.33).

Quadro VI. 33 - Modelo de referência MNL12.16BWA para a modelação com políticas de mobilidade

Variável	MNL12.16BWA		Obs.
	B	W	
ASC	-4,156	2,919	- Sinais coerentes
Lc	-1,748	-1,528	- Sinais coerentes
NC0	4,101	2,996	- Sinais e comportamento coerente
NC1	1,949	0,801	
NC2	1,712	0,464	
NC3	0,820	0,039***	
NC4a	-	-	
D2	0,005***	-2,444	- Parâmetros coerentes
Dx1			
R1			
R2			
R3			
R4			
R5			- Sinais coerentes. In1 perde significância e influencia muito semelhante
In1	0,093***	0,265***	
In2	0,436	0,420	
In3	0,490	0,396	
In4	-	-	
Id1	-	-	- Sinais coerentes. Comportamentos plausíveis
Id2	2,019	1,628	
Id3	2,158	1,600	
Id4	1,942	1,592	
Id5	2,358	1,632	
Sx	0,474	0,378	- Parâmetros coerentes
Mot			- Parâmetros coerentes
Til			
De1			
De2			
De3			
De4			
De5			- Perda de significância para W. Parâmetros plausíveis para B
nV	-0,138	-0,155	
Tr	0,426	0,192**	
Rg	0,266	0,352	
Es	0,342	0,212**	
Lz	-	-	- Comportamento coerente para B
Cs	0,323	0,461	
OV1	0,651	-1,106	
OV2	0,342	-1,215	
OV3	0,192*	-0,945	
OV4	-	-	- Comportamento coerente
PO1	-	-	
PO2	0,140**	0,431	
PO3	0,418	1,479	
Ftl1	0,007	0,010	- Sinais coerentes
TBA	-0,209	0,087***	- Sinal coerente e falta de significância para o W
Amostra	16197		- 1/6 da Amostra
p-R ²	0,522		- Ajustamento muito bom
LL	-8493,6		
Param.	54		

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

É possível constatar que:

- Lc: Apresentam sinal negativo coerente com o senso comum de que a existência de licença de condução torna o automóvel mais atrativo. Impacto semelhante em ambas as alternativas;
- Nci: Apresentam sinais positivos coerentes. O comportamento decrescente dos parâmetros com o aumento da taxa de motorização familiar é também coerente com o senso comum uma vez que aumenta a probabilidade de utilizar o automóvel. O impacto é significativamente mais acentuado no modo Bus do que no Walk;
- D2: A distância mais curta entre centróides não tem significância estatística para o modo Bus e mantém um impacto negativo muito acentuado para o modo pedonal. Estes resultados sugerem que a distância mais curta apenas é relevante para o modo pedonal, não permitindo diferenciar a atratividade dos modos motorizados.
- Ini: Os parâmetros do nível de instrução analfabeto não têm significância estatística provavelmente devida à purga deste nível da base de dados, nomeadamente de crianças até aos 5 anos. Os restantes níveis apresentam sinais positivos coerentes mas grandezas muito similares sugerindo que o nível de instrução não permite diferenciar distintas preferências por modo;
- Idi: Também as variáveis dos escalões etários apresentam sinal positivo coerente mas grandezas similares entre diferentes escalões. É possível constatar um comportamento enquadrável na teoria comportamental, mas a semelhança de valores para o modo Walk dificulta a diferenciação de preferências através das mesmas;
- Sx: As mulheres valorizam (ou utilizam mais) as alternativas como seria espectável. Essa valorização é muito semelhante mas sugere uma maior atratividade do modo Bus.
- nV: O aumento do número de viagens diárias tem um impacto negativo semelhante para as alternativas, sugerindo que o aumento das necessidades de mobilidade diárias potencia o uso do automóvel pela sua maior versatilidade;
- Motivo da Viagem: Os parâmetros revelam que viagens para a escola e para o trabalho a pé não têm significância estatística, provavelmente devido à purga destas viagens da base de dados. Também para estas variáveis os parâmetros apresentam valores muito similares. Aparentemente o modo Bus é mais valorizado para viagem para o trabalho enquanto o modo Walk tem mais utilidade nas viagens para compras e serviços.
- OVi (Razão entre a oferta de estacionamento e o total de viagens atraídas): Os parâmetros para o modo Bus apresentam um comportamento coerente de decréscimo de atratividade com o aumento da densidade de oferta face à

procura de estacionamento, uma vez que propicia o uso o automóvel. Para o modo pedonal, o comportamento é variável mas apresenta grandezas similares para as diferentes categorias. O sinal negativo é contrário ao esperado, uma vez que sugere uma diminuição da utilidade do modo pedonal com a diminuição da oferta de estacionamento. Este comportamento, como já foi referido, poderá dever-se a uma correlação desta variável com a localização da viagem na área urbana, uma vez que a maior utilização do modo pedonal está geralmente relacionada com zonas urbanas centrais onde, paralelamente, também há maior concentração de estacionamentos. Esta possibilidade leva à sugestão de uma ponderação prudente sobre a inclusão desta variável na utilidade do modo pedonal.

- POi (Percentagem de estacionamentos pagos): Os parâmetros apresentam sinal positivo e comportamento crescente coerente, verificando-se um maior impacto na utilidade do modo Walk. Estes resultados vêm corroborar o referido na variável anterior, já que o modo Walk é mais utilizado nas áreas centrais onde também se verifica que o estacionamento pago é mais frequente.
- Ftl1 (Frequência média horária diária): Verifica-se que os parâmetros têm sinal positivo coerente sugerindo um aumento da utilidade do Bus com a sua frequência. O modo Walk aparentemente também é mais utilizado, apesar de provavelmente estar mais relacionado com o facto de as maiores frequências acontecerem também nas áreas centrais onde o modo pedonal também é mais utilizado. Para ambas as alternativas, no entanto, o modelo revela que a sua influência é muito reduzida e semelhante;
- TBA (Razão tempo de viagem em Bus e tempo de viagem Auto): Os parâmetros são coerentes, com sinal negativo para o Bus e sem significância estatística para o modo Walk. Estes resultados sugerem que o aumento da diferença do tempo de viagem no modo Bus em relação ao modo Auto (geralmente nas áreas suburbanas e interurbanas) diminui a atratividade do modo Bus e que não tem significância para o modo pedonal.

As funções de utilidade para os três modos considerados serão assim:

$$\begin{aligned}
 U_B = & -4,156 + (-1,748 \times Lc) + (4,101 \times NC0 + 1,949 \times NC1 + 1,712 \times NC2 + 0,820 \times NC3) + (0,005 \times D2) \\
 & + (0,474 \times Sx) + (2,019 \times Id2 + 2,158 \times Id3 + 1,942 \times Id4 + 2,358 \times Id5) \\
 & + (0,093 \times \ln1 + 0,436 \times \ln2 + 0,490 \times \ln3) + (-0,138 \times nV) + (0,426 \times Tr + 0,266 \times Rg + 0,342 \times Es + 0,323 \times CS) \\
 & + (0,651 \times OV1 + 0,342 \times OV2 + 0,192 \times OV3) + (0,140 \times PO2 + 0,418 \times PO3) \\
 & + (0,007 \times Ftl1) + (-0,209 \times TBA) \\
 U_W = & 2,919 + (-1,528 \times Lc) + (2,998 \times NC0 + 0,801 \times NC1 + 0,464 \times NC2 + 0,039 \times NC3) + (-2,444 \times D2) \\
 & + (0,378 \times Sx) + (1,628 \times Id2 + 1,600 \times Id3 + 1,592 \times Id4 + 1,632 \times Id5)
 \end{aligned}$$

$$\begin{aligned}
& + (0,265 \times \ln 1 + 0,420 \times \ln 2 + 0,396 \times \ln 3) + (-0,155 \times nV) + (0,192 \times Tr + 0,352 \times Rg + 0,212 \times Es + 0,461 \times CS) \\
& + (-1,106 \times OV1 - 1,215 \times OV2 - 0,945 \times OV3) + (0,431 \times PO2 + 1,479 \times PO3) \\
& + (0,010 \times Ft1) + (0,087 \times TBA) \\
U_A = & 0
\end{aligned}$$

Os resultados das análises realizadas sugerem que as variáveis fundamentais continuam a ser a Licença de condução (Lc), a taxa de motorização familiar (NCi), a distância mais curta entre centróides (D2), principalmente para o modo pedonal, e o sexo do inquirido (Sx).

Como variáveis complementares, que introduzem alguma informação relevante mas com menor impacto nas escolhas, verificou-se o nível de instrução (Ini) e o grupo etário (Idi), geralmente em combinação, o número total de viagens diárias (nV) e o motivo da viagem.

As variáveis sobre políticas de mobilidade relacionadas com a gestão de estacionamento atendendo aos resultados também se consideraram complementares, uma vez que não melhoram significativamente o ajustamento global do modelo mas, ainda assim, apresentam valores coerentes para o modo Bus. As variáveis relacionadas com os serviços de TC têm um impacto bastante reduzido no ajustamento global, mas, ainda assim, apresentam valores coerentes para o modo Bus. Em estudos posteriores deverá ponderar-se a sua retirada da função de utilidade para o modo Walk.

O ajustamento global do modelo proposto é de $p-R^2 = 0,522$, correspondente a um coeficiente de regressão linear equivalente de cerca de 0,95, ou seja, muito elevado.

4. Avaliação da Consistência das Simulações dos Modelos através dos Resultados da Base de Dados

Para avaliar a consistência das estimativas realizadas pelos modelos considerados foram calculadas as repartições modais através dos modelos considerando os dados existentes na base de dados. Posteriormente estas foram comparadas com a repartição modal efetivamente verificada na respetiva amostra da base de dados. Todas as comparações foram realizadas com resultados extrapolados atendendo aos ponderadores da base de dados original. Os modelos e amostras consideradas foram as seguintes:

- MNL7.57: Modelo com ajustamento global elevado incluindo as variáveis socioeconómicas mais relevantes e as variáveis caracterizadoras da viagem e da organização das atividades diárias que permitem a utilização da amostra completa (95 426);
- MNL12.16: Modelo que melhor ajustamento global apresentou nas análises exploratórias. Inclui as variáveis socioeconómicas mais relevantes, as caracterizadoras das viagens, da organização das atividades diárias, as relacionadas

com as políticas de estacionamento e as caracterizadoras do serviço de Bus. Apenas aplicada a amostra com dados para todas as variáveis consideradas (17 005).

Os resultados dessa comparação podem ser verificados nas figuras VI.3 e VI.4.

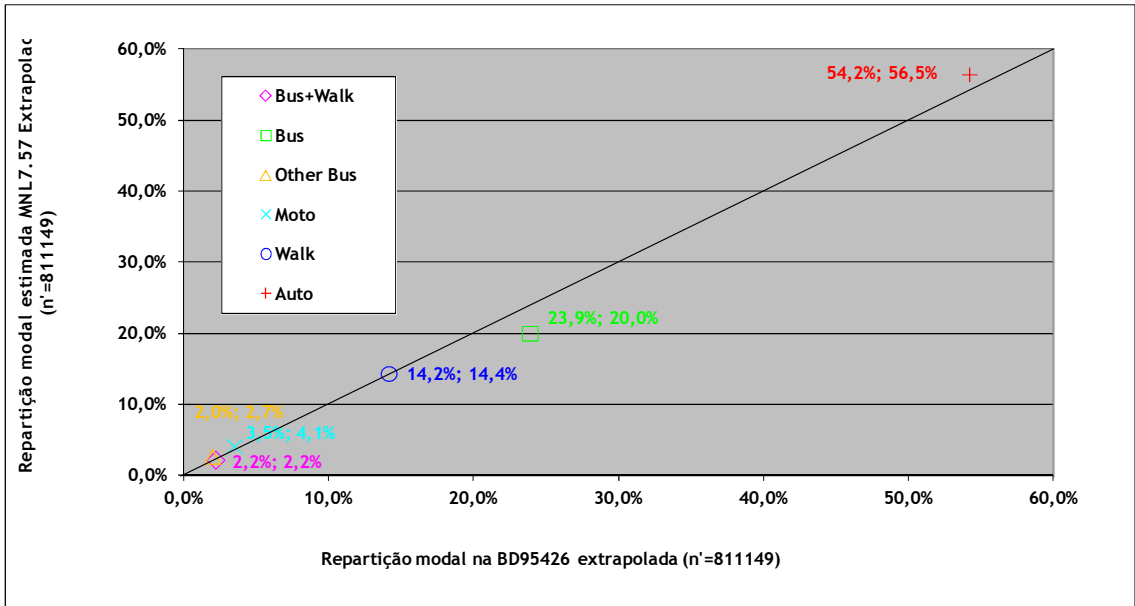


Figura VI. 3 - Diferenças entre a repartição modal verificadas na amostra e as estimativas obtidas pelo modelo MNL7.57

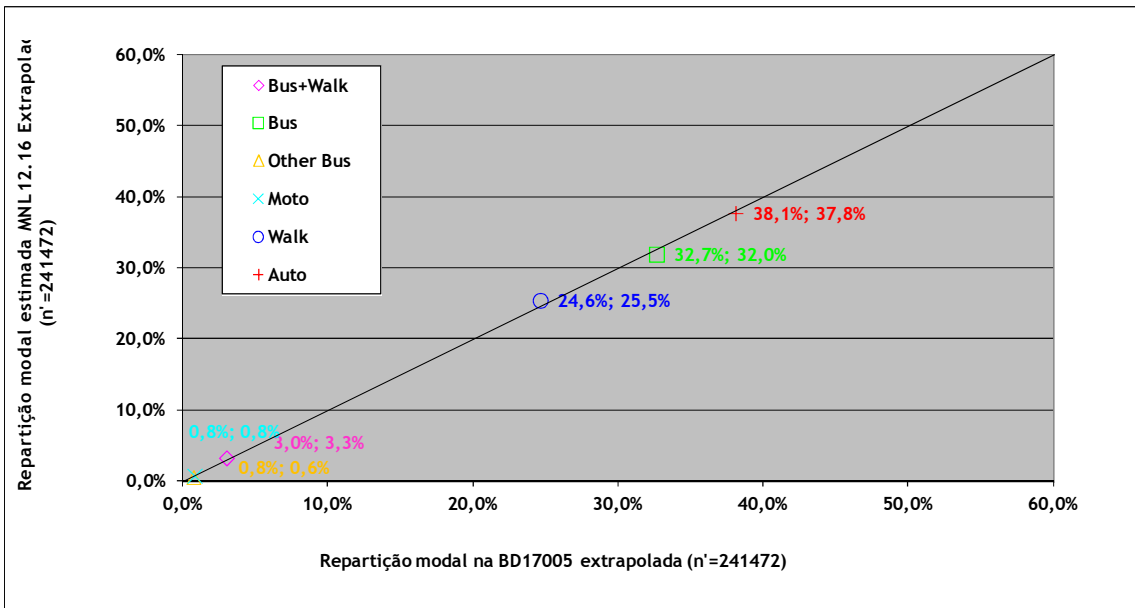


Figura VI. 4 - Diferenças entre a repartição modal verificadas na amostra e as estimativas obtidas pelo modelo MNL12.16

É possível constatar que apesar de apresentarem diferenças de grandeza, as mesmas são relativamente reduzidas, sugerindo que os modelos conseguem simular bastante bem as tendências existentes na base de dados. As diferenças mais significativas em termos globais verificam-se para os modos Auto e Bus, que em termos relativos se compreendem entre -16,1% e 4,2%, relativamente à repartição modal existente na base de dados. Verifica-se, ainda, que as menores diferenças são alcançadas com o modelo MNL12.16, resultado espectável atendendo a que foi um dos modelos que melhor ajustamento global apresentou.

5. Análises Adicionais Relativamente à Não Linearidade e Interdependência das Variáveis Explicativas/Independentes

Atendendo às simplificações subjacentes à fundamentação teórica dos modelos Logit Multinomiais, nomeadamente a impossibilidade de integrar facilmente correlações entre variáveis e erros, a assunção da independência entre alternativas relevantes, e a imposição de erros com distribuições independentes e igualmente distribuídas, sugerem a sua utilização, normalmente pertinente, para análises sobre eventuais comportamentos e relevância das variáveis em análise. No entanto, para a obtenção de modelos mais fidedignos, é necessária a estimação de modelos com fundamentação mais complexa. No sentido de explorar eventuais tendências não detetadas devido à falta de capacidade explicativa dos modelos exploratórios simplificados já desenvolvidos, realizaram-se algumas pesquisas e modelações da base de dados adicionais, combinando as variáveis consideradas mais relevantes.

5.1. Tendências Verificadas na Base de Dados Relativamente às Principais Variáveis Sociais: Sexo vs. Licença de Condução vs. Idade vs. Taxa de Motorização

Num primeira análise da base de dados procuraram-se eventuais relações entre as variáveis mais valorizadas: taxa de motorização, disponibilidade de licença de condução e afastamento entre as zonas de GA, atendendo ao sexo e ao escalão etário dos indivíduos. A quase exclusividade do uso do modo pedonal para distâncias inferiores a 1km levou à exclusão destas distâncias da pesquisa, realizando-se apenas a distinção entre distâncias entre 1 e 2km e distâncias superiores a 2km, procurando-se deste modo reduzir fortemente eventuais efeitos resultantes do impacto da variável “Distância da Deslocação”. Apesar de se ter optado pelo uso da taxa de motorização familiar *per capita* nos modelos sugeridos no ponto 3.2, fundamentalmente para não reduzir a amostra devido à falta de dados sobre todos os adultos em alguns agregados (amostra designada de NCA_b), os resultados exploratórios sugerem que a taxa de motorização familiar por adulto além de resultados similares também apresenta um potencial explicativo elevado (ver modelos MNL7.59_NCAb e MNL12.22_NCAb). Assim, na restante pesquisa realizada, optou-se por avaliar as possíveis relações com a taxa de motorização familiar por adulto no agregado, com vista à sua eventual substituição em estudos futuros. Os totais por escalão etário apresentados referem-se aos valores médios ponderados para as diferentes taxas de motorização familiar por adulto.

Os resultados dessa análise (apresentados nos Anexos 1 e 2 do Capítulo VI) foram agregados atendendo à similaridade de conveniência entre modos em Auto+Moto e Bus/Walk+Bus+Outro Bus (=All Bus) e são apresentados de forma gráfica nas figuras VI.5 e VI.6 para uma melhor interpretação.

Assim, para viagens entre 1 e 2 km é possível destacar que:

- O comportamento para ambos os sexos até aos 9 anos é muito similar, e verifica-se uma utilização do Auto+Moto ao mesmo nível das mulheres com licença de condução;
- As crianças de 0-9 anos, de ambos os sexos, utilizam muito mais o Auto e bastante menos o Walk que as restantes idades;
- A indisponibilidade de licença de condução, para ambos os sexos, leva a uma redução da utilização do Auto+Moto dos 60-80% para 15-35%, e a um ligeiro aumento do uso do modo Walk, exceto para as crianças de 0-9 anos;
- A existência ou não de licença de condução não afeta significativamente o uso de All Bus para ambos os sexos;
- As mulheres com licença de condução têm tendência para utilizar mais All Bus (exceto dos 20-34 anos) que os homens;
- Os homens com +65anos com licença de condução utilizam mais o Walk e menos o Auto que as mulheres da mesma idade;
- A existência de veículo no agregado leva a um aumento acentuado do uso do Auto quer de homens quer de mulheres;
- O modo Walk é uma alternativa relevante principalmente quando não há veículo disponível, com maior predominância nas mulheres;
- O uso do Auto sem licença de condução é superior nas mulheres que nos homens para taxas de motorização intermédias;

Para viagens superiores a 2 km é possível destacar que:

- Os comportamentos mostram tendências mais definidas que para distâncias inferiores, podendo dever-se a uma amostra substancialmente maior;
- É possível destacar três grupos de comportamento relativamente homogêneos:
 - Homens ou mulheres entre os 20-65 anos;
 - Homens ou mulheres com 10-19 e +65 anos;
 - Crianças entre os 0 e 9 anos;

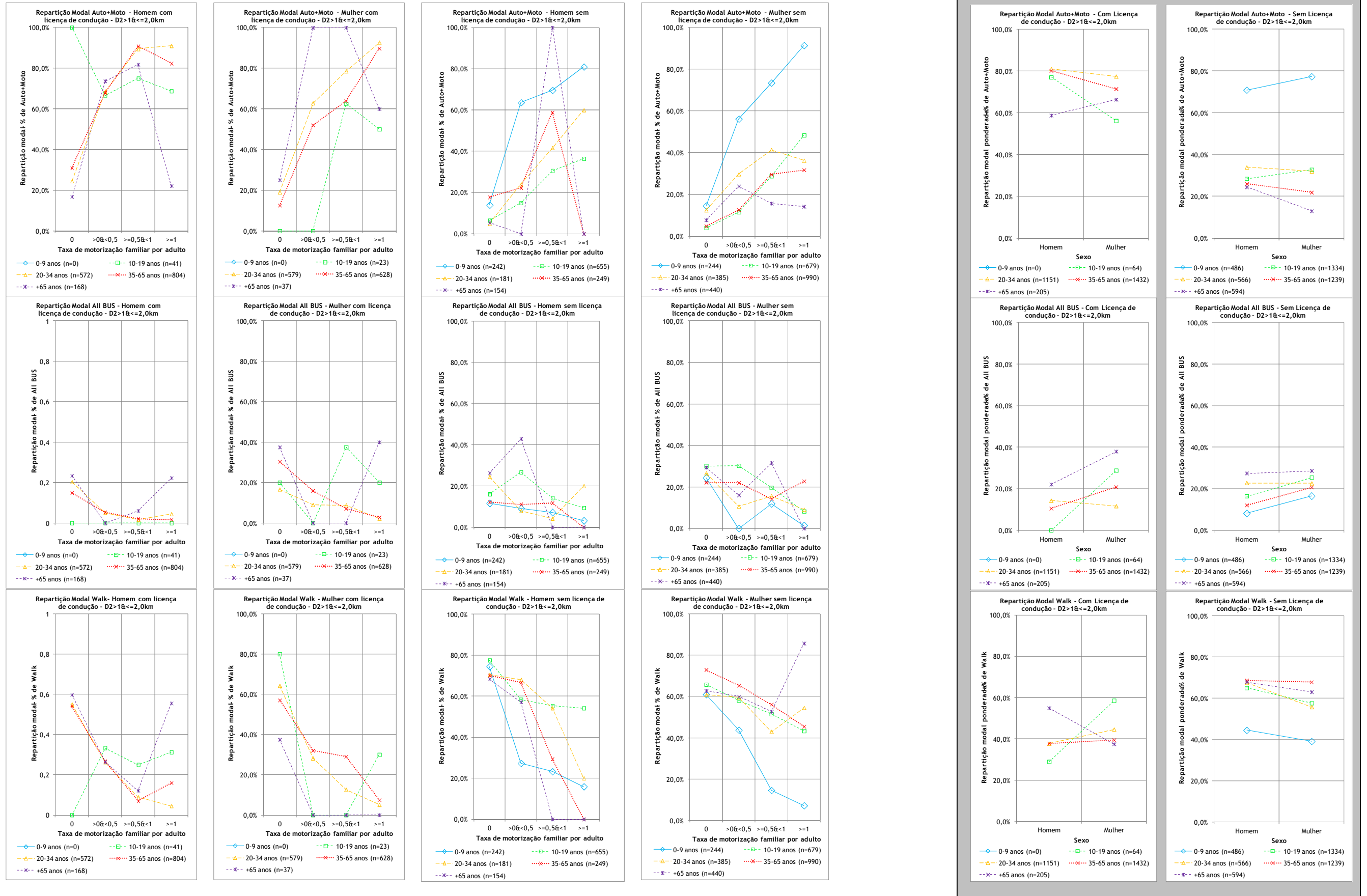


Figura VI. 5 - Gráficos com resultados da combinação das variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >1km e ≤2km

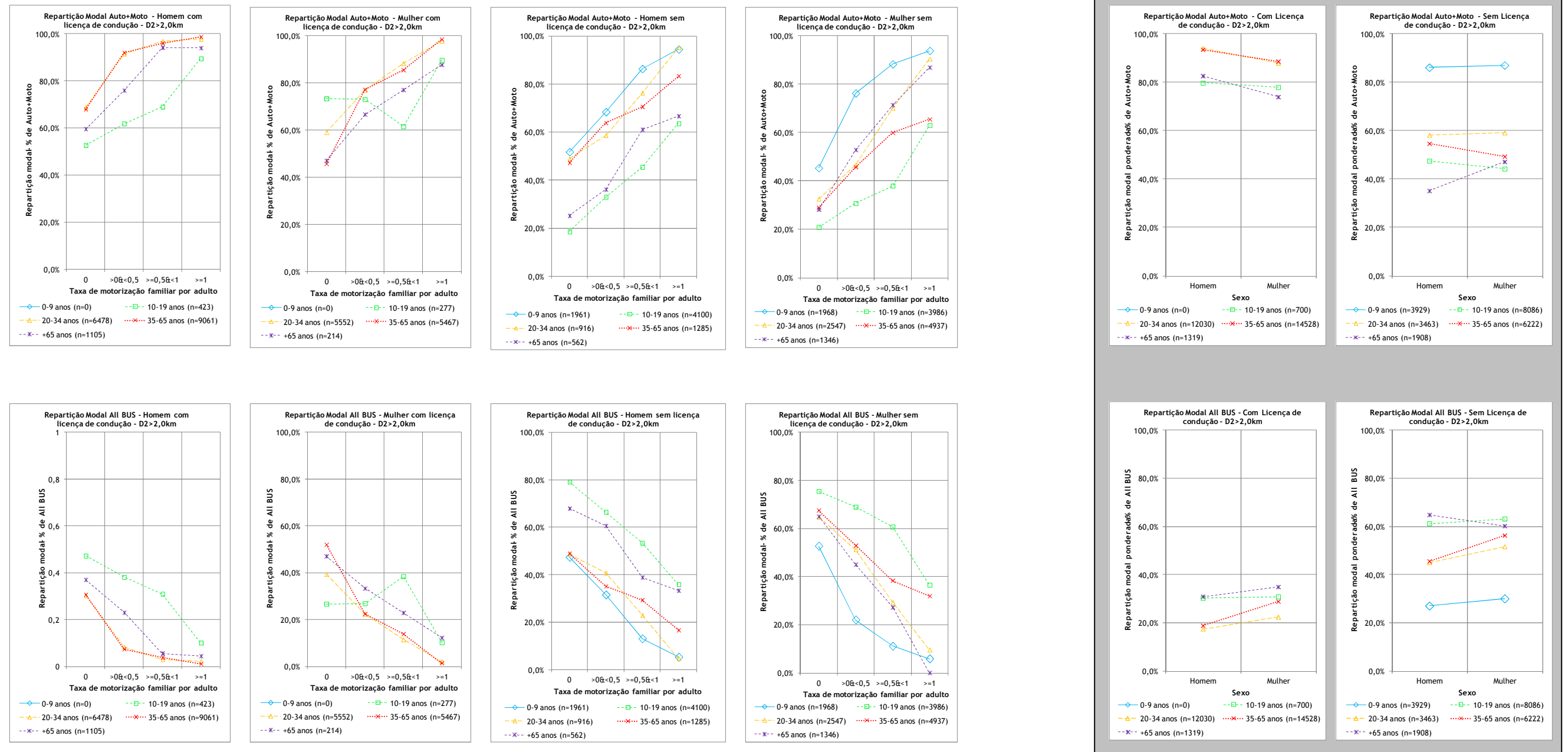


Figura VI. 6 - Gráficos com resultados da combinação das variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >2km

- Com licença de condução:

- Para idades dos 20 aos 65 anos, verifica-se um aumento significativo do uso do Auto+Moto quando há disponibilidade de pelo menos um automóvel no agregado. Este aumento é mais pronunciado nos homens logo a partir de taxas de motorização baixas e mais linear para as mulheres;
- Para idades entre os 10 e 19 e +65, apesar de algumas diferenças aparentemente seguem tendências similares. A utilização do Auto+Moto com licença de condução é menor que para o caso anterior e, aparentemente, com um crescimento mais linear com o aumento da taxa de motorização. As tendências para homens e mulheres são similares mas constata-se um menor uso do Auto+Moto pelas mulheres;

- A inexistência de licença de condução leva a comportamentos semelhantes entre homens e mulheres, mais concretamente na linearidade do crescimento do uso do Auto+Moto:

- Para idades dos 0-9 anos há uma maior utilização do automóvel, principalmente pelos homens, mas para taxas de motorização familiar perto de 1 os comportamentos são muito semelhantes;
- Para idades dos 20 aos 65 anos, os homens utilizam o Auto+Moto quase tanto como as crianças, mas as mulheres consideravelmente menos;
- Para idades dos 10 aos 19 e +65, o comportamento é destino, verificando-se uma maior utilização do Auto+Moto pelas mulheres mais velhas que pelos homens, e uma utilização semelhante do mesmo entre homens e mulheres adolescentes. Mais uma vez a evolução de uso do Auto+Moto pelos homens apresenta uma evolução mais rápida com o aumento da taxa de motorização que as mulheres;

Podem, assim, verificar-se comportamentos distintos para diferentes faixas etárias, consoante o sexo de quem se desloca. A existência de pelo menos um veículo por adulto no agregado potencia o uso do Auto tanto para homens como para mulheres de todos os escalões etários. As mulheres mais velhas sem carta de condução têm maior tendência a utilizar o automóvel que os homens nas mesmas circunstâncias. Esta constatação pode dever-se ao facto de não se conseguir distinguir condutores de passageiros. Sociologicamente o comportamento entre sexos é diferenciado e tradicionalmente as mulheres mais velhas, mesmo provenientes de agregados com meios económicos para terem acesso a automóvel, não costumavam obter licença de condução ou conduzir, pelo que a assunção de que estas mulheres utilizem o automóvel mas não como condutoras é bastante plausível. O comportamento para homens e mulheres menores que 20 anos é muito semelhante, não apresentando diferenças de comportamento muito significativas. Para os homens em idade ativa, com automóvel

disponível e licença de condução, verifica-se um aumento muito acentuado do uso do Auto+Moto logo a partir de taxas de motorização baixas.

Estes resultados sugerem uma tendência “machista” para a “apropriação” do primeiro veículo da família pelo homem do casal. Aparentemente, apenas quando há mais de um carro no agregado, é que as mulheres tendem a ser condutoras em vez de passageiras. Esta tendência justifica alguns dos resultados referidos anteriormente, podendo distorcer parte das conclusões a retirar dos modelos, pelo que se sugere uma análise mais aprofundada desta tendência em estudos futuros mais complexos.

A constatação destas tendências sugere que a utilização de variáveis isoladas nos modelos de escolha discreta pode não conseguir representar convenientemente as diferenças de comportamento dentro de determinados grupos, quer etários quer de sexo, que implicam efeitos de inter-relação e interdependência entre os mesmos. Nesse sentido, ponderou-se a possibilidade de implementar variáveis binárias, representativas de combinações das variáveis analisadas com comportamento relativamente homogêneo, e avaliar a sua aptidão para modelar esses comportamentos mais específicos. Simultaneamente procurou-se avaliar as vantagens associadas a eventuais ganhos de ajustamento global obtidos com a coerência dos resultados comparativamente ao aumento de complexidade, rigidez e especificidade, implícito nas variáveis a definir.

5.2. Modelos Socioeconómicos Adicionais

Como foi referido os resultados do ponto anterior sugerem tendências relativamente consistentes para grupos de determinadas condições sociais, cuja representação através de variáveis isoladas poderá ser limitativa e redutora. No sentido de avaliar a possibilidade de utilizar variáveis binárias de combinações entre variáveis socioeconómicas para identificar esses comportamentos diferenciados entre os grupos referidos, realizaram-se modelos exploratórios MNL e avaliaram-se os seus resultados em termos de coerência de sinal, grandeza e comportamento dos parâmetros e melhoria de ajustamento global.

Para testar de forma progressiva a eventual existência de interdependências entre várias das variáveis explicativas anteriormente utilizadas, sugerida pelos estudos preliminares apresentados no subponto anterior, foram realizadas 2 análises, considerando duas especificações distintas cujos resultados são posteriormente comparados com os do modelo incluindo todas as variáveis socioeconómicas e relacionadas com a organização das atividades (MNL7.57a):

- Variáveis binárias de possíveis combinações das variáveis socioeconómicas licença de condução (Lc), sexo (Sx), e Idade (Idi) (quadro VI.34);

- Variáveis binárias de combinações caracterizadoras de grupos sociais com comportamentos similares relativamente às variáveis anteriores, mas articuladas com taxas de motorização familiar por adulto (quadro VI.35);

Quadro VI. 34 - Variáveis binárias de combinação de variáveis socioeconómicas para o modelo MNL99.3a

Variável	Lc	Sx	ldb=1	ldb=2	ldb=3	ldb=4	ldb=5	NA0	NA1	NA2	NA3
X1	1	0	n.a.	1	0	0	0	n.a.	n.a.	n.a.	n.a.
X2	1	0	n.a.	0	1	0	0	n.a.	n.a.	n.a.	n.a.
X3	1	0	n.a.	0	0	1	0	n.a.	n.a.	n.a.	n.a.
X4	1	0	n.a.	0	0	0	1	n.a.	n.a.	n.a.	n.a.
X5	1	1	n.a.	1	0	0	0	n.a.	n.a.	n.a.	n.a.
X6	1	1	n.a.	0	1	0	0	n.a.	n.a.	n.a.	n.a.
X7	1	1	n.a.	0	0	1	0	n.a.	n.a.	n.a.	n.a.
X8	1	1	n.a.	0	0	0	1	n.a.	n.a.	n.a.	n.a.
X9	0	0	1	0	0	0	0	n.a.	n.a.	n.a.	n.a.
X10	0	0	0	1	0	0	0	n.a.	n.a.	n.a.	n.a.
X11	0	0	0	0	1	0	0	n.a.	n.a.	n.a.	n.a.
X12	0	0	0	0	0	1	0	n.a.	n.a.	n.a.	n.a.
X13	0	0	0	0	0	0	1	n.a.	n.a.	n.a.	n.a.
X14	0	1	1	0	0	0	0	n.a.	n.a.	n.a.	n.a.
X15	0	1	0	1	0	0	0	n.a.	n.a.	n.a.	n.a.
X16	0	1	0	0	1	0	0	n.a.	n.a.	n.a.	n.a.
X17	0	1	0	0	0	1	0	n.a.	n.a.	n.a.	n.a.
X18	0	1	0	0	0	0	1	n.a.	n.a.	n.a.	n.a.

Lc - 1= Tem; 0=Não tem Sx - 0=M, 1=F ldi - 1= Considerado; 0= Não Considerado;
NAi - 1= Considerado; 0= Não Considerado n.a. - Não aplicado

Quadro VI. 35 - Variáveis binárias de combinação de variáveis socioeconómicas com taxas de motorização familiar por adulto para o modelo MNL99.9b

Variável	Lc	Sx	ldb=1	ldb=2	ldb=3	ldb=4	ldb=5	NA0	NA1	NA2	NA3
Y10	1	0	n.a.	0	1	1	1	1	0	0	0
Y11	1	0	n.a.	0	1	1	1	0	1	0	0
Y12	1	0	n.a.	0	1	1	1	0	0	1	0
Y13	1	0	n.a.	0	1	1	1	0	0	0	1
W20	1	1	n.a.	0	1	1	1	1	0	0	0
W21	1	1	n.a.	0	1	1	1	0	1	0	0
W22	1	1	n.a.	0	1	1	1	0	0	1	0
W23	1	1	n.a.	0	1	1	1	0	0	0	1
W30	n.a.	n.d.	1	0	0	0	0	1	0	0	0
W31	n.a.	n.d.	1	0	0	0	0	0	1	0	0
W32	n.a.	n.d.	1	0	0	0	0	0	0	1	0
W33	n.a.	n.d.	1	0	0	0	0	0	0	0	1
W40	1	n.d.	n.a.	1	0	0	0	1	0	0	0
W41	1	n.d.	n.a.	1	0	0	0	0	1	0	0
W42	1	n.d.	n.a.	1	0	0	0	0	0	1	0
W43	1	n.d.	n.a.	1	0	0	0	0	0	0	1
W50	0	n.d.	0	1	1	1	1	1	0	0	0
W51	0	n.d.	0	1	1	1	1	0	1	0	0
W52	0	n.d.	0	1	1	1	1	0	0	1	0
W53	0	n.d.	0	1	1	1	1	0	0	0	1

Lc - 1= Tem; 0=Não tem Sx - 0=M, 1=F ldi - 1= Considerado; 0= Não Considerado; NAi - 1= Considerado; 0= Não Considerado
Considerado n.a. - Não aplicado n.d. - Não diferenciado

Tal como foi referido, a utilização da taxa de motorização familiar por adulto limita a amostra de agregados familiares que apresentava informação completa de todos os seus elementos e das suas viagens (amostra designada de NCA_b). Atendendo à reduzida amostra de viagens em certos modos, dificultando a convergência dos modelos MNL, apenas se centrou

a análise na base de dados de viagens realizadas através dos modos mais relevantes Bus, Walk e Auto (BWA), num total de 56 284 registos. Na normalização necessária aos modelos consideraram-se as variáveis que corresponderam à maior utilização do automóvel (X3 e Y13).

Os resultados da primeira análise com o modelo incluindo as variáveis do quadro VI.34 podem ser comparados com os resultados do mesmo modelo mas sem variáveis de combinação no quadro VI.36. É possível constatar que o modelo com combinação de binárias apresenta um ajustamento global melhor e estatisticamente significativo, com parâmetros similares para as variáveis da taxa de motorização familiar por adulto e para as restantes variáveis comuns.

Quadro VI. 36 - Modelos de combinações de variáveis socioeconómicos MNL99.3a_BWA_NCA_b

Variável	MNL7.57a_BWA_NCA_b		MNL99.3a_BWA_NCA_b		Obs.
	B	W	-3,873	2,692	
ASC	-2,783	3,348			- Sinais coerentes
Lc	-1,776	-1,668	3,252	2,864	- Sinal coerente
NA0	3,229	2,864	1,909	1,413	- Parâmetros estáveis
NA1	1,886	1,393	1,131	0,694	
NA2	1,003	0,684	-	-	
NA3	-	-	-0,506	-0,075***	
ln1	-0,698	-0,187**	-0,162	0,066***	- Parâmetros estáveis.
ln2	-0,291	0,001***	0,337	0,178	
ln3	0,274	0,167**	-	-	
ln4	-	-	-	-	
ld1	-	-	-	-	- Sinal e comportamento não linear coerente
ld2	1,967	1,699	-	-	
ld3	1,193	1,076	-	-	
ld4	1,331	1,386	-	-	
ld5	1,437	1,333	-	-	
Sx	0,371	0,184	-0,453	-0,087***	- Sinal coerente
Mot	-0,435	-0,063***	-0,015	-2,676	- Parâmetros estáveis.
D2	-0,014	-2,676	-0,228	-0,183	- Parâmetros estáveis.
nV	-0,231	-0,186	0,689	0,253	- Parâmetros estáveis.
Tr	0,652	0,234	0,471	0,315	- Parâmetros estáveis.
Rg	0,449	0,304	0,970	0,362	
Es	0,928	0,332	-	-	
Lz	-	-	0,157	-0,004***	
CS	0,154	-0,007***	2,178	1,280	
X1	-	-	-0,059***	-0,064***	
X2	-	-	-	-	- Análise de resultados difícil
X3	-	-	0,822	0,639	
X4	-	-	2,019	1,242	
X5	-	-	0,987	0,319	
X6	-	-	1,209	0,657	
X7	-	-	1,434	0,316***	
X8	-	-	1,067	0,691	
X9	-	-	3,071	2,407	
X10	-	-	2,374	1,853	
X11	-	-	2,680	2,516	
X12	-	-	3,512	2,745	
X13	-	-	1,073	0,571	
X14	-	-	2,978	2,231	
X15	-	-	2,461	1,736	
X16	-	-	2,660	2,214	
X17	-	-	2,180	1,641	
X18	-	-	-3,873	2,692	
Amostra	56284		56284		56284
p-R ²	0,530		0,536		0,536
LL	-23417,5		-23110,7		-23110,7
Param.	40		62		62
GL			22		
-2LL			613,6		
Chi ² (α=0,05)			33,9		- Hipótese nula dos modelos serem iguais pode ser recusada

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Os parâmetros das variáveis binárias, no entanto, têm interpretação menos imediata, uma vez que se referem a uma combinação de várias variáveis. No sentido de avaliar a coerência dos resultados realizaram-se gráficos com as repartições modais verificadas na base de dados utilizada para cada situação e gráficos com os valores dos seus parâmetros no modelo, facilitando a sua comparação e possibilitando a avaliação da coerência das suas relações (figuras VI.7 e VI.8).

É possível verificar que os parâmetros aparentemente seguem tendências similares às repartições modais verificadas para cada combinação definida por cada variável. Os parâmetros com comportamentos mais díspares estão relacionados com o grupo etário entre os 10 e 19 anos e +65 anos para o modo Walk, provavelmente devida à grande influência das restantes variáveis utilizadas na modelação, nomeadamente a distância entre a zona de GA das viagens.

Os resultados da segunda análise com o modelo incluindo as variáveis do quadro VI.35 podem ser comparados com os resultados do mesmo modelo mas sem variáveis binárias de combinação no quadro VI.37. É possível constatar que o modelo com combinação de binárias apresenta um ajustamento global melhor e estatisticamente significativo, com parâmetros similares para as variáveis comuns aos dois modelos.

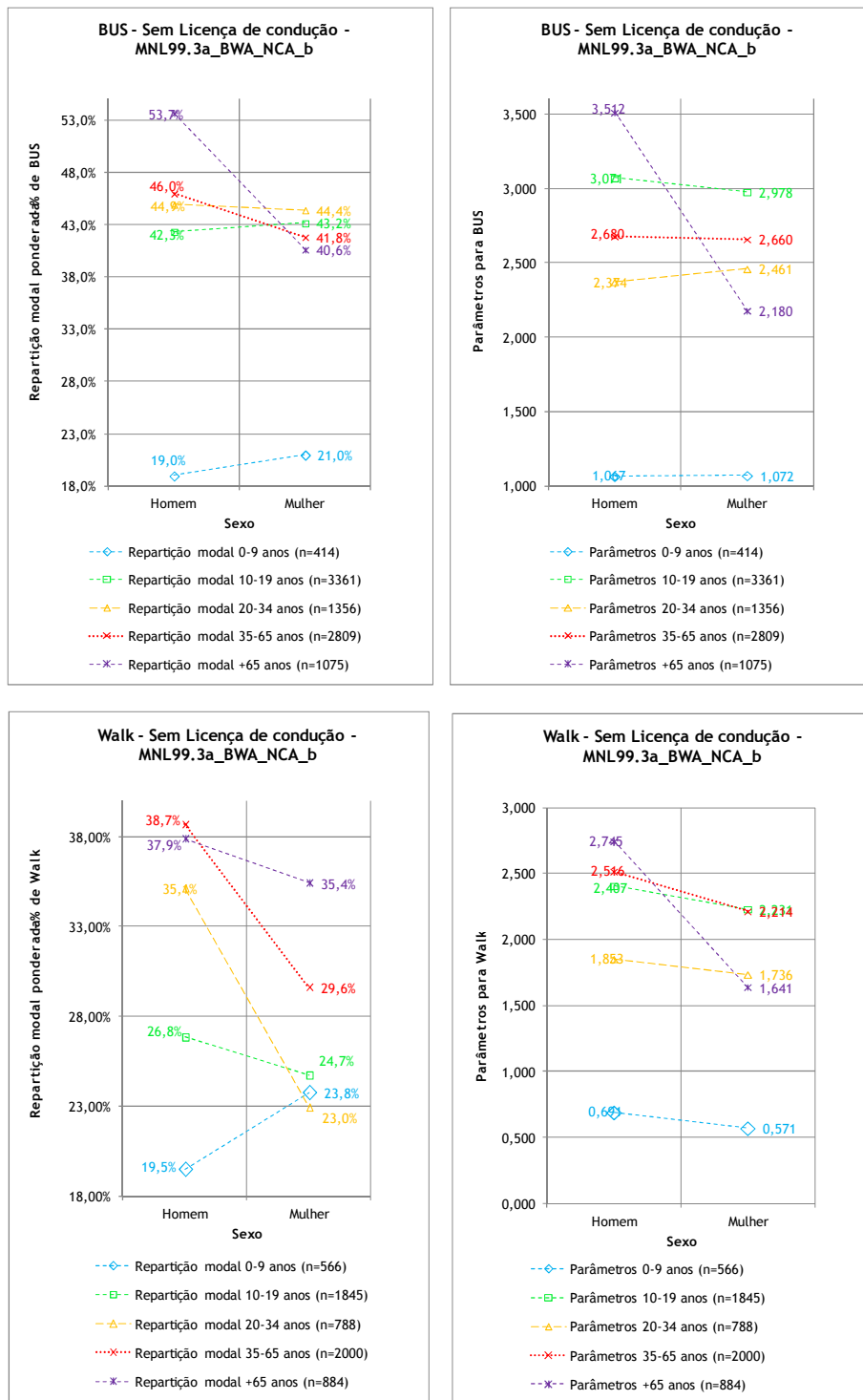


Figura VI. 7 - Gráficos com resultados do modelo MNL99.3a_BWA_NCA_b e repartições modais correspondentes dos modos Bus e Walk para indivíduos sem licença de condução

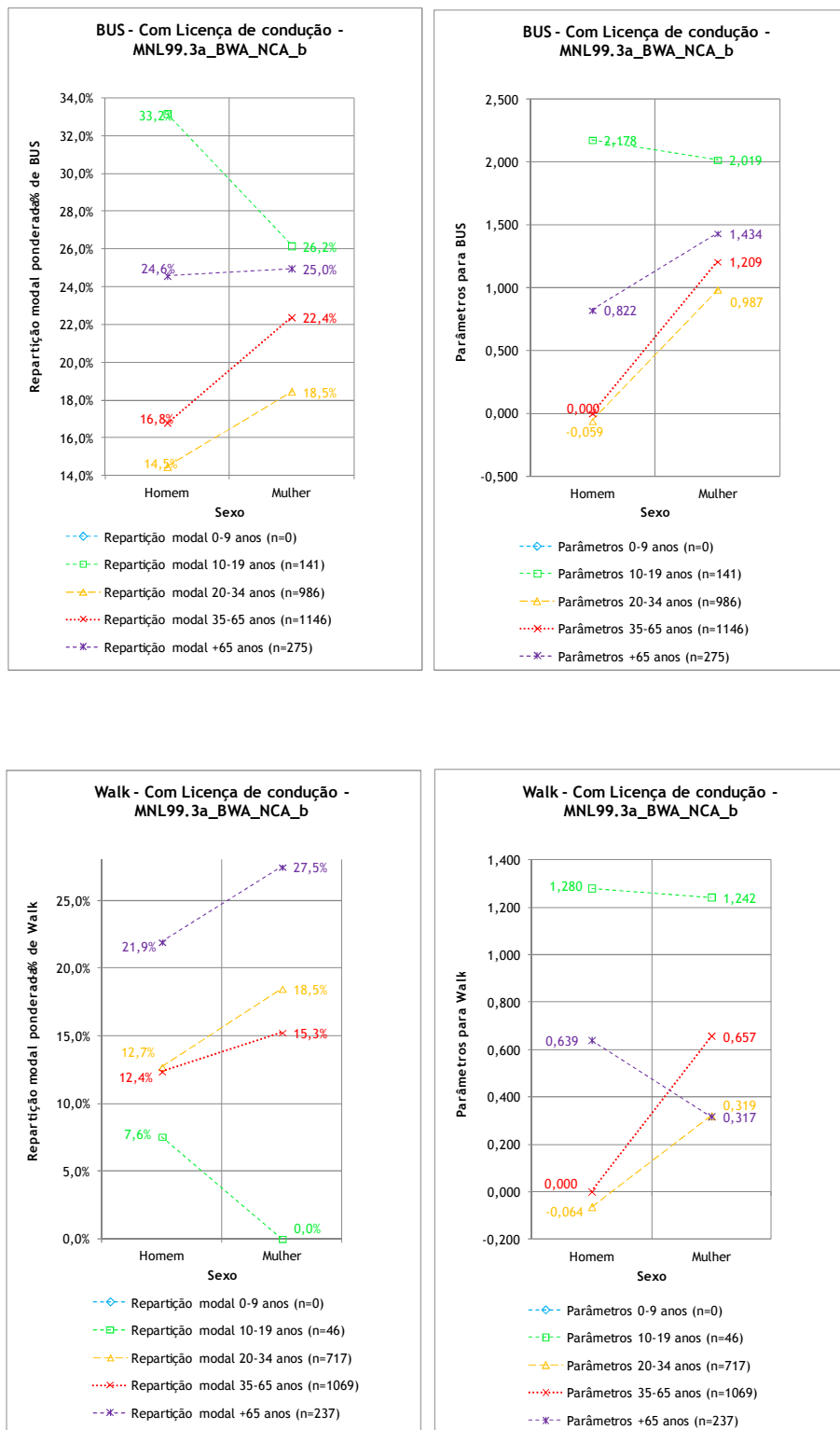


Figura VI. 8 - Gráficos com resultados do modelo MNL99.3a_BWA_NCA_b e repartições modais correspondentes dos modos Bus e Walk para indivíduos com licença de condução

Quadro VI. 37 - Modelos de combinações de variáveis socioeconômicos MNL99.9b_BWA_NCA_b

Variável	MNL7.57a_BWA_NCA_b		MNL99.9b_BWA_NCA_b		Obs.
	B	W	B	W	
ASC	-2,783	3,348	-3,715	3,189	- Sinais coerentes
Lc	-1,776	-1,668			
NA0	3,229	2,864			
NA1	1,886	1,393			
NA2	1,003	0,684			
NA3	-	-			
ln1	-0,698	-0,187**	-0,581	-0,101***	- Parâmetros consistentes.
ln2	-0,291	0,001***	-0,196	0,089**	
ln3	0,274	0,167**	0,268	0,130**	
ln4	-	-	-	-	
ld1	-	-			
ld2	1,967	1,699			
ld3	1,193	1,076			
ld4	1,331	1,386			
ld5	1,437	1,333			
Sx	0,371	0,184			
Mot	-0,435	-0,063***	-0,401	-0,068***	- Parâmetros estáveis.
D2	-0,014	-2,676	-0,016	-2,673	- Parâmetros estáveis.
nV	-0,231	-0,186	-0,225	-0,184	- Parâmetros estáveis.
Tr	0,652	0,234	0,630	0,237	- Parâmetros estáveis.
Rg	0,449	0,304	0,466	0,323	
Es	0,928	0,332	1,061	0,392	
Lz	-	-	-	-	
CS	0,154	-0,007***	0,126*	-0,006***	
Y10			3,700	2,647	- Análise de resultados difícil
Y11			1,792	1,005	
Y12			0,459	-0,074***	
Y13			-	-	
W20			4,273	2,857	
W21			2,885	1,317	
W22			2,282	0,852	
W23			0,205***	-0,230**	
W30			4,538	3,074	
W31			3,003	1,665	
W32			2,034	0,885	
W33			-0,269***	-0,009***	
W40			4,336	3,579	
W41			3,542	0,838	
W42			3,435	1,316	
W43			2,008	1,034	
W50			5,596	4,284	
W51			4,489	3,006	
W52			3,829	2,396	
W53			3,146	1,759	
Amostra	56284		56284		
p-R ²	0,530		0,537		- Modelo com variáveis binárias ligeiramente melhor mas praticamente igual ao modelo 99.3a_BWA_NCA_b
LL	-23417,5		-23054,5		
Param.	40		60		
GL			20		
-2LL			726		
Chi ² ($\alpha=0,05$)			31,4		- Hipótese nula dos modelos serem iguais pode ser recusada

* - p-valor >0,05 & <=0,1

** - p-valor >0,1 & <=0,25

*** - p-valor >0,25

Para a interpretação dos resultados das variáveis binárias de combinação, e de modo a avaliar a coerência das suas relações, os parâmetros para cada taxa de motorização familiar por adulto são apresentados graficamente na figura VI.9.

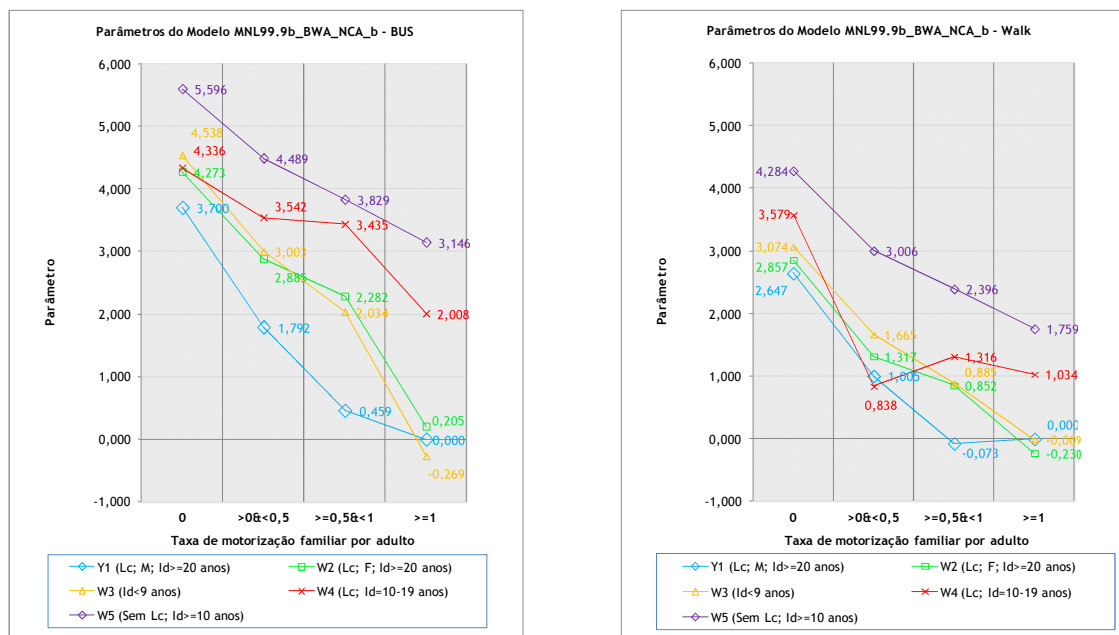


Figura VI. 9 - Gráficos dos parâmetros do modelo MNL99.9a_BWA_NCA_b com especificação binária de combinações de variáveis socioeconômicas

É possível constatar algumas tendências distintas para os grupos referidos, nomeadamente entre homens e mulheres adultos nas mesmas condições. Os homens apresentam uma depreciação das alternativas (nomeadamente do Bus) mais acentuada que as mulheres, mesmo para taxas de motorização baixas. Salienta-se, no entanto, a semelhança de comportamentos entre homens e mulheres encartados quando a taxa de motorização é superior a 1 veíc./adulto. Também é possível verificar que a inexistência de licença de condução ou a eventual inexperiência na condução dos jovens dos 18 a 19 anos com licença de condução levam a uma maior valorização das alternativas mesmo quando há mais de um automóvel disponível por adulto. Constata-se, ainda, uma tendência mais diferenciada para o Bus que para o Walk. Este último modo apresenta parâmetros com grandezas que sugerem uma menor valorização do mesmo relativamente ao Bus, e tendências muito semelhantes para quase todos os grupos exceto o W5 (sem licença de condução e com mais de 10 anos). Para o modo Bus é também possível verificar três conjuntos de tendências similares. O grupo dos inquiridos sem licença de condução e com idade superior a 10 anos e os jovens adolescentes com licença de condução (W4), com uma clara valorização do modo Bus. Outro grupo com tendências semelhantes inclui as mulheres com mais de 20 anos com licença de condução (W2) e as crianças com menos de 10 anos (W3), indiciando que estes dois grupos estão fortemente correlacionados e que estas crianças provavelmente estarão condicionadas em termos de transporte às escolhas das mães. O último grupo distinto é o dos homens com mais de 20 anos e com licença de condução (Y1).

Em geral é possível verificar que, com exceção dos homens com licença de condução e com mais de 20 anos, as tendências são aproximadamente lineares com um decréscimo mais

acentuado da utilidade na passagem de não ter veículos no agregado para a situação de ter pelo menos um veículo.

Estes resultados sugerem a possibilidade de substituir as variáveis socioeconômicas Lc, Sx e Id, por variáveis binárias de combinações entre as mesmas. O modelo com estas variáveis apresentou melhorias no ajustamento global com significância estatística e sem grandes alterações nos parâmetros das restantes variáveis. Este resultado vem sustentar a hipótese acima formulada de que esta especificação poderá introduzir mais alguma informação relevante no modelo, nomeadamente informação relacionada com a diferença de comportamento das mulheres em diferentes idades e entre homens e mulheres para baixas taxas de motorização e que aparentemente não consegue ser capturada pelas variáveis binárias que representam de forma independente diferentes características.

A agregação da variável NCA às variáveis anteriores melhorou apenas ligeiramente o ajustamento global, mas de forma estatisticamente significativa (ver quadro VI.38), podendo-se recusar a hipótese nula de ambos os modelos serem iguais. Os resultados sugerem a distinção de três grupos sociais, caracterizados pelas variáveis socioeconômicas analisados, cujos comportamentos seguem tendências similares.

Quadro VI. 38 - Comparação dos ajustamentos dos modelos com diferentes especificações para as variáveis binárias de combinação de variáveis socioeconômicas

Variável	MNL99.9b_BWA_NCA_b	MNL99.3a_BWA_NCA_b	Obs.
Amostra	56284	56284	
p-R ²	0,537	0,536	- Modelo com ajustamento global ligeiramente melhor
LL	-23054,5	-23110,7	
Param.	60	62	
GL		2	
-2LL		-112,4	- Hipótese nula dos modelos serem iguais pode ser recusada
Chi ² ($\alpha=0,05$)		6,0	

Apesar de ser claramente discutível o acréscimo de complexidade e de particularidade que a utilização destas variáveis binárias impõe, os resultados sugerem a existência de uma efetiva diferenciação de comportamentos, que poderá justificar a devida ponderação da sua utilização em modelos mais complexos ou detalhados.

6. Comparação dos Resultados com os de Estudos Internacionais

Para avaliar a coerência dos resultados alcançados nas análises exploratórias, compararam-se os resultados obtidos com os verificados nos estudos internacionais referidos no Capítulo III. Apesar das diferenças de especificação, de unidades utilizadas e de realidades socioculturais distintas, esta comparação deverá permitir balizar os valores obtidos e avaliar a sua coerência, nomeadamente ao nível dos sinais.

A partir do quadro VI.39 é possível constatar que a maioria dos estudos internacionais utilizou modelos Logit Multinomiais. Apesar da especificação dos parâmetros para os modelos Logit Mistos ser distinta das utilizadas nos modelos Logit Multinomiais (apresentam o valor médio considerado para a população em estudo mas também o seu desvio padrão atendendo à distribuição assumida para a variável) consideraram-se os valores médios de ambos para comparação. Os estudos analisados apresentam amostras entre 174 e 7770 inquiridos. Os modos de transportes considerados variaram consoante o tipo de viagem a modelar. Das variáveis referidas nos estudos, apenas se apresentam as mais comuns e as similares às consideradas no estudo:

- Autocarro (Bus);
- Comboio (Train);
- Condutor (Drive alone);
- Passageiro (Shared Drive);
- Peão (Walk).

É possível constatar que a maioria das variáveis foi considerada com especificação contínua ou como binária isolada. Algumas das variáveis entraram nos modelos tanto como genéricas (sendo constante para todos os modos) como específicas apenas para alguns modos. Os parâmetros das variáveis cuja formulação era contrária à utilizada no presente estudo (por exemplo, o sexo do inquirido ser 1 se fosse masculino e 0 se fosse feminino), foram ajustadas com sinal contrário.

Nas variáveis relacionadas com custos salienta-se que para os parâmetros apresentados se deve atender à sua definição, que podem resultar de dados divididos por 1 000, em cêntimos ou em divisas distintas, enquanto na base de dados do presente estudo os valores foram considerados em Euros.

Não foi possível encontrar estudos com a inclusão das seguintes variáveis:

- Nível de instrução;
- Número total de viagens diárias;
- Duração da estadia na atividade de destino;
- Oferta de estacionamento;

- Razão entre tempos de viagem;
- Disponibilidade diária de Moto;

No quadro VI.40 apresenta-se um resumo das gamas de valores verificados para os três modos mais relevantes, que são comparados com os resultados do presente estudo no quadro VI.41. Analisando os sinais e grandezas dos valores é possível constatar:

- *Rendimento*: O sinal negativo das variáveis contínuas revela que um aumento do rendimento diminui a utilidade do modo. É possível constatar que os parâmetros são coerentes para as alternativas ao automóvel, sendo sempre negativas, e que para o condutor a gama de valores acaba num valor positivo, indiciador que um aumento do rendimento aumenta a utilidade destes últimos. Salienta-se novamente a grande variedade de especificações verificadas e o facto de estar fortemente dependente da divisa analisada. Na especificação binária, os estudos sugerem uma influência dos rendimentos anuais variada atendendo ao modo em análise, apresentando-se muito particulares e de difícil comparação;
- *Sexo*: A maioria dos parâmetros das alternativas apresenta valores positivos indiciando que as mesmas são mais valorizadas/utilizadas pelas mulheres. O estudo que revelou um parâmetro negativo incorporou a variável como genérica, constante para todas as alternativas, e visava viagens para o trabalho;
- *Idade*: O sinal negativo indica que os inquiridos mais velhos consideram as alternativas menos atrativas que os mais novos. Este resultado apresenta-se pouco credível e resulta da prática comum nos estudos analisados de utilizar a especificação contínua ou discreta para esta variável, consideração que aparentemente carece de melhor fundamentação. O senso comum, corroborado pelos resultados obtidos no capítulo anterior, mostra que a especificação contínua desta variável é demasiado simplista, uma vez que a sua real influência claramente não é possível de definir numa função linear, atendendo à diferenciação de comportamentos verificada para distintas idades;

Quadro VI. 39 - Quadro com valores dos parâmetros obtidos de modelos discretos em estudo internacionais

Variável	Tipo	Unidade	b					Modelo	Amostra	Referência
			Bus	Train	Drive-Alone	Shared-ride	Walk			
Rendimento	Contínua	/\$CA1000	-	-1,172	-	-	-	MNL	2000	Bhat (1998a)
			-0,0090	-	-	-	-		1617	Bhat (2000)
		/\$US1000	-	-	-	-0,0022	-		5029	Koppelman&Bhat(2006)
			-	-	-	0,0004	-			
			-0,0053	-	-	-	-			
Rendimento anual individual	Discreta (valor maximo da gama)	/\$AU1000	-0,0078	-	-	-	-0,0097		1840	Hensher&Rose(2007)
Rendimento anual	Binária	60000 yuan	-1,9800	-	6,8700	-	-	MNL	986	Bao et al. (2010)
Baixo rendimento do agregado (<€45000/ano)	Binária	-	0,3990				-	MNL com paramentros não lineares	7770	Pozsgay &Bhat(2002)
Rendimento do agregado	Contínua	/\$US1000	-0,0114	-	-	-	-	MNL	3157	Koppelman&Bhat(2006)
			-	-	-	-0,0019	-			
			-	-	-	0,0011	-			
		\$US	-	-	-	-	-0,0075			
			-	-	0,00000454	-	-			
Rendimento familiar <7500\$			-	-	-0,0000572	-	-		771	Train (2003)
Rendimento familiar-7500\$ (>0\$ e <3000\$)			-	-	-0,0000543	-	-			
Rendimento familiar-7500\$ (>0\$ e <5000\$)			-	-	-	-	-			
Custo da Viagem	Contínua	Constante	\$CA	-3,1650			-	MNL	2000	Bhat (1998a)
			\$US cents	-0,0031			-		3000	Bhat (1998b)
			\$US	-0,4300			-		1617	Bhat (2000)
			€	-0,1320			-		1067	Cascetta&Papola(2003)
			\$UScent	-0,0038			-		3157	Koppelman&Bhat(2006)
			\$UScent	-0,0513			-		5029	
		\$AU	-0,1813	-	-	-	-		1840	Hensher&Rose(2007)
			-	-	-0,0992	-	-			
			-0,3056	-	-	-	-			
			-	-	-0,2064	-	-		1941	
		\$US	-0,0023			-	-		174	Munizaga et al. (2008)
		¥RMB	-2,0423 (0,0073N)			-2,0423 (0,0073N)	-	ML	664	Li et al (2010)
		yuan/km ou yuan/tempo	-9,28	-	-5,2700	-	-	MNL	986	Bao et al. (2010)
Custo da Viagem/Rendimento líquido		\$cent/min.	-0,0284	-	-0,0284	-0,0284	-		771	Train (2003)
Custo da viagem/salario		€	-0,19 (0,11LN)			-0,19 (0,11LN)	-	ML	1290	Yáñez et al. (2010)
Custo do estacionamento	Contínua	\$AU	-	-	-0,023	-	-	MNL	1840	Hensher&Rose(2007)
			-	-	-0,123	-	-	MNL	1941	
Sexo	Binária (0-M; 1-F)	-	-	1,281	-	-	-	MNL	2000	Bhat (1998a)
			-	-	-	0,620	-	ML	3000	Bhat (1998b)
			0,447	-	-	-	-	MNL	1941	Hensher&Rose(2007)
			-1,767			-	-	MNL	1840	
	Binária (1-M; 2-F)	-	0,09 (0,0072N)	-	-	-	0,09 (0,0072N)	ML	664	Li et al (2010)
			-0,97	-	3,250	-	-	MNL	986	Bao et al. (2010)
Idade	Contínua	-	-1,753				-	MNL com paramentros não lineares	7770	Pozsgay &Bhat(2002)
	Discreta (média da gama)	anos	-0,0152	-	-	-	-	MNL	1941	Hensher&Rose(2007)
Viagem sozinho			-	0,366	-	-	-	MNL	2000	Bhat (1998a)
Transporte de crianças	Binária	-	-	-	-	1,502	-	ML	3000	Bhat (1998b)
Dependentes (5-21 anos) no agregado			-0,689				-	MNL com paramentros não lineares	7770	Pozsgay &Bhat(2002)
Pessoas do agregado	Contínua	hab	0,129	-	-	-	-	MNL	3157	Koppelman&Bhat(2006)
			-	-	-	0,164	-			
			-	-	-	0,398	-			
			-	-	-	-	0,185			

Quadro VI. 39 - Quadro com valores dos parâmetros obtidos de modelos discretos em estudo internacionais (Continuação)

Variável	Tipo	Unidade	b					Modelo	Amostra	Referência	
			Bus	Train	Drive-Alone	Shared-ride	Walk				
Tempo de viagem no veículo	Constante	min.	-4,610	-	-	-	-	MNL	2000	Bhat (1998a)	
			-0,034	-	-	-	-		3157	Koppelman&Bhat(2006)	
			-0,055	-	-	-	-		1840	Hensher&Rose(2007)	
			-	-	-0,309	-	-		1840		
			-0,035	-	-	-	-		1941		
	Tempo dentro do veículo	Continua	min.	-	-	-0,035	-		-	ML	771
-	-			-0,064	-0,064	-					
Tempo a pé			-	-	-	-0,0689					
Tempo de viagem equivalente dentro do veículo			-0,040	-	-	-0,040			1290	Yáñez et al. (2010)	
LOG (Tempo de viagem equivalente dentro do veículo)			0,140	-	-	-	MNL com paramentros não lineares	7770	Pozsgay &Bhat(2002)		
			-1,965	-	-	-					
Tempo de viagem fora do veículo	Constante	min.		-3,453					2000	Bhat (1998a)	
Tempo de espera para Transbordo (Bus+Walk,Bus+Auto)	Continua		-0,0538	-	-	-			771	Train (2003)	
Tempo de viagem fora do veículo/distância					-0,236			MNL	3000	Bhat (1998b)	
Tempo total fora do veículo/distância	Continua	min/milha			-0,313				1617	Bhat (2000)	
				-0,239				3157	Koppelman&Bhat(2006)		
Tempo de viagem fora do veículo motorizado/distancia							-0,032			3000	Bhat (1998b)
		min			-0,047				1617	Bhat (2000)	
Tempo total de viagem	Continua	h	-	-	-1,221	-	-	MNL	1067	Cascetta&Papola(2003)	
			-0,241		-	-	-				
					-0,005					5029	Koppelman&Bhat(2006)
		min	-	-	-	-	-0,086			3157	Munizaga et al. (2008)
					-0,085					174	
			-0,070			-0,070				1290	
			-7,880	-	9,110	-	-	MNL	986	Bao et al. (2010)	
	Tempo de acesso	Continua	min	-0,06017	-	-	-	-	MNL	1840	Hensher&Rose(2007)
	-0,03040			-	-	-	-	1941			
Tempo de espera	Continua	min	-0,09226	-	-	-	-	MNL	1840	Hensher&Rose(2007)	
			-0,04628	-	-	-	-		1941		
			-0,09600	-	-	-0,096	-				1290
Automóveis/adultos do agregado			-1,1158	-	-	-	-		3000	Bhat (1998b)	
Automóveis/condutor			-	-	5,000	-	-	MNL			
			2,330	-	-	-	-		771	Train (2003)	
			-	-	-	2,380	-				
Número de Veículos per capita no agregado		veic./hab		2,988			-	MNL com paramentros não lineares	7770	Pozsgay &Bhat(2002)	
Automóveis/Trabalhador			-	-	0,859	-	-		1617	Bhat (2000)	
Veiculos do agregado per capita	Continua		-3,4400	-	-	-	-	MNL	3157	Koppelman&Bhat(2006)	
			-	-	-	-0,7150	-				
			-	-	-	-2,8400	-				
		-	-	-	-0,8970	-					
Número de condutores no agregado	Cond.	-	-	1,020	-	-		771	Train (2003)		
		0,990	-	-	-	-					
		-	-	-	0,872	-					
Número de automóveis		veic.	2,25	-	-	2,25	ML	1290	Yáñez et al. (2010)		
Agregado sem veiculos (Não tem Carro=1; Tem Carro =0)			2,4800	-	-	-	2,4800		3157	Koppelman&Bhat(2006)	
Disponibilidade diária de carro (Não tem Carro=1; Tem Carro =0)	Binária	-	-	-	-	-	5,279	MNL	9300	Müller et al. (2008)	
			0,553	-	-	-	-				
Licença de Condução	Binária	1 - Tem; 2 - Não tem	-0,46	-	1,12	-	-		986	Bao et al. (2010)	
Destino no CBD	Binária	-	1,9300	-	-	-	-	MNL	3157	Koppelman&Bhat(2006)	
			-	-	-	0,4090	-				
			-	-	-	-0,1250	-				
			-	-	-	-	0,1360				
			-	5,11	-	-	-				
OD Big City			-	-	-2,549	-	-		1067	Cascetta&Papola(2003)	
Distância	Continua	km	-	-	-	-	-4,376	MNL	9300	Müller et al. (2008)	
			-0,052	-	-	-	-				
			-10,381 (0,203LN)	-	-	-	-10,381 (0,203LN)		ML	664	Li et al (2010)
			0,33	-	2,08	-	-		MNL	986	Bao et al. (2010)
Frequência	Constante	Partidas/dia	-	-2,435	-	-	-		2000	Bhat (1998a)	
Regularidade	Continua	min.	-0,0318	-	-	-	-	MNL	771	Train (2003)	
Frequencia diária	Continua	veic/h	0,116	-	-	-	-		1067	Cascetta&Papola(2003)	
Sozinho no lazer	Binária	-			-0,391		-	MNL com paramentros nao lineares	7770	Pozsgay &Bhat(2002)	
Trabalhador	Binária	-			-0,0727		-	MNL com paramentros nao lineares	7770	Pozsgay &Bhat(2002)	
Viagem sistematica (Emprego, Escola)	Binária	-	-	-	-2,015	-	-	MNL	1067	Cascetta&Papola(2003)	
Viagem para o trabalho	Binária	-	0,0913 (0,0189N)	-	-	-	0,0913 (0,0189N)	ML	664	Li et al (2010)	
Ligação directa	Binária	-	1,224	-	-	-	-	MNL	1067	Cascetta&Papola(2003)	
Número de transbordos	Continua	-	-0,105	-	-	-	-	MNL	771	Train (2003)	
			-1,14	-	-	-1,14	-	ML	1290	Yáñez et al. (2010)	

Quadro VI. 40 - Quadro resumo das gamas de valores dos parâmetros analisados obtidos de modelos discretos em estudo internacionais

Variável	Especificação	Unidade	Gamas de valores de b		
			Bus	Drive-Alone	Walk
Rendimento	Continua	/1000	-0,0114; -0,0053	-0,0543; 0,00454	-0,0097; -0,0075
	Binária	< rendimento anual = XXX	-1,98; 0,3990	0,399; 6,87	-
Sexo	Binária	M=0; F=1	-1,767; 0,447	-1,767; 3,25	0,09
Idade	Continua	-	-1,753	-1,753	-
	Discreta	Média da gama	-0,0152	-	-
Tempo total de viagem	Continua	min	-7,88; -0,005	-0,085; 9,11	-0,086; -0,005
	Continua	h	-0,241	-1,221	-
Veículos do agregado per capita	Continua	veic./hab.	-3,440; 2,988	0,859; 5,00	-0,897
Licença de condução	Binária	Tem = 1; N tem = 2	-0,46	1,12	-
Disponibilidade diária de automóvel	Binária	Tem =0; N Tem =1	0,553; 2,480	-	2,480; 5,279
Distância	Continua	km	-0,052; 0,33	2,08	-4,376
Frequência diária	Continua	veic./h	0,116	-	-
Viagem pendulares (Emprego, escola)	Binária	N = 0; S = 1	-	-2,015	-

- *Tempo total de viagem*: O sinal negativo dos parâmetros indica que todas as alternativas ao automóvel perdem atratividade para viagens mais longas. Apenas o modo Auto tem uma valorização positiva num dos estudos;
- *Veículos do agregado per capita, ou taxa de motorização familiar*: O sinal negativo para as alternativas ao automóvel indicia que um aumento da possibilidade de utilizar o automóvel diminui a atratividade das alternativas. O valor positivo para este parâmetro provém de um estudo do modo de transporte para zonas de recreio, considerando a variável genérica para todos os modos motorizados.
- *Licença/Carta de Condução*: O sinal negativo para o modo Bus e positivo para o Condutor revelam-se contrários ao esperado. No estudo analisado possuir licença faz decrescer o valor da variável independente de 2 para 1, pelo que os sinais deveriam ser inversos. Estes resultados poderão dever-se a erros de transcrição para o artigo.
- *Disponibilidade diária de automóvel*: O valor positivo dos parâmetros indica, como seria espectável, que a indisponibilidade diária de automóvel aumenta a atratividade das alternativas;
- *Distância*: O sinal negativo dos parâmetros para o modo pedonal indica que o aumento da distância da viagem o torna substancialmente menos útil. Para o automóvel o parâmetro positivo revela que o aumento das distâncias aumenta a sua utilidade e para o modo Bus essa influência não é bem definida, mas evidentemente menos relevante;
- *Frequência diária*: O sinal positivo do parâmetro para os TC indica que tal como esperado um aumento da sua frequência diária aumenta a sua utilidade;
- *Viagens pendulares (Emprego, Escola)*: O sinal negativo do parâmetro para o automóvel indica que os transportes públicos são mais úteis para estas viagens mais regulares.

No quadro VI.41 apresenta-se o resumo dos parâmetros verificados nos estudos internacionais e os verificados nos modelos selecionados para as alternativas mais relevantes (Bus e Walk), para comparação e análise dos resultados.

Quadro VI. 41 - Quadro comparativo das gamas de valores dos parâmetros analisados obtidos de modelos discretos em estudo internacionais e nos modelos sugeridos no caso de estudo

Variável	Especificação	Gamas de valores Internacionais		MNL 7.57		MNL 12.16BWA	
		Bus	Walk	BUS	Walk	BUS	Walk
Rendimento	Contínua	-0,0114;-0,0053	-0,0097;-0,0075	-	-	-	-
	Binária	-1,98;0,3990	-	-	-	-	-
Sexo	Binária	-1,767;0,447	0,09	0,383	0,145	0,474	0,378
	Contínua	-1,753	-	-	-	-	-
Idade	Discreta	-0,0152	-	-	-	-	-
	Binárias	-	-	1,703;2,111	1,381;1,698	1,942;2,358	1,592;1,632
Tempo total de viagem	Contínua	-7,88;-0,005	-0,086;-0,005	-	-	-	-
	Contínua	-0,241	-	-	-	-	-
Veículos do agregado per capita	Contínua	-3,440;2,988	-0,897	-	-	-	-
	Binárias	-	-	0,627;1,725	-0,226;0,737	0,820;1,949	0,039;0,801
Licença de condução	Binária	-0,46	-	-1,935	-1,81	-1,748	-1,528
	Binárias	0,553;2,480	2,480;5,279	3,401	2,459	4,101	2,996
Disponibilidade diária de automóvel	Contínua	-0,052; 0,33	-4,376	0,001	-2,686	0,005	-2,444
	Contínua	0,116	-	-	-	0,007	0,01
Frequência diária	Contínua	-	-	-	-	-	-
	Binária	-	-	0,598;0,897	0,178;0,307	0,342;0,426	0,212;0,192

Do quadro VI.41 é assim possível verificar que:

- O parâmetro para a influência do sexo no modo Bus está próximo da gama de valores verificados nos estudos internacionais;
- Os parâmetros para as variáveis binárias sobre o escalão etário não podem ser comparados uma vez que as especificações que foram utilizadas foram distintas das utilizadas no presente estudo. É de referir que a utilização de variáveis contínuas/discretas sugere uma diminuição da atratividade do modo Bus com o aumento da idade. Este comportamento, com a eventual redução da tarifa para seniores, não parece ser credível, confirmado com os resultados da especificação com combinação de binárias;
- Também para os parâmetros sobre a taxa de motorização familiar não é possível uma comparação direta dos resultados devido à diferença de especificação. No entanto, o sinal negativo mais comum para as alternativas ao automóvel e o valor positivo para os condutores são tendências coerentes e corroboradas pelos resultados da especificação com combinação de binárias;
- Em relação à licença de condução, apesar da incoerência anteriormente referida, verifica-se que as grandezas dos parâmetros do presente estudo são significativamente superiores aos do estudo internacional;
- Para a disponibilidade diária de automóvel, verifica-se que para o modo Bus os parâmetros do estudo se apresentam bastante superiores aos internacionais, enquanto os para o modo Walk se apresentam próximos das gamas internacionais;
- Em relação à distância é possível verificar que para o modo Bus os parâmetros do estudo são compatíveis com os internacionais, enquanto os parâmetros para o modo Walk são menores em grandeza absoluta aos internacionais, mas coerentes em sinal e grandeza relativa à alternativa Bus;
- Para a frequência média diária, provavelmente devido a diferentes especificações na sua determinação, o parâmetro para o modo Bus é bastante mais reduzido que o verificado internacionalmente;
- Em relação ao motivo da viagem, mais concretamente para as viagens pendulares, não é possível avaliar os parâmetros, uma vez que nos estudos internacionais estes apenas foram determinados para as alternativas ao automóvel, modo de referência no presente estudo e, portanto, normalizado.

Salienta-se ainda que o estudo permitiu verificar que as variáveis nível de instrução, número total de viagens diárias, duração da estadia na atividade de destino, oferta de estacionamento, razão entre tempos de viagem e disponibilidade diária de Moto, se mostram significativas, com tendências coerentes e uma considerável capacidade explicativa

combinada, algo que não tinha sido identificado em nenhum dos estudos de referência analisados.

7. Conclusões

Na análise iterativa dos modelos com combinações das diferentes variáveis para os seis modos considerados verificou-se a possibilidade de obter ajustamentos globais bastante altos, na ordem dos $p-R^2=0,445$ a $0,460$. Estes pseudo- R^2 equivalem a R^2 das regressões na ordem dos $0,85$ a $0,88$. Atendendo à acentuada redução da amostra para a inclusão das variáveis sobre políticas de mobilidade, são sugeridos dois conjuntos de variáveis a utilizar, uma com variáveis sobre políticas de mobilidade e outra sem as mesmas. Esta abordagem permitiu o desenvolvimento de modelações especialmente detalhadas ao nível da análise da influência de um conjunto vasto de variáveis socioeconómicas.

O modelo sem variáveis sobre políticas de mobilidade (MNL7.57 com $p-R^2 = 0,445$) inclui as variáveis licença de condução, taxa de motorização familiar *per capita*, nível de instrução, grupo etário, sexo, disponibilidade diária de moto no agregado, distância mais curta entre os centróides de GA, número de viagens diárias e motivo da viagem. As variáveis mais valorizadas e consideradas fundamentais foram a licença de condução, a taxa de motorização familiar *per capita*, a distância mais curta e, pela sua estabilidade, o sexo. As restantes foram consideradas como complementares. Todas as variáveis apresentaram parâmetros com comportamentos e sinais coerentes ou pelo menos plausíveis. A variável que maior impacto apresenta é a indisponibilidade de automóvel no agregado, seguida da licença de condução. Apesar de manifestar um elevado impacto para certas idades nas alternativas ao automóvel, o escalão etário apresenta uma grande variedade de comportamentos. A variável sexo apresentou valores relativamente estáveis ao longo das iterações. Em relação à variável distância mais curta, os resultados sugerem a introdução de uma variável binária caracterizadora das viagens com distância inferior a 1 km apenas para a alternativa Walk. Este procedimento não foi aplicado nos modelos produzidos devido à impossibilidade de diferenciar as expressões das utilidades de cada alternativa atendendo à estrutura da base de dados, com informação apenas para o modo escolhido. A manutenção da combinação desta variável em modelos com mais variáveis, revelou problemas de convergência, provavelmente devido à inexistência de ocorrências suficientes para alguns modos (Outros Bus e Moto), dificultando a otimização do modelo para situações tão distintas. No entanto, na possibilidade de diferenciação das funções de utilidades dos diferentes modos, é recomendável a inclusão da variável binária DX1 na utilidade do modo Walk.

O modelo com variáveis sobre políticas de mobilidade com melhor ajustamento (MNL12.16 com $p-R^2 = 0,460$) inclui as variáveis licença de condução, taxa de motorização familiar *per capita*, distância mais curta entre os centróides de GA, nível de instrução, grupo etário, sexo, número de viagens diárias e motivo da viagem, bem como as variáveis relacionadas com a

oferta de estacionamento, a frequência média diária de Bus e a razão entre tempos de viagem por Bus e por Auto. Verificou-se uma acentuada redução da amostra disponível com dados sobre todas as variáveis incluídas, purgando a mesma das alternativas mais relevantes, optando-se por considerar apenas os três modos mais representativos Bus, Walk e Auto (BWA). Para o mesmo modelo MNL12.16BWA, esta limitação às alternativas mais relevantes, permitiu alcançar um ajustamento global de $p-R^2 = 0,522$. Constatou-se que as variáveis relacionadas com as políticas de mobilidade apresentam uma contribuição bastante moderada de informação relevante para o modelo, traduzida por um reduzido acréscimo de ajustamento com a sua inclusão. Esta constatação pode dever-se aos problemas com a fidelidade da informação recolhida, com a existência de erros ou por uma simples falta de relevância das mesmas nas escolhas modais dos cidadãos.

Todas as variáveis apresentaram parâmetros com comportamentos e sinais coerentes ou pelo menos plausíveis. Apenas nas variáveis sobre políticas de mobilidade e para o modo Walk se verificaram comportamentos algo contrários ao espetável. Este comportamento pode dever-se à falta de relação entre estas e o modo Walk. Surpreendentemente constata-se igualmente que a variável frequência média horária, segundo o modelo, pouca influência apresenta para a utilidade das alternativas. Em relação ao impacto das restantes diferentes variáveis, este segue as tendências verificadas para o modelo anterior.

Outra constatação relativamente estável ao longo das modelações foi o aparente comportamento antagónico do modo Outros Bus relativamente ao modo Bus. Estes resultados podem dever-se à sua difícil caracterização homogénea associada a amostras pequenas, ou devido à natureza operacional dos sub-modos considerados para a sua definição. Transportes escolares, excursões e serviços não regulares, têm características operacionais mais próximas do TI do que dos TC regulares, nomeadamente na oferta de serviços mais personalizados e geralmente em áreas periféricas, rurais ou em trajetos de longa duração, podendo justificar o seu comportamento mais próximo dos modos não coletivos.

A avaliação da validade dos modelos exploratórios obtidos foi realizada através da comparação das estimativas dos modelos com as repartições modais efetivamente verificadas nas bases de dados. Estas comparações foram realizadas com simulações e dados extrapolados. Foi possível constatar que apesar das estimativas dos modelos apresentarem algumas discrepâncias com os valores obtidos da base de dados, os erros são muito reduzidos. Foi também possível verificar que o uso do modelo com variáveis sobre políticas de mobilidade apresentava menores discrepâncias, corroborando a perceção de permitir a introdução de informação adicional relevante.

Apesar dos resultados promissores obtidos dos modelos combinados, postulou-se a hipótese da existência de eventuais tendências não detetadas devido à falta de capacidade explicativa dos mesmos, resultantes do princípio de independência de efeitos das várias variáveis

explicativas. Nesse sentido pesquisaram-se na base de dados eventuais influências de características socioeconômicas ainda não incorporadas nos modelos, nomeadamente o eventual efeito interligado da idade e do sexo, bem como do grau de acesso ao automóvel familiar, nas escolhas modais. Foi possível verificar algumas tendências relativamente homogêneas que se justifica serem testadas em modelos mais complexos, nomeadamente:

- As crianças até aos 9 anos aparecem condicionadas às opções modais dos pais e, particularmente, da mãe;
- Verifica-se uma certa homogeneidade de comportamentos para as mulheres com licença de condução e com mais de 20 anos e as crianças até aos 9 anos;
- Para os jovens até aos 20 anos aparentemente não há diferenciação de comportamento entre sexos;
- A falta de licença de condução parece levar a comportamentos muito semelhantes entre homens e mulheres em relação ao automóvel;
- Por outro lado, com licença de condução as mulheres utilizam menos o automóvel que os homens. Apenas com taxas de motorização familiares superiores a 1 veíc./adulto, se verifica uma aproximação do uso do automóvel entre homens e mulheres;
- Aparentemente, as mulheres utilizam bastante menos o veículo do agregado quando este é único, mesmo quando se ignora o efeito do motivo, traduzindo uma sociedade ainda algo “machista”, embora com tendência a esbater-se significativamente nos grupos etários mais jovens;
- Assim, as mulheres mais novas (20-34 anos) parecem mostrar tendência para utilizar mais o automóvel que as mulheres mais velhas (35-65 anos), aproximando-se do comportamento dos homens;
- Em termos de alternativa ao automóvel, a opção principal para as mulheres sem licença de condução parece ser o Bus enquanto para os homens nas mesmas condições a Moto ganha uma importância significativa;
- Parece verificar-se uma certa homogeneidade de comportamentos para os homens com licença de condução e com mais de 20 anos;
- Do mesmo modo transparece também uma certa homogeneidade de comportamentos para jovens dos 10 aos 19 anos com licença de condução e para homens e mulheres com mais de 10 anos e sem licença de condução;

A constatação destas tendências na análise qualitativa de possíveis interdependências entre variáveis levou à ponderação de as introduzir no modelo de escolha modal através de variáveis binárias para cada grupo. Esta análise revelou parâmetros relativamente coerentes e um ajustamento ligeiramente melhor comparativamente ao modelo com as variáveis

separadas. Em geral constatou-se que excetuando o comportamento dos homens com mais de 20 anos e com licença de condução, nos restantes grupos a influência do aumento da taxa de motorização familiar é relativamente linear, com um impacto mais preponderante (declives mais acentuados) para taxas de motorização mais baixas. Apenas o grupo dos homens com mais de 20 anos e com licença de condução apresenta um comportamento claramente não linear com uma redução drástica da utilidade das alternativas com a disponibilidade de um automóvel no agregado. No entanto, a dificuldade de interpretação dos resultados, a sua eventual especificidade para o presente caso sem garantia de ser possível a sua universalidade e o reduzido acréscimo de ajustamento alcançado, apenas sugerem a sua ponderação para modelações mais detalhadas no futuro.

O capítulo termina com uma comparação dos valores dos parâmetros estimados pelos modelos com a gama de valores verificados em outros estudos internacionais. Apesar de se verificarem importantes diferenças entre as especificações utilizadas, quer ao nível da tipologia de variável quer das unidades utilizadas, verificou-se que para a maioria das variáveis passíveis de comparação os resultados são coerentes e balizados pelos valores internacionais.

Foi, no entanto, possível verificar tendências que não corroboram as especificações habitualmente consideradas em estudos internacionais para certo tipo de variáveis. Atendendo aos resultados obtidos, foi possível constatar que, ao contrário do que geralmente se assume nos estudos internacionais, para a maioria das variáveis a especificação com combinação de binárias se apresenta como mais correta para introduzir os seus efeitos nos modelos de escolha modal. Com um ligeiro acréscimo de trabalho e de complexidade do modelo, embora também com algum risco de perda de generalidade, é possível representar de forma mais fidedigna a não linearidade do impacto de certos fatores nas escolhas modais.

Exemplo claro desta inconsistência verifica-se na utilização da especificação contínua/discreta para descrever o impacto da idade nas escolhas modais, cuja variabilidade de comportamentos claramente justifica o uso de uma especificação não linear. Apesar de não se terem encontrado estudos internacionais que considerem as variáveis nível de instrução ou motivos de viagens, os resultados obtidos também sugerem a sua inclusão através de especificação não linear.

Outra tendência verificada nos estudos internacionais, não corroborada pelos resultados obtidos, está relacionada com a eventual combinação de variáveis exógenas, eventualmente correlacionadas, nomeadamente o rendimento familiar e a taxa de motorização familiar. Constatou-se que, contrariamente ao assumido em muitos estudos internacionais onde em geral o rendimento do agregado é sempre assumido como relevante, foi a variável taxa de motorização familiar *per capita* ou por adulto que maior potencial revelou para introduzir indiretamente as condições económicas do agregado na modelação do processo de escolha modal. Estes resultados podem dever-se quer à maior dificuldade de definir com precisão o

real rendimento familiar comparativamente à taxa de motorização, quer a uma possível correlação entre ambas, dado que rendimentos familiares maiores tendem a originar taxas de motorização superiores.

Foi também possível verificar que, ao contrário do que é correntemente assumido noutras modelações, uma opção por uma simples variável binária para a distinção entre sexos não revela toda a informação pertinente sobre os seus comportamentos, uma vez que dependendo da idade o seu comportamento revela diferentes tendências. Assim sugere-se a combinação destas variáveis e, eventualmente, algumas outras nos modelos de forma a captar estas simbioses. Apesar da sua especificidade, a utilização de variáveis binárias definidoras de combinações de características sociais próprias de grupos com comportamentos homogêneos, revelou potencial para introduzir essa informação nos modelos de forma fidedigna.

Embora a teoria puramente económica considere que apenas fatores definidores do “produto” têm maior impacto na seleção/distinção entre alternativas, o sistema de transporte e a escolha do modo a utilizar aparentemente são um caso especial, em que as características sociais de cada sujeito se apresentam muito preponderantes. O estudo revelou que os fatores mais determinantes para a opção de não utilizar o automóvel são a sua (in)disponibilidade e a (im)possibilidade do seu uso. Com exceção das viagens de curta distância, particularmente as menores que 1km, mas até certo ponto as menores de 2km, para as restantes situações apenas uma grande dificuldade de utilizar o automóvel poderá levar ao uso de alternativas.

Os resultados relacionados com as políticas de mobilidade sugerem um potencial impacto das mesmas nas escolhas modais muito limitado, ou que a sua aplicação de forma isolada e descoordenada das políticas para os restantes subsistemas de transportes será de eficácia muito limitada. O seu potencial para auxiliar alterações de paradigmas no processo de decisão poderá, no entanto, aumentar quando integradas de forma coerente num quadro de políticas que visem tornar as alternativas ao automóvel mais atrativas, através das convenientes medidas complementares de restrição ao livre uso do automóvel.

Os grupos com maior apetência para utilizar as alternativas mais sustentáveis são as mulheres e os jovens até aos 18 anos, devido à presente dificuldade em aceder ao automóvel. Atendendo ao provável conhecimento que têm dos sistemas de transporte alternativos, deverão ser os grupos cujo comportamento deverá ser mais “acarinhado” e apoiado de forma a garantir que no futuro, mediante a disponibilização de automóvel, as suas opções se mantenham.

Por outro lado, o grupo dos homens entre os 20 e os 65 anos são os que mais utilizam o automóvel. Assim, parece justificar-se o estudo da sua motivação, de forma a avaliar quais as medidas com maior potencial para tornar o mesmo menos atrativo perante as alternativas.

Capítulo VII

Simulação de Cenários Futuros

1. Introdução

Do capítulo VI resultaram dois modelos de escolhas modais que demonstraram tendências coerentes e revelaram as variáveis com maior potencial para a definição das escolhas modais para viagens diárias. Embora haja necessidade de estudos mais aprofundados com modelos mais complexos, a utilização dos modelos obtidos para a simulação de cenários e avaliação de tendências permite já fornecer previsões válidas e relevantes e apontar caminhos para análises e modelações futuras. Assim, no presente capítulo apresentam-se as análises realizadas utilizando os modelos definidos anteriormente.

Numa primeira fase, e atendendo aos parâmetros obtidos nos modelos, definiram-se perfis de utilizadores-tipo com maior e menor propensão para a utilização dos modos mais relevantes. Atendendo a estes perfis realizaram-se simulações com os modelos propostos para avaliar a potencial repartição modal para os mesmos. Dada a natureza “limite” destes perfis, o procedimento foi também repetido para perfis intermédios mais realistas. Numa segunda fase realizou-se uma análise de sensibilidade dos modelos, testando modelações com variações isoladas de cada uma das diferentes variáveis, verificando o seu impacto nas previsões de repartições modais.

Após estes testes às respostas dos modelos, realizou-se um estudo de previsões para cenários futuros. Definiram-se quatro possíveis cenários considerando prováveis desenvolvimentos socioeconómicos e eventual vontade política de fomentar hábitos de mobilidade mais sustentável. Atendendo aos diferentes cenários pretendidos, foram definidos conjuntos exequíveis de valores para as diferentes variáveis, suscetíveis de caracterizarem cada cenário. Cada um dos cenários foi aplicado aos modelos para avaliação dos seus possíveis efeitos na repartição modal futura.

Atendendo à dificuldade de previsão de evolução futura das variáveis mais relevantes, taxa de motorização familiar e percentagem de encartados, e à facilidade de obtenção de dados relacionados com o Produto Interno Bruto e dos censos populacionais, o capítulo finaliza com a análise de eventuais relações entre as variáveis socioeconómicas com maior potencial explicativo e o rendimento médio *per capita* (que na análise no capítulo anterior não justificou a sua inclusão). Esta eventual relação deverá permitir a estimação indireta destas

variáveis mais relevantes através da evolução do rendimento médio *per capita*, permitindo assim, de forma objetiva e simples, a sua utilização em previsões futuras.

2. Análise da Potencial Influência das Características Socioeconómicas dos Indivíduos - Perfis Padrão

Dos resultados do capítulo anterior foi possível avaliar quais as características socioeconómicas mais relevantes e a sua importância relativa. Apesar de não serem características do “produto” em apreciação, a teoria comportamental sugere que as mesmas podem influenciar determinantemente as escolhas modais realizadas. Nos diversos modelos exploratórios tanto a existência de licença de condução como a taxa de motorização familiar representam parte dos fatores mais decisivos para as opções diárias. As restantes variáveis socioeconómicas, apesar de menos preponderantes, revelaram a introdução de informação relevante que permitiu a melhoria de ajustamento dos modelos. Atendendo à especificidade das variáveis binárias de combinação de variáveis sociais testadas no capítulo VI, caracterizadoras de grupos com comportamentos que se verificaram relativamente homogêneos, optou-se por não as utilizar nestas análises dada a diversidade de variações pretendidas para as mesmas.

Assim, realizou-se uma análise sobre a influência na repartição modal de diferentes combinações das variáveis socioeconómicas que potenciavam a atratividade de determinada alternativa. Se para as alternativas Bus e Walk essa análise implicava a simulação do modelo com as combinações dos parâmetros que se constatou serem os mais valorizados para esses modos, a análise para o Automóvel implicou a análise de diferentes combinações e a seleção das que, sendo teoricamente possíveis, apresentavam os valores extremos. Para esta análise optou-se por considerar os resultados do modelo com melhor ajustamento utilizando todas as variáveis socioeconómicas - MNL3.139. Foi possível definir perfis tipo de utilizadores que podem servir para balizar possíveis cenários extremos e fornecer dados para estipular outros perfis intermédios que expressem os objetivos de políticas futuras de mobilidade sustentável.

Assim, definiram-se seis perfis tipo de utilizador extremos para os três modos mais relevantes, atendendo aos resultados que sugerem uma maior ou menor utilização desses modos:

- Auto+: Indivíduos e viagens que apresentam características que normalmente estão associadas a uma elevada utilização do automóvel e grande relutância em mudar de escolha modal;
- Auto-: Indivíduos e viagens com características associadas a uma reduzida utilização do automóvel particular, quer voluntária, quer involuntariamente;

- Bus+: Indivíduos e viagens que apresentam características que normalmente estão associadas a uma elevada utilização do Bus;
- Bus-: Indivíduos e viagens com características que atualmente estão associadas a uma reduzida utilização do BUS;
- Walk+: Indivíduos e viagens que apresentam características que normalmente estão associadas a uma elevada utilização do modo Walk;
- Walk-: Indivíduos e viagens com características que atualmente estão associadas a uma reduzida utilização do modo Walk;

No quadro VII.1 apresentam-se as características consideradas para cada um dos perfis tipo extremos considerado.

Quadro VII. 1 - Perfis tipo extremos de utilizadores para os modos mais relevantes

Variável	AUTO+	AUTO- ^(a)	Bus+ ^(a)	Bus-	Walk+	Walk-
Lc	1	0	0	1	0	1
NC0	0	1	1	0	1	0
NC1	0	0	0	0	0	0
NC2	0	0	0	0	0	0
NC3	0	0	0	0	0	1
NC4a	1	0	0	1	0	0
R1	0	1	1	0	1	0
R2	0	0	0	0	0	0
R3	0	0	0	0	0	0
R4	0	0	0	0	0	0
R5	1	0	0	1	0	1
In1	0	0	0	1	1	1
In2	0	0	0	0	0	0
In3	0	1	1	0	0	0
In4	1	0	0	0	0	0
Id1	0	0	0	0	0	1
Id2	0	1	1	0	0	0
Id3	1	0	0	1	0	0
Id4	0	0	0	0	0	0
Id5	0	0	0	0	1	0
Sx	0	1	1	0	1	0
Mot	0	0	0	1	0	1

(a) - Mesma situação

É possível constatar que a situação que potencia o uso do modo Bus é, segundo o modelo, a mesma que é mais desfavorável ao uso do Automóvel. Os resultados da aplicação destes perfis ao modelo MNL3.139 são apresentados graficamente nas figuras VII.1 e VII.2. A figura VII.1 apresenta a utilidade relativa de cada modo para cada perfil e na figura VII.2 é possível verificar a estimativa de repartição modal obtida da aplicação do modelo com essas utilidades relativas. Destes resultados é possível verificar que o modelo sugere que em determinadas circunstâncias seria teoricamente possível reduzir drasticamente o uso do Automóvel (até

apenas 4% das viagens), potenciando o uso dos modos Bus (para cerca de 57%) e Walk (para cerca de 29%).

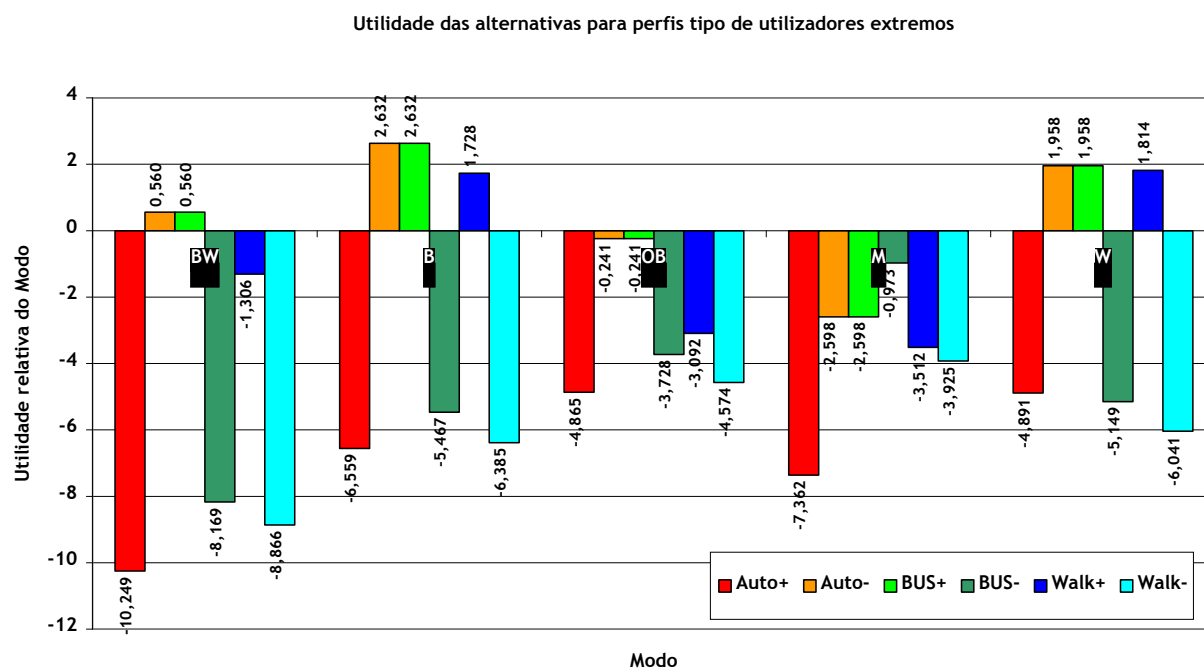


Figura VII. 1 - Utilidades parciais para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores extremos (MNL3.139).

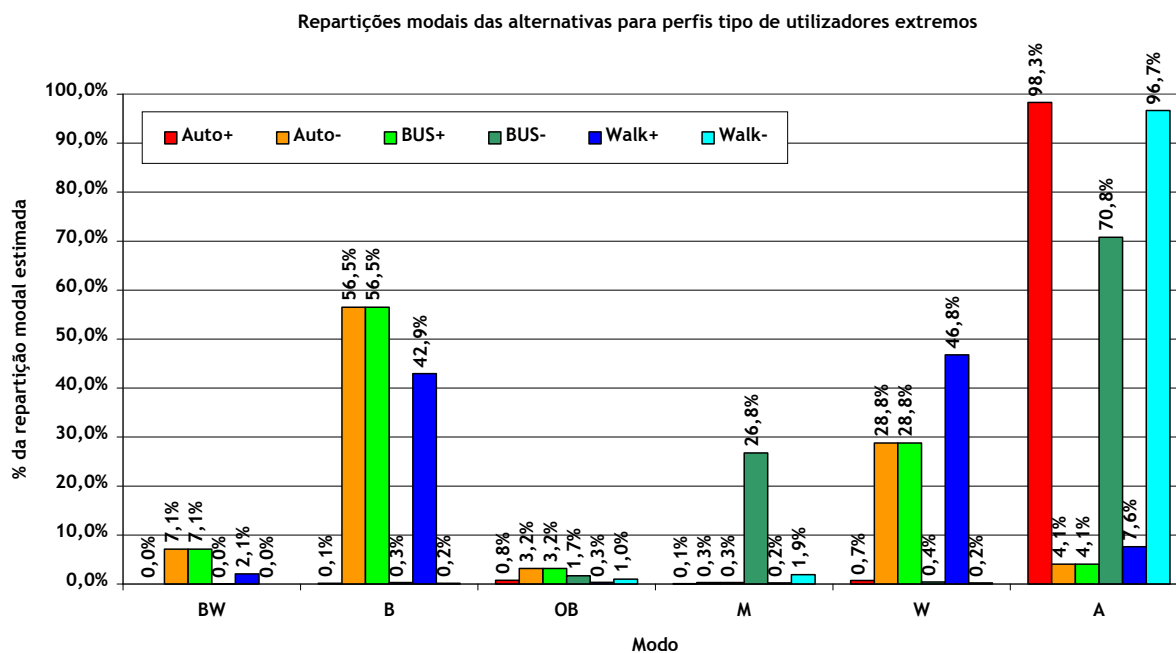


Figura VII. 2 - Repartições modais estimadas pelo modelo MNL3.139 para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores extremos.

Estas circunstâncias limite implicam sempre a inexistência tanto de licença de condução como de automóvel disponível no agregado. No entanto, esta combinação de fatores extremos

é de muito difícil obtenção na realidade, expresso pela falta de amostras obtidas da base de dados utilizada de 95 426 inquéritos, não permitindo a comparação das estimativas/previsões do modelo com resultados efetivos da base de dados e, portanto, irrealista para testes práticos.

Assim, no sentido de criar cenários mais realistas que permitem comparar as estimativas do modelo com os resultados da base de dados, definiram-se perfis tipo intermédios que apresentassem amostras com dados suficientes para permitir avaliar o ajuste do modelo. Definiram-se assim seis perfis tipo de utilizadores intermédios:

- Perfil 1: Homem, entre 35 e 65 anos, nível socioeconómico baixo, sem carro, com licença e com ensino básico, sem moto disponível, com potencial para utilizar menos o automóvel e mais as alternativas sustentáveis;
- Perfil 2: Homem, entre 35 e 65 anos, nível socioeconómico baixo, sem carro, com licença e com ensino básico, com moto disponível, para avaliar o possível impacto de ter alternativas de mobilidade particulares nos indivíduos com nível social mais baixo;
- Perfil 3: Mulher, entre 20 e 34 anos, nível socioeconómico baixo, sem carro, com licença e com ensino básico, sem moto disponível, de modo a comparar uma situação similar ao perfil 1 mas para mulheres;
- Perfil 4: Mulher, entre 35 e 65 anos, nível socioeconómico médio, com pelo menos 0,5 carros/membro do agregado familiar, sem licença e com ensino básico, sem moto disponível, para avaliar a situação social provavelmente mais comum para as mulheres;
- Perfil 5: Mulher, entre 35 e 65 anos, nível socioeconómico médio, com pelo menos 0,5 carros/membro do agregado familiar, com licença e com ensino básico, sem moto disponível, para comparar com o perfil 4 avaliando a diferença de dispor de licença de condução;
- Perfil 6: Homem, entre 20 e 34 anos, nível socioeconómico médio, com pelo menos 0,5 carros/membro do agregado familiar, com licença e com ensino básico, com moto disponível, para avaliar a situação social provavelmente mais comum do perfil 4 mas para homens.

No quadro VII.2 apresentam-se as características consideradas para cada um dos perfis tipo intermédios considerados. É de salientar a reduzida amostra disponível para os perfis 3 e 6.

Na figura VII.3 é possível comparar as estimativas do modelo com as repartições modais verificadas na BD para cada um dos perfis intermédios considerados. É possível constatar que na maioria dos casos as estimativas do modelo são corroboradas pelos resultados da base de dados, com exceção dos perfis com o modo Mota disponível - perfis 2 e 6. Nestes perfis,

verificam-se discrepâncias nas repartições modais para os modos Moto e Auto ou BUS, e no perfil 3 para o modo Moto e Walk. Estas discrepâncias podem dever-se a um mau comportamento do modelo em situações em que existem condições que potenciam o uso do modo Moto.

Quadro VII. 2 - Perfis tipo de utilizadores intermédios

Variável	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Perfil 5	Perfil 6
Lc	1	1	1	0	1	1
NC0	1	1	1	0	0	0
NC1	0	0	0	0	0	0
NC2	0	0	0	1	1	1
NC3	0	0	0	0	0	0
NC4a	0	0	0	0	0	0
R1	0	0	0	0	0	0
R2	1	1	1	0	0	0
R3	0	0	0	1	1	1
R4	0	0	0	0	0	0
R5	0	0	0	0	0	0
In1	0	0	0	0	0	0
In2	1	1	1	1	1	1
In3	0	0	0	0	0	0
In4	0	0	0	0	0	0
Id1	0	0	0	0	0	0
Id2	0	0	0	0	0	0
Id3	0	0	1	0	0	1
Id4	1	1	0	1	1	0
Id5	0	0	0	0	0	0
Sx	0	0	1	1	1	0
Mot	0	1	0	0	0	1
Amostra	824	359	208	587	532	190

Da análise das utilidades relativas de cada modo para os diferentes perfis (figura VII.4), verifica-se que o perfil 4 é o que mais potencia o uso do modo Bus, Walk e Bus+Walk, revelando o impacto que a falta de licença de condução e de ser mulher tem nas escolhas das alternativas. O perfil 2 potencia o uso de Moto, uma vez que está disponível como alternativa enquanto o Auto não. É possível verificar o elevado impacto no decréscimo das utilidades de quase todas as alternativas com a disponibilidade de automóvel, principalmente para os perfis 5 e 6, um pouco atenuado no perfil 4, provavelmente pela ausência de licença de condução.

É assim possível verificar que as condições socioeconómicas têm um elevado potencial para influenciar as escolhas modais diárias dos cidadãos. Corroborando os resultados dos modelos, a análise revelou que a possibilidade de utilização do automóvel particular, quer socialmente quer economicamente, se apresenta como o principal vetor de atuação para influenciar as escolhas modais dos cidadãos. A sua conjugação com medidas de políticas de mobilidade sustentável, será analisada nos pontos seguintes.

Repartições modais das alternativas para perfis tipo de utilizadores i

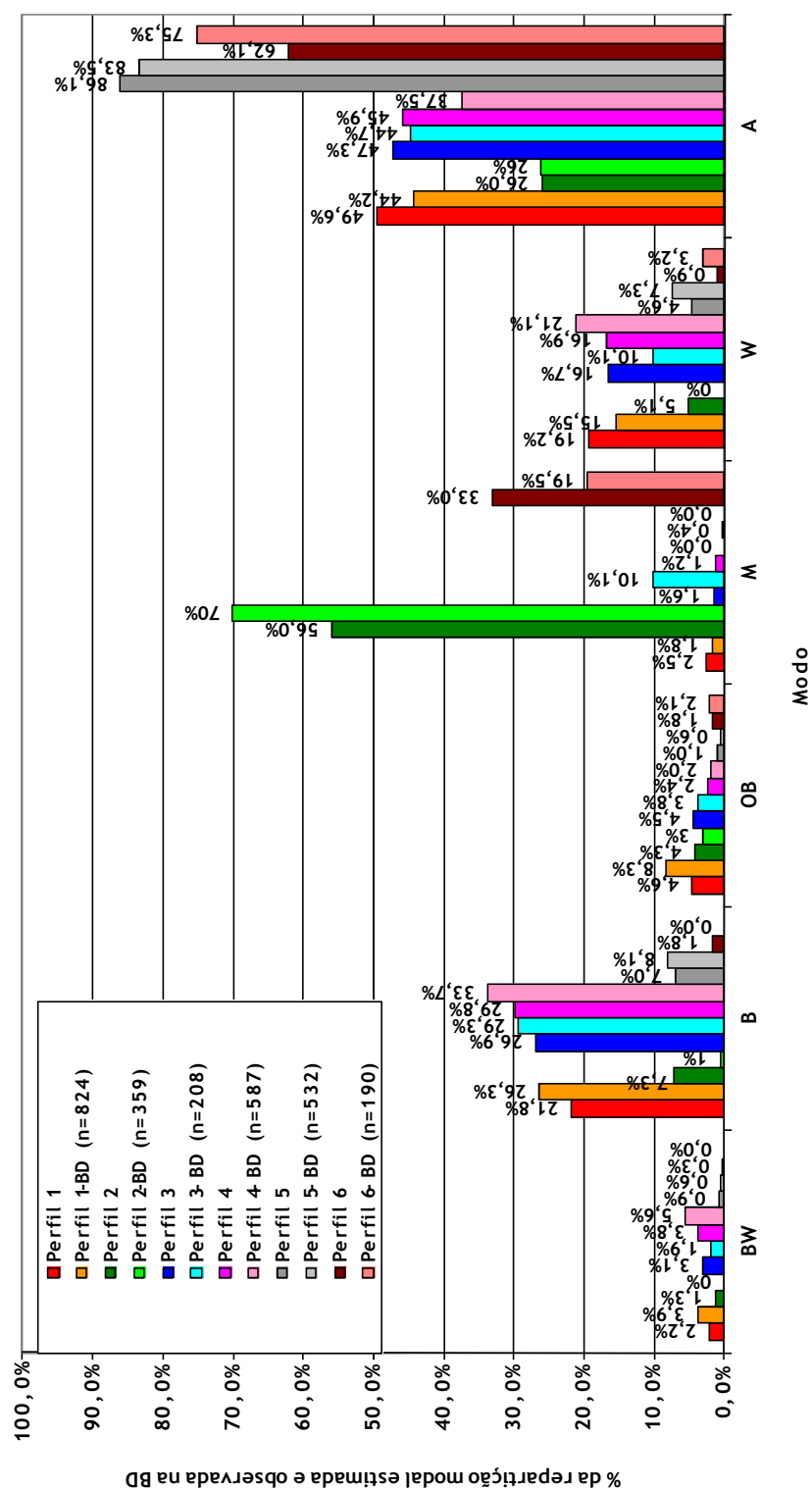


Figura VII. 3 - Comparação das Repartições modais reais da BD com as estimadas pelo modelo MNL3.139 para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores intermédios.

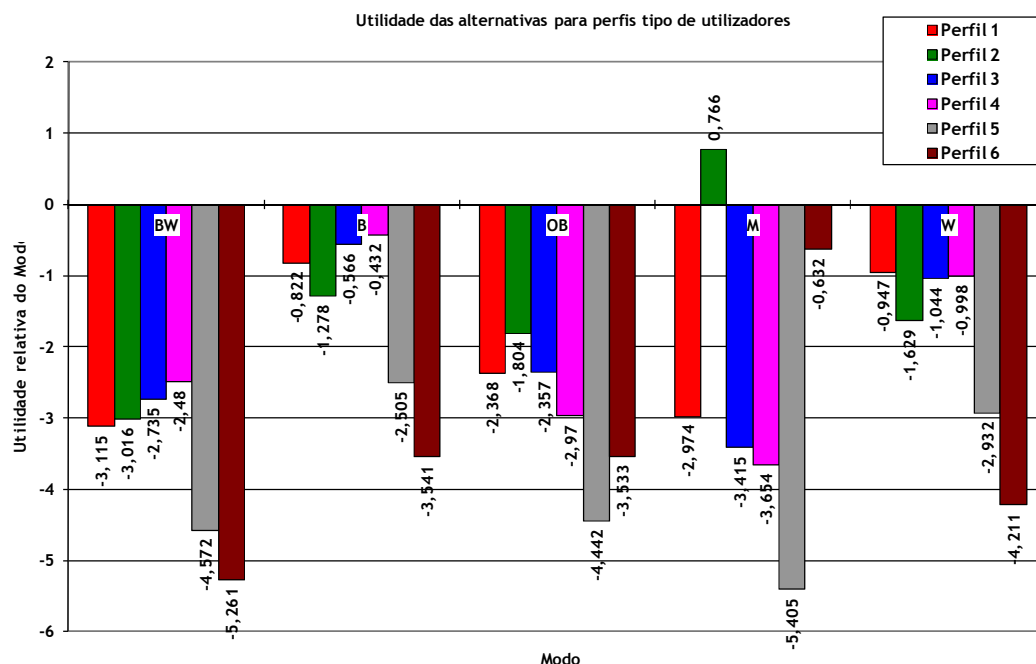


Figura VII. 4 - Utilidades parciais para os diferentes modos de transporte considerando os diferentes perfis tipo de utilizadores intermédios (MNL3.139).

3. Análise de Sensibilidade dos Modelos a Alterações Sociais e a Mudanças de Políticas de Mobilidade

Para avaliar o impacto de diferentes possíveis alterações nas variáveis explicativas consideradas mais relevantes nas previsões dos modelos, realizou-se uma análise de sensibilidade a diferentes variações isoladas em cada uma das variáveis mais propensas a sofrerem evoluções futuras. Para o efeito foram realizados simuladores em folhas de cálculo Excel® aplicados às bases de dados selecionadas devidamente ponderadas (através dos fatores de ponderação/extrapolação da base de dados original), para estimar repartições modais atendendo às variações pretendidas, introduzidas de forma relativamente aleatória (pseudoaleatória).

Na definição das gamas de valores a aplicar nas variações pretendidas considerou-se que:

- *Indisponibilidade de automóvel no agregado familiar* - NC=0 (NC0=1): No futuro esta variável apenas poderia diminuir (NC0=0), com a aquisição de pelo menos um automóvel para o agregado (passando, portanto, a ter taxa de motorização familiar maior que zero). Assim, nas simulações foram definidas percentagens de redução em relação à proporção atualmente verificada na base de dados em estudo;
- *Taxa de motorização familiar* (NC): Considerou-se que esta variável contínua poderia aumentar ou diminuir quer pela aquisição de mais automóveis quer pela redução/aumento do agregado familiar. As viagens de agregados sem automóvel (NC0=0) que passaram a ter veículo disponível não foram afetados dos coeficientes

de majoração/minoração e considerou-se que os agregados que se mantiveram sem automóvel disponível não sofreriam alterações no futuro;

- *Licença de Condução (Lc)*: Considerou-se que esta variável binária apenas poderia aumentar, ou seja, que no futuro parte das pessoas sem carta de condução a conseguiriam obter. Nos simuladores, considerou-se o aumento percentual pretendido de inquiridos com licença relativamente à situação da base de dados utilizada;
- *Disponibilidade diária de Moto no agregado (Mot)*: Considerou-se que esta variável binária apenas poderia aumentar, com a aquisição de Moto. Assim, nos simuladores também apenas se considerou o aumento percentual pretendido para esta variável, relativamente à base de dados inicial;
- *Sexo (Sx)*: Atendendo à repartição quase equitativa desta variável nas diferentes amostras que se foram considerando, optou-se por considerar que esta variável não variava no futuro. Esta assunção deve-se quer a não haver dados que relacionem a sua evolução com os dados em análise, quer por se considerar que numa amostra suficientemente grande a sua repartição tenderá a manter-se constante, embora a longo prazo o peso do sexo feminino numa população mais envelhecida tenda inexoravelmente a aumentar;
- *Escalão etário (ldb)*: Assumiu-se nas simulações que esta variável não sofria alterações, e que a passagem de registo entre escalões ao longo dos anos permitiria a manutenção das distribuições verificadas na base de dados atual. Não se estudou, assim, qualquer cenário para o continuado processo de envelhecimento populacional;
- *Nível de Instrução (ln)*: Foi assumida uma situação similar à da variável escalão etário, não sendo, assim, testado nenhum cenário de um esperado, e desejável, aumento progressivo dos níveis de literacia na população;
- *Motivo da viagem (Mt)*: Esta variável discreta foi definida sem alterações. Assumiu-se que as distribuições atuais se manteriam na mesma, considerando-se inviável estipular alguma regra para as alterações;
- *Número de viagens totais diárias (nV)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar;
- *Distância mais curta entre centróides (D2)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar. Um acréscimo desta distância poderá ser justificado com a expansão da área urbana em “mancha-de-óleo” ou com a extinção de atividades em determinadas zonas da área em estudo, enquanto a sua diminuição poderá advir do aumento de atividades/oportunidades nas zonas mais próximas dos espaços residenciais ou da densificação, fortuita ou resultado de uma

alteração deliberada das políticas de ordenamento do território, da ocupação dos espaços urbanos, levando a percursos menores para as mesmas atividades;

- *Percentagem de estacionamento pagos (PO)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar;
- *Razão entre a oferta de estacionamento e o total de viagens atraídas (OV)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar;
- *Frequência média horária diária dos BUS (Ftl1)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar;
- *Razão entre tempos de viagem por Bus e por Auto (TBA)*: Esta variável foi considerada de forma semelhante à taxa de motorização familiar. Um aumento desta variável poderá resultar de mais congestionamentos nas áreas centrais ou da deterioração dos veículos de transporte coletivos, enquanto uma diminuição poderá dever-se a uma redução dos congestionamentos em áreas centrais pela limitação do uso do automóvel particular ou devido a medidas de gestão de tráfego de discriminação positiva para os Bus;

No quadro VII.3 apresentam-se as variações propostas para as análises de sensibilidade das diferentes variáveis para ambos os modelos utilizados

Quadro VII. 3 - Gamas de variações utilizadas nas análises de sensibilidade dos modelos

Variável	Gamas de Variações	
	MNL7.57	MNL12.16
NC	-25%, -15%, +15%, +25%	-25%, -15%, +15%, +25%
NC0	-25%, -14%, -5,5%	-25%, -14%, -5,5%
Lc	+5,5%, +15,6%, +23,4%	+5,5%, +15,8%, +28%
Mot	+9,6%, +15,5%, +25,5%	-
nV	-25%, -15%, +15%, +25%	-25%, -15%, +15%, +25%
D2	-25%, -15%, +15%, +25%	-25%, -15%, +15%, +25%
PO	-	-50%, -25%, +25%, +50%
OV	-	-50%, -25%, +25%, +50%
Ftl1	-	-50%, -25%, +25%, +50%
TBA	-	-50%, -25%, +25%, +50%

A dificuldade de obter valores inteiros para as variáveis NC0, Lc e Mot na base de dados alterados, deve-se à forma pseudoaleatória como se pretendeu introduzir essas alterações. A alteração do valor da variável em intervalos regulares entre registos ordenados de forma aleatória, poderá não ser a mais realista para cenários futuros uma vez que possibilita

alterações diferenciadas destas variáveis para viagens realizadas por membros do mesmo agregado, e até do mesmo indivíduo.

Nas figuras VII.5 e VII.6 apresentam-se os resultados das simulações atendendo às diferentes gamas de variações consideradas para ambos os modelos.

É possível constatar que as alterações propostas na base de dados para as diferentes variáveis, apresentam os seguintes resultados:

- *Disponibilidade de automóvel no agregado familiar (NC0)*: O modelo MNL7.57 sugere que o aumento dos agregados com automóvel disponível diariamente tem uma influência moderada para os modos Bus e Auto, com uma evolução tendencialmente linear em ambos os casos embora inversa. O modelo MNL12.16 sugere a mesma tendência mas com um comportamento não linear para Bus e Auto;
- *Taxa de motorização familiar (NC)*: Segundo o modelo MNL7.57 apenas os modos Auto e Bus apresentaram sensibilidade às variações propostas para esta variável, mas de forma aparentemente não linear. A simulação através do modelo MNL12.16 corrobora as tendências verificadas para o modelo anterior;
- *Licença de Condução (Lc)*: O modelo MNL7.57 sugere que o aumento proposto dos inquiridos com licença de condução tem uma influência moderada para o modo Bus e mais acentuada para o Auto, com uma evolução tendencialmente linear. O modelo MNL12.16 sugere uma menor influência da variável nos modos Bus e Auto mas mantém uma propensão para um efeito tendencialmente linear;
- *Disponibilidade de diária de Moto no agregado (Mot)*: O modelo MNL7.57 sugere que as alterações propostas a esta variável apenas têm uma influência moderada no uso do modo Moto, e quase nula nas restantes. Para o modo Moto essa influência é aparentemente linear;
- *Número total de viagens do dia (nV)*: Atendendo aos resultados do modelo MNL7.57 as variações propostas a esta variável aparentemente afetam moderadamente os modos Bus e Auto e ligeiramente os restantes modos, de uma forma aparentemente linear. A análise com o modelo MNL12.16 apresenta resultados similares ao modelo anterior;
- *Distância mais curta entre centróides (D2)*: Segundo as simulações com o modelo MNL7.57, as variações propostas para esta variável afetam significativamente o modo Walk e moderadamente os modos Bus e Auto. Esta influência é aparentemente linear para os modos Bus e Auto e não linear para o modo Walk. O modelo MNL12.16 sugere uma influência mais acentuada para os modos Bus e Walk e menos acentuada para o modo Auto, com uma evolução tendencialmente não linear para os modos mais relevantes;

- *Porcentagem de estacionamentos pagos (PO)*: As simulações com o modelo MNL12.16 sugerem que as alterações propostas têm uma influência mais acentuada nos modos Walk e Auto, com um comportamento não linear;
- *Razão entre a oferta de estacionamentos e o total de viagens atraídas (OV)*: Segundo as simulações com o modelo MNL12.16, as alterações a esta variável apresentam uma influência mais acentuada no modo Bus e mais moderada para os modos Walk e Auto. Essa influência aparentemente não é linear;
- *Frequência média horária diária dos BUS (Ftl1)*: Atendendo aos resultados do modelo MNL12.16 as variações propostas para esta variável apresentam uma influência muito reduzida no uso de qualquer das alternativas disponíveis, aparentemente com uma evolução linear;
- *Razão entre tempos de viagem por Bus e por Auto (TBA)*: Segundo as simulações com o modelo MNL12.16 as alterações propostas para esta variável apresentam uma influência mais acentuada no modo Bus e mais moderada para os modos Walk e Auto. Essa influência é, aparentemente, linear;

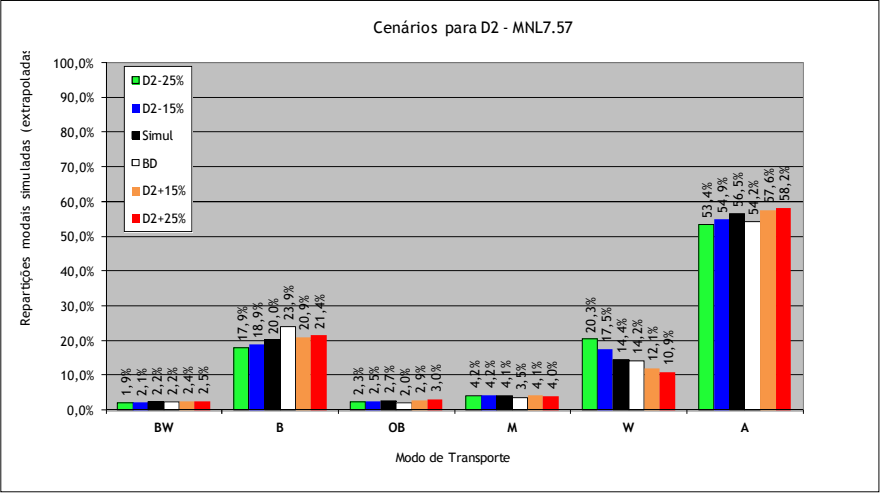
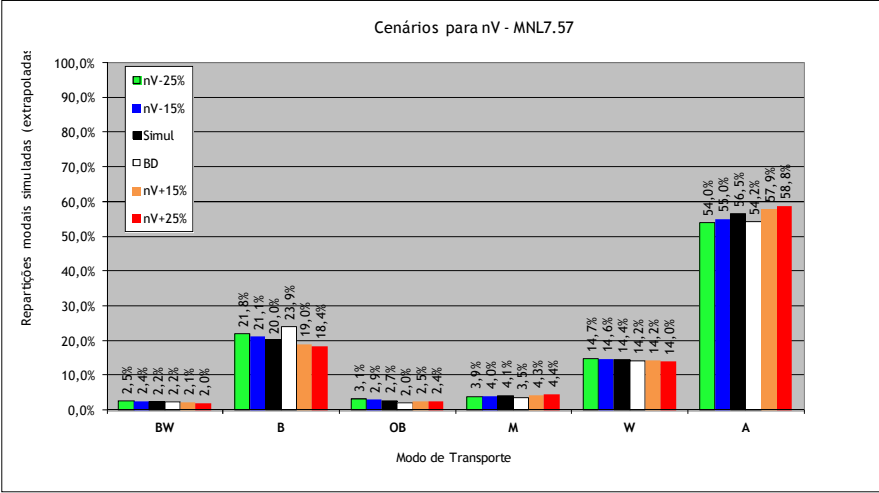
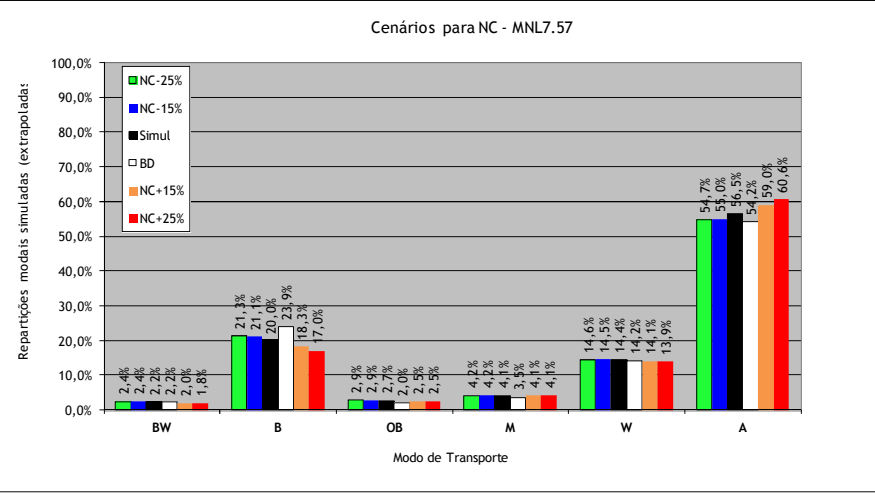
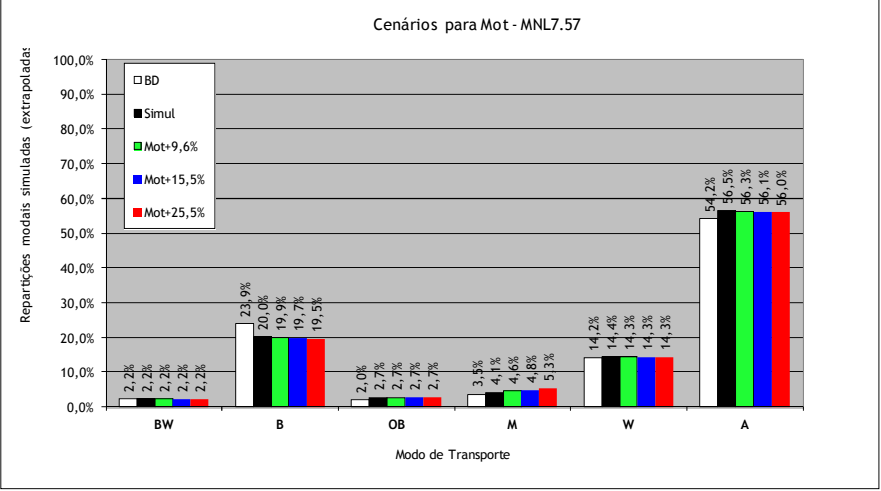
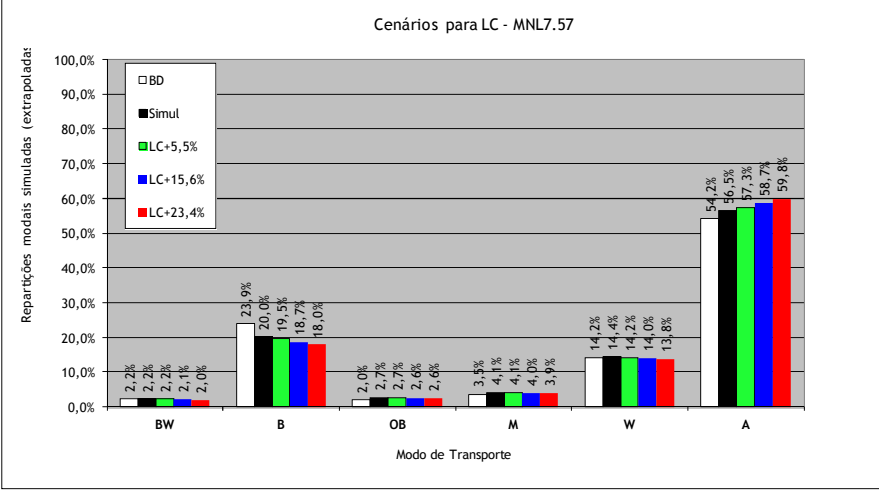
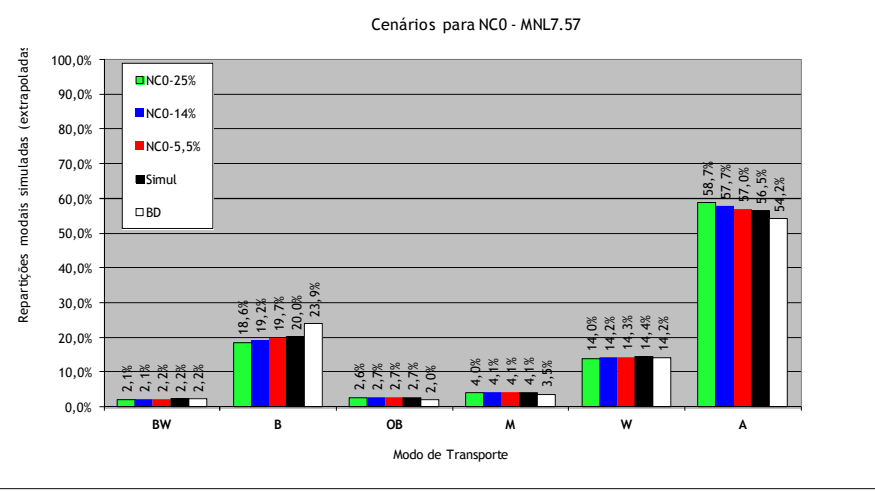


Figura VII. 5 - Análise de sensibilidade de variáveis com o modelo MNL7.57

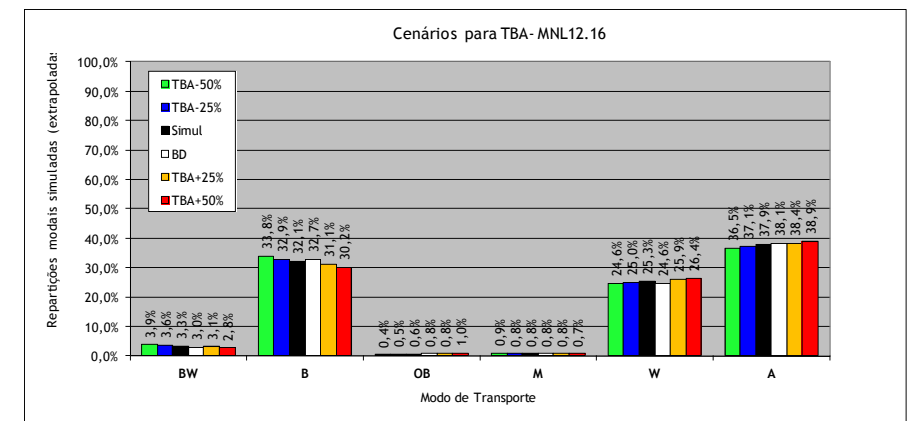
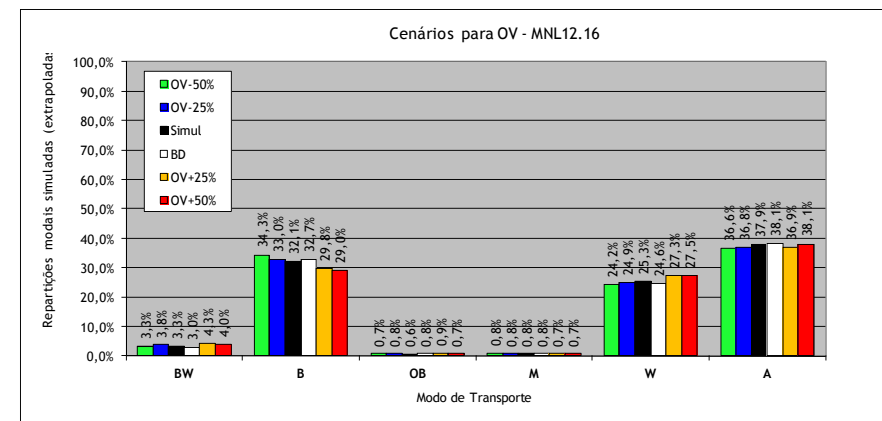
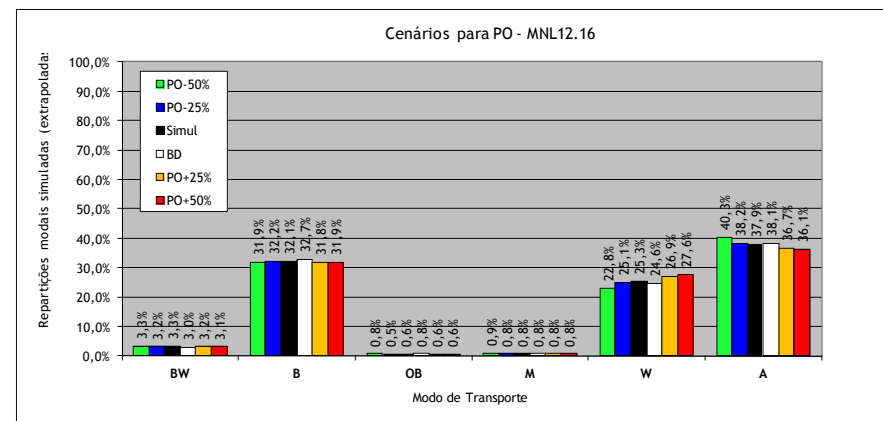
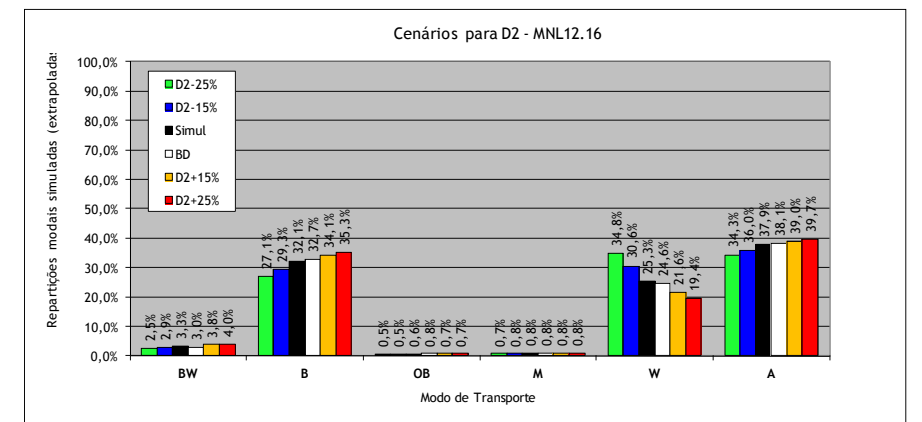
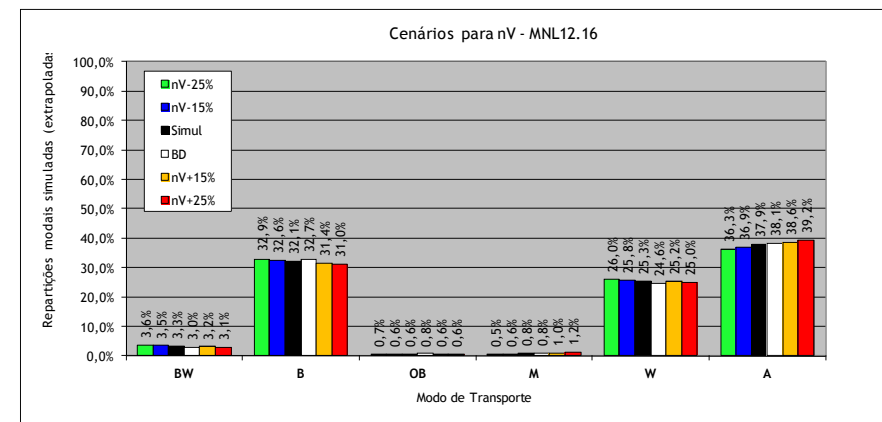
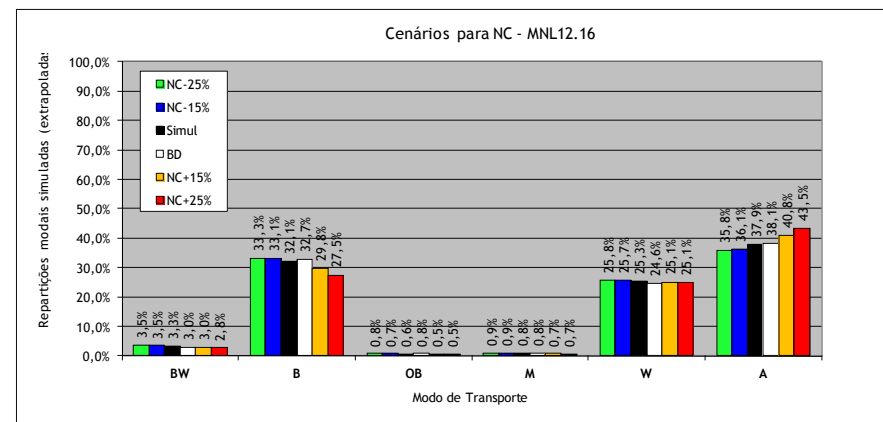
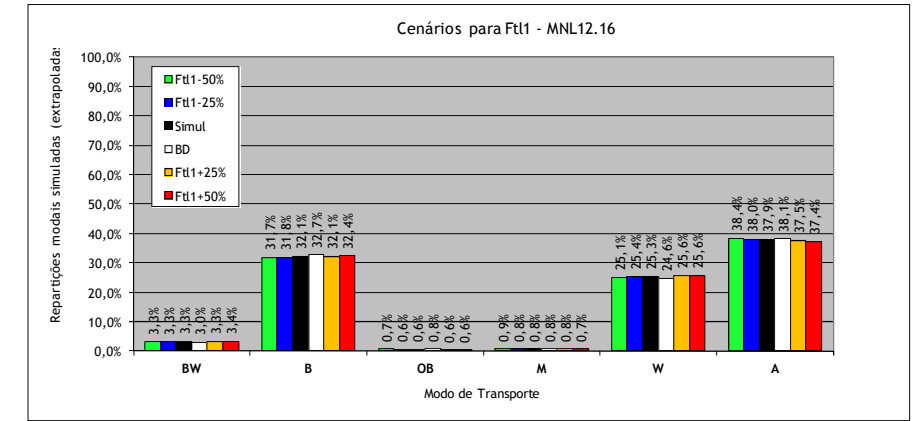
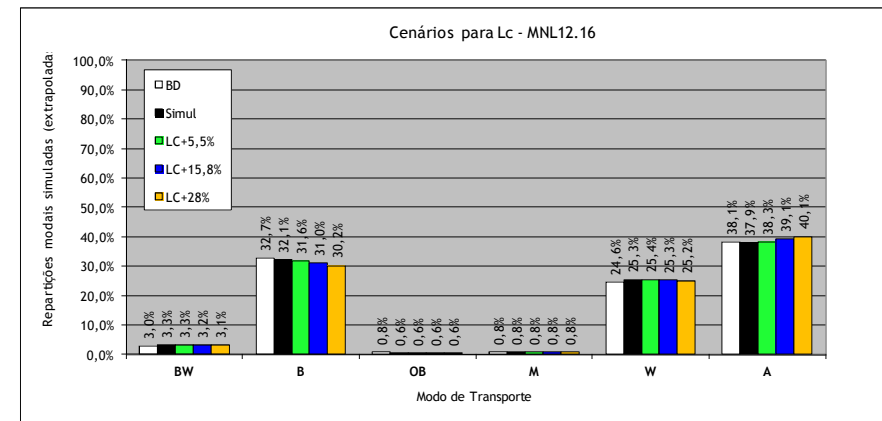
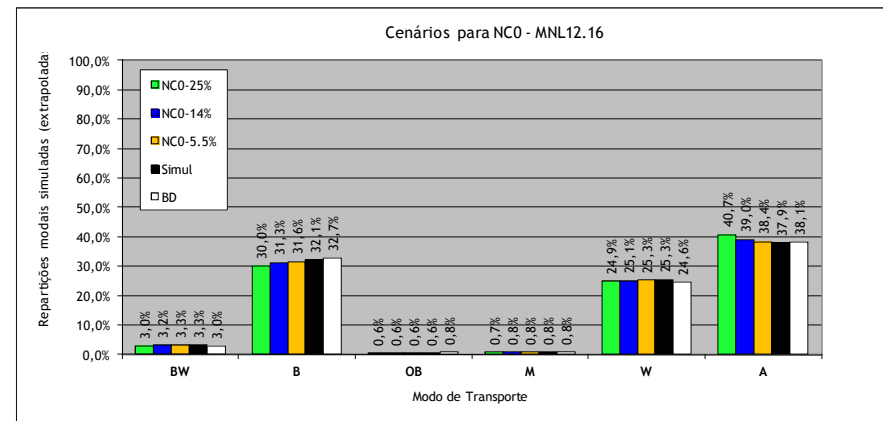


Figura VII. 6 - Análise de sensibilidade de variáveis com o modelo MNL12.16

4. Simulação de Repartições Modais Globais para Cenários Resultantes de Políticas de Mobilidade Específicas

Após as análises anteriores sobre a aplicabilidade dos modelos, promoveu-se a simulação de hipotéticos cenários futuros de repartição modal resultantes de alterações de ordem social credíveis, conjugadas com desejadas políticas de mobilidade mais sustentáveis.

Procurou-se, assim, através dos modelos, testar o potencial de alteração futura dos padrões de mobilidade resultantes quer de alterações credíveis verificadas ao nível da estrutura socioeconómica da sociedade, quer ao nível da adoção de políticas de mobilidade e/ou de ordenamento do território diferentes.

Foram, assim, definidos quatro hipotéticos cenários futuros, combinando dois possíveis cenários socioeconómicos e dois cenários de políticas de mobilidade e ordenamento do território:

- Cenário 1: Otimista Social e Ambientalmente Sustentável (OSAS);
- Cenário 2: Otimista Social e Ambientalmente Tradicional (OSAT);
- Cenário 3: Pessimista Social e Ambientalmente Sustentável (PSAS);
- Cenário 4: Pessimista Social e Ambientalmente Tradicional (PSAT);

Considerou-se que:

- *Cenário Otimista Social*: A evolução socioeconómica da população da área de estudo no futuro é favorável à melhoria das condições sociais e económicas e é propensa a uma maior vontade e disponibilidade/necessidade de aumentar a mobilidade;
- *Cenário Pessimista Social*: A evolução socioeconómica da população da área de estudo no futuro não é favorável e há estagnação ou deterioração das condições sociais e económicas, traduzindo-se em menores necessidades ou possibilidades de mobilidade;
- *Cenário Ambientalmente Sustentável*: A vontade política no futuro tenderá a fomentar o uso dos modos mais sustentável e a restringir a dependência em relação ao automóvel particular, quer através de políticas de ordenamento do território, quer através de políticas de mobilidade mais favoráveis aos modos alternativos ao automóvel;
- *Cenário Ambientalmente Tradicional*: A vontade política tenderá a manter as políticas atuais, sem grandes incentivos ou medidas para reduzir a dependência em relação ao automóvel particular.

Atendendo a estes hipotéticos cenários ponderaram-se diferentes evoluções para as variáveis que se consideraram passíveis de alterações no futuro, e realizaram-se previsões de escolha

modal utilizando os simuladores referidos. No quadro VII.4, apresentam-se as alterações consideradas para os vários cenários definidos.

Quadro VII. 4 - Alterações das variáveis a considerar nos hipotéticos cenários futuros

Variável	Cenário			
	Otimista Social		Pessimista Social	
	Ambientalmente Sustentável	Ambientalmente Tradicional	Ambientalmente Sustentável	Ambientalmente Tradicional
	OSAS	OSAT	PSAS	PSAT
NC	1,1 (+10%)	1,25 (+25%)	0,9 (-10%)	1 (0%)
NC0	-10%	-25%	0%	-6%
Lc	+10%	+22%	0%	+6%
Mot	+25%	+10%	0%	+6%
Rb**	-	-	-	-
Sx	0	0	0	0
ldb	0	0	0	0
In	0	0	0	0
nV	1,1 (+10%)	1,25 (+25%)	0,8 (-20%)	1,0 (0%)
D2	0,75 (-25%)	1,2 (+20%)	1,0 (0%)	1,1 (+10%)
PO*	1,5 (+50%)	1 (0%)	1,25 (+25%)	0,5 (-50%)
OV*	0,5 (-50%)	1,5 (+50%)	0,75 (-25%)	1,25 (+25%)
Ftl1*	1,5 (+50%)	0,75 (-25%)	1,25 (+25%)	0,5 (-50%)
TBA*	0,5 (-50%)	1,25 (+25%)	0,9 (-10%)	1,25 (+25%)

* - Não incluída no modelo MNL7.57; ** - Não incluída nos modelos MNL7.57 e MNL3.139

Atendendo a estas evoluções, na figura VII.7 apresentam-se as repartições modais verificadas na base de dados, as estimadas pelo modelo MNL12.16 para as condições existentes (extrapoladas) e as simuladas pela aplicação de cada cenário ao modelo. Na figura VII.8, apresenta-se as variações verificadas em relação à repartição na base de dados para cada um dos cenários simulados.

É de salientar que o modelo MNL7.57 apenas incorpora duas variáveis que não são socioeconómicas, pelo que a sua aplicação aos cenários com alterações de políticas de mobilidade e de ordenamento do território basicamente não reflete a influência das alterações nestas políticas. Assim, optou-se por apenas se aplicaram os cenários ao modelo MNL12.16.

A aplicação do modelo ao cenário OSAS não apresenta grandes diferenças para as alternativas modais menos relevantes em relação às repartições modais verificadas na simulação com a base de dados atual. No entanto, para os modos Bus, Walk e Auto verifica-se uma redução de 18% e um aumento de 6% e 15%, respetivamente.

O cenário OSAT apresenta resultados semelhantes ao cenário anterior para o modo Bus, sugerindo, no entanto, um decréscimo de 18% no uso do modo Walk e um aumento de 29% no uso do automóvel.

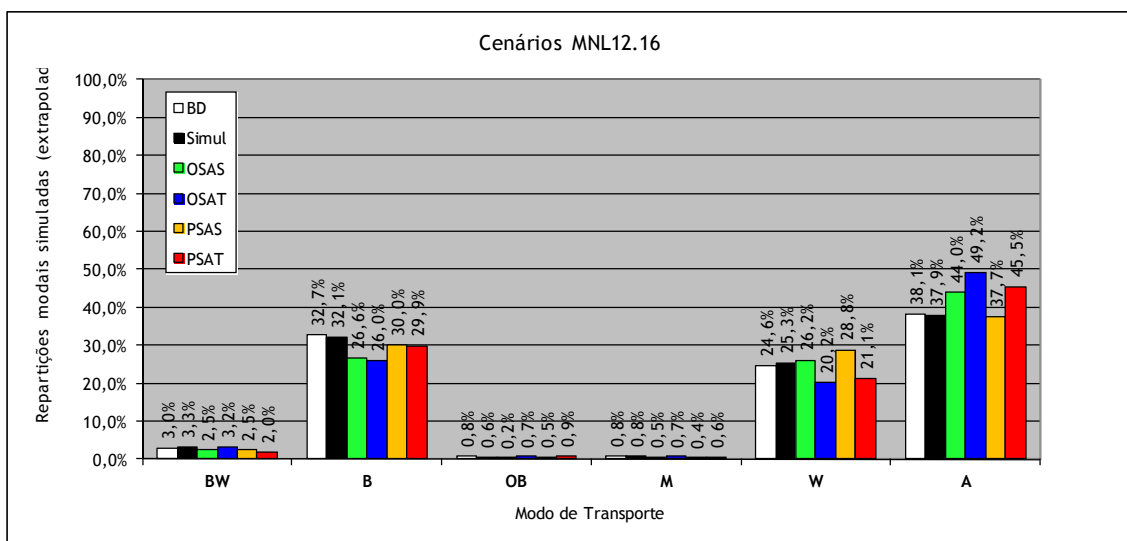


Figura VII. 7 - Resultados da aplicação de cada cenário ao modelo MNL12.16

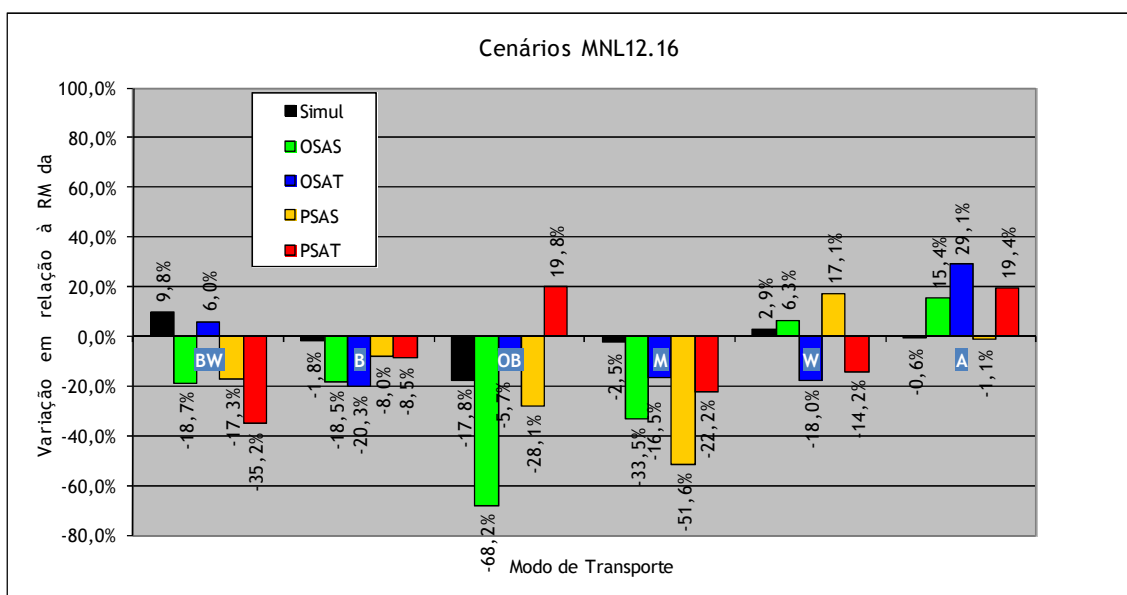


Figura VII. 8 - Variações nos resultados para cada cenário com o modelo MNL12.16, relativamente à repartição modal verificada na base de dados

O cenário PSAS sugere uma repartição modal para o modo Bus e Auto muito semelhante à existente atualmente (uma redução de apenas 8% e 1% respetivamente), mas um acréscimo de 17% no uso do modo Walk.

O cenário PSAT apresenta uma previsão de repartições modais que sugere uma ligeira diminuição do uso dos Bus (-8,5%), um moderado decréscimo no uso do Walk (-14%) e aumento do uso do Auto (19%).

Estes resultados indiciam que um ambiente otimista social leva sempre a um aumento moderado do uso do automóvel e que as medidas para promover políticas de mobilidade e ordenamento do território sustentáveis (com as combinações sugeridas), pouco efeito terão no aumento do uso de alternativas, com exceção do modo Walk, mostrando mesmo que a procura de BUS se apresenta como extremamente rígida e com uma tendência quase total de redução de procura. A combinação de um cenário pessimista social com medidas para promover um desenvolvimento ambientalmente sustentável, apenas apresentam potencial para manter a utilização do Auto ao nível atual.

5. Relação entre o Rendimento Mensal Médio *per Capita* e as Variáveis Explicativas mais Relevantes

Admitindo a validade de determinado modelo matemático para a simulação de uma determinada realidade, definida pela informação contida na base de dados utilizada na sua estimação, um dos objetivos deste seu desenvolvimento será a sua aplicação para previsão de repartições modais atendendo a alterações/evoluções no futuro. Para esta previsão há a necessidade de estimar a evolução futura das variáveis independentes utilizadas nos modelos. Se para algumas dessas variáveis essa previsão está relacionada com a evolução desejada para o sistema em análise e das opções de política que se pretendem fomentar, outras há cuja evolução é mais autónoma e resulta de uma conjugação de fatores externos alheios a estas políticas. Variáveis cuja evolução está fortemente relacionada com fatores externos à política de mobilidade e do território são as socioeconómicas. Se variáveis como o sexo, a repartição etária da população, do seu nível de instrução, da dimensão e do rendimento médio dos agregados, são periodicamente estudadas e sujeitas a processos de estimação da sua evolução futura por instituições públicas credíveis, como o INE, outras há cuja evolução previsível não é tão estudada nem conhecida. Exemplos destas últimas são as taxas de motorização familiar, quer de automóveis quer de motos, e a percentagem de encartados. Um conhecimento de eventuais relações entre as variáveis de previsão mais estudadas com as menos exploradas, e relevantes para o teste de políticas de mobilidade e do território, poderá auxiliar a estimativa de evolução futura das últimas e permitir a obtenção de dados de base para a modelação mais credíveis.

Assim, estudaram-se eventuais relações entre o rendimento mensal médio *per capita* de cada agregado e a sua taxa de motorização assim como com a percentagem de encartados para cada categoria.

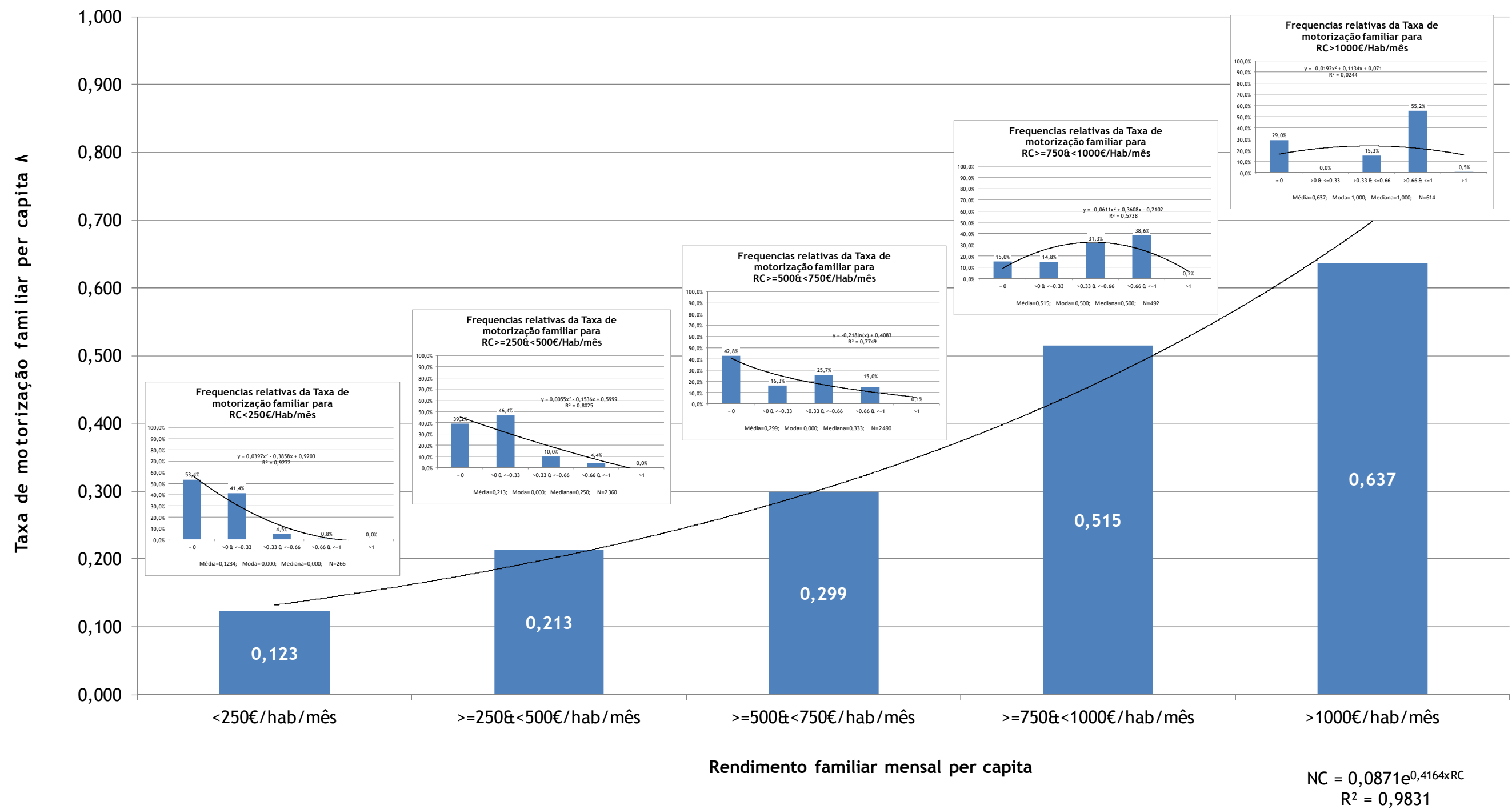


Figura VII. 9 - Evolução da taxa de motorização familiar *per capita* média com o rendimento mensal do agregado *per capita*

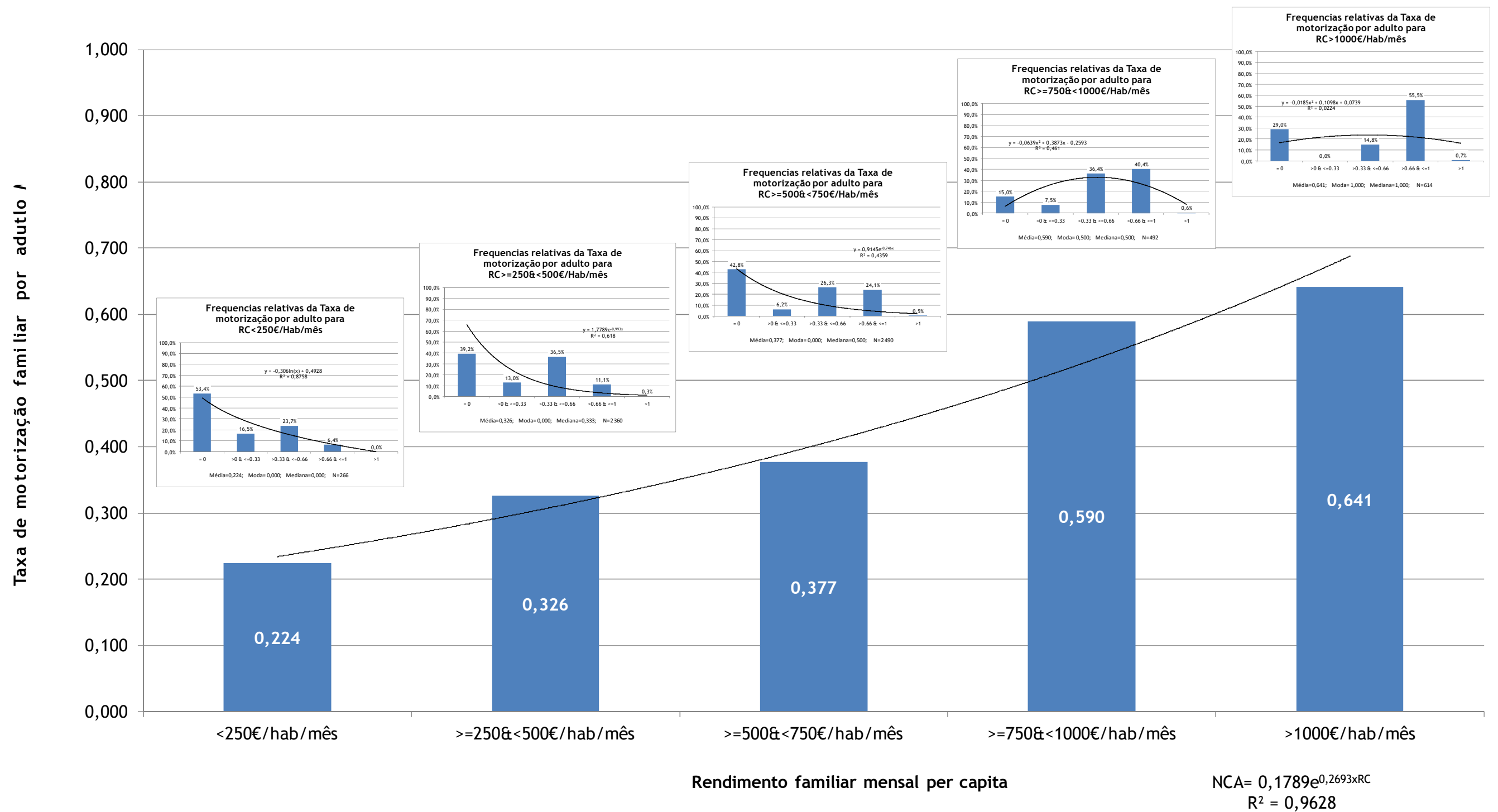


Figura VII. 10 - Evolução da taxa de motorização familiar por adulto média com o rendimento mensal do agregado *per capita*

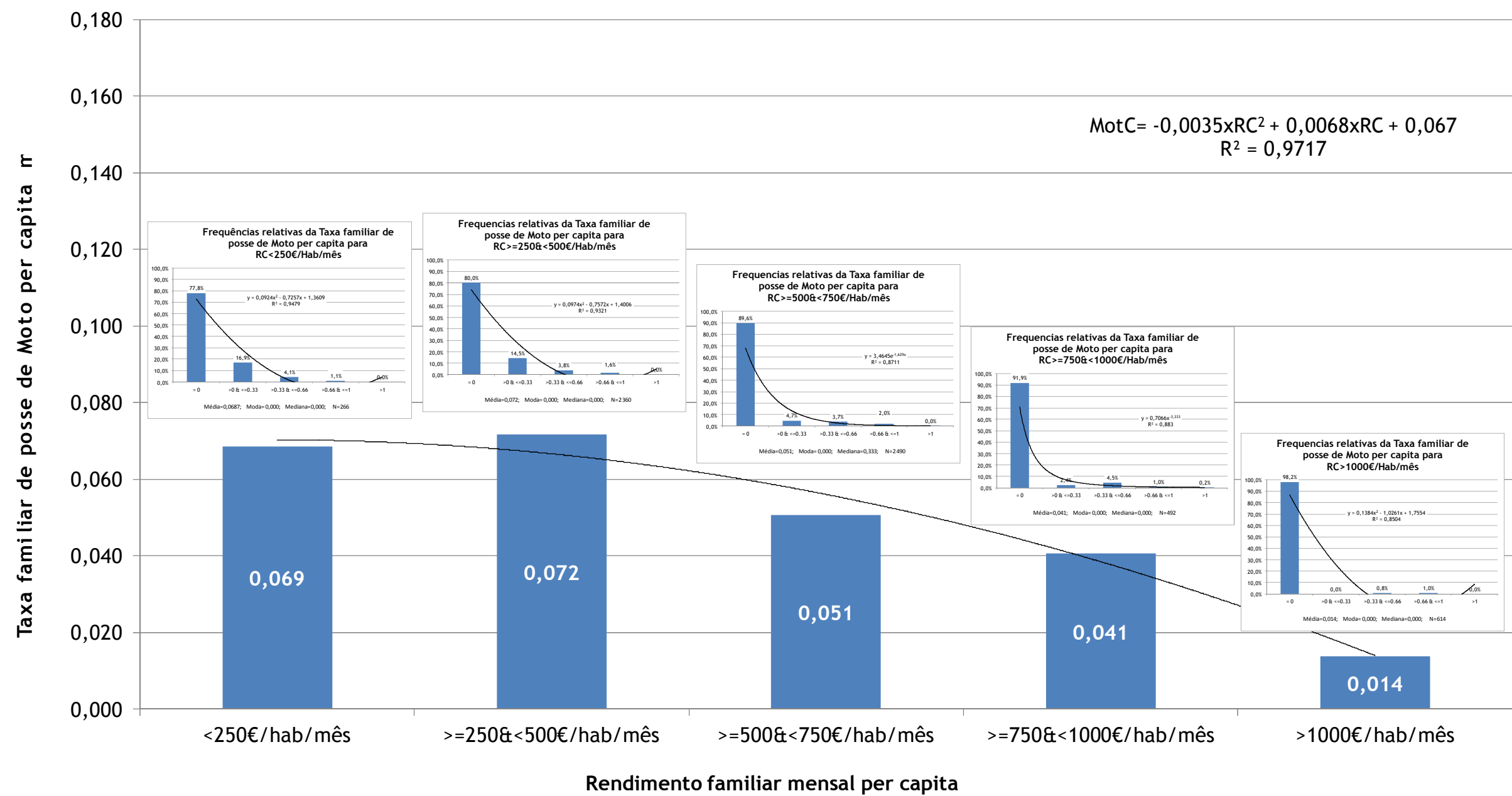


Figura VII. 11 - Evolução da taxa de motorização familiar de moto *per capita* média com o rendimento mensal do agregado *per capita*

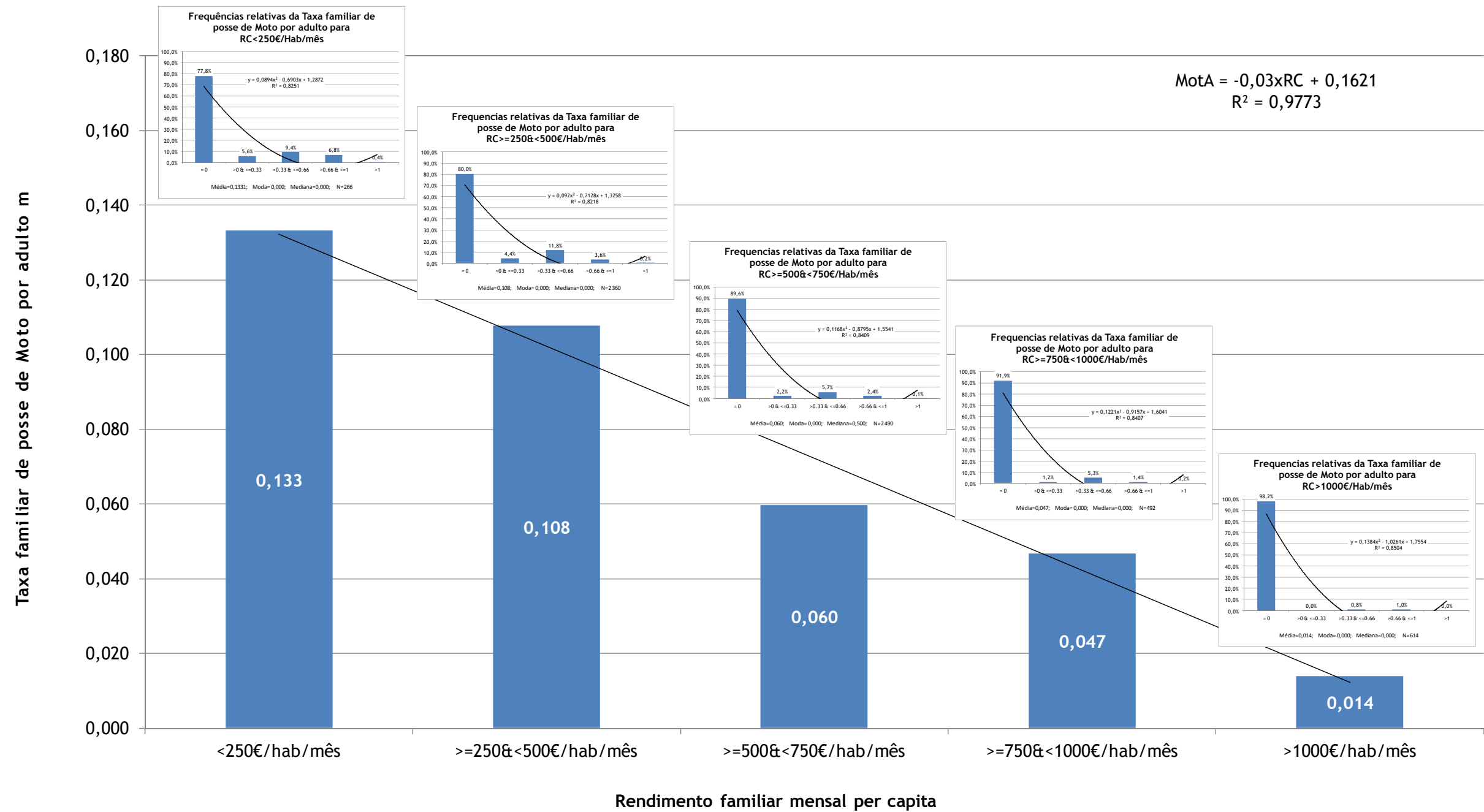


Figura VII. 12 - Evolução da taxa de motorização familiar de moto por adulto média com o rendimento mensal do agregado *per capita*

:

5.1. Relação Taxa de Motorização *versus* Rendimento Familiar

Atendendo à dificuldade de estimar evoluções futuras da taxa de motorização familiar, quer automóvel que de moto, procuraram-se eventuais relações entre as mesmas e o rendimento médio mensal *per capita* do agregado, cuja evolução se afigura de definição potencialmente mais simples. Para esta análise apenas se consideraram agregados familiares, uma vez que estas variáveis são comuns para todos os seus elementos. Para permitir a comparação entre os resultados para as taxas de motorização *per capita* e por adulto, apenas se considerou a amostra dos agregados com dados de ambas, reduzindo a amostra a 6 222 agregados familiares.

Salienta-se aqui novamente a dificuldade verificada na definição do rendimento médio mensal *per capita* do agregado, devido à falta de precisão dos reais valores, fundamentalmente relacionada com a relutância natural da maioria das pessoas em responder com exatidão a esta questão. Para uma maior probabilidade de obter dados é comum solicitar a resposta em gamas de valores, facto que retira precisão aos dados. No presente estudo, tal como já foi referido, optou-se por considerar o máximo de cada categoria e o valor de 2500€ para rendimentos mensais superiores a 2000€. Este facto deve ser devidamente ponderado na avaliação e utilização dos resultados obtidos.

Na figura VII.9 apresenta-se a evolução da taxa de motorização familiar *per capita* média, e na VII.10 a evolução da taxa de motorização familiar por adulto média, atendendo ao rendimento familiar mensal *per capita*. É possível constatar que o aumento do rendimento *per capita* propicia o aumento da taxa de motorização familiar com uma forma similar à função exponencial, assumindo-se que deverá ser limitada superiormente à semelhança de uma função logística. É também possível constatar que os valores médios da taxa de motorização *per capita* e por adulto são da mesma ordem de grandeza principalmente para as classes de rendimentos mais elevados.

Nas figuras VII.11 e VII.12, apresenta-se a evolução da taxa de motorização familiar de moto, quer *per capita* quer por adulto, atendendo também ao rendimento familiar mensal *per capita*. A evolução da taxa de motorização familiar de motos é inversa à de automóveis, uma vez que tem tendência a diminuir com o aumento da classe de rendimentos. Este comportamento sugere que provavelmente a tipologia de Moto existente nos agregados será maioritariamente de motociclos, mais baratos, e mais acessíveis a quem tem baixos rendimentos. A evolução da taxa de motorização *per capita* sugere um pico para a classe de 250 a 500€/mês/hab., coerente com o pressuposto do uso de Moto ser mais acentuado nas classes com baixos rendimentos mas suficientes para a aquisição do motociclo. A evolução da taxa de motorização de moto por adulto sugere um decréscimo tendencialmente linear, com valores de grandeza igual à taxa *per capita* a partir de 500€/mês/hab.

É de salientar nas taxas de motorização de moto a elevada percentagem de agregados sem moto, sempre acima dos 75%.

5.2. Relação Licença de Condução *versus* Sexo *versus* Rendimento Familiar

Uma outra relação difícil de avaliar no futuro é a percentagem de encartados na população. Se por um lado uma melhoria das condições sociais pode potenciar um maior anseio por obter uma licença de condução que possibilite um aumento da mobilidade pessoal, essas mesmas melhorias conjugadas com uma melhor formação/instrução poderá despertar uma maior consciência ambiental e diminuir a utilização dos modos menos sustentáveis. Além destes fatores económicos e de formação, poderá também haver alterações sociais no sentido de uma maior igualdade entre géneros.

Assim, no sentido de avaliar eventuais relações entre a percentagem de encartados com o aumento do rendimento familiar mensal *per capita*, analisou-se a sua distribuição na base de dados dos inquiridos de agregados com todos os adultos entrevistados ou que tenham realizado viagens, com idades entre os 20 e 65 anos (12 045 indivíduos). Optou-se igualmente pela distinção entre homens e mulheres na procura de eventuais situações diferenciadas. Os resultados desta análise são apresentados na figura VII.13.

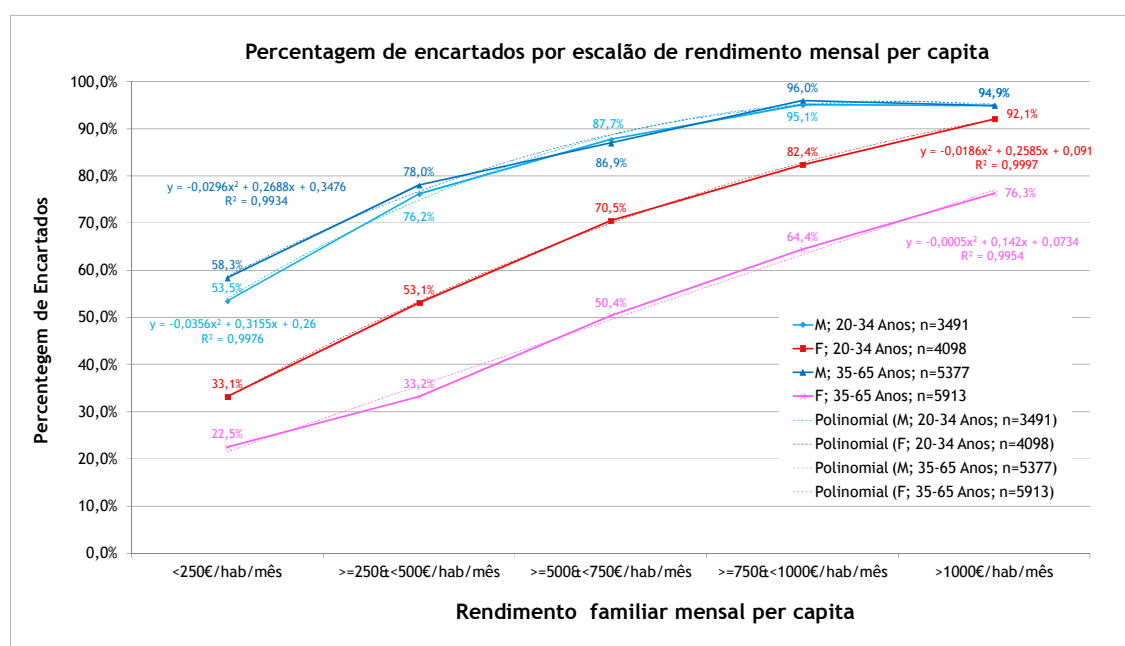


Figura VII. 13 - Evolução da percentagem de encartados por sexo com o rendimento familiar mensal *per capita*

Da figura VII.13 é possível constatar uma grande diferença de comportamento para homens e mulheres. Aparentemente não existe diferenciação de comportamento entre idades para os homens. Relativamente a estes verifica-se um aumento acentuado da percentagem de homens encartados mesmo com rendimentos entre 250 a 500€/mês/hab., e uma estabilização da mesma para rendimentos mensais *per capita* acima de 750€.

Nas mulheres verifica-se uma clara distinção de comportamentos entre as mais jovens e as mais velhas. As mulheres mais jovens apresentam uma tendência mais próxima dos homens mas com valores mais baixos, mas com um crescimento quase linear da percentagem de encartados com o rendimento. Apenas para rendimentos mensais *per capita* superiores a 1000€ a sua situação se mostra praticamente igual à dos homens.

As mulheres mais velhas apresentam uma percentagem de encartadas para todas as gamas de rendimentos sempre mais baixa que as mulheres mais jovens, com reduções de 11% a 20%, e uma relação também quase linear face ao rendimento. Este resultado sugere que estereótipos sociais relativamente à mulher se têm vindo a esbater mas continuam a existir. Note-se que nas idades mais avançadas mesmo entre as mulheres de situação económica mais favorável a percentagem que não tirou licença de condução é ainda significativa. Este facto pode dever-se a fatores sociais, esperando-se que as senhoras de determinados estratos sociais fossem conduzidas pelos maridos ou por motorista.

Atendendo a estes resultados, será prudente a assunção de uma evolução no sentido de uma maior utilização do automóvel pelas mulheres. A combinação da tendência verificada de aumento da percentagem de mulheres encartadas nas gerações futuras, aproximando-se da realidade masculina, com a possibilidade económica de aquisição de automóvel próprio e as alterações nas estruturas familiares tradicionais verificadas nas últimas décadas, tenderá a levar a uma muito maior igualdade de comportamentos entre ambos os sexos. Se do ponto de vista social esta igualdade é, obviamente, desejável, do ponto de vista de uma mobilidade sustentável, haverá a necessidade de implementar políticas que promovam a alteração dos hábitos dos homens (e eventualmente das mulheres no futuro) para uma aproximação ao comportamento atual das mulheres, evitando o inverso, cada vez menos sustentável.

6. Conclusões

Das simulações com os modelos definidos no Capítulo anterior foi possível constatar quais as características socioeconómicas mais relevantes para potenciar determinadas escolhas modais. Apenas restrições muito fortes ao uso do automóvel ou a existência de alternativas muito interessantes/convenientes (por exemplo grande proximidade dos destinos) poderá levar a repartições modais significativas para as alternativas mais sustentáveis. Segundo o modelo as situações limites de combinações favoráveis a um acentuado uso das alternativas ao automóvel são aparentemente extremamente improváveis, mas foi possível definir perfis

plausíveis com acentuadas utilizações das alternativas. Das simulações para estes perfis intermédios verificou-se alguma dificuldade de prever a utilização do modo Moto, apesar de haver uma variável explicativa que retrata o efeito identificado como resultante da existência de mota no agregado familiar. Tal deve-se, provavelmente, à grande especificidade de condições que levam à sua utilização (tornando a sua modelação em valores médios muito errónea).

Outra análise às previsões dos modelos definidos foi a avaliação da sensibilidade dos mesmos a variações isoladas de cada uma das variáveis. Foi possível verificar que a maioria das variáveis socioeconómicas tem um impacto moderado nas repartições modais e que esse impacto segue um comportamento tendencialmente linear. A variável caracterizadora de eventuais alterações nas políticas de ordenamento do território, distância entre centróides, revelou potencial impacto acentuado apenas para o modo Walk, com tendência não linear, e um impacto moderado nos restantes modos relevantes. Relativamente às variáveis sobre políticas de mobilidade verificou-se um impacto potencial modesto para quase todas as alternativas. Salienta-se o impacto da variável frequência diária média de Bus que não mostrou capacidade para influenciar qualquer alternativa. Este resultado poderá dever-se a uma deficiente captação da sua influência através da variável definida. Foi, no entanto, possível perceber tendências de uma influência positiva das variáveis relacionadas com os transportes coletivos, nomeadamente no uso do BUS.

Após esta análise de sensibilidade foram definidas alterações plausíveis das diferentes variáveis, coerentes com cenários futuros em termos de evolução socioeconómica e de possíveis políticas de mobilidade e ordenamento territorial sustentáveis. Estes cenários serviram de base para a realização de previsões com o modelo definido para a situação de inclusão de políticas de mobilidade (MNL12.16). Constatou-se que as alterações de cenários socioeconómicos revelaram mais impacto em relação à situação atual que as medidas de mobilidade para o modo BUS, cuja influência é muito reduzida. Verifica-se que os cenários otimistas sociais levam a uma redução do uso do Bus, enquanto cenários pessimistas sociais mantêm uma repartição semelhante à atual. Relativamente ao modo Walk, o modelo sugere um maior impacto das medidas de mobilidade e ordenamento territorial sustentáveis do que das condições socioeconómicas. A evolução das condições socioeconómicas sugere um ligeiro acréscimo na utilização deste modo para o cenário ambientalmente sustentável. Regista-se igualmente uma forte redução nas previsões da sua repartição modal com a manutenção das tendências atuais em termos de políticas de mobilidade. O modo mais sensível às alterações testadas nos cenários propostos foi, um pouco paradoxalmente, o Auto. Verificaram-se acréscimos acentuados de uso para o cenário otimista social e uma influência mais modesta para uma evolução mais pessimista. A manutenção de políticas de mobilidade tradicionais sugere a continuação do presente padrão de acentuado aumento do uso do Auto, verificando-

se no máximo a manutenção da repartição modal atual com a combinação de medidas de mobilidade e ordenamento sustentáveis aliadas a condições socioeconómicas desfavoráveis.

A formulação de possíveis cenários futuros revelou algumas tendências interessantes. No entanto, relembra-se que os resultados obtidos resultam de combinações específicas das variáveis em estudo, cujas evoluções nos diferentes cenários poderão ser contestadas e eventualmente alteradas. O estudo da evolução das variáveis para diferentes cenários futuros poderá revelar-se difícil no caso de variáveis cuja evolução não esteja bem documentada. No caso das variáveis socioeconómicas existem dados para a grande maioria delas, no entanto, para as variáveis mais relevantes como a taxa de motorização familiar e a percentagem de encartados, esses dados são mais escassos. No sentido de realizar estimativas credíveis para estas variáveis analisou-se a sua eventual relação com o rendimento *per capita*, cuja estimação tem mais informação disponível. Destas análises foi possível obter relações interessantes, com resultados coerentes. Verificou-se um crescimento com uma curva exponencial, com uma provável limitação máxima semelhante à função logística para a taxa média de motorização familiar *per capita*, bem como um crescimento quase linear da taxa média de motorização familiar por adulto com o rendimento *per capita*, para as gamas de rendimento analisadas. Ambas as análises revelaram ajustamentos muito bons. Em relação à taxa média de motorização familiar de moto *per capita*, verificou-se uma diminuição da mesma com o rendimento *per capita*, ajustada com uma expressão polinomial de 2º grau, enquanto para a taxa de motorização por adulto revela uma relação aproximadamente linear. Ambas as relações apresentam ajustamentos muito bons embora com significâncias limitadas devido ao pequeno número de pontos de base. Na possível aplicação a outros casos deve ter-se em conta a dificuldade e as simplificações assumidas na especificação do rendimento familiar mensal *per capita*, cujos dados de base foram recolhidos como gamas de valores.

O capítulo conclui com a análise de eventuais relações entre a percentagem de encartados com o rendimento *per capita*, considerando a idade e o sexo dos indivíduos. Esta análise mostra alguns resultados interessantes, úteis para ajudar a prever alterações dos padrões de escolha modal resultantes das alterações sociológicas que se vêm a verificar ao longo das últimas décadas e que tenderão a continuar. Foi possível constatar que o aumento dos rendimentos leva a uma maior percentagem de encartados, tanto dos homens como das mulheres. Em relação à evolução desta relação, verifica-se um aumento não linear acentuado para os homens logo a partir de rendimentos mais baixos com uma tendência assintótica perto dos 100%, enquanto a evolução para as mulheres é mais linear e diferenciada entre as mulheres mais velhas e as mais novas. Verifica-se que no grupo das mulheres mais velhas há uma tendência para haver uma menor percentagem de encartadas do que no grupo das mais novas para todos os escalões de rendimentos, mostrando as últimas uma situação mais próxima da que se observa nos homens. Em relação aos homens não se detetaram diferenças significativas com a idade. Todas as relações apresentaram ajustamentos muito bons com

relações polinomiais de 2º grau, embora com significâncias limitadas devido ao pequeno número de pontos de base.

A qualidade dos resultados sugere a expansão destas análises preliminares, nomeadamente através da consideração da influência eventual de outras variáveis.

Capítulo VIII

Conclusões

O estudo realizado teve como objetivo principal a avaliação das possíveis influências dos perfis socioeconómico dos cidadãos, bem como das características do território e dos sistemas de transporte, no processo de tomada de decisão na escolha do modo a utilizar em viagens diárias, nomeadamente em áreas urbanas. O conhecimento destas influências poderá auxiliar o entendimento dos potenciais impactos de diferentes estratégias de gestão da mobilidade na alteração para hábitos mais sustentáveis.

Atendendo às tendências internacionais em termos de políticas de mobilidade sustentável, é necessário encontrar metodologias para analisar de forma o mais possível objetiva e quantificada a influência que a qualidade dos sistemas de transportes alternativos ao automóvel particular, especialmente os transportes coletivos, tem ou poderá ter, na escolha do modo de transporte em viagens diárias. Simultaneamente justifica-se também avaliar o impacto que medidas que restrinjam o livre uso do automóvel particular podem ter nas opções modais. Apesar de existirem estudos internacionais que incluem separadamente partes destas análises, não se conhecem análises aprofundadas da sua combinação e os seus resultados são específicos para os cenários analisados, não sendo de aplicação generalizada direta. Verifica-se igualmente uma escassez de estudos mais aprofundados sobre a melhor forma de especificar diferentes variáveis para que transfiram informação fidedigna aos modelos sem aumentar demasiadamente a sua complexidade e especificidade. Com o presente estudo pretendeu-se, assim, colmatar estas lacunas e contribuir para quantificar os possíveis impactos de estratégias de incentivo ao uso do transportes coletivos e de restrição ao livre uso do automóvel particular, no processo de tomada de decisão. Neste estudo consideraram-se os aspetos relacionados com as características pessoais/sociais de cada indivíduo, com as características de cada viagem conjugando-as com características do sistema de mobilidade urbana disponível.

Para essa quantificação estudou-se uma possível formulação matemática optando-se por modelos de escolha discreta que, com uma base teórica consecutivamente consolidada, se apresentam como ferramentas potentes para a modelação das escolhas de indivíduos. Estes modelos, além de se basearem em teorias de comportamento, permitem modelar escolhas com base em variáveis contínuas mas também discretas, isto é, selecionadas de um conjunto finito de opções, provenientes dos tradicionais inquéritos de preferências reveladas e/ou declaradas. Dos diversos modelos de escolha discreta desenvolvidos desde os anos 70, o

modelo mais conhecido e vastamente utilizado é o Logit Multinomial. Apesar de contestado para problemas especialmente complexos devido às limitações inerentes à sua especificação, o seu uso para modelações exploratórias é justificado atendendo à facilidade de cálculo e de interpretação de resultados. Geralmente é aconselhada a realização de extensas análises exploratórias com este tipo de modelos, de forma a definir quais as variáveis, e qual a sua melhor especificação, a integrarem a formulação matemática a estimar, antes de, eventualmente, se passar a modelos mais complexos que mitiguem as limitações dos Logit Multinomiais.

O conhecimento dos diferentes impactos da cada uma das variáveis estudadas poderá servir de base para a tomada de decisões de forma mais fundamentada, potenciando a eficiência e eficácia das decisões políticas para os diferentes subsistemas que compõem o complexo sistema urbano.

O caso de estudo apresentado é, assim, uma vasta aplicação de modelos Logit Multinomiais para a análise das variáveis explicativas mais relevantes e a melhor forma de as incluir atendendo às suas diferentes especificações. Esta aplicação incidiu geograficamente em parte da AMP e utilizou dados pré-existentes recolhidos no inquérito à mobilidade realizado pelo INE no ano 2000. Com estes dados e através da incorporação de outros considerados relevantes relativamente à oferta de transportes coletivos na mesma área e das diferentes tipologias de oferta de estacionamento existentes na área urbana do Porto, organizou-se uma vasta base de dados para a estimação dos modelos recorrendo ao *software* de análise econométrica *NLogit 4*[®].

Numa primeira fase foram estimados modelos Logit multinomiais com a utilização de cada uma das variáveis incluídas isoladamente. Este procedimento foi repetido para as suas possíveis diferentes especificações. Para as variáveis que nos modelos isolados apresentaram resultados menos espectáveis, foram também analisados modelos combinados para a avaliação de possíveis situações simbióticas ou de correlações. Assim, foi possível identificar um conjunto alargado de variáveis socioeconómicas, variáveis políticas, eventuais correlações e/ou relações simbióticas, e a respetiva especificação, com potencial para fornecerem informação relevante ao modelo.

Na fase seguinte foram estimados modelos com as combinações das variáveis consideradas relevantes. Atendendo à metodologia recomendada por Koppelman e Bhat (2006), foi realizado um processo iterativo com o consecutivo teste de diferentes combinações de variáveis de forma a melhorar o ajustamento global do modelo. A partir desta análise foi possível estimar as combinações de variáveis mais promissoras observando as melhorias do ajustamento global e a coerência de sinais e comportamento dos diferentes parâmetros estimados. Unidos dos modelos que melhores resultados apresentaram realizaram-se

previsões cuja validade foi avaliada não só para o cenário explicitado pela base de dados utilizada, mas também para hipotéticos cenários futuros.

A possibilidade de existência de interdependências entre algumas das variáveis sociais selecionadas levou a uma análise qualitativa dessas relações na base de dados, averiguando eventuais tendências não captadas pelo modelo. Foi possível constatar alguns comportamentos relativamente homogêneos para determinados grupos sociais, sugerindo um estudo final para analisar a viabilidade de uma especificação alternativa de combinação dessas variáveis sociais.

Do vasto leque de análises realizadas foi possível confirmar/fundamentar cientificamente algumas tendências esperadas bem como constatar que determinadas assunções geralmente consideradas internacionalmente nestas modelações, aparentemente podem ser facilmente reajustadas para obter resultados mais fidedignos.

Do estudo foi possível verificar que a utilização de levantamentos pré-existentes, não especificados para determinado objetivo mas para outros mais abrangentes, levou à existência de algumas inconsistências que por vezes limitaram o pleno aproveitamento de todo o potencial de informação disponível. Atendendo ao objetivo do levantamento do INE, a análise da mobilidade na AMP, houve alguns aspetos metodológicos na definição das variáveis levantadas que deveriam ser revistos em trabalhos futuros. Dos problemas verificados destacam-se:

- A impossibilidade de definir em concreto o local de origem e de destino de viagens intrazonais (dentro da mesma zona de tráfego) quando as zonas são muito extensas, aconselhando claramente a utilização de zonamentos mais finos caracterizados em termos de homogeneidade de uso do solo, características dos sistemas de transportes e/ou características socioeconómicas da população residente;
- Em relação ao modo de transporte utilizado os inquéritos deverão, sempre que possível, avaliar também quais as alternativas disponíveis ao modo utilizado e as suas características operacionais apreendidas pelo inquirido.
- Atendendo ao impacto claramente distinto entre a utilização do automóvel como condutor ou como passageiros, deverá haver uma clara distinção entre ambos na definição do modo de transporte;
- Nos motivos da viagem as características operacionais de viagens para compras ou obtenção de serviços financeiros ou institucionais justificam a sua clara diferenciação;
- Na definição da tipologia de variáveis a utilizar, para uma análise mais coerente, é recomendável, sempre que possível, a escolha de uma especificação contínua para o

rendimento mensal do agregado, ou, pelo menos, a adoção de gamas de valores menos amplas, nomeadamente para os rendimentos superiores;

- Na aquisição de informação sobre o custo do estacionamento e da viagem em transportes coletivos, recomenda-se a inquirição desses custos (ou estimativa) efetivamente verificados para cada uma das viagens/atividades em análise, e não um valor total mensal/diário despendido pelo agregado.

Salientam-se ainda as dificuldades verificadas na obtenção de dados sobre as redes de transportes coletivos, nomeadamente as interurbanas e concessionadas a privados, bem com na caracterização da oferta de estacionamento fora do concelho do Porto. Estas dificuldades reforçam a importância de definir um processo de recolha sistemático e detalhado desta informação para sistemas de transportes complexos.

Apesar destas dificuldades foi possível definir um conjunto alargado de variáveis com potencial para introduzirem informação relevante na modelação das escolhas modais. Parte das variáveis consideradas foram baseadas em estudos de escolhas modais aplicados aos transportes apresentados na bibliografia internacional, sendo outras comuns a modelos genéricos de escolha discreta.

Atendendo à informação disponível considerou-se ainda que outras variáveis apresentavam potencial para puderem acrescentar informação relevante que justificasse algumas escolhas modais, nomeadamente o número total de viagens diárias, a duração da atividade, a disponibilidade diária de Moto no agregado, bem como adaptações de variáveis normalmente consideradas, como a razão entre tempos de viagem e variáveis relacionadas com a densidade de oferta de estacionamento.

Das modelações preliminares com variáveis isoladas, nas suas diferentes especificações, foi possível verificar a existência de um número alargado de variáveis com potencial capacidade explicativa das escolhas modais. Identificaram-se como particularmente relevantes: a taxa de motorização familiar, a existência de licença de condução e a distância mais curta entre centróides. São de salientar os resultados obtidos para a combinação destas variáveis que contribuem com informação que justifica cerca de 50% do ajustamento máximo alcançado, destacando-se a importância revelada pela combinação taxa de motorização *versus* licença de condução, cuja disponibilidade conjugada torna a utilização do automóvel quase plena.

As restantes variáveis menos proeminentes, mas cujos resultados sugerem que podem introduzir informação relevante aos modelos de escolha modal, são o rendimento mensal do agregado familiar, o sexo, o grupo etário, o nível de instrução, o número total de viagens diárias, a disponibilidade diária de moto no agregado, o motivo da viagem, o logaritmo do tempo médio de viagem, a duração da estadia/atividade, a densidade de estacionamentos

relativamente à procura, a percentagem de estacionamento pago, a razão de tempos médios de viagem entre Auto e Bus e a frequência média horária diária. Verificou-se que os parâmetros das variáveis sobre o nível de instrução revelavam comportamentos não expectáveis, nomeadamente para o nível analfabeto com uma exagerada valorização do modo Auto. A análise do seu comportamento combinado com a variável grupo etário revelou uma elevada correlação deste nível de instrução com o grupo etário dos 0-5 anos (transportados de Auto principalmente pelas mães).

Das análises foi também possível constatar que, na sua maioria, estas variáveis terão a sua influência melhor representada através das suas especificações definidas por combinações de variáveis binárias, apesar de tal poder implicar alguns riscos de menor generalização automática dos resultados. Estes resultados contrariam a tendência verificada na maioria dos estudos internacionais que utilizam preferencialmente a especificação contínua e/ou discreta. Esta opção generalizada aparentemente não tem fundamentação, pelo menos indiscriminadamente, uma vez que há variáveis que atendendo à sua natureza intrínseca dificilmente poderão ter a sua influência nas escolhas modais representada por uma aproximação linear. Exemplos claros deste tipo de variáveis são o grupo etário, o nível de instrução e o motivo da viagem.

Assume-se, ainda assim, que para outras variáveis, a simplicidade da especificação contínua/discreta possa continuar a justificar a sua utilização atendendo à reduzida melhoria de representatividade da especificação com combinação de binária. Das variáveis socioeconómicas e relacionadas com a organização das atividades analisadas apenas a duração média da viagem (genérica) e o número de viagens diárias revelaram melhores resultados com a especificação contínua. Verificou-se ainda que alguns conjuntos de variáveis, nomeadamente o escalão etário e o nível de instrução, revelavam a correção de incoerências no comportamento dos seus parâmetros em determinadas gamas, verificadas nas análises isoladas, com a sua combinação demonstrando um efeito tendencialmente simbiótico. A análise da melhor especificação para a variável taxa de motorização familiar revelou que a mesma aparentemente seria melhor representada pela combinação de binárias. Este resultado deveu-se provavelmente ao grande impacto revelado pela inexistência de automóvel no agregado, exibindo as restantes categorias maior viabilidade à sua especificação através de uma variável contínua.

Das variáveis relacionadas com políticas de mobilidade, constatou-se que a velocidade comercial equivalente, apesar de um elevado ajustamento global do modelo com especificação contínua, revelou que a informação que introduz no modelo é muito semelhante à introduzida pela distância mais curta entre centróides, variável primária utilizada para a definição da velocidade. Este facto foi comprovado pela elevada correlação verificada entre as mesmas. Esta constatação levou à sua eliminação como variável explicativa.

Em relação às variáveis relacionadas com a oferta de estacionamento, a dificuldade em definir a sua melhor especificação levou à opção de conciliar as suas especificações como combinação de binárias em separado para cada variável, uma vez que esta opção apresentou um dos melhores ajustamentos entre as diferentes especificações analisadas, com a vantagem de uma fácil interpretação dos seus parâmetros.

Complementarmente às análises anteriores avaliou-se, ainda, a estabilidade do modo de transporte ao longo das atividades do dia. Esta análise revelou que a maior parte das viagens ao longo do dia utilizava o modo escolhido na primeira viagem do dia. Foi possível constatar que as condições sociais de base dos utentes influenciam o modo escolhido para realizar as viagens, mas que a estabilidade verificada ao longo do dia é comum a todas as opções. Verificou-se ainda que um dos poucos fatores que se mostrou verdadeiramente capaz de “quebrar” esta fidelidade foi a distância de viagem, já que para pequenas distâncias a atratividade do modo Pedonal se revelou determinante. Estes resultados sugerem uma fidelidade, voluntária ou imposta, ao modo escolhido à saída de casa que indicia a importância de poder/conseguir influenciar as escolhas logo a partir da primeira viagem.

A constatação da relevância das variáveis licença de condução, taxa de motorização e distância de viagem, levou à decisão de as incluir numa modelação base a integrar qualquer modelo com combinação das restantes variáveis. Partindo deste modelo base realizou-se o processo iterativo de combinação das restantes variáveis analisando-se os respetivos ganhos de ajustamento global e coerência de sinais e comportamentos. Dos resultados do modelo base com a combinação das restantes variáveis consideradas relevantes nas modelações preliminares obtiveram-se duas modelações complementares:

- Modelo sem variáveis sobre políticas de mobilidade - MNL7.57 ($p\text{-}R^2 = 0,445$);
- Modelo com variáveis sobre políticas de mobilidade - MNL12.16BWA ($p\text{-}R^2 = 0,522$).

Esta distinção justificou-se fundamentalmente pela clara preponderância do impacto revelado pelas variáveis socioeconómicas e das variáveis relacionadas com a organização das atividades ao longo do dia, relativamente ao impacto das medidas de políticas de mobilidade. A existência de dados sobre as políticas de mobilidade apenas para as viagens com atração no concelho do Porto justificou a limitação da modelação para essas zonas às duas alternativas mais relevantes ao automóvel (Bus, Walk e Auto - BWA). Esta limitação implicou que o modelo com variáveis sobre políticas de mobilidade apenas possa ser utilizado para viagens atraídas para zonas urbanas centrais, onde essas medidas geralmente são mais comuns e incisivas, e geralmente realizadas pelos três modos mais relevantes. O modelo sem variáveis sobre políticas de mobilidade tem aplicação alargada também às viagens entre zonas suburbanas e interurbanas, onde as políticas de mobilidade analisadas são mais incipientes, aplicando-se a todos os modos selecionados.

Embora teoricamente relevantes, os resultados dos modelos mostraram que as medidas de política de mobilidade analisadas, apesar de apresentarem parâmetros com comportamento coerente atendendo à teoria comportamental, particularmente para o modo Bus, não revelaram impactos muito significativos nas escolhas modais comparativamente às gamas de variabilidade estudadas para estas variáveis. Estes resultados podem ser consequência de diversos fatores:

- Um desconhecimento efetivo dos cidadãos dos fatores analisados conjugado com uma falha na sua percepção em contextos reais;
- Uma efetiva reduzida influência das mesmas nas escolhas modais diárias dos cidadãos inquiridos quer consciente, quer inconscientemente;
- Pela acentuada redução da amostra devido à dificuldade verificada na recolha de informação sobre políticas de mobilidade;
- Devido à deficiente capacidade das variáveis selecionadas para representar o impacto das políticas de mobilidade ou a dificuldades na sua correta quantificação.

Os resultados para o modelo sem variáveis sobre políticas de mobilidade MNL7.57 sugerem que as variáveis fundamentais foram a Licença de condução (Lc), a taxa de motorização familiar (NCi), a distância mais curta entre centróides, principalmente para o modo pedonal (D2), e, devido à sua estabilidade ao longo das modelações, o sexo do inquirido (Sx). Como variáveis complementares, que introduzem alguma informação relevante mas com menor impacto nas escolhas, os resultados sugerem o nível de instrução (Ini) e o grupo etário (Idi), geralmente em combinação, a disponibilidade diária de Moto (Mot), o número total de viagens diárias (nV) e o motivo da viagem.

Todas as variáveis apresentaram parâmetros com comportamentos e sinais coerentes ou pelo menos plausíveis para determinados modos. A variável que maior impacto revelou foi a (in)disponibilidade de automóvel no agregado, seguida da licença de condução. Foi possível constatar que apesar de se ter verificado um elevado impacto de certas idades nas utilidades relativas das alternativas ao automóvel, que o escalão etário apresentou uma grande variedade de comportamentos. O nível de instrução atendendo à correlação com a idade, apresentou comportamentos coerentes, com uma tendência de maior valorização das alternativas ao automóvel pelos níveis de instrução superiores. A variável sexo apresentou valores relativamente estáveis ao longo das iterações. Em relação à variável distância mais curta, os resultados sugerem a introdução de uma variável binária caracterizadora das viagens a distâncias inferiores a 1 km apenas para a alternativa Walk. Na possibilidade de diferenciação das funções de utilidades dos vários modos, recomenda-se a inclusão da variável binária DX1 para distinguir essas viagens até 1 km apenas na utilidade do modo Walk.

O ajustamento global do modelo combinado MNL7.57 de $p-R^2 = 0,445$, ou seja, muito significativo, demonstrou uma capacidade muito elevada para explicação das escolhas modais a partir essencialmente de variáveis socioeconómicas, embora também com recurso a uma variável representativa da distribuição geográfica dos polos de GA de viagens.

Os resultados para o modelo com variáveis sobre políticas de mobilidade e apenas para três modos de transporte (MNL12.16BWA) sugeriram que as variáveis fundamentais continuavam a ser a Licença de condução (Lc), a taxa de motorização familiar (NCi), a distância mais curta entre centróides (D2), principalmente para o modo pedonal, e o sexo do inquirido (Sx). Como variáveis complementares, que introduzem alguma informação relevante mas com menor impacto nas escolhas, os resultados continuam a sugerir o nível de instrução (Ini) e o grupo etário (Idi), geralmente em combinação, o número total de viagens diárias (nV) e o motivo da viagem. As variáveis relacionadas com políticas de mobilidade nomeadamente com a filosofia de dimensionamento e regulação do acesso aos estacionamento, atendendo aos resultados, aparentemente são menos relevantes, uma vez que não melhoraram significativamente o ajustamento, apresentando, no entanto, valores totalmente coerentes para o modo Bus. Já as variáveis relacionadas com os serviços de transportes coletivos revelaram um impacto muito reduzido no ajustamento global, mas com valores também coerentes para o modo Bus.

O ajustamento global do modelo MNL12.16BWA proposto foi de $p-R^2 = 0,522$, correspondendo a um coeficiente de regressão linear equivalente de cerca de 0,95, ou seja, muito elevado. Note-se que parte significativa do acréscimo do $p-R^2$ deste modelo, face ao do modelo combinado anterior, se deveu essencialmente ao facto de se concentrar na explicação de escolhas entre apenas os três modos mais relevantes.

No sentido de avaliar a coerência e relevância dos resultados alcançados nas análises combinadas, compararam-se os resultados obtidos em ambos os modelos com os verificados nos estudos internacionais de referência.

As variáveis nível de instrução, número total de viagens diárias, duração da estadia na atividade de destino, oferta de estacionamento, razão entre tempos de viagem e disponibilidade diária de Moto, que não tinham, mesmo em diferentes especificações e distintas unidades e realidades socioculturais, sido identificadas como relevantes noutros estudos internacionais, foram-no no presente estudo. Os seus resultados revelaram-se estatisticamente significativos, com tendências coerentes e uma considerável capacidade explicativa quando combinadas.

Já para as restantes variáveis, comparando os sinais, grandezas e comportamento dos parâmetros foi possível constatar que a maioria revelou tendências similares ou enquadráveis com as verificadas internacionalmente.

Uma constatação relativamente estável ao longo das modelações foi o aparente comportamento oposto do modo Outros Bus relativamente ao modo Bus. Estes resultados, menos esperados, podem provavelmente dever-se à sua difícil caracterização resultante de amostras muito reduzidas, ou devido à sua natureza operacional relativamente variada, com características algo diferentes face aos TC regulares.

Para a avaliação da validade dos modelos obtidos foi estimada a repartição modal com os dados extrapolados que revelou previsões muito semelhantes às efetivamente verificadas nas bases de dados. Apenas para o modo Moto se verificaram discrepâncias indiciadoras de alguma dificuldade de explicitação comportamental do modelo para esta alternativa. Esta constatação pode dever-se à especificidade das condições que potenciam o uso deste modo, que provavelmente só poderão ser ultrapassadas mediante a diferenciação das funções de utilidade para cada modo de transporte distinto.

Além dos modelos apresentados, foram ainda realizadas análises adicionais para avaliar a existência de eventuais tendências não passíveis de ser reveladas pelos mesmos, como sejam interdependências entre conjuntos de variáveis ao nível do seu impacto modal. Estas análises centraram-se sobre eventuais relações entre as principais variáveis sociais (taxa de motorização familiar, licença de condução, sexo e idade), definidoras de grupos com comportamentos relativamente homogêneos. Foi possível verificar que:

- Existe uma certa homogeneidade de comportamentos para as mulheres com licença de condução e com mais de 20 anos e as crianças até aos 9 anos;
- As crianças até aos 9 anos, aparecem condicionadas às opções modais dos pais e, particularmente, da mãe;
- Para os jovens até aos 20 anos, aparentemente não há diferenciação de comportamento entre sexos;
- A falta de licença de condução parece levar a comportamentos muito semelhantes entre homens e mulheres em relação ao automóvel;
- Mulheres com licença de condução utilizam menos o automóvel que os homens. Apenas com taxas de motorização familiares superiores a 1 veíc./adulto, se verifica uma aproximação do uso do automóvel entre homens e mulheres;
- Aparentemente, as mulheres utilizam bastante menos o veículo do agregado quando este é único;
- As mulheres mais jovens (20-34 anos) parecem mostrar tendência para utilizar mais o automóvel que as mulheres mais velhas (35-65 anos), aproximando-se do comportamento dos homens;

- Em termos de alternativa ao automóvel, a opção principal para as mulheres sem licença de condução parece ser o Bus enquanto para os homens nas mesmas condições a Moto ganha uma importância significativa;
- Parece verificar-se uma certa homogeneidade de comportamentos para os homens com licença de condução e com mais de 20 anos;
- Verifica-se também uma certa homogeneidade de comportamentos para jovens dos 10 aos 19 anos com licença de condução e para homens e mulheres com mais de 10 anos e sem licença de condução;

Em geral constatou-se que, excetuando o comportamento dos homens com mais de 20 anos e com licença de condução, para os restantes grupos a influência do aumento da taxa de motorização familiar é relativamente linear, com um impacto mais preponderante (declives mais acentuados) para taxas de motorização mais baixas. Apenas o grupo dos homens com mais de 20 anos e com licença de condução apresenta um comportamento claramente não linear com uma redução drástica da utilidade das alternativas com a disponibilidade de um automóvel no agregado. Na bibliografia internacional consultada não foi possível encontrar quaisquer referências a tendências semelhantes que permitam balizar os resultados verificados. No entanto, parece confirmar-se, que em modelações mais detalhadas se deverá seriamente considerar formulações que permitam incorporar efeitos de correlação entre diferentes variáveis, no presente caso particularmente nas socioeconómicas.

Atendendo à globalidade dos resultados, verificaram-se algumas tendências pertinentes a considerar em processos de definição de políticas de mobilidade sustentável. Foi possível constatar que os modelos revelaram uma maior preponderância no processo de decisão das variáveis socioeconómicas e das variáveis relacionadas com a organização das atividades diárias, relativamente às variáveis relacionadas com os sistemas de transporte alternativos. Apesar de este facto poder ser imputado à deficiente definição ou captação da influência das mesmas, ou ser uma consequência da tradicional qualidade dos transportes coletivos disponibilizados à população, a diferença de grandeza dos parâmetros e a coerência de comportamentos e sinais sugerem alguma confiança nos resultados. Estes resultados, aliados à modesta influência relevada pelas variáveis relacionadas com os estacionamento, podem interpretar-se como uma clara manifestação da atual quase irrelevância da apreensão da qualidade dos serviços de TC no processo de escolha modal, sugerindo que só intervenções muito mais sistémicas e sistemáticas de promoção dos modos alternativos e de condicionamento do modo auto, poderão ser um impacto significativo.

Em relação às condições socioeconómicas, foi possível confirmar que o aumento do rendimento médio do agregado familiar apresenta tendência para um aumento da taxa de motorização familiar bem como da percentagem de encartados, tendo sido possível identificar relações analíticas promissoras. Esta combinação potencia o uso do automóvel de

forma tão acentuada que dificilmente, nas condições atuais, se vislumbra a possibilidade de alterações modais. Com exceção das viagens de curta distância, particularmente as menores que 1km, para as restantes situações apenas grandes restrições à utilização do automóvel poderão levar ao uso de alternativas. Estes resultados sugerem a grande dificuldade em tornar as alternativas ao automóvel mais atrativas, sem as convenientes medidas complementares de restrição ao livre uso do automóvel.

Foi ainda possível constatar que contrariamente ao assumido em muitos estudos internacionais, uma variável indireta potencialmente mais credível para introduzir as condições económicas do agregado na modelação do processo de escolha modal é a taxa de motorização familiar *per capita* ou por adulto. Este resultado pode dever-se à maior dificuldade de definir com precisão o real rendimento familiar comparativamente à taxa de motorização, associado à natural correlação entre ambas, dado que rendimentos familiares maiores tendem a originar taxas de motorização superiores.

Em relação ao nível de instrução, não foi possível encontrar tendências claras sobre a sua influência nas escolhas modais. Da análise do comportamento dos seus parâmetros estimados pelos modelos, níveis de instrução mais elevados parecem tender a valorizar mais a alternativa Bus que os níveis mais baixos, cuja capacidade financeira deveria tornar mais atrativa. Estes resultados sugerem, talvez, uma maior consciência ambiental com o aumento do nível de instrução, tendência a confirmar em estudos futuros.

Na avaliação do impacto da organização das atividades as variáveis que maior relevância apresentaram foram:

- Distância entre atividades: Verificou-se que para distâncias até 1km o modo pedonal, devido à sua enorme conveniência para este tipo de deslocações, apresenta uma hegemonia quase total. Esta relevância, embora atenuada, estende-se até aos 2km. A partir desta distância não se verificam grandes diferenças entre as atratividades das restantes alternativas. Note-se, no entanto, que os limites 1 e 2 quilómetros devem ser assumidos como aproximações devido à forma como as distâncias entre OD foram estimadas no presente estudo;
- Motivo da viagem: Verificou-se uma tendência para valorizar as alternativas ao automóvel, principalmente para as atividades de maior duração, trabalho e escola, provavelmente pela menor necessidade de mobilidade que as mesmas implicam;
- Número de atividades diárias: O seu aumento e correspondente maior necessidade de mobilidade, geralmente associado aos motivos de menor duração, como seria de esperar, apresentou tendência para diminuir a utilidade das alternativas ao automóvel, com exceção do modo Moto, esperando-se que, se fosse possível isolar o modo automóvel/condutor, este efeito fosse ainda mais evidente. Para o modo Moto o

modelo, provavelmente devido à mobilidade e conveniência semelhante ao automóvel, apresenta um impacto muito reduzido e sem significância estatística relativamente ao modo Auto.

Estes resultados sugerem um potencial de captação para as alternativas ao automóvel de uma parte significativa das viagens que decorrem nas áreas urbanas, nomeadamente as realizadas para viajar para o emprego e para a escola, cuja realização diária para o mesmo local poderá garantir um conhecimento das possíveis alternativas existentes ao automóvel e das restrições impostas nos habituais locais de destino, potenciando/viabilizando um maior uso das mesmas. A definição de estratégias concertadas que motivem a alteração destas viagens para modos alternativos mais sustentáveis logo a partir do domicílio, ou pelo menos antes de entrar nos centros urbanos, permitirá libertar parte da capacidade de tráfego e da oferta de estacionamento para viagens motivadas por atividades menos prolongadas mas também vitais para o desenvolvimento social e económico das áreas urbanas.

Numa avaliação global e prospetiva das implicações ao nível da definição eficaz de políticas sustentáveis de mobilidade urbana, o impacto potencial das diferentes variáveis ao nível da repartição modal revelada pelos modelos combinados sugere que apesar de necessárias e desejáveis, melhorias operacionais dos transportes alternativos, nomeadamente os coletivos, isoladamente não terão impactos significativos. Os resultados sugerem que apenas medidas abrangendo todo o sistema urbano e aplicadas de forma bem ponderada nomeadamente quanto a limitar o livre uso do automóvel, poderão alterar hábitos de mobilidade “viciados” no automóvel. Apesar de um impacto reduzido, os modelos apontam para que as políticas de restrição ao estacionamento tenham potencial para auxiliar nessa mudança, corroborando tendências verificadas em estudos internacionais.

As estratégias de incentivo ao uso de alternativas ao automóvel deverão assim, concertar o ordenamento do territorial com as políticas de mobilidade sustentável. A expansão desordenada, em “mancha de óleo”, e a baixa densidade das áreas urbanas inviabiliza sistemas de transportes coletivos eficientes e eficazes, havendo a necessidade de no planeamento do território se passar de um paradigma de expansão para novas áreas para uma compactação das existentes. Esta mudança de organização espacial deverá simultaneamente ser coordenada com os sistemas de transportes, permitindo densidades de usos maiores em áreas com sistemas de transportes já implementados ou onde se desejam implementar, dificultando o uso do automóvel nas mesmas. Haverá, no entanto, zonas em que essas restrições deverão ser relaxadas de forma a não impedir a mobilidade de parte da população, em parte garantida pelo automóvel.

Os modelos definidos, apesar dos problemas referidos na aquisição dos dados, revelaram tendências coerentes e motivadoras, com um aumento da atratividade da alternativa BUS com a diminuição da oferta de lugares de estacionamento disponíveis e o aumento da

percentagem de lugares pagos. Estes resultados podem, no entanto, ser enganadores uma vez que é nas áreas consolidadas que as redes de transportes coletivos são mais densas e também onde há maiores rácios procura potencial/oferta de lugares de estacionamento, levando a uma maior taxação dos mesmos para motivar a sua rotatividade. A implementação de medidas de restrição ao livre uso do automóvel abre uma janela de oportunidade para tornar os transportes coletivos mais competitivos mediante a disponibilização de um sistema que satisfaça as necessidades de mobilidade dos cidadãos e se aproxime da conveniência permitida (ou imposta) ao automóvel. Para que haja uma possibilidade de mudança no paradigma dos transportes em áreas urbanas, os operadores e/ou municípios terão de continuamente adaptar os seus sistemas atendendo às expectativas e anseios dos seus atuais e potenciais futuros clientes.

Dos resultados do presente estudo, foi também possível constatar determinadas tendências que deverão ser devidamente ponderadas na definição de políticas de mobilidade sustentáveis futuras. Apesar de atualmente ainda se verificar alguma diferenciação entre o comportamento de homens e de mulheres, em parte devido à tradicional atitude machista de ser o homem a transportar a família, a emancipação da mulher, o acesso à instrução e a empregos com maiores responsabilidades e remunerações, levaram a sua independência financeira e emocional, passando a ter estilos de vida indiferenciados dos homens e com tendência a semelhanças comportamentais. Esta tendência, aliada ao previsível aumento do acesso ao automóvel, poderá levar a uma ampliação dos problemas de tráfego atualmente existentes. Outro grupo social com um comportamento atual mais sustentável, são os adolescentes, com idade suficiente para serem autónomos mas ainda novos demais para poderem utilizarem individualmente o automóvel. Ambos os grupos, atendendo ao provável conhecimento que têm dos sistemas de transporte alternativos, serão os grupos cujas expectativas deverão ser mais “acarinhas” e apoiadas de forma a garantir que no futuro, mediante a disponibilização de automóvel, as suas opções se mantenham. No sentido contrário, com comportamentos menos sustentáveis, encontram-se os homens com idades entre os 20 e os 65 anos, cuja pouca apetência em relação aos transportes coletivos deve ser convenientemente estudada, de forma a avaliar quais as medidas com maior potencial para tornar o automóvel menos atrativo perante as alternativas. Este grupo é ainda mais importante se se atender à evolução da pirâmide etária nacional, cuja inversão perspetiva um potencial grupo de clientes a médio prazo cuja opinião em relação aos transportes coletivos interessa modificar/melhorar de forma progressiva.

Outro vetor de atuação das políticas de mobilidade sustentável deverá ser a implementação de sistemas alternativos inovadores. Apesar de existirem noutros países com algum sucesso, a introdução de sistemas inovadores em Portugal tem-se mostrado algo insípida. Existem exemplos, alguns com sucesso, outros não, em diversas cidades nacionais de implementação de sistemas de *Park&Ride*, de ciclovias, de linhas de *MiniBus*, de autocarros a Hidrogénio, vias

bus reservadas, entre outros. No entanto, sistemas alternativos comuns em muitos países europeus e americanos, de custos relativamente baixos, apresentam igualmente grande potencial para atrair os condutores nacionais. Como exemplo desses sistemas podem referir-se os sistemas de *Car-pooling*, *Car-sharing*, *Transport-on-Demand* (TOD), vias *High-occupancy-Vehicle* (HOV), ou transporte coletivos com possibilidade do embarque de bicicletas ou ciclomotores. Estes sistemas aliam a potencial redução do número de veículos a circular, promovendo a utilização de transportes “coletivos” particulares, com a disponibilização de alternativas com mobilidade e conveniência similar ao automóvel. Em função da importância revelada pela variável posse de carro, sugere-se que particular atenção deva ser dada às soluções do tipo “*Car-sharing*”, que ao evitar a “posse” de automóvel próprio, permita quebrar significativamente este processo de dependência que a posse de automóvel parece induzir.

Considerando os resultados obtidos no presente estudo, vislumbram-se linhas de investigação futuras potencialmente relevantes. No sentido de melhorar os modelos sugeridos, e atendendo aos seus resultados e aos das análises qualitativas realizadas, sugere-se a evolução natural dos mesmos com a aplicação das variáveis selecionadas a modelos mais complexos. O modelo de escolha discreta de referência atualmente para modelação clássica dos quatro passos é o Logit Misto. No entanto, e tal como foi possível verificar, existem variáveis como a distância mais curta entre zonas de origem e de destino, que apenas são relevantes para determinadas alternativas (no caso para o Walk). Esta constatação revela a pertinência de alterar a base de dados, transformando cada viagem em seis registos para cada um dos modos analisados. Esta transformação permitirá a formulação de funções de utilidade diferenciadas para cada modo. Após a “expansão” da base de dados, dever-se-á estudar quais as distribuições mais adequadas para caracterizar cada uma das variáveis e dos seus eventuais erros, de forma a poder defini-las para o modelo Logit Misto. Esta etapa deverá ser devidamente fundamentada através de testes com diferentes distribuições e crescentes quantidades de extrações para avaliar a estabilidade das simulações. A possibilidade de incorporar eventuais correlações entre alternativas e entre variáveis deverá ser um facto a considerar cuidadosamente, atendendo às tendências verificadas nas análises qualitativas com variáveis sociais.

Atendendo aos resultados destas análises, sugere-se também um estudo mais detalhado das amostras com comportamentos extremos, nomeadamente analisando eventuais padrões para os modos menos relevantes, ou para indivíduos com comportamentos muito diferentes dos normalmente esperados para as suas condições socioeconómicas.

O estudo poderá ainda alargar a análise de combinações a mais variáveis sociais, como o nível de instrução, discriminando mas detalhadamente as diferenças comportamentais entre diferentes grupos de cidadãos.

Dos resultados das simulações, mais concretamente para os perfis intermédios de utilizadores, a dificuldade de previsão da repartição modal para a Moto, sugere uma análise exploratória dos motivos dessa falta de capacidade explicativa. Esta análise poderá ser comparada com as previsões realizadas através de modelos mais complexos, no sentido de avaliar se são comportamentos de difícil modelação ou se se devem a dificuldades na capacidade explicativa dos modelos obtidos para esse modo em específico;

Atendendo às dificuldades verificadas no levantamento de dados sobre os sistemas de transportes existe a necessidade de em estudos futuros, provavelmente com dados mais recentes, encetar esforços no sentido de uma maior precisão e alargamento da gama de dados disponível. Atendendo a que parte dos resultados verificados poderão ter resultado de uma deficiente representação dos sistemas realmente implementados no local, sugere-se um esforço particular na aquisição de informação de qualidade sobre:

- Trajetos e horários de passagem nas diferentes paragens existentes ao longo das carreiras;
- Definição dos sistemas SIG da rede de transportes coletivos com entidades que permitam medições lineares e definir sentidos de circulação (*routes*) bem como o uso da aplicação *Network Analyst*. Esta aplicação poderá automatizar a análise de frequências de passagem bem como eventuais transbordos;
- Alargamento do levantamento de dados sobre a oferta de estacionamento e dos sistemas de TC pelo menos às áreas urbanas periféricas, nomeadamente às sedes de concelho, ou onde haja uma maior diversidade e quantidade de pares GA com informação fidedigna sobre os mesmos. Este alargamento deverá incluir a análise do tarifário existente e de eventuais condicionantes de acesso;
- Sugere-se a comparação dos resultados deste estudo com outros provenientes da sua eventual aplicação a outras áreas urbanas com realidades distintas, nomeadamente no que concerne à disponibilidade de várias alternativas mais sustentáveis ao automóvel (incluindo por exemplo modos ferroviários). Um caso de estudo com interesse poderá ser a evolução natural para a aplicação dos modelos, devidamente ajustados, à realidade na AMP depois da introdução do Metro do Porto e da reestruturação das carreiras dos STCP, no sentido de avaliar os seus impactos em termos de padrões de mobilidade e de valorização das diferentes variáveis explicativas adotadas;
- Sugere-se igualmente a aplicação da metodologia utilizada a outras áreas e bases de dados, também de menor dimensão, para avaliar a estabilidade de eventuais resultados comuns ou a deteção de especificidades culturais ou locais, resultantes da dimensão e da sua localização no país.

Capítulo IX

Referências Bibliográficas

- **Abreu e Silva, J. (2007);** “Poderão os padrões de uso do solo contribuir para uma mobilidade sustentável? Uma aplicação da modelação de equações estruturais à área Metropolitana de Lisboa”; Tese de Doutoramento em Engenharia do Território; Universidade Técnica de Lisboa - Instituto Superior Técnico; Maio.
- **Amador, Francisco; Gonzalez, Rosa; Ortúzar, Juan (2005);** "Preference heterogeneity and willingness-to-pay for travel time savings"; Transportation; Springer Netherlands; Vol. 32; Nº6; pag.627-647.
- **Arsénio, Elisabete; Bristow, Abigail; Nunes da Silva, Fernando; Pinelo, António; Wardman, Mark (2002);** “Externalidades ambientais dos transportes: novos modelos económicos com vista à sua avaliação”; Comunicação no 2º Congresso Rodoviário Português - As vias da segurança; Volume III; 18 a 20 de Novembro de 2002; Laboratório Nacional de Engenharia Civil; Lisboa.
- **Bao, Danewen, Deng, Wei; Gu, Shihui (2010);** “Impact of parking rates on resident travel behaviour”; Journal of Transportation Systems Engineering and Information Technology, Vol 10, Issue 3, July, page 80-85
- **Barbosa, José (2004);** “Noções sobre matrizes e sistemas de equações lineares”; Edições FEUP; Porto.
- **Ben-Akiva, Moshe; Lerman, Steven (1985);** "Discrete Choice Analysis: Theory and application to travel demand"; Massachusetts Institute of Technology; Cambridge.
- **Ben-Akiva, Moshe; Bolduc, Denis; Walker, Joan (2001);** “Specification, Identification and Estimation of the Logit Kernel (or Continuous Mixed Logit) Model”; Working paper; Massachusetts Institute of Technology; Cambridge.
- **Bhat, C.R. (1998a);** "Accommodating Variations in Responsiveness to Level-of-Service Variables in Travel Mode Choice Modelling", Transportation Research Part A, Vol. 32, No. 7, pp. 495-507.

- **Bhat, C.R. (1998b)**; "Accommodating Flexible Substitution Patterns in Multidimensional Choice Modelling: Formulation and Application to Travel Mode and Departure Time Choice", *Transportation Research Part B*, Vol. 32, No. 7, pp. 455-466.
- **Bhat, C.R. (2000)**; "A Multi-Level Cross-Classified Model for Discrete Response Variables", *Transportation Research Part B*, Vol. 34, No. 7, pp. 567-582
- **Carson, G.; Dix, M.; Callaghan, J.; Slevin, R. (1999)**; "Encouraging sustainable development by linking public transport accessibility, new parking standards and developer contributions"; *Traffic Engineering and Control*; July/August; page 370-379.
- **Cascetta, Ennio; Papola, Andrea (2003)**; "A joint mode-transit service choice model incorporating the effect of regional transport timetables"; *Transportation Research - Part B*, Vol. 37, page. 595-614
- **CEP - Centro de Estudos de Planeamento (1976)**; "Áreas de Influência dos centros urbanos: Fluxos rodoviários - Transporte de Passageiros"; Secretaria de Estado do Planeamento; Centro de Estudos de Planeamento; Estudo N°7; Volume I; Março; Lisboa
- **Chiou, Lesley; Walker, Joan (2006)**; "Identification and Estimation of Mixed Logit Models under Simulation Methods"; *Transportation Research Board Annual Meeting*; Paper N° 06-2787; 25 pages (CD-ROM);
- **COST 342 (2005)**; "Parking Policies and the Effects on Economy and Mobility - Report no COST Action 342", Chairman: Teun de Wit, CROW, Netherlands
- **Davidson, William; Donnelly, Robert; Vovsha, Peter; Freedman, Joel; Ruegg, Steve; Hicks, Jim; Castiglione, Joe; Picado, Rosella (2007)**; "Syntesis of first practices and operational research approaches in activity-based travel demand modelling"; *Transportation Research - Part A*; Elsevier Science; vol.41; p. 464-488.
- **Davidson, Peter (2008)**; "Introducing a 'new way'"; Guest Editor; *Traffic Engineering and Control*; Vol. 49, N°9; October; page. 322-323.
- **Domencich T.; MacFadden D. (1975)**; "Urban Travel Demand a behavioural analysis"; New York and Oxford: Elsevier.
- **European Commission (2011)**; "EU transport in figures: Statistical pocketbook 2011"; Publication Office of the European Union, Luxemburg

- **Foster, M. S.; Eastman, C. R. (1993);** “Parking and Public Transport - the effect on mode choice: a study of B1 developments”; Traffic Engineering and Control; October; page 480-489
- **Fricker, Jon; Whitford, Robert (2004);** “Fundamentals of Transportation Engineering - A multimodal systems approach”; Pearson Education - Prentice Hall; New Jersey.
- **Giannopoulos, G. (1990);** “Bus Planning and Operation in Urban Areas: A Practical Guide”; Avebury; Hampshire.
- **Gonçalves, Jorge; Seco, Álvaro; Mathias, Michael (2011);** “Análise da geração de viagens ao longo do dia, atendendo ao modo de transporte e ao motivo da viagem - A realidade Portuguesa”; ICEUBI2011 - INTERNATIONAL CONFERENCE ON ENGINEERING UBI2011;28-30 Novembro; University of Beira Interior; Covilhã; Portugal
- **Grava, S. (2003);** “Urban Transport Systems: Choices for communities”; McGraw-Hill Editors;
- **Guevara, C.; Ben-Akiva, M. (2009);** “Addressing Endogeneity in discrete choice models: Assessing Control-Function and Latent-Variable methods”; Working paper TSI-SOTUR-09-03; MITPortugal Program - Transportation Systems Focus Area; June http://www.mitportugal.org/index.php?option=com_docman&task=cat_view&gid=132&Itemid=108
- **Guimarães, Rui; Sarsfield Cabral, José (1997);** “Estatística”; Editora McGraw-Hill Portugal; Alfragide.
- **Hensher, David; Green, William (2003)** “The Mixed Logit model: The state of practice”; Transportation; Kluwer Academic Publishers; N°30; p. 133-176.
- **Hensher, David; Stopher, Peter, e Bullock, Philip (2003a)** “Service quality - developing a service quality index in the provision of commercial bus contracts.” Transportation Research - Part A; Pergamon -Elsevier Science; vol.37; p. 499-517.
- **Hensher, David; Rose, John; Green, William (2005);** “Applied Choice analysis - A Primer”; Cambridge University Press; Cambridge.
- **Hensher, David; Rose, John (2007);** "Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study"; Transportation Research - Part A, N°41, Elsevier Ltd, page 428-443.

- **Hess, Stephen; Polak, John (2003)**; “An alternative method to the scrambled Halton sequence for removing correlation between standard Halton sequence in high dimensions”; European Regional Science Conference; Jyväskylä; August; 25 pages.
- **INE - Instituto Nacional de Estatística (2002)**; “Inquérito à Mobilidade da População Residente 2000”; Instituto Nacional de Estatística, Direção Regional do Norte; financ. Direção Geral de Transportes Terrestres. - Porto
- **INE - Instituto Nacional de Estatística (2002a)**; “Censos 2001 - Resultados Provisórios”; Instituto Nacional de Estatística; Lisboa;
- **INE - Instituto Nacional de Estatística (2002b)**; “Estatísticas dos Transportes 2001”; Instituto Nacional de Estatística; Lisboa;
- **INE - Instituto Nacional de Estatística (2004)**; “As cidades em números - Informação estatística 2000-2002”; CD-ROM
- **ITE - Institute of Transport Engineering (1999)**; “Transportation Planning Handbook”; Institute of Transport Engineering; 2nd Edition.
- **Koppelman, F.; Bhat, C. (2006)**; “A Self Instructing Course in Mode Choice Modelling: Multinomial and Nested Logit Models”; U.S. Department of Transportation, Federal Transit Administration; June.
- **Krykova, Inna (2003)**; “Evaluating of path-dependent securities with low discrepancy methods”; Master in Science in Financial Mathematics Thesis; Worcester Polytechnic Institute; Massachusetts; December.
- **Li, Huamin; Huang, Haijun; Liu, Jianfeng (2010)**; “Parameter estimation of the Mixed Logit model and its application”; Journal of Transportation System Engineering and Information Technology, Vol. 10, Issue 5, October, page. 73-78
- **Louviere, Jordan; Street, Deborah (2000a)**; “Handbook of Transport Modelling: Chapter 8 - Stated Preference Methods”; Edited by David Hensher and Kenneth Button; Pergamon - Elsevier Science; Oxford.
- **Louviere, Jordan; Hensher, David; Swait, Joffre (2000b)**; “Stated Choice Methods: Analysis and application”; Cambridge University Press; Cambridge.
- **Louviere, J.; Train, K.; Ben-Akiva, M.; Bhat, C.; Brownstone, D.; Cameron, T.; Carson, R.; Deshazo, J.; Green, W.; Hensher, D.; Waldman, D. (2005)**; “Recent progress on endogeneity in Choice Modelling”; Marketing Letters; Volume 16; Number 3-4; December; page. 255-256

- **Montgomery, Douglas; Runger, George (2007);** “Applied Statistics and Probability for Engineers”; John Wiley and Sons; 4th Edition; Chichester.
- **Müller, S; Tscharaktschiew, S; Haase, K. (2008);** "Travel-to-school mode choice modelling and patterns of school choice in urban areas"; Journal of Transport Geography, N°. 16, page. 342-357
- **Munizaga, Marcela; Heydecker, Benjamin; Ortúzar, Juan de Dios (2000);** “Representation of heteroskedasticity in discrete choice models”; Transport Research - Part B; Pergamon-Elsevier Science; vol. 34; page. 219-240.
- **Munizaga, M., S. Jara-Diaz, P. Greeven, and C.R. Bhat (2008),** "Econometric Calibration of the Joint Time Assignment - Mode Choice Model," Transportation Science, Vol. 42, No. 2, pp. 208-219.
- **Meyer, M.; Miller, E. (2001);** “Urban Transportation Planning”; McGraw-Hill; 2nd Edition; New York.
- **Microsoft Corporation (2001);** “AutoRoute 2001”; CD-Rom
- **Olzewski, P.; Cher-sin Tan (1999);** “Walking Lessons: Pedestrians travel in Singapore”; Traffic Engineering and Control; October, vol. 40 N°10, page. 480-483;
- **Ortúzar, Juan; Willumsen, Luis (2001);** “Modelling Transport”; John Wiley & Sons; 3rd Edition; Chichester.
- **Pons, Joana M. S.; Bey, Joana M. P. (1991);** “Geografia de Redes y Sistemas de Transporte”; Editorial Síntesis; Madrid
- **Pozsgay, Michael; Bhat, Chandra (2002);** "Destination Choice Modelling for Home-Based recreational Trips; Analysis and Implications for Land-use, Transportation and Air Quality planning"; Transportation Research Record, Vol. 1777, page. 47-54
- **QUATTRO (1998);** “Quality approach in tendering urban public transport operations - Final Report - Synthesis and Recommendations”; European Commission under the Transport RTD Programme of de EU’s 4th Framework for Research, Technological Development and Demonstration.
- **Rigby, Dan; Burton, Mike (2006);** “Modelling Disinterest and Dislike: A Bounded Bayesian Mixed Logit Model of the UK Market for GM Food”; Environmental and Resource Economics; Springer Netherlands; Vol°33, N°4; page. 485-509; April.

- **Rodrigues, Carlos; Seco, Álvaro; Ribeiro, Paulo; Ribeiro, Eduardo (2005);** “Caracterização do Estacionamento na Cidade do Porto - Relatório Final”; Relatório Interno; Faculdade de Engenharia do Porto, Câmara Municipal do Porto; Dezembro
- **Seco, Álvaro; Gonçalves, Jorge (2006);** “Parking supply standards using a “matrix” approach: analysis of its potential and applicability”; Routes/Roads Magazine; Road World Association - PIARC; Nº 331; pag.72-81.
- **Seco, Álvaro; Gonçalves, Jorge (2007);** “The Quality of Public Transports: Relative importance of different performance indicators and their potential to explain Modal Choice”; Urban Transport and the Environment in the 21st Century - Urban Transport 2007 Conference; 3-5 September; Coimbra; WIT Press (Editor C. Brebbia); Vol. 96; page. 313-325.
- **Silva, M., Garrido, R.A., (2003);** “Estimation performance of low discrepancy sequences in stated preference”. In: Proceedings 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.
- **Train, Kenneth (2000);**”[Halton Sequences for Mixed Logit](#)”; [Economics Working Papers](#) E00-278; University of California at Berkeley.
- **Train, Kenneth (2003);** “Discrete Choice Methods with simulation”; Cambridge University Press; New York.
- **Train, Kenneth (2009);** “Discrete Choice Methods with simulation”; 2nd Edition; Cambridge University Press; New York. (por publicar, <http://elsa.berkeley.edu/books/choice2.html> acedido em 15-08-2009
- **Tyler, Nick (2002);** “Accessibility and the Bus System: from concepts to practice”; Thomas Telford Publishing; London.
- **VTPI - Victoria Transport Policy Institute (2007);** ; “Traffic Demand Management Encyclopaedia”; disponível on-line em Maio de 2007
<http://www.vtpi.org/tdm/index.php>
- **Vovsha, Peter; Freedman, Joel (2008a);** “What is an Activity-Based Model and how does it Work?”; Traffic Engineering and Control; Vol. 49, Nº9; October; page. 330-335.
- **Vovsha, Peter; Donnelly, Robert; Chiao, Kuo-Ann (2008b);** “Activity-Based model for the mega-region of New York”; Traffic Engineering and Control; Vol. 49, Nº9; October; page. 336-340.

- **Vuchic, V. (1981);** “Urban Public Transportation - Systems and Technology”; Prentice-Hall, Inc.; New Jersey.
- **Vuchic, V. (1999);** “Transportation for liveable cities”; Rutgers - State University of New Jersey;
- **Washington, Simon; Karlaftis, Matthew; Mannering, Fred (2003);** “Statistical and Econometric methods for transportation data analysis”; Chapman & Hall/CRC; Boca Raton.
- **Weatherall, R. (1997);** “What we need to know about walking”; Traffic Engineering and Control; July/Ago, vol. 38 N°7/8, page. 385-386;
- **Wegener, M. (1995);** “Chapter 11 - Accessibility and Development Impacts”; in “Transport and Urban Development”; Edited by Banister, D.; E&FN Spon; London
- **White, C.; Foster-Clark, J.; Lee, A.; Salter, S. (2000);** “Developing a strategy for Blackwater Valley”; Traffic Engineering and Control; February; page 58-59.
- **White, P. (2002);** “Public transport: its planning, management and operation”; SPON press; 4th Edition;
- **Walker, Joan (2002);** “The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification”; Transportation Research Record 1805; Transportation Research Board; page. 86-98.
- **Yañez, M.F.; Raveau, S.; Ortuzar, J.D. (2010);** " Inclusion of latent variables in Mixed Logit models: Modelling and forecasting"; Transportation Research Part A, N°44, page 744-753
- Site: http://www.puc-rio.br/marco.ind/spreads_list.html; acedido em 27-3-2009; Autoria: Marco Antonio Guimarães Dias, Senior Consultant (internal) and Senior Petroleum Engineer - Petrobras/Exploration & Production - Adjunct Professor (part time) - PUC-Rio

Anexos

Anexo 1 do Capítulo II

1. Amostragem Aleatória Simples

Neste tipo de amostragem, a mais comum, a amostra é retirada aleatoriamente da população pelo que cada indivíduo tem a mesma probabilidade de ser escolhido. Considera-se que as características da amostra são iguais às da população original. É este tipo de amostragem que os modelos Logit assumem como o utilizado na recolha de dados.

2. Amostragem Estratificada Generalizada

Esta abordagem distingue-se da anterior pelo facto de as observações serem retiradas aleatoriamente não de toda a população, mas de subgrupos (estratos) definidos atendendo a determinados valores do i e x (Ben-Akiva e Lerman, 1985), isto é, a amostra não é aleatória apenas em relação a determinadas variáveis independentes ou escolhas conhecidas. É particularmente útil quando se pretender modelar partes da população importantes mas em minoria (reduzido número em relação ao total da população). Esta metodologia deu origem a um subconjunto de estratégias de amostragem não aleatórias. A amostragem aleatória é um caso especial da amostragem estratificada em que o número de observações de cada estrato corresponde exatamente a proporção do tamanho do estrato em relação à população total.

3. Amostragem Exógena

Na amostragem exógena, os estratos são definidos em função de determinadas variáveis independentes e não das escolhas em si. Dividem-se as possíveis variáveis independentes x em subconjuntos exaustivos de $x_1, x_2, \dots, x_G$, e os pares (i, x) que tiverem $x \in x_G$ pertencem a esse estrato G . Um exemplo pode ser a estratificação da população por categorias de rendimentos.

4. Amostragem Baseada em Escolhas

Esta abordagem pode ser utilizada quando se pretende uma representatividade suficiente de um determinado resultado (por exemplo utilizadores de TC). Dividem-se o conjunto total de escolhas em subconjuntos exaustivos $C_1, C_2, \dots, C_G$, e os pares (i, x) que tiverem um $i \in C_g$, formam o estrato G . Num problema de escolhas modais corresponde à segmentação da população segundo o modo escolhido. Se a proporção de resultados na amostra não for igual à proporção dos resultados na população total então um ajustamento deve ser realizado na estimação. Este ajustamento é realizado nas constantes específicas de cada alternativa (geralmente serão $n-1$ constantes, com n igual ao número de alternativas), subtraindo-lhe $\ln(SF_i/PF_i)$ (Washington et al., 2003), onde SF_i é a fracção de observações da amostra com resultado i e PF_i é a fracção de observações da população total que escolheu i .

5. Amostragem Enriquecida

Ben-Akiva e Lerman (1985) definem esta abordagem como a agregação de amostras exógenas estratificadas com uma ou mais baseadas em escolhas. É uma estratégia atrativa de forma a incrementar o número de observações que escolhem uma determinada alternativa quando a sua população agregada é muito reduzida comparativamente a população total.

6. Amostragem Dupla

Esta estratégia consiste na utilização de informação obtida através de amostragens aleatórias para direcionar a aquisição de dados adicionais de forma a aumentar a amostra de componentes com pouca representação na população. Com este procedimento, aproveita-se a informação de uma amostragem para melhorar as amostragens seguintes, ou seja, utiliza-se a informação de uma primeira amostragem para estimar as proporções da população dos diferentes estratos a utilizar numa segunda amostragem mais apropriada para os objetivos específicos definidos. Outra maneira de aplicar esta estratégia pode ser a estimação de um modelo com uma amostra reduzida e utilizar os resultados para definir uma estratégia de amostragem posterior mais completa.

Anexo 1 do Capítulo III

Definição das expressões das variâncias e covariâncias entre erros (Train, 2003)

- Variância entre os 1º e 2º erros:

$$Var(\tilde{\varepsilon}_{n21}) = Var(\varepsilon_{n2} - \varepsilon_{n1}) = Var(\varepsilon_{n1}) + Var(\varepsilon_{n2}) - 2 \cdot Cov(\varepsilon_{n1}; \varepsilon_{n2}) = \sigma_{11} + \sigma_{22} - 2 \cdot \sigma_{12}$$

- Variância entre os 1º e 3º erros:

$$Var(\tilde{\varepsilon}_{n31}) = Var(\varepsilon_{n3} - \varepsilon_{n1}) = Var(\varepsilon_{n1}) + Var(\varepsilon_{n3}) - 2 \cdot Cov(\varepsilon_{n1}; \varepsilon_{n3}) = \sigma_{11} + \sigma_{33} - 2 \cdot \sigma_{13}$$

- Variância entre os 1º e 4º erros:

$$Var(\tilde{\varepsilon}_{n41}) = Var(\varepsilon_{n4} - \varepsilon_{n1}) = Var(\varepsilon_{n1}) + Var(\varepsilon_{n4}) - 2 \cdot Cov(\varepsilon_{n1}; \varepsilon_{n4}) = \sigma_{11} + \sigma_{44} - 2 \cdot \sigma_{14}$$

- Covariância dos erros entre os 2ºs e os 3ºs erros:

$$\begin{aligned} Cov(\tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n31}) &= E(\varepsilon_{n2} - \varepsilon_{n1}) \cdot (\varepsilon_{n3} - \varepsilon_{n1}) = E(\varepsilon_{n2} \cdot \varepsilon_{n3} - \varepsilon_{n2} \cdot \varepsilon_{n1} - \varepsilon_{n1} \cdot \varepsilon_{n3} + \varepsilon_{n1} \cdot \varepsilon_{n1}) = \\ &= \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} \end{aligned}$$

- Covariância dos erros entre os 2ºs e os 4ºs erros:

$$\begin{aligned} Cov(\tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n41}) &= E(\varepsilon_{n2} - \varepsilon_{n1}) \cdot (\varepsilon_{n4} - \varepsilon_{n1}) = E(\varepsilon_{n2} \cdot \varepsilon_{n4} - \varepsilon_{n2} \cdot \varepsilon_{n1} - \varepsilon_{n1} \cdot \varepsilon_{n4} + \varepsilon_{n1} \cdot \varepsilon_{n1}) = \\ &= \sigma_{24} - \sigma_{12} - \sigma_{14} + \sigma_{11} \end{aligned}$$

- Covariância dos erros entre os 3ºs e os 4ºs erros:

$$\begin{aligned} Cov(\tilde{\varepsilon}_{n31}, \tilde{\varepsilon}_{n41}) &= E(\varepsilon_{n3} - \varepsilon_{n1}) \cdot (\varepsilon_{n4} - \varepsilon_{n1}) = E(\varepsilon_{n3} \cdot \varepsilon_{n4} - \varepsilon_{n3} \cdot \varepsilon_{n1} - \varepsilon_{n1} \cdot \varepsilon_{n4} + \varepsilon_{n1} \cdot \varepsilon_{n1}) = \\ &= \sigma_{34} - \sigma_{13} - \sigma_{14} + \sigma_{11} \end{aligned}$$

Anexo 2 do Capítulo III

Tratamento de funções quadráticas, definição do tamanho do incremento e avaliação da concavidade no procedimento de *Newton-Raphson* (NR) para a maximização da função de verosimilhança (Train, 2003)

No caso de a função de verosimilhança logarítmica ser exatamente quadrática, o procedimento NR aproxima o máximo numa única iteração independentemente do valor inicial. Como exemplo apresenta-se o caso de apenas um parâmetro a estimar ($K=1$), onde a função de verosimilhança logarítmica pode ser expressa por:

$$LL(\theta) = a + b.\theta + c.\theta^2$$

O seu máximo será:

$$\frac{\partial LL(\theta)}{\partial \theta} = b + 2.c.\theta = 0 \Leftrightarrow \hat{\theta} = -\frac{b}{2.c}$$

O seu gradiente será $g_t = b + 2.c.\theta_t$ e a sua hessiana será $H_t = 2.c$, pelo que o procedimento de NR implica um incremento de:

$$\theta_{t+1} = \theta_t - \frac{g_t}{H_t} = \theta_t - \frac{b + 2.c.\theta_t}{2.c} = \theta_t - \frac{b}{2.c} - \theta_t = -\frac{b}{2.c} = \hat{\theta}$$

No entanto, a maioria das funções de verosimilhança logarítmica não são quadráticas pelo que necessitam de mais iterações para chegar ao máximo.

Em relação ao tamanho do incremento, podem existir situações em que incremento proposto pelo procedimento NR ultrapasse o valor máximo, necessitando de ajustes de modo a poder determiná-lo exatamente. Um possível ajustamento que evita aumentar o número de iterações será a inclusão de um valor de ajustamento do tamanho do incremento ξ :

$$\theta_{t+1} = \theta_t - \xi \cdot \frac{g_t}{H_t}$$

Este ξ poderá começar com $\xi = 1$, e se $LL(\theta_{t+1}) > LL(\theta_t)$ realiza-se o incremento e passa-se à iteração seguinte. Se no entanto $LL(\theta_{t+1}) < LL(\theta_t)$, então ξ tomará um valor inferior à unidade por exemplo 0,5, e verifica-se de novo se $LL(\theta_{t+1}) > LL(\theta_t)$. Se esta inequação não se verificar volta-se a reduzir o ξ até que se verifique. Se ξ tiver que ser muito reduzido, então o

processo incremental pouco acrescenta para encontrar o máximo e outro processo deve ser ponderado.

O procedimento no sentido contrário deve ser igualmente implementado, comparando os resultados do incremento com $\xi = 2$, se $LL(\theta_{t+1}) > LL(\theta_t)$ com $\xi = 2$ então experimenta-se com o dobro do valor anterior $\xi = 4$, se $LL(\theta_{t+1}) > LL(\theta_t)$ com $\xi = 4$, continua-se com $\xi = 8$. Se se encontrar um incremento em que $LL(\theta_{t+1}) < LL(\theta_t)$ será o incremento imediatamente anterior que deverá ser utilizado para a aproximação ao máximo. Este procedimento geralmente reduz o número de iteração a realizar para chegar ao máximo, sem ter que calcular tantas vezes o gradiente e a hessiana da função de verosimilhança logarítmica.

Quanto à concavidade, se a função de verosimilhança logarítmica for globalmente côncava o procedimento NR garante um aumento da função em cada incremento, no entanto se a função em algumas partes não for côncava, o procedimento pode falhar. Se numa determinada parte da função, onde se esteja a calcular o θ_t a função for convexa, então o procedimento NR direciona a iteração no sentido oposto à inclinação da função. Em zonas convexas a hessiana não fornece informação útil sobre o sentido do máximo (Train, 2003) uma vez que não é uma matriz positiva definida, podendo ponderar-se a alteração do seu sinal. No entanto, existem procedimentos mais fáceis para este tipo de situações que não passam pelo cálculo da hessiana em cada iteração, cujo processamento é geralmente mais exigente. A maioria das metodologias sugeridas por Train (2003) passa pela utilização de outras matrizes que não a hessiana, mais fáceis de determinar (ver Train (2003) pág. 196-202). A avaliação sobre a validade do valor máximo encontrado, nomeadamente se é um máximo local (apenas de parte da função) ou global (de toda a função sendo este o pretendido), pode realizar-se através da verificação da convergência do procedimento de NR para diferentes valores de partida. Se os máximos encontrados forem diferentes a função têm máximos locais e há necessidade de avaliar convenientemente qual o maior. Se os valores encontrados convergirem para o mesmo ponto, então esse será o máximo global pretendido.

Anexo 3 do Capítulo III

Definição de elasticidades agregadas

Ben-Akiva e Lerman (1985) referem a existência de elasticidades agregadas, onde a elasticidade não é mais que a média ponderada das elasticidades individuais utilizando as probabilidades de escolha como pesos. Assim considera que:

$$\bar{P}(i) = \frac{\sum_{n=1}^N P_n(i)}{N}$$

é a probabilidade de escolha da alternativa i esperada (cota de mercado) para o segmento de N indivíduos da população. E que:

$$E_{x_{jnk}}^{\bar{P}(i)} = \frac{\sum_{n=1}^N [P_n(i) \cdot E_{x_{jnk}}^{P_n(i)}]}{\sum_{n=1}^N P_n(i)}$$

é a Elasticidade agregada de um segmento de N elementos da população considerando a elasticidade desagregada $E_{x_{jnk}}^{P_n(i)}$;

Assim, a elasticidade direta agregada será:

$$E_{x_{jnk}}^{\bar{P}(i)} = \frac{\sum_{n=1}^N [P_n(i) \cdot [\delta_{ij} - P_n(j)] \cdot x_{jnk} \cdot B_k]}{\sum_{n=1}^N P_n(i)} = \frac{B_k}{N \cdot \bar{P}(i)} \cdot \sum_{n=1}^N [P_n(i) \cdot (\delta_{ij} - P_n(j)) \cdot x_{jnk}]$$

onde δ_{ij} é a função *Delta de Kronecker* que toma o valor 1 se $i = j$ e 0 de $i \neq j$. No caso de elasticidade cruzada agregada, $\delta_{ij} = 0$ e a expressão depende de i e j , pelo que se verifica que a IIA não necessita de se manter ao nível agregado.

Anexo 4 do Capítulo III

Variações dos Modelos GEV (Train, 2003)

1. Decomposição em Dois Logit Padrão

Sem perda de generalidade a componente observada da utilidade pode ser decomposta em duas parcelas: uma designada de W constante para todas as alternativas do nicho; e uma designada por Y que varia entre alternativas do mesmo nicho. Assim a utilidade será expressa por:

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj}, \text{ com } j \in B_k$$

Onde:

- W_{nk} : apenas depende de variáveis que descrevam o nicho k . Estas variáveis diferem entre nichos mas não entre alternativas do mesmo nicho.
- Y_{nj} : depende de variáveis que descrevam a alternativa j . Estas variáveis diferem entre alternativas do mesmo nicho, e pode ser definida simplesmente como $Y_{nj} = V_{nj} - W_{nk}$.

Com esta decomposição da utilidade a probabilidade do modelo Logit Hierárquico pode ser expressa como o produto de duas probabilidades Logit padrão, nomeadamente o produto da probabilidade de uma alternativa i pertencente ao nicho B_k ser definida com a probabilidade da alternativa i do nicho k ser a escolhida dado que uma alternativa de B_k foi escolhida, ou seja:

$$P_{ni} = P_{ni|B_k} \cdot P_{nB_k}$$

Onde:

- $P_{ni|B_k}$: é a probabilidade condicional de se escolher a alternativa i dado que se escolheu uma alternativa que pertence ao nicho B_k .
- P_{nB_k} : é a probabilidade marginal de escolher uma alternativa do nicho B_k (com a marginalidade considerada sobre todas as alternativa em B_k).

Com esta decomposição a probabilidade de escolha de uma alternativa i no nicho B_k , segundo o modelo Logit Hierárquico, assume a forma de Logit padrão e serão expressas por:

$$P_{nB_k} = \frac{e^{W_{nk} + \lambda_k \cdot I_{nk}}}{\sum_{l=1}^K e^{W_{nl} + \lambda_l \cdot I_{nl}}}$$

$$P_{nij|B_k} = \frac{e^{\frac{Y_{ni}}{\lambda_k}}}{\sum_{j \in B_k} e^{\frac{Y_{nj}}{\lambda_k}}}$$

$$I_{nk} = \ln \sum_{j \in B_k} e^{\frac{Y_{nj}}{\lambda_k}}$$

É usual designar a probabilidade marginal (escolha do nicho) como o modelo superior e a probabilidade condicional (escolha da alternativa dentro do nicho) como o modelo inferior, em analogia com o diagrama em árvore anteriormente apresentado. A quantidade I_{nk} liga o modelo superior com o inferior, sendo o logaritmo do denominador do modelo inferior, e $\lambda_k \cdot I_{nk}$ é a utilidade esperada que o indivíduo receba da escolha entre as alternativas do nicho B_k . Assim I_{nk} é geralmente chamado de *valor inclusivo* ou *utilidade inclusiva* do nicho B_k , ou atendendo a que é o logaritmo da soma dos exponenciais da utilidade representativa, termo da log-soma. A inserção do termo inclusivo I_{nk} no modelo superior como uma variável independente é justificada pela probabilidade de escolher o nicho B_k estar dependente da utilidade esperada que um indivíduo recebe desse nicho, incluindo a utilidade que recebe independentemente da alternativa que escolher no nicho (W_{nk}) e a utilidade esperada extra que recebe por poder escolher a melhor alternativa do nicho ($\lambda_k \cdot I_{nk}$).

Os parâmetros de um modelo com nichos podem ser estimados através de técnicas comuns da máxima verossimilhança logarítmica, substituindo as probabilidades de escolha na função de verossimilhança logarítmica de modo a obter uma função explícita dos parâmetros do modelo. Os valores que maximizarem a função são, em condições gerais, consistentes e eficientes (Train, 2003). Recomenda-se a utilização da estimação simultânea da máxima verossimilhança logarítmica de modo a assegurar que parâmetros comuns a diferentes alternativas sejam restringidos à igualdade onde quer que apareçam. Como a maximização numérica em algumas situações pode ser mais difícil, dado que a função de máxima verossimilhança logarítmica não é globalmente côncava ou próxima de uma função quadrática, Train (2003) apresenta alguns procedimentos para alterar o algoritmo de aproximação ao máximo ou para definir valores de partida.

2. Logit Hierárquico Sobreposto

Nos modelos Logit Hierárquicos até agora considerados, cada alternativa pertencia a apenas um nicho, que em certas situações pode ser uma restrição não apropriada. Considere-se o

exemplo da figura III.6, onde se agrupou o condutor com o passageiro uma vez que têm atributos não observados similares como o conforto, a privacidade, etc. O passageiro, no entanto, também tem atributos não observados similares aos dos TC como seja a dependência em relação a horários, uma vez que não pode realizar a viagem quando quer mas sim quando o condutor pode levar, tal como os passageiros dos TC estão limitados aos períodos estipulados nos horários. Seria assim útil que se pudesse colocar os passageiros em dois nichos diferentes eventualmente com erros correlacionados com os dos TC e erros correlacionados, em grau diferente, com os dos condutores. Vários modelos foram propostos nos anos 90 para modelar a modos com sobreposição de nichos, designados de modelos *Logit Hierárquicos cruzados*. Train (2003) refere ainda alguns outros modelos desenvolvidos assumindo diferentes formas de sobreposição, como o modelo de valores extremos generalizado ordenado, o Logit de pares combinatórios (PCL), ou o Logit hierárquico generalizado (GNL).

3. Logit de Pares Combinatórios (PCL)

Neste tipo de modelos, vistos como dos mais simples, cada par de alternativas é considerado como pertencente a um nicho. Como cada alternativa está emparelhada com cada uma das restantes alternativas, pertence a $J-1$ nichos. Um parâmetro λ_{ij} indica o grau de independência entre a alternativa i e j . Se $\lambda_{ij} = 1$ para todas as alternativas, o PCL torna-se num Logit padrão. A probabilidade de escolha com o PCL será:

$$P_{ni} = \frac{\sum_{j \neq i} e^{\frac{V_{ni}}{\lambda_{ij}}} \left(e^{\frac{V_{ni}}{\lambda_{ij}}} + e^{\frac{V_{nj}}{\lambda_{ij}}} \right)^{\lambda_{ij}-1}}{\sum_{k=1}^K \sum_{l=k+1}^J \left(e^{\frac{V_{nk}}{\lambda_{kl}}} + e^{\frac{V_{nl}}{\lambda_{kl}}} \right)^{\lambda_{kl}}}$$

A soma em numerador implica todos $J-1$ nichos em que a alternativa i está presente, e para os restantes nichos, o valor a somar é igual ao numerador da probabilidade Logit Hierárquica (LH), pelo que o modelo é similar ao LH mas permite que a alternativa i esteja em mais de um nicho (Train, 2003). O denominador dos PCL também assume a forma expressa no LH, sendo a soma em todos os nichos da soma dos $e^{(V/\lambda)}$ pertencentes a cada nicho, elevados ao λ apropriado. Se este λ estiver entre 0 e 1 para todos os pares i e j , então o modelo é consistente com a maximização da utilidade para todos os níveis dos dados. Train (2003) refere estudos que mostraram que o PCL estimava melhores resultados que o LH ou o Logit padrão.

Como a escala e o nível da utilidade são imateriais, dos parâmetros da matriz de covariância apenas $(J(J-1)/2)-1$ podem ser estimados por modelos de escolha discreta. Assim, os modelos PCL terão $J(J-1)/2$ parâmetros λ (um para cada alternativa emparelhada com todas as outras mas que i emparelhado com j é igual a j emparelhado com i), excedendo o número de

parâmetros identificáveis por uma unidade e implicando a imposição de restrições a pelo menos um λ . Esta restrição pode ser imposta, por exemplo, através da sua normalização impondo que um deles seja igual à unidade, que permite a normalização automática da utilidade.

4. Logit Hierárquico Generalizado (GNL)

Uma forma geral para o problema da sobreposição de nichos será a introdução de um parâmetros de alocação α_{ik} que reflita qual a proporção em que uma alternativa pertence a um nicho B_k . Assim, não só a alternativa pode pertencer a diferentes nichos como pode ter importâncias diferentes em cada um. Este parâmetro tem ser que maior ou igual a zero e se for igual a zero ($\alpha_{ik} = 0$) indica que a alternativa i simplesmente não existe nesse nicho k . Para facilitar a sua interpretação a soma de todos os parâmetros de alocação de uma alternativa nos diferentes nichos deve ser 1.

Um parâmetro λ_k é definido para cada nicho e tem a mesma função que nos modelos Logit Hierárquicos, ou seja, indicar o grau de independência entre alternativas no mesmo nicho. Valores de λ_k maiores traduzem maior independência e menos correlação.

Assim a probabilidade de escolha a alternativa i pelo indivíduo n será:

$$P_{ni} = \frac{\sum_k (\alpha_{ik} \cdot e^{V_{ni}})^{1/\lambda_k} \cdot \left(\sum_{j \in B_k} (\alpha_{jk} \cdot e^{V_{nj}})^{1/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^K \left(\sum_{j \in B_l} (\alpha_{jl} \cdot e^{V_{nj}})^{1/\lambda_l} \right)^{\lambda_l}}$$

A fórmula é semelhante à probabilidade de escolha pelo Logit Hierárquico, com exceção do numerador que é a soma de todos os nichos onde a alternativa i existe multiplicado pela sua proporção em cada nicho α_{ik} . Se cada alternativa apenas existir num único nicho k , $\alpha_{jk} = 1$ para $j \in B_k$, e será zero para os restantes, tornando o modelo o Logit Hierárquico. Se além disso, $\lambda_k = 1$ para todos os nichos, todos independentes e sem correlação, então o modelo torna-se no Logit padrão. Para facilitar a interpretação das probabilidades do GNL, a fórmula anterior pode ser decomposta em:

$$P_{ni} = \sum_k P_{ni|B_k} \cdot P_{nk}$$

Onde a probabilidade de escolha do nicho k será:

$$P_{nk} = \frac{\left(\sum_{j \in B_k} (\alpha_{jk} \cdot e^{V_{nj}})^{1/\lambda_k} \right)^{\lambda_k}}{\sum_{l=1}^K \left(\sum_{j \in B_l} (\alpha_{jl} \cdot e^{V_{nj}})^{1/\lambda_l} \right)^{\lambda_l}}$$

E a probabilidade de escolher i no nicho k é:

$$P_{ni|B_k} = \frac{(\alpha_{ik} \cdot e^{V_{ni}})^{1/\lambda_k}}{\sum_{j \in B_k} (\alpha_{jk} \cdot e^{V_{nj}})^{1/\lambda_k}}$$

5. Logit Heteroscedástico

Existem situações em que o analista, em vez de tentar calcular as correlações entre alternativas, pode desejar apenas que as variâncias entre erros de diferentes alternativas sejam diferentes, isto é, que sejam heteroscedásticas.

Train (2003) apresenta um tipo de modelos desenvolvido por outros autores designado de *modelos de valores extremos heteroscedásticos* (HEV), semelhantes aos modelos Logit mas nos quais se permite diferentes variâncias para diferentes alternativas. A utilidade é expressa como:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

Onde ε_{nj} tem distribuições de valores extremos independentes com variância $(\theta_j \cdot \pi)^2/6$ diferente para cada alternativa j . Considera-se que não existe correlação entre os erros as alternativas, mas a variância dos seus erros é diferente de alternativa para alternativa. Para definir a escala da utilidade normaliza-se a escala de uma alternativa como $\pi^2/6$, a variância de uma distribuição de valores extremos padrão, ficando as variâncias das restantes estimadas em relação à alternativa normalizada. Assim, a probabilidade de escolha com um Logit HEV será:

$$P_{ni} = \int \left[\prod_{j \neq i} e^{-e^{-\frac{(V_{ni} - V_{nj} + \theta_j \cdot \omega)}{\theta_j}}} \right] \cdot e^{-e^{-\omega}} \cdot e^{-\omega} \cdot d\omega$$

Onde $\omega = \varepsilon_{ni} / \theta_i$.

O integral não tem uma fórmula definida, mas pode ser aproximado por simulação, e como $e^{-e^{-\omega}} \cdot e^{-\omega}$ é a densidade de probabilidades para valores extremos, a expressão transforma-se

num integral do que está entre parêntesis reto sobre a densidade de valores extremos. A sua simulação pode ser realizada da seguinte forma:

1. Retira-se um valor aleatório com distribuição de valores extremos ω . Se o integral for unidimensional, ou seja, tem apenas uma variável, o procedimento pode ser realizado através de uma função uniforme padrão da sua função inversa (ver Train (2003) pág. 209 e 210).
2. Para este valor de ω calcula-se a expressão entre parêntesis reto.
3. Repete-se este procedimento várias vezes e calcula-se a média.

Esta média é uma aproximação à probabilidade de escolha P_{ni} . Train (2003) refere estudos que demonstraram que sendo o integral unidimensional, o seu cálculo pode ser realizado efetivamente através de quadratura em vez de simulação.

6. Formulação Geral para Modelos GEV

Train (2003) apresenta uma metodologia para a obtenção de um modelo GEV que melhor se adapte às circunstâncias específicas de um determinado problema.

Considere-se uma função G que depende de $e^{Y_j} = Y_i$ para todas as j alternativas, ou seja, $G = (Y_1, \dots, Y_j)$. Se a sua derivada em relação à alternativa i for $G_i = \frac{\partial G}{\partial Y_i}$, e se esta satisfizer as seguintes condições:

1. $G \geq 0$ para todos os valores positivos de $Y_j \forall j$;
2. G é homogénea de grau 1, implicando que a multiplicação dos seus elementos por uma proporção ρ aumenta a própria função na mesma proporção, isto é:

$$G(\rho.Y_1, \dots, \rho.Y_j) = \rho.G(Y_1, \dots, Y_j)$$

3. A função tende para infinito se todas as suas componentes tendem para infinito:

$$Y_j \rightarrow \infty \forall j \Rightarrow G \rightarrow \infty$$

As derivadas cruzadas parciais de G mudam de sinal de uma determinada forma, isto é:

$$\begin{aligned}
G_i &\geq 0; \forall i \\
G_{ij} &= \frac{\partial G_i}{\partial Y_j} \leq 0; \forall j \neq i \\
G_{ijk} &= \frac{\partial G_{ij}}{\partial Y_k} \geq 0; i \neq j \neq k \\
&\vdots
\end{aligned}$$

(Nota: se as derivadas têm que ser maior ou igual a zero ou menor ou igual a zero, serem iguais a zero também verifica esta condição, incluindo assim equações de 1º grau ou lineares!)

Se as condições anteriores se verificarem, então o modelo de escolha discreta pode ser definido como:

$$P_{ni} = \frac{Y_{ni} \cdot G_i}{G}$$

Sendo consistente com a teoria da maximização da utilidade. Qualquer modelo derivado desta forma é um modelo de valores extremos generalizado. Train (2003) refere que as propriedades apresentadas têm pouco significado do ponto de vista da teoria económica apresentando pouca informação sobre o modo como G deve ser explicitado para que o modelo seja adequado ao problema a resolver. Por outro lado, a abordagem puramente matemática permite a geração de modelos que partindo de teorias económicas poderiam não ser desenvolvidas.

Anexo 5 do Capítulo III

Metodologia genérica para a definição da matriz M de diferença de erros (Train, 2003)

Para o cálculo da matriz de diferenças de erros, Train (2003) sugere uma metodologia genérica através da definição de uma matriz de identidade (apenas com 1's na diagonal) de dimensão $J-1$ com uma coluna extra de -1 adicionada na i -énésima coluna. Esta coluna extra torna a matriz com tamanho $J-1$ por J . Num exemplo de $J = 4$ alternativas e $i = 3$, ter-se-ia a seguinte matriz ($J-1 \times J = 3 \times 4$):

$$M_3 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Esta matriz pode ser utilizada para transformar a matriz de covariância dos erros na matriz de covariância da diferença entre erros, uma vez que $\tilde{\Omega}_i = M_i \Omega M_i^T$. Considerando um exemplo com três alternativas com erros $\langle \varepsilon_{n1}, \varepsilon_{n2}, \varepsilon_{n3} \rangle$ e matriz de covariância:

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

Se se pretender definir as diferenças em relação à 2ª alternativa, então a matriz de covariância das diferenças será $\langle \tilde{\varepsilon}_{n12}, \tilde{\varepsilon}_{n32} \rangle$:

$$\tilde{\Omega}_2 = \begin{pmatrix} \varepsilon_{n1} - \varepsilon_{n2} \\ \varepsilon_{n3} - \varepsilon_{n2} \end{pmatrix} = \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2.\sigma_{12} & \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} \\ \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} & \sigma_{33} + \sigma_{22} - 2.\sigma_{23} \end{pmatrix}$$

Que também pode ser obtida através da transformação referida:

$$\begin{aligned} \tilde{\Omega}_2 &= M_2 \Omega M_2^T = \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \sigma_{13} - \sigma_{12} & \sigma_{23} - \sigma_{22} & \sigma_{33} - \sigma_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} \sigma_{11} - \sigma_{12} - \sigma_{12} + \sigma_{22} & -\sigma_{12} + \sigma_{22} + \sigma_{13} - \sigma_{23} \\ \sigma_{13} - \sigma_{12} - \sigma_{23} + \sigma_{22} & -\sigma_{23} + \sigma_{22} + \sigma_{33} - \sigma_{23} \end{pmatrix} = \\ &= \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2.\sigma_{12} & \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} \\ \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} & \sigma_{22} + \sigma_{33} - 2.\sigma_{23} \end{pmatrix} \end{aligned}$$

Anexo 6 do Capítulo III

Metodologia para a normalização de modelos Probit (Train, 2003)

Considere-se como exemplo a expressão usual para a utilidade $U_{nj} = V_{nj} + \varepsilon_{nj}$, e um modelo com quatro alternativas $j=1, \dots, 4$. O vetor dos erros $\langle \varepsilon_{n1}, \varepsilon_{n2}, \varepsilon_{n3}, \varepsilon_{n4} \rangle$ tem distribuição normal com média igual a zero e matriz de covariância Ω (matriz simétrica):

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \cdot & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \cdot & \cdot & \sigma_{33} & \sigma_{34} \\ \cdot & \cdot & \cdot & \sigma_{44} \end{pmatrix}$$

Pode verificar-se que a matriz tem 10 elementos para determinar que representam as variâncias (σ_{ii}) e as covariâncias (σ_{ij}) entre os erros. Em geral pode dizer-se que as matrizes de covariância terão $J(J+1)/2$ elementos distintos a determinar.

Como o *nível de utilidade* não é relevante, apenas se necessitam saber as diferenças entre utilidades e Train (2003) recomenda realiza-las a respeito da primeira alternativa, ou seja, realizar a diferença de cada utilidade em relação à utilidade da primeira alternativa considerada. Assim as diferenças entre erros serão:

$$\tilde{\varepsilon}_{nj1} = \varepsilon_{nj} - \varepsilon_{n1}; j = 2, 3, 4$$

E o vetor de diferenças de erros será:

$$\langle \tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n31}, \tilde{\varepsilon}_{n41} \rangle$$

Salienta-se que os erros são em relação à 1ª alternativa e não para a 1ª alternativa.

A matriz de covariância passará a ser:

$$\tilde{\Omega}_1 = \begin{pmatrix} \theta_{22} & \theta_{23} & \theta_{24} \\ & \theta_{33} & \theta_{34} \\ & & \theta_{44} \end{pmatrix}$$

Onde os erros θ se relacionam com os erros originais σ da seguinte forma:

$$\begin{aligned}
\theta_{22} &= \sigma_{22} + \sigma_{11} - 2.\sigma_{12}, \\
\theta_{33} &= \sigma_{33} + \sigma_{11} - 2.\sigma_{13}, \\
\theta_{44} &= \sigma_{44} + \sigma_{11} - 2.\sigma_{14}, \\
\theta_{23} &= \sigma_{23} + \sigma_{11} - \sigma_{12} - \sigma_{13}, \\
\theta_{24} &= \sigma_{24} + \sigma_{11} - \sigma_{12} - \sigma_{14}, \\
\theta_{34} &= \sigma_{34} + \sigma_{11} - \sigma_{13} - \sigma_{14}
\end{aligned}$$

Esta transformação pode ser obtida utilizando o procedimento da matriz M explicada no anexo 5 do Capítulo III, $\tilde{\Omega}_1 = M_1 \Omega M_1$.

Para definir a *escala da utilidade*, um dos elementos da diagonal da matriz anterior tem que ser normalizado (dividindo todos os elementos da matriz por esse elemento, ele próprio fica igual à unidade). Uma opção será o elemento do canto superior esquerdo, originando a seguinte matriz (Train, 2003; Ortúzar e Willumsen, 2001):

$$\tilde{\Omega}_1^* = \begin{pmatrix} 1 & \theta_{23}^* & \theta_{24}^* \\ & \theta_{33}^* & \theta_{34}^* \\ & & \theta_{44}^* \end{pmatrix}$$

Onde os erros θ^* se relacionam com os erros originais σ da seguinte forma:

$$\begin{aligned}
\theta_{33}^* &= \frac{\sigma_{33} + \sigma_{11} - 2.\sigma_{13}}{\sigma_{22} + \sigma_{11} - 2.\sigma_{12}}, \\
\theta_{44}^* &= \frac{\sigma_{44} + \sigma_{11} - 2.\sigma_{14}}{\sigma_{22} + \sigma_{11} - 2.\sigma_{12}}, \\
\theta_{23}^* &= \frac{\sigma_{23} + \sigma_{11} - \sigma_{12} - \sigma_{13}}{\sigma_{22} + \sigma_{11} - 2.\sigma_{12}}, \\
\theta_{24}^* &= \frac{\sigma_{24} + \sigma_{11} - \sigma_{12} - \sigma_{14}}{\sigma_{22} + \sigma_{11} - 2.\sigma_{12}}, \\
\theta_{34}^* &= \frac{\sigma_{34} + \sigma_{11} - \sigma_{13} - \sigma_{14}}{\sigma_{22} + \sigma_{11} - 2.\sigma_{12}}
\end{aligned}$$

Assim, pode constatar-se que a matriz de covariância fica com apenas 5 elementos em comparação com os 10 originais. Cada θ^* é função dos σ originais, mas uma vez que apenas há 5 equações (os θ^*) com 10 variáveis para estimar (os σ) o sistema é impossível de resolver para todas as σ a partir das θ^* , ou seja, não é possível estimar todas as σ . De uma forma geral, um modelo normalizado com J alternativas e uma matriz de covariância não restringida terá $[(J-1)J/2]-1$ parâmetros de covariância e, portanto, o mesmo número de parâmetros identificáveis. O que à partida parece uma restrição do modelo é na realidade uma consequência da normalização, uma vez que os parâmetros que não podem ser estimados apenas permitem variâncias e covariâncias que se devem à escala e ao nível da utilidade e portanto não têm efeito no comportamento. Dito de outra forma, as únicas variâncias e

covariâncias que contêm informação relevante sobre os erros independentemente da escala e nível da utilidade são as que são possíveis de estimar e apenas esse número de parâmetros pode ser identificável (Train, 2003). Às restantes variância e covariâncias devem ser aplicadas restrições (geralmente dando-lhes um determinado valor).

Anexo 7 do Capítulo III

Metodologia para a definição de padrões de substituição com e sem restrições (Train, 2003)

1. Covariância completa ou Padrões de substituição sem restrições

Uma matriz completa para os componentes não observados (erros) da utilidade tem a forma Ω e a sua normalização em relação à escala e nível torna-a em Ω^*_1 . A estimação dos elementos da matriz normalizada permite representar qualquer padrão de substituição, pois a normalização da escala e nível não restringe a sua aplicabilidade, apenas elimina alguns aspetos irrelevantes para o comportamento. No entanto, os parâmetros estimados θ^* não permitem interpretar as reais relações entre as variâncias e covariâncias originais σ . Portanto, a estimação da matriz de covariância completa permite ao modelo representar qualquer padrão de substituição mas os parâmetros estimados θ^* são ininterpretáveis.

2. Covariância estruturada ou Padrões de substituição restringidos

A imposição de uma estrutura para a matriz de covariância torna a estimação mais interpretável, no entanto, reduz a sua capacidade de representar outros padrões de substituição. Ainda assim, se a estrutura representar realmente o comportamento do decisor, mesmo com a matriz de covariância restringida, o modelo poderá simular os reais padrões de substituição, desde que a estrutura seja apropriada para a situação específica a modelar. Como exemplo, apresenta-se uma situação adaptada de Train (2003) para um problema sobre transportes com quatro alternativas. Considere-se um modelo sobre a escolha do local de estacionamento: um com uma tarifação crescente no tempo e três com tarifas fixas. A parte não observada da utilidade é constituída por duas componentes, uma relacionada com a disponibilidade para pagar designada por r_n comum a todas as alternativas e uma outra relacionada com todos os outros fatores não observados designada por η_{nj} , resultando na expressão:

$$\varepsilon_{nj} = -r_n \cdot d_j + \eta_{nj}$$

Onde $d_j = 1$ para tarifas variáveis e 0 para as tarifas fixas, e o seu sinal negativo indica que o aumento da tarifação implica uma diminuição na utilidade local de estacionamento. Assumindo que r_n tem distribuição normal entre os condutores com variância σ , e que $\eta_{nj} \forall j$ é normalmente IID com média igual a zero e variância ω , então a matriz de covariância para $\varepsilon_n = \langle \varepsilon_{n1}, \varepsilon_{n2}, \varepsilon_{n3}, \varepsilon_{n4} \rangle$ será:

$$\Omega = \begin{pmatrix} 0 & 0 & 0 & 0 \\ . & \sigma & \sigma & \sigma \\ . & . & \sigma & \sigma \\ . & . & . & \sigma \end{pmatrix} + \omega \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ . & 1 & 0 & 0 \\ . & . & 1 & 0 \\ . & . & . & 1 \end{pmatrix}$$

O modelo já está normalizado em relação ao nível da utilidade mas necessita de normalização em relação à escala. Assim, a matriz de covariância das diferenças entre erros será:

$$\tilde{\Omega}_1^* = \begin{pmatrix} \sigma & \sigma & \sigma \\ . & \sigma & \sigma \\ . & . & \sigma \end{pmatrix} + \omega \cdot \begin{pmatrix} 2 & 1 & 1 \\ . & 2 & 1 \\ . & . & 2 \end{pmatrix}$$

Como a matriz não tem menos parâmetros que Ω , indica que a matriz está normalizada para o nível (Train, 2003), pelo que para a normalizar em relação à escala se define $\sigma + 2 \cdot \omega = 1$, tornando a matriz de covariância em:

$$\tilde{\Omega}_1^* = \begin{pmatrix} 1 & \theta & \theta \\ . & 1 & \theta \\ . & . & 1 \end{pmatrix}$$

Onde $\theta = (\sigma + \omega) / (\sigma + 2 \cdot \omega)$. Os parâmetros σ e ω não são estimáveis a partir de θ , mas este fornece alguma informação sobre a variância na utilidade devido à disponibilidade para pagar relativamente à variância devido a todos os restantes fatores não observados. Como exemplo, se θ for estimado com o valor 0,75 poderá ser interpretado como uma indicação de que a variância na utilidade devido à disponibilidade para pagar é o dobro da variância na utilidade atribuível aos restantes fatores não observados, pois:

$$\begin{aligned} \theta = 0,75 &\Leftrightarrow \frac{\sigma + \omega}{\sigma + 2 \cdot \omega} = 0,75 \Leftrightarrow \sigma + \omega = 0,75 \cdot \sigma + 1,5 \cdot \omega \Leftrightarrow 0,25 \cdot \sigma = 0,5 \cdot \omega \Leftrightarrow \\ &\Leftrightarrow \sigma = 2 \cdot \omega \end{aligned}$$

Dito de outra forma, $\theta=0,75$, significa que a disponibilidade para pagar o estacionamento tem um peso de 2/3 na variância dos fatores não observados da utilidade, ou que a variância da disponibilidade para pagar é de $\sigma=2$. Como o modelo original já estava normalizado em relação ao nível, não houve necessidade de o normalizar com a diferença entre erros, enquanto a normalização da escala poderia igualmente ter sido feita com $\omega=1$ na matriz de covariância original (Train, 2003).

Anexo 8 do Capítulo III

Metodologia a inclusão de dados recolhidos ao longo do tempo por preferências declaradas (Train, 2003)

Se um indivíduo enfrenta uma escolha entre J alternativas em cada T períodos temporais ou situações, e se as alternativas puderem mudar no tempo, então J e T podem divergir para diferentes decisores. Assim, a utilidade que um indivíduo n obtém da alternativa j no período t será:

$$U_{njt} = V_{njt} + \varepsilon_{njt}$$

Geralmente é expectável que os erros sejam correlacionados ao longo do tempo assim como com outras alternativas, uma vez que os fatores não observados podem persistir no tempo. Se o vetor dos erros para todas as alternativas ao longo de todos os períodos for:

$$\varepsilon_n = \langle \varepsilon_{n11}, \dots, \varepsilon_{nJ1}, \varepsilon_{n12}, \dots, \varepsilon_{nJ2}, \dots, \varepsilon_{n1T}, \dots, \varepsilon_{nJT} \rangle$$

Então a sua matriz de covariância será Ω com dimensão $JT \times JT$.

Se a sequência de alternativas, uma para cada período T , for $i = \{i_1, \dots, i_T\}$, a probabilidade de um indivíduo realizar essa sequência de escolhas será:

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni,t} > U_{nj,t} \forall j \neq i_t, \forall t) = \text{Prob}(V_{ni,t} + \varepsilon_{ni,t} > V_{nj,t} + \varepsilon_{nj,t} \forall j \neq i_t, \forall t) = \\ &= \int_{\varepsilon_n \in B_n} \varphi(\varepsilon_n) \cdot d\varepsilon_n \end{aligned}$$

Onde $B_n = \{\varepsilon_n: V_{ni,t} + \varepsilon_{ni,t} > V_{nj,t} + \varepsilon_{nj,t} \forall j \neq i_t, \forall t\}$ e $\varphi(\varepsilon_n)$ é a densidade normal conjunta com media igual a zero e variância Ω . Comparando o integral com o de apenas uma situação de escolha, apenas se expandiu para que tivesse JT dimensões em vez das J .

Por vezes é mais conveniente trabalhar com as diferenças de utilidade, onde a probabilidade da sequência i acontecer é a probabilidade de todas as diferenças de utilidades para cada alternativa nesse espaço temporal serem negativas, quando calculadas em relação à alternativa i desse período. Isto é:

$$P_{ni} = \text{Prob}(\tilde{U}_{nji,t} < 0 \forall j \neq i_t, \forall t) = \int_{\tilde{\varepsilon}_n \in \tilde{B}_n} \varphi(\tilde{\varepsilon}_n) \cdot d\tilde{\varepsilon}_n$$

Onde:

$$\tilde{U}_{nji_t} = U_{njt} - U_{ni_t}$$

$$\tilde{\varepsilon}_n' = \langle (\varepsilon_{n11} - \varepsilon_{ni_11}), \dots, (\varepsilon_{nJ1} - \varepsilon_{ni_11}), \dots, (\varepsilon_{n1T} - \varepsilon_{ni_1T}), \dots, (\varepsilon_{nJT} - \varepsilon_{ni_1T}) \rangle$$

Este integral tem dimensão $(J-1).T$, e a densidade $\phi(\tilde{\varepsilon}_n)$ é normal e calculado a partir de Ω conjuntamente com a matriz de covariância.

Anexo 9 do Capítulo III

Exemplo do método de simulação GHK simples e com estimação da máxima verosimilhança para a probabilidade de escolha (Train, 2003)

1. Simulador GHK

O simulador GHK funciona com diferenças de utilidade, começando a simulação da probabilidade P_{ni} subtraindo a utilidade da alternativa i às utilidades de cada uma das restantes alternativas. É importante salientar que a utilidade de uma alternativa diferente é subtraída dependendo da probabilidade que se pretende estimar: para P_{ni} , U_{ni} é subtraída de todas as outras utilidades, enquanto para P_{nj} , U_{nj} é que é subtraída. Este aspeto é crítico para a implementação da simulação. Como exemplo apresenta-se a descrição do procedimento com três alternativas para que a diferença de utilidades possa ser representada graficamente.

Considere-se um modelo com a especificação comportamental em utilidades habitual $U_{nj} = V_{nj} + \varepsilon_{nj}$, com $j=1, 2, 3$, e o vetor dos erros com distribuição normal com média zero e covariância Ω . Assuma-se que o modelo já foi normalizado na escala e nível de utilidade, de modo a que os parâmetros de Ω sejam identificáveis (Ω pode ser uma função paramétrica dos dados por exemplo para atender a variações de gosto). Se se pretender simular a probabilidade da 1ª alternativa P_{n1} reescrevemos o modelo em diferenças de utilidade subtraindo a utilidade da alternativa 1:

$$U_{nj} - U_{n1} = (V_{nj} - V_{n1}) + (\varepsilon_{nj} - \varepsilon_{n1}) \Leftrightarrow \tilde{U}_{nj1} = \tilde{V}_{nj1} + \tilde{\varepsilon}_{nj1}; \text{ com } j = 2, 3$$

O vetor $\tilde{\varepsilon}_{nj1} = \langle \tilde{\varepsilon}_{n21}, \tilde{\varepsilon}_{n31} \rangle$ tem distribuição normal com média igual a zero e covariância $\tilde{\Omega}_1$ derivada de Ω . Se L_1 for o fator de Choleski¹⁰ (ou Cholesky) de $\tilde{\Omega}_1$, que tem dimensão 2x2 no exemplo, então este será uma matriz diagonal inferior com a forma:

$$L_1 = \begin{pmatrix} c_{aa} & 0 \\ c_{ab} & c_{bb} \end{pmatrix}$$

Então as diferenças entre erros originais que são correlacionados podem ser expressas como uma função linear de desvios padrão normais não correlacionados:

¹⁰ - Fator de Choleski ou Cholesky: O fator Choleski de uma matriz simétrica positiva A é L tal que $L \cdot L^T = A$, ou seja, L é uma matriz triangular inferior com diagonal positiva. No caso da matriz a decompor ser Ω , o fator de Choleski também é designado raiz quadrada generalizada de Ω ou desvio padrão generalizado de ε . Sendo $K = 1$ (número de erros a estimar) e a variância igual a s^2 , o fator de Choleski é s , ou seja, o desvio padrão do erro ε .

$$\begin{aligned}\tilde{\varepsilon}_{n21} &= c_{aa} \cdot \eta_1 \\ \tilde{\varepsilon}_{n31} &= c_{ab} \cdot \eta_1 + c_{bb} \cdot \eta_2\end{aligned}$$

Onde η_1 e η_2 têm distribuição normal *IID* $N(0,1)$. Ambos os erros $\tilde{\varepsilon}_{n21}$ e $\tilde{\varepsilon}_{n22}$ estão correlacionados pois ambos dependem de η_1 . Assim as utilidades serão expressas como:

$$\begin{aligned}\tilde{U}_{n21} &= \tilde{V}_{n21} + c_{aa} \cdot \eta_1 \\ \tilde{U}_{n31} &= \tilde{V}_{n31} + c_{ab} \cdot \eta_1 + c_{bb} \cdot \eta_2\end{aligned}$$

A probabilidade de escolha da alternativa 1 será uma função unidimensional da distribuição cumulativa normal padrão:

$$\begin{aligned}P_{n1} &= \text{Prob}(\tilde{U}_{n21} < 0 \text{ e } \tilde{U}_{n31} < 0) = \text{Prob}(\tilde{V}_{n21} + c_{aa} \cdot \eta_1 < 0 \text{ e } \tilde{V}_{n31} + c_{ab} \cdot \eta_1 + c_{bb} \cdot \eta_2 < 0) = \\ &= \text{Prob}(\tilde{V}_{n21} + c_{aa} \cdot \eta_1 < 0) \times \text{Prob}(\tilde{V}_{n31} + c_{ab} \cdot \eta_1 + c_{bb} \cdot \eta_2 < 0 \mid \tilde{V}_{n21} + c_{aa} \cdot \eta_1 < 0) = \\ &= \text{Prob}\left(\eta_1 < -\frac{\tilde{V}_{n21}}{c_{aa}}\right) \times \text{Prob}\left(\eta_2 < -\frac{(\tilde{V}_{n31} + c_{ab} \cdot \eta_1)}{c_{bb}} \mid \eta_1 < -\frac{\tilde{V}_{n21}}{c_{aa}}\right) = \\ &= \Phi\left(-\frac{\tilde{V}_{n21}}{c_{aa}}\right) \times \int_{\eta_1=-\infty}^{-\frac{\tilde{V}_{n21}}{c_{aa}}} \Phi\left(-\frac{(\tilde{V}_{n31} + c_{ab} \cdot \eta_1)}{c_{bb}}\right) \cdot \phi(\eta_1) \cdot d\eta_1\end{aligned}$$

Onde $\Phi(\cdot)$ é a distribuição cumulativa normal padrão calculada para o ponto entre parêntesis e $\phi(\cdot)$ é a densidade normal truncada¹¹. A primeira parcela calcula-se como função cumulativa normal padrão no ponto $-(\tilde{V}_{n21}/c_{aa})$ recorrendo a programas informáticos de estatística. A segunda parcela é um integral que os programas estatísticos não conseguem integrar pelo que se tem de usar a simulação para uma aproximação. É este o aspeto mais relevante do procedimento GHK, simular a aproximação ao integral em P_{n1} (Train, 2003).

O procedimento de simulação será o seguinte atendendo a que se trata de um integral sobre uma variável com distribuição normal truncada de η_{21} até $-(\tilde{V}_{n21}/c_{aa})$:

1. Calcula-se $k = \Phi(-\tilde{V}_{n21}/c_{aa})$
2. Retira-se aleatoriamente um valor para η_1 , designado por η_1^r , de uma distribuição normal padrão truncada em $-\tilde{V}_{n21}/c_{aa}$, que se consegue da seguinte forma:
 - 2.1. Retira-se aleatoriamente um valor de uma distribuição normal padrão e uniforme designado por μ^r

¹¹ - Distribuição normal truncada, significa uma distribuição de probabilidades de uma variável aleatória com distribuição normal, com limite superiores e/ou inferiores.

2.2. Calcula-se $\eta_1^r = \Phi^{-1}\left(\mu^r \cdot \Phi\left(-\frac{\tilde{V}_{n21}}{c_{aa}}\right)\right)$

3. Calcula-se $g^r = \Phi\left(-\frac{(\tilde{V}_{n31} + c_{ab} \cdot \eta_1^r)}{c_{bb}}\right)$

4. A probabilidade simulada para este valor aleatório será: $\tilde{p}_{n1}^r = k \times g^r$

5. Repetem-se os passos anteriores R vezes e faz-se a média dos resultados. Será esta

média que simulará a probabilidade: $\tilde{p}_{n1} = \left(\frac{1}{R}\right) \sum_r \tilde{p}_{n1}^r$

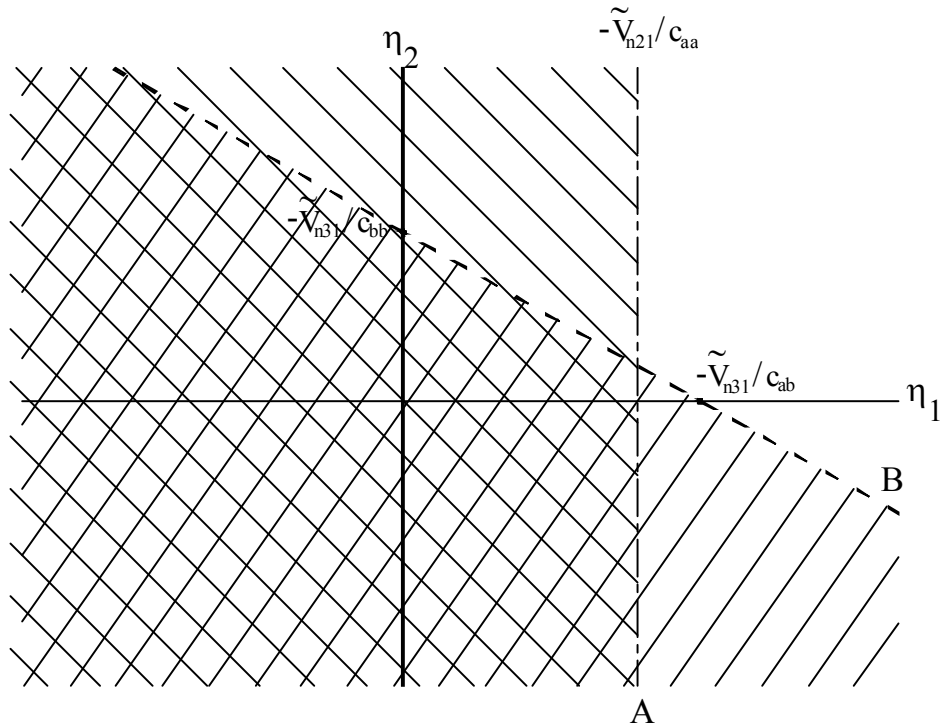


Figura A. 1 - Exemplo da probabilidade da alternativa 1 no espaço dos erros independentes η_1 e η_2 (adaptado de Train, 2003)

Na figura A.1 apresenta-se uma possível representação da probabilidade de escolha da alternativa 1 no espaço dos erros independentes η_1 e η_2 . O eixo dos XX' tem os valores de η_1

e o dos YY' os valores de η_2 . A linha A representa a primeira condição $\eta_1 = -\frac{\tilde{V}_{n21}}{c_{aa}}$, e a linha B

a condição $\eta_2 = -\frac{(\tilde{V}_{n31} + c_{ab} \cdot \eta_1)}{c_{bb}}$. Como o espaço de possíveis soluções é menor que η_1 e η_2 ,

será a densidade de massa da área preenchida com quadriculado que definirá a probabilidade de escolha da alternativa 1. Note-se que a linha B intercepta o eixo do YY' quando $\eta_1 = 0 \Rightarrow$

$\eta_2 = -\frac{\tilde{V}_{n31}}{c_{bb}}$, e o eixo do XX' quando $\eta_2 = 0 \Rightarrow \eta_1 = -\frac{\tilde{V}_{n31}}{c_{ab}}$, e que tem inclinação $\frac{\partial \eta_2}{\partial \eta_1} = -\frac{c_{ab}}{c_{bb}}$. A

probabilidade é a massa da área quadriculada multiplicada pela proporção dessa faixa que está abaixo da linha B.

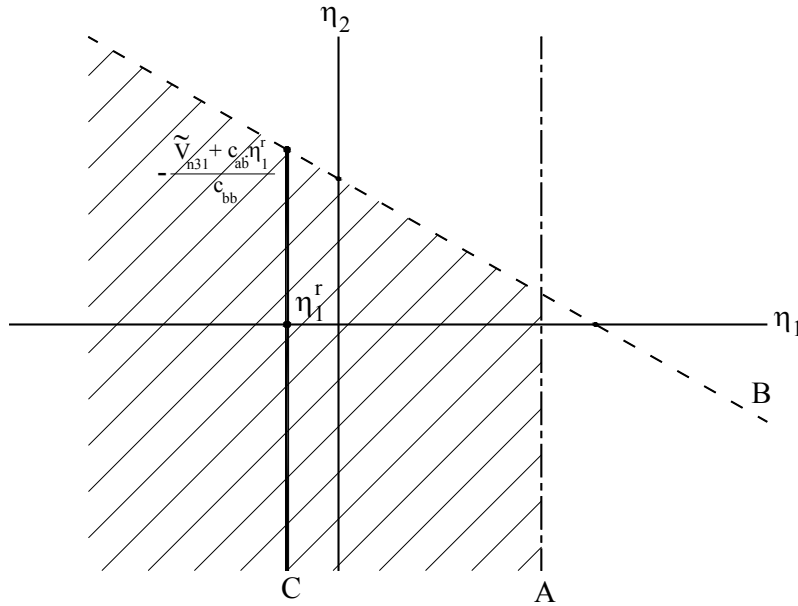


Figura A. 2 - Probabilidade de η_2 estar no limite correto atendendo a η_2^r (adaptado de Train, 2003)

Considerando a figura A.2, a área sombreada à esquerda da linha A tem massa igual $\Phi\left(-\frac{\tilde{V}_{n21}}{c_{aa}}\right)$ e quando η_1 toma o valor de η_1^r a probabilidade de η_2 estar abaixo da linha B é a

proporção de massa da linha C abaixo da linha B, que é calculada através de $\Phi\left(-\frac{\tilde{V}_{n31} + c_{ab} \cdot \eta_1^r}{c_{bb}}\right)$. A porção da área sombreada abaixo da linha B será a média dos valores

de $\Phi\left(-\frac{\tilde{V}_{n31} + c_{ab} \cdot \eta_1^r}{c_{bb}}\right)$ para os valores de η_1 à esquerda da linha A, que pode ser simulada

através da escolha aleatória de vários η_1 com valores à esquerda de A, calculando para todos a sua proporção abaixo de B e calculando a sua média. A probabilidade de escolha da alternativa 1 será a multiplicação desta média pela massa da área à esquerda de A

$$\Phi\left(-\frac{\tilde{V}_{n21}}{c_{aa}}\right).$$

A generalização do modelo para mais alternativas é relativamente simples e pode ser encontrada em Train (2003) da página 130 à 133.

2. Simulador GHK com Estimação da Máxima Verosimilhança

Existem vários aspetos a atender quando se utiliza o simulador GHK com estimação da máxima verosimilhança (Train, 2003):

- A função de verosimilhança logarítmica utiliza a probabilidade de um indivíduo escolher uma alternativa e como indivíduos diferentes fazem escolhas diferentes, existe a necessidade de calcular as probabilidades para as diferentes alternativas. O simulador GHK, no entanto, calcula a diferença entre as utilidades em relação à alternativa em análise, implicando o cálculo das diferenças de utilidades em relação a todas as alternativas a estudar.
- A matriz de covariância $\tilde{\Omega}_i$ utilizada no simulador GHK varia consoante a alternativa escolhida, todas calculadas a partir da matriz de covariância dos erros originais Ω . Assim há a necessidade de assegurar que os elementos das várias matrizes de covariância $\tilde{\Omega}_i \forall i$ sejam consistentes uma vez que derivam da mesma origem.
- É necessário assegurar que os parâmetros estimados por máxima verosimilhança impliquem matrizes de covariância $\tilde{\Omega}_i \forall i$ que sejam definidas positivas (análoga a uma matriz simétrica positiva de números reais).
- E também será necessário assegurar que o modelo esteja normalizado para a escala e nível da utilidade para que os parâmetros sejam identificáveis (isto é, estimáveis).

Train (2003) sugere o seguinte procedimento. Para assegurar que o modelo é identificável começa-se pela matriz de covariância das diferenças de utilidades escalonadas com as diferenças realizadas em função da primeira alternativa (canto superior esquerdo igual à unidade e o restos dos elementos divididos pela variância da primeira alternativa) designada por $\tilde{\Omega}_1$. Para assegurar que esta matriz de dimensão $(J-1) \times (J-1)$ é definida positiva parametriza-se o modelo com o seu fator de Choleski, ou seja, com uma matriz triangular inferior $(J-1) \times (J-1)$ com o elemento do canto superior esquerdo igual à unidade:

$$L_1 = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ c_{21} & c_{22} & 0 & \dots & \dots & 0 \\ c_{31} & c_{32} & c_{33} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Os elementos c_{lk} deste fator são os parâmetros identificáveis e portanto estimados pelo modelo. A matriz original dos J erros não diferenciados é criada a partir de L_1 , adicionando-

lhe uma linha superior e uma coluna à esquerda de zeros, resultando no fator de Choleski da matriz de covariância dos erros originais.

$$L = \begin{pmatrix} 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & c_{21} & c_{22} & 0 & \dots & \dots & 0 \\ 0 & c_{31} & c_{32} & c_{33} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Implicando que $\Omega = L.L^T$, permitindo derivar $\tilde{\Omega}_j$ para qualquer j .

Expressando a utilidade em forma de vetor com as várias alternativas empilhadas $U_n = V_n + \varepsilon_n$ e erros com distribuição normal $N(0, \Omega)$, para a função de verosimilhança logarítmica necessitamos do cálculo das P_{nj} . Utilizando a matriz M do anexo 5 do capítulo III as diferenças entre utilidades podem ser calculadas através das seguintes expressões:

$$\begin{aligned}\tilde{U}_{ni} &= M_i \cdot U_n \\ \tilde{V}_{ni} &= M_i \cdot V_n \\ \tilde{\varepsilon}_{ni} &= M_i \cdot \varepsilon_n\end{aligned}$$

E a matriz de covariância das diferenças entre erros $\tilde{\varepsilon}_{ni}$ é calculada como $\tilde{\Omega}_i = M_i \Omega M_i^T$, sendo o seu fator de Cholesky L_i . A utilidade para cada indivíduo é expressa como $\tilde{U}_{ni} = \tilde{V}_{ni} + L_i \eta_n$, onde η_n é um vetor de dimensão $(J-1)$ com desvios IID e distribuição normal padrão. Será a esta expressão que se aplicará o simulador GHK para o cálculo das probabilidades a utilizar na estimação da máxima verosimilhança logarítmica.

Com este procedimento os requisitos enumerados no início do ponto ficam verificados:

- O modelo está normalizado para a escala e nível uma vez que foi definido com recurso ao fator de Choleski L_1 da matriz de covariância de diferenças de erros escalonada $\tilde{\Omega}_1$.
- Cada $\tilde{\Omega}_i$ é consistente com as restantes $\tilde{\Omega}_j$ para $j \neq i$, uma vez que foram determinadas a partir da mesma matriz Ω construída a partir de L_1 .
- Cada matriz $\tilde{\Omega}_i$ é definida positiva para quaisquer valores dos parâmetros pois todas as matrizes que resultem do produto de uma matriz triangular inferior por si mesma são definidas positivas, e $\tilde{\Omega}_1 = L_1.L_1^T$ pelo que é definida positiva. Como todas as restantes matrizes $\tilde{\Omega}_j$ são construídas para serem consistentes com $\tilde{\Omega}_1$, são também definidas positivas.

Anexo 10 do Capítulo III

Caracterização das distribuições mais comuns para a estimação de modelos Logit Mistos

1. Distribuição Uniforme

Uma variável com distribuição uniforme tem um desvio padrão entre a (*média-s*) e (*média+s*) de (Hensher *et al.*, 2005):

$$\sigma = \frac{(\text{Limite superior} - \text{Limite inferior})}{2\sqrt{3}} = \frac{s}{\sqrt{3}}$$

A estratégia para simular uma extração do parâmetro com esta distribuição é:

$$(\text{Estimação da média do parâmetro} + s \cdot (2 \cdot u - 1))$$

Onde u é a variável aleatória com distribuição uniforme padrão (entre 0 e 1). Como a sua distribuição é uniforme entre 0 e 1, $2 \cdot u - 1$ é uniforme entre -1 e +1 que multiplicado por s dá uma distribuição uniforme entre $-s$ e $+s$ da média. A amplitude s (“spread”) pode ser calculado através do desvio padrão multiplicado por $\sqrt{3}$.

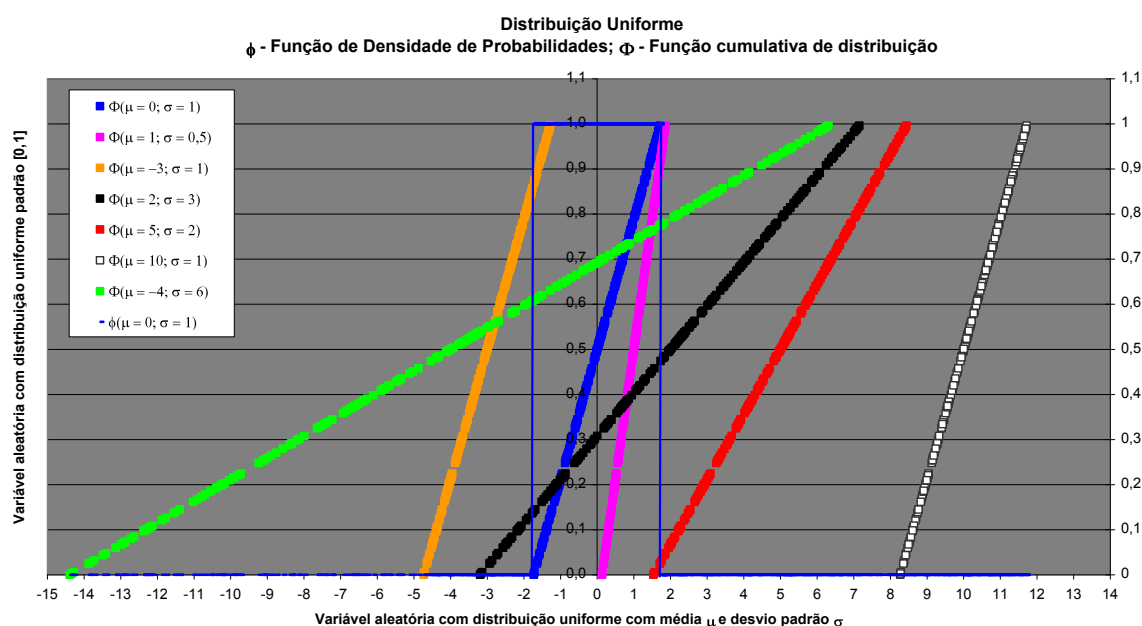


Figura A. 3 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição uniforme, para várias médias e desvios padrão.

2. Distribuição Triangular

Numa distribuição triangular a função de densidade de probabilidades assemelha-se a uma tenda com um pico no centro e com declives lineares de ambos os lados (neste caso será uma distribuição triangular simétrica). Se c for o centro e s a amplitude, a densidade começa em $c-s$ aumenta linearmente até ao seu valor máximo em c e decresce linearmente até zero no ponto $c+s$. A função será nula abaixo de $c-s$ e acima de $c+s$, tem média e moda igual a c (no caso de não ser simétrica o ponto c será a moda), e o seu desvio padrão será a amplitude dividida por $\sqrt{6}$, ou seja, a amplitude s será o desvio padrão multiplicado por $\sqrt{6}$. A altura no pico da função será $1/s$, para que ambos os lados do gráfico tenham área igual a $0,5 \times s \times 1/s = 1/2$, ou seja, um total de 1 para a função densidade. A inclinação da função terá o valor de $1/s^2$.

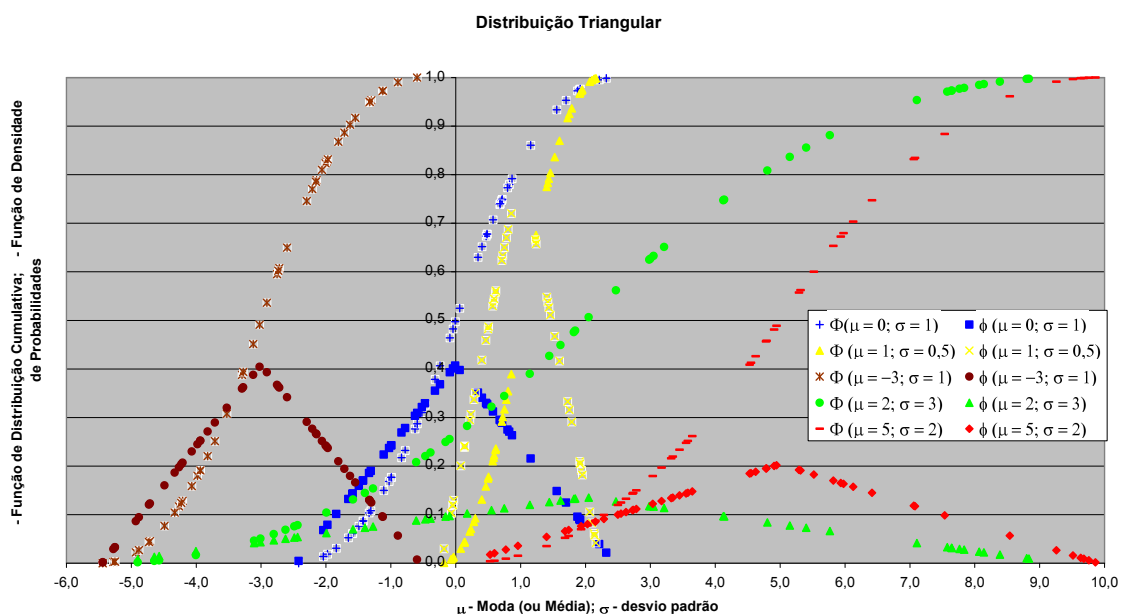


Figura A. 4 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição triangular (simétricas), para várias médias/modas e desvios padrão.

Considerando um exemplo com moda/média (usa-se a média no caso de uma distribuição simétrica) $\bar{x}$, um valor mínimo $min.=\bar{x}-s$ e máximo $max.=\bar{x}+s$, a distribuição cumulativa num ponto z pode ser definida como:

$$\begin{cases} P(x < z) = \frac{(z - \text{Min})^2}{(\bar{x} - \text{Min})(\text{Max} - \text{Min})}; \text{ se } z \leq \bar{x} \\ P(x < z) = 1 - P(x > z) = \frac{(\text{Max} - z)^2}{(\text{Max} - \bar{x})(\text{Max} - \text{Min})}; \text{ se } z > \bar{x} \end{cases}$$

3. Distribuição Lognormal

Esta distribuição é bastante popular, porque atendendo ao Teorema do Limite Central (TLC), se uma variável aleatória resulta da soma de um elevado número de efeitos aleatórios provocados por causas independentes, positivas e negativas, a distribuição dessa variável tenderá a tornar-se normal à medida que o número de observações aumenta. No entanto, se essas observações afetarem o atributo de forma multiplicativa, alterando o valor de x em proporções aleatoriamente distribuídas em vez de quantidades absolutas, então a aplicação do TLC aplicado a $y = \ln(x)$ tenderá a apresentar uma distribuição normal (Hensher *et al.*, 2005; Guimarães e Sarsfield Cabral, 1997). Este tipo de distribuição é apelativa uma vez que está limitada a um domínio de valores positivos, mas apresenta geralmente uma cauda para a direita bastante longa o que pode ser desvantajoso em determinadas situações, como a *disponibilidade para pagar*, devido à grande proporção de valores “não razoáveis” que permite o que põe em causa a sua adequação em situações mais específicas. Rigby e Burton (2006) referem ainda a sua desadequação para situações de probabilidade de escolha nula, por exemplo para modelar situações em que parte da população é indiferente a um determinado atributo enquanto o resto da população tem uma avaliação positiva ou negativa em relação ao mesmo.

Se uma variável y tiver uma distribuição normal $N(\mu_y, \sigma_y^2)$, então x terá uma distribuição Lognormal $LN(\mu_y, \sigma_y^2)$, cujas função densidade de probabilidade e função de distribuição cumulativa serão (Montgomery e Runger, 2007):

$$f(x) = \frac{1}{x \cdot \sigma_y \cdot \sqrt{2 \cdot \pi}} \cdot e^{\left[-\frac{(\ln x - \mu_y)^2}{2 \cdot \sigma_y^2} \right]}; \text{ com } 0 < x < \infty$$

$$F(x) = \Phi \left[\frac{\ln(x) - \mu_y}{\sigma_y} \right]; \text{ que tem a particularidade de } F(x) = 0 \text{ para } x \leq 0$$

Hensher *et al.* (2005) refere que na estimação de parâmetros, evidências empíricas sugerem que a consideração de um atributo na expressão da utilidade associado a um parâmetro aleatório de distribuição lognormal que se espera que venha a ter uma média estimada de sinal negativo, resulta geralmente num modelo que não converge ou que o faz com estimação de médias exageradamente grandes. Em situações destas recomendam que se inverta o sinal do atributo antes da estimação, ou seja, que se defina o atributo com valores negativos em vez de impor uma inversão do sinal do parâmetro estimado. Um parâmetro lognormal positivo para o negativo de um atributo é o mesmo que o parâmetro lognormal negativo desse mesmo atributo.

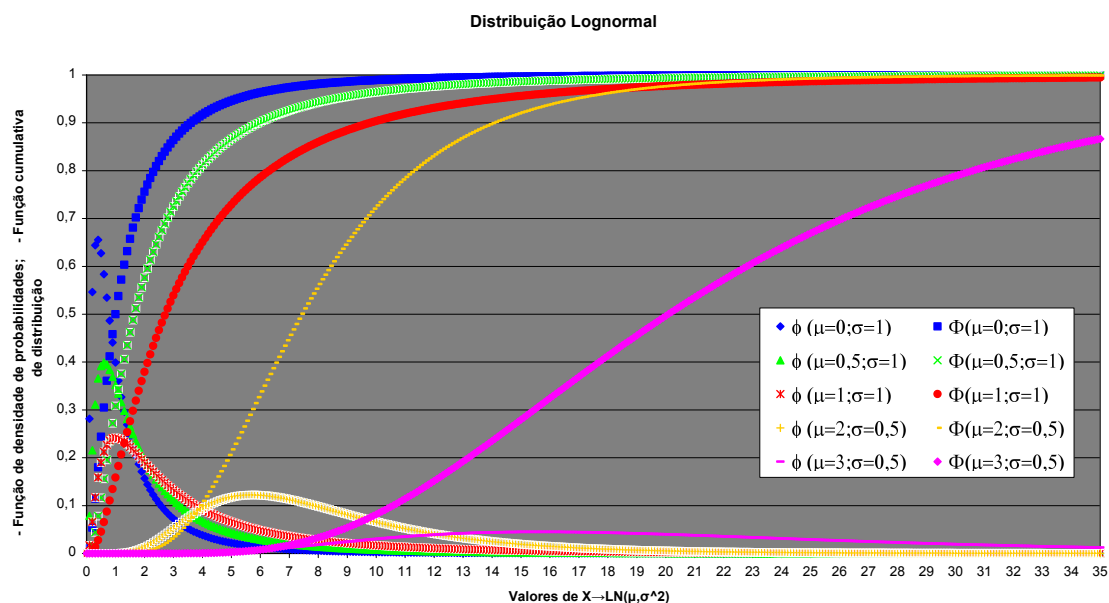


Figura A. 5 - Exemplos de funções de distribuição cumulativa de variáveis aleatórias com funções de densidade de distribuição Lognormal, para várias médias e desvios padrão.

4. Imposição de Restrições às Distribuições

Todas as distribuições têm pontos fortes e pontos fracos, estando as fraquezas geralmente associadas ao desvio padrão e à amplitude, que nos seus pontos extremos podem incluir alterações não aceitáveis de sinal nas distribuições simétricas, caudas demasiado longas, etc. Um solução sugerida por Hensher *et al.* (2005) é condicionar a amplitude ou o desvio padrão de determinada distribuição da cada parâmetro aleatório como uma função da sua média, sugerindo-se uma expressão do seguinte tipo para distribuições normais:

$$\theta_{q,n} = \theta_n + \sigma.v_{q,n}$$

Onde $v_{q,n}$ é a variável aleatória. Se a restrição aplicada for que o desvio padrão deverá ser igual à média, a expressão será:

$$\theta_{q,n} = \theta_n + \theta.v_{q,n}$$

Outro exemplo poderá ser a imposição de que o desvio padrão seja o produto do coeficiente de variação z (que pode ter qualquer valor positivo) pela média:

$$\theta_{q,n} = \theta_n + z.\theta.v_{q,n}$$

Será conveniente que z varie entre 0 e 1 para que o desvio padrão não seja superior à média, o que geralmente leva a comportamentos inaceitáveis dos parâmetros estimados.

Hensher *et al.* (2003) referem que na maioria dos estudos empíricos se têm obtido simulações da média e de variância comparáveis para as distribuições normal, uniforme e triangular. Apenas na distribuição Lognormal estes valores variam de forma substancial, mas referem que é possível verificar que a média das distribuições normais, uniformes e triangulares é geralmente balizada pela média e moda da distribuição Lognormal.

5. Comparação de Distribuições Empíricas com Distribuições Analíticas

Escolher uma distribuição analítica que tenha propriedades comportamentais desejáveis não é tarefa fácil, uma vez que distribuições reais bi ou multimodais não permitem a utilização das distribuições mais populares (Hensher *et al.*, 2003). Esta incerteza leva a que uma abordagem empírica possa ser útil, implicando que se estabeleça uma estimacão única (média) para os parâmetros de cada indivíduo da amostra e que se defina a forma da distribuição (o calculo do desvio padrão ou da amplitude não é suficiente para a definição da sua forma). O estabelecimento da real distribuição de forma empírica é assim um grande desafio devido ao enviesamento que pode existir nos dados reais (quer revelados, quer declarados). Dados que não tenham situações suficientemente diferentes de forma a abranger o espectro de possíveis condições, levantam problemas na eficiente estimacão assintótica dos parâmetros.

Hensher *et al.* (2003) sugerem a utilização de um estimador de densidade Kernel para descrever a distribuição de um parâmetro de forma não parametrizada, ou seja, sem nenhuma assunção para a distribuição analítica subjacente. Este estimador é uma modificação de um histograma usado para descrever graficamente a distribuição de uma amostra de observações. A densidade Kernel é mais vantajosa que os histogramas pois assume uma distribuição contínua ao contrário do histograma que as assume descontínuas, e ao contrário dos histogramas a sua forma não depende do tamanho e localização dos limites das gamas de valores assumidos.

Uma das principais decisões a tomar implica o tamanho de cada categoria de valores. Categorias muito estreitas melhoram o ajustamento da função à real distribuição mas implica a possibilidade de existirem categorias com poucos ou nenhuns dados tornando o histograma numa função imprevisível e incorreta. O estimador de densidade Kernel tem uma forma suavizada uma vez que indica, para cada ponto, a proporção da amostra que está à sua volta. Este estimador é definido por uma função ponderada designada de *função Kernel*, cujas ponderações serão tanto menores quanto mais afastadas estiverem as amostras observadas do ponto selecionado. Para um único atributo, a função Kernel é calculada através:

$$f(z_j) = \frac{1}{n} \cdot \sum_{i=1}^n \frac{K\left[\frac{(z_j - x_i)}{h}\right]}{h}; j = 1, \dots, M$$

A função é calculada para um conjunto específico de M valores de interesse z_j , correspondentes à porção de cada categoria do atributo. Cada valor necessita da soma de toda a amostra de n valores x_i .

O principal componente do cálculo é o núcleo (*Kernel*) ou função ponderadora $K[.]$ que pode tomar diversas formas. Se o núcleo for normal, os pesos variam de $\varpi(0) = 0,399$ quando $x_i = z_j$ até valores próximos de zero quando x_i for muito afastado de z_j , medindo a proporção da amostra de valores que está perto do atributo z_j .

A outra componente importante do cálculo é o parâmetro suavizante ou a largura de banda h , de forma a assegurar uma boa resolução na definição do gráfico. Este parâmetro de largura de banda é análogo à largura das barras nos histogramas, pelo que larguras de banda muito estreitas melhoram a resolução (aumentam o detalhe) mas tornam os estimadores de densidades *kernel* instáveis. Uma expressão comum para esta largura de banda (utilizada inclusivamente no programa *Nlogit*) será (Hensher *et al.*, 2003):

$$h = 0,9 \cdot \frac{Q}{n^{0,2}}; \text{ onde } Q = \min(\text{desvio padrão}, \text{amplitude}/1,5)$$

O número de pontos a calcular tem de ser especificado e o conjunto de categorias z_j (qualquer que seja o número de pontos) será determinado através:

$$z_j = z_L + j \times \left[\frac{(z_U - z_L)}{M} \right]; j = 1, \dots, M \text{ e } z_L = \min(x) - h \text{ até } z_U = \max(x) + h;$$

Esta metodologia produz uma matriz $M \times 2$ em que a primeira coluna contém os valores de z_j e a segunda os valores de $f(z_j)$, cuja representação gráfica é a função de densidade estimada. Para uma melhor interpretação da distribuição analítica que melhor se aproxima da distribuição empírica dos dados reais Hensher *et al.* (2003) sugerem a análise das propriedades das distribuições obtidas com o estimador Kernel, mais concretamente a sua *Kurtose*¹² e o *Coefficiente de Assimetria*¹³.

¹² - *Kurtose* (ou *Curtose*) é um valor que permite avaliar a forma da distribuição. Uma distribuição com um valor elevado da kurtose tem maior concentração de dados perto do valor médio e nas caudas e menor concentração de dados nos intervalos entre essas zonas. Uma menor kurtose significa uma distribuição mais suave, com crescimentos e decrescimentos mais homogêneos ao longo da mesma, ou seja, uma distribuição mais achatada. A kurtose está geralmente associada ao momento centrado de quarta ordem (m_4), podendo recorrer-se a uma estatística em que se subtrai três vezes o quadrado do momento de segunda ordem de uma distribuição normal ($3 \cdot m_2^2$), que corresponde ao momento de quarta ordem de uma distribuição normal. O resultado pode ser interpretado, consoante seja superior, igual ou inferior a zero, como uma kurtose superior, igual ou inferior à da população normal de referência. (Guimarães e Sarsfield Cabral, 1997).

¹³ - *Coefficiente de Assimetria* é calculado como a razão entre o momento centrado de terceira ordem e o cubo do desvio padrão da amostra. É uma medida adimensional da simetria de uma distribuição, sendo igual a zero para uma distribuição simétrica, positivo se for assimétrica à direita e negativo se for assimétrica à esquerda (Guimarães e Sarsfield Cabral, 1997).

Anexo 11 do Capítulo III

Procedimentos empíricos para a identificação de modelos Logit Mistos (Walker, 2002)

Nos exemplos apresentados, o modelo será especificado como:

$$U_n = X_n \cdot \theta + \eta_n + v_n$$

Onde:

U_n : é um vetor de $(J \times 1)$ que corresponde às utilidades das J alternativas disponíveis apreendida pelo indivíduo n ;

X_n : é uma matriz de $(J \times K)$ atributos ou variáveis observadas que descrevem o indivíduo n e as J alternativas disponíveis;

θ : é um vetor de $(K \times 1)$ parâmetros ou coeficientes relacionados com os atributos X_n ;

η_n : é um vetor $(M \times 1)$ que corresponde à parte flexível das perturbações ou erros, semelhante aos erros dos modelos Probit, podendo ou não ter distribuição normal,;

v_n : é um vetor $(J \times 1)$ de variáveis de valores extremos aleatórios IID com parâmetros de localização nulo, parâmetro de escala $\mu > 0$, e com variância igual a $\pi^2 / (6 \cdot \mu^2)$;

J : número de alternativas disponíveis;

K : número de atributos observados incluindo constantes específicas binárias e genéricas, assim como atributos específicos de cada alternativa;

M : número de fatores latentes de distribuição das várias variáveis e os seus correspondentes pesos;

Os parâmetros desconhecidos serão assim θ , μ e os que caracterizam a distribuição de η . Os X_n são observados enquanto η_n e v_n não, e para a determinação da identificação do modelo utiliza-se a matriz de covariância como:

$$\text{Cov}(U_n) = \Omega_n = \text{Cov}(\eta_n) + \left(\frac{\pi^2}{6 \cdot \mu^2} \right) \cdot I_J$$

Onde I_J é uma matriz identidade ($J \times J$) e a matriz de covariância para a diferença entre utilidades é designada por $Cov(\Delta U_n) = \Omega_{n\Delta}$.

1. Bases para a Identificação

Segundo Walker (2002), todas as modelações de escolhas discretas requerem restrições de identificação para poderem ser determinadas, de modo a estimar um conjunto único de parâmetros que maximize a função de verosimilhança. Sem uma solução única os erros padrão dos parâmetros estimados não podem ser determinados e as ferramentas estatísticas utilizadas nos testes de hipótese, por exemplo a estatística *t de Student*, e os intervalos de confiança não estão disponíveis. Existe assim a necessidade de estabelecer restrições ao modelo de modo a que se possa ter uma solução única. No entanto, a quantidade de possíveis conjuntos de restrições a realizar é infinito, pelo será aí que reside a maior dificuldade no problema de identificação. Exemplo destas restrições nos modelos Logit multinomiais seria a definição da sua escala restringindo a variância dos erros ($\mu = 1$ ou $\mu = 2$) ou restringindo um dos parâmetros sistêmicos (um determinado $\theta = 1$ ou $\theta = 2$). Assim, nos modelos Logit mistos também haverá necessidade de escolher onde realizar as restrições, podendo optar-se pelo vetor dos parâmetros θ ou pelos parâmetros das distribuições dos erros η e ν . Walker (2002) sugere que as restrições impostas aos parâmetros θ sejam semelhantes às realizadas para modelos Logit multinomiais padrão e que a escala seja restringida nos erros, permitindo focar a atenção apenas na identificação da distribuição do vetor de erros no modelo misto. Esta abordagem sugere a divisão do modelo Logit misto em duas partes, uma em que o Logit é tratado tradicionalmente e outra mais complexa implícita na definição da distribuição dos erros.

Uma ferramenta geralmente utilizada para identificar determinado modelo é a avaliação da sua matriz de informação, isto é, a matriz de segundas derivadas esperadas da função de verosimilhança logarítmica. No entanto, esta apenas está disponível depois de já se ter imposto regras convencionais de identificação e estimado o modelo, estimação esta computacionalmente complexa. Assim, existe a necessidade de encontrar outras formas mais expeditas para identificação do modelo Logit misto, sugerindo Walker (2002) a avaliação da matriz de covariância da diferença entre utilidades.

O primeiro passo da identificação do modelo pretendido é assumir uma estrutura coerente para a distribuição dos erros ou perturbações, nomeadamente se se pretende que a matriz de covariâncias de diferenças entre utilidades não tenha restrições (tipo modelo Probit) ou que tenha diferentes matrizes de covariância atendendo a heteroscedaticidade ou à hierarquização. Após esta definição existem três verificações a realizar, além da normalização, para que o modelo possa ser identificado e portanto estimável: a condição da sua *ordem* (“*order condition*”); da sua *característica* (“*rank condition*”); e a condição de ser *definida positiva* (“*positive definiteness condition*”).

A verificação da condição de ordem é necessária à identificação do modelo mas a da sua característica é suficiente, uma vez que se a última é verificada a primeira automaticamente também o é (Walker, 2002). O facto de a condição de ordem ser de fácil verificação justifica mesmo assim a sua aplicação, uma vez que permite a deteção imediata de um problema de identificação.

1.1. Condição de Ordem da Matriz

Na determinação da ordem do modelo é conveniente separar a matriz de covariância dos erros nos que são constantes na amostra (parte específica da alternativa) e nos que variam entre indivíduos da amostra (por exemplo quando se têm parâmetros aleatórios). A condição de ordem apenas se aplica à parte dos erros que são específicos da cada alternativa. Tal como já foi referido, a ordem de uma matriz define o limite superior do número de parâmetros possíveis de estimar atendendo ao número de alternativas disponíveis, e se o modelo não incluir elementos específicos do decisor a condição de ordem define que haverá $s = J(J-1)/2-1$ parâmetros dos erros (específicos da cada alternativa) estimáveis, em que J é o número de alternativas. Esta expressão refere o número de elementos únicos da matriz simétrica de covariância das diferenças entre utilidades menos um de modo a estabelecer a escala. Assim pode concluir-se que:

- Modelos com apenas 2 alternativas não permitem a estimação de nenhum termo da matriz de covariância específico de cada alternativa;
- Modelos com 3 alternativas permitem a identificação de 2 termos;
- Modelos com 4 alternativas permitem a identificação de 5 termos;
- Modelos com 5 alternativas permitem a identificação de 9 termos;
- etc.

Nos modelos em que os erros têm parte não específica da cada alternativa e variam, por exemplo erros com parâmetros aleatórios, será possível estimar mais que s parâmetros devido à informação adicional determinada das variações da matriz de covariância entre os indivíduos, pelo que o limite de parâmetros estimáveis é elevado comparativamente ao tamanho da amostra e, portanto, não constitui uma condicionante (Ben-Akiva *et al.*, 2001).

Apesar de a condição de ordem fornecer rapidamente um teste ao número de parâmetros identificáveis, Ben-Akiva *et al.* (2001) refere que frequentemente o número de parâmetros estimáveis é inferior a s devido à estrutura de covariância de erros escolhida, pelo que, apesar de necessária, não é suficiente para a identificação do modelo.

1.2. Condição da Característica da Matriz

A condição de ordem apenas conta o número de células da matriz de covariância de diferenças de utilidades mas ignora a sua estrutura interna. A condição da característica, no entanto, utiliza a característica da matriz de covariância das diferenças entre utilidades, ou seja, o número de linhas ou equações linearmente independentes. Esta condição estabelece o número de parâmetros que podem ser estimados uma vez que apenas uma quantidade de parâmetros igual ao número de equações linearmente independentes pode efetivamente ser estimado. Para a definição da característica de uma matriz existe a necessidade de recorrer a algumas transformações da matriz de covariâncias entre diferenças de utilidades original, nomeadamente a vectorização dos seus elementos únicos; o cálculo da matriz Jacobiana desse vetor; e sua transformação numa matriz escalonada (de escada de linhas) reduzida. O número de linhas da matriz escalonada reduzida não nulas é a característica da matriz original.

No anexo 13 do capítulo III apresenta-se um exemplo da determinação da característica de uma matriz de covariâncias entre diferenças de utilidade de um modelo heteroscedástico com 4 alternativas.

1.3. Condição da Definição Positiva

As condições de ordem e da característica referidas nos pontos anteriores são geralmente utilizadas para a identificação de modelos Probit puros, mas uma vez que parte do modelo Logit misto é um Probit também podem ser utilizados para o Logit misto (Ben-Akiva *et al.*, 2001). No entanto, não são suficientes uma vez que apesar de úteis têm diferenças importantes. Recorde-se que uma restrição é sempre necessária de modo a estabelecer a escala do modelo, que num Probit puro é realizado impondo um valor positivo pelo menos a um elemento da matriz de covariância, geralmente a unidade, designado de σ_r elemento *restringido*. Num modelo Logit misto, no entanto, a definição da escala é realizada como num Logit padrão, restringindo o valor do seu parâmetro de escala μ . Uma vez que a escala é definida com μ , a normalização de σ_r passa a ser uma restrição pura para a identificação do modelo. Ben-Akiva *et al.* (2001) refere que para facilmente se avaliar a existência de diferenças estatísticas entre o modelo Probit puro e o modelo Logit misto, o elemento normalizado σ_r deve ser restringido a zero para que o Probit puro seja um caso especial do modelo Logit misto especificado. Outra diferença a considerar é o facto de num modelo Probit puro o elemento da matriz de covariância a restringir para definir a escala ser arbitrário, facto que não se verifica num modelo Logit misto equivalente devido à sua estrutura.

Outro aspeto particular dos modelos Logit mistos decorre da sua estrutura aditiva de perturbações. Antes de a identificação ser imposta, existe um número infinito de soluções corretas para determinado problema, que as condições anteriores reduzem a uma destas

possíveis soluções. Esta solução, no entanto, pode não maximizar a função de máxima verossimilhança, ou seja, não é a correta, pelo que existe a necessidade de verificar a sua validade, determinando um conjunto de normalizações aceitáveis. O motivo de necessidade de normalizações adicionais provem da infinidade de possíveis soluções que resultam da hipotética estrutura de covariância assumida, de modo a estabelecer a existência de uma solução única que não altere a estrutura do modelo, nomeadamente a sua matriz de covariância de diferenças de utilidades. A condição de definição positiva permite verificar se a normalização escolhida é válida e se os restantes parâmetros estimados (portanto, não normalizados) conseguem replicar a estrutura do modelo subjacente. Ben-Akiva *et al.* (2001) refere que as normalizações comuns, e aparentemente óbvias (por exemplo, $\sigma = 1$ ou $\mu = 1$), podem não ser válidas para os modelos Logit mistos pois a estrutura do modelo impede a matriz de covariância de diferenças de utilidades subjacente seja recuperada. Se as normalizações forem válidas, os resultados do modelo mediante a estimação com essas mesmas normalizações serão idênticos.

Ben-Akiva *et al.* (2001) refere duas questões incontornáveis na definição positiva do modelo:

- Existe uma infinidade de possíveis normalizações a implementar de forma a identificar o modelo. Contudo todas as normalizações válidas para uma determinada especificação resultam em matrizes de covariâncias idênticas, ou seja, os parâmetros estimados com uma determinada normalização (por exemplo, $\sigma_{11} = 0$) podem ser convertidos nos parâmetros estimados impostos por uma normalização diferente (por exemplo, $\sigma_{11} = 1$), desde que ambas sejam válidas.
- A matriz de covariância do modelo Logit misto pode ser expressa como:

$$\text{Cov}(U_n) = \Omega_n = \text{Cov}(\eta_n) + \left(\frac{\pi^2}{6 \cdot \mu^2} \right) \cdot I_J = F_n \cdot T \cdot T' \cdot F_n' + \left(\frac{\pi^2}{6 \cdot \mu^2} \right) \cdot I_J$$

Onde:

- F_n : Matriz de $(J \times M)$ de pesos dos fatores incluindo parâmetros fixos e/ou desconhecidos, podendo ser um função de covariantes;
- T : Matriz de $(M \times M)$ triangular inferior de parâmetros desconhecidos em que $T \cdot T' = \text{Cov}(\eta_n)$, designada de factorização Cholesky da matriz de covariância de η_n .

Assim, segundo Ben-Akiva *et al.* (2001), a $\text{Cov}(\eta_n)$ é por construção semi-definida positiva, uma vez que $F_n \cdot T$ também pode ser igual a zero.

Se a normalização a realizar não conseguir satisfazer as duas questões anteriores, os parâmetros estimados serão inconsistentes e haverá perda de precisão do modelo. Assim, as

questões anteriores conseguem impor restrições aos possíveis conjuntos de normalizações a utilizar em modelo Logit mistos. Como a segunda questão é satisfeita pelo modo como a matriz de covariâncias da parte flexível dos erros é definida, o problema passa a cingir-se a encontrar um conjunto de normalizações que satisfaça a primeira questão. Esta satisfação pode ser dificultada devido à covariância dos erros IID de valores extremos *Gumbel* (os v_n), que pode necessitar de ser definida negativa, contrariando a segunda questão e tornando tal normalização inválida.

2. Restrições a Aplicar para a Identificação de Modelos Logit Mistos

A maioria dos programas informáticos para tratamento estatístico de modelos de escolha discreta permite a estimação de resultados mesmo que o modelo não seja identificável (Walker, 2002), necessitando que a sua aplicação seja antecedida de uma correta validação suplementar do mesmo por parte do analista.

É comum considerar que se um modelo Probit Puro necessita de restrições para ser possível de identificar, um modelo Logit misto análogo, devido à sua porção semelhante ao Probit, também necessita. No entanto, Walker (2002) apresenta um estudo empírico para a estimação de um modelo Probit multinomial com duas análises sem restrições e uma com restrições mas todos com valores iniciais diferentes. Todos os modelos apresentam resultados iguais (valor da função de máxima verosimilhança logarítmica iguais) mas estimações para os parâmetros de valores diferentes. Nas duas análises sem restrições estimaram-se mais parâmetros do que as condições de ordem e de característica sugerem e a diferença das suas estimações deveu-se a uma deslocação da escala dos seus valores. Estas diferenças são uma evidência de um modelo não identificado, corroboradas pelo facto da matriz Hessiana ser singular¹⁴ e as respetivas estatísticas de t não puderem ser determinadas. Na terceira análise

¹⁴ - Uma matriz quadrada é singular se o seu determinante for igual a zero ou se não tiver uma matriz inversa. Uma matriz com determinante igual a zero não tem uma solução única. Pode ser determinada da seguinte forma:

$$\text{Det}(A) = |A| = \sum_{i=1}^k a_{ij} \cdot C_{ij}$$

$$C_{ij} = (-1)^{i+j} \cdot M_{ij}$$

Em que M_{ij} é a menor da matriz A , resultante da eliminação em A da linhas i e coluna j . Ou seja:

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix} = a_{11} \cdot \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2k} \\ a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} - a_{12} \cdot \begin{vmatrix} a_{21} & a_{23} & \cdots & a_{2k} \\ a_{k1} & a_{k3} & \cdots & a_{kk} \end{vmatrix} + \cdots \pm a_{1k} \cdot \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2(k-1)} \\ a_{k1} & a_{k2} & \cdots & a_{k(k-1)} \end{vmatrix}$$

Outra forma de calcular o determinante de uma matriz quadrada é o produto dos elementos da matriz diagonal principal resultante da aplicação da *Eliminação de Gauss* ou do método de *Condensação da Matriz* para obter uma matriz triangular utilizando as seguintes regras (Barbosa, 2004):

- Se a matriz A_{i+1} resultar de A_i através da troca entre duas linhas ou colunas: $\det(A_{i+1}) = - \det(A_i)$;
- Se a matriz A_{i+1} resultar de A_i através da multiplicação de uma linha ou coluna por um valor c : $\det(A_{i+1}) = c \cdot \det(A_i)$;

houve uma restrição de um elemento da matriz de covariâncias de diferenças entre utilidades (mais especificamente a variância $\sigma_{11} = 1$), permitindo ao modelo ser identificado, apresentando uma matriz hessiana não singular, estatísticas t para todos os parâmetros (exceto o que foi restringido) e um valor da função de máxima verossimilhança logarítmica igual ao valor dos modelos não restringidos. Apesar de existirem programas estatísticos que permitem a detecção deste tipo de erro, pode acontecer que não se perceba imediatamente esta situação devendo para isso atender-se a algumas pistas: como a existência de erros padrão muito grandes ou a instabilidade dos parâmetros para valores iniciais diferentes ou quando o número de extração para simulação aumenta, como um sinal de a matriz hessiana é próxima de singular (Walker, 2002).

Num estudo de um modelo Logit misto equivalente ao Probit anterior, Walker (2002) e Ben-Akiva *et al.* (2001) referem existirem evidências empíricas que o primeiro não necessita de tantas restrições para ser identificável. Apresenta resultados em que o modelo Logit misto sem restrições têm resultados equivalentes, sendo estáveis tanto com o aumento de extrações, como com diferentes valores iniciais. A análise com restrições (sucessivamente 0, 1, 4 e 10) mostrou ligeiras alterações no resultado mas provam que não são necessárias as mesmas restrições do modelo Probit. Apesar disso em ambos os artigos se sugere a normalização pois a diferença entre os resultados não apresenta valores que justifiquem a alteração da metodologia de identificação apresentada anteriormente.

Outra assunção comum é que as restrições para identificação de modelos Probit podem ser aplicados nos modelos Logit mistos. Esta questão deve ser analisada de duas formas:

- Num modelo Logit misto é desejável impor restrições que reduzam o modelo a um Logit multinomial puro, ou seja, restringir os elementos da parte Probit a zero;
- Devido à sua estrutura aditiva, as perturbações do modelo Logit misto leva a complicações que não existem no Probit, nomeadamente no que se refere ao número de restrições a impor devido às condições de ordem e de característica, que podem ser maiores do que as desejáveis, necessitando-se assim de verificar a condição da definição positiva do mesmo.

No entanto, a primeira questão referida não é tão simples como parece uma vez que provas empíricas sugerem que a escolha do elemento das perturbações da parte Probit a anular não é arbitrária (Walker, 2002). Assim, Ben-Akiva *et al.* (2001) sugere que se escolha a alternativa, ou uma delas, com variância mínima para alternativa de base, e portanto a anular. Esta alternativa com variância mínima pode ser depreendida através do modelo sem restrições (e portanto não identificável).

- Se a matriz A_{i+1} resultar de A_i através da soma do múltiplo de uma linha a outra linha ou de uma coluna a outra coluna: $\det(A_{i+1}) = \det(A_i)$;
 Outra forma de cálculo é através da Regra de Sarrus, que consiste na diferença entre os produtos das diagonais principais e os produtos das diagonais secundárias.

Outro aspeto a ter em atenção prende-se com o facto de, ao contrário do Probit, a imposição da restrição a uma alternativa com o valor de 1, não resolver o problema de identificação do modelo Logit misto análogo. Ben-Akiva *et al.* (2001) apresenta um estudo empírico com 10 000 observações sintéticas (não reais) para três alternativas, cujos verdadeiros parâmetros são conhecidos (aliás foram utilizados para gerar os dados sintéticos). A aplicação do modelo com três hipóteses para o parâmetro da alternativa com variância mínima (respetivamente 0, 1 e 2) apresenta resultados da função de máxima verosimilhança logarítmica quase idênticos e que determina que todas as restrições são válidas. No entanto, a análise restringindo o parâmetro da alternativa com máxima variância apresenta valores diferentes consoante o valor atribuído na restrição. Nesta análise, a restrição ao valor zero leva a uma perda significativa da precisão do modelo e a conclusões erradas em relação a heteroscedaticidade do modelo, e só restrições com valores acima de 2,25 é que permitem modelos estimados com a precisão correta.

Em relação a modelos Logit Mistos hierárquicos, Walker (2002) e Ben-Akiva *et al.* (2001) referem que neste tipo de modelos todos os parâmetros dos nichos podem ser estimados quando existem três ou mais nichos e se pelo menos um dos nichos tiver mais de uma alternativa.

Walker (2002) refere em relação a modelos Logit mistos com parâmetros aleatórios (variações de gosto), que estudos empíricos mostram que a identificação de parâmetros sistémicos não é semelhante à identificação de termos aleatórios. Refere que para modelos heteroscedásticos a definição inicial da parte sistémica é arbitrária ao contrário da parte das perturbações. Assim, refere que para parâmetros aleatórios distribuídos independentemente, especificados como variáveis categóricas, a identificação dos parâmetros é semelhante à situação dos modelos hierárquicos, isto é, se a variável tiver duas categorias apenas um termo das perturbações pode ser estimado, e se as categorias forem três ou superiores todos os termos das perturbações são estimáveis.

Outro aspeto a atender na modelação de Logit mistos relaciona-se com o facto de amostras reduzidas não permitirem a estimação de um elevado número de parâmetros aleatórios. Walker (2002) apresenta resultados de estudos empíricos em que fracas rotinas de otimização, valores iniciais errados, problemas relacionados com a escala, ou um número insuficiente de extrações nas simulações podem esconder problemas de identificação. Para atender a este aspeto sugere que se verifique a estabilidade da estimação dos parâmetros para sucessivos incrementos nas extrações não referindo valores de referência mas salientando que depende drasticamente do tipo de dados e da especificação do modelo. Outra verificação sugerida passa pela alteração dos valores iniciais de estimação adotando particular atenção a situações não convexas que podem levar a conclusões erradas sobre a aproximação ao melhor ajuste. É conveniente verificar se os parâmetros se afastam dos valores iniciais durante a estimação para descartar problemas de escala e a análise da matriz

de informação (valores esperados da matriz de segundas derivadas da função de máxima verosimilhança logarítmica) no ponto de convergência pode evidenciar problemas, nomeadamente se os erros padrão tiverem valores extremamente elevados.

3. Outros Aspetos Relevantes na Especificação do Modelo Logit Misto

3.1. Seleção dos Parâmetros Aleatórios

Em relação à seleção dos parâmetros aleatórios a utilizar, os autores referem que estes podem providenciar um vasto conjunto de informações sobre as preferências, podendo definir o seu grau de heterogeneidade através do desvio padrão dos parâmetros e das interações entre a média estimada do parâmetro e os critérios determinísticos de segmentação (atendendo aos atributos das alternativas, aos indicadores socioeconómicos de contextualização, bem como aos descritores dos métodos de recolha de informação e dos instrumentos utilizados). Afirmam ainda que os mesmos servem de base para a determinação de correlações entre alternativas e entre situações de escolha. Assim, recomendam que parte considerável do tempo despendido na estimação seja empregue para a definição dos atributos das alternativas através de parâmetros aleatórios, investigando a possibilidade de considerar diferentes distribuições para cada atributo, especialmente se o sinal do mesmo for importante (por exemplo, o custo nunca pode ter um sinal positivo pois não seria expectável que um indivíduo tivesse mais propensão para escolher uma alternativa mais cara).

A definição de um conjunto apropriado de parâmetros aleatórios requer atenção no número e tipo de extrações, nos pressupostos da distribuição e na possibilidade de situações de escolha correlacionadas. Os autores sugerem a estimação de um modelo Logit multinomial padrão como um passo essencial para começar a definir os possíveis atributos e a sua forma funcional determinística, reconhecendo, no entanto, as suas limitações na seleção inicial dos atributos a parametrizar aleatoriamente. Um método sugerido em (Hensher *et al.*, 2003 e 2005) sugere o teste com *Multiplicadores de Lagrange* como base estatística para aceitar ou rejeitar a inclusão de parâmetros fixos no modelo. Nestes testes definem-se variáveis artificiais para a alternativa i do género:

$$z_{in} = (x_{in} - \bar{x}_i)^2, \text{ onde } \bar{x}_i = \sum_j x_{jn} \cdot P_{jn}$$

E P_{jn} é a probabilidade de escolha da alternativa j pelo indivíduo n , determinado pelo modelo Logit multinomial padrão (MNL). Estas variáveis são depois utilizadas numa nova estimação do modelo MNL rejeitando-se a hipótese nula dos coeficientes não aleatórios dos atributos x se os coeficientes das variáveis artificiais forem significativamente diferentes de zero. O teste da significância conjunta das variáveis z pode ser determinado através de testes estatísticos de razão de verosimilhança.

Um método alternativo sugerido pelos mesmos autores assume todos os parâmetros como aleatórios e avalia os seus desvios padrão estimados utilizando para os parâmetros individualmente um teste de t com base nula (assimptótico) e para avaliar a sua contribuição conjunta na melhoria do modelo o teste da razão de verossimilhança. Este último método, no entanto, pode ser muito exigente quando se pretende estimar uma grande quantidade de variáveis dependentes.

3.2. Seleção da Distribuição dos Parâmetros Aleatórios

Um aspeto de grande importância para a estimação de modelos Logit mistos refere-se à possível influência das distribuições assumidas para os parâmetros aleatórios. Estas distribuições podem ter várias formas sendo as mais populares, como já foi referido, a distribuição normal, a lognormal, a triangular e a uniforme.

A distribuição lognormal é geralmente utilizada quando o parâmetro necessita de limitações em relação ao sinal (não negativo) e a distribuição uniforme com limites (0,1) é útil quando se tem variáveis binárias.

As distribuições são essencialmente aproximações arbitrárias do perfil de comportamento real das variáveis em análise, que se consideram como empiricamente próximas dos valores reais. No entanto, muitas distribuições utilizadas têm pelo menos dois inconvenientes: problemas com o sinal e com o comprimento da(s) cauda(s). Distribuições truncadas ou restringidas parecem ser soluções com grande potencial para minorar os problemas referidos, sugerindo Rigby e Burton. (2006) as seguintes:

- Distribuições *normais restringidas* (“censored normal”) superiormente a zero em que se substitui os parâmetros θ distribuídos normalmente $N(b, \sigma)$ pela transformada $t_n = \min(0, \theta)$, com densidade abaixo de zero idêntica à densidade normal de b . A estimação envolve a identificação de b e σ , e portanto de t_n , e assim a proporção da população agrupada a zero e a proporção abaixo de zero.
- Distribuições de *Johnson S_B limitadas* superior (l) e inferiormente (u). A transformação $t_n = l + (u - l) \cdot (\exp(\theta) / (1 + \exp(\theta)))$, produz uma distribuição entre u e l com a forma, média e variância determinada pela média e variância da distribuição normal de θ .

3.3. Seleção do Número de Extrações para as Simulações

O cálculo de probabilidades de escolha através do modelo Logit misto é realizado recorrendo a integrações por simulação de Monte Carlo que se apresentam como o melhor método de cálculo de integrais de multidimensionais, principalmente se o número de dimensões for muito elevado. Este cálculo envolve a geração de sequências pseudoaleatórias que tentam simular extrações de uma distribuição uniforme com um intervalo unitário (Hensher *et al.*, 2005). Apesar destas sequências pseudoaleatórias não se distinguirem das extrações de uma

distribuição uniforme, a sua dispersão no intervalo unitário não é uniforme, pelo que houve a necessidade de introduzir sequências baseadas na teoria de números que aproximem a integração de Monte Carlo de forma mais célere. A sequência atualmente mais utilizada em modelação de transportes é a designada de *Halton* (designadas de sequências inteligentes) caracterizada no anexo 12 do Capítulo III.

O número de extrações necessárias para assegurar um conjunto estável de estimações para os parâmetros varia muito significativamente. Em geral pode assumir-se que quanto maior a complexidade da estimação (número de parâmetros aleatórios a estimar, tratamento de heterogeneidades de preferências em torno da média, correlação de atributos e alternativas, etc.), maior o número de extrações necessárias. Como exemplo, Hensher *et al.* (2005) refere que para um modelo de escolha com três alternativas e um a dois parâmetros aleatórios, sem correlação entre atributos e sem a decomposição da heterogeneidade em torno da média, seriam necessárias pelo menos 25 extrações inteligentes para produzir um modelo estável, e recomendam o valor de 100 extrações como valor de referência. Sugerem, no entanto, que o melhor será a estimação do modelo com uma gama de extrações (25, 50, 100, 500, 1000, 2000) e avaliar a estabilidade e precisão das estimações.

Uma vez que um número de extrações elevado assegura, em princípio, maior probabilidade de chegar a um conjunto de parâmetros estáveis e precisos, seria de esperar que se realizassem sempre em elevado número. No entanto, a estimação de modelos com muitos parâmetros aleatórios, com grandes amostras e milhares de extrações, mesmo com o potencial de processamento dos computadores atuais, leva muitas horas, pelo que na prática é mais eficiente tentar encontrar parâmetros estáveis com o menor número de extrações possíveis. Hensher *et al.* (2005) referem estudos que mostram simulações de parâmetros aleatórios com 100 números de Halton com variância inferiores a simulações com 1000 número pseudoaleatórios, e que com 125 extrações de Halton se conseguiam erros de simulação metade dos conseguidos com 1000 extrações pseudoaleatórias e menores que os conseguidos com 2000 pseudo-extrações. Concluem assim que o procedimento de estimação com sequências de Halton podem ser até 10 vezes mais rápidas que utilizando sequências pseudoaleatórias.

Apesar das vantagens apresentadas, as sequências de Halton não são perfeitas, referindo Hensher *et al.* (2003 e 2005) um estudo em que a uniformidade das mesmas era interrompida para sequências de grande dimensão (acima de 25) devido a correlações na mesma.

3.4. Heterogeneidade de preferência em redor da média de um parâmetro aleatório

O modelo Logit multinomial (ou Logit condicional) assume uma homogeneidade de preferências entre indivíduos, pelo que apenas um vetor de estimativas de parâmetros fixos é

modelado para os atributos considerados na escolha. Isto implica que todos os indivíduos têm o mesmo gosto ou preferência para determinado atributo. Variáveis socioeconómicas podem ser incluídas como interações entre atributos ou como interações entre constantes específicas das alternativas, e podem estimar-se modelos para diferentes subconjuntos de dados (segmentos), mas estas abordagens continuam a ser débeis na representação de heterogeneidade de preferências (Colombo *et al.*, 2008).

Amador *et al.* (2005) referem que para capturar a heterogeneidade de preferências individuais se podem utilizar duas abordagens:

- Utilizar uma especificação em que os parâmetros dos atributos sejam uma função das características socioeconómicas individuais observadas (como a idade, sexo, rendimento, posse de veículos, etc.), permitindo modelar variações de gosto sistémicas e identificar as causas de variabilidade para diferentes medidas de disponibilidade para pagar;
- Tentar captar variações de gosto aleatórias através da especificação de modelos Logit mistos. Esta abordagem permite obter parâmetros da população e individuais, sendo os últimos determinados pela combinação de estimativas de máxima verosimilhança simuladas da população total com métodos de inferência *Bayesiana*.

Ambas as abordagens podem ser utilizadas num único modelo permitindo a incorporação de heterogeneidade não observada bem como variações de gosto sistémicas.

Rigby e Burton (2006) referem que os modelos Logit mistos são uma ferramenta analítica poderosa para incorporar diversas preferências em estudos de escolhas apresentando-se mais apelativos pois podem ser mais reveladores que estimativas segmentadas pontuais. Estes permitem incluir explicitamente uma gama de atitudes em relação aos atributos dentro da população, especialmente importante quando não é a atitude média que importa identificar mas sim o tamanho do grupo disponível para aceitar ou não determinada opção.

Seria pouco usual não encontrar um considerável grau de variabilidade na maneira com as pessoas avaliam determinados atributos, pelo que um tratamento incorreto da heterogeneidade de preferências, principalmente em dados obtidos por preferências declaradas, pode levar a resultados de estimação enganadores.

Os modelos Logit mistos incorporam a heterogeneidade de preferências através da componente sistemática da utilidade, V , ou em alternativa em termos da componente aleatória da utilidade, ε . A formulação típica dos modelos Logit mistos decompõe-se numa componente não observada, numa componente de heterogeneidade de preferências e numa componente determinística:

$$U_{njt} = \theta \cdot X_{njt} + \eta_n \cdot X_{njt} + \varepsilon_{njt}$$

Onde para o individuo n que escolhe a alternativa j no instante t , X_{njt} é um vetor de atributos observados, θ é um vetor de parâmetros associados aos atributos observados, η_n é um vetor de k (número de atributos) desvios padrão e ε_{njt} é um termo aleatório não observado, independente dos restantes termos e com distribuição IID de *Gumbel*.

Hensher *et al.* (2005) sugerem a utilização da seguinte expressão para a utilidade com K atributos num modelo Logit misto:

$$\begin{aligned} U_{njt} &= (\theta + \Delta \cdot z_n + \Gamma \cdot v_n) \cdot x_{njt} + \varepsilon_{njt} = (\theta + \Delta \cdot z_n + \eta_n) \cdot x_{njt} + \varepsilon_{njt} = \\ &= \sum_{k=1}^K [(\theta_k + \delta'_k \cdot z_n + \eta_{nk}) \cdot x_{njtk}] + \varepsilon_{njt} \end{aligned}$$

Em que η_{nk} é o termo aleatório cuja distribuição entre os indivíduos depende geralmente de parâmetros subjacentes, z_n são covariâncias observados, v_n é um vetor de variáveis aleatórias não correlacionadas com matriz de variância e covariância de ε com diagonal conhecida e Γ é a matriz triangular inferior que permite variâncias e correlações entre os parâmetros já que $var[\theta_n] = \Gamma \varepsilon \Gamma'$.

Segundo Hensher *et al.* (2005) a introdução de uma interação entre a estimação da média do parâmetro aleatório e a covariância é equivalente a revelar a presença ou ausência de heterogeneidade de preferências em volta da média do parâmetro estimada. O termo $\delta'_k \cdot z_n$ da expressão da utilidade tem essa função e se a interação não tiver significância estatística, com base nas covariâncias observadas z_n , conclui-se que à uma ausência de heterogeneidade nas preferências em relação à média, ou pelo menos que não foi possível revelar a sua presença no termo usado. Neste caso, todas as fontes de heterogeneidade nas preferências são representadas na análise pelo desvio padrão estimado.

3.5. Inclusão de Situações de Escolha Correlacionadas

Situações de escolha correlacionadas, uma violação da assunção de independência de observações considerada nas estimações clássicas de modelos de escolha, podem surgir quando se realizam várias observações de um mesmo indivíduo, por exemplo com preferências declaradas. Esta correlação pode provir de várias fontes, como a utilização de descritores socioeconômicos que não variam entre situações de escolha, atendendo a um determinado indivíduo e/ou sequência de situações de escolha.

Hensher *et al.* (2003) referem que a heterogeneidade é um tipo especial de correlação entre situações de escolha que ainda não são bem compreendidas. Sugerem que desde que se reconheça que a heterogeneidade de preferências deve incluir todas as alternativas em

relação a todas as situações de escolha, a correlação é automaticamente incorporada na respetiva matriz de covariância de parâmetros aleatórios.

3.6. Correlação entre Parâmetros

Hensher *et al.* (2005) referem que todos os conjuntos de dados, independentemente do número de situações de escolha por indivíduo selecionado, podem apresentar efeitos não observados que sejam correlacionados entre alternativas de uma dada situação de escolha. Uma forma de incorporar esta situação é a permissão de correlação dos parâmetros aleatórios dos atributos comuns às diferentes alternativas. Este procedimento resulta numa matriz de covariâncias com estimativas dos elementos não diagonais que identificam a dependência de um atributo em relação a outro, dentro de cada alternativa e entre alternativas (dependendo se o atributo é específico da alternativa ou genérico).

Pode dizer-se que cada indivíduo utiliza as mesmas preferências para avaliar as utilidades relativas em cada situação de escolha ou período de tempo, o que em termos de comportamento implica que as preferências aleatórias induzem correlações entre alternativas e entre situações de escolha. Os modelos Logit mistos em geral permitem incluir (Hensher *et al.*, 2005):

- Correlações entre alternativas e não entre situações de escolha, assumindo que os parâmetros θ_{tn} são *IID* (não apresentam covariâncias, mas apenas variâncias) entre situações de escolha;
- Correlações entre situações de escolha mas não entre alternativas fixando todos os parâmetros exceto as que representam constantes específicas de cada alternativa, e assumindo que os parâmetros destas constantes são *IID* entre alternativas mas iguais para cada indivíduo entre situações de escolha.

3.7. Disponibilidade para Pagar (DpP)

A seleção da distribuição de cada parâmetro individualmente é um desafio cuja complexidade aumenta quando o interesse foca a razão entre parâmetros aleatórios, como por exemplo o cálculo de estimativas da DpP (geralmente determinadas pelo quociente entre um dado atributo e um que considere o custo ou preço, considerando os restantes constantes). Hensher *et al.* (2003) referem que a razão entre dois parâmetros com distribuição triangular tem uma distribuição descontínua com ambas as distribuições apresentando singularidades se o seu conjunto imagem não for forçado a excluir o zero. Ambas as situações têm média e variância infinita. A razão entre duas distribuições normais apresenta os mesmos problemas com singularidade quando o denominador for igual a zero.

No cálculo de estimativas da DpP com base em parâmetros aleatórios pode utilizar-se toda a informação sobre a distribuição ou apenas a sua média e desvio padrão. Apesar de mais

complicada é preferível a primeira metodologia, recorrendo a simulações com extrações da matriz de covariância estimada para os parâmetros (Hensher *et al.*, 2005).

Em relação ao tipo de distribuição a utilizar para definir a DpP, Hensher *et al.* (2005) referem que o facto da distribuição Lognormal ter uma cauda muito extensa, levou à sugestão de remoção dos últimos percentis de modo a assegurar que a média e o desvio padrão representassem de forma mais fidedigna a maioria dos indivíduos. No entanto, dados empíricos têm mostrado que a remoção dos dois percentis mais elevados leva a alterações significativas da média e do desvio padrão, pelo que também se sugere a utilização de distribuições restringidas. A título de sugestão, Hensher *et al.* (2005) sugerem a utilização de uma distribuição triangular limitada inferiormente para a DpP, uma vez que elimina a longa cauda da distribuição Lognormal ao mesmo tempo que assegura o seu correto sinal de forma coerente com o comportamento.

3.8. Existência de Variáveis Endógenas

Nos modelos de escolha discreta assume-se que um grupo de indivíduos tem um conjunto de alternativas à disposição e que cada um metodicamente escolhe a alternativa da qual recebe maior utilidade. A utilidade de cada alternativa depende dos seus atributos e estes estão condicionados às características de cada indivíduo e a um termo específico de erros. Quando um analista quer replicar o comportamento de um indivíduo está limitado pela informação disponível, ou seja, a um conjunto limitado de atributos. Se esses atributos forem os mais relevantes nas opções individuais e a amostra for suficientemente grande a modelação permite estimar os “verdadeiros” coeficientes. No entanto, se alguns dos atributos relevantes não forem observados os coeficientes corretos do modelo só serão estimáveis se estes atributos não observados não estiverem correlacionados com os observados (Guevara e Ben-Akiva, 2009).

Assim, a endogeneidade, ou não ortogonalidade, em modelos de escolha discreta ocorre quando a parte sistemática da utilidade está correlacionada com o termo dos erros (Train, 2009; Louviere *et al.*, 2005; Guevara e Ben-Akiva, 2009). Este problema é virtualmente inevitável uma vez que a modelação é feita sem total conhecimento de todos os atributos relevantes ou da sua inter-relação. Louviere *et al.* (2005) definem a endogeneidade como todos os efeitos que não são exógenos, ou seja, a endogeneidade equivale a erros de especificação, que quando ignorados podem levar a enviesamentos dos coeficientes estimados e das conclusões inferidas.

Como exemplo de variáveis ou situações em que se pode verificar endogeneidade, Train (2009) apresenta o preço de um bem ou serviço que pode ser influenciado por atributos que são impossíveis de observar ou medir, ou a inter-relação entre escolhas de um indivíduo, em que uma escolha é determinada por outras, por vezes, de forma pessoal e não intuitiva. Existem exemplos no dia-a-dia em que com duas opções aparentemente similares, diferindo

apenas no seu custo (e qualquer outro atributo não observado), a escolha de determinados indivíduos recai na opção mais cara, atitude aparentemente irracional se a endogeneidade não for considerada.

Louviere *et al.* (2005) refere que tradicionalmente a modelação de escolhas assume um determinado conjunto de opções exógenas do qual se assume que todos os indivíduos escolhem com total informação. Na realidade, no entanto, cada indivíduo pode optar pela forma como adquire e utiliza a informação sobre o conjunto de escolhas, um processo que torna o real conjunto de escolhas endógeno. Para corrigir esta endogeneidade há a necessidade de modelar o modo como cada indivíduo considera a informação do conjunto de escolhas disponíveis, podendo algumas das assunções dos modelos de escolha associadas a expectativas racionais revelarem-se inaceitáveis.

Guevara e Ben-Akiva (2009) consideram que a variável preço é frequentemente o principal motivo do problema de endogeneidade numa função de procura quer pela omissão de atributos, quer devido a erros nas variáveis utilizadas, à determinação simultânea ou a enviesamentos na seleção das amostras. Além destes fatores, Louviere *et al.* (2005) alertam que a endogeneidade pode também advir de interações entre decisores (quando a escolha não é realizada apenas por um indivíduo), de comportamentos estratégicos (o formato das respostas pode influenciar a predisposição para revelar as suas verdadeiras preferências), ou de múltiplas escolhas discretas ou contínuas (quando uma escolha depende de outras anteriores). Este último problema é relevante devido ao facto de na modelação tradicional se considerar que a definição de um conjunto de alternativas exaustivo e mutuamente exclusivo seria suficiente para uma correta modelação. No entanto, se todas estas alternativas não forem substitutivas, como acontece frequentemente, a procura de várias opções é simultânea e a endogeneidade entre alternativas existe.

A inclusão da endogeneidade na modelação de escolhas é uma área em desenvolvimento que apenas a partir de meados desta década começou a ser abordada. Várias estratégias têm sido sugeridas entre as quais se destacam (Train, 2009; Louviere *et al.*, 2005; Guevara e Ben-Akiva, 2009):

- A *abordagem BLP*: retira a endogeneidade de modelos de escolha não lineares para uma regressão linear, permitindo a utilização de métodos de variáveis instrumentais tradicionais. É uma abordagem aplicável em segmentos da população homogéneos em termos de atributos não observados (e preço). É particularmente útil porque a distribuição dos seus erros em torno da sua média condicional (geralmente igual a zero) não necessita de ser conhecida ou estimada (Louviere *et al.*, 2005). A teoria subjacente assenta na ideia de que constantes podem ser incluídas no modelo de escolha de modo a captar os efeitos médios dos atributos (observados e não observados) (Train, 2009). A sua aplicabilidade é limitada pelo elevado número de constantes a determinar, mas

Train (2009) refere a utilização de um procedimento designado de “contração” para minorar o problema;

- *A abordagem da função de controlo:* A ideia base da abordagem é a construção de uma variável ou função de controlo que capte a parte do valor dos erros esperados, condicionados aos atributos observados, que não é zero. Se esta função de controlo for adicionada como uma variável explicatória no modelo, o problema de endogeneidade fica resolvido (Guevara e Ben-Akiva, 2009). A inclusão de uma nova variável induz uma nova componente de erros, cuja distribuição deve ser conhecida ou empiricamente inferida de modo a calcular as probabilidades de escolha (Louvier *et al.*, 2005). Geralmente a abordagem é realizada em duas fases. Numa primeira fase, a variável independente endógena (por exemplo o preço) é estimada por regressão atendendo a variáveis exógenas. Esta estimação é posteriormente utilizada para criar a função de controlo a incluir no modelo como variável independente (Train, 2009).
- *A abordagem da máxima verosimilhança total:* É uma abordagem semelhante à da função de controlo, uma vez que também capta a média condicional não nula dos fatores não observados, mas combina os dois passos da abordagem anterior num só de modo a realizar uma estimação conjunta (Train, 2009; Guevara e Ben-Akiva, 2009). Pode resumir-se que a abordagem da máxima verosimilhança requer a especificação da distribuição conjunta dos erros não observados (ε_n) e a parte não observada da variável endógena (μ_n) enquanto a função de controlo requer a especificação da distribuição condicional de (ε_n) atendendo a (μ_n) (Train, 2009). Se esta distribuição de erros conjunta for possível de especificar corretamente a abordagem da máxima verosimilhança é mais eficiente que a abordagem da função de controlo, uma vez que é a máxima verosimilhança de todos os parâmetros. A abordagem da função de controlo é mais geral, uma vez que pode se aplicada a qualquer distribuição conjunta consistente com uma distribuição condicional específica. Guevara e Ben-Akiva (2009) apresentam um estudo comparativo em que esta abordagem apresentou bons resultados na correta estimação de parâmetros e na correção da endogeneidade.

Anexo 12 do Capítulo III

Caracterização das Sequências de Halton

As sequências de Halton são a sequência de baixa discrepância multidimensional mais básica. A primeira dimensão da sequência de Halton (com base no número primo 2) é a sequência de *Van der Corput* de base 2 e a segunda dimensão é a sequência de *Van der Corput* de base 3. A dimensão S de uma sequência de Halton é a sequência de *Van der Corput* utilizando o s -ésimo número primo como base. Considere-se o número primo 3 para definir a base da sequência, assumindo um intervalo entre 0 e 1 faz-se a sua divisão pelo número 3. Os pontos de divisão tornam-se nos primeiros dois elementos da sequência de Halton, $1/3$ e $2/3$. Divide-se cada divisão por 3, e obtêm-se os elementos seguintes da sequência, ou seja: $1/9$, $4/9$ (que é $1/9$ acima de $1/3$), $7/9$ (que é $1/9$ acima de $2/3$), $2/9$, $5/9$ (que é $2/9$ acima de $1/3$), e $8/9$ (que é $2/9$ acima de $2/3$) que corresponde a dividir a unidade em nove partes. Cada uma destas nove partes é então dividida em três, resultando na sequência: $1/27$, $10/27$, $19/27$, $4/27$, $13/27$, $22/27$, $7/27$, $16/27$ (ou seja, adicionando $1/9$ a zero e aos números da sequência anterior), e $2/27$, $11/27$, $20/27$, $5/27$, $14/27$, $23/27$, $8/27$, $17/27$, $26/27$ (que resultam da adição de $2/9$ a zero e aos elementos da sequência anterior). Cada um dos 27 espaços é então dividido em três, e assim por diante até se ter uma sequência com um número suficiente de valores. As sequências podem igualmente ser calculadas com base no valor 2 ($1/2$, $1/4$, $3/4$, $1/8$, $5/8$, $3/8$, ...), ou por 5 ($1/5$, $2/5$, $3/5$, $4/5$, $1/25$, $6/25$, $11/25$, ...), ou qualquer outro valor. *As sequências com base em números primos resultam em sequências de Halton* (Train, 2000). O número de ordem do número primo base é designado de dimensão da sequência, pelo que a base 2 será a sequência de dimensão 1, a de base 3 será a sequência de dimensão 2, e assim por diante (a sequência de dimensão 50 terá o quinquagésimo número primo, ou seja, 229). Quanto maior a base, maior será o ciclo e mais tempo de cálculo necessitará. Num problema s -dimensional cada sequência de Halton é realizada com base num número primo diferente, ou seja, os s primeiros números primos são utilizados.

O comprimento de cada sequência é determinado pelo número de observações e pelo número de extrações que o analista decidir utilizar. Com N observações e R extrações por observação é comum utilizar sequências com $(N \times R) + 10$ elementos. Na prática é comum excluir as primeiras extrações (Train (2000) recomenda a exclusão dos 10 primeiros elementos) de forma a melhorar a uniformidade em sequências de grande dimensão uma vez que os primeiros elementos têm tendência para serem correlacionadas entre sequências de Halton de diferentes números primos. Após a exclusão dos primeiros elementos devem utilizar-se os R elementos seguintes para a primeira observação, os segundos R elementos para a segunda observação e assim sucessivamente. A título de exemplo Train (2000) refere um modelo Logit misto especificado com três coeficientes normalmente distribuídos, em que se criaram sequências de Halton para os três primeiros números primos: 2, 3 e 5. Uma sequência de trios

é definida em que o primeiro termo é da sequência de Halton com base no número primo 2, o segundo termo é da sequência com base em 3 e o terceiro da sequência com base em 5:

$$\begin{aligned} &\langle \frac{1}{2}, \frac{1}{3}, \frac{1}{5} \rangle \\ &\langle \frac{1}{4}, \frac{2}{3}, \frac{2}{5} \rangle \\ &\langle \frac{3}{4}, \frac{1}{9}, \frac{3}{5} \rangle \\ &\langle \frac{1}{8}, \frac{4}{9}, \frac{4}{5} \rangle \\ &\vdots \end{aligned}$$

Para todo os valores da sequência de Halton escolhida, calcula-se o inverso da função cumulativa de distribuição assumida, resultando em extrações Halton da distribuição pretendida. Para o exemplo será o inverso da função cumulativa normal padrão que resultará na seguinte sequência:

$$\begin{aligned} &\langle 0,00, -0,43, -0,84 \rangle \\ &\langle -0,67, 0,43, -0,25 \rangle \\ &\langle 0,67, -1,22, 0,25 \rangle \\ &\langle -1,15, -0,14, 0,84 \rangle \\ &\vdots \end{aligned}$$

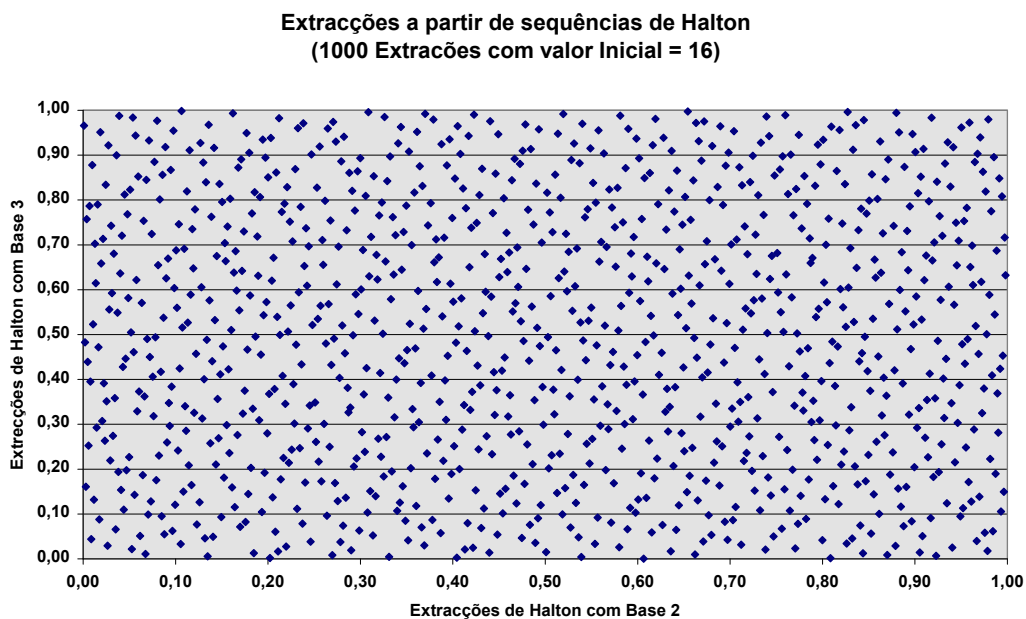


Figura A. 6 - Exemplo de 1000 extrações de sequências de Halton de base 2 e base 3 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009).

As figuras A.6 e A.7 apresentam sequências de números obtidas que utilizando sequências de Halton (com base 2 e 3) e extrações aleatórias (pseudo) padrão de uma distribuição uniforme,

onde se pode verificar a diferença entre homogeneidades das extrações, verificando-se uma cobertura mais homogênea por parte das sequências de Halton.

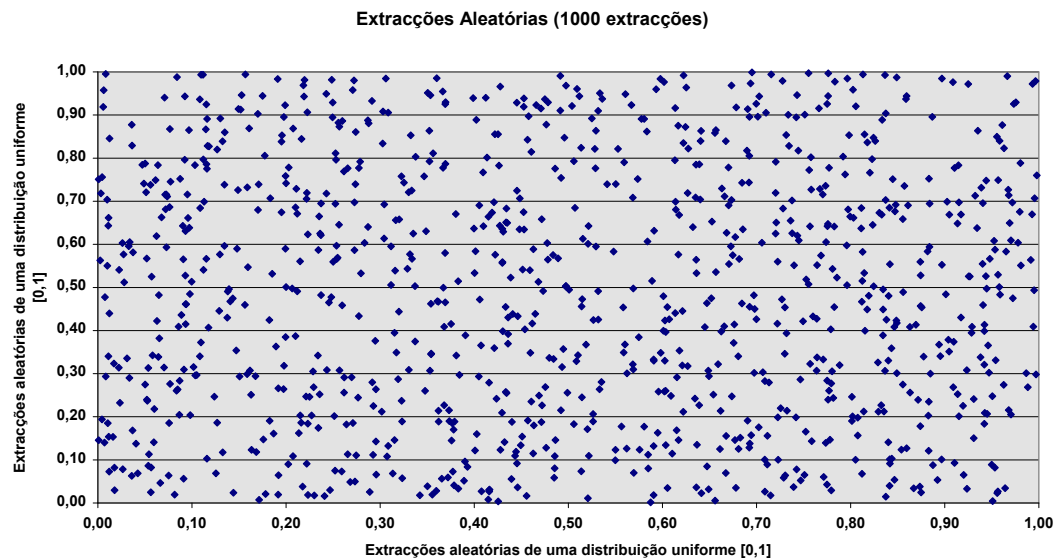


Figura A. 7 - Exemplo de 1000 extrações de sequências aleatórias de distribuição uniforme $[0,1]$ utilizando o Microsoft Office Excel© 2003.

Uma das principais desvantagens das sequências de Halton é a sua degradação para grandes dimensões, dado que com o aumento do número de base aumenta espaço a preencher. Uma sequência com base em números primos elevados tem um ciclo de grande extensão que precisa de um grande número de pontos para percorrer todo o intervalo entre $[0,1]$, dificultando o preenchimento de forma uniforme do espaço do hipercubo e levando a correlação entre sequências. Devido a esta possibilidade de correlações entre sequências de grande dimensão (igualmente existentes nas primeiras extrações das sequências de pequena dimensão) e de uma má cobertura do espaço de integração, é corrente evitar sequências com dimensão superior a 6 a 8, existindo evidências de que as mesmas apresentam resultados insatisfatórios a partir da dimensão 14 a 16 (ver figuras A.8 e A.9), que correspondem aos primos 43 ou 53 (Hess e Polak, 2003; http://www.puc-rio.br/marco.ind/quasi_mc.html#Halton acessado em 27-3-2009; Krykova, 2003). Hess e Polak (2003) referem que os problemas de correlação e de discrepância ocorrem quando a razão entre os dois números primos que definem a sequência é próxima de um valor inteiro (especialmente 1), quando os comprimentos dos ciclos utilizados são semelhantes (com uma razão próxima de 1) ou múltiplos quase exatos um do outro (próximos de um qualquer inteiro).

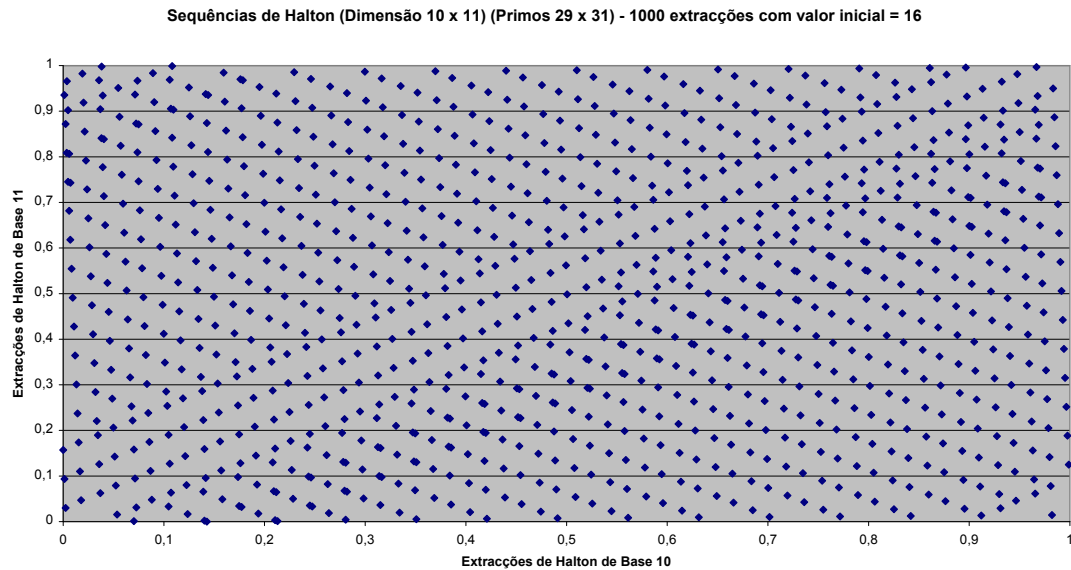


Figura A. 8 - Exemplo de correlações em 1000 extracções de seqüências de Halton de base 10 e base 11 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009).

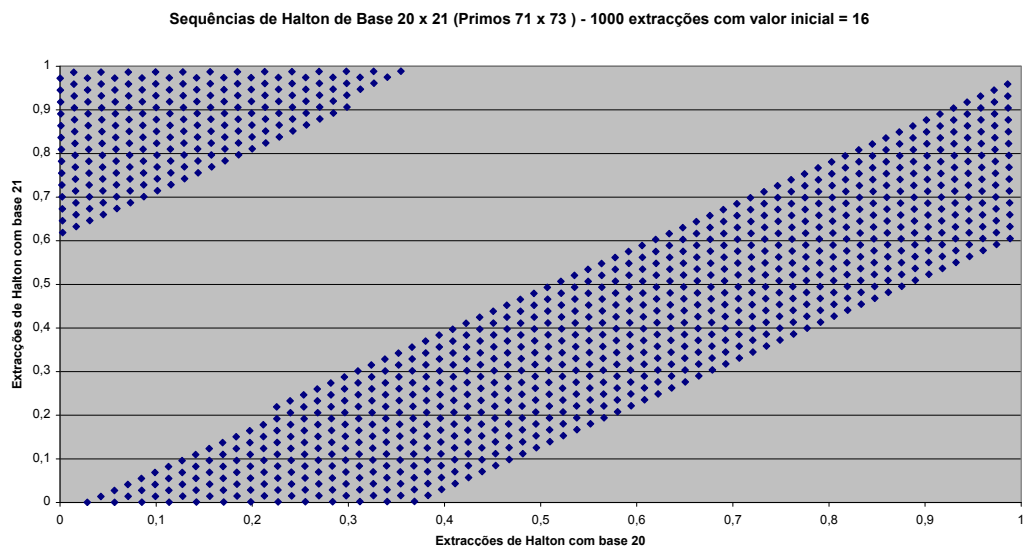


Figura A. 9 - Exemplo de correlações em 1000 extracções de seqüências de Halton de base 20 e base 21 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009).

Como forma de contornar este problema com as seqüências de Halton (especialmente de elevada dimensão), Chiou e Walker (2006) e Hess e Polak (2003) referem estudos que revelaram que seqüências de Halton baralhadas (“*Shuffled*”) ou aleatórias (baralham aleatoriamente ou pseudo-aleatoriamente a posição dos valores de uma seqüência de Halton)

tinham melhor comportamento que as sequências de Halton tradicionais ou que as sequências de Halton mexidas (“*Scrambled*”) (permutando os valores da sequência de Halton através de uma função inversa (Hess e Polak, 2003)), recomendando o seu uso para evitar os problemas anteriores. As sequências de Halton aleatórias são definidas através do desfasamento aleatório de sequências originais. Train (2003) sugere a fórmula:

$$s_n = \text{mod}(s_0 + \varphi)$$

onde:

- s_n : novo elemento da sequência de Halton aleatória
- s_0 : elemento da sequência de Halton original
- φ : extração aleatória de uma distribuição uniforme padrão entre 0 e 1
- *mod* : operador que fornece a parte decimal do resultado da operação entre parêntesis

A fórmula anterior equivale à deslocação de φ em relação ao elemento original. Se este deslocamento ultrapassar a unidade, a parte decimal passa a ser o novo elemento da sequência. Esta nova sequência tem as propriedades da sequência original de cobertura do espaço amostral e de correlação negativa sobre observações uma vez que a disposição relativa de cada elemento se mantém. A vantagem é que os seus valores são agora pseudoaleatórios e permitem que sequências de baixa dimensão possam ser utilizadas múltiplas vezes aumentando a sua possibilidade de utilização (em teoria se houverem N extrações haverá $N!$ possíveis combinações de valores sem se repetirem sequências de valores (Hess e Polak, 2003), ver figuras A.10 e A.11). Segundo Hess e Polak (2003) a utilização repetida de sequências de Halton aleatórias de baixa dimensão em sequências multidimensionais (utilizando a sua componente unidimensional) permite que as diferentes dimensões beneficiem de cobertura similar à componente unidimensional original. Assim há que ter em atenção que sequências com fraca cobertura inicial nunca poderão levar a boas coberturas multidimensionais. Train (2003) alerta, no entanto, para que o analista não fique demasiado entusiasmado com estes métodos uma vez que a precisão da estimação pode ser sempre melhorada com o aumento do número de extrações, cabendo ao mesmo a decisão sobre investir na aprendizagem e codificação de novos métodos de extração é mais proveitoso (em termos de restrições temporais) que simplesmente processar o modelo com mais extrações.

Além dos problemas de correlação, as sequências de Halton com poucas extrações podem esconder problemas de identificação dos modelos. Chiou e Walker (2006) referem que em alguns dos estudos que realizaram, problemas de identificação dos modelos apenas eram revelados, a partir de 2000 extrações de Halton. Estes problemas resultam em modelos com matriz Hessiana singular e/ou em estimações para os parâmetros invulgarmente elevados

traduzidos por gráficos de máxima verosimilhança simulada sem máximos e praticamente planos. Tal como já foi referido, a não identificação de um modelo leva a que a utilização de outro conjunto de extracções e/ou de parâmetros iniciais, resulte em conclusões diferentes.

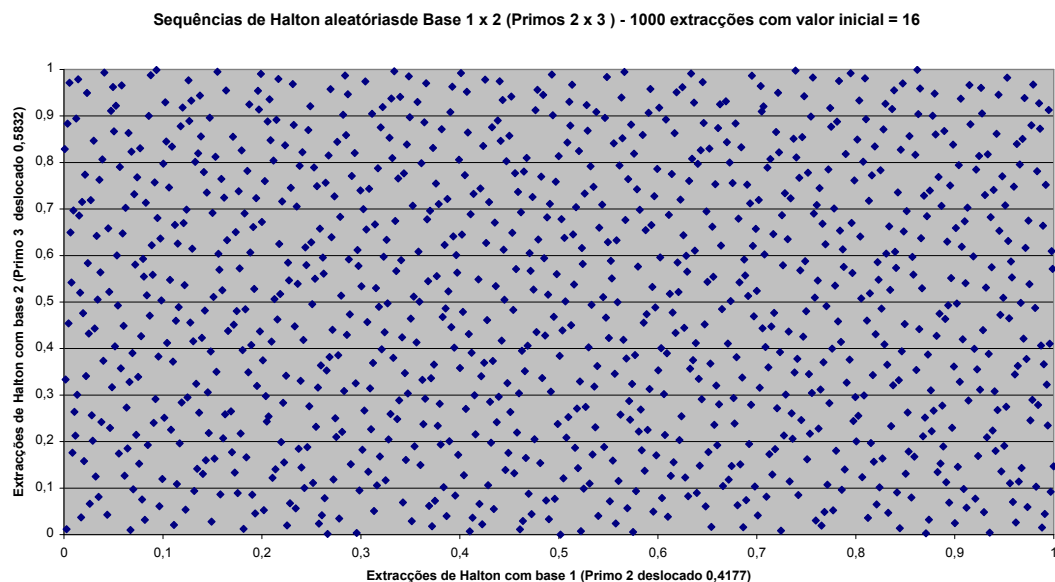


Figura A. 10 - Exemplo de 1000 extracções de sequências de Halton aleatórias de base 2 e base 3 deslocadas respetivamente 0,4177 e 0,5832 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009).

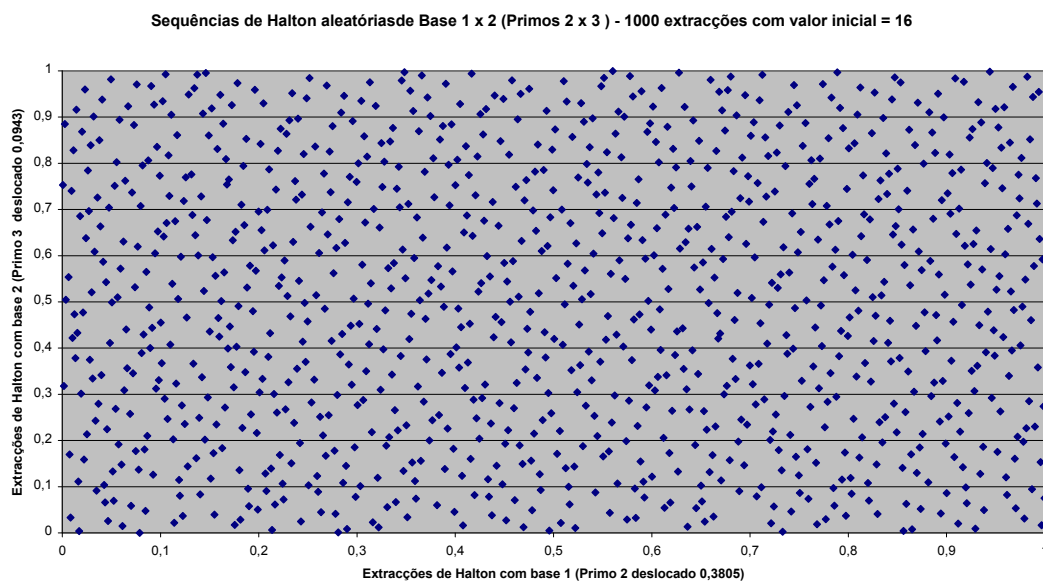


Figura A. 11 - Exemplo de 1000 extracções de sequências de Halton aleatórias de base 2 e base 3 deslocadas respetivamente 0,3805 e 0,0943 (quasi-random-multi_dim.xls - http://www.puc-rio.br/marco.ind/spreads_list.html acedido em 27-3-2009).

Portanto, outra forma de analisar a identificabilidade do modelo estimado será a estabilidade dos parâmetros estimados à medida que se aumenta o número de extrações, bem como a estabilidade dos seus erros padrão. Os mesmos estudos sugerem que a utilização de sequências de Halton aleatórias revela problemas de identificação com um menor número de extrações comparativamente a outros métodos.

Anexo 13 do Capítulo III

Exemplo da Determinação da Característica de uma Matriz de Covariâncias entre Diferenças de Utilidade de um Modelo Heteroscedástico com quatro Alternativas

Considere-se uma matriz de covariâncias entre diferenças de utilidade de um modelo heteroscedástico com 4 alternativas com os seguintes elementos:

$$\text{Cov}(U) = \Omega = \begin{bmatrix} \sigma_{11} + \frac{\pi^2}{6.\mu^2} & 0 & 0 & 0 \\ 0 & \sigma_{22} + \frac{\pi^2}{6.\mu^2} & 0 & 0 \\ 0 & 0 & \sigma_{33} + \frac{\pi^2}{6.\mu^2} & 0 \\ 0 & 0 & 0 & \sigma_{44} + \frac{\pi^2}{6.\mu^2} \end{bmatrix}$$

Onde $\sigma_{ii} = (\sigma_i)^2 = \text{variâncias}$.

A matriz de covariância das diferenças entre utilidades (considerando no caso, a diferença em relação à alternativa 4, isto é, subtraindo a utilidade da alternativa 4 às restantes) será assim (recordar a forma simplificada de definir esta matriz utilizando a matriz M exemplificada no anexo 5 do capítulo III):

$$\text{Cov}(\Delta U) = \Omega_{\Delta} = \begin{bmatrix} \sigma_{11} + \sigma_{44} + \frac{2.\pi^2}{6.\mu^2} & \dots & \dots \\ \sigma_{44} + \frac{\pi^2}{6.\mu^2} & \sigma_{22} + \sigma_{44} + \frac{2.\pi^2}{6.\mu^2} & \dots \\ \sigma_{44} + \frac{\pi^2}{6.\mu^2} & \sigma_{44} + \frac{\pi^2}{6.\mu^2} & \sigma_{33} + \sigma_{44} + \frac{2.\pi^2}{6.\mu^2} \end{bmatrix}$$

A característica desta matriz pode ser determinada recorrendo à sua vectorização, que consiste em colocar todos os seus elementos únicos (que não se repetem) num vetor:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{bmatrix} \Rightarrow \text{Vec}(A) = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{21} \\ a_{22} \\ \vdots \\ a_{nn} \end{bmatrix}$$

O que no exemplo corresponde a (Ben-Akiva *et al.*, 2001, Walker, 2002):

$$\text{Vec}(\Omega_{\Delta}) = \begin{bmatrix} \sigma_{11} + \sigma_{44} + \frac{2 \cdot \pi^2}{6 \cdot \mu^2} \\ \sigma_{22} + \sigma_{44} + \frac{2 \cdot \pi^2}{6 \cdot \mu^2} \\ \sigma_{33} + \sigma_{44} + \frac{2 \cdot \pi^2}{6 \cdot \mu^2} \\ \sigma_{44} + \frac{\pi^2}{6 \cdot \mu^2} \end{bmatrix}$$

A determinação da característica da matriz original é obtida através da matriz Jacobiana da matriz vetorizada que é função dos parâmetros a estimar (σ_{11} , σ_{22} , σ_{33} , σ_{44} , $\pi^2/6/\mu^2$):

$$\text{Jacob}(\text{Vec}(\Omega_{\Delta})) = \begin{bmatrix} \frac{\partial \text{Vec}(\Omega_{\Delta 1})}{\partial \sigma_{11}} & \frac{\partial \text{Vec}(\Omega_{\Delta 1})}{\partial \sigma_{22}} & \frac{\partial \text{Vec}(\Omega_{\Delta 1})}{\partial \sigma_{33}} & \frac{\partial \text{Vec}(\Omega_{\Delta 1})}{\partial \sigma_{44}} & \frac{\partial \text{Vec}(\Omega_{\Delta 1})}{\partial \left(\frac{\pi^2}{6 \cdot \mu^2} \right)} \\ \frac{\partial \text{Vec}(\Omega_{\Delta 2})}{\partial \sigma_{11}} & \frac{\partial \text{Vec}(\Omega_{\Delta 2})}{\partial \sigma_{22}} & \frac{\partial \text{Vec}(\Omega_{\Delta 2})}{\partial \sigma_{33}} & \frac{\partial \text{Vec}(\Omega_{\Delta 2})}{\partial \sigma_{44}} & \frac{\partial \text{Vec}(\Omega_{\Delta 2})}{\partial \left(\frac{\pi^2}{6 \cdot \mu^2} \right)} \\ \frac{\partial \text{Vec}(\Omega_{\Delta 3})}{\partial \sigma_{11}} & \frac{\partial \text{Vec}(\Omega_{\Delta 3})}{\partial \sigma_{22}} & \frac{\partial \text{Vec}(\Omega_{\Delta 3})}{\partial \sigma_{33}} & \frac{\partial \text{Vec}(\Omega_{\Delta 3})}{\partial \sigma_{44}} & \frac{\partial \text{Vec}(\Omega_{\Delta 3})}{\partial \left(\frac{\pi^2}{6 \cdot \mu^2} \right)} \\ \frac{\partial \text{Vec}(\Omega_{\Delta 4})}{\partial \sigma_{11}} & \frac{\partial \text{Vec}(\Omega_{\Delta 4})}{\partial \sigma_{22}} & \frac{\partial \text{Vec}(\Omega_{\Delta 4})}{\partial \sigma_{33}} & \frac{\partial \text{Vec}(\Omega_{\Delta 4})}{\partial \sigma_{44}} & \frac{\partial \text{Vec}(\Omega_{\Delta 4})}{\partial \left(\frac{\pi^2}{6 \cdot \mu^2} \right)} \end{bmatrix} =$$

$$\text{Jacob}(\text{Vec}(\Omega_{\Delta})) = \begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Para a determinação da característica desta matriz pode recorrer-se ao método da eliminação de *Gauss-Jordan* de forma a reduzi-la à matriz escalonada (em escada de linhas) reduzida (única para cada matriz), que corresponde à matriz em que (Barbosa, 2004):

- Todas as linhas não nulas estão acima das linhas nulas;
- O pivô da cada linha está sempre exatamente à direita do pivô da linha acima;
- O pivô de cada linha não nula é a unidade;
- O primeiro elemento não nulo de uma linha é o seu pivô.

O método de eliminação de Gauss-Jordan consiste na tentativa de, após adicionar uma matriz identidade à direita da matriz a reduzir, obter uma matriz identidade na posição da matriz

reduzida, através de operações elementares com os valores das linhas. Estas operações elementares incluem troca de posição de linhas completas; multiplicação dos valores de uma linha por um escalar não nulo e substituindo uma linha pela sua soma com uma linha multiplicada por um escalar não nulo (Barbosa, 2004). A matriz jacobiana anterior já está na sua forma escalonada reduzida pelo que se pode concluir que a sua característica é igual a 4, significando que tem 4 equações linearmente independentes. Como uma das equações tem que ser utilizada para definir a escala (definindo o denominador das restantes) apenas 3 dos 5 parâmetros podem ser estimados (existe assim a necessidade de restringir dois deles).

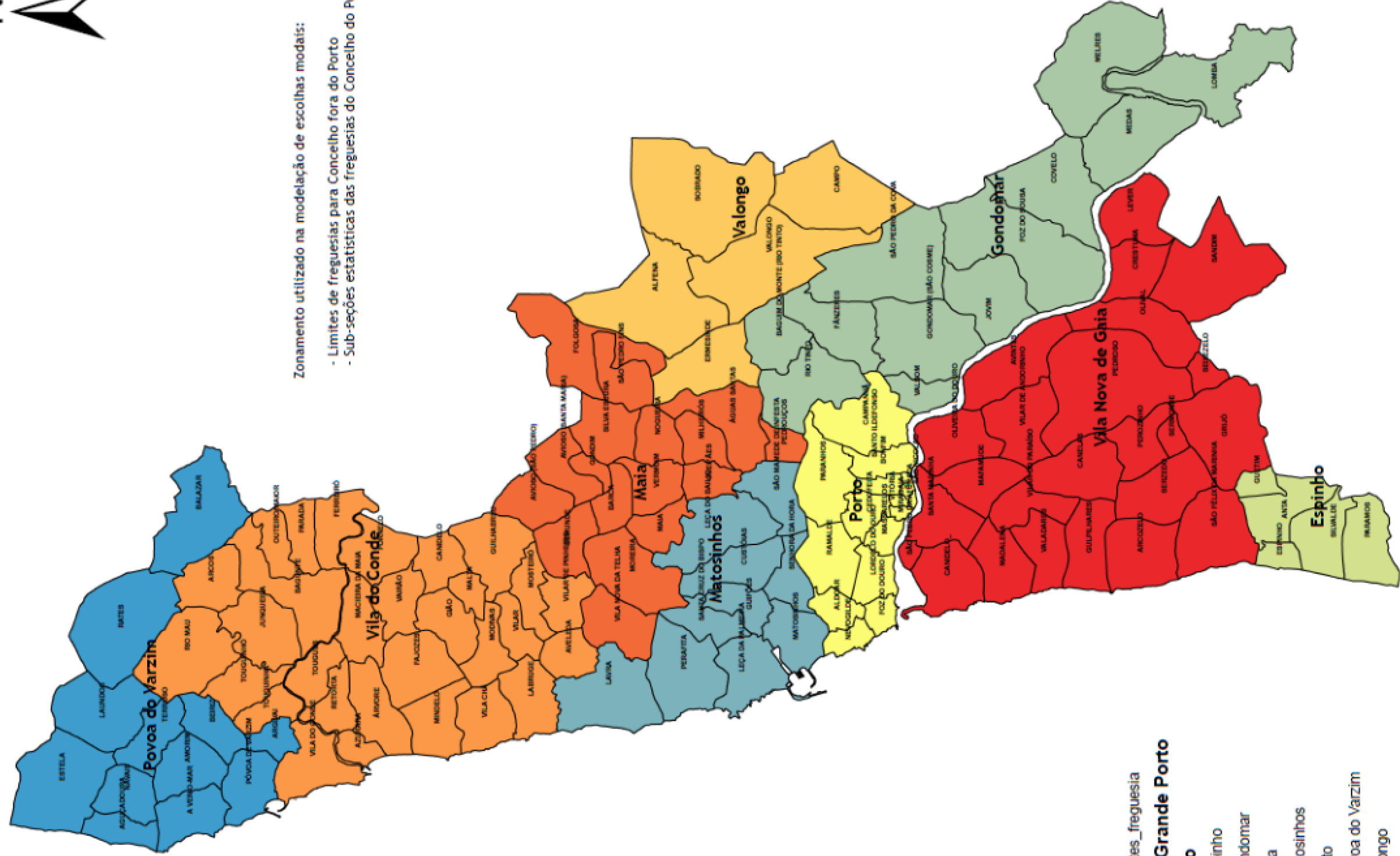
Como exemplo de um escalonamento para a sua forma reduzida de uma matriz apresenta-se o seguinte exemplo:

$$\begin{aligned}
 & \left[\begin{array}{ccc|cccc} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \Rightarrow L_2 = L_2 - L_1 \Rightarrow \left[\begin{array}{ccc|cccc} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \Rightarrow L_1 = L_1 - L_4 \Rightarrow \\
 & \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \Rightarrow L_2 = L_2 + L_3 \Rightarrow \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array} \right] \Rightarrow L_4 = L_4 - 2.L_3 \Rightarrow \\
 & \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 \end{array} \right] \Rightarrow L_2 \leftrightarrow L_3 \Rightarrow \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 \end{array} \right]
 \end{aligned}$$

Esta última matriz está na sua forma escalonada reduzida e como tem apenas duas linhas não nulas, apenas tem duas equações linearmente independentes e assim apenas um parâmetro pode ser calculado (a outra equação definirá a escala).

Anexo 1 do Capítulo IV

Figura A. 12 - Freguesias dos municípios da NUT III - Grande Porto consideradas no zonamento para o caso de estudo



Zonamento utilizado na modelação de escolhas modais:

- Limites de freguesias para Concelho fora do Porto
- Sub-seções estatísticas das freguesias do Concelho do Porto

Legenda

limites_freguesia

Nut III - Grande Porto

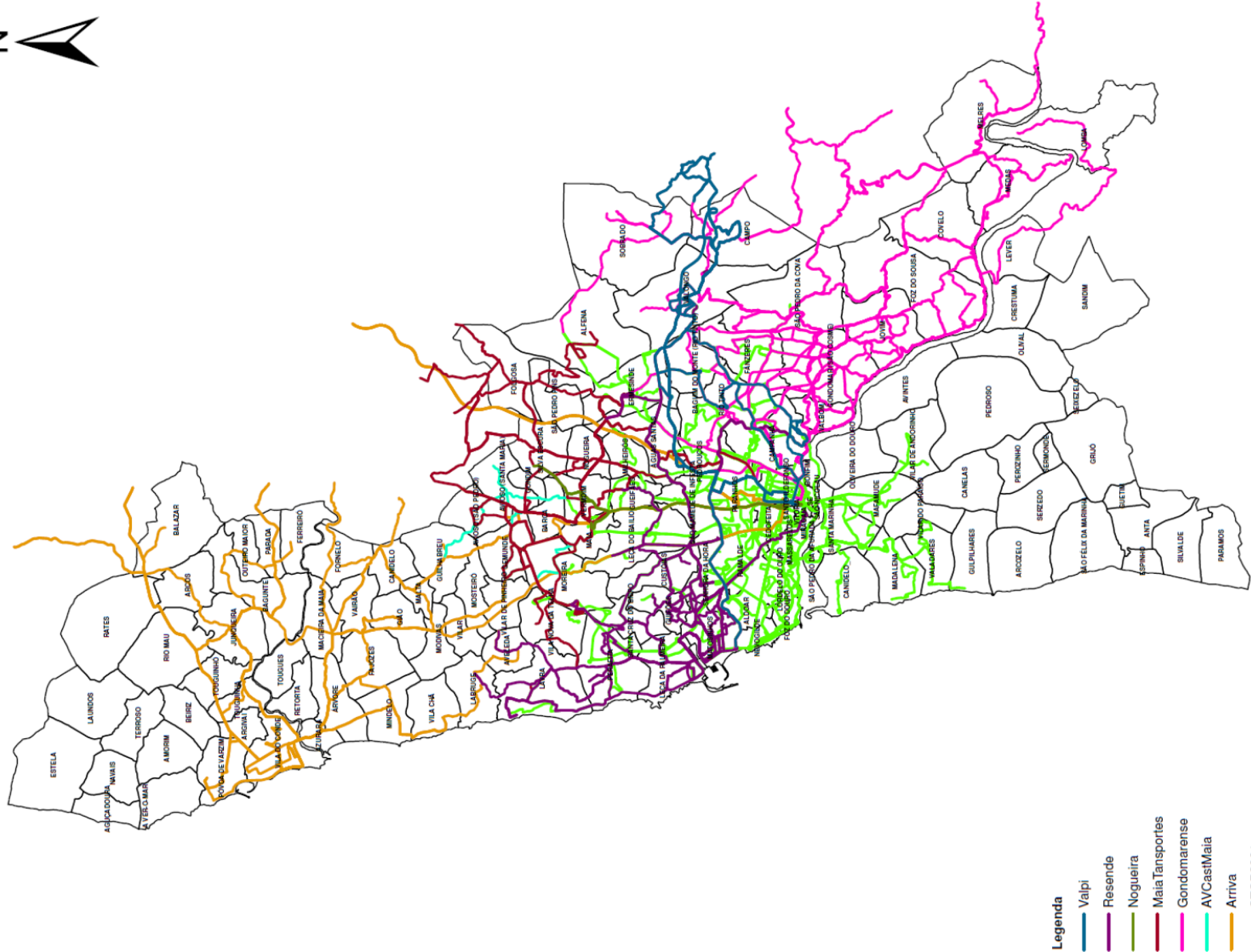
Concelho

Espinho
Gondomar
Maia
Matosinhos
Porto
Póvoa do Varzim
Valongo
Vila Nova de Gaia
Vila do Conde



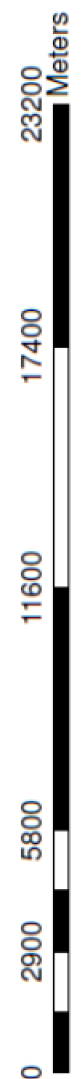
Anexo 2 do Capítulo IV

**Figura A. 13 - Rede de Transportes Coletivos utilizada na análise
do caso de estudo**



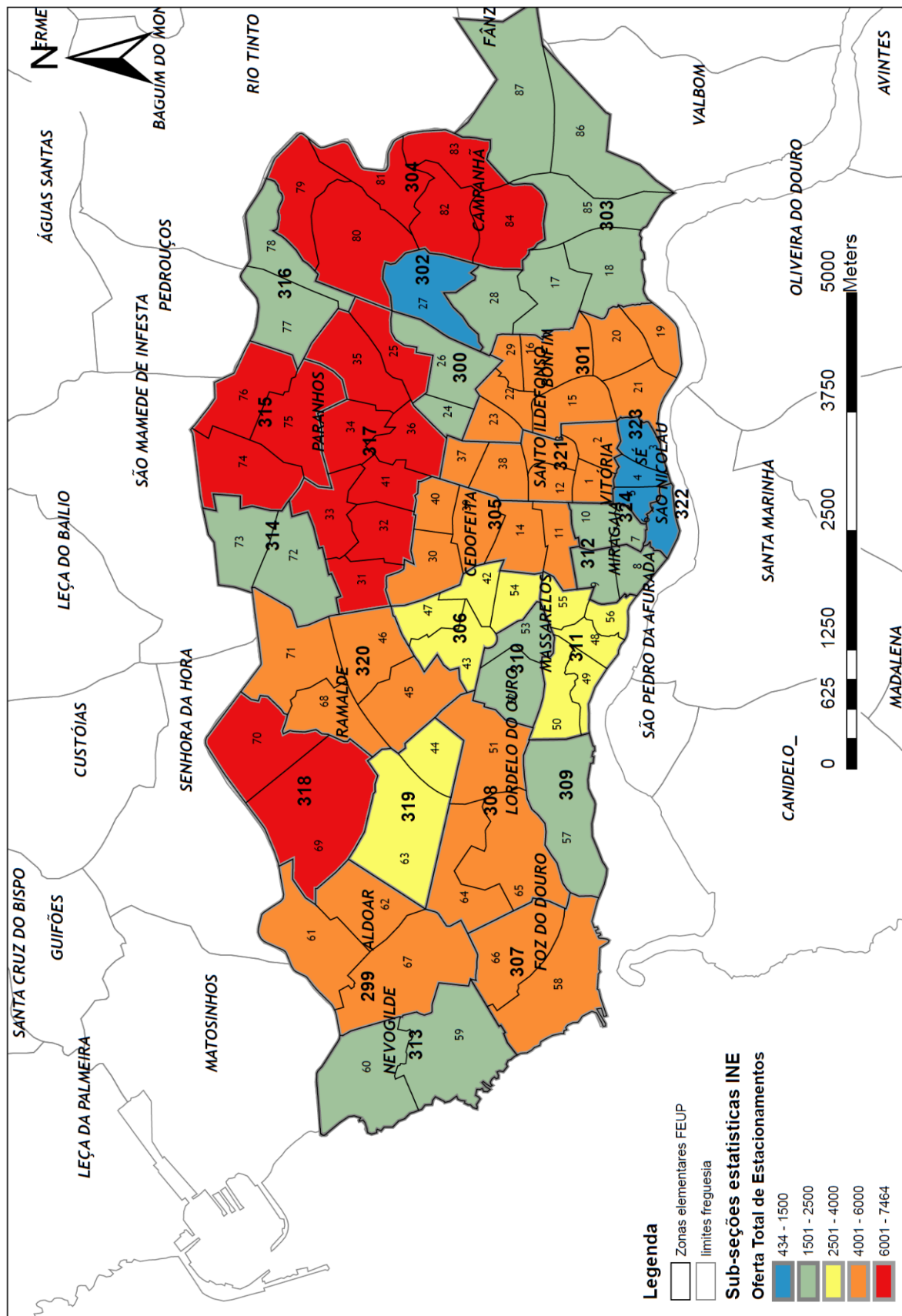
Legenda

- Valpi
- Resende
- Nogueira
- MaiaTansportes
- Gondomarenses
- AVCastMaia
- Arriva
- STCP2001
- limites_freguesia



Anexo 3 do Capítulo IV

Figura A. 14 - Oferta total de estacionamento nas subseções estatísticas do concelho do Porto em 2000 (baseado em Rodrigues *et al.*, 2005)



Anexo 1 do Capítulo VI

Quadro A. 1 - Resultados combinando as variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA >1km e $\leq$ 2km

Licença de condução	Sexo	NCA	Idade	Bus+Walk	%	BUS	%	Outro Bus	%	Moto	%	Walk	%	Auto	%	Total
		-0	0-9	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	0
			10-19	0	0,0%	0	0,0%	0	0,0%	6	100,0%	0	0,0%	0	0,0%	6
			20-34	1	0,8%	20	16,9%	3	2,5%	8	6,8%	65	55,1%	21	17,8%	118
			35-65	3	1,7%	19	10,5%	5	2,8%	28	15,5%	88	54,1%	28	15,5%	181
			+65	4	0,0%	25	23,4%	0	0,0%	0	0,0%	64	59,8%	18	16,8%	107
			Total	4	1,0%	64	15,5%	8	1,9%	42	10,2%	227	55,1%	67	16,3%	412
			0-9	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	0
			10-19	0	0,0%	0	0,0%	0	0,0%	0	0,0%	1	33,3%	2	66,7%	3
			20-34	0	0,0%	4	3,4%	2	1,7%	4	3,4%	31	26,3%	77	65,3%	118
			35-65	0	0,0%	7	5,3%	0	0,0%	6	4,6%	35	26,7%	83	63,4%	131
			+65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	5	26,3%	14	73,7%	19
			Total	0	0,0%	11	4,1%	2	0,7%	10	3,7%	72	26,6%	116	64,9%	271
			0-9	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	0
			10-19	0	0,0%	0	0,0%	0	0,0%	4	25,0%	4	25,0%	8	50,0%	16
			20-34	4	1,6%	0	0,0%	0	0,0%	4	1,6%	22	8,9%	217	87,9%	247
			35-65	0	0,0%	4	1,1%	4	1,1%	14	3,8%	26	7,0%	325	87,1%	373
			+65	2	6,1%	0	0,0%	0	0,0%	0	0,0%	4	12,1%	27	81,8%	33
			Total	6	0,9%	4	0,6%	4	0,6%	22	3,3%	56	8,4%	577	86,2%	669
			0-9	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	-	0,0%	0
			10-19	0	0,0%	0	0,0%	0	0,0%	0	0,0%	5	31,3%	11	68,8%	16
			20-34	0	0,0%	3	3,4%	1	1,1%	0	0,0%	4	4,5%	81	91,0%	89
			35-65	0	0,0%	2	1,7%	0	0,0%	0	0,0%	19	16,0%	88	82,4%	119
			+65	0	0,0%	2	22,2%	0	0,0%	0	0,0%	5	55,6%	2	22,2%	9
			Total	0	0,0%	7	3,0%	1	0,4%	0	0,0%	33	14,2%	192	82,4%	233
			Total 0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			Total 10-19	0	0,0%	0	0,0%	0	0,0%	10	70,0%	10	29,0%	21	61,4%	41
			Total 20-34	5	1,5%	27	13,4%	6	2,0%	16	4,6%	122	37,8%	396	80,4%	572
			Total 35-65	3	1,7%	32	7,6%	9	2,0%	48	10,7%	178	37,8%	534	78,8%	804
			Total +65	2	6,1%	27	23,3%	0	0,0%	0	0,0%	78	54,9%	61	58,8%	168
			Total	10	0,6%	86	5,4%	15	0,9%	74	4,7%	388	24,5%	1012	63,8%	1585
Term		-0	0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			10-19	0	0,0%	0	0,0%	1	20,0%	0	0,0%	4	80,0%	0	0,0%	5
			20-34	0	0,0%	21	16,7%	1	0,0%	10	7,9%	81	64,3%	14	11,1%	126
			35-65	7	5,2%	33	24,4%	1	0,7%	3	2,2%	77	57,0%	14	10,4%	135
			+65	2	8,3%	7	29,2%	0	0,0%	0	0,0%	9	37,5%	6	25,0%	24
			Total	9	3,1%	61	21,0%	2	0,7%	13	4,5%	171	59,0%	34	11,7%	290
			0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			10-19	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			20-34	2	2,6%	5	6,4%	0	0,0%	0	0,0%	22	28,2%	49	62,8%	78
			35-65	2	4,0%	6	12,0%	0	0,0%	0	0,0%	16	32,0%	26	52,0%	50
			+65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	6	100,0%	6
			Total	4	3,0%	11	8,2%	0	0,0%	0	0,0%	38	28,4%	81	60,4%	134
			0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			10-19	3	37,5%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	5	62,5%	8
			20-34	1	0,4%	19	8,0%	1	0,4%	7	2,9%	30	12,6%	180	75,6%	238
			35-65	2	0,7%	16	5,9%	1	0,4%	6	2,2%	78	29,0%	166	61,7%	269
			+65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	2	100,0%	2
			Total	6	1,2%	35	6,8%	2	0,4%	13	2,5%	108	20,9%	353	68,3%	517
			0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			10-19	0	0,0%	2	20,0%	0	0,0%	0	0,0%	3	30,0%	5	50,0%	10
			20-34	0	0,0%	3	2,2%	0	0,0%	0	0,0%	7	5,1%	127	92,7%	137
			35-65	1	0,6%	4	2,3%	0	0,0%	0	0,0%	13	7,5%	156	89,7%	174
			+65	0	0,0%	2	40,0%	0	0,0%	0	0,0%	0	0,0%	3	60,0%	5
			Total	1	0,3%	11	3,4%	0	0,0%	0	0,0%	23	7,1%	291	89,3%	326
			Total 0-9	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			Total 10-19	3	37,5%	2	20,0%	1	20,0%	0	0,0%	7	58,6%	10	56,3%	23
			Total 20-34	3	1,8%	48	11,3%	1	0,4%	17	5,9%	140	44,6%	370	77,4%	579
			Total 35-65	12	3,9%	59	16,7%	2	0,6%	9	2,2%	184	39,5%	362	71,1%	628
			Total +65	2	8,3%	9	31,6%	0	0,0%	0	0,0%	9	37,5%	17	66,5%	37
			Total	20	1,6%	118	9,3%	4	0,3%	26	2,1%	340	26,8%	759	59,3%	1269
		-0	0-9	0	0,0%	5	11,6%	0	0,0%	0	0,0%	32	74,4%	6	14,0%	43
			10-19	0	0,0%	33	14,3%	4	1,7%	4	1,7%	179	77,5%	11	4,8%	231
			20-34	0	0,0%	30	24,6%	0	0,0%	3	2,5%	86	70,5%	3	2,5%	122
			35-65	1	0,5%	25	11,7%	0	0,0%	34	15,9%	150	70,1%	4	1,9%	214
			+65	0	0,0%	38	26,2%	0	0,0%	5	3,4%	99	68,3%	3	2,1%	145
			Total	1	0,1%	131	17,4%	4	0,5%	46	6,1%	546	72,3%	27	3,6%	755
			0-9	0	0,0%	1	9,1%	0	0,0%	0	0,0%	3	27,3%	7	63,6%	11
			10-19	4	6,7%	11	18,3%	1	1,7%	0	0,0%	35	58,3%	9	15,0%	60
			20-34	0	0,0%	2	8,0%	0	0,0%	4	16,0%	17	68,0%	2	8,0%	25
			35-65	0	0,0%	2	11,1%	0	0,0%	4	22,2%	12	66,7%	0	0,0%	18
			+65	0	0,0%	3	42,9%	0	0,0%	0	0,0%	4	57,1%	0	0,0%	7
			Total	4	3,3%	19	15,7%	1	0,8%	8	6,6%	71	58,7%	18	14,9%	121
			0-9	0	0,0%	6	4,8%	3	2,4%	4	3,2%	29	23,2%	83	66,4%	125
			10-19	8	3,0%	22	8,2%	8	3,0%	11	4,1%	148	55,2%	71	26,5%	268
			20-34	0	0,0%	1	4,2%	0	0,0%	0	0,0%	13	54,2%	10	41,7%	24
			35-65	0	0,0%	2	11,8%	0	0,0%	2	11,8%	5	29,4%	8	47,1%	17
			+65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	2	100,0%	2
			Total	8	1,8%	31	7,1%	11	2,5%	17	3,9%	195	44,7%	174	39,9%	436
			0-9	0	0,0%	0	0,0%	2	3,2%	0	0,0%	10	15,9%	51	81,0%	63
			10-19	2	2,1%	2	2,1%	5	5,2%	4	4,2%	32	54,2%	31	32,3%	96
			20-34	0	0,0%	0	0,0%	2	20,0%	6	60,0%	2	20,0%	0	0,0%	10
			35-65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			+65	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0
			Total	2	1,2%	2	1,2%	9	5,3%	10	5,9%	64	37,9%	82	48,5%	169
			Total 0-9	0	0,0%	12	8,0%	5	2,7%	4	3,2%	74	44,5%	147	69,3%	242
			Total 10-19	14	3,9%	68	12,6%	18	3,3%	19	3,6%	414	65,0%	122	25,2%	655
			Total 20-34	0	0,0%	33	23,0%	2	20,0%	2	33,2%	118	67,5%	15	29,3%	181
			Total 35-65	1	0,5%	29	11,6%	0	0,0%	40	16,3%	167	68,6%	12	32,0%	249
			Total +65	0	0,0%	41	27,4%	0	0,0%	5	3,4%	103	67,8%	5	41,2%	154
			Total	15	1,0%	183	12,4%	25	1,7%	81	5,5%	876	59,1%	301	20,3%	1481
Não tem		-0	0-9	1	2,4%	7	17,1%	2	4,9%	2	4,9%	25	61,0%	4	9,8%	41
			10-19	23	8,6%	45	16,7%	13	4,8%	4	1,5%	177	65,8%	7	2,6%	269

Anexo 2 do Capítulo VI

Quadro A. 2 - Resultados combinando as variáveis licença de condução, taxa de motorização familiar por adulto, sexo e escalão etários para viagens com distância entre a zona de GA > 2km

Licença de condução	Sexo	NCA	Idade	Bus+Walk	%	BUS	%	Outro Bus	%	Moto	%	Walk	%	Auto	%	Total
	M	-0	0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
			10-19	0	0.0%	17	47.2%	0	0.0%	12	33.3%	0	0.0%	7	19.4%	36
			20-34	12	1.7%	147	21.4%	49	7.1%	110	16.0%	6	0.9%	82	52.8%	686
			35-65	25	2.2%	268	23.3%	60	5.2%	196	17.0%	15	1.3%	387	51.0%	1151
			+65	14	3.3%	140	33.0%	3	0.7%	20	4.7%	14	3.3%	23	55.0%	424
			Total	51	2.2%	572	24.9%	112	4.9%	338	14.7%	35	1.5%	1189	51.8%	2297
			0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
			10-19	0	0.0%	13	31.0%	3	7.1%	7	16.7%	0	0.0%	19	45.2%	42
			20-34	4	0.4%	64	6.4%	14	1.4%	66	6.6%	2	0.2%	84	84.9%	996
			35-65	4	0.3%	64	5.2%	23	1.9%	90	7.3%	6	0.5%	1039	84.7%	1226
	>0.58<1	+65	3	3.0%	20	20.0%	0	0.0%	4	4.0%	1	1.0%	72	72.0%	100	
		Total	11	0.5%	161	6.8%	40	1.7%	167	7.1%	9	0.4%	1976	83.6%	2364	
		0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		10-19	3	2.4%	33	26.2%	3	2.4%	18	14.3%	0	0.0%	69	54.8%	126	
		20-34	9	0.3%	55	1.6%	39	1.1%	133	4.4%	3	0.1%	3210	92.5%	3469	
		35-65	13	0.3%	84	1.8%	89	1.9%	158	3.3%	3	0.1%	4416	92.7%	4763	
		+65	2	0.4%	23	5.1%	0	0.0%	3	0.7%	1	0.2%	418	93.3%	447	
		Total	27	0.3%	195	2.2%	131	1.5%	332	3.8%	7	0.1%	8113	92.1%	8805	
		0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		10-19	0	0.0%	20	9.1%	2	0.9%	18	8.2%	1	0.5%	178	81.3%	219	
	>=1	20-34	0	0.0%	17	1.3%	13	1.0%	38	2.9%	0	0.0%	1259	94.9%	1327	
		35-65	0	0.0%	17	0.9%	6	0.3%	9	0.5%	1	0.1%	1888	98.3%	1921	
		+65	0	0.0%	6	4.5%	0	0.0%	0	0.0%	2	1.5%	126	94.0%	134	
		Total	0	0.0%	60	1.7%	21	0.6%	65	1.8%	4	0.1%	3561	95.8%	3601	
		Total 0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		Total 10-19	3	2.4%	83	27.1%	8	3.8%	55	16.8%	1	0.5%	273	70.5%	423	
		Total 20-34	25	1.0%	283	13.0%	115	3.7%	367	8.1%	11	0.5%	5677	89.4%	6478	
		Total 35-65	42	1.4%	433	15.6%	178	2.9%	453	10.0%	25	0.9%	7930	89.9%	9061	
		Total +65	19	3.0%	189	27.3%	3	0.7%	27	4.2%	18	2.8%	849	81.2%	1105	
		Total	89	0.5%	988	5.8%	304	1.8%	902	5.3%	55	0.3%	14729	86.3%	17067	
	F	-0	0-9	1	3.3%	7	23.3%	0	0.0%	10	31.3%	0	0.0%	12	40.0%	30
			20-34	13	2.2%	194	33.4%	22	3.8%	92	15.8%	9	1.5%	251	43.2%	581
			35-65	44	8.3%	222	42.0%	9	1.7%	47	8.9%	12	2.3%	195	36.9%	529
			+65	5	7.4%	25	36.8%	2	2.9%	0	0.0%	4	5.9%	32	47.1%	68
			Total	63	5.2%	448	37.1%	33	2.7%	149	12.3%	25	2.1%	460	40.6%	1208
			0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
			10-19	0	0.0%	7	26.9%	0	0.0%	0	0.0%	0	0.0%	19	73.1%	26
			20-34	11	1.4%	147	19.3%	12	1.6%	39	5.1%	5	0.7%	547	71.9%	761
			35-65	10	2.1%	87	18.0%	11	2.3%	12	2.5%	2	0.4%	380	74.7%	482
			+65	1	16.7%	1	16.7%	0	0.0%	0	0.0%	0	0.0%	4	66.7%	6
	>0.58<1	Total	22	1.7%	242	19.0%	23	1.8%	51	4.0%	7	0.5%	930	72.9%	1275	
		0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		10-19	6	5.8%	32	30.8%	2	1.9%	6	5.8%	0	0.0%	38	55.8%	104	
		20-34	18	0.7%	258	9.9%	23	0.9%	112	4.3%	5	0.2%	2198	84.1%	2614	
		35-65	17	0.7%	285	12.3%	21	0.9%	69	3.0%	14	0.6%	1918	82.5%	2324	
		+65	0	0.0%	17	20.5%	2	2.4%	0	0.0%	0	0.0%	64	77.1%	83	
		Total	41	0.8%	592	11.6%	48	0.9%	187	3.6%	19	0.4%	4238	82.7%	5125	
		0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		10-19	2	1.7%	10	8.5%	0	0.0%	13	11.1%	0	0.0%	92	78.6%	117	
		20-34	0	0.0%	27	1.7%	6	0.4%	11	0.7%	2	0.1%	1350	97.1%	1596	
	>=1	35-65	3	0.1%	23	1.1%	5	0.2%	11	0.5%	0	0.0%	2090	98.0%	2132	
		+65	0	0.0%	7	12.3%	0	0.0%	0	0.0%	0	0.0%	50	87.7%	57	
		Total	5	0.1%	67	1.7%	11	0.3%	35	0.9%	2	0.1%	3782	96.9%	3902	
		Total 0-9	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	
		Total 10-19	9	4.6%	56	25.4%	2	1.9%	29	17.7%	0	0.0%	181	68.2%	277	
		Total 20-34	42	1.4%	626	19.0%	63	2.0%	254	8.4%	21	0.9%	4546	84.8%	5552	
		Total 35-65	74	5.4%	617	23.3%	46	1.3%	139	4.7%	28	1.3%	4563	87.1%	5467	
		Total +65	6	8.9%	50	27.4%	4	2.7%	0	0.0%	4	5.9%	150	74.0%	214	
		Total	131	1.1%	1349	11.7%	115	1.0%	422	3.7%	53	0.5%	9440	82.0%	11510	
		0-9	11	4.4%	95	37.8%	13	5.2%	17	7.8%	2	0.8%	113	45.0%	251	
	M	-0	10-19	102	9.9%	534	52.0%	175	17.1%	80	7.8%	24	2.3%	111	10.8%	1026
			20-34	21	3.6%	229	39.4%	33	5.7%	222	38.2%	14	2.4%	62	10.7%	581
			35-65	27	2.9%	364	39.7%	58	6.3%	333	38.5%	35	3.8%	81	8.8%	918
			+65	27	6.6%	241	59.1%	9	2.2%	70	17.2%	28	6.9%	33	8.1%	408
			Total	188	5.9%	1463	45.9%	288	9.0%	742	23.3%	103	3.2%	400	12.6%	3184
			0-9	3	3.9%	13	17.1%	8	10.5%	0	0.0%	0	0.0%	32	68.4%	76
			10-19	26	7.4%	151	43.0%	56	16.0%	35	10.0%	2	0.6%	81	23.1%	351
			20-34	9	5.0%	57	29.6%	11	6.1%	30	27.9%	1	0.6%	55	30.7%	179
			35-65	6	3.0%	51	25.2%	14	6.9%	81	40.1%	2	1.0%	48	23.8%	202
			+65	10	10.6%	45	47.9%	2	2.1%	18	19.1%	3	3.2%	16	17.0%	94
	>0.58<1	Total	54	6.0%	313	34.7%	91	10.1%	184	20.4%	8	0.9%	252	27.9%	902	
		0-9	4	0.4%	79	8.0%	46	4.7%	20	2.0%	5	0.5%	828	84.3%	982	
		10-19	80	4.3%	672	36.3%	235	12.7%	159	8.6%	22	1.2%	681	36.8%	1849	
		20-34	4	3.0%	14	10.4%	13	9.6%	31	23.0%	1	0.7%	72	53.3%	135	
		35-65	4	2.6%	27	17.6%	14	9.2%	54	35.3%	0	0.0%	54	35.3%	133	
		+65	3	5.6%	14	25.9%	4	7.4%	18	33.3%	0	0.0%	15	27.8%	54	
		Total	95	3.0%	886	25.4%	312	9.8%	282	8.9%	28	0.9%	1650	52.0%	3173	
		0-9	0	0.0%	7	1.1%	28	4.3%	2	0.3%	1	0.2%	614	94.2%	652	
		10-19	17	1.9%	226	27.0%	61	7.0%	42	4.8%	4	0.5%	354	58.8%	874	
		20-34	0	0.0%	1	4.8%	0	0.0%	5	23.8%	0	0.0%	15	71.4%	21	
	>=1	35-65	0	0.0%	2	16.7%	0	0.0%	8	66.7%	0	0.0%	2	16.7%	12	
		+65	0	0.0%	2	33.3%	0	0.0%	0	0.0%	0	0.0%	4	66.7%	6	
		Total	17	1.1%	248	15.8%	89	5.7%	37	3.6%	5	0.3%	149	73.4%	1565	
		Total 0-9	18	3.4%	194	23.0%	95	5.1%	39	4.0%	8	0.5%	1607	84.8%	1961	
		Total 10-19	225	7.0%	1593	40.9%	527	13.8%	316	8.0%	52	1.6%	1387	42.1%	4100	
		Total 20-34	34	3.9%	297	36.2%	57	6.7%	308	34.8%	16	2.2%	204	35.6%	916	
		Total 35-65	37	2.9%	444	36.6%	86	6.9%	496	38.8%	37	3.7%	185	20.5%	1285	
		Total +65	40	7.5%	302	55.7%	15	3.6%	106	20.2%	31	6.5%	68	18.0%	562	
		Total	354	4.0%	4280	32.1%	780	8.8%	1265	14.3%	144	1.6%	3421	39.1%	8824	
	M	-0	0-9	7	2.7%	107	41.2%	23	8.8%	17	6.5%	5	1.9%	101	38.8%	260
			10-19	92	8.6%	543	50.9%	170	15.9%	35	5.2%	39	3.7%	168	15.7%	1067
			20-34	86	9.1%	494	52.2%	33	3.5%	173	18.3%	25	2.6%	136	14.4%	947
			35-65	143	7.0%	1158	56.6%	81	4.0%	221	10.8%	71	3.5%	373	18.2%	2047
			+65	64	6.7%	531	55.4%	28	2.9%	21	2.2%	65	6.8%	250	26.1%	959
			Total	392	7.4%	2833	53.7%	335	6.3%	487	9.2%	205	3.9%	1028	19.5%	5280
			0-9	0	0.0%	8	13.6%	5	8.5%	4	6.8%	1	1.7%	47	69.5%	59
			10-19	22	6.3%	164	46.7%	56	16.0%	6	1.7%	1	0.3%	102	29.1%	351
			20-34	17	4.6%	159	43.1%	13	3.5%	45	12.2%	8	2.2%	127	34.4%	369
			35-65	53	6.2%	371	43.1%	31	3.6%	48	5.6%	11	1.3%	546	40.2%	860
	>0.58<1	+65	9	6.4%	49	35.0%	5	3.6%	3	2.1%	3	2.1%	1	50.7%	140	
		Total	101	5.7%												

