

Universidade do Minho
Escola de Engenharia

Harley Alexánder Largo Amado

Modelo de procura de transporte público em autocarro
para territórios de baixa densidade

Julho de 2018
27 de julho de 2018



Universidade do Minho
Escola de Engenharia

Harley Alexnder Largo Amado

Modelo de procura de transporte pblico em autocarro
para territrios de baixa densidade populacional

Dissertao de Mestrado
Mestrado Integrado em Engenharia Civil

Ciclo de Estudos Integrados Conducentes ao
Grau de Mestre em Engenharia Civil

Trabalho efetuado sob a orientao do Professor Doutor
Paulo Jorge Gomes Ribeiro e da Eng.^a Claudia Patricia
Baez Trujillo

Julho de 2018

AGRADECIMENTOS

Todos mis logros tanto profesionales como personales serán siempre dedicados a Dios.

Agradecer en primer lugar a mi madre Mariela Peñaloza, porque sin ella este sueño llamado ingeniería civil no sería relevante para mi vida, gracias por enseñarme a dar siempre lo mejor de mí sin importar cansancio, críticas o sufrimiento. Para ella mi más profundo respeto y admiración.

Importante agradecer a mis compañeros de “batalla”, ellos que a lo largo de estos 6 años académicos estuvieron ahí con una sonrisa, un apoyo, un consejo o simplemente con la presencia. Ellos hacen parte de este logro, y como se dijo una vez “algunos más que otros pero ninguno menos”. A mi familia y no necesariamente de sangre, espero sepan cuánto los amo; para ellos esta vida y la otra, gracias por estar siempre a mi lado en cada momento de mi vida. Ustedes son una de mis mayores bendiciones.

Es importante mencionar a todas las personas que han marcado mi vida y que han aportando una semilla a lo largo de mi recorrido académico y con esto no solamente me refiero a estudiantes, sino también a profesores y empleados de las dos universidades a la que tuve el privilegio de pertenecer, no se imaginan las enseñanzas que le han dejado a este hombre, gracias por permitirme tener el placer de conocerlos y aprender.

Gracias a la Ing. Claudia Baez, por su mentoría en este proyecto investigativo, pero estaré más agradecido por ser mi primera mentora en mi vida académica y profesional, una mujer con un corazón enorme y siempre dispuesta para hacer las cosas de la mejor manera posible, una buena enseñanza para toda la vida.

Por último, quiero dar una mención especial al profesor Paulo Ribeiro, por instruirme a lo largo de esta investigación, enseñarme lo espectacular labor que es investigar y por estar siempre pendiente de cualquier inquietud e invertir gran parte de su tiempo para poder culminar este proyecto satisfactoriamente, a usted mil gracias.

RESUMO

Os territórios de baixa densidade são atualmente alvo de particular atenção em vários domínios do desenvolvimento territorial, económico e social de regiões e países, nomeadamente no setor da mobilidade e transportes, como é o caso de Portugal.

A compreensão dos principais fatores que influenciam a procura de transporte público em autocarro para os territórios de baixa densidade em Portugal Continental, bem como o desenvolvimento do modelo que permita estimar o número de viagens em autocarro geradas nesses territórios assume particular interesse no processo de planeamento de transportes à luz da entrada em vigor em Portugal da Lei nº 52/2015. Deste modo, nesta dissertação é apresentado um modelo de regressão linear múltipla que sustenta a análise e possibilita a estimação da procura em função de variáveis sócio-económicas e demográficas. Para desenvolver este modelo foi usado a ferramenta estatística SPSS (*Statistical Package for the Social Sciences*).

Com base na pesquisa bibliográfica selecionaram-se 11 variáveis que potencialmente terão influência na procura de transporte público por autocarro, tendo sido apenas utilizadas 6 dessas variáveis dada a disponibilidade de informação existente para os municípios portugueses. Em Portugal estão classificados 165 municípios como territórios de baixa densidade que servirão de base para calibrar e validar o modelo encontrado, tendo por base a informação existente nos Censos 2011 do Instituto estatístico nacional (INE).

O modelo de procura de transporte em autocarro para os municípios de baixa densidade é caracterizado pelas seguintes variáveis explicativas: número de pessoas sem nível de escolaridade, densidade populacional e o poder de compra. O ajuste para o modelo encontrado foi de 87%, tendo sido determinado por amostra de 114 municípios, já que outros municípios foram retirados por serem considerados *outliers*, observações influentes ou, então, utilizados para a respetiva validação. Do processo de validação foi possível constatar que o modelo apresentou um bom ajuste à realidade com cerca de 80% dos resultados para os municípios a apresentarem um erro inferior ao 50%, ou seja, o modelo apresenta uma boa explicação da variável do número de viagens para o transporte público.

Palavras-chave: Territórios de baixa densidade, Modelos de procura de transporte, Transporte Público, Autocarro

ABSTRACT

Low-density regions are currently receiving special attention in various areas of territorial, economic, and social development in regions and countries, particularly in the mobility and transport sector, such as Portugal.

The understanding of the main factors influencing the demand for public bus transportation to the low-density areas in Portugal, as well as the development of the model that allows estimating the number of bus trips generated in these territories takes a special interest in the planning process. transport considering the entry into force in Portugal of Law 52/2015. Thus, this thesis presents a multiple linear regression model that supports the analysis and enables the estimation of demand as a function of socio-economic and demographic variables. To develop this model was used the statistical tool SPSS (Statistical Package for the Social Sciences).

Based on the research, eleven variables were selected that will potentially influence the demand for public transportation by bus, with only six of these variables being used given the availability of information for the Portuguese municipalities. In Portugal, 165 municipalities are classified as low-density territories that will serve as a basis for calibrating and validating the model found, based on the information existing in the 2011 National Statistics Institute (INE) Census.

The bus transportation demand model for low density municipalities is characterized by the following explanatory variables: number of people without education level, population density and purchasing power. The adjustment for the model found was 87%, and was determined by a sample of 114 municipalities, since other municipalities were withdrawn because they were considered outliers, influential observations or, then, used for their validation. From the validation process it was possible to verify that the model presented a good adjustment to the reality with about 80% of the results for the municipalities presenting an error lower than 50%, that is, the model presents a good explanation of the variable of the number of trips for public transport.

Keywords: Low-density territories, Transportation search models, Public transport, Bus

ÍNDICE

AGRADECIMENTOS	2
RESUMO	3
ABSTRACT	5
ÍNDICE.....	7
ÍNDICE DE FIGURAS	10
ÍNDICE DE QUADROS	12
1 INTRODUÇÃO.....	14
1.1 Enquadramento	14
1.2 Objetivos.....	17
1.3 Estrutura.....	17
2 ESTADO DE ARTE.....	19
2.1 Territórios de baixa densidade	19
2.1.1 A mobilidade nos TBD.....	30
2.2 Modelos de procura de transporte.....	32
2.2.1 Modelos de procura de transporte	37
2.2.2 Modelo clássico de quatro etapas	46
2.2.2.1 Modelos de Geração de Viagens.....	48
2.2.2.2 Modelos de Distribuição de Viagens	53
2.2.2.3 Repartição modal.....	57
2.2.3 Outros métodos e modelos para a estimação da procura.....	60
3 METODOLOGIA.....	63
3.1 Enquadramento metodológico	63
3.2 Definição de variáveis	65
3.3 Os dados e a sua obtenção	67
3.3.1 Seleção da amostra	68
3.4 Metodologia para o desenvolvimento do Modelo de Regressão Linear para estimar a procura de passageiros de autocarro em TBD	71
3.4.1 Definição da escala das variáveis	71
3.4.2 Análise dos dados	71
3.4.3 Análise univariada dos dados	72
3.4.3.1 Estudo da simetria e do achatamento	72
3.4.3.2 Estudo dos valores atípicos - “outliers”	74
3.4.3.3 Normalidade	75
3.4.4 Transformação de variáveis.....	77
3.4.5 Análise bivariada dos dados	79
3.4.6 Modelo de Regressão Linear Múltipla	82
3.4.6.1 Objetivos	82
3.4.7 A formulação geral do modelo	82
3.4.8 Hipóteses do modelo	83
3.4.9 Estimação do modelo	85

3.4.10	Diagnóstico do modelo estimado	86
3.4.10.1	Hipóteses de Homocedasticidade (variância constante)	86
3.4.10.2	Diagnóstico de independência.....	87
3.4.10.3	Diagnóstico de normalidade.....	87
3.4.10.4	Diagnóstico da multicolinearidade.....	88
3.4.11	Análises de Outliers e observações influentes.....	89
3.5	Validação do modelo	90
4	CASO DE ESTUDO	92
4.1	Análise dos dados e variáveis para o MRLM	94
4.1.1	Análise Univariada	94
4.1.2	Análise dos valores extremos	96
4.1.3	Seleção das variáveis significativas (análise bivariada).....	97
4.2	Modelo de procura de transporte para os territórios de baixa densidade de Portugal continente – regressão linear múltipla	101
4.2.1	Introdução.....	101
4.2.2	Estimação do modelo de Regressão Linear Múltipla	102
4.2.3	Diagnóstico do modelo do ponto de vista estatístico	109
4.2.3.1	Hipóteses de Homocedasticidade.....	109
4.2.3.2	Diagnóstico de independência.....	109
4.2.3.3	Diagnóstico de normalidade.....	110
4.2.3.4	Diagnóstico da multicolinearidade.....	112
4.2.4	Análise dos outliers e observações influentes	114
4.2.4.1	Resíduos	114
4.2.4.2	Leverage	116
4.2.4.3	Valor estimado ajustado	117
4.2.4.4	Distância de Cook	118
4.2.4.5	Df Beta	119
4.2.4.6	Df Fit	122
4.2.4.7	Conclusão da análise de resíduos e valores influentes.....	123
4.3	Estimação do novo modelo – MRLM - para estimar viagens de autocarro em TBDs 124	
4.4	Estimação de modelos estratificados	131
4.4.1	Seleção de grupos de municípios	131
4.4.2	Grupo 1 - densidade populacional entre 0 e 50 hab/km ²	132
4.4.3	Grupo 2 - densidade populacional entre 50 e os 100 hab/Km ²	134
4.4.4	Grupo 3 - Densidade populacional maior a 100 hab/Km ²	135
4.4.5	Validação do modelo – Análise de resíduos e observações influentes.....	136
5	ANALEASES DE RESULTADOS	138
5.1	Validação do modelo	138

5.2	Análise de resultados para o modelo de estimação de viagens em autocarro para os TBDS em Portugal continental.	142
5.3	Análises de resultados por regiões.....	146
5.3.1	Região Norte.....	146
5.3.2	Região Centro.....	148
5.3.3	Região Sul.....	150
5.3.4	Comparações de resultados das três regiões.....	153
5.4	Resultados para os municípios considerados outliers e influentes.....	155
6	CONCLUSÕES E DESENVOLVIMENTOS FUTUROS.....	157
6.1	Trabalhos futuros.....	161
7	REFERÊNCIAS.....	162
	ANEXO A.....	7-A
	ANEXO B.....	7-A
	ANEXO C.....	7-A
	ANEXO D.....	7-B
	ANEXO E.....	7-H

ÍNDICE DE FIGURAS

Figura 1 Limites das NUTS 2013.....	26
Figura 2 - Classificação por freguesias para os Territorios de Baixa Densidade	29
Figura 3. Processo de análise da procura em planejação de transporte. Fonte: (Cal Mayor & Asociados, 2005).	36
Figura 4 - Exemplos de modelos físicos. Fonte:	38
Figura 5. Clasificação e composição dos modelos abstratos. Fonte: Elaboração propia	39
Figura 6. Variação do erro em função da dificuldade do modelo. Fonte:(Ortúzar, 2015)	45
Figura 7 – Gráficos de verificação da normalidade para a variável desempregados: a)Normal Q-Q plots b)Detrended Normal Q-Q plot.....	76
Figura 8 – Transformações das distribuições Fonte: (Rummet, 1970).....	78
Figura 9 – a) análise da variância com $Y=Zre$ e $X=Pre$ b) Análise da variância com $Y=Sresid$ e $X=Zpred$	86
Figura 10 – Caixa de bidoges para a variável de desemprego.....	96
Figura 11 – Gráficos de análise da variância.....	109
Figura 12 - Gráfico Normal Q-Q Plot, teste à normalidade	111
Figura 13 - Gráfico Detrended Normal Q-Q plot, teste à normalidade.....	112
Figura 14 – Distribuição dos resíduos	112
Figura 15 – Resíduos estandardizados	114
Figura 16 – Resíduos estudantizados	115
Figura 17 – Resíduos estudantizados deleted.....	116
Figura 18 – Gráfico de Leverage.....	117
Figura 19 – Valores estimados ajustados	118
Figura 20 – Distância de Cook	119
Figura 21 – DfBetas estandardizado para a constante.....	120
Figura 22 – DfBeta estandardizado para densidade	120
Figura 23 – DfBeta estandardizado para sem nivel de escolaridade	121
Figura 24 – DfBeta estandardizado para poder de compra	121
Figura 25 – DfFit estandardizado	122
Figura 26 – Gráficos de análises de variância	126
Figura 27 – Gráfico Normal Q-Q plot, teste à normalidade.....	128
Figura 28 – Gráfico Detrended Normal Q-Q plot, teste à normalidade	128
Figura 29 - Histograma da densidade populacional	132

Figura 30 – Gráfico de distribuição do erro	140
Figura 31 – Distribuição do modelo Portugal continental.....	143
Figura 32 – Distribuição do erro pelos municípios da região norte	148
Figura 33 – Distribuição dos municípios na região Centro	150
Figura 34 – Distribuição dos municípios na região Sul.....	152
Figura 35 – Distribuição das regiões para Portugal Continental	153
Figura 36 – Caixas de bidoges para regiões	154
Figura 37 – Distribuição de outliers e municípios influentes	155

ÍNDICE DE QUADROS

Tabela 1 – Critérios para a determinação das áreas rurais e urbanas Fonte: (IBGE, 2017).....	24
Tabela 2. Classificação do território em NUTS segundo a população dos países da União Europeia. Fonte: (Instituto Nacional de Estatística, 2007).....	26
Tabela 3 - Composição de cada nível de NUTS.....	27
Tabela 4. Matriz de distribuição. Fonte: (Ortúzar, 2000).....	53
Tabela 5 - Quociente de simetria utilizado pelo SPSS (Pestana & Gageiro, 2014).....	74
Tabela 6 – Medidas estatísticas para todos os tipos de variáveis (adaptado de Pestana & Gageiro, 2014).....	79
Tabela 7 – Intensidade de correlação através do valor do R de Pearson.....	81
Tabela 8 – Tabela de decisão em função de d_L e d_U Fonte: (Rodrigues, 2012).....	87
Tabela 9 – Municípios para validação.....	90
Tabela 10 – Folha tipo análises de variáveis.....	95
Tabela 11 – Dados dos municípios considerados valores aberrantes “Outliers”.....	97
Tabela 12 – Testes de Normalidade para as variáveis em análise.....	98
Tabela 13 – Excerto da tabela com os coeficientes de R Pearson (anexo B).....	99
Tabela 14 – Excerto da tabela com os coeficientes de Ró de Spearman (Anexo B).....	99
Tabela 15 – Nivel de correlação com base no Ró de Spearman.....	100
Tabela 16 – Sumário do modelo.....	103
Tabela 17 – Resultados de correlação.....	104
Tabela 18 – Resultado dos coeficientes.....	105
Tabela 19 – Sumário do modelo.....	106
Tabela 20 – Correlações do modelo.....	106
Tabela 21 – Resultados de Anova para o modelo.....	107
Tabela 22 – Resultados dos coeficientes para o modelo.....	108
Tabela 23 – Verificação da independência.....	110
Tabela 24 – Teste à normalidade dos resíduos.....	111
Tabela 25 – Tolerância e VIF.....	113
Tabela 26 – Municípios Outliers e influentes.....	123
Tabela 27 – Sumário do modelo.....	125
Tabela 28 - Resultados de Anova.....	125
Tabela 29 - Resultado dos coeficientes.....	125
Tabela 30 – Verificação da independência.....	127

Tabela 31 – Teste à normalidade dos resíduos	127
Tabela 32 – Correlações entre as variáveis	129
Tabela 33 – Tolerância e VIF	129
Tabela 34 – Sumário do modelo – Grupo 1	133
Tabela 35 – Resultados de coeficientes – Grupo 1	133
Tabela 36 – Sumário do modelo – grupo 2	134
Tabela 37 – Resultados dos coeficientes – grupo 2	134
Tabela 38 – Sumário do modelo – grupo 3	135
Tabela 39 – Resultados dos coeficientes – grupo 3	136
Tabela 40 – Excerto de síntesis de diagnóstico	137
Tabela 41 – Análise dos resultados da validação (real vs modelado)	139
Tabela 42 – Resultados para os municípios utilizados na validação do modelo	140
Tabela 43 – Resultados para Portugal continental	144
Tabela 44 – Resultados para os 8 municípios com erro superior a 75%	145
Tabela 45 – Resultados para os 8 municípios com erro superior a 75%	147
Tabela 46 – Distribuição do erro (resultados modelados vs reais) para a região Norte	147
Tabela 47 – Resultados para os 3 municípios com um erro superior a 90%	148
Tabela 48 – Distribuição do erro (resultados modelados vs reais) para a região Centro	149
Tabela 49 – Resultados para os 3 municípios com um erro superior a 90%	151
Tabela 50 – Distribuição do erro (resultados modelados vs reais) para a região Sul	151
Tabela 51 – Análises de Outliers	156

1 INTRODUÇÃO

1.1 Enquadramento

O sistema de transportes, como um elemento basilar da economia nacional e mundial, é dinâmico e integrado por uma série de elementos relacionados entre si, esta interação visa prestar um serviço que permita o deslocamento eficiente, económico e seguro de pessoas e bens.

Um dos setores deste sistema é o transporte público rodoviário, que ao longo do tempo tem sofrido alterações significativas, nomeadamente em termos de oferta, procura, gestão e operação das empresas e tipo de autoridades que intervêm neste setor. Em muitas situações a oferta não acompanha a evolução da procura, e vice-versa, criando desequilíbrios estruturais, que reduzem a qualidade do serviço e consequentemente a atração de novos utentes (passageiros).

Em certas condições o transporte público tradicional não se consegue afirmar e constituir uma solução de mobilidade alternativa ao uso do automóvel. Um fator que afeta significativamente o desempenho do transporte público é a forma com são, ou foram, planeadas e projetadas as cidades, sobretudo em situações onde a tendência de crescimento das áreas urbanas consistiu na criação de zonas residências periféricas pouco densas, tendo como consequência a degradação da qualidade da oferta do transporte público. Assim, menos passageiros conduzirão a menores receitas, levando os operadores a tendencialmente prestarem um serviço de pior qualidade, que por sua vez irá conduzir a uma redução do número de passageiros, entrando-se desta feita num ciclo vicioso de espiral negativa, com uma previsível rotura de todo o sistema de transportes públicos, mais propriamente em autocarro.

O problema da entrada numa espiral destrutiva do serviço de transporte público não assume a mesma relevância em todos os tipos de território, prevendo-se que o impacto seja maior em territórios deprimidos e com pouca vitalidade económica, como é o caso dos territórios de baixa densidade.

Deste modo, segundo Domingues (2009), os territórios de baixa densidade, sobretudo em zonas rurais, apresentam problemas associados com a dificuldade de acesso a emprego, aos serviços e outro tipo de bens por parte da população residente, particularmente se este acesso for

assegurado por um sistema de transporte público regular em autocarro. Por outro lado, este serviço, em muitas situações, também é responsável por estabelecer as principais ligações efetivas entre populações dispersas em aldeias e as atividades e serviços centralizados em meios urbanos de maior relevância sócio-económica e territorial.

O conjunto de características associadas às áreas de baixa densidade que deve ser tomado em conta na conceção e exploração de um serviço de transportes públicos por autocarro compreende as seguintes aspetos comuns a este tipo de território (Costa, 2009) número de habitantes reduzido, proporção de idosos elevados, nível de rendimentos individual baixo, densidade populacional baixa, e densidade das atividades de serviço e comerciais baixa.

De acordo com Costa (2009) os três pilares fundamentais que sustentam o sistema de transporte de modo que as viagens se efetuem de uma forma mais eficiente, mais segura e com menor impacte ambiental são o ordenamento do território, que condiciona fortemente o padrão de viagens a realizar, o apoio ao transporte público coletivo (de massas) e a restrição ao transporte individual.

Assim, para determinar a viabilidade do serviço de transporte público por autocarro é preciso avaliar a relação entre a procura e a respetiva oferta, através da comparação das necessidades de mobilidade da população com a capacidade oferecida pelo próprio sistema de transportes, tendo em vista a definição de medidas para o melhoramento do sistema de transporte no seu conjunto.

A oferta faz referência ao serviço proporcionado pelo deslocamento rápido das pessoas numa região urbana. Ortuzar & Willumsen (1994) diz que uma característica particular da oferta de transporte é que é um serviço e não um bem. O serviço de transporte tem que ser “consumido” no mesmo momento e local em que é “produzido”, já que de forma contrária perde seu benefício. Portanto, é importante que a oferta de transporte se ajuste continuamente à procura. A caracterização da oferta é baseada numa recolha de informação destinada a averiguar as condições em que a necessidade de transporte das populações é satisfeita. Esta informação não inclui apenas a cobertura temporal e espacial da rede, mas também os operadores de transporte e o material circulante usado (Costa, 2009).

A procura em transportes públicos que interessa conhecer refere-se não apenas ao volume de passageiros atual, mas também à procura potencial (futura), sendo que a informação sobre o comportamento humano em resposta a mudanças no sistema de transporte representa-se por meio de funções de procura. Com elas tenta-se prever o comportamento de um indivíduo o do um grupo de indivíduos, perante uma situação mutável do sistema de transporte.

Para determinar a procura futura do sistema de transporte usam-se os modelos de procura. Segundo Ortuzar & Willumsen (1994) um modelo pode ser definido como uma representação simplificada do mundo real. Concentram-se vários elementos considerados relevantes para a sua análise de acordo com o ponto de vista específico. O modelo tem uma estrutura definida pelo conjunto de expressões básicas com as quais se estimam as principais variáveis do problema analisado de acordo com a forma funcional (as equações principais do modelo podem ser lineares, exponenciais, logarítmicas, ou outras) (Triana Camargo, 2005).

Nos estudos de caracterização da procura é possível constatar que os sistemas digitais e a computação avançaram bastante, e o processo de cálculo já não representa um problema na simulação do sistema de transporte. Na atualidade, as principais limitações de modelação da procura de um sistema de transportes públicos por autocarro são, sobretudo, de carácter humano e técnico, uma vez que requerem profissionais altamente qualificados para o efeito (Ortuzar & Willumsen, 1994).

Por outro lado, atualmente, em Portugal com a entrada em vigor da Lei nº 52/2015, de 9 de junho, que aprovou o Regime Jurídico do Serviço Público do Transporte de Passageiros (RJSPTP) tornou ainda mais premente a necessidade de assegurar o transporte nos territórios de baixa densidade. A introdução da necessidade de se garantir serviços mínimos de transporte público a praticamente todo o território terá um forte impacto na forma como se planeia, organiza e gere um serviço público de transportes por autocarro. Assim, surge a necessidade de se elaborarem modelos de procura de transporte em autocarro para os territórios de baixa densidade, uma vez que a procura é um dos principais fatores no planeamento do sistema de transporte públicos por autocarro, sobretudo quando a informação sobre a mesma é muito escassa.

1.2 Objetivos

A presente dissertação tem como principal objetivo a determinação de um modelo que permita estimar o número de viagens em autocarro em Portugal Continental, em função de um conjunto de variáveis explicativas que influenciam a procura deste meio de transporte para um território de características muito específicas, como são os territórios denominados de baixa densidade populacional.

Para a produção do modelo de procura é necessário cumprir o seguinte conjunto de objetivos secundários:

- levantamento dos principais variáveis utilizadas em modelos de procura de transporte por autocarro presentes na bibliografia, particularmente para territórios considerados de baixa densidade;
- caracterizar a forma como a procura pode ser avaliada no modelo clássico de transportes;
- identificar os principais modelos de geração e viagens em autocarro;
- definir uma metodologia para analisar a geração de viagens em territórios de baixa densidade populacional para o caso português através da realização de estudo de caso.

1.3 Estrutura

Com o objetivo desta dissertação constituir um documento útil para aqueles que se interessem pelo tema que aqui será abordado, esta tese foi organizada e estruturada em 7 capítulos que serão sucintamente descritos de seguida.

Para além do primeiro capítulo, onde é realizada uma introdução e enquadramento geral do tema e os objetivos desta dissertação, no capítulo 2 pretende-se realizar a apresentação da pesquisa bibliográfica subjacente ao estado da arte. Neste será abordado a temática dos territórios de baixa densidade, nomeadamente a definição, a forma como são classificados, a mobilidade e os respetivos problemas. Para além disso, faz-se uma abordagem e análise dos modelos de procura de transporte e suas características, com particular enfoque no modelo de quatro etapas e outros métodos e modelos para a determinação da procura de transporte.

No capítulo 3 apresenta-se a metodologia seguida para desenvolver o modelo de procura que será aplicado no caso de estudo. A metodologia centra-se em vários aspetos, como a definição

das variáveis, a obtenção dos dados, a seleção da amostra, o desenvolvimento de um modelo de regressão linear múltipla até à respetiva validação estatística do modelo que passa pela obrigatória análise de resíduos.

No capítulo 4 é desenvolvido o caso de estudo, sendo apresentadas as análises para os dados e variáveis e estimação do modelo de regressão linear múltipla para determinar o número de viagens em autocarro para municípios portugueses classificados como sendo de baixa densidade.

Já, no capítulo 5 são apresentados os principais resultados obtidos a partir da aplicação do modelo e é realizada uma validação desses resultados através da quantificação do desvio/ erro entre os valores modelados e reais. Estes resultados são analisados com recurso a tabelas e gráficos que pretendem sintetizar e expor os resultados de uma maneira mais simples e objetiva.

O último capítulo tem como principal propósito apresentar de uma forma sintética as conclusões mais relevantes da dissertação em termos gerais e sobretudo dos resultados encontrados com o desenvolvimento do caso de estudo para Portugal: No final do capítulo apresenta-se um conjunto de potenciais trabalhos futuros que decorreram do desenvolvimento deste trabalho.

2 ESTADO DE ARTE

2.1 Territórios de baixa densidade

Conceito

O território é um elemento base da Geografia, Planeamento Territorial e dos Transportes, cujo conceito se encontra envolto de muita discussão, devido às diversas perspetivas e formas de entendimento da sua utilização, coincidindo em alguns elementos e divergindo noutros, nomeadamente em termos de usos normativos, usos disciplinares, perspetivas epistemológicas e procuras aplicadas. Por esta razão, serão apresentadas algumas definições em diferentes momentos, com seus correspondentes autores, no sentido de contextualizar e familiarizar o leitor, com o conceito de território.

Tizon (1995) define o território como sendo o ambiente de vida, de ação, e de pensamento de uma comunidade, associado a processos de construção de identidade ou de acordo com Pecqueur (2001) o território corresponde ao “encontro de atores sociais, num espaço geográfico dado, que procura identificar e resolver um problema comum”. Já, Llanos-Hernández (2010) apresenta o território como “um conceito teórico e metodológico que explica e descreve o desenvolvimento espacial das relações sociais que estabelecem, os seres humanos, nos âmbitos: cultural, social, político ou económico”. Gottmann (2012) refere ainda o território como sendo “uma porção do espaço geográfico que coincide com a extensão espacial da jurisdição de um governo, onde o território é o suporte do corpo político organizado sob uma estrutura de governo. Descreve a arena espacial do sistema político desenvolvido num Estado nacional, ou uma parte deste, que é dotada de certa autonomia”.

Um conceito mais completo e complexo do território é apresentado por Velásquez (2012), no livro “Como entender o território”, onde refere que o território é um tecido complexo de espaços, locais e tempos específicos, circunscritos dinamicamente que, articula uma matriz multidimensional de condições e circunstâncias, de dinâmicas e processos, de sistemas abertos e duráveis, de configuração, representação, reprodução e apropriação das potências, energias e elementos objetivos e subjetivos em complexa relação que, funciona como uma estrutura conformada das perceções, ações e relações dos sujeitos e dos seus atores a curto e longo prazo.

Com base nas definições apresentadas anteriormente, pode-se afirmar que o território não é só uma porção de “terra” delimitada pela sua complexidade biofísica (condições ambientais, biodiversidade), sendo, também, um espaço construído socialmente, ou seja, que apresenta características históricas, económicas, sociais, culturais e políticas. A definição de um conceito claro, preciso e pouco diverso, depende das alterações realizadas às formas de planeamento e ordenamento territorial, tendo em conta que estes são um instrumento para atingir os objetivos da coesão territorial. As orientações políticas da UE, no espaço europeu, para o planeamento e ordenamento do território, tornam-se decisivas na alteração do conceito de território, sobretudo, para um desenvolvimento social e económico da base territorial (Tsukamoto, 2017).

Configuração de um território

Deste modo, o território ultrapassa a mera definição de um espaço da soberania de um estado ou de um eleitorado político-administrativo. A sua configuração consiste na associação dos elementos físicos, biológicos, ecológicos, sociais e simbólicos, que são manifestados, distribuídos, localizados e implantados (espacialmente e temporalmente), devendo contemplar a complexidade como estes elementos são organizados e se relacionam (Velásquez, 2012).

Outro conceito de configuração territorial é fornecido por Santos (citado em (Velásquez, 2012)), a partir de uma perspetiva dialética e complexa, em que diz que a configuração territorial é dada pelo conjunto formado pelos sistemas naturais existentes em um determinado país ou em uma área determinada pelos agregados que o Homem sobrepõem nestes sistemas naturais. Além disso, acrescenta que a configuração territorial não é apenas um espaço, pois sua realidade tem um carácter material, mas o espaço reúne, para além dos aspetos físicos, a vida que anima esses mesmo espaços. Santos termina a sua abordagem conceptual, destacando que a configuração territorial, que também se pode denominar por configuração geográfica com uma existência material própria associada ao forte carácter social dos habitats.

Representação do território

O território pode ser representado de uma forma unidimensional ou multidimensional, parcial ou total, pelo facto de cada tipo de representação ter associados vários atores que analisam e avaliam certas características, poderes e significados para o território, com base em distintas visões, interpretações e interesses.

Algumas representações do território podem vir de matrizes religiosas, cosmológicas, políticas ou económicas, que atuam como mapas mentais para definir, ordenar, sacralizar, historiar, projetar e controlar espaços e pessoas. Tais representações são cobertas e realizadas por sujeitos, ou atores sociais, de caracteres diferentes, que refletem os seus interesses nos mecanismos de apropriação e transformação desse mesmo território, tornando-os um elo que articula relações e vínculos económicos, políticos, sociais e culturais no processo e dinâmica social territorial.

Apropriação do território

O processo de construção e representação do território diz respeito à aprovação que os diversos atores fazem do mesmo. Essa aprovação é, ao mesmo tempo, objetiva e subjetiva, e corresponde ao processo de apropriação do território, sendo mais do que um simples exercício económico e político. Especificamente, a apropriação pode ser imaterial, social, política e física através de práticas espaciais próprias, que têm como ponto de partida o valor que é atribuído ao território pelos diversos grupos de interesse, *e.g.*, instrumental (relacionado com as suas potencialidades económicas e, ou geopolíticas) e cultural (relacionado com aspetos simbólico-expressivos), e como chegada um conjunto de ações que a sociedade coletiva objetiva e põe em prática para poder a melhorá-lo, transformá-lo e enriquecê-lo (Velásquez, 2012).

Estruturação e desenvolvimento do território

O território é construído a partir de processos económicos, sociais, políticos e culturais que o configuram e estruturam, correspondendo a uma configuração espacial organizada assente na utilização ou controlo dos seus recursos ou elementos naturais, mas sobretudo alicerçado em objetivos de administração e exercício de poder (Velásquez, 2012).

Por conseguinte, pode-se concluir que o território está estruturado e organizado, através das relações entre os seres humanos e os demais elementos que o compõem, desde os marcos espaciais (*e.g.*, povoamento, padrões urbanísticos e produção), das questões de mobilidade de pessoas, bens e serviços (quotidiana e circunscrita, imigração e emigração), que moldam os territórios em torno do Homem, de modo que estes sejam valorizados, representados, construídos, apropriados e transformados pelo Homem e para o Homem, numa dinâmica social.

Assim, uma classificação dos territórios terá que ter em conta quais são as características mais relevantes que os distingue e torna únicos. Segundo Tsukamoto (2017), os territórios diferenciam-se pelas suas características físicas e as relações socioeconómicas que se estabelecem sobre o uso e a transformação do solo. O potencial territorial depende das suas características biofísicas, que condicionam os seus recursos endógenos (dimensão material), sujeitos a alterações de natureza antrópica. As relações funcionais, os diferentes modelos de produção, consoante as necessidades e as expectativas geradas em diferentes ciclos económicos sobre os recursos contribuem para a diversidade territorial.

Em Portugal, segundo o sistema integrado de nomenclaturas estatísticas – INE (2017), existem 2.882 freguesias no Continente e 210 nas regiões independentes de Açores e Madeira, sendo que, com base em critérios tales como a densidade populacional e a quantidade de população residente, definidos pela Tipologia de áreas urbanas (2014), são categorizadas, como territórios rurais, 2445 freguesias, segundo TIPAU (2014). Deste modo, em termos numéricos este o território rural domina o panorama do território português, pelo menos em termos administrativos.

Segundo Silva (2010), a tradicional oposição entre campo e a cidade perdeu grande parte do seu significado, sendo que o “rural” não é mais um conceito fechado. O “rural” transcende o sector agropecuário, em favor da multifuncionalidade (que exige uma abordagem a novos desafios, dimensões e oportunidades), estruturando-se num complexo sistema de relações com o urbano através de um conjunto de relações de dependência e complementaridade, onde à provisão de alimentos se junta um conjunto de elos de ligação muito fortes, associados à prestação de um conjunto de bens e serviços, onde emergem como essenciais os recursos naturais, os espaços de descanso e lazer e elementos de preservação da memória, da identidade e da cultura.

A formalização da distinção entre campo e cidade só foi necessária a partir do Século XIX, quando a urbanização, até então um processo lento, se acelerou com a Revolução Industrial. No século seguinte, foi reconhecida como um fenómeno transformador e um dos principais vetores de transformação do século. IBGE (2017) aponta que, a expansão do regime capitalista no último século transformou profundamente o sistema de relações cidade-campo de forma mais profunda que a longa evolução dos séculos anteriores. Além disso, é importante salientar que,

a urbanização é simultaneamente um resultado e uma condição do processo de difusão do capital.

As atividades económicas, as interações sociais e culturais, bem com os impactos ambientais e sociais estão cada vez mais concentrados nas cidades, isto traz uma alteração na configuração espacial dos municípios em termos da definição dos limites geográficos e administrativos referentes a zonas urbanas e zonas rurais. É importante referir que esses limites são, sobretudo, definidos com objetivos fiscais que enquadram os domicílios sem considerar necessariamente as características territoriais e sociais do município (Velásquez, 2012)

Por conseguinte, O debate sobre o rural e o urbano admite muitas interpretações e é composto de muitos elementos analíticos que compõe um mosaico vasto e diverso de interpretações. Contudo, é reconhecida a necessidade de parâmetros que diferenciem os espaços e os avaliem em termos estatísticos e geográficos para a produção de políticas públicas e estudos mais aprofundados. A Divisão Estatística das Nações Unidas (United Nations Statistics Division - UNSD) atesta que as diferenças existentes nas definições de rural e urbano entre as nações ainda não permite uma abordagem aplicável a todos os países e, por esse motivo, essa instituição não recomenda nenhuma tipologia específica (IBGE, 2017).

Os critérios definidos para delimitar as zonas rurais de urbanas são diferentes em alguns países e continentes, dificultando a sua comparação, nomeadamente em relação a alguns fatores, tais como: as distribuições populacionais e o ordenamento territorial, entre outros. A tabela seguinte apresenta a comparação entre alguns critérios definidos em alguns países e órgãos internacionais para a definição de áreas rurais e urbanas.

Tabela 1 – Critérios para a determinação das áreas rurais e urbanas

Fonte: (IBGE, 2017)

País / Organização	Critérios para a determinação das áreas rurais e urbanas					
	Tamanho da população	Densidade demográfica	Oferta de serviços	Participação da agricultura	Divisão administrativa	Aglomeração de habitações
Austrália		X	X			X
Brasil					X	
Chile	X			X		
Colômbia					X	
Estados Unidos	X					
França	X					
México	X					
País de Gales						X
OCDE		X				
União Europeia		X				

Seguidamente, apresentam-se algumas das metodologias de organismos internacionais para a classificação dos territórios, sendo usadas, atualmente, em Portugal. Cada metodologia exposta, não será alvo de análise pormenorizada, uma vez que não é objetivo do projeto, no entanto, serão apresentados os principais critérios para a respetiva classificação.

Definição da Tipologia Regional da Organização para a cooperação e Desenvolvimento Económico - OCDE

A OCDE elabora e avalia políticas para o desenvolvimento, económico, social e político dos seus países-membros. A necessidade de dados e diagnósticos mais precisos sobre as potencialidades e necessidades de investimentos nas áreas rurais motivou a criação de uma classificação rural-urbana dos territórios desses países. A partir dessa classificação, torna-se possível avaliar as condições socioeconómicas de regiões geograficamente passíveis de comparação, neste caso, áreas rurais, intermédias ou urbanas.

A OCDE elegeu a densidade demográfica como critério principal de classificação dos territórios. Ao longo do tempo, foram feitos melhoramentos e adaptações das tipologias, tendo sido adicionados dois critérios: a quantidade e proporção de população que vive em grandes centros urbanos, e a acessibilidade definida pelo tempo de deslocamento entre os centros urbanos e as áreas não urbanas (OCDE, 2011).

A classificação do território em áreas urbanas e rurais elaboradas pela OCDE é denominada Tipologia Nível 3 - TL3, sendo que a classificação final proposta apresenta cinco classes:

- Predominantemente rural remota;
- Predominante rural próxima a uma cidade;
- Intermédia remota;
- Intermédia próxima a uma cidade; e,
- Predominantemente urbana.

Tipologia territorial utilizada em Inglaterra e no País de Gales

A Inglaterra e o País de Gales possuem uma tipologia territorial denominada “Rural-Urban Classification” – RUC. O critério fundamental adotado é morfológico, sendo mais específica a identificação e caracterização dos aglomerados urbanos, medidos pela disposição agregada ou dispersa de habitações, prédios e áreas construídas.

Da classificação resulta uma metodologia e um processamento de dados, bem como, dois tipos de territórios: um urbano com quatro unidades, e as áreas rurais com seis classes. As áreas urbanas classificam-se em: grande conurbação; pequena conurbação; cidade compacta e cidade dispersa. Por outro lado, as áreas rurais são classificadas como: a cidade pequena e subúrbio compacto; urbanização pouco disperso; lugares/habitações isoladas pouco dispersas; cidade pequena e subúrbio dispersos; urbanização dispersa e, por último, lugar/habitações isoladas dispersas.

Tipologia territorial utilizada na União Europeia

A Nomenclatura das Unidades Territoriais para fins Estatísticos – NUTS foi criada pelo EUROSTAT com os Institutos Nacionais de Estatística dos diferentes países da União Europeia (UE) para efeitos de análise estatística dos dados territoriais e populacionais, com base numa divisão coerente e estruturada do território económico comunitário. Trata-se de uma regionalização adotada pela legislação do Parlamento Europeu desde 2003, o qual tem como objetivo amenizar o efeito das mudanças de estruturas territoriais e administrativas dos estados-membros para fins de divulgação estatística. As NUTS aplicam-se apenas no nível regional, criando uma hierarquia de três níveis considerando a organização interna dos países (INE, 2007). O primeiro critério para delimitação dos três níveis da NUTS é a dimensão, conforme se apresenta na tabela seguinte.

Tabela 2. Classificação do território em NUTS segundo a população dos países da União Europeia. Fonte: (Instituto Nacional de Estatística, 2007)

Nível	Mínimo	Máximo
NUTS I	3 milhões	7 milhões
NUTS II	800 mil	3 milhões
NUTS III	150 mil	800 mil

Aplicando esta classificação para Portugal e tendo em conta a modificações feitas no ano de 2013, apresenta-se na Figura 1 o mapa do território português de acordo com a classificação preconizada na Tabela 2, i.e., dividido pelas respetivas três NUTS.

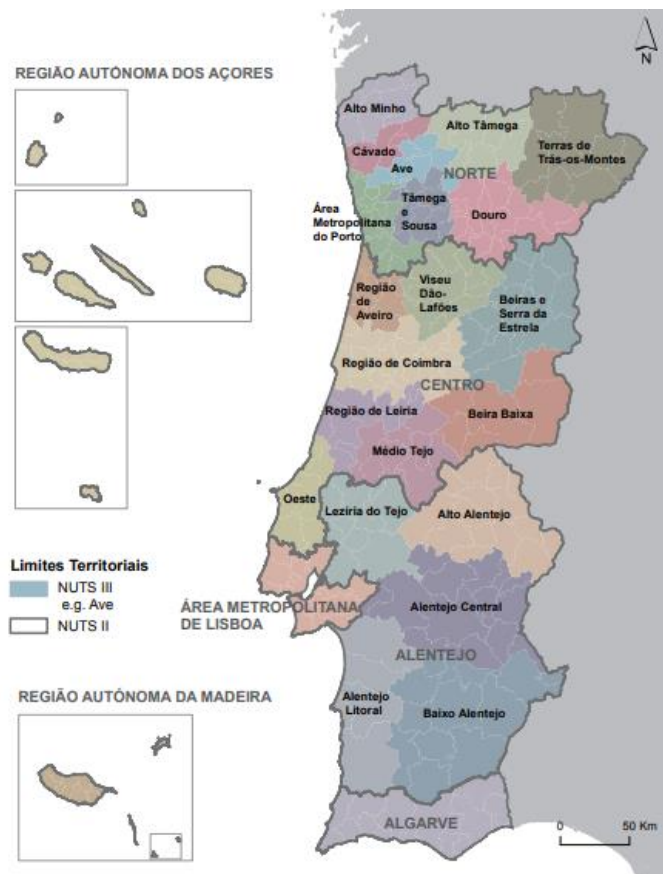


Figura 1 Limites das NUTS 2013

Fonte: (Instituto Nacional de Estatística, 2015)

Tabela 3 - Composição de cada nível de NUTS

Fonte: elaboração própria

NUTS I	NUTS II	NUTS III
Esta classificação contém três subníveis:	Esta classificação contém sete subníveis:	Foi nesta classificação onde se apresento maior mudanças das NUTS 2002 à NUTS 2013, passaram de 30 para 25 unidades territoriais, agora designadas de “unidades administrativas”(Comtemporâneo, 2017)
<ul style="list-style-type: none"> • Continente; • Região Autónoma de Açores; • Região Autónoma de Madeira. 	<ul style="list-style-type: none"> • Norte; • Centro; • Área Metropolitana de Lisboa; • Alentejo; • Algarve; • Região Autónoma de Açores 	

Desde há pelo menos cinco centenas de anos que se tem verificado em Portugal alguma regularidade na distribuição da população pelo seu território, dando origem a territórios com características demográficas e económicas diferenciadas (Santos, 2016). Em Portugal a Área Metropolitana de Lisboa e do Porto têm 2,8 milhões e 1,8 milhões de habitantes, respetivamente, concentrando cerca de 45% do total da população residente no continente. Por outro lado, verifica-se um acentuado desequilíbrio na ocupação do território em Portugal, com uma forte concentração da população residente no litoral e em algumas capitais de distrito do interior (Ministro adjunto-República Portuguesa, 2016). Apesar disso, constata-se que mais de um terço da população vive em zonas rurais, que estão frequentemente associadas à presentas zonas de baixa densidade, que estende ao longo do inferior fronteiroço Norte e o Alentejo interior.

Segundo Ramos & Bento (2012), um Território de Baixa Densidade é “um território marcadamente rural, que apresenta um clima severo, com uma população envelhecida e em perda, de povoamento escasso e disperso, funcionalmente periférico e com uma acessibilidade decrescente aos principais serviços e bens públicos, economicamente marginal e dependente de uma agricultura de subsistência e/ou das prestações sociais públicas”.

O conceito de Território de Baixa Densidade varia de acordo com as orientações das políticas públicas de uma região ou país. Alguns indicadores que permitem definir um território de baixa densidade além da densidade populacional são o PIB per capita, o índice de dependência de

idosos e o índice de envelhecimento da população, a taxa de fecundidade e a variação populacional (Tsukamoto, 2017).

Por outro lado, para Silva (2010), a designação de Territórios de Baixa Densidade advém não apenas de aspetos demográficos, mas também económicos (escassez e fraca diversidade de atividades económicas, elevada taxa de desemprego), urbanos (insuficiente dimensão da maioria dos seus centros urbanos, mesmo os mais importantes), institucionais, (reduzido leque de entidades com atribuições e competências de proximidade) e relacionais (fracas redes de parcerias e deficientes taxas de participação e envolvimento da população).

Já, para Santos (2016) os TBDs caracterizam-se pelo envelhecimento populacional, a escassez de atividades económicas, a baixa qualificação dos recursos humanos, a exiguidade do mercado de trabalho, a falta de dinamismo institucional e dos serviços de apoio às atividades, entre outras.

Em suma, é possível concluir que não existe uma classificação técnico-administrativa única para o conceito de Território de Baixa Densidade, tendo vindo a ser adotados diferentes critérios, centrados ora na densidade populacional, ora no rendimento per capita de cada concelho ou da NUTS 3 a que o município pertence (CIC Portugal 2020, 2015).

Ramos & Bento (2012) apresentam uma metodologia para classificar os TBDs do território de Portugal Continental, tendo sido utilizados vários critérios, onde para além da densidade populacional, é considerado o perfil territorial, demográfico, de povoamento, socioeconómico e de acessibilidade do diversos territórios, uma vez que os mesmos defendem que só ter em conta um critério para classificar territórios é insuficiente e limitado.

Ramos & Bento (2012) apresentam um procedimento para classificar o território em TBDs através da normalização dos indicadores selecionados para os territórios de baixa densidade, logo são agregados dando origem a índices parciais (perfis) que são o intermedio para a determinação do índice global para o território em análise, este índice é calculado de duas formas: a) o índice global resulta da agregação do valor de cada índice parcial obtido, através de média dos respetivos indicadores com igual ponderação, e b) o índice global resulta da média ponderada da densidade populacional (com um peso de 50%) e dos restantes perfis (com ponderação igual a 10% cada, significando, por isso, que os resultados refletem uma visão onde

se assume a mesma importância dos vários perfis para a identificação dos territórios de baixa densidade). O resultado deste exercício apresenta-se na Figura 2, onde a referida classificação é realizada ao nível das freguesias, as quais são classificadas como TBDs quando o valor do índice global apresenta um valor inferior ao da mediana para todas as freguesias de Portugal Continental, representando a mediana então o valor de referência para este índice.

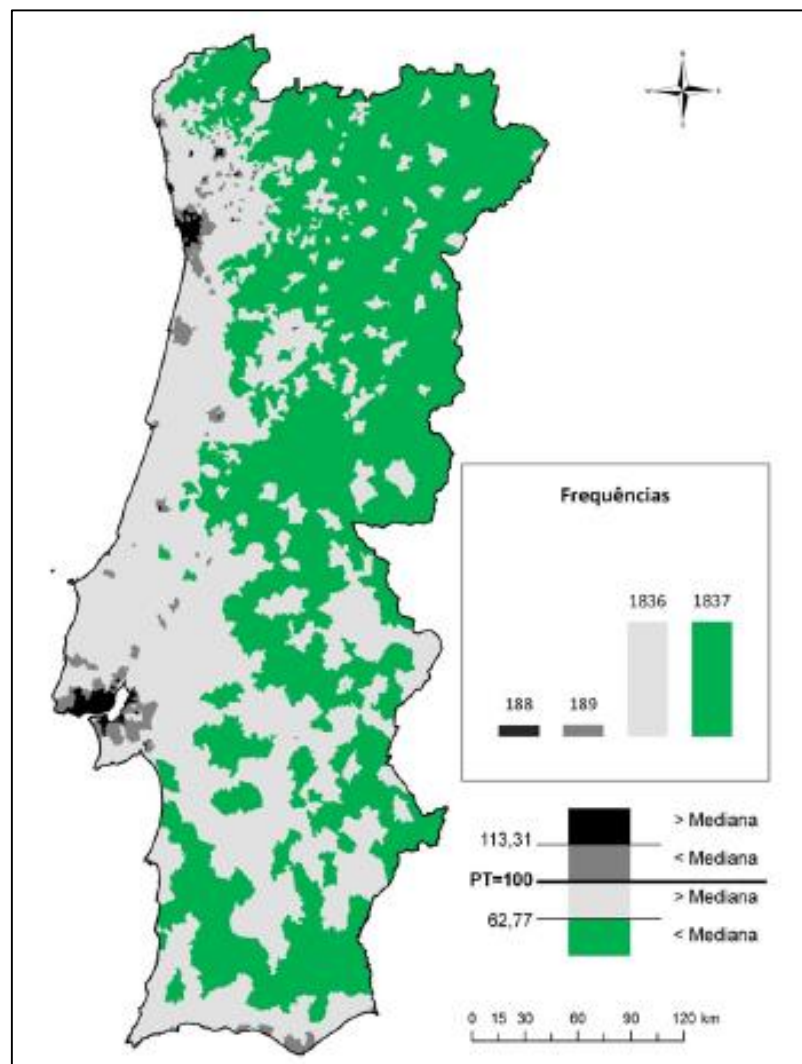


Figura 2 - Classificação por freguesias para os Territórios de Baixa Densidade

Fonte: (Ramos & Bento, 2012)

2.1.1 A mobilidade nos TBD

A problemática da mobilidade e da equidade no acesso da população a bens e serviços, conhece expressão maior no espaço rural e áreas de baixa densidade. Deste modo, a privação destas mais-valias (ou défice de acesso) constitui um elemento catalisador do agravamento das desigualdades e de fenómenos de exclusão social, os quais assumem uma natureza multidimensional. A mobilidade no espaço rural e em áreas de baixa densidade urbana apresenta um conjunto de particularidades que colocam ao planeamento dos transportes, em articulação com o planeamento do território, o desafio de satisfazer as necessidades de deslocação da população de forma eficiente e financeiramente sustentável, visando garantir o acesso a bens, equipamentos e serviços em condições de equidade (Fernandes, *et al.*, 2009).

Os operadores de transportes públicos a operar em áreas rurais de baixa densidade têm mostrado, de um modo geral, alguma dificuldade em desenvolver soluções de serviços de transporte público inovadoras (entenda-se mais flexíveis), adaptadas às necessidades específicas de mobilidade dessas populações. Esta dificuldade de adaptação dos operadores decorre não apenas da dificuldade em implementar soluções alternativas à oferta tradicional de serviços de transporte regular decorrente de uma estrutura e *praxis* organizacional e de planeamento e gestão da oferta por vezes demasiado rígida, como também das dificuldades inerentes à necessidade de investimento que esta adaptação pressupõe, num contexto de baixa rentabilidade financeira dos serviços de transporte público, ou ao próprio enquadramento legal (Fernandes, *et al.*, 2009).

A evolução do sistema de povoamento que consiste num despovoamento das zonas rurais, tem consequências para a sustentabilidade e qualidade de vidas das suas populações, já que pode tornar inviável um acesso “normal” das populações a bens e serviços essenciais. Por outro lado, o despovoamento é o resultado de uma organização funcional do espaço urbano e das dinâmicas de expansão urbana que tem vindo a verificar em diversos territórios, nomeadamente nos países mais desenvolvidos, que poderá conduzir a um aumento nas principais deslocações diárias, especialmente para a aquisição de bens e serviços.

Existe um conjunto de fatores, tales como o nível de emprego, o ganho medio mensal, entre outros, que atuam como vetores indutores da privação ou aumento da dificuldade de acesso a bens e serviços por parte da população residente nas áreas de baixa densidade, nomeadamente a redução, ou extinção, da oferta de serviços de transporte público, que podem conduzir a uma redução da procura e ao encerramento de locais com funções comerciais, uma vez que se pode verificar um declínio da massa crítica nestes sectores de atividade.

Segundo (Fernandes, *et al.*, 2009), é possível identificar um conjunto de características comuns, que possibilitam a determinação de um padrão transversal aos territórios rurais e de baixa densidade, que influenciam a mobilidade das suas populações:

- Tendência de concentração da oferta de equipamentos, serviços e bens nos aglomerados urbanos de “nível superior”;
- Ocorrência de alterações socioeconómicas de cariz estrutural induzidas por transformações nas bases produtivas locais;
- Estruturas etárias caracterizadas por acentuados níveis de envelhecimento da população; baixos níveis de rendimento que se traduzem em níveis de poder de compra igualmente baixos;
- Estruturas de povoamento pautadas pela dispersão ou concentração da população em aglomerados de pequena dimensão mais ou menos espalhados pelo território;
- Redes de transporte escolar com custos de operação relativamente elevados, justificados pela necessidade de assegurar a deslocação diária dos estudantes para estabelecimentos de ensino, por vezes distantes das áreas de residência.

Segundo UITP (2004), viajar em transporte público nunca foi tão difícil. Nos países da União Europeia, durante as três décadas, entre 1970 do ano 2000, a quota modal de automóveis aumentou 4,5% (de 73,8 para 78,3%), ao mesmo tempo que a quota modal relativa ao transporte público diminuiu 8,7% (de 24,6% para 15,9%). Domingues (2009) realça que esta redução se deve a importantes alterações sociais e políticas, correspondendo a mudanças no estilo de vida.

Em suma, a mobilidade dos Territórios de Baixa Densidade apresenta vários problemas, que assentam em importantes fraquezas, como o envelhecimento da sua população residente, economias locais débeis, deficientes acessibilidades, que afetam o seu desenvolvimento e produzem um ciclo vicioso depressivo onde o declínio das dinâmicas socioeconómicas se tornam evidentes.

2.2 Modelos de procura de transporte

O transporte ajuda a promover a saúde económica e a qualidade de vida de um território. O sistema de transportes, não só proporciona a mobilidade de pessoas e bens, como também, influi nos padrões de crescimento e atividade económica ao proporcionar melhores níveis de acessibilidade aos territórios.

O transporte pode ser definido como a deslocação de pessoas e bens, desde um local de origem a outro de destino, utilizando uma variedade de modos e veículos com diversas fontes de energia e de diferentes características, acompanhados de um processo no qual interatuam diversos componentes e que o investigador/ técnico deve poder simplificar através de modelos, os quais devem ter uma suficiente aproximação ao sistema real de acordo com os recursos que possui (Triana Camargo, 2005).

O sistema de transportes faz referência a todos os componentes da oferta e procura do transporte, seja na cidade ou na sua envolvente. Neste sentido, o mesmo inclui a infraestrutura de transporte, assim como, a tecnologia disponível, os utentes e as organizações. É preciso distinguir os elementos que compõem um sistema desta natureza e perceber que, o mesmo, é afetado por uma série de elementos exógenos, como políticas, estruturas financeiras, normatividades, as autoridades de trânsito e transporte, e a comunidade como elemento participativo de primeira ordem.

O sistema de transportes refere-se a um sistema mobilidade multimodal e serviços que são fornecidos numa determinada área. O sistema pode incluir qualquer um, ou todos, os serviços relativos a rodovias, ferrovias, ciclovias, e caminhos pedestres. Estes podem partilhar o mesmo direito de passagem, ou serem parcial ou totalmente separados (Webster & Bly, 1981).

O primeiro conceito que tem que ficar claro antes de apresentar a procura de transporte é o planeamento do mesmo, por conseguinte, planear em termos gerais significa antecipar a direção da ação que tem que seguir-se para alcançar uma situação desejada. Assim, o planeamento converte-se na capacidade e definição racional das prioridades, objetivos, metas, estratégias e dos recursos necessários.

O planeamento de transportes encarrega-se pela determinação da procura, dos polos geradores de viagem e dos modos utilizados pelos utentes, para fornecer a alternativa de prestação do serviço de transporte. Além disso, o planeamento aborda diversos problemas e aplica procedimentos que mudam de acordo com a escala e o objeto, i.e., com o nível em que é preciso desenvolver a solução e tipo da mesma. Assim, um dos problemas que vulgarmente se apresentam no processo de planeamento de transportes é o de prever (“adivinhar”) o futuro do funcionamento e evolução do próprio sistema. Este problema geralmente resolve-se através da utilização e desenvolvimento de cenários, os quais representam várias alternativas de possibilidades de ocorrência no futuro, devendo integrar o nível de incerteza associado à previsão das variáveis que intervêm na formulação e resolução do problema.

Tendo em conta a importância que tem o planeamento de transportes no serviço proposto aos utentes, é necessário conhecer quais são os principais objetivos que se pretende alcançar para manter um funcionamento eficaz e eficiente do transporte público. Segundo Triana Camargo (2005), os principais objetivos do planeamento do transporte são:

- Utilizar de uma forma otimizada a infraestrutura dos diferentes modos de transportes disponíveis para abranger toda a procura de transporte de uma região;
- Prever as mudanças que se apresentam na procura de transportes como consequência das modificações no sistema de transporte;
- Avaliar cenários futuros do sistema de transportes, que estejam relacionados ao pronóstico das principais variáveis socioeconómicas utilizadas para determinar a procura de transporte;
- Analisar a relação entre o sistema de transporte e o sistema socioeconómico de uma região.

Deste modo, o planeamento do transporte pressupõe uma previsão dos fluxos no sistema de transportes para diferentes situações, sendo portanto pertinente conhecer o comportamento humano para poder prever a procura de transporte. Este comportamento armazena como informação e completo respeito do que uma pessoa planeia e como ele a “computa”. Ter acesso a este tipo de informação detalhada pode-se considerar como afetação da privacidade de dados. Além disso, a recolha da informação pode aumentar, de uma maneira considerável, a complexidade do trabalho. Uma forma de recolher alguma informação do comportamento das utentes ou da rotina das pessoas, é através da análise dos dados censitários, inquéritos (preferências reveladas ou preferências declaradas), entre outros métodos de recolha de informação pessoal, com o intuito dos planeadores de transporte obterem informação do comportamento atual da população em relação ao transporte. Balmer (2007) afirma, com

certeza, que se tiverem este tipo de informação, os planeadores precisam de modelos, programas e ferramentas para responder a uma grande variedade de perguntas, tais como: o número de passageiros numa linha de transportes públicos, a facilidade de modificação nas frequências do transporte entre outras perguntas que ajudam a uma melhorar a tomada de decisão.

Descrever cada pessoa (utente do sistema) individualmente é de facto, o alvo desejado no planeamento de transportes. A razão para isso é bastante óbvia: se alguém pode gerar e monitorar de maneira sintética cada pessoa numa determinada área de interesse durante um certo período de tempo, torna-se possível extrair um vasto conjunto de informação agregada sobre esse cenário.

As decisões das pessoas sobre as viagens que devem realizar como parte das suas atividades cotidianas levam diretamente a uma “procura” ou um “desejo”. A informação sobre o comportamento humano em resposta a mudanças no sistema de transporte representa-se através de funções de procura. Com elas tenta-se prever o comportamento de um individuo, ou de um grupo de indivíduos, perante alterações no funcionamento do sistema (Triana Camargo, 2005).

Segundo Ortúzar (2015), em seu livro “Os modelos de procura de transporte”, as características mais relevantes da procura de transporte são :

- A procura de transporte é derivada, ou seja, as viagens são geradas pela necessidade de realizar atividades (por exemplo, trabalho, compras, recreação) no destino;
- A procura de transporte está localizada no espaço, a espacialidade da procura pode produzir problemas de descoordenação que afeta o equilíbrio oferta-procura (autocarros quase vazios pelas ruas e possíveis passageiros a espera em outras áreas);
- A procura de transporte é dinâmica e há poucas horas disponíveis para realizar as atividades; ao nível da oferta, o facto do transporte ser um serviço previamente definido tem como consequência a impossibilidade de fazer reservas (stock) para que sejam utilizadas em períodos de maior procura. Por conseguinte, se o serviço de transporte não se consome quando é produzido, perde-se. Isso tende a gerar problemas nos períodos de ponta em que existe uma maior procura e desequilíbrios em relação aos períodos fora de ponta com requisitos mais baixos.

No passado, o processo de planeamento do transporte visava garantir a provisão da oferta para satisfazer a procura. Com este enfoque contruíram-se os grandes projetos rodoviários dentro

das cidades e fora delas (*e.g.*, as autoestradas), os sistemas de transporte coletivo de grandes massas (*e.g.*, autocarros, *Bus Rapid Transit* - BRT, metro, comboio). Mas, devido às crises económicas, à introdução dos computadores pessoais, à falta de recursos para realizar grandes obras, este enfoque têm vindo a mudar, no sentido de uma otimização, a nível económico e de serviço, do sistema de transportes atual, bem como, operar da melhor maneira possível as diferentes infraestruturas disponíveis, nomeadamente a rodoviária. Juntamente com esta mudança enfatiza-se o planeamento a curto prazo, assim como, as limitações do planeamento a longo prazo, permitindo o desenvolvimento de métodos de atualização de informação por procedimentos mais simples, em contraposição a uma nova e completa recompilação de todos os dados, quer no tempo, quer no espaço (Cal Mayor & Asociados, 2005).

Tal como referido anteriormente, a oferta de um sistema de transportes representa um dos seus componentes principais. Em primeiro lugar a oferta, que em termos gerais representa toda a infraestrutura onde o funcionamento do sistema assenta, quer seja coletivo ou individual, cuja utilização se representa pela procura (massa crítica ou utentes do serviço).

Oferta do sistema de transporte

A oferta faz referência à infraestrutura que proporciona o serviço para o movimento, que se quer rápido e eficiente das pessoas num certo território, urbano ou rural. A oferta de transporte projeta-se para atender determinado nível de procura, portanto, em horas de maior confluência de utentes, apresentam-se os níveis mais baixos de serviços. Nas restantes horas do dia consideram-se as menos rentáveis pelos concessionários do serviço de transporte público, mas, para os utentes, podem ser as mais cómodas (Cal Mayor & Asociados, 2005).

A oferta, pode-se distinguir por tangível e intangível, sendo que a tangível se pode dividir em física através dos seus elementos imóveis (*e.g.*, estradas, estacionamento, estações, portos), também por elementos móveis (*e.g.*, veículos). Já, oferta de carácter intangível se prende mais com as questões do serviço oferecido, como as regras de operação dos modos, os horários, as rotas ou percursos.

Procura do sistema de transporte

A procura diz respeito à forma como o sistema é utilizado pelos diversos utentes, que se classificam segundo as diversas categorias de transporte, sobretudo entre mercadorias e passageiros. As abordagens usadas para estimar a procura de transporte podem classificar-se em desagregadas e agregadas, que analisam a procura em termos individuais e grupos de indivíduos, respetivamente. O comportamento da procura desagregada refere-se a consumidores simples (e.g., indivíduos, residências, empresas), consideradas como uma unidade de decisão. Já, a procura agregada encontra-se relacionada com grupos de consumidores que respondem, ou são diretamente afetados, a mudanças nas condições futuras.

A avaliação da procura potencial de transporte deve considerar aspetos socioeconómicos, de uso do solo, origem e destino das viagens e respetiva distribuição, volumes de tráfego, entre outros elementos.

Na Figura 3 apresenta-se o procedimento para avaliar a procura, ao nível do processo de planeamento de transporte. Assim, este processo deve compreender duas etapas básicas: i) estabelecer a procura para um nível e qualidade de serviço, ou seja, a sua abordagem tem em conta o estudo das necessidades dos utentes; ii) elaborar um plano de ação com a capacidade de satisfazer a procura, ou seja, um estudo dos modos.

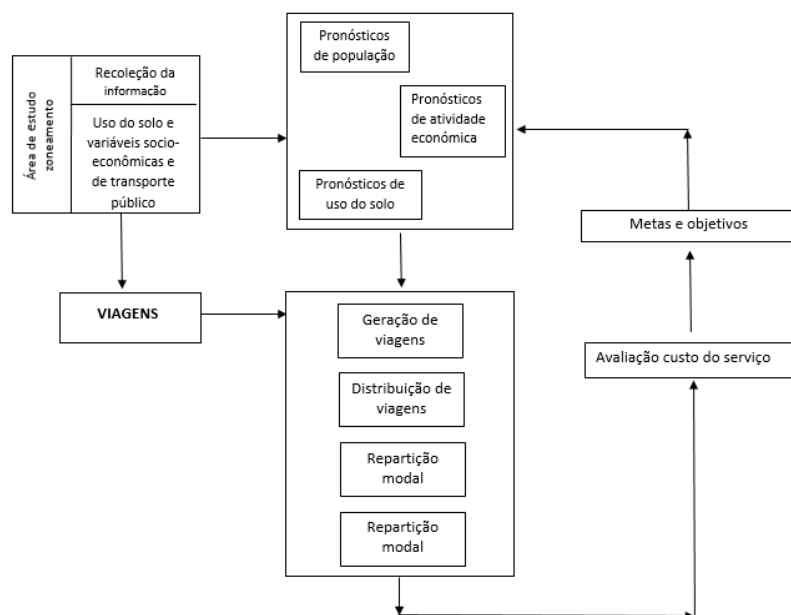


Figura 3. Processo de análise da procura em planeação de transporte. Fonte: (Cal Mayor & Associados, 2005).

Por outro lado, importa referir que, a análise da procura de transporte deve considerar a dimensão espacial, *i.e.*, o tamanho da região, que poderá variar entre áreas pequenas, médias e metropolitanas, uma vez que a disponibilidade e a respetiva obtenção de informação para os diferentes territórios é muito distinta. Assim, para Triana Camargo (2005) algumas das características gerais que permitem classificar os diferentes tipos de territórios para determinar a procura dos mesmos são as seguintes: ganho mensal, propriedade de veículo (conhecido como taxa de motorização), tamanho do domicílio, ocupação das famílias, localização geográfica, propósito da deslocação, população residente, uso do solo, presença de zonas industriais ou de plataformas logísticas (dado estarem associadas a importantes locais de atração ou produção de viagens e, por isso, requerem especial atenção).

2.2.1 Modelos de procura de transporte

Em princípio é importante definir o conceito de modelo, o qual corresponde a uma representação da realidade, uma abstração que se utiliza para alcançar maior certeza conceptual o comportamento dessa realidade, reduzindo a variedade e complexidade a níveis que permitam perceber e especificar de maneira adequada as análises a realizar (Ortúzar, 2015). Já para Triana Camargo (2005), um modelo é definido como um conjunto de expressões matemáticas ou de procedimentos matemáticos e físicos que relacionam variáveis que permitam explicar e solucionar um problema, em qualquer disciplina. Os modelos aplicam-se em diversas ocasiões para justificar de um a forma racional a escolha da melhor solução para a resolução de problemas.

Os processos de tomada de decisão têm sido uns dos grandes promotores do desenvolvimento de modelos em todas as áreas do conhecimento, geralmente associados à complexidade dos acontecimentos em estudo, os quais podem ser de índole social, natural, físicos, químicos, etc. Além disso, a quantidade de informação que tem de ser processada para dar solução aos problemas através de modelos pode ser muito extensa e vasta. Porém importa destacar que o fundamento principal de um modelo é, como tinha dito anteriormente, simplificar os processos, ou seja, deve-se fazer uma análise somente com aquelas variáveis que realmente influenciam, em maior proporção, o problema em análise, especialmente em previsões para o futuro.

Os modelos podem ser divididos em dois grupos, nomeadamente os físicos e os abstratos. Assim, para ter uma melhor perspetiva do que cada de estes grupos de modelos, apresenta-se de seguida no que consistem e alguns exemplos que auxiliem a sua compreensão.

Modelos físicos

Estos modelos são adequados para analisar certos problemas tangíveis e materiais, através da criação de modelos à escala reduzida, que permitem a mediação e observação do comportamento modelado de uma forma próxima da realidade. Contudo, têm como desvantagens a associadas ao aumento da escala, que não permite avaliar várias singularidades associadas às dimensões, nomeadamente as questões de pormenor. Alguns dos exemplos mais comuns para entender este tipo de modelos são as maquetas arquitetónicas, os tuneis de vento, os canais ou barragens na engenharia hidráulica, entre outros. Na Figura 4, apresentam-se alguns dos exemplos anteriormente citados.



Figura 4 - Exemplos de modelos físicos. Fonte:

(1. <http://www.alarcomaquetas.com/galeria-maquetas.php>

2. (<https://pt.dreamstime.com/imagens-de-stock-modelo-da-represa-image25397594>)

Modelos abstratos

Nestes modelos, a situação é representada através de símbolos e não por mecanismos físicos, tornando-se muitos mais uteis para o planeador, uma vez que, mudam a sua atenção para aspetos mais abrangentes do projeto, que contemplam a representação de relações funcionais e os processos básicos que, de alguma maneira, estão ligados aos sistemas analisados. A sua utilização é muito mais habitual do que os modelos físicos, porque permitem investigar o comportamento do sistema e avaliar as reações a determinados estímulos. Exemplos disso são os modelos utilizados para avaliar determinadas ações de certos agentes e respetivas

consequências no mundo que os rodeia, permitindo que se tomem decisões que se julgam mais adequadas, em virtude do necessário suporte técnico-científico que os modelos lhes conferem.

Dentro dos modelos abstratos, há um grupo de modelos, chamados matemáticos (Figura 5), que assumem particular relevância para o desenvolvimento desta dissertação. Deste modo, este tipo de modelos será analisado de uma forma mais detalhada, especialmente em termos das suas características e composição.

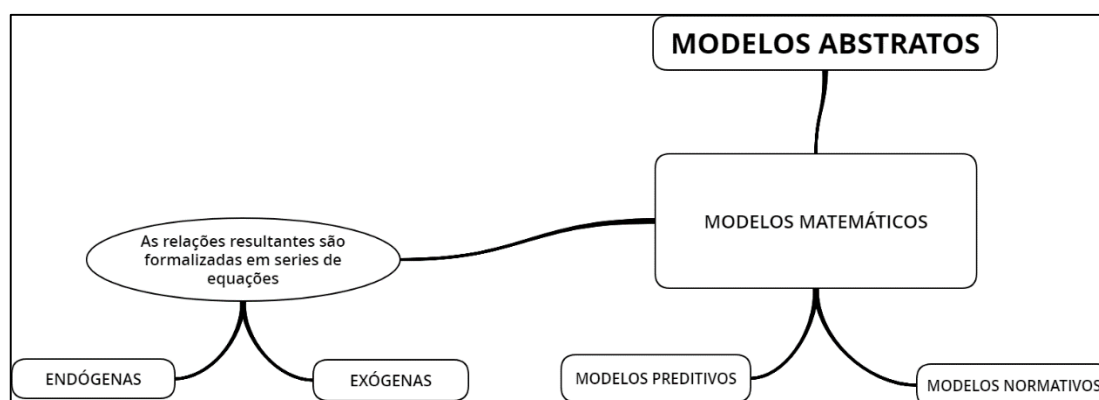


Figura 5. Classificação e composição dos modelos abstratos. Fonte: Elaboração própria

Modelos matemáticos

Os modelos são compostos por uma série de equações que expressam relações, predições, variáveis e parâmetros. Como se observa na Figura 5, há dois tipos de variáveis: em primeiro lugar, estão as variáveis chamadas exógenas - as quais são independentes e explicativas. Os seus valores numéricos determinam-se fora do modelo e, além disso, possuem dois tipos de “sub – variáveis” dentro da sua composição - este é o caso das variáveis políticas (são variáveis baixo controlo do moderador) e as variáveis adicionais que, como seu nome indica, não são do controlo do modelar e devem ser tidas em conta para a situação em que é estudado o modelo. Por outro lado, estão as variáveis endógenas ou dependentes. O valor numérico destas variáveis resulta da operação do modelo. É importante sublinhar que, o problema básico da modelação, nasce na estimação das variáveis endógenas, uma vez que, dependem da qualidade e precisão das variáveis exógenas e estas, por sua vez, dependem das suposições explícitas ou implícitas que foram realizadas para sua especificação. Tem de se ter em conta as especificações das variáveis, ou seja, indicar as variáveis que podem ser utilizadas e a forma em como se

consideram no modelo, do mesmo modo, pode-se determinar o intervalo de restrição de valores das mesmas (Ortúzar, 2015).

Uma classificação para os modelos matemáticos, que diz respeito à disponibilidade da informação é dada por Pereira (2014), onde considera os modelos de “caixa branca” (os quais têm acesso a toda a informação disponível) e os de “caixa preta” (onde não existe informação disponível, tendo o planejador que efetuar todo o trabalho de pesquisa e obtenção de dados). Também os modelos podem-se classificar em modelos determinísticos ou estocásticos. Nos deterministas é conhecida toda a informação, sendo possível obter uma solução para o problema em análise. Para os modelos estocásticos a informação necessária não é de todo conhecida, logo existe um potencial aumento da incerteza, onde as variáveis se caracterizam por apresentarem um comportamento aleatório.

Segundo Ortúzar (2015), dentro dos modelos matemáticos, distinguem-se dois tipos de modelos: preditivos e os normativos. Os primeiros, os modelos preditivo, procuram determinar uma causalidade entre as variáveis, com o intuito de formar relações funcionais que tenham constância no tempo. Os segundos modelos, os modelos normativos, que se constroem com o propósito de produzir estimações sobre o comportamento do sistema em relação aos objetivos definidos, pelo que devem incorporar metas e restrições.

Os modelos matemáticos são um elemento fundamental para se obter uma solução para os problemas relacionados com a tomada de decisões, que incluem fatores intangíveis e que, de alguma maneira, não são facilmente quantificáveis. Segundo Triana Camargo (2005), o principal fator intangível, e que é difícil de quantificar, é a presença dos fatores humanos na maior parte das tomadas de decisões, que contribui de forma singular para um aumento do problema, não sendo possível de modelar de uma forma 100% fidedigna do ponto de vista matemático

Segundo Teixeira (2002), um modelo matemático tem que cumprir um conjunto mínimo de objetivos, que são:

- Obter respostas sobre o que acontecerá num mundo físico;
- Influir na experimentação ou observação posteriores;
- Promover o progresso e a compreensão conceptual;
- Auxiliar ao estabelecimento de um conjunto de normas justificadas da situação física.

Até ao momento, nesta dissertação, já se abordou o conceito de modelo, das diferentes classificações de modelos que são considerados importantes para o desenvolvimento deste trabalho, bem como, dos objetivos e da constituição dos modelos, numa forma mais geral. Porém, a construção de um modelo pressupõe uma avaliação dos diferentes objetivos e campos de aplicação, dos erros que devem ser tidos em conta no momento da sua estruturação, que representa uma das partes mais complexas, em termos conceituais, da elaboração de um modelo a ter em conta no desenvolvimento desta dissertação, que se aborda de seguida.

Construção de um modelo

Para estruturar ou desenhar um modelo considera-se um conjunto de etapas que se têm que desenvolver com o intuito de prover o planeador com um instrumento de análises que lhe auxiliem o processo de tomada de decisão, de uma forma mais assertivas. Não existe total clareza sobre a forma como nasce ou se gera um modelo. Seguidamente, apresenta-se uma serie de tópicos que se tem de considerar no desenvolvimento de um modelo, com base em Triana Camargo (2005), Ortúzar (2015) e Flores Reyna (2017).

1º - Definição do problema: nesta etapa deve-se identificar, compreender e descrever o problema. Para além disso, deve-se estabelecer sua abrangência. Nesta fase é necessário estabelecer e definir os objetivos do projeto, os quais são, muitas vezes, difíceis de quantificar.

2º - Desenvolvimento e construção do modelo: consiste em expressar o problema de uma forma matemática. Para expor o problema, é preciso lembrar que existem variáveis cujos valores ainda não se conhecem. É comum tratar este tipo de variáveis como de “controle”, já que se possuem controlo sobre os seus valores.

Importa salientar que, a obtenção dos dados pode, algumas vezes, demorar mais tempo que o desenvolvimento do modelo, mas, isto assegura, de alguma maneira, uma qualidade superior da solução. Assim, a qualidade dos resultados depende diretamente da confiabilidade e precisão dos dados de entrada.

3º - Solução do modelo: nesta etapa obtêm-se os valores numéricos das variáveis. A forma como se chega a esses valores, depende do tipo de modelo resultante e, deste modo, é possível determinar a solução e a técnica mais correta para obter a respetiva solução. Esta etapa permite

fazer a análise da sensibilidade da procura, a qual consiste em obter informação adicional sobre a solução encontrada, quando o modelo está exposto a variações nos parâmetros.

4º - Calibração do modelo: o conceito de calibração consiste no processo de atribuir valores empíricos às relações entre variáveis e valores constantes do modelo, isto é realizado com o objetivo de desenvolver o modelo para a realidade para a qual foi em princípio criado. A calibração inclui a realização de provas de ajuste de alguns parâmetros de modo que os dados reais sejam semelhantes aos resultados obtidos pelos modelos. As provas de ajuste são feitas por meio de procedimentos estatísticos convencionais.

5º - Validação do modelo: após se fazer a calibração, é realizada a validação do modelo. É importante ter cuidado com os valores usados para fazer isto, já que não podem ser os mesmos usados na calibração. A validação tem como objetivo verificar se o modelo cumpre com a função para o qual foi desenhado, ou seja, se o modelo explica uma proporção razoável do comportamento do sistema em estudo, isto é, se os seus resultados são aceitáveis, podendo-se então afirmar que o modelo cumpre com os objetivos para o qual foi desenvolvido. Uma das razões para fazer a validação reside no facto de um modelo matemático não poder captar todas as limitações de um problema real, ou que se tenha omitido ou simplificado outros aspetos do problema em estudo.

6º - Modificação do modelo: se durante a validação ou calibração se verificar que a solução não é viável, devem-se identificar as limitações omitidas durante a formulação do problema original para determinar se algumas limitações originais são incorretas e, portanto, se é necessário produzir alguma modificação. Em caso afirmativo, tem que voltar a etapa de formulação do problema e fazer cuidadosamente as modificações necessárias para se obterem valores com um nível maior de precisão em relação à realidade do problema.

Zonamento

As Zonas de Análises de Transporte (ZAT), utilizam-se para juntar os domicílios (casas) individuais, gabinetes e outros lugares de trabalho ou serviços, em grupos mais fáceis de trabalhar do ponto de vista da modelação. As duas dimensões chave num zonamento são: o tamanho e o número de zonas. É importante salientar que, o tamanho das zonas corresponde a diferentes níveis de detalhe, tendo em conta os resultados esperados do processo de planeamento do sistema de transporte. Há uma relação entre estas duas dimensões que

corresponde a um maior número de zonas numa área determinada, menor será o seu tamanho. Por outras palavras, pode-se definir que com um menor tamanho dos ZAT, se poderá chegar a uma maior precisão nas análises, mas, isto aumenta o custo da coleta de informação e o desenvolvimento do processo de planeamento (Kriger et al., 2016); (Ortuzar & Willumsen, 1994); (Triana Camargo, 2005).

O número e tamanho das zonas em que se divide uma área depende basicamente de dois fatores, (Ortúzar, 2015), que são:

- **Caráter do estudo:** se o estudo é estratégico, vão ser eleitas menos zonas e maiores; se é detalhado, o número de zonas vão ser maior mas de menor tamanho;
- **Recursos disponíveis:** um maior número de pessoas implica mais dados e à partida uma maior precisão, mas, deste todo ponto de vista, é mais custoso.

Uma forma de definir as ZAT de uma maneira mais geral foi apresentada por Kriger et al. (2016) e corresponde áreas homogéneas em termos de uso de solo, *e.g.*, os bairros residenciais, CBD – “Central Business Districts” e áreas industriais. As ZAT, também se podem caracterizar e distinguir pelos principais seu polos geradores de tráfego, *e.g.*, as universidades, hospitais, centros comerciais e aeroportos. Uma última característica, mas não menos importante, são os limites geográficos naturais (rios, montanhas) ou feitos pelo homem (estradas, ferrovias, entre outros).

Os critérios de zonamento, que segundo Ortúzar (2015) tinham sucesso ao longo do tempo por demonstrar que são uteis e importantes, são os seguintes:

- i. O zonamento deve ser compatível com outras divisões administrativas, em especial com zonas definidas pelos CENSOS, sendo este um critério fundamental;
- ii. As zonas devem ter um tamanho ótimo tal que a suposição de que todas suas atividades se concentram no “centroide” da mesma e que isto não ocasione um grande erro, além disso, a existência de muitas zonas pequenas tem a vantagem de poder agregar no futuro mais zonas, dependendo dos projetos.
- iii. As zonas devem ser, se possível, homogéneas em relação ao uso do solo e/ou a composição da população; quando não existe homogeneidade, não se deve agrupar as zonas definidas nos CENSOS;
- iv. Os limites das zonas devem ser compatíveis com “cordões” internos e “linhas telas”, e com zonamentos de estudos anteriores; também é importante que os limites não sejam

- definidos em vias importantes da rede viária primária, já que isso dificultará, consideravelmente, a posterior atribuição de viagens às zonas;
- v. A forma das zonas deve permitir uma fácil determinação de seu centroide; como este critério tem menos importância que os anteriores, resulta alguma zona com forma irregular, o seu centroide localizar-se-á no centro obtido pelas “densidades populacionais” e não pela sua geometria - centro geográfico;
 - vi. Não é necessário dividir a área de estudo em zonas de igual tamanho.

Ortúzar (2015) refere que todos os critérios apresentados no parágrafo anterior têm que ser alteradas sempre que as características topográficas, do uso do solo e da população da área de estudo o exijam.

Erros de modelação

A maioria dos procedimentos estatísticos utilizados na estimação de modelos assumem, implícita ou explicitamente, que os dados e a forma funcional do modelo são conhecidos. Na prática não se cumprem essas condições, mesmo assim, se forem cumpridas, existirão erros nas previsões dos modelos devido à imprecisão dos valores estimados para as variáveis exógenas.

Ter em conta os diferentes tipos de erros, a forma como se podem apresentar ao longo do processo de modelação e como se podem reduzir, é um tema importante para esta dissertação, onde se faz uma pequena descrição dos erros presentes desde a amostra até à validação, ou mesmo até ao funcionamento do modelo. Desta forma, os principais erros na modelação segundo (Alonsi, 1968; Tardiff, 1979; Horowitz, 1981; Transporte & Triana, 2005; Ortúzar, 2015; Flores Reyna, 2017) são os seguintes:

- *Erros de medição, codificação e digitação dos dados:* Estes erros aumentam com o refinamento ou sofisticação das variáveis que se devem medir. Uma forma de reduzir este tipo de erro é através de um maior investimento na supervisão e treino dos operadores responsáveis pela recolha de dados, além da sua respetiva verificação.
- *Erros da amostra:* este tipo de erro tem sua origem na consideração de amostras em vez da população completa. Para reduzir o erro é preciso aumentar amostra, tendo em conta que este incorrerá num aumento de custos no projeto;
- *Erros de especificação:* apresentam-se na omissão de uma variável relevante, numa forma funcional errada, na presença de um hábito ou inércia no comportamento. Estes erros têm subjacente que nenhum modelo representará a realidade de um modo

completamente preciso. Um modelo com maior sofisticação pode reduzir este tipo de erros através de um maior investimento na recolha e processamento dos dados;

- *Erros de calibração, validação e previsão:* os primeiros erros têm como origem o uso de técnicas de calibração parcialmente imprecisas, os segundos erros resultam de falhas na previsão das variáveis independentes do modelo;
- *Erros de transferência:* estes erros ocorrem quando se utiliza um modelo desenvolvido para uma área específica e, ou época, diferente da utilizada na modelação, sem realizar os respetivos ajustes.

Analisando as etapas básicas de construção de um modelo e os diferentes tipos de erros que se podem observar nesse processo, levanta-se a dúvida se é melhor desenhar um modelo complexo que requeira mais precisão nos dados recolhidos, com maiores amostras e mais custos, i.e., um modelo cuja estrutura matemática possui mais variáveis, um maior número de equações e operações mais sensíveis ao erro., ou se pelo contrário, ter um modelo mais simples que represente o comportamento de uma forma satisfatória sem aumentar o investimento no projeto, i.e., cujos dados sejam mais fáceis de obter, mas ao mesmo tempo pondo em causa a especificação do modelo, por outras palavras a qualidade da abstração do problema da vida real é baixa. Para resolver esta dúvida Ortúzar (2015) faz uma análise e uma comparação destas duas possibilidades, tendo concluído que o equilíbrio entre a complexidade e a simplicidade, resulta como se pode observar na Figura 6, que o *erro geral (E)* mais baixo não corresponde ao ponto de maior complexidade do erro de medição, mas a um valor intermédio. Logo, o modelo mais otimizado não é o mais complexo, mas aquele que com menores custos soluciona o problema em estudo.

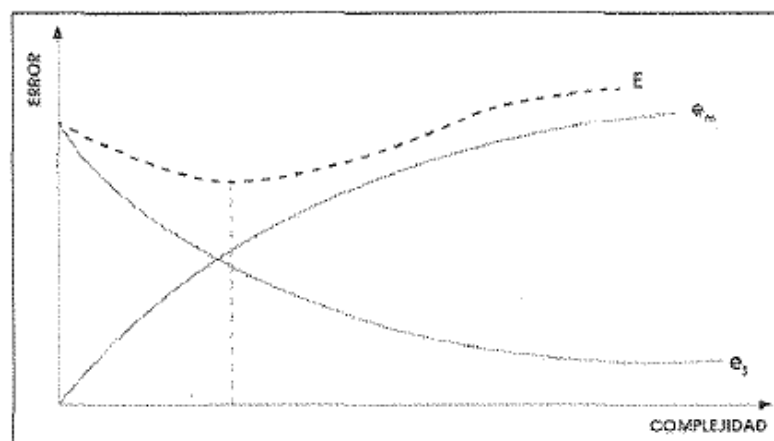


Figura 6. Variação do erro em função da dificuldade do modelo. Fonte:(Ortúzar, 2015)

2.2.2 Modelo clássico de quatro etapas

A primeira referência sobre uma modelação de um sistema de transportes surgiu em meados dos anos setenta, tendo sido denominada como o “modelo de clássico de transportes” ou “método das quatro etapas”. Neste tipo de modelos a área em estudo é dividida num conjunto de zonas espacialmente contíguas em termos de geração e de atração de viagens. A procura é estimada utilizando um processo sequencial de quatro etapas, que compreendia i) a geração de viagens, ii) a distribuição de viagens, iii) a repartição modal, concluindo-se com a iv) afetação das viagens aos diferentes itinerários. Aparentemente esta metodologia não foi alterada desde sua origem, mas é algo ilusório, já que têm sido feitos importantes avanços na especificação dos seus diversos submodelos, e ainda mais importante prende-se com a definição clara da forma como estes se devem relacionar e interatuar para garantir uma solução única para o problema complexo de garantir um certo equilíbrio entre oferta e procura implícito na estrutura do modelo de quatro etapas (Teixeira, 2002).

Um problema que se apontava ao modelo clássico era a falta de interação entre o sistema de transportes com as questões de ordenamento do território, dado que apresentava características “fechadas”, sendo considerado como unidirecional uma vez que não permitia qualquer “feedback” sobre os custos das viagens no processo de geração de viagens. Atualmente, o enfoque do modelo clássico começa por considerar uma rede multimodal de transportes, um zonamento da área de estudo, e a recolha e codificação dos dados tanto para a calibração e validação dos modelo, como para o seu uso para a realização de previsões (Ortúzar, 2015).

O modelo de 4-passos foi pioneira no planeamento de transportes e é ainda a mais usada para esse fim. Assim, torna-se relevante explicar cada uma destas etapas e definir a que etapas corresponde, ou em que se pode enquadrar, o modelo a desenvolver nesta dissertação para estimar o número de viagens em autocarro para Territórios de Baixa Densidade.

Geração de viagens

O objetivo desta etapa é estimar o número de viagens geradas (desde uma origem) e atraídos (em direção a um destino) por diferentes zonas dentro de uma área de estudo, isto é, à origem faz-se corresponder o número total de viagens que aí se iniciam, e a cada destino associa-se o número total de viagens que aí terminam. Apesar de existirem diferentes tipos de viagens, tendo em conta o seu propósito, são habitualmente definidas as viagens casa-trabalho, casa-escola, e

de uma certa origem para outros fins, que incluem as viagens para consultas médicas, compras, lazer, entre outras mas cujo destino ou origem não inclui uma casa. Uma condicionante para a determinação da produção, ou atração, das viagens nas áreas de estudo é dada pelas características socioeconómicas e de uso do solo (Kriger et al., 2016; Fischer, 1992).

Uma forma muito habitual para representar um modelo de geração de viagens é através de uma equação de regressão linear múltipla em que as variáveis explicativas poderão ser, entre outras, a população, o emprego por sector de atividade e a taxa de motorização.

Distribuição modal

Nesta etapa procede-se à distribuição do número total de viagens geradas em cada origem pelos vários destinos. A distribuição é estruturada em duas fases. A primeira refere-se à definição dos usos e zonas do território de estudo, i.e., as viagens são distribuídas de acordo com a magnitude e intensidade das atividades e o tipo de uso de solo. Por exemplo, as viagens de casa para o trabalho seriam distribuídas apenas para as zonas onde há emprego. A segunda corresponde à entrada das características da rede de transporte, i.e., definição da função da acessibilidade relativa de uma determinada zona, que é medida como uma função do tempo de viagem (congestionamento) e do custo (tarifas de trânsito, taxas de estacionamento, portagens, entre outros).

Um dos modelos mais usados para definir esta etapa é o modelo de Gravitacional, que integra a dimensão e crescimento dos diversos usos de solo nas zonas de origem e de destino e a respetiva impedância relativa entre as zonas para distribuir viagens, através da utilização de variáveis de custo.

Repartição modal

Na etapa correspondente à repartição modal, as viagens distribuídas na etapa anterior são alocadas aos diferentes modos de transporte disponíveis. Normalmente, a alocação é entre carros, peões e ciclistas (modos ativos) e o transporte público (autocarros), mas alguns modelos também diferenciam entre os modos de transporte público, os veículos de alta ocupação (HOV – veículos com dois ou mais ocupantes) e veículos de um único ocupante (SOV – veículos em que o único ocupante é o condutor) (Kriger et al., 2016).

Afetação de viagens à rede

Esta etapa consiste em atribuir, ou carregar de acordo com cada modo, as viagens à rede de transportes respetiva. Segundo Kriger et al. (2016) esta etapa traduz a procura (expressa como

o número de viagens por modo X, para todos os fins combinados entre a zona i e a zona j) em volumes de tráfego para um determinado eixo (link) de estrada e conseqüentemente em número de passageiros para uma rota de um serviço de autocarro, ou de transportes públicos similares.

O processo de afetação pode tomar-se extremamente simples no caso de redes pouco densas e com baixos níveis de tráfego, em que a escolha do percurso, dada a origem, o destino e o modo de transporte é quase “óbvia”, e algumas vezes, única.

Uma vez apresentado o modelo de 4 etapas, de uma maneira mais geral, tendo em conta que cada uma das etapas apresenta um vasto conjunto de investigações e o seu estudo aprofundado, representaria por si só, um outro trabalho de investigação e pesquisa que não se coaduna com o âmbito desta dissertação. De acordo com os objetivos desta dissertação concluí-se que a modelação da procura associada ao sistema de transporte por autocarro em Territórios de Baixa Densidade encontra-se, sobretudo, associado com as primeiras três etapas do modelo clássico de transportes, sendo então apresentado com mais detalhe essas três etapas nos pontos seguintes.

2.2.2.1 Modelos de Geração de Viagens

A geração de viagens procura estimar o número total de viagens que são produzidas (P_i) ou atraídas (A_j) por cada uma das zonas que compõem a área de estudo (vetor de viagens atraídas e produzidas). Geralmente, a geração de viagens produzidas é medida pela unidade – “casa” dentro de um determinado zonamento na qual se dividiu a área estudada.

Segundo Ortúzar (2015), as viagens da população que interessam integrar no modelo são aquelas que se realizam em modos de transporte motorizados. No caso de viagens realizados pelos peões (os quais têm que ter um percurso maior a 300 metros) também são considerados, porém não devem ser considerados e modeladas viagens de peões com menos de 5 anos (crianças). Estas viagens são classificadas em dois tipos: “*home based*” (HB) e “*non home based*” (NHB), de acordo com a sua partida ou chegada, i.e., as viagens HB (home based) são as viagens que têm como origem ou destino uma casa (domicílio), sendo as restantes classificadas como NHB, e.g., uma viagem que seja de casa para o centro da cidade é classificada com uma viagem tipo HB, sendo que uma viagem do centro da cidade para um centro comercial é classificada como uma viagem NHB. Atualmente, em muitos contextos

urbanos é possível observar a existência de um outro tipo, frequente, de viagens que recebem o nome de “tours” que correspondem a um tipo específico de NHB, mas que são caracterizadas e reguladas por padrões de utilização e procura distintos dos da população residente, mostrando mais uma vez que a procura de transporte é dependente fortemente do tipo de atividades e propósitos das viagens.

O total de viagens (P_i e A_j) devem estar classificado e caracterizado, tendo em conta outros aspetos importantes, de modo a se obter melhores modelos de geração. Uma classificação pelos tipos de viagens é fundamental, nomeadamente no tipo HB, onde estas dependem muito da atividade que se vai realizar de e para casa, e.g., trabalho, compras estudo, social e recreativo, entre outros. Já, para o tipo NHB, comumente não existe uma classificação fixa, dado que habitualmente só chegam a valores entre os 15% e 20% das viagens (Ortúzar, 2015).

É comum que as viagens do tipo HB que tenham como principal propósito a deslocação para o trabalho ou para a escola se realizem diariamente e apresentem um padrão regular, assumindo vulgarmente a designação de viagens pendulares ou de comutação. As demais viagens realizam-se correntemente de uma maneira mais esporádica (Cal Mayor & Asociados, 2005).

Por outro lado, importa distinguir os diferentes períodos de análise, para as classes HB e NHB, entre hora de ponta (peak) e fora das horas ponta (off-peak). Em cada período, a proporção de viagens muda de acordo com os distintos propósitos. Outra classificação potencialmente importante segunda é por tipo de pessoa (Ortúzar, 2015), já que o comportamento dos indivíduos com respeito a as viagens depende de vários fatores, nomeadamente das suas características socioeconómicas, sendo proposta a seguinte classificação:

- Nível de salario (baixo, médio e alto);
- Posse de um veículo (e.g., 0, 1 e 2 ou mais veículos em casa);
- Tamanho da casa (pequena, média e grande).

Segundo Ortúzar (2015), é necessário ter cuidado com o número de classificações feitas, já que isto pode influir na quantidade de modelos por calibrar, aumentando assim o compromisso, ajustes e agregações.

Os modelos utilizados na geração de viagens correspondem à aplicação de procedimentos estatísticos convencionais para definir as equações de estimação do número de viagens geradas.

Assim, para é possível constatar que para dados que provenham de inquéritos domiciliários e de informação socioeconómica os modelos mais usados são os de regressão linear simples e, ou múltipla, e pelo método da classificação cruzada.

Modelos de geração de viagens obtidos por análises de regressão linear simples e múltipla

Nos casos em que a geração de viagens se possa expressar como uma função linear de uma só variável independente, pode-se utilizar a regressão linear simples para encontrar o modelo que determine o melhor ajuste possível aos dados analisados. Os indicadores estatísticos utilizados nomeadamente para estabelecer a precisão do ajuste obtido são o coeficiente de correlação (R) e o coeficiente de determinação (R^2) (Cal Mayor & Asociados, 2005).

Em alguns casos, a geração das viagens são função de duas ou mais variáveis independentes. Nestes casos pode-se aplicar a regressão linear múltipla, a qual se considera como uma extensão da regressão linear simples. Uma vez que que nesta dissertação se procurará modelar a procura de transporte público (autocarro) para territórios de baixa densidade e sabendo que esta será, certamente, função de mais de uma variável, este tipo de regressão será analisado com mais detalhe.

Análises da regressão linear múltipla

A análise de regressão linear múltipla (RLM) consiste em relacionar um conjunto de variáveis independentes com uma única variável dependente, predizendo as mudanças da mesma com respeito as outras variáveis.

Os modelos estatísticos baseiam-se em pressupostos que permitem efetuar uma melhor abordagem dos problemas, onde se inclui a RLM. Deste modo, serão apresentados os pressupostos que se consideram numa análise deste tipo, tendo em conta apenas os aspetos mais relevantes para a RLM (Herrera, 2014)

Pressupostos a considerar numa análise de RLM

- *Linearidade:* é necessário avaliar por dois motivos principais, o primeiro motivo prende-se com a necessidade de se avaliar se a variável independente está linearmente relacionada com a variável dependente e o outro motivo é quando se exclui uma variável relevante do modelo ou, caso contrário, se considera uma variável insignificante. Uma forma recomendada por Herrera (2014) consiste em fazer uma série de análises parciais

bivariadas entre a variável dependente e cada uma das variáveis independentes, de modo a se poder observar se a relação é linear, quadrática ou de outro tipo.

- *Independência*: num modelo de RLM ótimo tem que se verificar a independência dos seus resíduos. O teste estatístico usado para medir essa independência é o Durbin-Watson (1951), podendo o seu valor variar entre 0 e 4, sendo que 2 é o valor que representa a independência. Para Herrera (2014) pode-se considerar independência dos resíduos quando os valores se encontram entre 1,5 e 2,5.

A estatística do teste é dada por:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (1)$$

Onde se tem que $e_i = Y_i - Y_l$, Y_i é o valor da variável explicativa, Y_l é o valor estimado da variável segundo o modelo.

- *Homocedasticidade*: indica que para cada combinação de valores das variáveis independentes a variância dos resíduos deve ser constante. Para realizar esta análise usa-se o método gráfico, que consiste em elaborar e contrapor os gráficos de resíduos tipificados e de prognósticos tipificados. Nesta análise gráfica tem de se observar uma nuvem de pontos sem nenhum padrão definido;
- *Normalidade*: os resíduos da regressão linear devem apresentar uma distribuição Norma, nomeadamente com média de zero. Para se observar a normalidade dos resíduos podem-se utilizar os histogramas de resíduos tipificados e os gráficos de probabilidade normal.

Segundo Ortúzar (2015), para $f(Y/X)$ corresponde à distribuição de probabilidade de Y para um valor dado de X. O modelo de regressão linear pressupõe que:

- i. A distribuição $f(Y/X)$ tem a mesma variância σ^2 para todos os níveis experimentais da variável independente X;
- ii. A esperança $E(Y)$ encontra-se numa linha reta, denominada como a verdadeira reta de regressão linear (ou da população), está reta é dada por:

$$E(Y) = \alpha + \beta X$$

Os parâmetros populacionais α e β , que definem a reta devem ser estimados de uma amostra;

- iii. As variáveis aleatórias Y são estatisticamente independentes, i.e., um valor de Y_1 não afeta o valor de Y_2 . As diferenças (desvios) de Y em relação ao valor esperado ($E(Y)$) dá origem ao erro e . Assim de acordo com a expressão definida em ii), resulta:

$$Y_i = \alpha + \beta X_i + e_i$$

- iv. A variável Y tem uma distribuição normal. Para além disso, importa referir que não há limites para a quantidade de variáveis no modelo, desde que estas não estejam relacionadas entre si, ou seja, não exista multicolinealidade dos X_i .

Modelo de classificação cruzada

Uma outra forma de modelar a procura de transporte é através do desenvolvimento de modelos obtidos por a classificação cruzada, também designados modelos de categorias, este método é o mais usado para prever a produção das viagens com uma origem/destino num domicílio/ casa, i.e., viagens do tipo HB.

Na classificação cruzada obtêm-se taxas de produção média de viagens para cada uma das combinações em que sejam ordenadas as casas. Habitualmente, as análises das taxas de produção de viagens é feita para toda a área de estudo e não para cada uma das zonas.

Se é conhecida a distribuição das casas, em função dos parâmetros utilizados na classificação cruzada, é possível estimar com uma boa precisão a produção total de viagens do tipo HB (home based) em cada uma das zonas analisadas do transporte. Para além disso, importa realçar que para se poder aplicar o modelo de classificação cruzada é fundamental levantar a informação certa e atualizada das casas em cada zona de análise, do mesmo modo tem que se contar com os prognósticos dos atributos das casas definidos no âmbito da classificação cruzada, de modo a se poder realizar análises de cenários futuros para o planeamento de transportes (Flores Reyna, 2017).

O método baseia-se na estimação da resposta (número de viagens produzidos, por cada Casa, com um propósito/ atividade específica) em função dos atributos da casa. Este é o pressuposto básico das taxas de geração de viagens, que se encontram estáveis para determinados períodos de tempo e características socioeconómicas das Casas/ Domicílios, e uma das desvantagens apresentadas pelo modelo, faz referência ao estabelecimento de moradias em cada um dos estratos no futuro.

Algumas das vantagens apresentadas para o modelo de classificação cruzada são (Cal Mayor & Asociados, 2005):

- Os grupos de classificação cruzada são independentes do zonamento da área de estudo;
- Não é necessário atender a pressupostos das relações consideradas na geração de viagens, nomeadamente em relação aos modelos de regressão linear, linear múltipla;
- As relações podem mudar em cada tipo, um exemplo de isto é o efeito do tamanho da casa para ter um veículo poder ser diferente.

2.2.2.2 Modelos de Distribuição de Viagens

Nesta segunda etapa do modelo clássico é realizada a distribuição do número de viagens geradas em cada origem (1ª etapa) pelos vários destinos da área de estudo. O resultado desta atividade é conhecido como a estruturação da matriz origem (O) e destino (D) das viagens. As matrizes O/D podem ser desagregadas por tipo de veículo (sem e com veículo; veículos pesados, autocarros, etc.), de acordo com o propósito da viagem (trabalho, compras, lazer, etc.) e por hora do dia (ponta da manhã, fora dos períodos de ponta, total diário). É importante sublinhar que a desagregação não é sempre a mesma, já que depende dos objetivos do estudo e da disponibilidade dos dados. Na Tabela 4 apresenta-se a estrutura da matriz e o hipotético preenchimento das células da mesma com as diferentes viagens entre origens e destinos.

Tabela 4. Matriz de distribuição. Fonte: (Ortúzar, 2000)

Atracciones Producciones	1	2	...	j	...	n	$\sum_j V_{ij}$
1	V_{11}	V_{12}	...	V_{1j}	...	V_{1n}	O_1
2	V_{21}	V_{22}	...	V_{2j}	...	V_{2n}	O_2
...
...
...
i	V_{i1}	V_{i2}	...	V_{ij}	...	V_{in}	O_i
...
...
...
n	V_{n1}	V_{n2}	...	V_{nj}	...	V_{nn}	O_n
$\sum_i V_{ij}$	D_1	D_2	...	D_j	...	D_n	$\sum_i V_{ij} = V$

A nomenclatura para a tabela 4 é a seguinte:

v_{ij} = Viagens atuais entre as zonas i e j

V_{ij} = Viagens futuras entre as zonas i e j

O_i e D_j = Viagens futuras produzidas na zona i e atraídas na zona j

n = número total de zonas na área de estudo

Alguns dos modelos para determinar a forma como as viagens são distribuídos pela matriz O/D (e.g., Tabela 4) são expostos a seguir, tendo em conta que só se apresentarão os principais aspectos de cada método.

I. **Modelo de fator de crescimento:** este modelo assume que em princípio existe uma matriz de viagens base v . Com este modelo pretende-se estimar uma matriz de viagens para um tempo de projeto X , tendo por base os cálculos de fatores de crescimento, sendo para esse efeito necessário realizar contagens de alguns dados das ZAT. O método a utilizar para determinar o fator de crescimento depende dos dados disponíveis, as possíveis metodologias são:

a) **Método do fator uniforme:** tem como principal vantagem o facto de não ser iterativo, uma vez que assenta no pressuposto de que as viagens em toda a área de estudo aumentarão de uma forma uniforme, resultando em:

$$V_{ij} = F v_{ij}$$

b) **Método do fator média** - tem por base a seguinte equação:

$$V_{ij}^{(1)} = \frac{F_i^{(1)} + F_j^{(1)}}{2} \cdot v_{ij} \quad (2)$$

Onde se tem que $F_i^1 = \frac{P_i^f TM_i^f Y_i^f}{P_i^d TM_i^d Y_i^d}$ e é a mesma equação para F_j^1 , este processo é iterativo, sendo que a iteração chega a seu fim quando os valores de F_i^n e F_j^n são aproximadamente iguais. Este método não tem uma precisão muito boa.

c) **Modelo de Detroit:** este é um modelo iterativo e apresenta resultados similares ao do fator média. O fator de crescimento geral da área (F) e o número de viagens entre pares O/D são calculados com base nas seguintes expressões.

$$V_{ij} = v_{ij} F_{ij} \quad (3)$$

$$\text{nde } F_{ij} = \frac{F_i F_j}{F} \quad \text{e} \quad F = \frac{\sum_{ij} V_{ij}}{\sum_{ij} v_{ij}}$$

d) **Modelo de Fratar:** também é um modelo iterativo, mas que requer um menor número de iterações. O cálculo das viagens entre pares O/D é dado por:

$$V_{ij} = v_{ij} F_i F_j \left(\frac{L_i + L_j}{2} \right) \quad (4)$$

$$\text{Onde } L_i = \frac{\sum_j v_{ij}}{\sum_j v_{ij} F_j} \quad \text{e} \quad L_j = \frac{\sum_i v_{ij}}{\sum_i v_{ij} F_i}$$

Segundo Ortúzar (2000) refere que as vantagens dos modelos de fator de crescimento residem na menor dificuldade de compreensão e aplicação dos mesmos, apresentando um equilíbrio relativamente rápido. Para além disso, os seus resultados são mais precisos a curto prazo e quando são aplicados a áreas com uma estrutura de desenvolvimento estável. As principais desvantagens, que se consideram mais importantes, prendem-se com o facto destes modelos requerem como entrada os mesmos dados que modelos mais avançados (dados de inquéritos de origem- destino) e produzirem resultados menos precisos. Uma das aplicações mais comuns deste tipo de modelo é a previsão, a curto prazo e em áreas estáveis, dos níveis de tráfego que cruzam uma área, sendo também muito aplicados em atualizações dos dados de inquéritos de origem-destino.

II. **Modelo gravitacional:** a formulação deste modelo baseia-se na Física de Newton e tem aplicabilidade em várias zonas com áreas urbanas e/ou conjunto de cidades para determinar o número de viagens, entre pares de O/D, através da seguinte equação:

$$V_{ij} = \frac{KO_iD_j}{d_{ij}^n} \quad (5)$$

Onde d é a distância entre as duas áreas; O são as produções da zona (i) e D são as atrações da zona (j).

A aplicação deste modelo, ao longo dos anos, revelou que a distância (d) pode ser importante em casos interurbanos, mas não o é para estimar viagens no interior das cidades, nomeadamente quando existem congestionamentos. Assim, a distância (d) tem vindo a ser substituída por funções de custo - C_{ij} (custo geral de viagem), que pode incluir elementos como o tempo, custo monetário da viagem (associado à distancia), entre outros parâmetros.

Alguns dados complementares para este modelo (Cal Mayor & Asociados, 2005) são expostos a seguir:

- Distribuição de frequência da duração ou da extensão das viagens. Esta informação pode ser obtida de análises estatísticas dos dados coletados em inquéritos às habitações, realizando uma frequência para cada um dos propósitos de viagem considerados no desenvolvimento dos modelos básicos;
- Distância entre todo os pares de centroides das zonas de análise do transporte (ZAT);
- Custos das viagens realizadas na região de estudo. Estes dados servem para determinar as funções que definem a utilidade negativa (impedância) ou o custo generalizado das viagens realizadas entre duas ZAT.

III. **A abordagem de maximização da entropia:** primeiro é importante fazer referência ao significado de entropia, já que este termo é usado em diferentes áreas do conhecimento. Para o caso desta dissertação é importante realçar que é um conceito subjetivo que pode ser utilizado como ferramenta para contruir modelos. Neste caso a entropia não esta

associada ao sistema, estando pelo contrario associada à informação que o analista tem do sistema. Neste sentido é possível considerar dois tipos entropia:

- Entropia como probabilidade (relacionada com a noção de incerteza)
- Entropia como estadigrafo de uma distribuição de probabilidade.

a) **Entropia como probabilidade:** em princípio considera-se que um sistema é composto por componentes separáveis e distinguíveis, sendo organizado em três estados: i) estado micro - que corresponde a uma descrição completa e específica em que cada um dos componentes é único; ii) estado meso - neste tem-se uma descrição não definida mas não complexa; iii) estado macro – que tem como propósito analisar as restrições totais do sistema, i.e., as origens (O_i) e os destinos (D_j). O importante que se tem que ter em consideração é que muitos estados micro produzem um estado meso, e muitos estados meso produzem um estado macro.

A essência deste método, é a suposição que todos os estados micro são igualmente prováveis (sempre e quando não exista informação que refira o contrário). As únicas restrições que são feitas ao sistema são a nível macro.

b) **Entropia de uma distribuição de probabilidade:** este método é mais flexível que o anterior já que define a probabilidade como um conjunto subjetivo que no fundo expressa ignorância do modelador.

2.2.2.3 Repartição modal

Nesta etapa é possível obter a percentagem de utilização de cada modo de transporte disponível, isto está relacionad com o número de viagens inter e intrazonais. Os modelos de repartição modal quantificam a quantidade e o custo do sistema de transporte existente ou proposto. São os modelos que têm maior importância dentro do planeamento de transportes, porque afeta a eficiência geral do sistema de transportes nas áreas de estudo, uma vez que oferece diferentes possibilidades de realização de viagens em vários modos de transporte.

Os modelos utilizados nesta etapa são conhecidos como modelos de eleição discreta os também conhecidos modelos desagregados, estando assentes nas eleições individuais, i.e., postulam que

a probabilidade de escolher um modo de transporte é função das características socioeconómicas do indivíduo e da atratividade do modo em comparação com os outros modos.

As principais propriedades dos modelos desagregados são (Ortúzar, 2015):

- Os modelos desagregados (MD) são estimados através da utilização de observações do comportamento dos indivíduos como dados de entrada;
- Os MD são probabilísticos. Como entregam a probabilidade de escolher cada alternativa e não indicam qual é a escolhida, devem fazer uso de conceitos probabilísticos básicos, como a média e a probabilidade conjunta;
- Para as variáveis explicativas, que se incluem nos modelos, são estimados coeficientes de uma forma explícita;
- Os MD baseiam-se em teorias de eleição individual e não constituem analogias de nenhum tipo, mas importa referir que estas teorias também permitem gerar modelos agregados.

O uso da teoria da utilidade aleatória tem vindo a ser a abordagem mais usado pelos modelos de eleição discreta. Esta teoria postula que os indivíduos fazem parte de uma certa população homogénea Q , atuam de modo racional e têm informação perfeita, i.e., que sempre escolhem o percurso (caminho) que lhes permite maximizar a sua utilidade pessoal (este tipo de “espécie” é denominada *homo economicus*), sujeito às suas restrições ambientais. Estas podem ser legais, sociais, físicas ou orçamentais (e.g., tempo e dinheiro).

Para apresentar de uma maneira geral o conteúdo da teoria da utilidade aleatória é importante realçar que se tem um conjunto de alternativas disponíveis que é denominado por $\underline{A} = \{A_1, \dots, A_i, \dots, A_N\}$, onde cada alternativa $A_i \in \underline{A}$ e tem associada uma utilidade U_{iq} para o indivíduo q . Como o modelador é um observador do sistema e não possui informação completa em relação a todos os elementos que cada indivíduo tem em conta no momento de realizar sua eleição, ele supõe que U_{iq} pode-se representar por dois componentes:

- Uma parte mensurável, sistemática ou representativa, \overline{U}_{iq} , que é função dos atributos medidos \underline{x} ;
- Uma porção aleatória ε_{iq} que representa a idiosincrasia e os gostos particulares de cada indivíduo, além de erros de medição e observação por parte do modelador.

$$U_{iq} = \overline{U}_{iq} + \varepsilon_{iq} \quad (6)$$

Isto permite explicar duas irracionalidades; a primeira que dois indivíduos iguais escolham diferentes alternativas e a segunda que um indivíduo escolha uma alternativa que não é a melhor (segundo a perspectiva dos atributos escolhidos pelo modelador).

Um dos modelos que tem maior popularidade, pela sua simplicidade, é chamado **Modelo Logit Múltiplo** (também conhecido com MNL), que se caracteriza por ser gerado a partir de funções independentes e identicamente distribuídas (IID) Gumbel (também conhecida como Weibull ou de valor extremo Tipo I). A forma geral do modelo e que tem mais uso na prática, é aquela onde não é preciso que todos os indivíduos podam ter todas as alternativas a sua disposição, sendo dada pela seguinte equação:

$$P_i = \frac{\exp(\beta \overline{U}_{iq})}{\sum_{A_j \in \Delta(q)} (\exp(\beta \overline{U}_{jq}))} \quad (7)$$

A propriedade mais importante deste modelo e que é apontada como a principal causa para uma maior utilização, é a denominada “independência de alternativas irrelevantes (IAA)” e que se expressa em que o cociente em la equação (7) é uma constante, independente das utilidades das demais alternativas.

$$\frac{P_i}{P_j} = \exp\{\beta(\overline{U}_i - \overline{U}_j)\} \quad (8)$$

Em princípio esta propriedade foi considerada como uma vantagem, já que permite tratar o problema de prever o impacto de opções novas, no futuro, sem ser necessário re-calibrar o modelo, mas isto foi mudando ao longo do tempo e na atualidade é mais importante ter atenção à falha do modelo quando há alternativas que estão correlacionadas, ou seja, se alguns grupos de alternativas são mais parecidos entre si do que outros.

O modelo anteriormente exposto tem problemas quando há correlações entre alternativas, mas para isso e outros problemas relativos à complexidade dos mesmos, Ortuzar (2015) propõe o modelo de Logit mixto, o qual permite tratar matrizes gerais de covariância e adicionar tópicos

importantes. O Logit mixto requer igualmente um processo de estimação complexo e sua interpretação é também um problema difícil de resolver devido a problemas de identificabilidade e de confusão.

2.2.3 Outros métodos e modelos para a estimação da procura

Modelos de tendência

Este tipo de modelos associa a variação da procura ao longo do tempo assumindo que os fatores que afetam a mesma se manterão no futuro, i.e., representando uma tendência similar, alguns dos modelos mais comuns segundo Islas Riveira et al.(2002) são:

✓ *Tendência Linear*: baseia-se que o comportamento se assume como uma linha reta tipo:

$$Y = m * t + b \quad (9)$$

Y: procura de transporte (viagens, passageiros, toneladas, ton-Km, etc)

t: tempo (ano, período, antiguidade, etc)

m, b: parâmetros que se obtêm pelo método dos mínimos quadrados.

É importante esclarecer que só se pode usar este modelo quando o crescimento é realmente linear (i.e., taxa de crescimento constante), que o coeficiente de correlação e os demais testes estatísticos sejam satisfatórios. Este método só representa uma aproximação à tendência da procura de transporte, uma vez que o tempo não é a variável que explica, de uma forma muito precisa, o comportamento da procura de transporte (dado que existem mais variáveis que afetam o seu comportamento).

✓ *Tendência geométrica*: tem um padrão de crescimento a uma taxa constante, sendo que a sua composição se representa da seguinte forma:

$$Y_f = Y_p + (1 + r)^t \quad (10)$$

Y_f: Procura futura

Y_p: Procura atual (presente)

r: taxa de crescimento por período, ano, entre outros;

t: tempo (ano, período, etc)

- ✓ *Tendência exponencial modificada*: assume que a procura tende a aumentar até um limite predeterminado, que se pode considerar como o nível de saturação ou a máxima capacidade, sendo calculada através da seguinte equação:

$$Y = K + a * b^t \quad (11)$$

Onde K é o limite de crescimento, as demais variáveis correspondem de mesma maneira como já se tinha definido anteriormente.

- ✓ *Curva de Goempertz*: neste caso a procura é representada por um comportamento em forma de “S”, por outras palavras tem um crescimento lento e estendido ao início seguido por uma fase de grande crescimento, seguindo-se uma fase de diminuição da taxa de crescimento e, por último, uma fase de saturação com uma menor taxa de crescimento. A procura é dada pela seguinte equação:

$$\text{Log}(Y) = \log K + b^t * \log a \quad (12)$$

- ✓ *Curva logística*: tem um comportamento semelhante ao de *Goempertz*, mas neste as mudanças de fase são mais suaves, com a procura a ser determinada pelas seguinte expressão:

$$Y = \frac{K}{1 + b * e^{-a*t}} \quad (13)$$

Onde e é o logaritmo neperiano e a demais variáveis são idênticas às das curvas anteriores.

A modelação da procura de transporte constitui uma parte importante do processo de planeamento de transportes, sendo tanto mais significativa quanto maior for a abrangência da perspectiva com que os mesmos a analisam, com implicações em maiores exigências naturalmente ditadas pela complexidade dos trabalhos a desenvolver. Os modelos devem pois responder a novas necessidades de planeamento, e no final deverão contribuir para a avaliação e conceção das opções técnicas mais ajustadas aos diversos contextos (Teixeira, 2002).

Simulações com a matriz matéria prima-produto (entrada-saída)

É uma tabela ou arranjo que apresenta as relações interindustriais de um país ou região em um determinado período, compreendendo as relações industriais, como compras e vendas, que os diferentes setores da economia fazem uns com os outros, por exemplo, o setor agrícola vende e compra produtos para outros setores. Também as compras e vendas podem ocorrer dentro do próprio setor.

Se existirem n setores, essas relações intersectoriais podem ser representadas numa matriz quadrada de dimensão $N \times N$ (A_{ij}), onde os elementos representam as vendas realizadas pelo setor i para o setor j , ou vice-versa.

$$\begin{bmatrix} (A_{ij}) & B \\ C & D \end{bmatrix}$$

Onde (A_{ij}) é a matriz de dimensão $n \times n$ que representa os fluxos intersectoriais; B é a matriz de $n \times p$ como a procura final para cada um dos n setores; C a matriz de $l \times n$ que representa todos os requisitos que possuem os diferentes componentes da procura final da matéria-prima primária.

Esta simulação, embora tenha um grande potencial teórico, tem três limitações importantes, segundo Islas Riveira et al.(2002) a quais são:

- Comportamento de previsão em quantidades, não em preços. Isso é importante porque, ao simular uma mudança na economia, são feitos ajustes nas duas variáveis: os preços não permanecem fixos ou proporcionais.
- Gerenciar setores, não regiões. Se considerarmos que o transporte é essencialmente um fenômeno espacial e que a demanda é melhor medida em toneladas-quilômetro ou passageiro-quilômetro, é necessário transformar a procura estimada em valor agregado em toda a economia em alguma unidade de procura, localizada geograficamente.
- A proporção de matéria-prima é constante e determinista, ou seja, assume-se que as tecnologias de produção são constantes e conhecidas. Esta suposição é, dentro de um contexto de frequente mudança tecnológica e incerteza, difícil de aceitar.
- No entanto, com a implementação de melhorias recentes do método de entrada-saída, e/ou combinadas com outras técnicas, elas são consideradas uma das melhores opções devido aos seus custos relativamente baixos e simplicidade.

3 METODOLOGIA

3.1 Enquadramento metodológico

Para realizar um planeamento adequado do sistema de transportes deve-se ter em conta o nível de detalhe, escala e período do tempo para o qual se espera que as ações sejam efetivas e devidamente implementadas sem que seja necessário algum tipo de redefinição. Tendo em consideração estas características fundamentais é necessário definir uma proposta metodológica para cumprir com os objetivos definidos. Assim, Cal Mayor & Asociados (2005) apresenta as diferentes escalas do planeamento do transportes da seguinte forma:

- Planeamento a curto prazo: neste nível o período é inferior a 5 anos, sendo considerado adequado para o planeamento estratégico, no qual se analisa o problema de transportes de uma maneira muito específica considerando o maior número de fatores que influenciam o problema;
- Planeamento a médio prazo: o período que decorre entre 5 e 15 anos, verificando-se que neste nível os problemas são analisados de uma maneira embora específica, mas com menos detalhe que o curto prazo;
- Planeamento a longo prazo: para um período entre 15 anos e 30 anos. Neste tipo de planeamento o nível de detalhe é inferior aos dos outros períodos, sendo necessário realizar periodicamente ajuste ou alterarem-se os planos a longo prazo para ter em conta mudanças inesperadas na população, desenvolvimento, rendimento e atividades sociais e económicas.

No âmbito deste trabalho pretende-se definir uma metodologia para estimar a procura do transporte em autocarro nos territórios que sejam considerados TBD, ou que tenham comportamentos similares. Esta estimação será usada para escolher e analisar as estratégias para o planeamento e implementação do serviço de transporte público nesses territórios. Assim, tendo em conta que os TBDs têm comportamentos dinâmicos e complexos, entende-se que devem ser implementadas um conjunto de medidas que se adequama às três escaals apresentadas anteriormente.

Ao longo do estado de arte encontraram-se algumas recomendações e dicas para ter em conta ao momento de desenvolver um estudo para o planeamento do transporte, especialmente para

qa produção de um modelo de procura de transporte. Por conseguinte, se terá como base os resultados da pesquisa feita na secção anterior para o desenvolvimento da metodologia associada à construção de um modelo de procura de transporte. Algumas das recomendações qualitativas, serão ainda implementas ao longo da aplicação em cada uma das etapas do caso de estudo.

A base de um modelo de procura de transporte é seu modelo matemático, com isto se pode estimara a procura que terá cada zona de análise. Na secção anterior estudou-se e pretendeu-se determinar qual é a técnica ou modelo matemático mais apropriado para esse efeito, i.e., que tem os resultados mais ajustados e precisos, ou seja, com o erro de previsão mais baixo.

Segundo Cal Mayor & Asociados (2005), Ortuzar & Willumsen (1994) e Teixeira (2002), o método da regressão linear múltipla é um dos mais usados dado que a sua utilização aprsenta geralmente valores similares ao comportamento real. Por conseguinte, e seguindo com a recomendação dos diferentes autores, o modelo (método) matemático que será desenvolvido para a determinar da procura de transporte em autocarros nos territórios de baixa densidade será a regressão linear múltipla RLM.

A RLM pressupõe uma caracterização e tratamento incial das variáveis que serão utulizadas no modelo. Assim, segundo Pestana & Gageiro (2014), é necessário identificar as escalas ou níveis de medidas das variáveis para a escolha do tratamento estatístico adequado,s endo identificadas quatro escalas ou níveis de variáveis: nominal, ordinal, intervalar e rácio.

Numa escala nominal os elementos são atributos ou qualidades. Os números servem apenas para identificar ou categorizar os elementos, um exemplo de este tipo de variáveis pode ser a identificação das pessoas através do seu número de cartão de cidadão. Outras variáveis de escala nominal incluem, por exemplo, o sexo, a reça, a religião.

Numa escala ordinal apenas se pode distinguir diferentes graus de um atributo ou variável, existindo, portanto, entre eles uma relação de ordem. Os números são associados de modo que a relação de ordem se mantenha, como por exemplo, os escalões de rendimento, as classes etárias, as ordens de preferências.

A escala intervalar usa números para classificar os elementos, sendo feita de forma que, a igual diferença entre os números, corresponda a igual diferença nas quantidades do atributo medido. O zero é um valor arbitrário e não representa a ausência da característica medida.

Uma escala rácio só difere de uma escala intervalar, porque o zero tem existência real, denotando ausência da característica medida.

3.2 Definição de variáveis

A definição das variáveis é um dos trabalhos mais importantes no desenvolvimento de um modelo, uma vez que será com base na informação que elas contenham que será definida a procura de transporte. As variáveis exógenas são a “pedra angular” do modelo e possuem informação socioeconómica e demográfica da área à qual será aplicado o modelo.

Para determinar as variáveis que serão analisadas no estudo do modelo, foi necessário em abordar o tema dos territórios de baixa densidade e determinar quais eram os problemas de mobilidade e suas principais características intrínsecas, através de pesquisas bibliográficas e científicas de diversos autores, empresas e organizações governamentais e não-governamentais, que tenham realizado algum tipo de investigação, ou projeto, no tema de modelos de procura de transporte para os TBD ou territórios com características similares.

Segundo Litman (2017), os fatores demográficos, geográficos e económicos podem afetar a procura das viagens. Alguns exemplos destes fatores são o número de pessoas (residentes, empregados e visitantes), o nível de emprego, a escolaridade, a idade/ ciclo de vida, os estilos de vida, entre outros aspetos culturais.

As variáveis demográficas como a idade, a escolaridade e o ciclo de vida tem efeitos diferenciais e interativos sobre o comportamento de viagem, Kattiyapornpong (2006) sublinha um grupo de pessoas que apresentam um comportamento diferenciado com influência na procura de transporte é o dos jovens entre 20 e os 24 anos, cuja escolha modal para a viagem está relacionada com a escolaridade, a etapa de vida e a idade. Curtis & Perkins (2006) estudou o impacto das variáveis demográficas no comportamento das viagens e encontrou algumas relações relevantes entre o comportamento da viagem, incluindo a eleição do modo e variáveis como a idade, o sexo, a propriedade do veículo, a composição da residência e os rendimentos.

As pessoas de diferentes grupos de idade apresentam um padrão de viagens muito distinto, isto ocorre porque tendem a realizar atividades distintas, e.g., para as pessoas mais velhas as restrições físicas afetam muito o o tipo de viagens e modos de transporte utilizados. O estado laboral é considerado um fator importante para decidir o padrão da viagem, Best & Lanzendorf (2005) demonstraram que os trabalhadores a tempo parcial têm uma probabilidade mais alta de viajar com maior frequência do que os trabalhadores a tempo integral, já que eles tem um número de atividades mais variada ao longo do dia.

Algumas entidades governamentais elaboraram diversos estudos para determinar quais são os grupos de pessoas mais afetadas pela falta de serviços de transporte público, já que uma das políticas governamentais prende-se com a oferta de serviço público, nomeadamente no setor dos transportes, de modo que as pessoas se possam deslocar para os diferentes destinos. Um exemplo disso é o governo de Nova Zelândia, onde através do manual de planeamento, programação e financiamento da agência de transporte e da lei de gestão do transporte público tem vindo a classificar os seguintes grupos pessoas:

- Pessoas com idade avançada (idosos);
- Jovens e crianças;
- Pessoas que não tem veículos com motor;
- Pessoas com problemas de mobilidade (desapacitados);
- Pessoas com um rendimento salarial baixo;
- Grupos minoritários e novos imigrantes;
- Pessoas em áreas não atendidas por transporte público.

Os distintos grupos etários e o sexo são de facto uns dos fatores que afetam o comportamento e os padrões de viagens, isto ocorre porque cada grupo tende a participar em diferentes tipos de atividades. Dieleman (2002) concluiu que as famílias com salários mais elevados têm maior níveis de propriedade e uso do veículo motorizados (automóvel), para além das famílias com filhos apresentarem uma maior probabilidade de utilização regular do automóvel.

A forma urbana pode influenciar consideravelmente os comportamentos das viagens. Segundo Carey & Curtins (2006) os territórios de baixa densidade produzem viagens mais longas e com maior dependência do veículo privado como principal meio de transporte. O uso do solo, a

maior densidade e o zonamento misto apresentam maior probabilidade de comportamentos sustentáveis para os padrões de viagens.

Wang et al. (2014) determinou que a procura nos transportes públicos não convencionais é maior nos territórios de baixa densidade populacional e com menor propriedade de veículos. Entretanto, a universidade de Loughborough em 2006, identificou possíveis mercados para um sistema de transporte não convencional para território de baixa densidade (condado de Wiltshire), que são: as pessoas que não podem utilizar transporte público, pessoas sem transporte próprio, pessoas desempregadas, pessoas inativas.

Com base na pesquisa bibliográfica é possível concluir que há vários fatores socioeconómicos, demográficos e características das famílias que afetam as viagens dos indivíduos não só nos territórios de baixa densidade como em qualquer território onde seja necessário utilizar um sistema de transporte público. Assim, da análise do estado da arte, as variáveis que são determinadas como influentes para o comportamento dinâmico das viagens nos territórios de baixa densidade são apresentadas a seguir:

- i. Pessoas entre os 15 e 24 e 55 ou mais de idade;
- ii. Número de viagens das mulheres;
- iii. Número de pessoas desempregadas;
- iv. Número de pessoas sem carta de condução;
- v. Número de veículos por habitação;
- vi. Ganho médio mensal por família;
- vii. Número de viagens para fins de compras ou sociais;
- viii. Tempo de viagem mais longo;
- ix. Nível de escolaridade;
- x. Tempo de viagem caminhando (a pé) mais longo;
- xi. Número de moradias com uma só pessoa.

3.3 Os dados e a sua obtenção

Após uma análise das diferentes bases de dados territoriais, dos sistemas de transportes e mobilidade portuguesas foi possível verificar uma grande escassez de informação disponível para as 11 variáveis apresentadas no ponto 3.2. Assim, decidiu-se que a escala para a qual existia

informação com qualidade e reativa abundância era a dos municípios, já que descendo para o nível da freguesia não existiam dados de mobilidade. Por conseguinte, a obtenção dos dados para os municípios classificados como território de baixa densidade populacional ficou muito dependente dos dados presentes nos Censos realizados no ano 2011 pelo Instituto Nacional de Estatística (INE). Os censos são fontes de informação para a caracterização da população e dos territórios, sendo de muita importância no planeamento e intervenção social, económico, ambiental, entre outras.

É importante realçar que a base de dados do censo do 2011 teve uma importância significativa na escolha das variáveis para analisar, uma vez que é única fonte aberta que permite obter os dados de uma maneira livre e rápida.

3.3.1 Seleção da amostra

A seleção da amostra para definir o modelo de procura para os territórios de baixa densidade foi definida com base no total de municípios definidos por CIC Portugal 2020 (2015). Destes municípios foram selecionados 85% através do método “random”, uma vez que os 15% dos dados que não serão selecionados para a criação do modelo serão utilizados na validação do modelo. Os municípios são apresentados na tabela seguinte, é importante sublinhar que o ID do município é definido pelo INE (2011).

Tabela 5 – Municípios de amostra

ID_MUNICIPIO	Nome_Município	ID_MUNICIPIO	Nome_Município
1215	<i>Sousel</i>	1213	<i>Ponte de Sor</i>
1509	<i>Santiago do Cacém</i>	616	<i>Tábua</i>
1816	<i>São Pedro do Sul</i>	1707	<i>Murça</i>
1210	<i>Marvão</i>	1710	<i>Sabrosa</i>
407	<i>Mirandela</i>	1818	<i>Sernancelhe</i>
912	<i>Seia</i>	1610	<i>Vila Nova de Cerveira</i>
1806	<i>Mangualde</i>	902	<i>Almeida</i>
501	<i>Belmonte</i>	615	<i>Soure</i>
1007	<i>Castanheira de Pêra</i>	1713	<i>Vila Pouca de Aguiar</i>
606	<i>Góis</i>	804	<i>Castro Marim</i>
1706	<i>Montalegre</i>	1701	<i>Alijó</i>
710	<i>Redondo</i>	1802	<i>Carregal do Sal</i>
1814	<i>Santa Comba Dão</i>	706	<i>Montemor-o-Novo</i>
713	<i>Viana do Alentejo</i>	304	<i>Cabeceiras de Basto</i>

ID_MUNICIPIO	Nome_Municipio	ID_MUNICIPIO	Nome_Municipio
910	<i>Pinhel</i>	704	<i>Estremoz</i>
213	<i>Serpa</i>	408	<i>Mogadouro</i>
903	<i>Celorico da Beira</i>	1605	<i>Paredes de Coura</i>
1411	<i>Ferreira do Zêzere</i>	914	<i>Vila Nova de Foz Côa</i>
906	<i>Gouveia</i>	401	<i>Alfândega da Fé</i>
913	<i>Trancoso</i>	815	<i>Vila do Bispo</i>
1813	<i>Resende</i>	701	<i>Alandroal</i>
1409	<i>Coruche</i>	1815	<i>São João da Pesqueira</i>
1606	<i>Ponte da Barca</i>	1808	<i>Mortágua</i>
1817	<i>Sátão</i>	1810	<i>Oliveira de Frades</i>
1820	<i>Tarouca</i>	207	<i>Cuba</i>
803	<i>Aljezur</i>	709	<i>Portel</i>
601	<i>Arganil</i>	702	<i>Arraiolos</i>
711	<i>Reguengos de Monsaraz</i>	1809	<i>Nelas</i>
410	<i>Vila Flor</i>	214	<i>Vidigueira</i>
1302	<i>Baião</i>	1208	<i>Fronteira</i>
403	<i>Carrazeda de Ansiães</i>	210	<i>Moura</i>
208	<i>Ferreira do Alentejo</i>	310	<i>Terras de Bouro</i>
1505	<i>Grândola</i>	1002	<i>Alvaiázere</i>
209	<i>Mértola</i>	1203	<i>Avis</i>
901	<i>Aguiar da Beira</i>	1201	<i>Alter do Chão</i>
202	<i>Almodôvar</i>	404	<i>Freixo de Espada à Cinta</i>
1204	<i>Campo Maior</i>	1206	<i>Crato</i>
305	<i>Celorico de Basto</i>	1202	<i>Arronches</i>
1821	<i>Tondela</i>	707	<i>Mora</i>
1801	<i>Armamar</i>	1211	<i>Monforte</i>
504	<i>Fundão</i>	1811	<i>Penalva do Castelo</i>
1212	<i>Nisa</i>	412	<i>Vinhais</i>
508	<i>Proença-a-Nova</i>	411	<i>Vimioso</i>
1804	<i>Cinfães</i>	1003	<i>Ansião</i>
1413	<i>Mação</i>	1008	<i>Figueiró dos Vinhos</i>
1807	<i>Moimenta da Beira</i>	911	<i>Sabugal</i>
714	<i>Vila Viçosa</i>	703	<i>Borba</i>
611	<i>Oliveira do Hospital</i>	617	<i>Vila Nova de Poiares</i>
212	<i>Ourique</i>	905	<i>Fornos de Algodres</i>
509	<i>Sertã</i>	511	<i>Vila Velha de Ródão</i>
311	<i>Vieira do Minho</i>	1408	<i>Constância</i>
1803	<i>Castro Daire</i>	1711	<i>Santa Marta de Penaguião</i>
1209	<i>Gavião</i>	908	<i>Manteigas</i>
1013	<i>Pedrógão Grande</i>	506	<i>Oleiros</i>

ID_MUNICIPIO	Nome_Municipio
614	<i>Penela</i>
1501	<i>Alcácer do Sal</i>
607	<i>Lousã</i>

ID_MUNICIPIO	Nome_Municipio
712	<i>Vendas Novas</i>
1205	<i>Castelo de Vide</i>
1603	<i>Melgaço</i>

3.4 Metodologia para o desenvolvimento do Modelo de Regressão Linear para estimar a procura de passageiros de autocarro em TBD

3.4.1 Definição da escala das variáveis

Segundo Pestana & Gageiro (2014), é necessário identificar as escalas ou níveis de medida das variáveis para a escolha do tratamento estatístico adequado. Há quatro escalas ou níveis de variáveis: nominal, ordinal, intervalar e de rácio.

Numa escala nominal os elementos são atributos ou qualidades. Os números servem apenas para identificar ou categorizar os elementos, um exemplo deste tipo de variáveis pode ser a identificação das pessoas através do seu número de cartão de cidadão. Outras variáveis de escala nominal incluem, por exemplo, o sexo, a raça, a religião.

Numa escala ordinal apenas se podem distinguir diferentes graus de um atributo ou variável, existindo, portanto, entre eles uma relação de ordem. Os números são associados de modo que a relação de ordem se continue a manter, como por exemplo, os escalões de rendimento, as classes etárias, as ordens de preferências.

A escala intervalar utiliza números para classificar os elementos é feita de forma que seja a igual diferença entre os números corresponda a igual diferença nas quantidades do atributo medido. O zero é um valor arbitrário e não representa a ausência da característica medida.

3.4.2 Análise dos dados

A análise dos dados pode ser univariada, bivariada ou multivariada, consoante o número de variáveis tratadas em simultâneo. Na análise univariada cada variável é tratada isoladamente, e deve ser o primeiro passo na exploração dos dados. Na análise bivariada estabelecem-se relações entre duas variáveis. Na análise multivariada estabelecem-se relações entre mais de duas variáveis (Pestana & Gageiro, 2014).

3.4.3 Análise univariada dos dados

A estatística descritiva é o primeiro passo para a análise estatística e a caracterização da amostra em estudo. Este tipo de estatística estuda as características das unidades observadas ou experimentais, como por exemplo, pessoas, cidades, famílias, entre outras. Consiste na determinação de um conjunto de estatísticas de caracterização amostral para os diferentes tipos de variáveis existentes.

O processo selecionado (Ribeiro, 2011) para a descrição estatística das variáveis corresponde à realização de uma síntese exploratória, através da utilização do software SPSS, para cada variável utilizada no estudo, onde estão representados os seguintes elementos:

- a. Quadro com os resultados descritivos;
- b. Quadro com os valores extremos;
- c. Histograma e a respetiva curva da distribuição dos dados;
- d. Diagrama caule & folhas (complementar ao estudo do histograma),
- e. Resultados da aplicação dos testes de normalidade da variável.

3.4.3.1 Estudo da simetria e do achatamento

As medidas de assimetria e achatamento caracterizam o aspeto da distribuição em torno do valor médio e são caracterizadas pelos respetivos coeficientes de assimetria (g_1) e de achatamento (g_2). O seu conhecimento indica o grau de enviesamento e de achatamento da distribuição em estudo, em relação à distribuição da curva normal correspondente. Os coeficientes de assimetria (g_1) e achatamento (g_2) são dados por:

$$g_1 = \frac{N^2 m^3}{(N-1)(N-2)s^3}, \text{ com } m^3 = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{N} \quad (14)$$

$$g_2 = \frac{N^2(N+1)m^4}{(N-1)(N-2)(N-3)s^4} - 3 * \frac{(N-1)^2}{(N-2)(N-3)}, \text{ com} \quad (15)$$

$$m^4 = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{N}$$

$$g_2 = \frac{N^2(N+1)m^4}{(N-1)(N-2)(N-3)s^4} - 3 * \frac{(N-1)^2}{(N-2)(N-3)}, \text{ com} \quad (16)$$

$$m^4 = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{N}$$

- m^3 : terceiro momento em torno da média\
 m^4 : Quarto momento em torno da média
 s Desvio padrão
 N : Número total de observações
 X_i Valor da variável para a observação i
 \bar{X} : Valor médio da variável X

Segundo Maroco (2003), os resultados provenientes do SPSS em relação aos coeficientes de assimetria e achatamento devem apresentar valores próximos de zero, dentro do intervalo entre $[-0,5; 0,5]$, isto para que a distribuição seja considerada simétrica e mesocúrtica (característica de uma distribuição normal).

Pestana & Gageiro (2014) indicam que os resultados obtidos do programa SPSS devem ser afetados da estimação do erro amostral – *Std Error g_1* – para os respectivos coeficientes, de forma a considerar variação dos valores dos coeficientes de assimetria e achatamento, em amostras retiradas do mesmo universo. O procedimento para o coeficiente de achatamento é o mesmo.

$$Std Error g_1 = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (17)$$

O quociente a utilizar com os dados do SPSS, dever ser interpretados com a Tabela 6

Tabela 6 - Quociente de simetria utilizado pelo SPSS (Pestana & Gageiro, 2014)

$g_t / Std. error g_t$				
Assimétrica negativa	Não rejeitar a simetria		Assimétrica positiva	
< -1,96	-1,96	0	1,96	> 1,96

Ribeiro (2011) sublinha que este estudo permite identificar problemas relacionados com as variáveis que não sigam uma distribuição teórica normal, o qual permite identificar numa primeira fase se é possível assumir o valor médio das variáveis como representativo da variável, quando esta apresenta uma ligeira assimetria.

3.4.3.2 Estudo dos valores atípicos - “outliers”

O outlier, é denominado como um valor extremo, atípico ou aberrante. Os valores extremos atípicos podem representar casos que apresentam características particulares em relação à variável de estudo, ou resultar de erros cometidos durante a introdução dos dados.

Os outliers, são compostos por dois casos, os moderados e os severos. Os valores extremos aberrantes moderados são caracterizados por estarem situados abaixo dos valores do primeiro quartil, ou acima do terceiro quartil, a uma distância entre 1,5 e 3 amplitudes inter-quartil, e os severos a mais de 3 amplitudes inter-quartil (Pestana & Gageiro, 2014).

A verificação da existência de observações extremas aberrantes é realizada pelo SPSS, e pode ser observada recorrendo à análise dos diagramas de “caule & folhas” e de “caixa de bigodes”.

«*Diagrama de caule & folhas*; agrupa as observações em intervalos, representando o comprimento de cada linha as observações individuais e a sua frequência. Cada valor é dividido em duas componentes: o caule e folhas.»

«*Caixa de bigodes*; a caixa ou retângulo situa os quartis da distribuição, nos extremos de cada bigote posicionam-se as observações mínima e máxima. Todas as observações sitas para além dos bigodes chamam-se aberrantes ou outliers»

3.4.3.3 Normalidade

Uma distribuição normal é também designada por distribuição contínua de probabilidade ou função densidade de probabilidade, a equação que caracteriza-la é:

$$Y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \text{ com } -\infty \leq x \leq +\infty \quad (18)$$

m: Média

s: Desvio padrão

A área total definida pela equação anterior e o eixo dos X é unitária, deste modo, a área sob a curva, compreendida entre as duas coordenadas X=a e X=b, em que a<b representa a probabilidade de X estar situado entre a e b.

A função densidade de probabilidade da normal centrada e reduzida é dada por:

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ com } -\infty \leq x \leq +\infty \quad (19)$$

Para verificar que a variável seja segue a distribuição teórica normal, se faz o teste de Kolmogorov- Smirnov (K-S), com a correção de significância de Lilliefors, para amostras com dimensões superiores a 30 casos.

▪ Teste de Kolmogorov- Smirnov

O principal pressuposto do teste de Kolmogorov- Smirnov (K-S) é a consideração de uma distribuição populacional contínua, sendo a estatística do teste definida rigorosamente quando se conhecem a forma e os parâmetros da função de densidade de probabilidade.

O teste de K-S consiste na análise do ajuste entre a função de distribuição populacional teórica, F0(x), e a função de distribuição empírica, F (x). A estatística de teste, denominada D, corresponde ao supremo do valor absoluto da diferença F(x) e F0(x).

As hipóteses “nula” e “alternativa” são formuladas nos seguintes termos:

H0: F(x) = F0(x) para todos os valores da variável X (ou seja, a função de distribuição da população da qual provém a amostra é idêntica a uma função que se assume conhecida)

H1: $F(x) \neq F_0(x)$ para algum valor da variável X.

Uma vez especificado o nível de significância do teste, o valor D é comparado com o respectivo valor crítico e, em função do resultado, H_0 é ou não rejeitada. Note-se que a estatística D é calculada com base no módulo da diferença entre $F(x)$ e $F_0(x)$, não se distinguindo entre valores positivos ou negativos.

O teste de K-S é um teste exato, onde o risco (α) está rigorosamente definido, assim como a função teórico $F_0(x)$, através dos seus parâmetros estatísticos. Contudo, os parâmetros da função $F_0(x)$ são estimados e definidos com base nos valores de amostras. De modo a corrigir essa inexatidão, o SPSS utiliza uma correção designada por correção de Lilliefors.

A análise da normalidade das variáveis só fica completa, com a análise e verificação da existência de possíveis observações que se afastem da normalidade, através dos gráficos: Normal e Detrend Q-Q plots. No gráfico Normal Q-Q plot, os valores devem distribuir-se ao longo da linha reta oblíqua para a distribuição ser considerada normal. Para o gráfico Detrend Normal Q-Q plot, as observações devem encontrar-se distribuídas aleatoriamente para a distribuição ser considerada normal (Pestana & Gageiro, 2014). A seguir se apresenta um exemplo dos gráficos anteriormente ditos para a variável de desemprego.

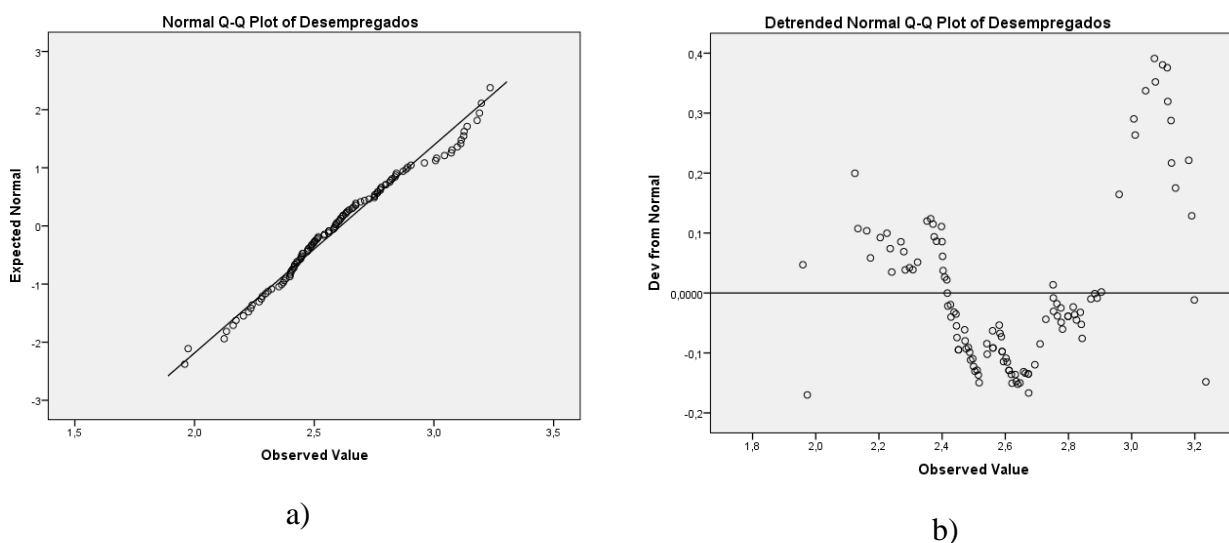


Figura 7 – Gráficos de verificação da normalidade para a variável desempregados: a)Normal Q-Q plots b)Detrended Normal Q-Q plot

▪ Teste de Shapiro-Wilk

Em relação ao domínio das pequenas amostras ($N < 30$), O SPSS produz o teste de Shapiro-Wilk (S-W), que pressupõe uma metodologia de cálculo diferente da usada no teste de K-S, mas para o qual análise dos resultados é semelhante.

As fases que compõe o teste de S-W são expostos a seguir com base (Pestana & Gageiro, 2005);

- i. Formulação de hipóteses a testar, que podem originar dois tipos de testes, os bilaterais e unilaterais, consistindo na seguinte formulação:

- Bilateral: $H_0: \mu=0$ e $H_1: \mu \neq 0$
- Unilateral $H_0: \mu=0$ e $H_1: \mu > (ou <) 0$

H_0 representa a hipótese nula e é considerada como a mais restritiva, devendo apenas rejeitada quando se tem um bom grau de fiabilidade do teste. H_1 é designada por hipótese alternativa, que complementa ou indica o sentido da hipótese, quando a hipótese é rejeitada

- ii. Cálculo da estatística de teste
- iii. Decisão: consiste em rejeitar a hipótese nula, se os valores de estatística de teste estiverem situados fora do intervalo definido pelo nível de significância do teste (α). No caso do teste bilateral, a rejeição corresponde a um intervalo, cujos valores críticos (VC) são definidos entre $(1 - \alpha/2)$ e $(1 + \alpha/2)$, para uma distribuição simétrica.

Em resumo a análise da significância do teste K-S realizado através do valor da probabilidade de significância p-value. Para não rejeitar a hipótese da distribuição ser normal, o valor da significância tem que ser superior a 0,05 ($\alpha=5\%$).

3.4.4 Transformação de variáveis

O estudo univariado anteriormente apresentado para todas as variáveis dos dados iniciais, permite conhecer os resultados dos testes de normalidade para as várias variáveis. De modo a resolver o problema derivado da falta de normalidade e consequente inadequada aplicação dos métodos estatísticos já referidos, optou-se por transformar matematicamente as variáveis de modo a torná-las normais.

Ribeiro (2011) realça que nem todas as variáveis podem ser transformadas de modo a ajustarem-se a uma distribuição normal e que nesses casos há duas opções, eliminar a variável ou mudar o modo de contabilizá-las.

O problema de falta de normalidade das variáveis é um problema clássico, da fase de tratamento de dados amostrais. Por quanto podem-se encontrar vários exemplos de possíveis transformações matemáticas das variáveis, com o objetivo de as normalizar e de homogeneizar as suas variâncias, tais como as apresentadas por Reis (2003), citado em Ribeiro (2011), as quais são baseadas sobretudo na observação da distribuição das variáveis dos dados originais (histogramas).

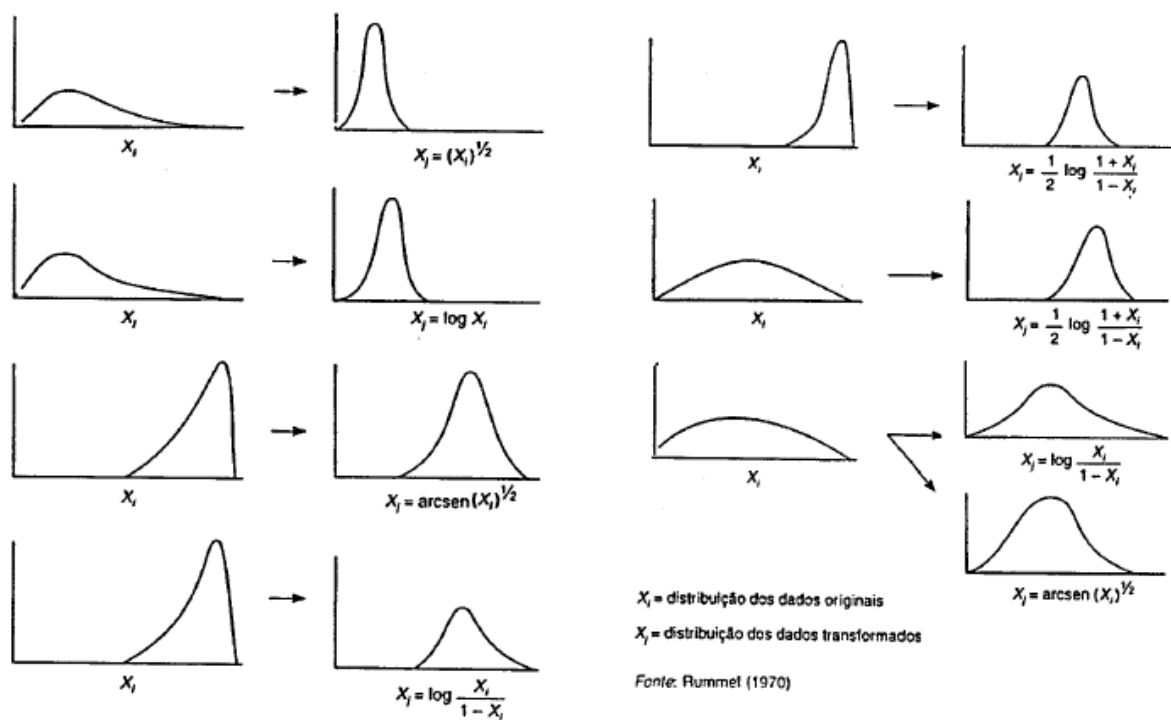


Figura 8 – Transformações das distribuições Fonte: (Rummet, 1970)

Os exemplos apresentados no esquema anterior, podem não ser suficientes para obter variáveis normais. Isto quer dizer que não existe uma regra ou método analítico, que permita introduzir os dados e obter uma variável normal; o que existe é um conjunto de regras baseadas na experiência sobre estes assuntos (Ribeiro, 2011).

3.4.5 Análise bivariada dos dados

Nesta secção se realiza o tratamento de dados e a seleção das variáveis que de acordo com a suas características estatísticas são consideradas influentes para o projeto a desenvolver. A escolha das variáveis definitivas será feita com base no estudo das relações entre a variáveis independentes (Xs) e dependente (ou, de resposta - Y).

Em princípio é importante ter em conta a escala das variáveis envolvidas no estudo das relações, assim é aplicada a medida, ou coeficiente estatístico apropriado. Deste modo, o SPSS dispõe de um conjunto de medidas que pode ser utilizado e que podem ser encontradas na tabela seguinte.

Tabela 7 – Medidas estatísticas para todos os tipos de variáveis (adaptado de Pestana & Gageiro, 2014)

Tipo	Qualitativa		Quantitativa
Escala	Nominal	Ordinal	Intervalar/ Rácio
Nominal	Chi-Quadrado Fisher Odds Ratio		
Ordinal	Chi-Quadrado Kolmogorov-Smirnov	Kappa of Cohen Ró de Spearman	
Intervalar/ Rácio	Eta	Eta	R de Pearson Ró de Spearman

Para fazer um análise de dados e variáveis é relevante definir o tipo de variáveis que foram determinadas no estado da arte, para poder ter certeza que medidas de estatísticas são necessárias para os estudos de definição da Regressão Linear. Deste modo, as variáveis escolhidas com base na literatura são de tipo intervalar/rácio (este tipo de variáveis foi exposto na anterior secção) e tipo quantitativo.

O primeiro nível é o estabelecimento e avaliações das relações entre as variáveis quantitativas, por conseguinte, a medida a usar é o R de Pearson para avaliar a intensidade de associação entre as variáveis quantitativas.

As medidas que serão utilizadas para medir a intensidade da relação entre as variáveis quantitativas, são expostas seguidamente (Guimarães & Cabral, 2010; Ribeiro, 2011; Pestana & Gageiro, 2014)):

Coefficiente de correlação linear R de Pearson

É um dos coeficientes mais utilizados para medir a associação linear entre duas variáveis. Porém, por ser um método paramétrico, pressupõe que os dados sejam provenientes de duas populações normais e que exista uma associação linear entre duas variáveis.

O R de Pearson é dado pela seguinte expressão:

$$R = \frac{\frac{\sum(X_1 - \bar{X})(Y_1 - \bar{Y})}{n-1}}{\sqrt{\frac{\sum(X_1 - \bar{X})^2}{n-1}} \sqrt{\frac{\sum(Y_1 - \bar{Y})^2}{n-1}}} = \frac{COV(X, Y)}{S_x S_y} \quad (20)$$

COV (X,Y) covariância entre as variáveis X e Y

\bar{X}, \bar{Y} média das variáveis X e Y

S_X, S_Y desvio padrão das variáveis X e Y

Uma vez este coeficiente de correlação deriva da divisão de uma covariância pelos desvios padrão das variáveis, é eliminado o problema de estabelecer a correlação entre variáveis com unidades de medida diferentes, pelo que se conclui ser uma medida de associação independente das escalas de medida.

Embora não seja muito comum referir-se o coeficiente R de Pearson, este tem um determinado significado geométrico e trigonométrico, podendo ser interpretado como o co-seno de um ângulo. Isto é, se os pontos estiverem alinhados na mesma direção, o R assume valor unitário, mas se as distribuições forem ortogonais, ou independentes, R assume o valor nulo.

Salienta-se ainda que o coeficiente de correlação R de Pearson corresponde ao coeficiente de correlação de uma linha reta a uma nuvem de pontos de duas variáveis X e Y, que represente bem a relação entre as duas variáveis, determinada através da aplicação do método dos mínimos quadrados. Deste modo, o valor de R de Pearson apresenta a qualidade de ajuste de uma reta de regressão para definir a relação entre essas variáveis, e geralmente é adequado para coeficientes

de correlação elevados. O R de Pearson assume valores no intervalo de -1 a 1, onde o sinal indica o sentido da tendência da regressão linear simples entre as variáveis.

No entanto, este coeficiente não permite comparações diretas do grau de associação entre as relações encontradas, devendo para isso ser utilizado o coeficiente de determinação R^2 , como os apresentados por (Pestana & Gageiro, 2014), os quais uma correlação forte obriga a que pelo menos 50% da variação de uma variável seja explicada pela outra.

Tabela 8 – Intensidade de correlação através do valor do R de Pearson

R de Pearson		Designação verbal da intensidade de correlação
(Pestana & Gageiro, 2003)		
0,0	0,19	Muito fraca
0,2	0,39	Fraca
0,4	0,69	Moderada
0,7	0,89	Forte
0,9	1,0	Muito forte

Coefficiente Ró de Spearman

Este coeficiente é não paramétrico, logo não exige nenhum pressuposto sob a forma da distribuição das variáveis, que as variáveis sejam normais. Substituindo na expressão do R de Pearson, os valores das observações de X e Y pelas respectivas ordens r_x e r_y , obtém-se a expressão do Ró de Spearman:

$$Ró_{Spearman} = \frac{\sum_{i=1}^n (r_{X_1} - r_{\bar{X}})(r_{Y_1} - r_{\bar{Y}})}{\sqrt{(\sum_{i=1}^n (r_{X_1} - r_{\bar{X}})^2 \sum_{i=1}^n (r_{Y_1} - r_{\bar{Y}})^2)} \quad (21)$$

No entanto, este coeficiente está sobretudo vocacionado para medir a intensidade de relação entre variáveis ordinais. Como não é sensível a valores extremos aberrantes, nem a assimetrias da distribuição e não exige que as variáveis sejam normais, é utilizado frequentemente como uma excelente alternativa o R de Pearson, sendo que a sua interpretação é análoga a este último.

3.4.6 Modelo de Regressão Linear Múltipla

3.4.6.1 Objetivos

O objetivo principal é explicar e prever a evolução de fenómenos, com base num conjunto de hipóteses, que expressam relações lógicas entre as grandezas que dele fazem parte. Do mesmo modo é importante sublinhar que o modelo não é mais do que uma aproximação aos aspetos essenciais da realidade em estudo, pelo que deve responder às questões para que foi construído.

Para cumprir o objetivo de previsão é importante que o modelo tenha flexibilidade, ou seja, é possível admitir a existência de desvios entre os valores observados e os propostos sem que sejam postos em causa. Uma forma de tornar os modelos flexíveis, consiste em incluir variáveis aleatórias, ditas residuais, que representam a influência dos efeitos que não foram explicitamente considerados. Deste modo, as variáveis que se pretendem explicar, tornam-se também aleatórias visto dependerem igualmente das variáveis residuais (Pestana & Gageiro, 2005).

3.4.7 A formulação geral do modelo

Quando existem várias variáveis exógenas o modelo designa-se por Modelo de Regressão Linear Múltipla (MRLM), e é representado genericamente por :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_j X_{ji} + \dots + \beta_k X_{ki} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (22)$$

n: dimensão da amostra;

k: Número de variáveis exógenas observáveis adicionadas da constante, onde X's e Y são variáveis observáveis;

ε_i : Variável exógena não observável e aleatória, a qual inclui todas as influencias em Y que não são explicados por X's;

β_k : são parâmetros do modelo, isto é, grandezas que nele assumem sempre o mesmo valor.

Não é possível encontrar uma função que traduza exatamente as relações entre as variáveis representativas, porque o grande número de influências sobre a variável dependente obriga a selecionar as variáveis mais relevantes em detrimento de outras que pouco contribuam para a explicação de Y . Estas últimas variáveis são agrupadas numa única variável designada por residuo ε_i , com comportamento aleatório.

3.4.8 Hipóteses do modelo

I. Forma e caracterização das variáveis

O modelo de regressão linear múltipla (MRLM) analisa a relação entre duas ou mais variáveis exógenas, com uma variável endógena quantitativa. Este tipo de variáveis são introduzidas no modelo através da criação de variáveis artificiais, também conhecidas como variáveis dummies, possibilitam a inclusão sob a forma aditiva, de variáveis qualitativas exógenas no modelo de regressão, assinalando a presença ou ausência de determinada característica»

No MRLM, por existirem mais do que duas variáveis exógenas, podem ocorrer problemas de elevada multicolinearidade, tornando necessário adoptar certos cuidados para a correta interpretação dos coeficientes do modelo.

II. Valor esperado nulo da variável residual

Para cada observação, o valor esperado de cada variável aleatória é zero, $E(\varepsilon_i) = 0$, para $i = 1, 2, \dots, n$, o que significa que, em média as diferenças positivas entre o valor observado e o valor assumido pela parte principal do modelo se compensam com as diferenças negativas, de modo que as nuvens dos pontos potencialmente observáveis se distribuem de um modo equilibrado de um e do outro lado do hiperplano que representa a parte principal do modelo.

III. Homocedasticidade

A variância de cada variável aleatória é finita e sempre a mesma para qualquer observação (homoscedasticidade), $Var(\varepsilon_i) = \sigma^2$, para $i = 1, 2, \dots, n$, o que significa que a dispersão das observações à volta da reta de regressão é constante.

Assim, não existem observações para as quais se considere uma influência mais intensa das condicionantes incluídas nessa variável aleatória que determinem uma maior flutuação dos possíveis valores observáveis na variável endógena.

A hipótese da homoscedasticidade significa que os elementos da diagonal da matriz $Var(E)$ são iguais podendo, portanto, ser representados por um único símbolo σ^2 .

IV. **Ausência de autocorrelação**

As variáveis aleatórias referentes a duas observações diferentes não estão correlacionadas entre si, sendo a sua covariância zero, isto é

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0, \text{ para } i \neq j, \text{ com } i, j = 1, 2, \dots, n$$

Assim, mesmo que se pudesse conhecer o valor ou sinal assumido por um lado ε_i isso não acrescentaria nenhuma informação sobre os valores ou sinais assumir pelas outras variáveis aleatórias.

V. **Os ε_i 's são independentes e identicamente distribuídos com distribuição normal N**

Admite-se que os resíduos são independentes e estão identicamente distribuídos tendo uma distribuição normal com média zero e variância constante, $N(0; \sigma^2)$.

Esta hipótese permite determinar a distribuição dos estimadores, dos testes estatísticos (t, F) e calcular os intervalos e regiões de confiança.

VI. **Não existe perfeita multicolinearidade**

Admite-se que as variáveis exógenas não estão perfeitamente correlacionadas umas com as outras. O número de observações n tem de ser superior ao número de variáveis exógenas adicionadas do termo independente, que se representam por k , que por sua vez tem de ser igual à característica da matriz X , representada por $r(X)$.

O modelo de regressão é uma das técnicas estatísticas mais populares devido ao seu elevado poder explicativo e à facilidade da sua interpretação e de utilização em programas informáticos, nomeadamente no SPSS.

As variáveis incluídas no modelo podem ser exógenas ou endógenas. As exógenas são determinadas por causas externas ao modelo e nele são introduzidas de forma a contribuir para a explicação das variáveis endógenas, ou seja, aquelas que o modelo procura interpretar.

3.4.9 Estimação do modelo

Nesta seção apresentam-se os parâmetros/ variáveis a utilizar no modelo com base na informação obtida da amostra, i.e., determina-se a equação do modelo¹ a qual estará conformada de maneira geral por:

$$VGA = \hat{B}_0 + \hat{B}_1(Dsgd) + \hat{B}_2PdC + \hat{B}_3DP + \hat{B}_4IM + \hat{B}_5GM + \hat{B}_6SnE + \hat{B}_7ES + \hat{B}_8IP \quad (23)$$

VGA: Número de viagens em autocarro;

Dsgd: Número de pessoas desempregadas;

PdC: Poder de compra;

DP: Densidade Populacional;

IM: Número de mulheres maiores a 15 anos;

IP: Número de pessoas entre os 15 e 24 anos e maiores a 50 anos;

GM: Ganho médio mensal;

SnE: Número de pessoas analfabetas;

ES: Número de pessoas com educação superior.

Um modelo com muitas variáveis explicativas será mais difícil de interpretar já que os efeitos de algumas das variáveis serão pequenos, por isso será melhor a determinação de um modelo com menos variáveis, com efeitos mais fortes e que tenha um erro preditivo menor do que o modelo com todas as variáveis.

A estimação do modelo é feita no software SPSS, este permite escolher um método de seleção das variáveis com o qual se poda chegar ao “melhor modelo”. Por tanto e com base em Pestana & Gageiro (2005), o método escolhido é o stepwise, o qual permite resolver problemas de multicolinearidade. Consiste em entrar a variável explicativa que apresenta maior coeficiente de correlação com a variável independente. Em seguida, calculam-se os coeficientes de correlação parcial para todas as variáveis que não fazem parte da primeira equação de regressão, para que, a próxima variável a entrar, seja a que apresenta maior coeficiente de correlação parcial. Estima-se a nova equação de regressão e analisa-se se uma das duas variáveis independentes deve ser excluída do modelo. No final, se ambas variáveis apresentam valores *t* significativos, novos coeficientes de correlação são calculados para as variáveis que não entraram. O processo finda, quando chegue à situação em que nenhuma variável deva ser acrescentada à equação (Rodrigues, 2012).

¹ As variáveis apresentadas na equação podem variar dependendo do modelo

3.4.10 Diagnóstico do modelo estimado

Nesta secção se analisará separadamente as hipóteses da homocedasticidade, da independência e da normalidade das variáveis aleatórias residuais, bem como a hipótese de multicolinearidade entre as variáveis independentes

3.4.10.1 Hipóteses de Homocedasticidade (variância constante)

A condição de homocedasticidade verifica o pressuposto que os erros devem ter variância constante. A variância ser constante equivale a supor que não existe observações incluídas na variável residual cuja influência seja mais intensa na variável dependente.

Segundo Pestana & Gageiro (2014) e Rodrigues (2012) há duas técnicas para verificar para verificar que os resíduos são homoscedásticos; a primeira é com o gráfico dos resíduos estandardizados ($Y=Z_{resid}$) com os valores estimados de Y ($X=Pre_1$), na figura 9 (a), se apresenta um exemplo da gráfica. O gráfico tem que mostrar que os resíduos mantem uma amplitude aproximadamente constante em relação ao eixo horizontal zero.

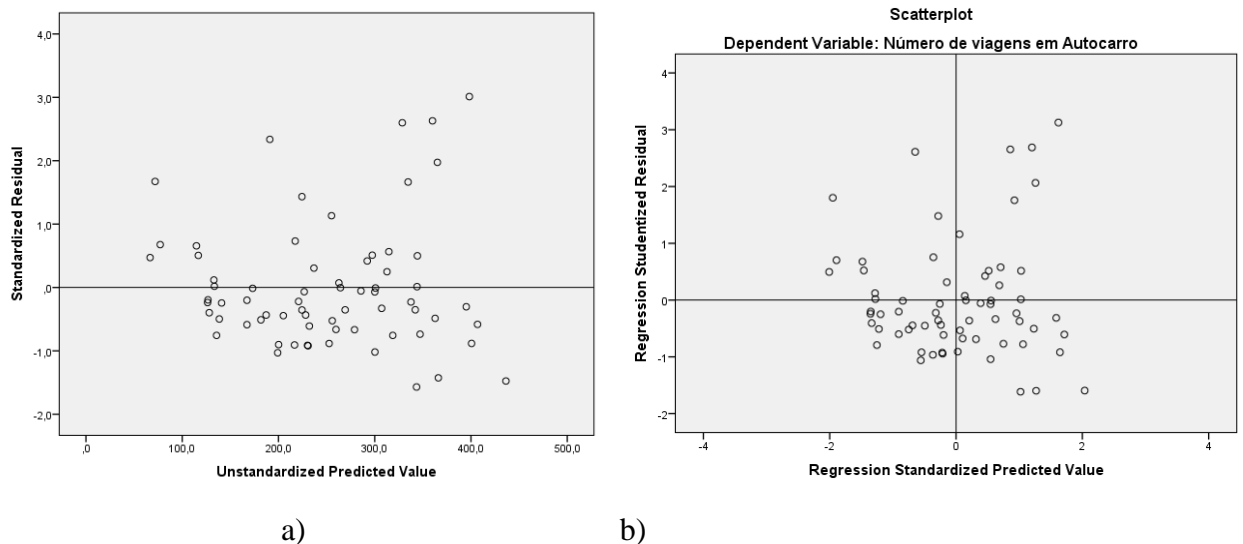


Figura 9 – a) análise da variância com $Y=Z_{re}$ e $X=Pre$ b) Análise da variância com $Y=S_{resid}$ e $X=Z_{pred}$

A segunda técnica consiste em analisar o gráfico que relaciona os resíduos estudentizados (S_{resid}) com os resíduos estandardizados (Z_{pred}), tendo em conta que os resíduos

estudentizados são os resíduos divididos pelo seu desvio padrão, que variam de ponto para ponto de acordo com a distância de cada observação X_i à média, refletindo com maior precisão a eventual violação da homocedasticidade.

3.4.10.2 Diagnóstico de independência

A análise da covariância nula ou da não autocorrelação (ρ) dos resíduos, pode ser feita através do testes de Durbin – Watson (DW), cujas hipóteses são:

H_0 : $\rho=0$, onde ρ é a autocorrelação dos resíduos

H_a : $\rho \neq 0$

A interpretação do teste (DW) é apresentada por Pestana & Gageiro (2014) e corresponde:

- Para valores próximos de 2, não existe autocorrelação dos resíduos;
- Para valores próximos de 0 significa uma autocorrelação positiva;
- Para valores próximos de 4 existe uma autocorrelação negativa.

Para verificar se o valor do teste está na zona de não rejeição da hipóteses nula, se compara o valor com os valores críticos d_L e d_U da tabela de Durbin- Watson para um nível de significância de 0,05 e para o número de variáveis do modelo (K) com o número de dados da amostra (n). Seguidamente se apresenta a tabela de decisão em função de d_L e d_U :

Tabela 9 – Tabela de decisão em função de d_L e d_U Fonte: (Rodrigues, 2012)

Zona de Rejeição e de não-rejeição de H_0					
dw	$[0; d_L[$	$[d_L; d_U[$	$[d_U; 4 - d_U[$	$[4 - d_U; 4 - d_L[$	$[4 - d_L; 4]$
Decisão	Rejeitar H_0	Nada se pode concluir	Não rejeitar H_0	Nada se pode concluir	Rejeitar H_0
	Auto-correlação positiva		Os resíduos são independentes		Auto-correlação negativa

3.4.10.3 Diagnóstico de normalidade

Nesta fase analisa-se se os resíduos tem um comportamento normal. As técnicas usadas para determinar o cumprimento desse pressuposto são:

- Teste de Kolmogorov- Smirnov;
- Teste de Shapiro- Wilk;

- Gráfico normal Q-Q plot;
- Gráfico DetreDED Normal Q-Q plot.

Na seção 3.2.1.3 já se abordaram as características associadas a estes elementos de avaliação da normalidade de uma variável.

3.4.10.4 Diagnóstico da multicolinearidade

O termo de multicolinearidade é utilizado para expressar a correlação fortemente entre mais de duas variáveis independentes. Rodrigues (2012) define a multicolinearidade como a existência de relação linear entre uma das variáveis independentes e as restantes. O termo de colinearidade é utilizado para expressar a existência de correlação elevada entre duas variáveis independentes.

O método de estimação *Stepwise* permite detetar a multicolinearidade, pela observação de modificações significativas nos coeficientes estimados. De facto, se as variáveis forem independentes cada coeficiente obtém-se à custa das observações de apenas uma variável, pelo que se outra variável for adicionada ou retirada do modelo não se verificam modificações nas estimativas dos restantes parâmetros do modelo. Quando tal não acontece aumenta o grau de multicolinearidade e maiores são as modificações nas estimativas dos coeficientes em consequência das alterações das variáveis introduzidas no modelo (Pestana & Gageiro, 2015)

A intensidade da multicolinearidade é analisada essencialmente através dos pontos seguintes:

- i. Correlação entre as variáveis independentes: recorre-se à matriz das correlações. Quando os coeficientes de correlação entre as variáveis independentes são elevados (superiores em termos absolutos a 0,9) indicam a possibilidade de existência da multicolinearidade.
- ii. Tolerância e VIF: a tolerância mede o grau em que uma variável X é explicada por todas as outras variáveis independentes. Por tanto a tolerância da variável X_a mede a proporção da sua variação que não é explicada pelas restantes variáveis independentes, esta tolerância varia entre zero e um, em quanto mais próxima estiver de zero, maior será a multicolinearidade e o modo contrario entre mais próxima estiver de um menor será a multicolinearidade.

O inverso da tolerância designa-se por VIF (Variance Inflation Factor), quanto mais próxima de zero estiver VIF, menor será a multicolinearidade. O valor habitualmente considerado como o limite acima do qual existe multicolinearidade é 5 (Pestana & Gageiro,2014).

3.4.11 Análises de Outliers e observações influentes

Nesta seção, serão identificados os outliers entre as variáveis independentes (X 's) bem como os casos (X_i, Y_i) que exercem influência nos coeficientes estimados da regressão.

Como já se tinha dito na seção 3.2.1.2, os outliers são observações extremas que segundo Rodrigues (2012), se encontram de tal forma afastadas da maioria dos dados que surgem dúvidas sobre se estas poderão ou não ter sido geradas pelo modelo proposto para explicar essa maioria dos dados.

As observações aberrantes não são detetadas exclusivamente pela análise dos resíduos, sendo necessário utilizar outras estatísticas para esse efeito, nomeadamente (Pestana & Gageiro, 2014):

- I. Estatísticas usadas para a identificação dos outliers entre as variáveis independentes:
 - Resíduos estandardizados;
 - Resíduos estudantizados;
 - Leverage;
 - Valores estimados ajustados.

- II. Estatísticas usadas para identificar os casos influentes:
 - Resíduos estudantizados eliminados;
 - A distância de Cook;
 - DfBetas estandardizados;
 - DfFit estandardizados;
 - Rácio de covariância.

3.5 Validação do modelo

A validação do modelo de estimação da procura de transporte para os territórios de baixa densidade, visa verificar que o modelo cumpra com a sua função, em outras palavras, se o modelo suministra uns resultados razoáveis e aceitáveis do comportamento da procura cumpre com o objetivo. Para realizar esta verificação foram escolhidos o 15% dos municípios totais (equivale a 23 dados), a seleção dos municípios será feito aleatoriamente entre os 151 municípios.

É importante realçar que fazer a verificação do diagnóstico do modelo estimado (apresentado na seção 3.4.9) e corroborar cumprimento de cada um deles para um intervalo de confiança do 5%, faz parte da avaliação do modelo. Por conseguinte a validação do modelo tem dois procedimentos igualmente importantes.

Para cada um dos municípios vai-se calcular o número estimado de viagens em autocarro (Yestimado) usando a equação do modelo de RLM, para depois com base no valor real do número de viagens em autocarro, determinar o percentagem de error para cada estimação. Com base neste cálculo vai realizar-se a frequência dos erros para os dados de validação, os quais permitem definir se o modelo é relevante para a estimação das viagens de autocarro nos territórios de baixa densidade.

Seguidamente se apresenta na **Error! Reference source not found.** os municípios escolhidos para fazer o procedimento de validação com a análise dos erros:

Tabela 10 – Municípios para validação

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada
0904	<i>Figueira de Castelo Rodrigo</i>	6260	246	1598	12	66,35	243
0405	<i>Macedo de Cavaleiros</i>	15776	592	3503	23	74,67	576
0117	<i>Sever do Vouga</i>	12356	531	2260	95	74,08	591
0809	<i>Monchique</i>	6045	238	1490	15	61,26	266
1417	<i>Sardoal</i>	3939	160	752	43	73,92	138
0409	<i>Torre de Moncorvo</i>	8572	321	2119	16	63,83	369
0406	<i>Miranda do Douro</i>	7482	269	1778	15	75,79	225
1712	<i>Valpaços</i>	16882	758	4550	31	60,13	893

0909	<i>Meda</i>	5202	202	1348	18	62,1	244
1702	<i>Boticas</i>	5750	422	1659	18	59,91	315
1407	<i>Chamusca</i>	10120	474	2384	14	73	350
0104	<i>Arouca</i>	22359	1403	4774	68	69,49	997
0505	<i>Idanha-a-Nova</i>	9716	375	3077	7	68,61	484
0510	<i>Vila de Rei</i>	3452	110	916	18	65,39	142
1705	<i>Mondim de Basto</i>	7493	656	1939	44	58,76	460
1708	<i>Peso da Régua</i>	17131	808	3645	181	84,52	1065
1709	<i>Ribeira de Pena</i>	6544	325	1971	30	57,33	430
0507	<i>Penamacor</i>	5682	222	1705	10	59,94	297
1822	<i>Vila Nova de Paiva</i>	5176	211	1347	30	61,25	287
0201	<i>Aljustrel</i>	9257	422	2024	20	88,92	201
1819	<i>Tabuaço</i>	6350	267	1474	47	55,85	406
0612	<i>Pampilhosa da Serra</i>	4481	111	1417	11	65,3	213
0206	<i>Castro Verde</i>	7276	391	1578	13	102,41	5

4 CASO DE ESTUDO

É importante sublinhar que uma das definições mais completas de territórios de baixa densidade é apresentada por Ramos & Bento (2012), indicando que são territórios marcadamente rurais com índices elevados de população envelhecida e um povoamento escasso e disperso. Para além disso os TBDs apresentam um decréscimo nos níveis de acessibilidade a serviços e bens públicos. Em relação ao sistema económico, os TBDs têm uma tendência a ser pouco dinâmico e dependente da agricultura.

Ao longo do tempo, as pessoas residentes nos TBD's têm sentido um crescente aumento na dificuldade de acessibilidade e mobilidade da população residente. Este problema pode estar associada à privatização do setor do transporte e da mobilidade e, ou à extinção da oferta de serviços de transporte público. Outros fatores apontados como tendo influência na degradação da mobilidade nos TBD's é a concentração da oferta de equipamentos, bens e serviços nos aglomerados urbanos mais importantes e, ainda, devido aos baixos níveis de rendimento (poder de compra baixo) das populações residentes.

Não há uma metodologia clara para classificar e determinar se um território é de baixa densidade, já que esta determinação depende das políticas internas de cada país ou região. No caso específico de Portugal a metodologia elaborada por Ramos & Bento (2012), com base em parâmetros associados ao perfil demográfico, de povoamento, socioeconómico e de acessibilidade, além da densidade populacional, com um peso na definição de um TBD em 50%, indica que 165 municípios podem ser considerados de baixa densidade, tal como é apresentado em CIC Portugal 2020 (2015). Este valor equivale a 59% dos municípios totais no continente. Estes municípios encontram-se localizados e distribuídos em maior quantidade no interior do país.

Para definir as variáveis que podem explicar o comportamento de mobilidade em autocarro (principal e, na maioria dos municípios, o único transporte público de massas) dos TBD em Portugal, foi necessário elaborar um levantamento sobre a bibliografia existente sobre os TBDs, os seus principais problemas e, sobretudo, sobre os modelos de procura de transporte públicos convencionais mais usados, com o objetivo de analisar e comparar as variáveis encontradas e definir para o caso do Portugal, aquelas que podiam ser mais importantes e influentes. As variáveis escolhidas são apresentadas na seção 3.2 deste trabalho. É importante realçar que

algumas variáveis que são consideradas, no estado da arte, muito relevantes para definir um modelo de estimativa da procura em autocarro para TBDs não podem ser usadas no caso português, porque na maior parte dos casos a informação necessária (dados) não pode ser facilmente recolhida, sobretudo em bases de dados abertas, nomeadamente informação gerida por instituições de mobilidade e transportes, quer a nível nacional, quer a nível local, como por exemplo a informação sobre a propriedade e registo automóvel e o número de pessoas com carta de condução por município.

Por outro lado, também foram analisados os modelos/ métodos para estimar a procura de passageiros em autocarro na área do planeamento de transportes, tendo-se verificado que o mais usado ao nível mundial com bons resultados na prática é o modelo matemático de regressão linear múltipla (Ortuzar, 2015), sobretudo quando a referência é o modelo de 4 passos, mais especificamente nas fases 7 etapas de geração e atração de viagens, distribuição de viagens, repartição modal.

Assim, para a elaboração do modelo de geração de viagens em autocarro, no âmbito desta dissertação, foi usado o modelo matemático de regressão linear múltipla, já que este permite analisar a relação e, sobretudo, estimar o número de viagens em autocarro com base num conjunto de variáveis independentes selecionadas, segundo o estado da arte, como influentes: Além disso, é o modelo mais usado para determinar a primeira etapa do modelo de quatro passos como foi apresentado na seção 2.2.2.1.

Importa destacar que no âmbito desta dissertação será elaborado uma proposta de modelo de geração de viagens em autocarro seria importante em futuros trabalhos comparar os resultados obtidos com os da aplicação das três primeiras fases do modelo clássico, geração de viagens de autocarro num município e respetiva distribuição entre vários municípios, ou outro tipo de territórios, e por fim com a repartição entre modos, com particular incidência no transporte por autocarro nas fases de planeamento para os territórios de baixa densidade.

Deste modo, neste capítulo será apresentado o processo de elaboração do modelo de regressão linear múltipla (MRLM) que permita estimar o número de viagens em autocarro para TBDs em Portugal, através de uma análise estatística aprofundada em termos univariados, bivariados e geração e respetiva calibração e validação do MRLM.

4.1 Análise dos dados e variáveis para o MRLM

4.1.1 Análise Univariada

A seguir se apresenta um exemplo de uma estrutura de análise “folha” tipo de uma das variáveis usadas nas análises estatísticas deste trabalho, nomeadamente “o número de pessoas sem trabalho – desempregados”. Este procedimento foi realizado para todas as potenciais variáveis que foram selecionadas para pertencerem ao MRLM. As estatísticas apresentadas foram obtidas com o recurso a um programa de análise estatística de dados, cujo nome é *Statiscal Package for the Social Sciences* – SPSS versão (IBM SPSS Statistics 24).

Na Figura seguinte apresenta-se o exemplo da folha tipo com um vasto conjunto de informação, mais propriamente:

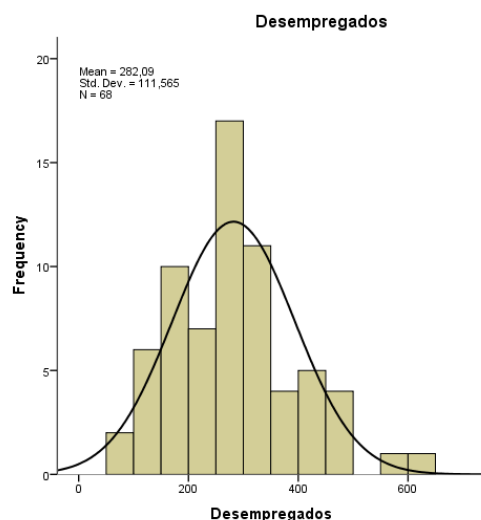
- quadro “estatística”: observam-se as medidas de tendência central, pretendendo-se caracterizar o valor da variável que acontece com mais frequência, das quais as mais vulgares e que foram utilizadas, são a média aritmética, a mediana e moda. As medidas de dispersão utilizadas foram a variância, o desvio padrão, o coeficiente de variação (razão entre o desvio padrão e a média), o intervalo de variação (a diferença entre o valor máximo e mínimo) e a amplitude inter-quartis (engloba mais de 50% dos dados da amostra, que corresponde à diferença entre o valor do 3º e 1º quartil).
- no quadro “valores extremos”: são apresentados 5 valores mais altos e mais baixos da variável analisada;
- os diagramas: histograma e caule e folhas apresentam como é realizada a distribuição da frequência dos dados e permite visualizar em princípio o ajuste da variável a uma distribuição normal. Caule e folhas permite complementar a análise da frequência dos dados, proporcionando ainda o levantamento do número de observações para os valores extremos.
- o resultados dos testes de Normalidade da variável: K-S e S-W, permite avaliar se os dados da variável apresentam ou não um comportamento similar ao de uma distribuição normal.

Tabela 11 – Folha tipo análises de variáveis

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent [%]	N	Percent [%]	N	Percent [%]
Desempregados	68	100	0	0	68	100

Estatística	
Média	282.09
Mediana	269.50
Moda	174a
Desvio padrão	111.565
Variância	12446.679
Enviesamento (assimetria)	0.638
Std. Error do enviesamento	0.291
Curtose (achatamento)	0.370
Std. Error da Curtose	0.574
Amplitude Total	19182
Mínimo	91
Máximo	604

Valores extremos				
		ID		Valor
		Desempregados	Mais altos	1
2	78			582
3	52			494
4	53			472
5	56			461
Mais baixos	1		9	91
	2		34	94
	3		7	102
	4		19	130
	5		4	133



Frequência Caule & Folhas

```

2.00      0 . 99
6.00      1 . 033344
10.00     1 . 6677788999
7.00      2 . 0012334
17.00     2 . 5555666677889999
11.00     3 . 00111222344
4.00      3 . 6889
5.00      4 . 00023
4.00      4 . 5679
2.00 Extremes (>=582)
    
```

Largura do caule: 100
Cada Folha: 1 case(s)

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Desempregados	0.072	68	,200 [*]	0.969	68	0.090

Por último, importa referir que a análise univariada das outras variáveis consideradas para o modelo MRLM são apresentadas no Anexo A.

4.1.2 Análise dos valores extremos

Os valores extremos, nomeadamente os que se designam por “outliers”, são valores da variáveis que se encontram geralmente muito longe dos valores médios. Para alguns autores, quando uma observação tem um valor igual ou superior a 3 desvios-padrão do valor médio, esta pode ser considerada um outlier. É importante saber se os outliers estão, ou não, a ser contabilizados no modelo, uma vez que estes valores podem ter um peso significativo nos resultados do estudo global, dependendo do seu valor (Ribeiro, 2011).

Para fazer estas análises usa-se o gráfico chamado caixa de bigodes (“box plot”) o qual situa os quartis da distribuição. Nos extremos de cada bigode posicionam-se as observações mínimas e máximas. Os valores ou observações fora de esta caixa chama-se aberrantes ou outliers.

Na Figura 10 apresenta-se o exemplo para a variável de desemprego, onde se pode observar uma quantidade considerável de dados que são considerados outliers. Para ter uma melhor ideia das análises é necessário classificar as observações aberrantes em severas (identificadas com um asterisco - *) ou moderadas (identificadas com um círculo - °) dependendo de sua frequência e afastamento às outras observações.

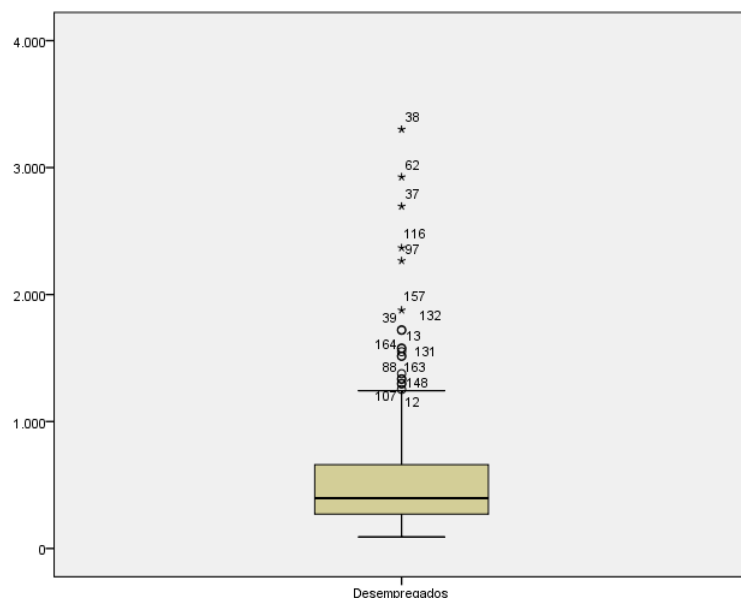


Figura 10 – Caixa de bigodes para a variável de desemprego

Para decidir as observações (municípios) são aberrantes (outliers) na base de dados de territórios de baixa densidade, foi necessário fazer uma análise prévia para cada variável e identificar os valores dos outliers, que são considerados severos, para assim determinar a frequência com a qual estes valores (municípios) se repetem em todas as variáveis em estudo para o MRLM. Esta análise foi feita em duas ocasiões, porque as variáveis continuavam apresentar dados aberrantes nas caixas de bigodes, o que afeta o comportamento da variável para o modelo matemático a analisar. O resultado dessas análises apresenta-se na Tabela 12.

Tabela 12 – Dados dos municípios considerados valores aberrantes “Outliers”

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada
0705	<i>Évora</i>	56596	1743	10227	43	116,39	1619
0502	<i>Castelo Branco</i>	56109	2119	10845	39	97,4	1843
0802	<i>Alcoutim</i>	2917	88	899	5	68,53	75
0307	<i>Fafe</i>	50633	2905	9984	231	74,75	2473
0205	<i>Beja</i>	35854	1428	7412	31	107,06	1118
1812	<i>Penedono</i>	2952	118	718	22	60,71	149
0907	<i>Guarda</i>	42541	1888	7701	60	96,25	1337
0313	<i>Vila Verde</i>	47888	3788	10939	209	66,45	2632
0204	<i>Barrancos</i>	1834	21	416	11	64,96	28
1714	<i>Vila Real</i>	51850	3167	9542	137	100,81	1906
0503	<i>Covilhã</i>	51797	3382	9910	93	87,84	1913
0402	<i>Bragança</i>	35341	751	6933	30	97,97	1085
0203	<i>Alvito</i>	2504	83	619	10	68,01	41
0708	<i>Mourão</i>	2663	110	717	10	70,38	44

4.1.3 Seleção das variáveis significativas (análise bivariada)

De acordo com as variáveis que melhor explicam o comportamento da procura em autocarro devidamente selecionadas no estado da arte e com a respetiva tipologia estatística, i.e., intervalar/rácio de caráter quantitativo, foram, então, utilizadas as medidas estatísticas o R de Pearson (contínua) e o Ró de Spearman (ordinal), para determinar a intensidade de associação entre todas as variáveis selecionadas que poderiam integrar o MRLM, ou seja, uma análise de pares de correlações lineares que se designou por análise bivariada.

Antes de analisar a relações entre as variáveis é necessário atender às hipóteses subjacentes à medida estatística do R de Pearson, a qual assume que as variáveis têm uma distribuição normal. Por tanto, para verificar esta hipótese é preciso aplicar o teste de normalidade de Kolmogorov-Smirnov, uma vez que se trata de uma amostra com mais de 50 casos, cujo resultado se apresenta na Tabela da Figura 12.

Tabela 13 – Testes de Normalidade para as variáveis em análise

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Densidade	,160	151	,000	,822	151	,000
Edade_menor21	,178	151	,000	,852	151	,000
Edade_mayor65	,132	151	,000	,845	151	,000
Edade_entre_15e24	,168	151	,000	,853	151	,000
Edade_mayor50	,138	151	,000	,839	151	,000
Idade_15e24_mayor50	,146	151	,000	,844	151	,000
Desempregados	,193	151	,000	,785	151	,000
Edade_mulheremayor15	,159	151	,000	,843	151	,000
Sem_nivel_de_escolaridade_completo	,164	151	,000	,867	151	,000
Ensino_básico	,154	151	,000	,854	151	,000
Ensino_secundario	,164	151	,000	,791	151	,000
Ensino_póssecundario	,149	151	,000	,822	151	,000
Ensino_superior	,191	151	,000	,713	151	,000
Ganho_medio_mensual	,149	151	,000	,774	151	,000
Poder_de_compra	,111	151	,000	,931	151	,000
Movimentos_pend_autocarro	,169	151	,000	,843	151	,000

a. Lilliefors Significance Correction

Da análise dos resultados apresentados na Tabela 13 é possível concluir que todas as variáveis analisadas não cumprem com a hipótese fundamental da medida estatísticas de R de Pearson, a qual diz que as variáveis têm que ter um comportamento similar à distribuição normal, por outras palavras que as variáveis sejam consideradas normais. Para que uma variável seja normal a significância do teste tem que ser maior a 0.05. Assim, o valor do coeficiente de R de Pearson assume o valor indicativo da associação na amostra mas que poderá não se verificar na população de todos os TBDs.

Da análise dos resultados das tabelas anteriormente expostas, pode-se sublinhar que o 53,7% das relações entre as variáveis são fortes e muito fortes. Outro aspecto que se pode realçar é que as duas variáveis com características económicas, neste caso o “ganho médio mensal” e o “poder de compra” têm um nível de relação com a variável número de viagens em autocarro de muito fraco de acordo com a classificação na Tabela 8.

Tabela 16 – Nível de correlação com base no Ró de Spearman

	Densidade	# de pessoas entre 15e24 e maior a 50	Desempregado	# de mulheres maior a 15	Sem nível de escolaridade	Ensino Superior	Ganho medio mensal	Poder de compra	# Viagens em autocarro
Densidade									
# de pessoas entre 15e24 e maior a 50	moderada								
Desempregado	moderada	Forte							
# de mulheres maior a 15	moderada	Muito forte	Muito forte						
Sem nível de escolaridade	moderada	Muito forte	Muito forte	Muito forte					
Ensino Superior	moderada	Forte	Forte	Muito forte	Muito forte				
Ganho medio mensal	muito fraca	muito fraca	Fraca	Fraca	muito fraca	Fraca			
Poder de compra	muito fraca	Fraca	Fraca	Fraca	Fraca	moderada	moderada		
# Viagens em autocarro	moderada	Forte	moderada	moderada	moderada	Forte	muito fraca	muito fraca	

Na Tabela 16, se apresenta a designação verbal da intensidade de correlação (com base na Tabela 8), ali se pode observar como as variáveis com características económicas se relacionam com as outras além de ter uma noção de que variáveis podem ser mais explicativas com a variável dependente, como é o caso da idade das pessoas entre os 15 e 24 mais as maiores a 50 anos e variável de ensino superior.

A escolha das variáveis que entrarão na análise do modelo MRLM de procura de transporte em autocarro para os TBD em Portugal, foram escolhidas com base nas relações entre a variável depende (nº de viagens de autocarro) e as variáveis independentes (ajustadas do estado da arte

aos dados disponíveis em Portugal) através da análise do nível e correlação linear entre estas e variável dependente (que deve ser o mais forte possível) e entre si (que deve ser o mais fraco possível). É importante sublinhar que as variáveis com características económicas não apresenta uma forte relação com a avrai´vel dependente, no entanto o seu uso não é rejeitado. Seguidamente, apresentam-se as variáveis independentes selecionadas para serem utilizadas no MRLM para estimar as viagens em autocarro em TBD:

- Densidade;
- Número de pessoas entre os 15 e 24 anos e maiores a 50 anos;
- Número de pessoas desempregadas;
- Número de mulheres maiores a 15 anos;
- Número de pessoas sem nível de educação;
- Número de pessoas com educação superior;
- Ganho medio mensual;
- Poder de compra.

4.2 Modelo de procura de transporte para os territórios de baixa densidade de Portugal continente – regressão linear múltipla

4.2.1 Introdução

Constituindo o transporte uma actividade cuja natureza tem como característica fundamental traduzir-se num serviço que não é procurado pela sua utilidade imediata e direta, mas antes se caracteriza pela capacidade de acesso que proporciona a outros bens e serviços, a estimação da sua procura afigura-se como uma tarefa complexa, já que tem de tomar em consideração a organização das atividades que dão origem a esse transporte, como é o caso das residências, locais de emprego, locais de lazer, atividades económicas de um modo geral.

A estimação da procura para os territórios de baixa densidade assume uma importância relativa semelhante à da estimação associada a um território urbano com uma densidade populacional muito elevada, já que ao ter uma menor quantidade de massa crítica, o planeamento do sistema de transportes tem que ser feito de uma maneira muito mais dinâmica e ajustada às reais necessidades dessas populações, nomeadamente tendo em conta que o comportamento de

aquele território pode ser muito diferente dos restantes. Para os territórios de baixa densidade que não apresentam um serviço de transporte público de massas, geralmente por autocarro, a estimação da procura é fundamental para definir como se pode satisfazer as necessidades dessa população, permitindo dessa maneira analisar se é preciso um transporte com tempos fixos ou se é possível um transporte flexível, ou adoptar outro tipo de serviços que se configurem fiáveis e viáveis.

Embora se tenha elaborado um estudo bivariado das relações na secção anterior, para a selecção das variáveis mais relevantes, i.e., que possuam alguma influência no comportamento das viagens geradas nos municípios dos TBD e analisar as relações entre as mesmas, importa referir que os resultados encontrados descrevem apenas o tipo de relação linear existente entre duas variáveis, sendo possível constatar a existência de algumas variáveis com uma maior influência na geração de viagens e outras com uma possível influência baixa.

Por conseguinte, é necessário conhecer a relação entre uma variável dependente (número de viagens em autocarro) e um conjunto de variáveis independentes (variáveis demográficas, territoriais, de mobilidade e económicas). Neste caso, tal como se advogou nas secções anteriores, foi seleccionado para este estudo o modelo de regressão linear múltipla - MRLM, que será apresentado e discutido detalhadamente nos pontos seguintes deste capítulo.

4.2.2 Estimação do modelo de Regressão Linear Múltipla

É importante realçar o mencionado na secção anterior, mais especificamente no ponto 3.3, onde se apresenta a metodologia para a escolha da amostra (municípios) para definir o modelo de regressão linear múltipla, que permita estimar a procura de transporte em autocarro nos territórios de baixa densidade para Portugal continental.

Assim, a amostra é constituída por 128 municípios escolhidos aleatoriamente, ou seja, corresponde ao 85% do total de municípios classificados como TBD (i.e., o total de dados sem *outliers*), entendendo-se que este valor permitirá obter uma base de dados robusta para que a estimação do modelo fosse o mais confiável possível.

Na estimação do MRLM, o método de selecção de variáveis usado no software SPSS foi o stepwise, cuja finalidade é seleccionar as variáveis independentes que tenham maior relação com a variável dependente, ou seja, o que selecciona as variáveis que explicam da melhor forma a variabilidade do número de viagens em autocarro.

Em princípio, inicia-se a estimação do modelo, vinculando aquelas variáveis que foram consideradas significativas segundo o análise bivariada das correlações realizada no ponto 4.1.3. Na Tabela 17 **Error! Reference source not found.** apresenta-se a informação sumária do modelo que foi determinado tendo como “input” as 8 variáveis consideradas significativas. Nesta tabela pode-se analisar o ordem como foram inseridas e contempladas as variáveis no modelo de RLM, tendo sido produzidos 4 MRLM com um ajuste de 90,5%.

Tabela 17 – Sumário do modelo

Model Summary ^e										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	,887 ^a	,787	,785	163,213	,787	416,960	1	113	,000	
2	,933 ^b	,871	,868	127,705	,084	72,575	1	112	,000	
3	,946 ^c	,895	,892	115,805	,024	25,202	1	111	,000	
4	,952 ^d	,905	,902	110,152	,011	12,684	1	110	,001	1,899

a. Predictors: (Constant), Sem_nivel_de_escolaridade_completo

b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade

c. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra

d. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra, Desempregados

e. Dependent Variable: Movimentos_pend_autocarro

No entanto, no processo de diagnóstico do modelo, foi encontrado um incumprimento da hipótese de multicolinearidade entre as variáveis independentes do MRLM, ou seja, não se verifica que as variáveis são realmente independentes entre si. Para verificar está hipótese utilizam-se os procedimentos² associados à matriz de correlação entre as variáveis e a determinação da tolerância ou do valor do parâmetro VIF.

Analisando a Tabela 18 é possível observar que a variáveis do “número de desempregados” tem um coeficiente de correlação superior ao 0,9 com a variável do “número de pessoas sem qualquer nível de escolaridade”, indica é possível existência de multicolinearidade. Corroborando o anterior dito se analisa o procedimento da tolerância ou VIF, apresentado na

⁹² Procedimentos mencionados na seção 3.4.10.4

Tabela 19 onde se pode observar que o valor do VIF é superior a 5 o qual verifica a existência de multicolinearidade entre duas variáveis.

Por tanto, se conclui que este modelo não pode ser usado para estimar a procura de transporte já que não cumpre com uma das hipóteses da RLM. Analisando o sumário do modelo se observa que a diferença em termos de explicação dos modelos 3 e 4 é de 0,01 e entre os R quadrados (qualidade de ajuste do modelo) é de 0,006, além que o modelo 3 cumpre com a hipótese de multicolinearidade já que o valor do VIF é menor a 5 e as correlações não há valor superior a 0,9. Deste modo, conclui-se que o modelo escolhido para continuar com os análises para se estimar um modelo de regressão linear múltipla é o modelo 3, cujas variáveis são: o número de pessoas sem nível de escolaridade, a densidade populacional e o poder de compra.

Tabela 18 – Resultados de correlação

		Correlations				
		Movimentos_ pend_autocar ro	Sem_nivel_d e_escolarida de_completo	Densidade	Poder_de_co mpra	Desemprega dos
Pearson Correlation	Movimentos_ pend_autocar ro	1,000	,887	,656	,005	,885
	Sem_nivel_de_escolarid ade_completo	,887	1,000	,447	,225	,932
	Densidade	,656	,447	1,000	-,043	,501
	Poder_de_compra	,005	,225	-,043	1,000	,227
	Desempregados	,885	,932	,501	,227	1,000
Sig. (1-tailed)	Movimentos_ pend_autocar ro	.	,000	,000	,480	,000
	Sem_nivel_de_escolarid ade_completo	,000	.	,000	,008	,000
	Densidade	,000	,000	.	,325	,000
	Poder_de_compra	,480	,008	,325	.	,007
	Desempregados	,000	,000	,000	,007	.
N	Movimentos_ pend_autocar ro	115	115	115	115	115
	Sem_nivel_de_escolarid ade_completo	115	115	115	115	115
	Densidade	115	115	115	115	115
	Poder_de_compra	115	115	115	115	115
	Desempregados	115	115	115	115	115

Tabela 19 – Resultado dos coeficientes

		Coefficients ^a									
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-57,408	29,714		-1,932	,056					
	Sem_nivel_de_escolaridade_completo	,213	,010	,887	20,420	,000	,887	,887	,887	1,000	1,000
2	(Constant)	-119,040	24,349		-4,889	,000					
	Sem_nivel_de_escolaridade_completo	,178	,009	,742	19,536	,000	,887	,879	,664	,800	1,250
	Densidade	3,677	,432	,324	8,519	,000	,656	,627	,290	,800	1,250
3	(Constant)	315,370	89,305		3,531	,001					
	Sem_nivel_de_escolaridade_completo	,190	,009	,792	22,094	,000	,887	,903	,681	,740	1,351
	Densidade	3,348	,397	,295	8,438	,000	,656	,625	,260	,778	1,285
	Poder_de_compra	-6,331	1,261	-,161	-5,020	,000	,005	-,430	-,155	,923	1,083
4	(Constant)	377,868	86,740		4,356	,000					
	Sem_nivel_de_escolaridade_completo	,127	,019	,530	6,532	,000	,887	,529	,192	,131	7,647
	Densidade	2,970	,392	,261	7,575	,000	,656	,585	,222	,721	1,386
	Poder_de_compra	-6,738	1,205	-,171	-5,592	,000	,005	-,470	-,164	,915	1,093
	Desempregados	,259	,073	,300	3,561	,001	,885	,322	,104	,121	8,232

a. Dependent Variable: Movimentos_pend_autocarro

Com base na análise realizada anteriormente, repete-se o processo e volta-se a introduzir ao SPSS as 3 variáveis do MRLM apresentado na Tabela 19, ou seja, aquelas que apresentaram maior poder explicativo do comportamento da variável dependente e que não apresentavam multicolinearidade entre elas. Seguidamente, os resultados desta análise são apresentados de forma detalhada.

Na Tabela 20, apresenta-se o sumário do modelo, sendo possível observar que a primeira variável que entra no modelo corresponde às pessoas que não têm qualquer nível de escolaridade (analfabetos), isto acontece porque é a variável com o maior coeficiente de correlação (o valor pode ver-se na Tabela 21). O sumário do modelo também apresenta o coeficiente de correlação múltipla, R para a variável VGA - número de viagens em autocarro. O R quadrado (“R Squared”) ou o coeficiente da determinação constitui uma medida da qualidade do ajuste e permite concluir, que aproximadamente o 84% do comportamento da variável é tido em consideração pelo modelo 3 (definitivo para o procedimento de stepwise).

Tabela 20 – Sumário do modelo

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	,847 ^a	,718	,716	209,506	,718	321,006	1	126	,000	
2	,903 ^b	,815	,812	170,602	,096	65,019	1	125	,000	
3	,914 ^c	,836	,832	161,105	,021	16,171	1	124	,000	1,513

a. Predictors: (Constant), Sem_nivel_de_escolaridade_completo

b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade

c. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra

d. Dependent Variable: Movimentos_pend_autocarro

As variáveis não estão, na sua generalidade, perfeitamente correlacionadas, pelo que se introduz um erro quando se utiliza uma regressão linear para estudar e descrever uma relação que se considere linear “imperfeita” (Ribeiro, 2012). O “Std. Error of the Estimate” fornece, de certa forma, uma média ou avaliação global da exactidão da discrepância dos resultados do teste, relativamente à curva de regressão estimada. Assim, no caso apresentado, os resultados reais variam entre cerca de 161,105 “pontos” acima ou abaixo do valor estimado.

Tabela 21 – Correlações do modelo

Correlations

		Movimentos_pend_autocarro	Sem_nivel_de_escolaridade_completo	Densidade	Poder_de_compra
Pearson Correlation	Movimentos_pend_autocarro	1,000	,847	,566	,074
	Sem_nivel_de_escolaridade_completo	,847	1,000	,321	,307
	Densidade	,566	,321	1,000	-,050
	Poder_de_compra	,074	,307	-,050	1,000
Sig. (1-tailed)	Movimentos_pend_autocarro	.	,000	,000	,202
	Sem_nivel_de_escolaridade_completo	,000	.	,000	,000
	Densidade	,000	,000	.	,287
	Poder_de_compra	,202	,000	,287	.
N	Movimentos_pend_autocarro	128	128	128	128
	Sem_nivel_de_escolaridade_completo	128	128	128	128
	Densidade	128	128	128	128
	Poder_de_compra	128	128	128	128

A Tabela 22 apresenta o valor de todos os testes F, os quais têm associado uma significância de 0,000, que é inferior a 0,05, o que leva à rejeição da nulidade dos coeficientes do parâmetro das variáveis independentes consideradas em cada modelo. Para o caso do modelo 3, o teste F permite rejeitar a hipótese nula dos coeficientes da regressão serem todos nulos, logo, é possível

escrever a expressão do modelo com os respectivos valores dos B_i , extraídos da tabela dos coeficientes.

Tabela 22 – Resultados de Anova para o modelo

		ANOVA ^a				
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14089831,54	1	14089831,54	321,006	,000 ^b
	Residual	5530480,326	126	43892,701		
	Total	19620311,87	127			
2	Regression	15982192,24	2	7991096,120	274,561	,000 ^c
	Residual	3638119,627	125	29104,957		
	Total	19620311,87	127			
3	Regression	16401908,43	3	5467302,810	210,647	,000 ^d
	Residual	3218403,436	124	25954,866		
	Total	19620311,87	127			

a. Dependent Variable: Movimentos_pend_autocarro

b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo

c. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade

d. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra

Agora é importante saber se todas as variáveis independentes influem de igual modo para o modelo e além saber qual é o efeito na previsão ou estimação da variável dependente.

Na Tabela 23 apresentam-se os coeficientes não estandardizados, os seus desvios padrões, os coeficientes estandardizados (Beta), os testes t e as tolerâncias de cada variável, para as várias etapas do procedimento *stepwise*. É importante sublinhar que os coeficientes B não são validos para avaliar o nível de importância no modelo das diversas variáveis independentes associadas, para poder fazer esta análise tem-se que utilizar os coeficientes estandardizados, já que estes são independentes do valor das variáveis e permitem estabelecer comparações entre si.

O teste t apresentada na Tabela 23 permite testar as hipóteses nulas de inexistência de uma relação linear entre Y com cada uma das variáveis X, bem como com as constantes consideradas nos modelos 1, 2 e 3. O teste é usado para decidir se certas variáveis independentes podem ser ou não eliminadas da equação de regressão. Em qualquer dos modelos apresentados seguidamente os testes t têm associado um nível de significância inferior a 0,05, pelo que se conclui que cada uma das variáveis independentes têm poder explicativo no número de viagens em autocarro.

Através do estudo comparativo da importância das variáveis utilizando os coeficientes Beta, podem ser retiradas algumas conclusões úteis sobre a magnitude das relações entre as variáveis independentes e o indicador de desempenho com a influência em simultâneo de todas as outras variáveis independentes.

Tabela 23 – Resultados dos coeficientes para o modelo

		Coefficients ^a									
Model		Unstandardized Coefficients		Standardized Coefficients		Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t		Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-23,620	35,133		-,672	,503					
	Sem_nivel_de_escolaridade_completo	,203	,011	,847	17,917	,000	,847	,847	,847	1,000	1,000
2	(Constant)	-116,505	30,841		-3,778	,000					
	Sem_nivel_de_escolaridade_completo	,178	,010	,742	18,245	,000	,847	,853	,703	,897	1,115
	Densidade	3,630	,450	,328	8,063	,000	,566	,585	,311	,897	1,115
3	(Constant)	321,880	112,838		2,853	,005					
	Sem_nivel_de_escolaridade_completo	,191	,010	,798	19,531	,000	,847	,869	,710	,792	1,263
	Densidade	3,344	,431	,302	7,757	,000	,566	,572	,282	,872	1,146
	Poder_de_compra	-6,435	1,600	-,156	-4,021	,000	,074	-,340	-,146	,881	1,135

a. Dependent Variable: Movimentos_pend_autocarro

Assim, o **modelo final** estimado corresponde ao modelo 3 e é dado por:

$$VGA = 321,88 + 0,191 * SnE + 3,344 * DP - 6,435 * PdC \quad (24)$$

Onde: VGA: número de viagens em autocarro; Intervalos Domínio
 SnE: número de pessoas sem nível de educação; [645;8221]
 DP: densidade populacional; [6;166]
 PdC: poder de compra. [56,08;105,7]

É importante sublinhar que o modelo estimado apresentado na equação (24), poderá sofrer alguma alteração, nomeadamente nos coeficientes, já que no momento da análise dos resíduos poderão ser eliminados os valores considerados outliers, dado o potencial desvio que poderão provocar entre os valores reais e os modelados.

4.2.3 Diagnóstico do modelo do ponto de vista estatístico

Nesta seção vão analisar-se separadamente as hipóteses da homocedasticidade, da independência e da normalidade das variáveis aleatórias residuais e a hipótese de multicolinearidade entre as variáveis independentes.

4.2.3.1 Hipóteses de Homocedasticidade

A homocedasticidade verifica o pressuposto que os erros devem ter variância constante. Os processos para analisar esta hipótese corresponde à análise das gráficas a) e b) na Figura 11.

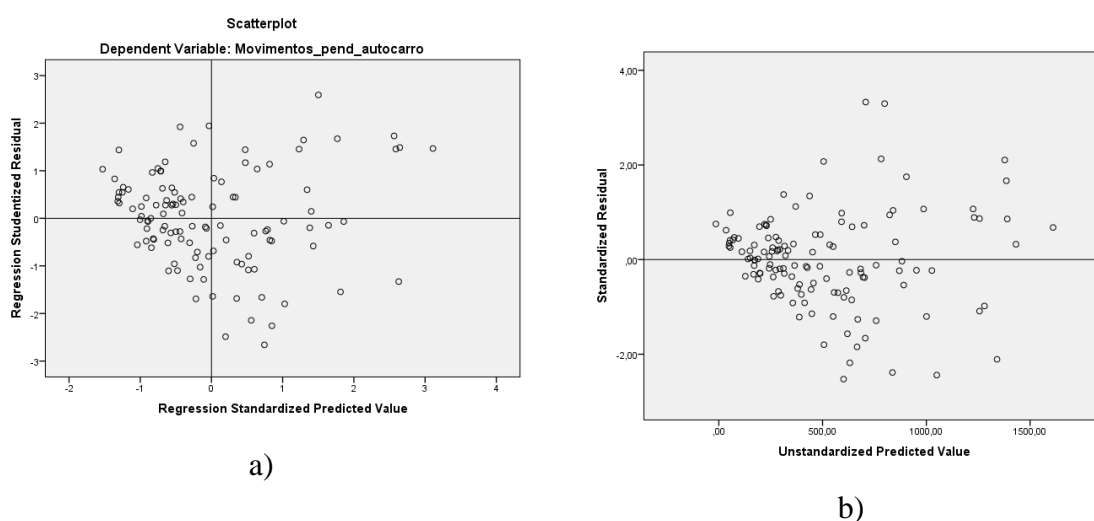


Figura 11 – Gráficos de análise da variância

Ambos os gráficos mostram que os resíduos mantêm uma amplitude aproximadamente constante em relação ao eixo horizontal zero, ou seja, não se evidenciam tendências crescentes ou decrescentes, pelo que não se rejeita a hipótese da homocedasticidade.

4.2.3.2 Diagnóstico de independência

Para determinar se os resíduos cumprem com a independência entre eles e necessário verificar que o valor do teste de Darwin-Watson (valor na Tabela 20) este dentro da região de não rejeição da hipótese nula, ou seja, quer dizer que os resíduos são independentes entre si.

Com base na Tabela 9, a região de não rejeição é determinada pelo valor de $[d_U; 4 - d_U]$. Para determinar o valor d_U é necessário a tabela do teste D-W para um nível de significância do 5%, além de ter o número de dados (n) e número de variáveis utilizados no modelo.

Tabela 24 – Verificação da independência

k	n	D-W
3	128	1,513
Não rejeitar a H_0		
[1,638;2,362]		

Com base na Tabela 24, o valor do teste D-W está na zona de não rejeição da hipótese nula. Assim, conclui-se que não existe autocorrelação entre os resíduos. É importante ter em conta que a violação desta hipótese leva a que as estimativas dos parâmetros do modelo pareçam mais precisas, pois o erro padrão da regressão tem um valor inferior quando existe autocorrelação, originando intervalos de confiança para B_i de menor amplitude do que realmente são (Gageiro & Pestana, 2014).

4.2.3.3 Diagnóstico de normalidade

Com este diagnóstico verifica-se se os resíduos apresentam um comportamento normal. Para esta análise vão ser usados os quatro procedimentos citados na seção da metodologia para avaliar a hipótese de normalidade.

A Tabela 25 apresenta o teste de K-S com a correção de Lilliefors e o teste de Shapiro-Wilk. É possível concluir que para um nível de significância de 5%, os resíduos não apresentam um comportamento normal, já que sua significância é inferior a 0,05. Mas é importante realçar que o valor está perto do 0,05, pelo que se poderá assumir que os resíduos apresentam um comportamento normal. No entanto, para se poder ter uma decisão mais correta serão analisados os restantes dois procedimentos.

Tabela 25 – Teste à normalidade dos resíduos

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	,080	128	,045	,968	128	,004

a. Lilliefors Significance Correction

Na Figura 12 e Figura 13 mostram que as observações se dispõem à volta das retas respectivamente oblíqua e horizontal, indicando a não violação da normalidade. Estes gráficos também permitem analisar as observações que se desviam da normalidade. Este é o caso dos municípios identificados com o código 609 e 613, que corresponde a Miranda do Corvo e Penacova, respetivamente, eles apresentam um comportamento mais afastado da normalidade, o que os torna candidatos a outliers, este análise será feita na seção seguinte.

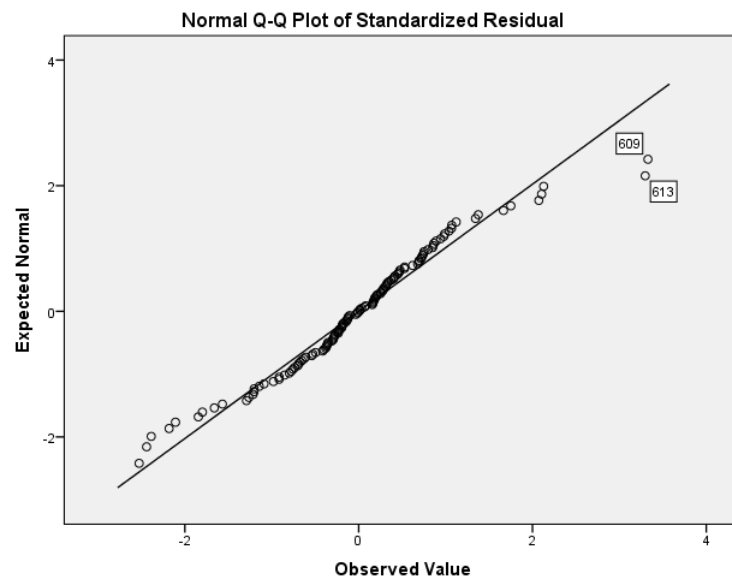


Figura 12 - Gráfico Normal Q-Q Plot, teste à normalidade

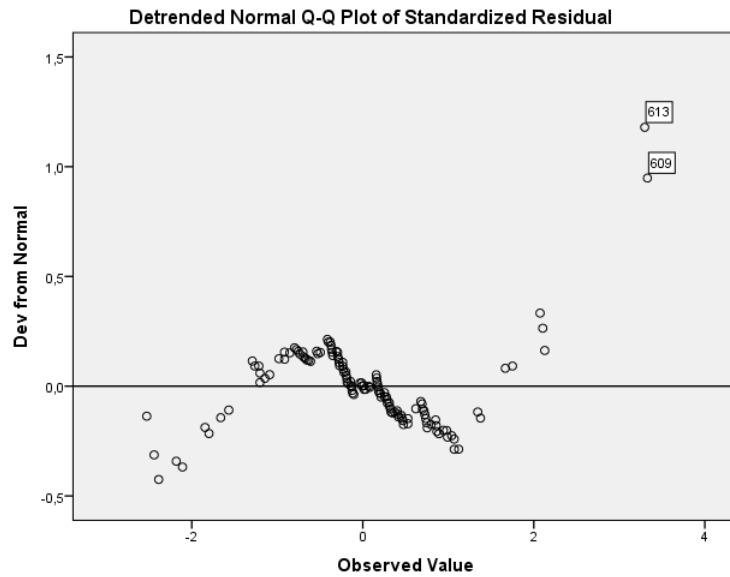


Figura 13 - Gráfico Detrended Normal Q-Q plot, teste à normalidade

Para concluir a análise da hipótese de normalidade, apresenta-se o histograma dos resíduos (Figura 14), o qual permite visualizar que a distribuição dos valores é similar à normal.

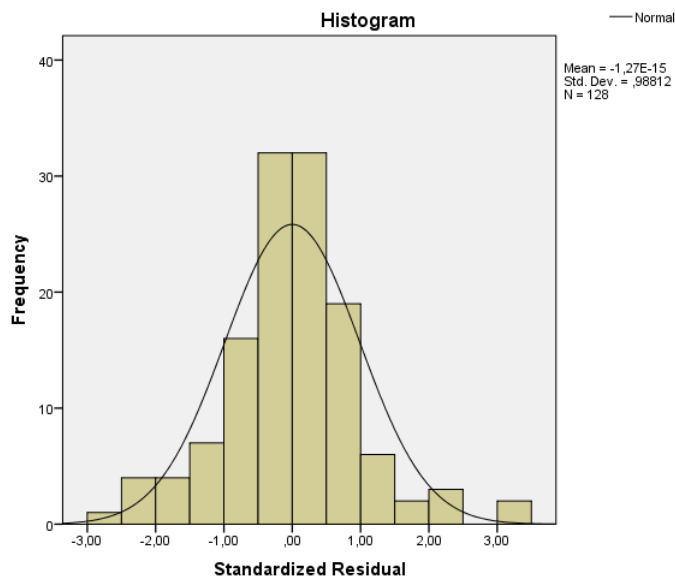


Figura 14 – Distribuição dos resíduos

4.2.3.4 Diagnóstico da multicolinearidade

O modelo de regressão linear múltipla assume que as variáveis independentes são realmente independentes, isto é, não estão correlacionadas. Para verificar este pressuposto analisam-se os

dois procedimentos referidos na secção da metodologia, dentro dos quais se apresenta a análise de correlação das variáveis independentes através da matriz de correlações do modelo e do VIF.

A verificação da multicolinearidade através da Tabela 21 é uma forma preliminar de análise. Quando os coeficientes de correlação entre as variáveis independentes são elevados (superiores a 0,9) indicam a possibilidade de existência da multicolinearidade. Portanto, como o modelo não apresenta uma correlação entre variáveis independentes com um valor superior a 0,9, sendo a correlação mais elevada de 0,32, indicando que não deverá existir multicolinearidade.

O segundo procedimento é o análise da Tabela 26, cujo valores apresentados são a tolerância e o VIF, a tolerância mede o grau em que uma variável X é explicada por todas as outras variáveis independentes. Se o valor da tolerância estivesse muito próximo de zero seria um indicador da existência de multicolinearidade. Porém, no modelo em análise não se apresenta nenhum caso com valores próximos a zero.

Outro indicador para avaliar a multicolinearidade é o valor do VIF (Variance Inflation Factor), que é o inverso da tolerância, neste caso quanto mais próximo de zero menor será a multicolinearidade. Outra forma de analisar os resultados do VIF é observando que não exista valores superiores a 5, para se concluir da não existência de multicolinearidade.

Tabela 26 – Tolerância e VIF

Coefficients ^a			
Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	Sem nível de escolaridade	1,000	1,000
2	(Constant)		
	Sem nível de escolaridade	0,897	1,115
	Densidade	0,897	1,115
3	(Constant)		
	Sem nível de escolaridade	0,792	1,263
	Densidade	0,872	1,146
	Poder de compra	0,881	1,135

a. Dependent Variable:
Movimentos_pend_autocarro

4.2.4 Análise dos outliers e observações influentes

Os Outliers não são detectadas exclusivamente pela análise dos resíduos, mas necessitam de outras estatísticas para o efeito (Rodrigues, 2012). Nesta seção serão identificados os outliers entre as variáveis independentes e os casos que exercem influência nos coeficientes estimados da regressão. Tendo em conta que as observações influentes são aquelas que individualmente, ou em conjunto com as outras observações, demonstram ter mais impacto do que as restantes no cálculo dos estimadores. A seguir serão apresentados as análises estatísticas necessárias para determinar os outliers entre as variáveis independentes.

4.2.4.1 Resíduos

Os resíduos usados para a identificação dos outliers no MRLM são habitualmente os estandardizados, estudantizados e estudantizados deleted. Estes devem cumprir pelo menos uma das siguientes condições afastadas pelo Gageiro e Pestana (2014).

- Resíduos estandardizados terem valores absolutos superiores a 3;
- Resíduos estudantizados terem valores absolutos superiores a 2;
- Resíduos estudantizados deleted terem valores absolutos superiores a 2.

Na Figura 15 identifica-se a presença de dois municípios cujo resíduo estandardizado é superior a 3, são os municípios Mirando do Corvo (609) e Penacova (613).

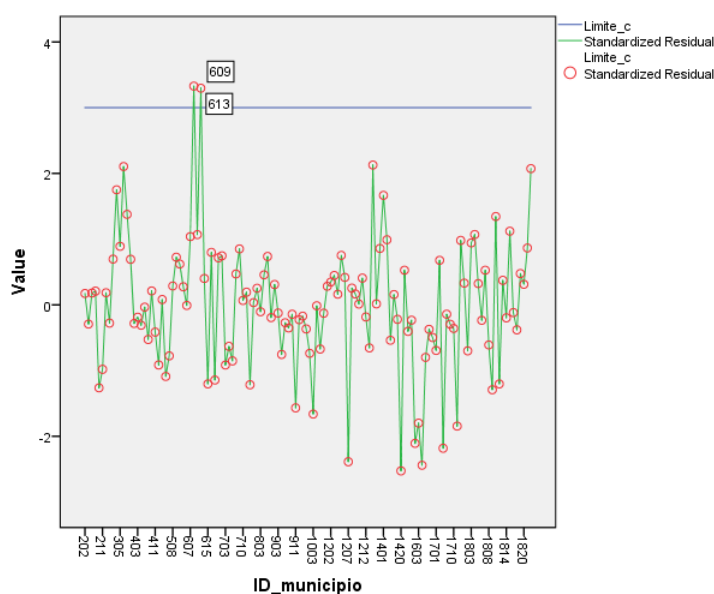


Figura 15 – Resíduos estandardizados

As Figura 16 e Figura 17, identificam um conjunto de municípios como outliers, caso existam municípios com valores aberrantes inferiores e superiores aos esperados. É expectavel que estes municípios influenciam o modelo estimado e, por isso, irão aparecendo ao longo da análise. A análise dos dois gráficos apresentam os mesmos municípios como outliers.

Se um Outlier for influente vai interferir sobre a função de regressão, o que significa que a inclusão, ou não, desse ponto modifica substancialmente os valores ajustados. Portanto, tem-se em conta todos os municípios que estão fora das retas que definem os limites de -2 e 2.

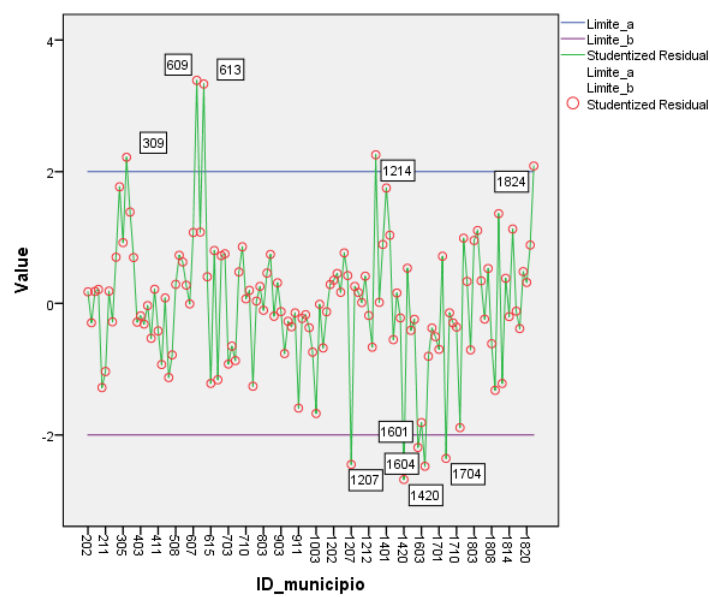


Figura 16 – Resíduos estudantizados

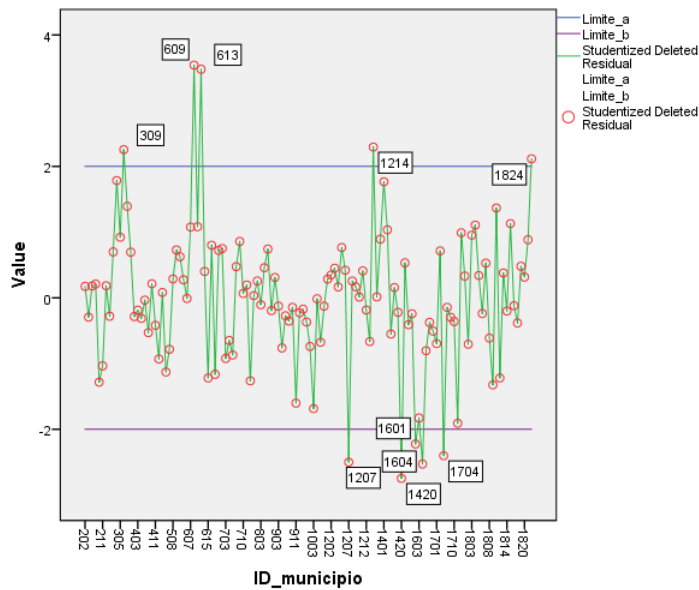


Figura 17 – Resíduos estudentizados deleted

4.2.4.2 Leverage

É uma medida da influência de uma observação na qualidade do ajustamento feito, que varia entre zero (quando não tem qualquer influência no ajustamento) até $(n - 1)/n$. Um valor alto de Leverage indica que a observação poderá ser influente.

Para determinar os municípios influentes calcula-se o valor de Leverage para uma amostra, considerada grande (128 municípios), cuja equação é:

$$lev > \frac{2(k + 1)}{n}, n > 30 \quad (25)$$

k: número de variáveis independentes

n: número de dados ou observações

Com base na equação anterior, o limite para o leverage é $lev = 0.0625$.

Na Figura 18, apresenta-se um conjunto de 12 municípios que apresentam um valor acima do limite, mas é importante atender que nem todos os valores acima do limite são considerados influentes e desta forma tenham que ser eliminados do modelo estimado. Só se devem considerar para esse efeito aqueles municípios cujo valor de Leverage seja muito alto, um

exemplo de um valor que não se considerará influente é o município 1601, uma vez que não é muito influente para o modelo estimado.

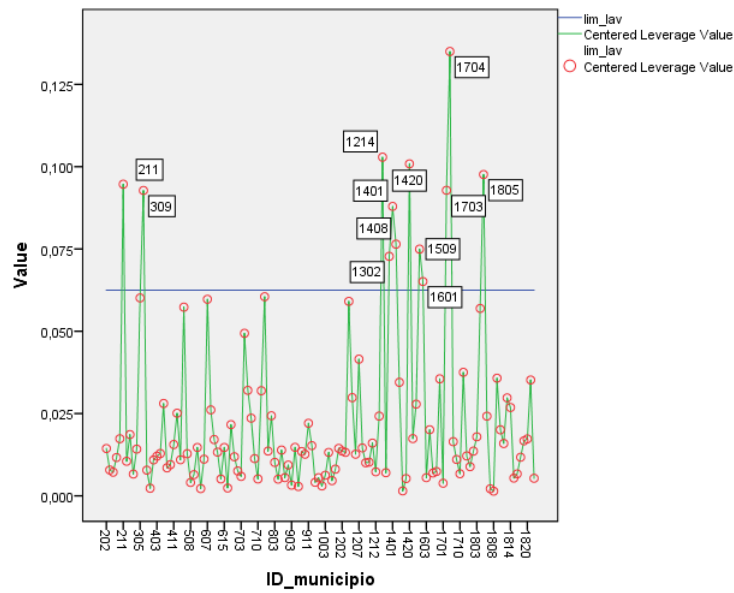


Figura 18 – Gráfico de Leverage

4.2.4.3 Valor estimado ajustado

O valor estimado ajustado é o valor atribuído a um caso quando este não está incluído no cálculo dos coeficientes da regressão. A diferença entre o valor estimado ajustado e os valores estimados permite identificar as observações influentes que se afastam da tendência central.

Na Figura 19 é possível observar que há um grupo de possíveis municípios que podem ser considerados influentes para o modelo, mas a diferenças não são muito significativas, logo não se será selecionado nenhum outlier (município) de acordo com a análise deste teste.

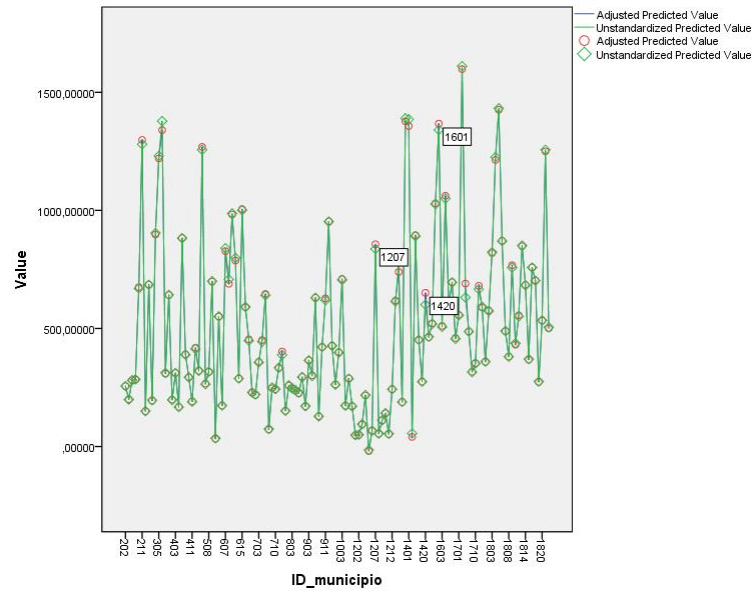


Figura 19 – Valores estimados ajustados

4.2.4.4 Distância de Cook

A distância de Cook permite avaliar a variação dos resíduos de todas as observações, quando um caso é excluído do cálculo dos coeficientes da regressão, sendo a medida mais importante que permite avaliar a influência dos diversos casos, à qual corresponde uma medida global estandardizada das diferenças dos coeficientes da regressão.

Uma observação é considerada influente quando o valor desta distância é superior ao quociente entre 4 e $(n-k-1)$, onde k é o número de variáveis e n o número de municípios ou observações. Seguidamente, apresenta-se a equação com o limite para realizar a respetiva análise.

$$Cook > \frac{4}{n - k - 1} \quad Cook = 0,0323 \quad (26)$$

Na Figura 20 é possível constatar que há 5 municípios com uma influência significativa nos coeficientes de regressão, sendo importante realçar que aunque os outros municípios que estão por encima do limite não tenham uma influência alta em comparação com os 5 primeiros têm que ser tidos em conta para determinar sua influência nos coeficientes.

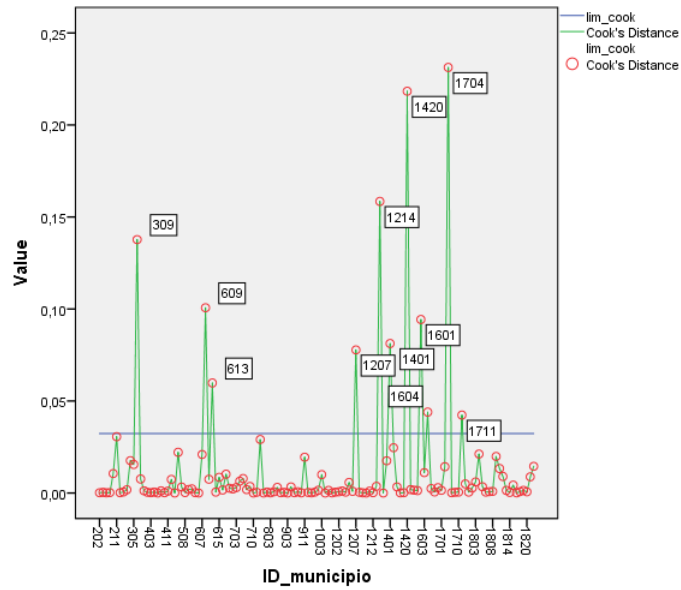


Figura 20 – Distância de Cook

4.2.4.5 Df Beta

A diferença nos valores Beta representam a mudança nos coeficientes de regressão quando se exclui uma observação particular. Existe um DfBeta para cada parâmetro do modelo, incluindo a constante.

São consideradas observações influentes aqueles municípios com valores de Dfbetas maiores que $2/\sqrt{n}$, para amostras de grande dimensão (maiores a 30). Neste caso, resulta um limite de Dfbeta de $\pm 0,1768$

Na Figura 21 verifica-se que os municípios identificados com os códigos 613, 712, 1207 e 1420 têm um efeito significativo no valor estimado para a constante do modelo.

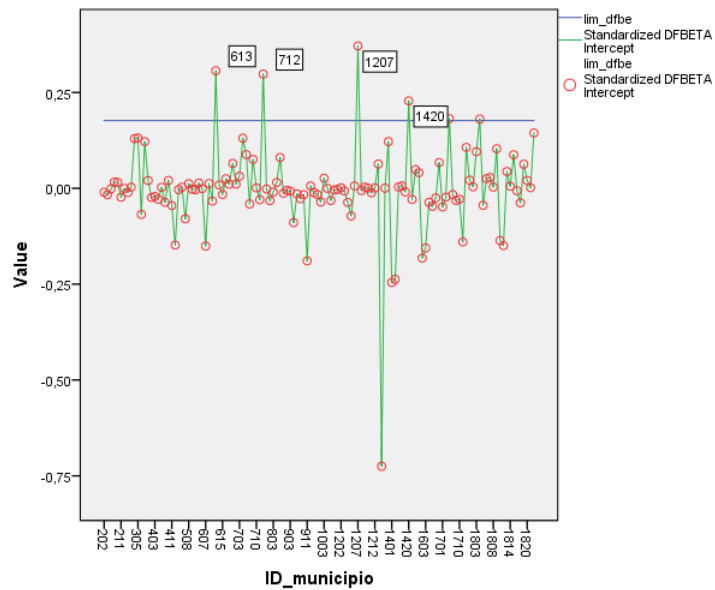


Figura 21 – DfBetas estandardizado para a constante

Na Figura 22 observa-se que os municípios com os códigos 309, 609 e 607 têm influência na estimativa do coeficiente associado à variável - densidade.

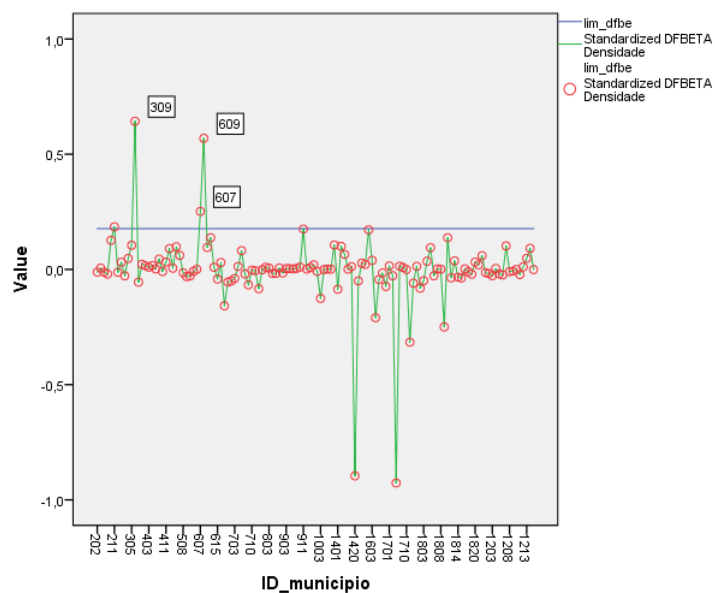


Figura 22 – DfBeta estandardizado para densidade

Na Figura 23 mostra-se que os municípios 1401, 1420, 1703, 1704 têm uma forte influência no coeficiente da variável “sem nível de escolaridade”. Já, o município 1804 embora se encontre acima do limite é possível concluir que não terá muita influência sobre o coeficiente desta variável, não sendo portanto considerado.

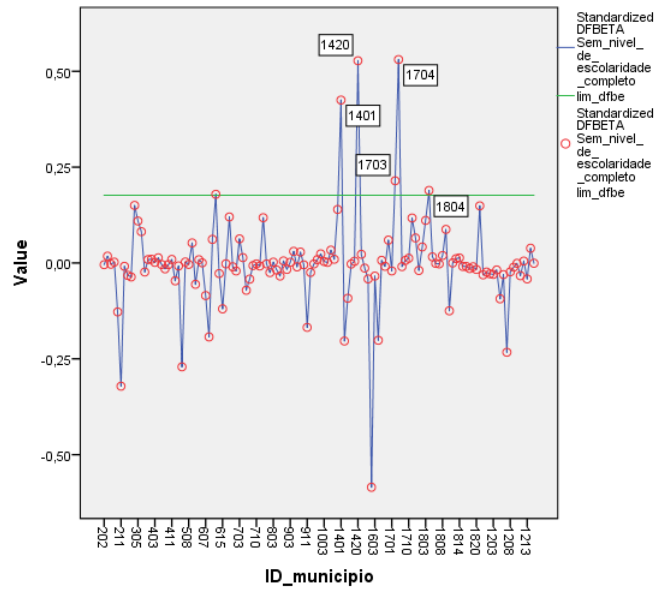


Figura 23 – DfBeta estandardizado para sem nivel de escolaridade

Na Figura 24 observa-se que o município 1214 é o caso que apresenta uma maior influência sobre o coeficiente da variável de poder de compra, sendo que os outros dois municípios (1408 e 1601) não são considerados influentes dada a proximidade do valor limite.

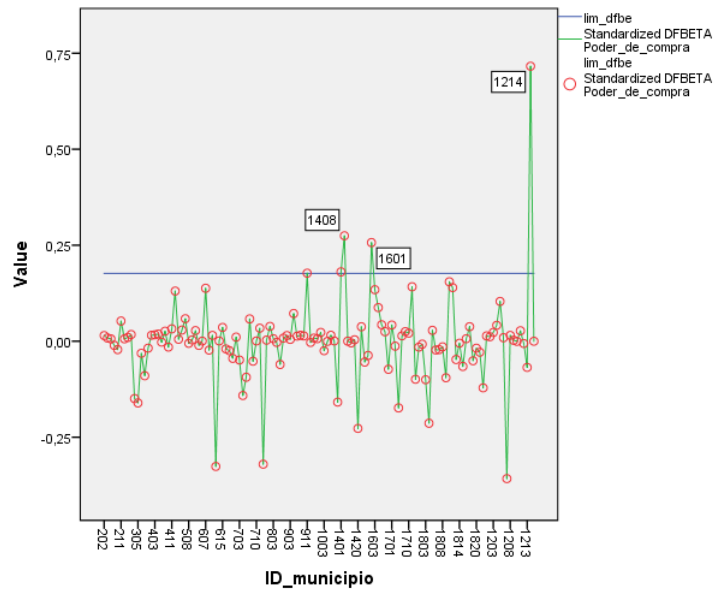


Figura 24 – DfBeta estandardizado para poder de compra

4.2.4.6 Df Fit

O Df Fit é uma medida de influência equivalente à distância de Cook, que mede o efeito de cada observação no ajustamento global do modelo estimado. O seu cálculo obtém-se pela diferença entre o valor estimado de Y e o correspondente valor obtido após a exclusão de uma observação particular.

Habitualmente, são consideradas observações influentes, os valores que obedecem à seguinte desigualdade, onde n e k têm a mesma correspondência com a expressão de DfBeta. Neste caso, o valor dos limites do DfBeta é de $\pm 0,359$.

$$|DfFit| > 2 * \sqrt{\frac{k + 1}{n - k - 1}} \quad (27)$$

Na Figura 25 apresentam-se os municípios que são considerados influentes para o ajuste global do modelo, todos os municípios identificados fora do intervalo definido pelos valores limite são considerados influentes para o modelo.

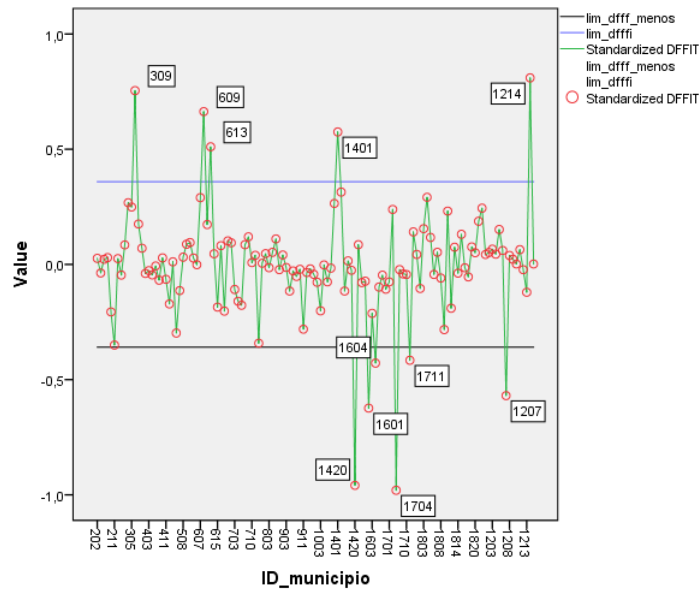


Figura 25 – DfFit estandardizado

4.2.4.7 Conclusão da análise de resíduos e valores influentes

O objetivo principal da análise de resíduos e dos valores influentes é o de eliminar aqueles casos, cujos valores se considera que estão a afetar com maior influência o comportamento do modelo estimado. Assim, é necessário realizar a análise de cada município que seja identificado e considerado influente no modelo, bem como a respetiva frequência em cada parte da análise, i.e., em cada parâmetro estatístico avaliado no ponto 4.2.4.

No caso dos *Outliers*, que foram determinados no ponto 4.2.4.1, foi considerado que todos deverão ser eliminados da estimação, ou seja, na elaboração de um novo modelo - MRLM, que tenha em linha de conta os resultados da análise do modelo 3 da Tabela 20, uma vez que estes valores estarão sempre a influenciar negativamente o nível de ajuste do modelo de regressão linear.

Assim, importa identificar os 14 municípios que serão eliminados da estimação do novo modelo – MRLM - para determinar a previsão do número de viagens em autocarro nos TBDs portugueses, que se apresentam na Tabela 27 com seu respectivo código de identificação do SPSS que é mostrado nos diversos gráficos apresentados nos pontos anteriores desta secção.

Tabela 27 – Municípios Outliers e influentes

ID_MUNICIPIO	Nome_Municipio
1805	<i>Lamego</i>
1703	<i>Chaves</i>
0211	<i>Odemira</i>
1401	<i>Abrantes</i>
0309	<i>Póvoa de Lanhoso</i>
1601	<i>Arcos de Valdevez</i>
1214	<i>Portalegre</i>
1824	<i>Vouzela</i>
0613	<i>Penacova</i>
0609	<i>Miranda do Corvo</i>
1604	<i>Monção</i>
1207	<i>Elvas</i>
1704	<i>Mesão Frio</i>
1420	<i>Vila Nova da Barquinha</i>

4.3 Estimação do novo modelo – MRLM - para estimar viagens de autocarro em TBDs

Após se terem determinado e eliminado aqueles municípios que foram considerados outliers ou observações influentes através da análise dos resíduos do MRLM apresentado no ponto 4.2, neste ponto será apresentado um novo modelo de estimação da procura de transporte em autocarro para os TBD em Portugal, ou seja, um novo modelo que resulta de um refinamento do modelo do ponto 4.2 em termos de valores da amostra, uma vez que se matirão as mesmas variáveis.

O número de municípios eliminados da base de dados por cumprir com as características mencionadas na seção 4.2.4 foram 14. Portanto, a base de dados para estimar o novo modelo ainda se considera suficientemente grande e representativa e robusta para análise deste assunto nos TBDs português, i.e., permitirá ter um MRLM estimado com o 70% dos dados totais.

Os resultados da análise estatística do modelo são apresentados desde a Tabela 28 até a Tabela 30, cuja análise é semelhante à realizada no ponto anterior, verificando-se que se mantém a primeira variável a entrar no modelo “número de pessoas que não têm escolaridade”, isto acontece porque o valor de correlação desta variável com o número de viagem em autocarro é de 0,86, ou seja, a maior de todas. O coeficiente R quadrado (“R Squared”) teve um aumento do 3% em relação ao modelo 3 do ponto 4.2, isto quer dizer que um 87% do comportamento da variável é tido em consideração pelo novo “modelo 3”.

O modelo continua a ser significativo na globalidade já que mantém o seu valor de significância em 0,00 para o teste F (no modelo 3 da Tabela 29), também se verifica a significância nos restantes parâmetros estimados e testes realizados (significância dos testes t na Tabela 30 inferiores a 0,05), verificando-se, porém, um aumento da correlação entre as variáveis assim como a explicação da variabilidade total do número de viagens em autocarro para os TBDs.

Tabela 28 – Sumário do modelo

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	,860 ^a	,740	,738	162,994	,740	319,227	1	112	,000	
2	,916 ^b	,840	,837	128,631	,099	68,834	1	111	,000	
3	,933 ^c	,871	,868	115,840	,031	26,867	1	110	,000	1,754

a. Predictors: (Constant), Sem_nivel_de_escolaridade_completo

b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade

c. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra

d. Dependent Variable: Movimentos_pend_autocarro

Tabela 29 - Resultados de Anova

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8480941,334	1	8480941,334	319,227	,000 ^b
	Residual	2975518,947	112	26567,133		
	Total	11456460,28	113			
2	Regression	9619858,967	2	4809929,483	290,701	,000 ^c
	Residual	1836601,314	111	16545,958		
	Total	11456460,28	113			
3	Regression	9980387,474	3	3326795,825	247,920	,000 ^d
	Residual	1476072,806	110	13418,844		
	Total	11456460,28	113			

a. Dependent Variable: Movimentos_pend_autocarro

b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo

c. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade

d. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Densidade, Poder_de_compra

Tabela 30 - Resultado dos coeficientes

Coefficients^a

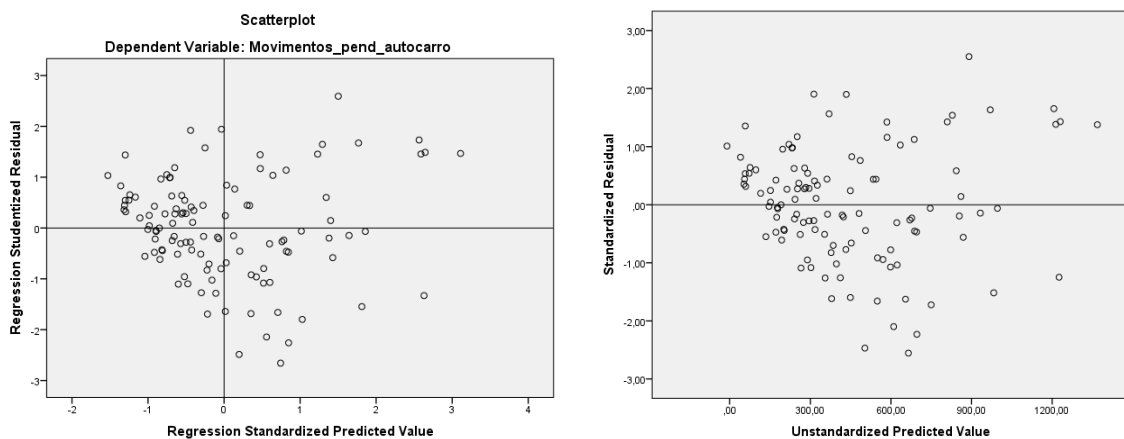
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-42,703	31,232		-1,367	,174					
	Sem_nivel_de_escolaridade_completo	,205	,011	,860	17,867	,000	,860	,860	,860	1,000	1,000
2	(Constant)	-112,325	26,037		-4,314	,000					
	Sem_nivel_de_escolaridade_completo	,173	,010	,727	17,620	,000	,860	,858	,670	,848	1,179
	Densidade	3,779	,456	,342	8,297	,000	,626	,619	,315	,848	1,179
3	(Constant)	339,660	90,297		3,762	,000					
	Sem_nivel_de_escolaridade_completo	,185	,009	,780	20,239	,000	,860	,888	,693	,789	1,268
	Densidade	3,335	,419	,302	7,957	,000	,626	,604	,272	,813	1,230
	Poder_de_compra	-6,541	1,262	-,185	-5,183	,000	-,060	-,443	-,177	,919	1,088

a. Dependent Variable: Movimentos_pend_autocarro

Tal como no ponto 4.2 verificou-se o cumprimento das hipóteses do MRLM, sendo importante realçar que na verificação das hipóteses não se vai entrar em detalhes do procedimento o estatístico usado já que na seção 4.2.3, mas é apresentada uma síntese dos resultados obtidos.

- Hipótese de homocedasticidade

Nos gráficos apresentados na Figura 26 é possível observar que os resíduos mantêm uma amplitude quase constante em relação ao eixo horizontal (zero), ou seja, não mostram tendência crescente ou decrescente. Por esta razão, a hipótese de uma variância constante dos resíduos não é rejeitada. Por outro lado, dado que os valores têm uma dispersão aleatória em relação ao valor zero o fenómeno pode igualmente ser considerado linear.



a) b)
Figura 26 – Gráficos de análises de variância

- Diagnóstico de independência

A análise da covariância nula ou da não autocorrelação dos resíduos é feita através do teste de Durwin – Watson. Com base na Tabela 9, a região de não rejeição é determinada pelo valor de $[d_U; 4 - d_U]$. Para determinar o valor d_U é necessário a tabela do teste D-W para 5% de significância, além de ter o número de dados (n) e número de variáveis utilizados no modelo(k).

Por tanto o valor de teste de D-W para o novo modelo é 1,754 (Tabela 28). Seguidamente, se apresenta a tabela com a análise da região de não rejeição e o novo valor para o teste.

Tabela 31 – Verificação da independência

k	n	D-W
3	114	1,754
Não rejeitar a Ho		
[1,621;2,379]		

Com base na Tabela 31, o valor do teste D-W está na zona de não rejeição da hipótese nula, concluindo-se que não existe autocorrelação entre os resíduos.

- Diagnóstico de normalidade

A Tabela 32 apresenta o teste de K-S com a correcção de Lilliefors e o teste de Shapiro Wilk, com base nesta tabela é possível concluir que para um nível de significância de 5%, os resíduos apresentam um comportamento normal, já que sua significância é maior a 0,05.

Tabela 32 – Teste à normalidade dos resíduos

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	,056	114	,200 [*]	,991	114	,646

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Para corroborar o apresentado na tabela anterior, se analisará as Figura 27 e Figura 28 as quais permitem concluir que as observações se dispõem à volta das retas obliqua e horizontal, respectivamente. Para além disso, não se observam valores com comportamentos afastados dessas retas.

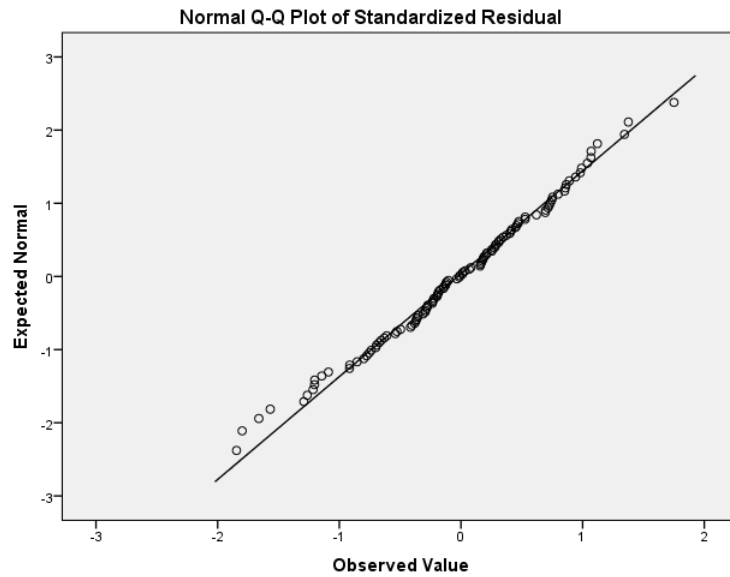


Figura 27 – Gráfico Normal Q-Q plot, teste à normalidade

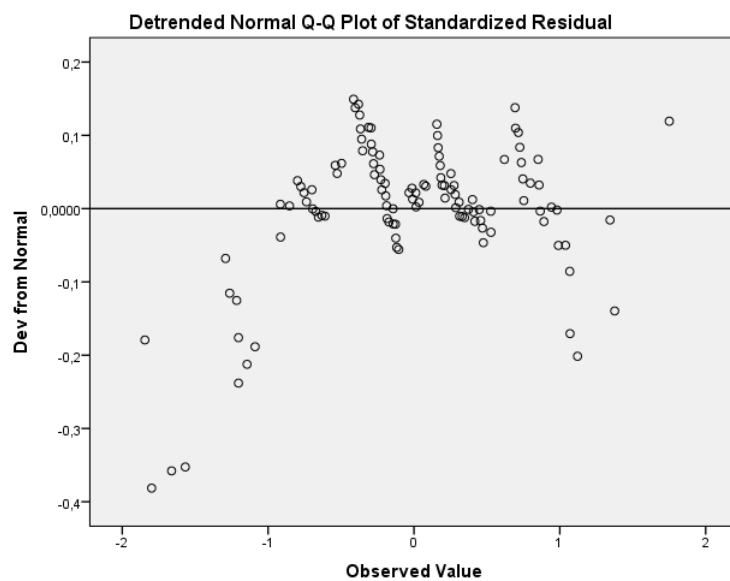


Figura 28 – Gráfico Detrended Normal Q-Q plot, teste à normalidade

- Diagnóstico de multicolinearidade

Para verificar o pressuposto que as variáveis são realmente independentes, analisa-se a matriz de correlações das variáveis independentes e o valor do VIF. Analisando a matriz de correlações, apresentada na Tabela 33, conclui-se que as variáveis não apresentam multicolinearidade, uma vez que não há correlações maiores que 0,90 entre as variáveis independentes.

Tabela 33 – Correlações entre as variáveis

		Correlations			
		Movimentos_ pend_autocar ro	Sem_nivel_d e_escolarida de_completo	Densidade	Poder_de_co mpra
Pearson Correlation	Movimentos_ pend_autoc arro	1,000	,860	,626	-,060
	Sem_nivel_de_escolarid ade_completo	,860	1,000	,390	,202
	Densidade	,626	,390	1,000	-,106
	Poder_de_compra	-,060	,202	-,106	1,000
Sig. (1-tailed)	Movimentos_ pend_autoc arro	.	,000	,000	,263
	Sem_nivel_de_escolarid ade_completo	,000	.	,000	,016
	Densidade	,000	,000	.	,130
	Poder_de_compra	,263	,016	,130	.
N	Movimentos_ pend_autoc arro	114	114	114	114
	Sem_nivel_de_escolarid ade_completo	114	114	114	114
	Densidade	114	114	114	114
	Poder_de_compra	114	114	114	114

Da análise da Tabela 34 verifica-se que nos três modelos não há um valor de VIF superior a 5, logo não existe multicolinearidade entre as variáveis do novo modelo.

Tabela 34 – Tolerância e VIF

Coefficients ^a				
Model		Correlations	Collinearity Statistics	
			Part	Tolerance
1	(Constant)			
	Sem_nivel_de_escolaridade_c ompleto	0,860	1,000	1,000
2	(Constant)			
	Sem_nivel_de_escolaridade_c ompleto	0,670	0,848	1,179
	Densidade	0,315	0,848	1,179
3	(Constant)			
	Sem_nivel_de_escolaridade_c ompleto	0,693	0,789	1,268
	Densidade	0,272	0,813	1,230
	Poder_de_com pra	-0,177	0,919	1,088

Após ser efetuado a verificação de todas as hipóteses subjacentes a um modelo de regressão linear múltipla e analisado que o modelo apresenta um ajuste melhor do que o modelo analisado no ponto 4.2, conclui-se que a exclusão dos 14 municípios se revelou positiva e necessária para aumentar a qualidade do ajuste. Assim, o modelo final proposto para a estimação das viagens em autocarro para os territórios de baixa densidade em Portugal continental é dado por:

$$VGA = 339,66 + 0,185 * SnE + 3.335 * DP - 6,541 * PdC \quad (28)$$

Onde:

VGA: número de viagens em autocarro

SnE: número de pessoas sem nível de escolaridade, com SnE a variar entre [645;6772];

DP: densidade populacional, com DP a variar entre [6; 127];

PdC: poder de compra, com PdC a variar entre [56,08; 95,96].

4.4 Estimação de modelos estratificados

Nesta seção se estimarão os modelos para os grupos definidos de acordo com a estratificação da variável densidade populacional. Estes grupos serão definidos de acordo com os resultados da análise das estatísticas descritivas dessa variável através da formação de grupos de municípios mais homogêneos e que desta forma apresentassem um comportamento diferenciado entre esses grupos para as restantes variáveis, nomeadamente na geração das viagens em autocarro. Por conseguinte, o objetivo desta análise é determinar se a densidade populacional de um município classificado como TBD influi diretamente no comportamento da geração de viagens em autocarro, além analisar que variáveis independentes tem uma maior explicação para cada um desses grupos.

A metodologia para estimar o MRLM para cada grupo, assim como o procedimento para selecionar as variáveis que entram ao modelo é igual à apresentada na seção 4.1.3. Para a seleção da parte da amostra que será usada como validação para cada grupo de municípios, também foi utilizada a mesma metodologia aplicada na seção anterior, que consistia em selecionar o 15% do número total municípios de cada grupo de uma forma aleatória.

Os resultados da análise do diagnóstico do ponto de vista estatístico para cada grupo, vai ser apresentado em conjunto nesta seção 4.4. Porém, as tabelas e gráficos necessários para sustentar o diagnóstico e respetivas conclusões são apresentado no anexo C. Do mesmo modo, também serão apresentados no anexo C os gráficos correspondentes à análise de outliers e observações influentes para cada grupo. É importante sublinhar que os modelos apresentados nos pontos 4.4.2 até ao 4.4.4., para cada grupo de municípios, são os definitivos ou finais, isto quer dizer que já foi retirado os outliers e as observações influentes.

4.4.1 Seleção de grupos de municípios

Tendo por base as principais características em que assenta a definição de TBD assume-se que poderá existir um comportamento mais homogêneo para territórios com níveis populacionais semelhantes, ou ainda melhor, para níveis de população no território semelhante, ou seja, para territórios de idêntica densidade populacional.

Assim, da análise da estatística descritiva de todas as variáveis selecionadas ANEXO A é possível observar que o comportamento da variável densidade populacional apresenta um espectro de frequências bastante alargado, como se pode constatar na Figura 29.

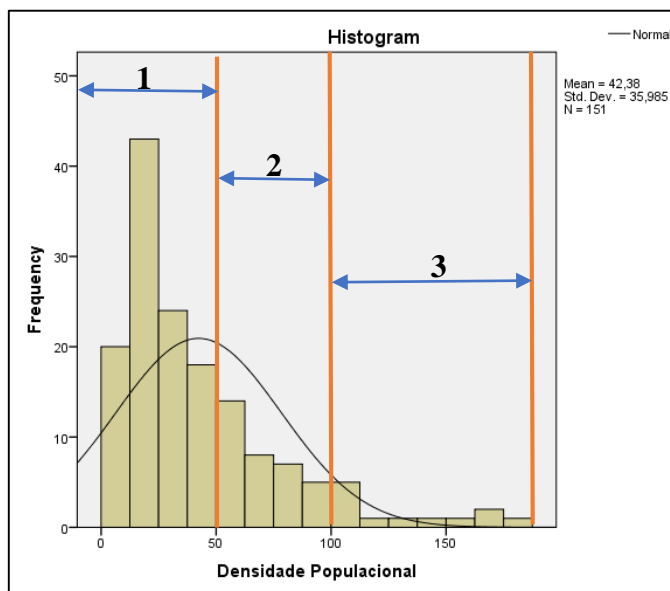


Figura 29 - Histograma da densidade populacional

Por outro lado, de acordo com a designação deste tipo de territórios os municípios apresentam valores de densidade populacional consideravelmente baixos. Assim, para se estimar a procura de transporte público em autocarro, foram propostos 3 grupos de densidade gradativa de acordo com a análise do histograma apresentado na Figura 29, tendo sido definidas as seguintes classes:

1. $0 \leq \text{Densidade Populacional} \leq 50$
2. $50 < \text{Densidade Populacional} \leq 100$
3. $\text{Densidade Populacional} > 100$

4.4.2 Grupo 1 - densidade populacional entre 0 e 50 hab/km²

O número de municípios com uma densidade populacional inferior a 50 hab/Km² é de 80, ou seja, 53% da totalidade dos municípios considerados de baixa densidade. Na Tabela 35 apresenta-se o sumário do MRLM verificando-se que 63% do comportamento do número de viagens em autocarro é tido em consideração pelo MRLM nº 2, para além do coeficiente ajustado (Adjusted R Square), que determina a qualidade do ajustamento feito, apresentar um valor de 0,61, permitindo-se considerar que este MRLM apresenta um nível de ajuste moderado.

Tabela 35 – Sumário do modelo – Grupo 1

Model Summary ^c											
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson	
						F Change	df1	df2			
1	,746 ^a	,556	,549	62,481	,556	73,903	1	59	,000		
2	,790 ^b	,625	,612	57,930	,069	10,633	1	58	,002	1,878	

a. Predictors: (Constant), Sem_nivel_de_escolaridade_completo
 b. Predictors: (Constant), Sem_nivel_de_escolaridade_completo, Desempregados
 c. Dependent Variable: Movimentos_pend_autocarro

Na Tabela 36 é possível analisar a influência de cada variável no modelo e saber o efeito na previsão ou estimação da variável dependente. O valor de significância para o teste t é menor que 0,05 para ambas variáveis, isto quer dizer que têm poder explicativo no número de viagens em autocarro. Entretanto, para o valor da constante a significância do teste t é consideravelmente superior a 0,05 pelo que a utilização deste modelo para fins de inferência deve ser cauteloso.

Tabela 36 – Resultados de coeficientes – Grupo 1

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-6,340	28,102		-,226	,822					
	Sem_nivel_de_escolaridade_completo	,165	,019	,746	8,597	,000	,746	,746	,746	1,000	1,000
2	(Constant)	-28,291	26,911		-1,051	,297					
	Sem_nivel_de_escolaridade_completo	,121	,022	,547	5,427	,000	,746	,580	,436	,636	1,572
	Desempregados	,316	,097	,329	3,261	,002	,659	,394	,262	,636	1,572

a. Dependent Variable: Movimentos_pend_autocarro

O MRLM estimado para os municípios com uma densidade populacional menor a os 50 hab/Km² é dado por:

$$VGA = -28,291 + 0,121 * SnE + 0,316 * Dsgd \quad (29)$$

Onde:

VGA: número de viagens em autocarro;

SnE: número de pessoas sem nível de educação, com SnE a variar entre [645;2200];

Dsgd: número de pessoas desempregadas, com Dsdg a variar entre [91;494].

4.4.3 Grupo 2 - densidade populacional entre 50 e os 100 hab/Km²

O número de municípios que se encontram nesta categoria é de 49, isto equivale ao 32% dos municípios totais. Na Tabela 37 constata-se que 43% do comportamento da variável dependente (número de viagens em autocarro) é tido em consideração pelo MRLM nº 2, que apresenta um coeficiente ajustado de R (Adjusted R Square) de 0,40, ou seja, verifica-se um ajuste que não é muito forte.

Tabela 37 – Sumário do modelo – grupo 2

Model Summary ^c										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	,557 ^a	,311	,290	157,922	,311	15,330	1	34	,000	
2	,658 ^b	,432	,398	145,469	,122	7,071	1	33	,012	1,800

a. Predictors: (Constant), Edade_mulheremayor15

b. Predictors: (Constant), Edade_mulheremayor15, Poder_de_compra

c. Dependent Variable: Movimentos_pend_autocarro

Na Tabela 38 verifica-se que as variáveis explicativas são diferentes do MRLM encontrado para o Grupo 1, i.e.. as mulheres com idade superior a 15 anos apresentam uma maior influência na previsão ou estimação da variável dependente, seguindo-se o poder de compra da população residente. O valor de significância para o teste de inferência estatística (teste t) é menor a 0,05 para ambas variáveis, isto quer dizer que se rejeita a hipótese nula, o que significa que as variáveis tem poder explicativo para a variável dependente e apresentam valores não nulos.

Tabela 38 – Resultados dos coeficientes – grupo 2

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	53,172	120,735		,440	,662					
	Edade_mulheremayor15	,081	,021	,557	3,915	,000	,557	,557	,557	1,000	1,000
2	(Constant)	472,659	193,019		2,449	,020					
	Edade_mulheremayor15	,092	,019	,634	4,723	,000	,557	,635	,619	,954	1,048
	Poder_de_compra	-6,519	2,452	-,357	-2,659	,012	-,221	-,420	-,349	,954	1,048

a. Dependent Variable: Movimentos_pend_autocarro

O modelo estimado para os municípios do Grupo 2, com uma densidade populacional entre 50 e os 100 hab/Km² é dado por:

$$VGA = 472,66 + 0,92 * IM - 6,519 * PdC \quad (30)$$

Onde:

VGA: número de viagens em autocarro;

IM: número de mulheres maiores a 15 anos, com IM a variar entre [4148; 8052]

PdC: poder de compra, com PdC a variar entre [58,25; 95,96]

4.4.4 Grupo 3 - Densidade populacional maior a 100 hab/Km²

O número de municípios que se encontram nesta categoria é de 22, isto equivale ao 15% dos municípios totais. Na Tabela 39 constata-se que 76% do comportamento da variável dependente (número de viagens em autocarro) é tido em consideração pelo modelo nº 2, apresentando um coeficiente ajustado (Adjusted R Square) de 0,73, o qual permite ser considerado um ajuste forte. Este valor de ajuste tende a ser influenciado pela dimensão da amostra e a dimensão da amostra é considerada como pequena já que segundo Gageiro & Pestana (2015) é menor a 30.

Tabela 39 – Sumário do modelo – grupo 3

Model Summary ^c										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	,812 ^a	,660	,639	192,634	,660	31,046	1	16	,000	
2	,873 ^b	,763	,731	166,116	,103	6,516	1	15	,022	2,341

a. Predictors: (Constant), Log_desempregados

b. Predictors: (Constant), Log_desempregados, Poder_de_compra

c. Dependent Variable: Movimentos_pend_autocarro

Na Tabela 40 é possível analisar a influência de cada variável no modelo para o grupo 3 e saber o efeito na previsão ou estimação da variável dependente. Assim, tal como para Grupo 2 as variáveis são diferentes das do Grupo 1 e 2, tendo sido neste caso necessário recorrer à transformação da variável “nº de desempregados” através da utilização de um logaritmo. Importa ainda realçar que para além do nível e importância das variáveis no MRLM ser muito diferente com o “nº de desempregados” a representar uma intensidade em módulo de cerca de 3 vezes a do “poder de compra”, ainda se verifica que o contributo das variáveis é de sinal contrário, ou

seja, o número de viagens aumenta com o nº de desempregados, e diminui com o aumento do poder de compra.

Tabela 40 – Resultados dos coeficientes – grupo 3

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-4939,930	1096,135		-4,507	,000					
	Log_desempregados	1947,775	349,571	,812	5,572	,000	,812	,812	,812	1,000	1,000
2	(Constant)	-4817,072	946,466		-5,090	,000					
	Log_desempregados	2110,213	308,092	,880	6,849	,000	,812	,870	,861	,957	1,045
	Poder_de_compra	-8,169	3,200	-,328	-2,553	,022	-,146	-,550	-,321	,957	1,045

a. Dependent Variable: Movimentos_pend_autocarro

O modelo estimado para os municípios do Grupo 3, com uma densidade populacional maiores a 100 hab/Km², é dado por:

$$VGA = -4817,07 + 2110,213 \text{ Log}(Dsgd) - 8,169 * PdC \quad (31)$$

Onde:

VGA: número de viagens em autocarro;

Dsgd: número de pessoas desempregadas, com Dsgd a variar entre [711; 2368]

PdC: poder de compra; com PdC a variar entre [56,08;105,70]

4.4.5 Validação do modelo – Análise de resíduos e observações influentes

No anexo C é apresentado a tabela com a síntese das hipóteses do MRLM para cada grupo, além é apresentado as análises dos resíduos e observações influentes para cada um e as observações que tem que se tirar dos modelos. Posto que já foi analisado e apresentado que parâmetro se tem que calcular para verificar o cumprimento dos testes e hipóteses é elaborada a Tabela 41 a qual é um excerto da tabela apresentada no anexo C.

Tabela 41 – Excerto de síntesis de diagnóstico

	Condição	GRUPO 1	
		k = 2	n = 68
		Valor/Intervalo	Estado
Linearidade	As variáveis independentes apresentam relação linear com a variável dependente	SnE: R=0,75 ; Dsgd:R=0,66	Duas variáveis apresentam
Homocedasticidade	Os gráficos: Studentized residual Vs. Standardized Predicted; Standardized Residual Vs. Unstandardized Predicted não apresentam tendências crescentes ou decrescentes	-	Cumpe condição
Independência	O valor do teste de Darwin-Watson tem que estar dentro da região de não rejeição da hipótese nula	[1,668;2,332]	D-W = 1,878 cumpe
Normalidade	Os resíduos apresentam um comportamento normal, isto é que a significância do teste de Kolmogorov-Smirnov ou Shapiro-Wilk seja superior a 0,05	K-S: sig > 0,05	sig = 0,041 Não cumpre
Multicolinearidade	As variáveis independentes não estão correlacionadas entre si. Analisa-se o valor do VIF(Variance Inflation Factor)	VIF < 5	VIF - SnE = 1,5 ; VIF - Dsgd = 1,5

5 ANALESES DE RESULTADOS

5.1 Validação do modelo

Para além da validação estatística do MRLM apresentada no ponto 4, referente ao caso de estudo, nesta seção pretende-se analisar os resultados da aplicação do MRLM definido para estimar a procura de transporte em autocarro, no ponto 4.3, para um conjunto de municípios previamente selecionadas da população dos municípios portugueses classificados como Territórios de Baixa Densidade.

O número de municípios utilizados neste processo de validação foram 23, que se apresentam na Tabela 43, com a informação das variáveis dependentes e independentes que caracterizam o MRLM.

Na seção 3.5 apresentou-se o procedimento para realizar a validação do modelo, o qual consiste em estimar o número de viagens em autocarro para cada município utilizando a equação do MRLM determinada na seção 4.3 e comparar esses resultados com o número real de viagens em autocarro, obtidas nos CENSOS 2011, determinando para esse efeito a percentagem de erro para cada município.

Na Tabela 42 apresenta-se a frequência (quantidade) de municípios cujo desvio corresponde a um determinado intervalo de percentagem de erro. Esses intervalos foram definidos com uma amplitude de 5%, podendo mesmo ultrapassar uma taxa superior a 100%. Para além disso, apresentam-se os valores das percentagens acumuladas de municípios por intervalos de erro, sendo possível constatar que 83% dos municípios apresenta um erro igual ou inferior a 50%. E que o 91% dos municípios têm um erro igual ou inferior a 55%.

Tabela 42 – Análise dos resultados da validação (real vs modelado)

Dados de validação	Intervalos de erro (%)	Número de municípios	Percentagem acumuladas de municípios
	0 - 5	2	9%
	5 - 10	0	9%
	10 - 15	4	26%
	15 - 20	2	35%
	20 - 25	2	43%
	25 - 30	5	65%
	30 - 35	3	78%
	35 - 40	1	83%
	40 - 45	0	83%
	45 - 50	0	83%
	50 - 55	2	91%
	55 - 60	0	91%
	60 - 65	0	91%
	65 - 70	0	91%
	70 - 75	0	91%
	75 - 80	0	91%
	80 - 85	0	91%
	85 - 90	0	91%
	90 - 95	1	96%
95 - 100	1	100%	
>100	0	100%	

Na Figura 30 apresenta-se o gráfico com o número de municípios por intervalo de erro, verificando-se que há dois municípios que apresentam um erro de 92% e 99% , Pampilhosa da Serra e Castro Verde, respectivamente. De modo apresentar uma justificação plausível para que se tenham verificado tais resultados foi necessário realizar uma análise dos valores encontrados para esses municípios em relação ao conjunto das variáveis independentes do MRLM, nomeadamente a densidade populacional, o poder de compra e o número de pessoas sem nível de escolaridade. Desta análise foi possível concluir que os 2 municípios apresentam um alto poder de compra em relação aos demais, apesar da variável do poder de compra não ser a mais influente no modelo estimado de acordo com o valor do teste t^3 (Tabela 30).

³ Entre mais cercano o valor do teste t esté de 2 menor é sua influência no modelo. (Gageiro & Pestana, 2015)

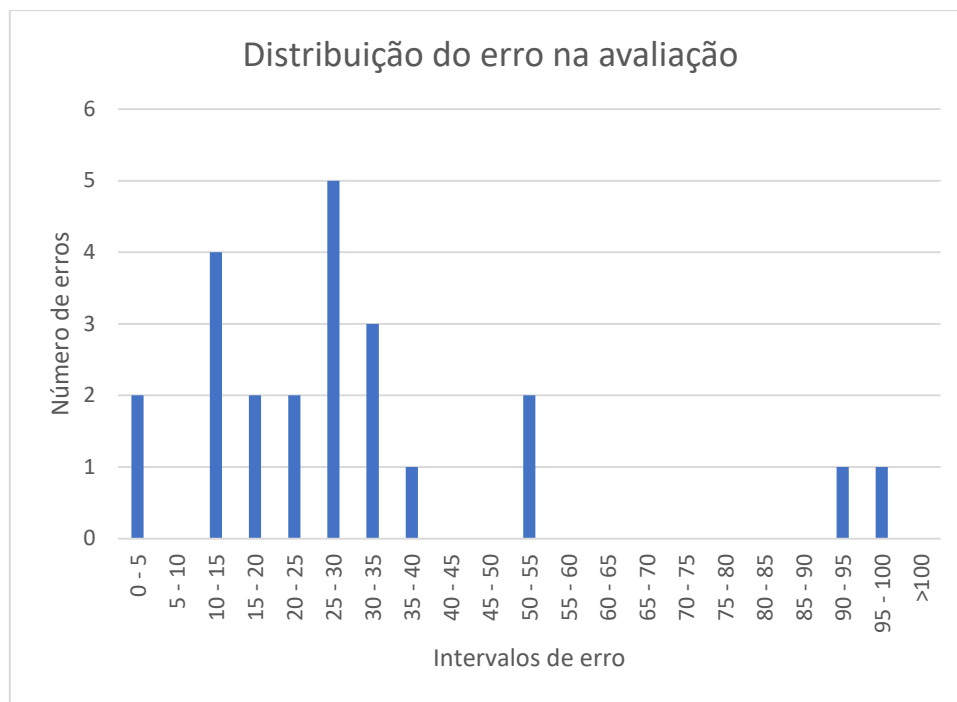


Figura 30 – Gráfico de distribuição do erro

Tabela 43 – Resultados para os municípios utilizados na validação do modelo

ID	Município	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Erro [%]
0904	<i>Figueira de Castelo Rodrigo</i>	6260	246	1598	12	66,35	243	1
0405	<i>Macedo de Cavaleiros</i>	15776	592	3503	23	74,67	576	3
0117	<i>Sever do Vouga</i>	12356	531	2260	95	74,08	591	11
0809	<i>Monchique</i>	6045	238	1490	15	61,26	266	12
1417	<i>Sardoal</i>	3939	160	752	43	73,92	138	14
0409	<i>Torre de Moncorvo</i>	8572	321	2119	16	63,83	369	15
0406	<i>Miranda do Douro</i>	7482	269	1778	15	75,79	225	16
1712	<i>Valpaços</i>	16882	758	4550	31	60,13	893	18
0909	<i>Meda</i>	5202	202	1348	18	62,1	244	21
1702	<i>Boticas</i>	5750	422	1659	18	59,91	315	25
1407	<i>Chamusca</i>	10120	474	2384	14	73	350	26
0104	<i>Arouca</i>	22359	1403	4774	68	69,49	997	29
0505	<i>Idanha-a-Nova</i>	9716	375	3077	7	68,61	484	29
0510	<i>Vila de Rei</i>	3452	110	916	18	65,39	142	29

1705	<i>Mondim de Basto</i>	7493	656	1939	44	58,76	460	30
1708	<i>Peso da Régua</i>	17131	808	3645	181	84,52	1065	32
1709	<i>Ribeira de Pena</i>	6544	325	1971	30	57,33	430	32
0507	<i>Penamacor</i>	5682	222	1705	10	59,94	297	34
1822	<i>Vila Nova de Paiva</i>	5176	211	1347	30	61,25	287	36
0201	<i>Aljustrel</i>	9257	422	2024	20	88,92	201	52
1819	<i>Tabuaço</i>	6350	267	1474	47	55,85	406	52
0612	<i>Pampilhosa da Serra</i>	4481	111	1417	11	65,3	213	92
0206	<i>Castro Verde</i>	7276	391	1578	13	102,41	5	99

5.2 Análise de resultados para o modelo de estimação de viagens em autocarro para os TBDs em Portugal continental.

O número total de municípios denominados como territórios de baixa densidade segundo CIC Portugal 2020 (2015) são 165 para Portugal continental. Assim, tendo em conta a análise da estatística descritiva para cada variável utilizada neste estudo que resultou do estado da arte, foram eliminados 14 municípios por serem considerados outliers, cujo procedimento foi apresentado na seção 4.1.2. É importante destacar que mesmo assim, serão apresentados e analisados os resultados da aplicação do MRLM definido pela equação (15), do ponto 4.3, ou seja, será determinada o erro (%) na seção 5.4 para os 14 municípios considerados outliers numa primeira fase de análise estatística.

Para a seleção das variáveis que foram analisadas no modelo de regressão múltipla foi necessário explorar aquelas que podem ser consideradas mais influentes para a geração de viagens tanto para o transporte público em autocarro convencional, quer mesmo para serviços associados à exploração do transporte flexível. Uma vez selecionadas as variáveis, com base no estado da arte, foi preciso realizar o levantamento da informação referente a essas variáveis no contexto dos 165 municípios portugueses classificados como TBDs, tendo sido esta procura realizada junto de entidades governamentais e não governamentais. Ainda assim importa destacar que a existência ou não de dados foi uma das principais restrições no processo de seleção de variáveis explicativas a utilizar no MRLM.

Segundo Ortuzar (2015), Cal & Mayor (2005) e Transport Planning and Traffic Engineering (2006), a regressão linear múltipla é uns dos modelos matemáticos mais usados para a determinação da geração de viagens no planeamento de transporte.

O modelo para a estimação de viagens em autocarro apresentado foi elaborado com base na metodologia apresentada por Gageiro & Pestana (2015) e Ribeiro (2011). É importante enfatizar que o modelo cumpre todas as hipóteses e verifica todos os testes de inferência estatística (seção 4.3). O número total de municípios com que modelo foi estimado foi 114, devido à análise de outliers e observações influentes que obrigou a retirar 14 municípios e à seleção de um conjunto de cerca de 20 municípios para efetuar a respetiva validação apresentada no ponto 5.1.

O modelo que define a geração de viagens em autocarro para os territórios de baixa densidade foi apresentado em 4.3 e é dado pela expressão (32):

$$VGA = 339,66 + 0,185 * SnE + 3.335 * DP - 6,541 * PdC \quad (32)$$

Onde:

VGA: número de viagens em autocarro

SnE: número de pessoas sem nível de educação; com SnE a variar entre [645;6772]

DP: densidade populacional; com DP a variar entre [6;127]

PdC: poder de compra; com PdC a variar entre [56,08;95,96]

No ANEXO D apresenta-se a tabela com a informação relativa aos 114 municípios utilizados na calibração do modelo e mais os 23 municípios da validação, isto tem como objetivo analisar o comportamento do modelo estimado de viagens para todos os territórios de baixa densidade e de esta maneira concluir se o modelo é relevante para estimar o número de viagens em autocarro.

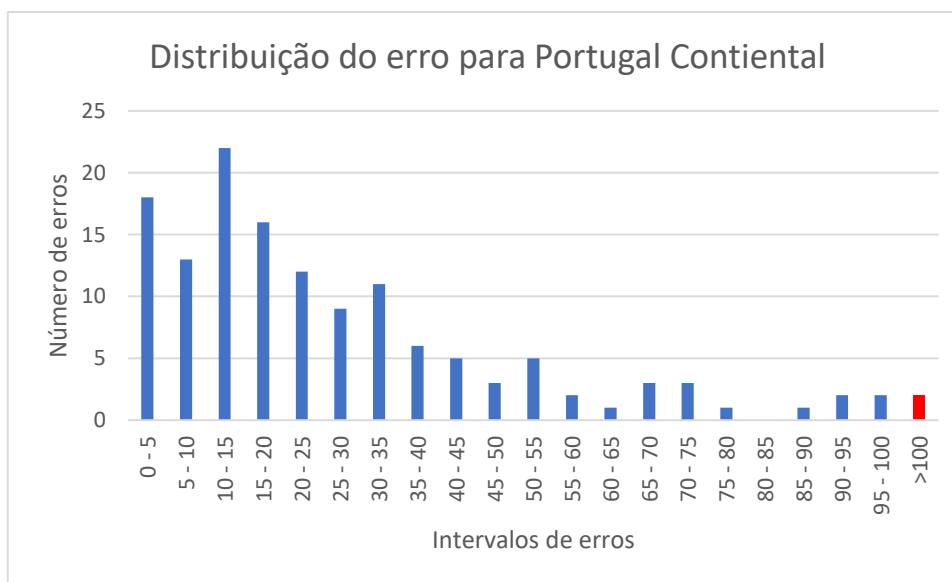


Figura 31 – Distribuição do modelo Portugal continental

A Figura 31 apresenta a distribuição do número de municípios pelos intervalos de erro. Da figura pode-se concluir que a maioria dos resultados apresentam um erro inferior a 35%, i.e., um desvio do valor em relação ao valor real inferior a 35%. No entanto, alguns resultados apresentam um erro superior a 100%, sendo necessária a realização de uma análise isolada de modo a identificar a possível causa do erro.

A Tabela 44 permite analisar de uma maneira mais específica a quantidade de municípios apresentados na figura. Assim, verifica-se que 74% dos municípios, o que equivale a 101 municípios, apresentam uma percentagem de erro inferior ao 35%. Esta tabela também permite determinar a linha crítica a qual corresponde uma percentagem de erro inferior a 50% à qual correspondem cerca de 84% dos municípios dos TBDs.

Tabela 44 – Resultados para Portugal continental

Portugal Continental	Intervalos de erro (%)	Número de municípios	Percentagem acumuladas de municípios
	0 - 5	18	13%
	5 - 10	13	23%
	10 - 15	22	39%
	15 - 20	16	50%
	20 - 25	12	59%
	25 - 30	9	66%
	30 - 35	11	74%
	35 - 40	6	78%
	40 - 45	5	82%
	45 - 50	3	84%
	50 - 55	5	88%
	55 - 60	2	89%
	60 - 65	1	90%
	65 - 70	3	92%
	70 - 75	3	94%
	75 - 80	1	95%
	80 - 85	0	95%
	85 - 90	1	96%
	90 - 95	2	97%
95 - 100	2	99%	
>100	2	100%	

Da análise dos resultados das variáveis independentes para os 8 municípios que apresentam um erro superior a 75%, é possível concluir que eles não apresentam um padrão semelhante a todos os municípios que permita indicar qual o principal motivo para se verificar um erro/ desvio tão elevado. Nestes casos o desvio pode resultar de um efeito combinado das variáveis independentes que está subjacente ao próprio modelo de regressão linear múltipla.

Tabela 45 – Resultados para os 8 municípios com erro superior a 75%

ID	Nome_Município	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1711	Santa Marta de Penaguião	7356	369	1973	106	60,23	666	80
0908	Manteigas	3430	71	645	28	63,91	135	90
0506	Oleiros	5721	139	1624	12	63,52	266	91
0612	Pampilhosa da Serra	4481	111	1417	11	65,3	213	92
0712	Vendas Novas	11846	192	2646	53	95,96	380	98
0206	Castro Verde	7276	391	1578	13	102,41	5	99
1205	Castelo de Vide	3407	107	793	13	82,58	-10	110
1603	Melgaço	9213	217	2443	39	64,17	502	131

Em suma, apesar de alguns desvios serem relevantes na globalidade é possível concluir que as variáveis independentes: densidade populacional, poder de compra e número de pessoas sem nível de escolaridade permitem estimar com alguma razoabilidade a quantidade de viagens em autocarro para os territórios de baixa densidade em Portugal continental.

5.3 Análises de resultados por regiões

Tendo por base o mesmo modelo apresentado no ponto 5.2 para Portugal continental importa perceber se o modelo pode ser aplicado de forma indiferenciada em todo o país, ou se existe um ajuste melhor, o pior, às diferentes regiões do país.

Assim, de acordo com a divisão administrativa de Portugal Continental em Unidades Territoriais para Fins Estatísticos de Nível II (NUTS II), Portugal encontra-se dividido nas regiões do Alentejo, Algarve, Centro, Lisboa e Vale do Tejo e do Norte, sendo que para a análise de TBDs, no âmbito desta dissertação, será apenas apresentada uma análise para região Norte, Centro e a Sul (Algarve e Alentejo), com a região Centro a apresentar 52 municípios de baixa densidade (a maior das três), a região Sul com 44 municípios e a região Norte com 41 municípios classificados como de baixa densidade.

Importa referir que no ANEXO E se apresentam os dados de todos os municípios classificados por região com o respetivo valor das viagens em autocarro estimadas para cada município e compração com o valor real através da apresentação do erro correspondente.

5.3.1 Região Norte

Na Tabela 47 é possível observar que para os 41 municípios de baixa densidade desta região, o modelo estima as viagens para o 88% dos municípios com um erro inferior a 50%. Por outro lado, os municípios que apresentam um desvio maior, i.e., que apresentam um erro superior a 60%, são Santa Marta de Penaguião e Melgaço. Estes dois municípios representam claramente uma exceção em relação ao padrão global. Da análise da informação das variáveis independentes para estes 2 municípios não foi possível definir que variável apresentava um valor mais disruptivo com influência direta no erro associado ao processo de estimação do número de viagens, sendo que o valor modelado é quase o dobro do real.

Tabela 46 – Resultados para os 8 municípios com erro superior a 75%

ID	Município	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1711	Santa Marta de Penaguião	7356	369	1973	106	60,23	666	80
1603	Melgaço	9213	217	2443	39	64,17	502	131

Com base na Figura 32 é possível observar, de uma maneira mais fácil, o comportamento dos municípios e como o modelo MRLM utilizado para estimar as viagens em autocarro apresenta uma maior quantidade de municípios com um erro inferior a 35%, o qual verificando na anterior tabela corresponde ao 83% dos municípios.

Tabela 47 – Distribuição do erro (resultados modelados vs reais) para a região Norte

Região Norte	Intervalos de erro (%)	Número de municípios	Porcentagem acumuladas de municípios
	0 - 5	3	7%
	5 - 10	4	17%
	10 - 15	7	34%
	15 - 20	5	46%
	20 - 25	6	61%
	25 - 30	5	73%
	30 - 35	4	83%
	35 - 40	0	83%
	40 - 45	2	88%
	45 - 50	0	88%
	50 - 55	2	93%
	55 - 60	1	95%
	60 - 65	0	95%
	65 - 70	0	95%
	70 - 75	0	95%
	75 - 80	1	98%
	80 - 85	0	98%
	85 - 90	0	98%
	90 - 95	0	98%
95 - 100	0	98%	
>100	1	100%	

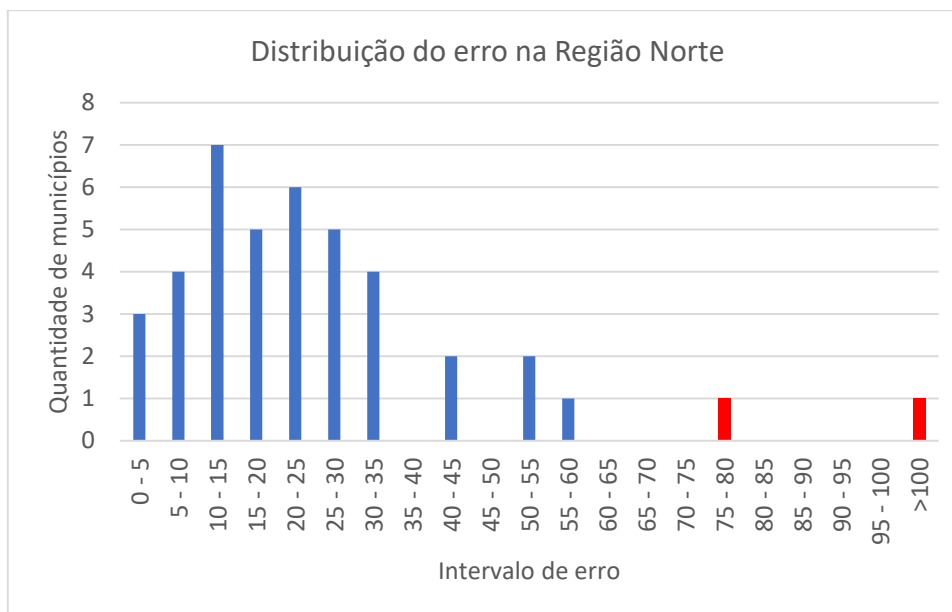


Figura 32 – Distribuição do erro pelos municípios da região norte

5.3.2 Região Centro

A região centro conta com 52 municípios de baixa densidade, é a região com um maior número de municípios com esta classificação. Com base na Tabela 49 é possível constatar que para 81% dos municípios o erro na estimação do número de viagens é menor o igual a 50%. Uma das diferenças mais notórias em relação à região Norte é que tem um maior número de municípios com um erro muito baixo entre 0% e 15% (a região Norte tem 14 municípios enquanto que a região centro tem 22).

Há três municípios que contam com um erro superior a 90%, nomeadamente o de Manteigas, Oleiros e Pampilhosa da Serra. Estes municípios tem uma característica similar, uma vez que todos apresentam um valor de densidade populacional a variar entre 11 a 28 habitantes/Km².

Tabela 48 – Resultados para os 3 municípios com um erro superior a 90%

ID	Município	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0908	Manteigas	3430	71	645	28	63,91	135	90
0506	Oleiros	5721	139	1624	12	63,52	266	91
0612	Pampilhosa da Serra	4481	111	1417	11	65,3	213	92

Tabela 49 – Distribuição do erro (resultados modelados vs reais) para a região Centro

Região do Centro	Intervalos de erro (%)	Número de municípios	Percentagem acumuladas de municípios
	0 - 5	9	17%
	5 - 10	6	29%
	10 - 15	7	42%
	15 - 20	7	56%
	20 - 25	4	63%
	25 - 30	2	67%
	30 - 35	3	73%
	35 - 40	2	77%
	40 - 45	1	79%
	45 - 50	0	79%
	50 - 55	1	81%
	55 - 60	1	83%
	60 - 65	1	85%
	65 - 70	2	88%
	70 - 75	3	94%
	75 - 80	0	94%
	80 - 85	0	94%
	85 - 90	1	96%
90 - 95	2	100%	
95 - 100	0	100%	
>100	0	100%	

Com base na Figura 33 é possível observar que a região do Centro tem um maior número de municípios com um erro menor ou igual a 50%, o que indica que o modelo apresenta uma estimação aceitável, analisando a mesma figura pode-se observar que a distribuição do erro é mais prolongada pelas várias classes dos intervalos, ou seja, para uma percentagem de erro entre 50% e 100% se encontram cerca de 20% dos municípios.

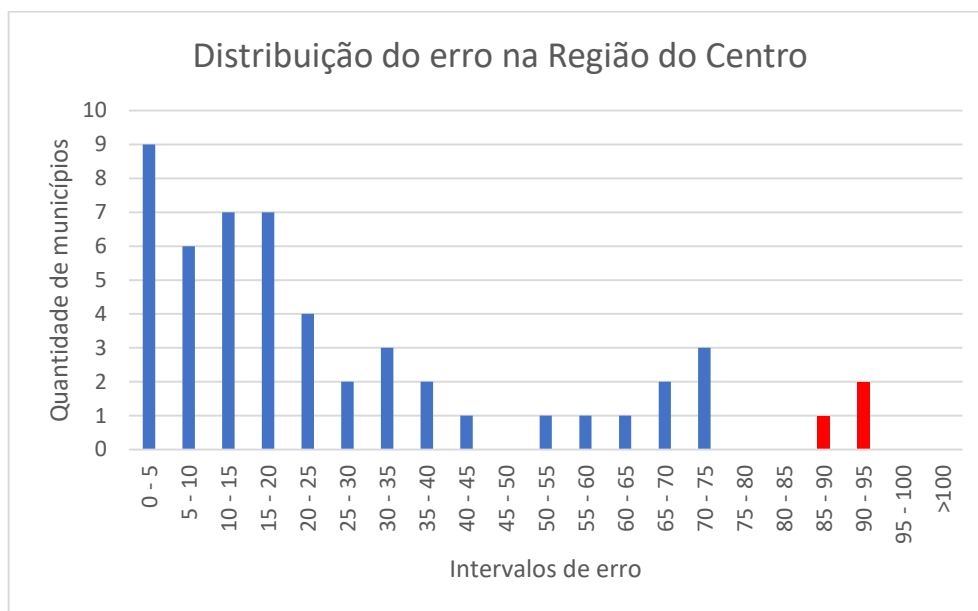


Figura 33 – Distribuição dos municípios na região Centro

5.3.3 Região Sul

A região Sul inclui a região do Alentejo e do Algarve, uma vez que os municípios de baixa densidade para o Algarve são apenas 4, o que não permitira desenvolver uma análise robusta para esta região de uma forma isolada. Com esta agregação, o número de municípios de baixa densidade desta região é no conjunto 44. O erro da estimação do número de viagens para esta região é apresentada na Tabela 51, sendo possível concluir que para 86% dos municípios da região o erro é menor que 50%, ou seja, para os restantes 24% dos municípios o MRLM não efectua uma boa estimação do número de viagens em autocarro.

Os três municípios que apresentaram uma percentagem de erro acima dos 90% foram o Castro Verde, Vendas Novas e Castelo de Vide. É importante sublinhar que os 4 municípios da região de Algarve apresentaram um comportamento similar à região de Alentejo o que permite que para futuros trabalhos se possa considerar conjuntamente estas duas regiões.

Tabela 50 – Resultados para os 3 municípios com um erro superior a 90%

ID	Município	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
0712	Vendas Novas	11846	192	2646	53	95,96	380	98
0206	Castro Verde	7276	391	1578	13	102,41	5	99
1205	Castelo de Vide	3407	107	793	13	82,58	-10	110

Tabela 51 – Distribuição do erro (resultados modelados vs reais) para a região Sul

Região Sul	Intervalos de erro (%)	Número de municípios	Porcentagem acumuladas de municípios
	0 - 5	6	14%
	5 - 10	3	20%
	10 - 15	8	39%
	15 - 20	4	48%
	20 - 25	2	52%
	25 - 30	2	57%
	30 - 35	4	66%
	35 - 40	4	75%
	40 - 45	2	80%
	45 - 50	3	86%
	50 - 55	2	91%
	55 - 60	0	91%
	60 - 65	0	91%
	65 - 70	1	93%
	70 - 75	0	93%
	75 - 80	0	93%
	80 - 85	0	93%
	85 - 90	0	93%
	90 - 95	0	93%
95 - 100	2	98%	
>100	1	100%	

Na Figura 34 apresentam-se as estimações da quantidade de viagens em autocarro para a região Sul, que apresentam um comportamento mais concentrado ao longo do intervalo entre 0% e 55%, sendo que este intervalo corresponde a 91% dos municípios da região.

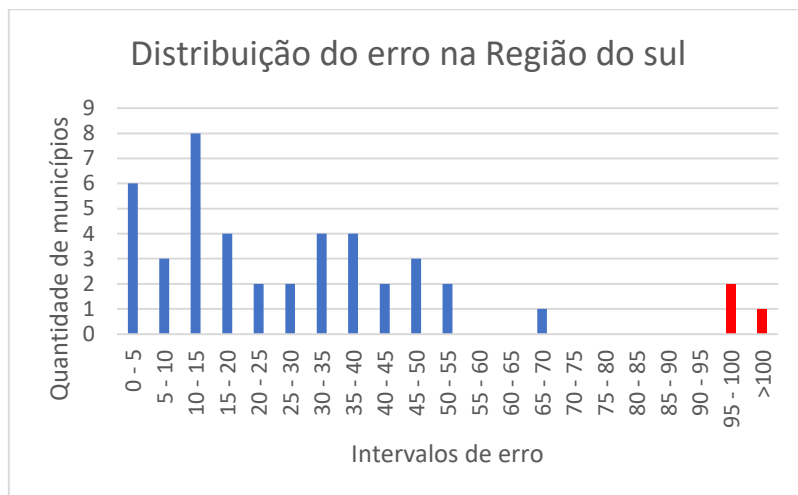


Figura 34 – Distribuição dos municípios na região Sul

5.3.4 Comparações de resultados das três regiões

Nesta seção pretende-se realizar um conjunto de análises dos resultados para as três regiões do Portugal continental (Norte, Centro e Sul (Alentejo e Algarve)), tendo como objetivo de perceber em que região o modelo apresenta as melhores estimativas do número de viagens em autocarro, de modo se poder indicar em que região o MRLM pode ser utilizado para futuros trabalhos.

A Figura 35 foi elaborada com base nas Tabela 47, Tabela 49 e Tabela 51. A análise da Figura permite identificar que a região apresenta maior quantidade de dados com uma percentagem de erro inferior a 50%, que é a região do Norte com 88%, seguindo-se a região Sul com 86% e a região Centro com 79% dos municípios. Para além disso, constat-se que para um erro inferior a 25% a região Centro apresenta 63% dos municípios, seguindo-se a região Norte com 61% e já mais afastada a região Sul com 52%. Em termos gerais, a região onde o modelo de estimação do número de viagens em autocarro apresenta uma estimação mais ajustada, ou seja, com menor erro é a região Norte.

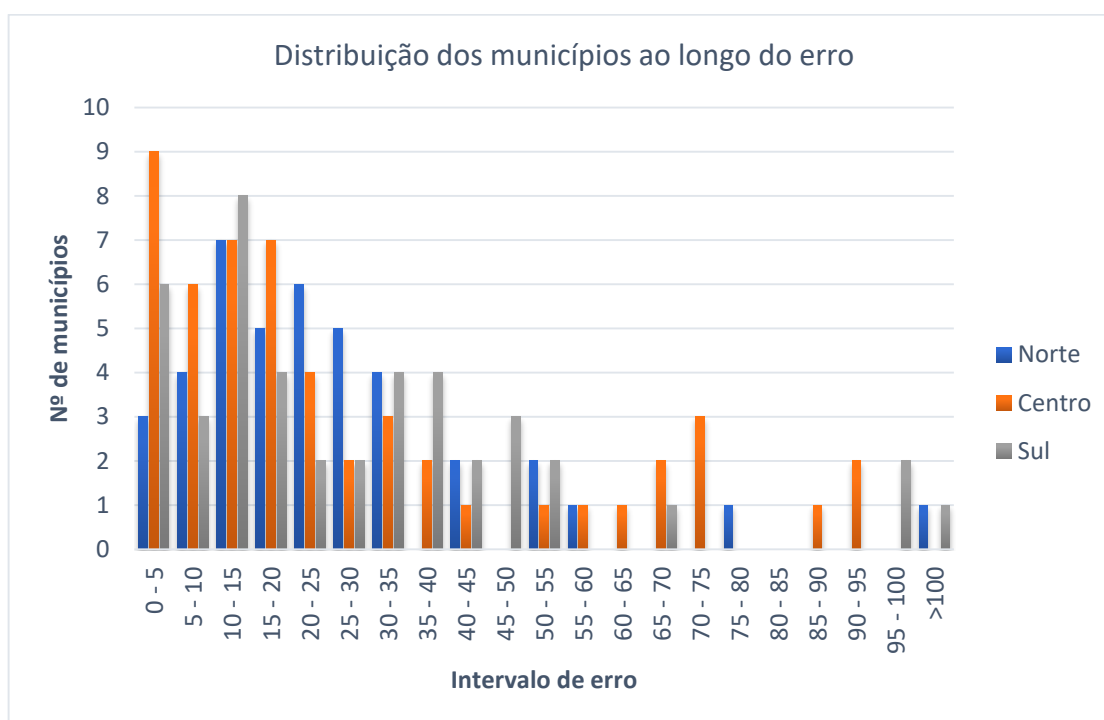


Figura 35 – Distribuição das regiões para Portugal Continental

O diagrama das caixas de bigodes apresentadas na Figura 36, permitem comparar as distribuições do percentagem de erro para cada região, além disso permite identificar os

municípios que são considerados outliers (moderados – círculos; severos - asterisco). Por conseguinte, a região Norte apresenta um comportamento mais uniforme em comparação com as outras regiões, posto que os 50% dos municípios encontram-se num intervalo de erro muito menor do que as outras regiões. Por outro lado, a dispersão do erro na Região Norte é muito inferior à das outras regiões tal como se tinha percebido na análise à Figura 36. Importa, ainda referir que o erro médio, considerando a mediana, é muito semelhante às três regiões com ligeira supremacia da região Sul.

Nas três regiões é possível observar os diferentes tipos de municípios considerados outliers, corroborando as análises dos resultados deste ponto 5.3, e.g., para região Norte tem-se Melgaço na categoria de severo, com um erro superior a 100%, e Santa Marta de Penaguião, com um erro de 80%.

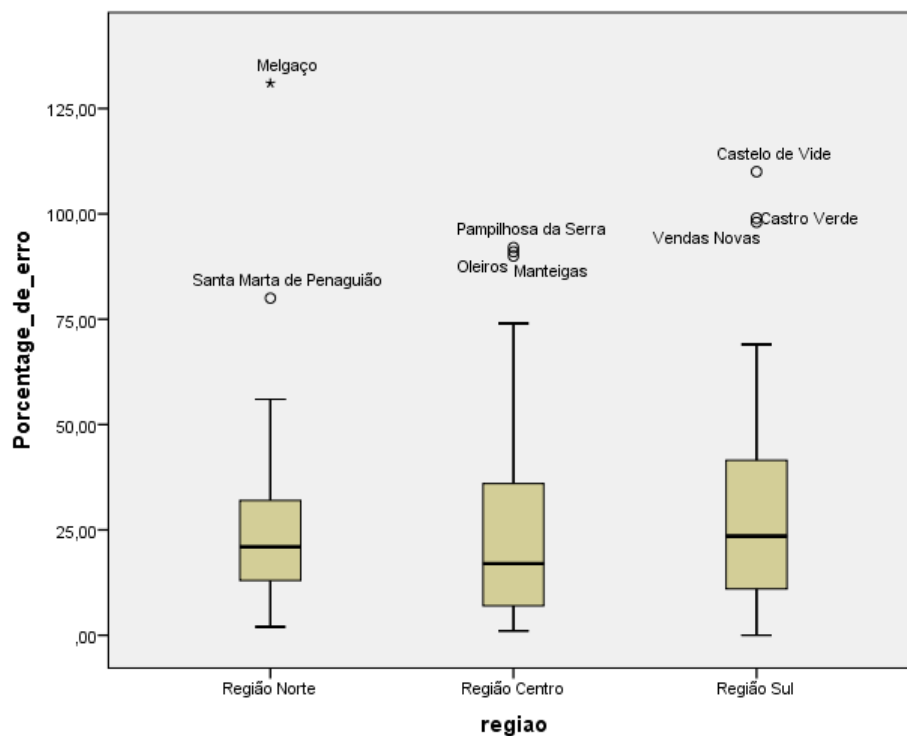


Figura 36 – Caixas de bidoges para regiões

5.4 Resultados para os municípios considerados outliers e influentes

Esta seção visa determinar e analisar o resultado de estimação do número de viagens em autocarro aqueles municípios que foram eliminados, por ser considerados como outliers ou influentes, ao longo do procedimento de análise da estatística descritiva (seção 4.1.2) e da análise de resíduos (seção 4.2.4) para o modelo de regressão linear múltipla.

A Figura 37 apresenta as distribuições do erro para os outliers e os municípios considerados influentes. Com base nesta figura e nos resultados da Tabela 52 é possível determinar que 82% dos municípios outliers apresentam um erro igual ou inferior a 50%. No entanto, importa referir que para se poder fazer uma análise sobre se estes municípios podem ser considerados “verdadeiramente” outliers, ter-se-ia que considerar um modelo MRLM com todos os casos possíveis 165 e comparar com os resultados do modelo obtido nesta dissertação, mas esta análise não faz parte dos objetivos desta dissertação pelo que poderá configurar um potencial trabalho futuro.

Para além disso, da análise da Figura 37 é possível constatar que dois municípios têm uma percentagem de erro superior a 100%, mais propriamente o município são Mesão Frío e Vila Nova da Barquinha.

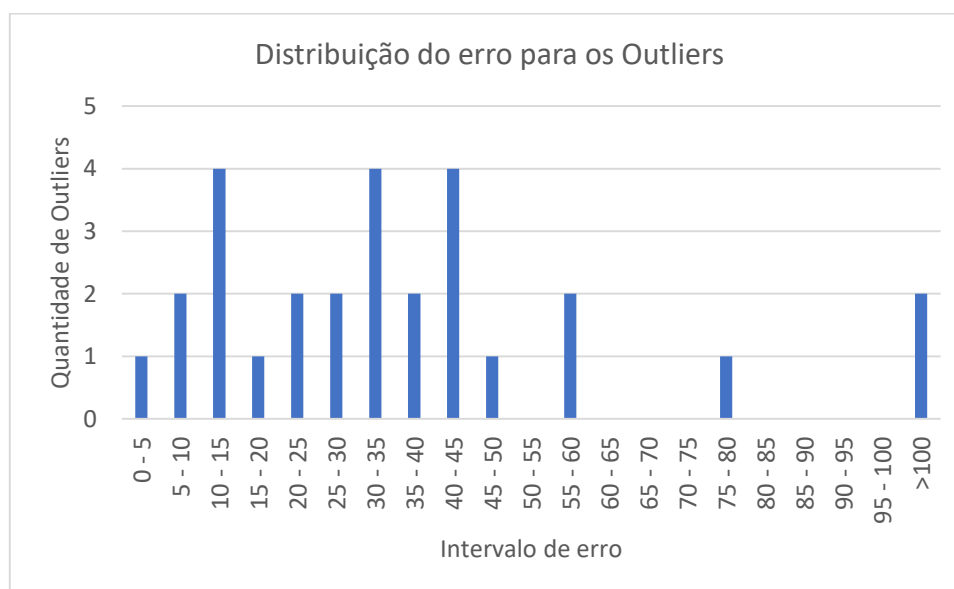


Figura 37 – Distribuição de outliers e municípios influentes

Tabela 52 – Análises de Outliers

Outliers e observações influentes	Intervalos de erro (%)	Número de municípios	Porcentagem acumuladas de municípios
	0 - 5	1	4%
	5 - 10	2	11%
	10 - 15	4	25%
	15 - 20	1	29%
	20 - 25	2	36%
	25 - 30	2	43%
	30 - 35	4	57%
	35 - 40	2	64%
	40 - 45	4	79%
	45 - 50	1	82%
	50 - 55	0	82%
	55 - 60	2	89%
	60 - 65	0	89%
	65 - 70	0	89%
	70 - 75	0	89%
	75 - 80	1	93%
	80 - 85	0	93%
	85 - 90	0	93%
	90 - 95	0	93%
95 - 100	0	93%	
>100	2	100%	

6 CONCLUSÕES E DESENVOLVIMENTOS FUTUROS

Os territórios de baixa densidade (TBD) são atualmente alvo de particular atenção em vários domínios do desenvolvimento territorial, económico e social de regiões e países, nomeadamente no setor da mobilidade e transportes, como é o caso de Portugal.

Os TBD são caracterizados por apresentarem características demográficas, socioeconómicas e de acessibilidade muito específicas, geralmente associados a uma população muito envelhecida e com baixo poder de compra, fruto de uma economia de certa forma débil e centrada no sector primário. Por outro lado, são territórios onde a oferta de serviço e bens de primeira necessidade não estão sempre disponíveis, ou encontram-se a grandes distâncias para a maioria dessas populações, ou seja, a garantia de acesso a esses bens e serviços só é conseguido através da utilização de veículos motorizados, vulgo o automóvel privado.

Deste modo, os TBDs configuram territórios pouco atrativos para a população e sectores empresariais, conduzindo a um ciclo vicioso de abandono dos mesmos e consequentemente a um agravamento dos problemas sociais e desenvolvimento sustentável desses territórios. Este facto, torna-se particularmente evidente no setor da mobilidade e transportes, onde o baixo poder de compra reduz de uma forma efectiva a acessibilidade nesses territórios por falta de infraestruturas e serviços de transporte ajustados às necessidades da população residente.

O transporte público em TBD é, em termos globais, muito deficitário ou inexistente, onde apesar dos índices de riqueza da população não serem em termos gerais muito altos a utilização do veículo automóvel domina o padrão das deslocações nos TBD.

Atualmente, em Portugal com a entrada em vigor da Lei nº 52/2015, de 9 de junho, que aprovou o Regime Jurídico do Serviço Público do Transporte de Passageiros (RJSPTP) tornou ainda mais premente a necessidade de assegurar o transporte nos territórios de baixa densidade, que até hoje, muitos casos, ou é muito deficitário ou não apresenta oferta de transporte público de massas, estando muitas vezes reduzida ao serviço de Táxis e de algumas Instituições Particulares de Solidariedade Social (IPSS). A introdução da necessidade de se garantir serviços mínimos de transporte público em lugares com mais de 40 habitantes poderá alterar o paradigma da mobilidade do serviço público de massas por autocarro de uma forma nunca vista anteriormente, obrigando as entidades gestoras de transporte público e as próprias operadoras

de transporte a planearem um serviço para áreas de maior risco e onde o retorno económico e financeiro poderá ser muito baixo para que se justifique a introdução deste tipo de serviço. Assim, surge a necessidade de se elaborarem modelos de procura de transporte em autocarro para os TBD, que foi o principal objetivo desta dissertação, já que a procura é um dos principais fatores com maior influência no planeamento do sistema de transporte, sobretudo quando a informação sobre a mesma é muito escassa.

Deste modo, nesta dissertação criou-se um modelo de procura de transporte público rodoviário de passageiros para os territórios baixa densidade populacional em Portugal Continental, classificados dessa forma de acordo com CIC Portugal 2020 (2015).

A criação do modelo de procura em autocarro para os TBDs pressupôs a realização de uma pesquisa bibliográfica sobre quais as variáveis que poderiam ter mais influência na geração de viagens para os TBDs, quer fossem territórios urbanos ou rurais. Para além disso, foi preciso identificar os fatores mais relevantes que influenciam a mobilidade dos territórios de baixa densidade. Por último, pesquisou-se sobre a técnica de modelação matemática e o modelo mais usado para determinar a geração de viagens nos modelos de procura de transporte público.

De uma forma sintética, com base na pesquisa bibliográfica, foi possível determinar um conjunto de 11 variáveis (apresentadas na seção 3.2) que apresentam uma importante influência na procura de transporte por autocarro, porém, verificou-se que um dos principais problemas relacionados com muitas dessas variáveis prendia-se com a falta de informação disponível não permitindo então a respetiva análise e utilização no caso de estudo, nomeadamente algumas como a “propriedade de veículos por domicílio” e o “número de viagens realizadas por pessoas do sexo feminino”. Por conseguinte, as variáveis que apresentaram uma influência significativa com o número de viagens em autocarro e para as quais não foram encontrados grandes constrangimentos ao nível da aquisição da informação foram as seguintes:

- Dsgd: Número de pessoas desempregadas;
- PdC: Poder de compra;
- DP: Densidade Populacional;
- IM: Número de mulheres maiores a 15 anos;
- IP: Número de pessoas entre os 15 e 24 anos e maiores a 50 anos;
- GM: Ganho médio mensal;
- SnE: Número de pessoas analfabetas;
- ES: Número de pessoas com educação superior.

Já, em relação ao modelo matemático para a estimação do número de viagens em autocarro foi possível concluir e seleccionar, com base no estado de arte, que o modelo de regressão linear múltipla (MRLM) é o mais utilizado, uma vez que permite perceber e caracterizar como é explicado o comportamento de uma variável dependente em função de um conjunto de variáveis independentes. Para além disso, deste modelo ser o mais usado pelos modeladores na área de planeamento do transporte a facilidade no uso e qualidade no resultados obtidos na maioria dos casos face a modelos mais complexos e uma grande vantagem do MRLM.

Os resultados da modelação do MRLM permitirão identificar que o número de viagens em autocarro para os municípios de baixa densidade para Portugal Continental pode ser explicado, em cerca de 80%, por três variáveis, cujos dados se podem obter diretamente dos Censos. As 3 variáveis explicativas foram: o número de pessoas sem nível de escolaridade, a densidade populacional e o poder de compra.

Da análise do MRLM foi possível concluir que a variável com maior poder explicativo, logo com maior influência, no número de viagens em autocarro nos TBDs é o número de pessoas que não tem escolaridade, ou seja, populações menos instruídas habitualmente estão relacionadas com menores índices de riqueza e um poder de compra mais baixo utilizando os transportes públicos como forma de deslocação nos territórios de baixa densidade.

Por outro lado, importa referir que a variável do poder de compra apresenta um valor negativo para seu coeficiente no MRLM, sendo possível concluir um aumento de está variável poderá conduzir a uma diminuição do número de viagens em autocarro nos municípios dos TBDs.

Com as variáveis de número de pessoas sem nível de escolaridade, densidade populacional e poder de compra o modelo apresentou um ajuste de 87%, tendo em conta que o número (114) de municípios com o qual o modelo foi determinado corresponde a 69% de todos os municípios classificados como TBD, permitindo assegurar uma certa robustez ao MRLM encontrado. Importa referir que o modelo não entrou com toda a população dos municípios em virtude da análise estatística conduzida ao longo do caso de estudo, com a eliminação prévia de 14 valores extremos (outliers), a retirada de 15% (23 municípios) para a realização da validação do modelo, não do ponto de vista estatístico, mas através da comparação dos resultados do modelo com os reais.

Assim, após a validação estatística de MRLM e concluída a calibração do mesmo com a obtenção da respetiva equação que permite estimar o número de viagens de autocarro em TBD, procedeu-se à validação do modelo, com a comparação dos resultados do modelo com os reais, através do cálculo do erro, ou desvio relativo, i.e., a diferença entre o valor modelado e o valor real a dividir pelo valor real, chegando-se à conclusão que o modelo pode ser utilizado para estimar o número de viagens em autocarro para os territórios de baixa densidade dos municípios de Portugal Continental, uma vez que 50% dos municípios apresentam um erro inferior a 20% e o 84% apresenta um erro inferior a 50%.

Para além disso, realizou-se uma análise do ajuste do MRLM por regiões de acordo com a classificação NUTSII, tendo sido possível concluir que a região que apresentou menores desvios/ erro para a estimação do número de pessoas que viajam em autocarro foi a região Norte, uma vez que 84% dos municípios apresentavam um erro inferior ao 50%, mas em termos gerais o modelo ajustou bem para as restantes regiões onde existem TBD, apesar de na região Centro se verificarem frequências mais elevadas para intervalos de erro superiores a 50%.

Por outro lado, para se perceber melhor a relevância e impacto da retirada dos valores extremos na criação do MRLM procedeu-se à análise dos resultados do modelo para os outliers. Desta forma, foi possível concluir que não se verifica um erro significativo, pelo contrario, mostra-se que 82% dos outliers têm um erro inferior a 50%, ou seja, apesar de serem considerados outliers numa análise univariada o feito combinado dos mesmo num MRLM pode-se entrar em processos compensatórios e não se configurarem como verdadeiros outliers. Em suma, o modelo pode ter em conta os outliers eliminados, já que à partida estes não afectam os valores regressão.

Por último, ainda se procedeu a criação de modelos MRLM estratificados segundo a densidade populacional, com o intuito e no pressuposto de quanto mais homogéneos fossemos territórios, melhor seria o ajuste dos MRLM, no entanto tal não se verificou. Assim, é possível concluir que um MRLM global (Portugal Continental) apresenta um ajuste melhor do que vários MRLM estratificados, sobretudo porque os MRLM são muito dependentes da dimensão da amostra e com a estratificação ocorre uma forte redução no número de casos, logo a qualidade do ajuste pode piorar pois torna-se mais sensível a flutuações dos valores das variáveis independentes.

6.1 Trabalhos futuros

Após a conclusão dos trabalhos emergiram da análise dos resultados um vasto conjunto de trabalhos que poderiam complementar o trabalho produzido, que consistem no seguinte:

1 – Determinação da procura potencial de passageiros de autocarro através da utilização das 3 etapas do modelo clássico de 4 etapas e comparar os resultados obtidos com o MRLM obtido nesta dissertação.

2 – Melhorar o atual modelo MRLM com a tentativa de introduzir outras variáveis, nomeadamente “número de proprietários de veículos”, número de pessoas habilitadas a conduzir automóvel e o número de viagens realizadas por pessoas do sexo feminino, tal como sugerido pelo estado da arte.

3 – Estudar a possibilidade de refinar a escala do modelo de procura em autocarro para os TBDs, aumentando a escala para o domínio das freguesias, sendo necessário um trabalho exaustivo de recolha de informação fidedigna sobre a mobilidade, população e território.

4 – Uma vez que o MRLM estima o número de viagens de autocarro em TBD e atendendo à análise estatística desenvolvida nesta dissertação, prevê-se que possa existir um melhor ajuste com os mesmos dados, mas com outras variáveis, utilizando modelos mais apropriados para explicar o comportamento de variáveis dependentes que derivam de contagens (não são funções contínuas) através da aplicação de modelos de regressão assentes em distribuições binomial negativa e de Poisson.

7 REFERÊNCIAS

- Balmer, M. (2007). *Travel demand modeling for multi-agent transport simulations*. ETH Zurich,
- Best, H., & Lanzendorf, M. (2005). Division of labour and gender differences in metropolitan car use: an empirical study in Cologne, Germany. *Journal of Transport Geography*, 13(2), 109-121.
- Cal Mayor, & Asociados. (2005). *Manual de Planeación y Diseño para la Administración del Tránsito y el Transporte. Planeación del transporte urbano*. Bogota, Colombia.
- Comtemporâneo, B. d. D. P. (2017). O que são NUTS? Retrieved from <https://www.pordata.pt/O+que+sao+NUTS>
- Costa, A. d. (2009). Manual de planeamento das Acessibilidades e da Gestão Viária. Transportes Públicos. In: Porto, Comissão de Coordenação e Desenvolvimento do Norte.
- Curtis, C., & Perkins, T. (2006). *Travel Behaviour: A review of recent literature*. Perth, WA: Urbanet, Curtin University of Technology.
- Domingues, A. E. D. (2009). *Transporte Público em Territórios de Baixa Densidade-O Caso de Melgaço–Alto Minho*.
- Fernandes André de Sousa, & Milvia, J. F. F. (2009). *A Problemática da Mobilidade em Espaço Rural e Áreas de Baixa Densidade Urbana: o caso dos concelhos de Mértola e Ourique*. Paper presented at the Anais do I Congresso de Desenvolvimento Regional de Cabo Verde.
- Fischer, M. M. (1992). The Demand for Passenger Transport.
- Flores Reyna, C. O. (2017). Análisis de la metodología clásica del modelo de planificación del transporte urbano desde el ámbito de la movilidad sostenible.
- Gottmann, J. (2012). A evolução do conceito de território. *Boletim Campineiro de Geografia*, 2(3), 523-545.
- Guimarães, R. C., & Cabral, J. A. S. (2010). Estatística. *Centro*, 29(9).
- Herrera, J. A. C. (2014). *Formulación y análisis de modelos de demanda agregada de validaciones y viajes de Transantiago*. Universidad de Chile Santiago de Chile
- IBGE, I. B. d. G. e. E.-. (2017). *Classificação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação* (Vol. 11).
- Instituto Nacional de Estatística, I. (2007). Divisão administrativa. Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_cont_inst&INST=6251013&xlang=pt
- Instituto Nacional de Estatística, I. (2015). *As novas unidades territoriais para fins estatísticos*.
- Kattiyapornpong, U. (2006). Understanding travel behavior using demographic and socioeconomic variables as travel constraints.
- Krigger, D., Pande, A., & Wolshon, B. (2016). Forecasting Travel Demand. *Traffic Engineering Handbook: Institute of Transportation Engineers, Seventh Edition*, 177-202.
- Litman, T. (2017). *Understanding transport demands and elasticities*: Victoria Transport Policy Institute.
- Llanos-Hernández, L. (2010). El concepto del territorio y la investigación en las ciencias sociales. *Agricultura, sociedad y desarrollo*, 7(3), 207-220.
- Luis Ramos, N. A., Ricardo Bento. (2012). *Determinação do conceito e mapeamento dos territórios de muito baixa densidade em Portugal Continental*. Retrieved from
- Maroco, J. (2003). *Análise estatística com utilização do SPSS*.

- Miranda, C. (2013). *Concepções da ruralidade contemporânea: as singularidades brasileiras*: Instituto Interamericano de Cooperação para a Agricultura-IICA.
- Ortúzar, J. d. D. (2015). *Modelos de demanda de transporte*: Alfaomega.
- Ortuzar, J. d. D., & Willumsen, L. G. (1994). Transport Modelling,”. In: John Wiley & Sons Ltd., England.
- Pecqueur, B. (2001). Qualité et développement territorial: l'hypothèse du panier de biens et de services territorialisés. *Économie rurale*, 261(1), 37-49.
- Pereira, J. A. d. O. (2014). Desenvolvimento e validação de modelo de procura de transporte ferroviário urbano.
- Pestana, M. H., & Gageiro, J. N. (2005). *Descobrimo a regressão: com a complementaridade do SPSS*.
- Pestana, M. H., & Gageiro, J. N. (2014). *Análise de dados para ciências sociais: a complementaridade do SPSS. 6 edição*
- República, P. (2016). *Programa de Coesão Territorial* Lisboa: Unidade de Missão para a Valorização do Interior.
- Ribeiro, P. J. G. (2011). *Processo de Selecção de Indicadores Ambientalmente Sustentáveis de Gestão de Tráfego*. Universidade do Porto,
- Santos, M. B., Maria da Saudade. (2016). Os territórios do território português: caracterização e perspetivas para TERRITÓRIOS DE BAIXA DENSIDADE NO HORIZONTE 2030.
- Silva, A. (2010). Estratégias de eficiência colectiva em territórios de baixa densidade: reflexões a propósito do Minho-Lima e do Tâmega. *XII Colóquio Ibérico de Geografia*, 3-8.
- Teixeira, V. F. R. (2002). *Um modelo de procura de transporte público rodoviário de passageiros: aplicação a Portugal e à região do Algarve*.
- TIZON, P. (1995). Le territoire au quotidien. *DI MEO, G. Les territoires du quotidien. Paris: L' harmattan*, 17-34.
- Transporte, B. A. M. S. d. T. y., & Triana, W. F. C. (2005). *Manual de planeación y diseño para la administración del tránsito y el transporte*: Alcaldía Mayor.
- Tsukamoto, M. G. P. M. (2017). *Governança Multiníveis em Territórios de Baixa Densidade: As Comunidades Intermunicipais do Alto Alentejo e da Beira Baixa*.
- UITP, C. B. (2004). Porqué es importante desarrollar un transporte público de baja demanda? Retrieved from <http://uitp.org/mos/corebrief/CB-Low-demand-es.pdf>.
- Velásquez, M. E. S. (2012). *Cómo entender el territorio? :* Editorial Cara Parens, Universidad Rafael Landívar.
- Webster, F., & Bly, P. (1981). The demand for public transport: Part I. the changing environment in which public transport operates. *Transport Reviews*, 1(4), 323-351.
- Balmer, M. (2007). *Travel demand modeling for multi-agent transport simulations*. ETH Zurich,
- Best, H., & Lanzendorf, M. (2005). Division of labour and gender differences in metropolitan car use: an empirical study in Cologne, Germany. *Journal of Transport Geography*, 13(2), 109-121.
- Cal Mayor, & Asociados. (2005). *Manual de Planeación y Diseño para la Administración del Tránsito y el Transporte. Planeación del transporte urbano*. Bogota, Colombia.
- Comtemporâneo, B. d. D. P. (2017). O que são NUTS? Retrieved from <https://www.pordata.pt/O+que+sao+NUTS>
- Costa, A. d. (2009). Manual de planeamento das Acessibilidades e da Gestão Viária. Transportes Públicos. In: Porto, Comissão de Coordenação e Desenvolvimento do Norte.
- Curtis, C., & Perkins, T. (2006). Travel Behaviour: A review of recent literature. *Perth, WA: Urbanet, Curtin University of Technology*.

- Domingues, A. E. D. (2009). *Transporte Público em Territórios de Baixa Densidade-O Caso de Melgaço–Alto Minho*.
- Fernandes André de Sousa, & Milvia, J. F. F. (2009). *A Problemática da Mobilidade em Espaço Rural e Áreas de Baixa Densidade Urbana: o caso dos concelhos de Mértola e Ourique*. Paper presented at the Anais do I Congresso de Desenvolvimento Regional de Cabo Verde.
- Fischer, M. M. (1992). The Demand for Passenger Transport.
- Flores Reyna, C. O. (2017). Análisis de la metodología clásica del modelo de planificación del transporte urbano desde el ámbito de la movilidad sostenible.
- Gottmann, J. (2012). A evolução do conceito de território. *Boletim Campineiro de Geografia*, 2(3), 523-545.
- Guimarães, R. C., & Cabral, J. A. S. (2010). Estatística. *Centro*, 29(9).
- Herrera, J. A. C. (2014). *Formulación y análisis de modelos de demanda agregada de validaciones y viajes de Transantiago*. Universidad de Chile Santiago de Chile
- IBGE, I. B. d. G. e. E.-. (2017). *Classificação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação* (Vol. 11).
- Instituto Nacional de Estatística, I. (2007). Divisão administrativa. Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_cont_inst&INST=6251013&xlang=pt
- Instituto Nacional de Estatística, I. (2015). *As novas unidades territoriais para fins estatísticos*.
- Kattiyapornpong, U. (2006). Understanding travel behavior using demographic and socioeconomic variables as travel constraints.
- Kruger, D., Pande, A., & Wolshon, B. (2016). Forecasting Travel Demand. *Traffic Engineering Handbook: Institute of Transportation Engineers, Seventh Edition*, 177-202.
- Litman, T. (2017). *Understanding transport demands and elasticities*: Victoria Transport Policy Institute.
- Llanos-Hernández, L. (2010). El concepto del territorio y la investigación en las ciencias sociales. *Agricultura, sociedad y desarrollo*, 7(3), 207-220.
- Luis Ramos, N. A., Ricardo Bento. (2012). *Determinação do conceito e mapeamento dos territórios de muito baixa densidade em Portugal Continental*. Retrieved from
- Maroco, J. (2003). *Análise estatística com utilização do SPSS*.
- Ministro adjunto-República Portuguesa , R. P. (2016). *Programa Nacional para a Coesão Territorial Lisboa: Unidade de Missão para a Valorização do Interior*.
- Ortúzar, J. d. D. (2015). *Modelos de demanda de transporte*: Alfaomega:.
- Ortuzar, J. d. D., & Willumsen, L. G. (1994). Transport Modelling,”. In: John Wiley & Sons Ltd., England.
- Pecqueur, B. (2001). Qualité et développement territorial: l'hypothèse du panier de biens et de services territorialisés. *Économie rurale*, 261(1), 37-49.
- Pereira, J. A. d. O. (2014). Desenvolvimento e validação de modelo de procura de transporte ferroviário urbano.
- Pestana, M. H., & Gageiro, J. N. (2005). *Descobrimos a regressão: com a complementaridade do SPSS*.
- Pestana, M. H., & Gageiro, J. N. (2014). Análise de dados para ciências sociais: a complementaridade do SPSS. *6 edição*
- Ribeiro, P. J. G. (2011). *Processo de Seleção de Indicadores Ambientalmente Sustentáveis de Gestão de Tráfego*. Universidade do Porto,
- Santos, M. B., Maria da Saudade. (2016). Os territórios do território português: caracterização e perspetivas para TERRITÓRIOS DE BAIXA DENSIDADE NO HORIZONTE 2030.
- Silva, A. (2010). Estratégias de eficiência colectiva em territórios de baixa densidade: reflexões a propósito do Minho-Lima e do Tâmega. *XII Colóquio Ibérico de Geografia*, 3-8.

- Teixeira, V. F. R. (2002). *Um modelo de procura de transporte público rodoviário de passageiros: aplicação a Portugal e à região do Algarve*.
- TIZON, P. (1995). Le territoire au quotidien. *DI MEO, G. Les territoires du quotidien. Paris: L 'harmattan*, 17-34.
- Transporte, B. A. M. S. d. T. y., & Triana, W. F. C. (2005). *Manual de planeación y diseño para la administración del tránsito y el transporte*: Alcaldía Mayor.
- Tsukamoto, M. G. P. M. (2017). *Governança Multiníveis em Territórios de Baixa Densidade: As Comunidades Intermunicipais do Alto Alentejo e da Beira Baixa*.
- UITP, C. B. (2004). Porqué es importante desarrollar un transporte público de baja demanda? Retrieved from <http://uitp.org/mos/corebrief/CB-Low-demand-es.pdf>.
- Velásquez, M. E. S. (2012). *Cómo entender el territorio?* : Editorial Cara Parens, Universidad Rafael Landívar.
- Webster, F., & Bly, P. (1981). The demand for public transport: Part I. the changing environment in which public transport operates. *Transport Reviews*, 1(4), 323-351.

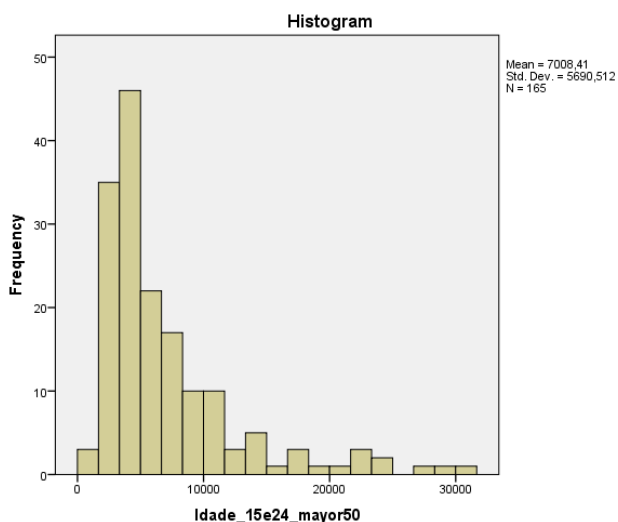
Use the "Insert Citation" button to add citations to this document.

ANEXO A

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Idade 15e24 mayor50	165	100,0%	0	0,0%	165	100,0%

Estatística	
Média	7008,41
Mediana	4885,00
Desvio padrão	5690,512
Variância	32381921,365
Enviesamento (assimetria)	2,035
Std. Error do enviesament	0,189
Curtose (achatamento)	4,288
Std. Error da Curtose	0,376
Amplitude	29266
Mínimo	933
Máximo	30199

Extreme Values				
			Case Number	Value
Idade 15e24 mayor50	Highest	1	37	30199
		2	38	28573
		3	62	28222
		4	127	24751
		5	19	24468
Idade 15e24 mayor50	Lowest	1	6	933
		2	5	1448
		3	65	1470
		4	139	1757
		5	161	1834



Idade_15e24_mayor50 Stem-and-Leaf Plot

Frequency	Stem & Leaf
1.00	0 . 9
7.00	1 . 4478899
20.00	2 . 00011122333456677999
36.00	3 . 001112222233445555567777888999999999
20.00	4 . 00012244555566688889
14.00	5 . 11235666778999
12.00	6 . 011356666688
7.00	7 . 0023367
10.00	8 . 0000136788
6.00	9 . 002368
8.00	10 . 01457899
4.00	11 . 3378
.00	12 .
4.00	13 . 2679
2.00	14 . 35
14.00	Extremes (>=16520)

ANEXO B

Correlations^c

		Movimentos_pend_autocar	Densidade	População	Edade_menor21	Edade_maior65	Edade_entre_15e24	Edade_maior50	Idade_15e24_maior50	Desempregados	Edade_mulheremaior15	Sem_nivel_de_escolaridade_completo	Ensino_básico	Ensino_secundario	Ensino_póssecundario	Ensino_superior	Ganho_medio_mensual	Poder_de_compra
Movimentos_pend_autocar	Pearson Correlation	1	,560**	,872**	,911**	,751**	,912**	,801**	,832**	,813**	,863**	,850**	,894**	,781**	,719**	,737**	,000	,095
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,998	,247
Densidade	Pearson Correlation	,560**	1	,397**	,494**	,218**	,483**	,288**	,330**	,387**	,389**	,333**	,426**	,361**	,341**	,338**	-,031	,011
	Sig. (2-tailed)	,000		,000	,000	,007	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,703	,897
População	Pearson Correlation	,872**	,397**	1	,977**	,953**	,979**	,980**	,991**	,929**	,999**	,976**	,996**	,966**	,898**	,927**	,092	,348**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,263	,000
Edade_menor21	Pearson Correlation	,911**	,494**	,977**	1	,872**	,994**	,917**	,943**	,941**	,968**	,948**	,980**	,936**	,871**	,891**	,088	,335**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,281	,000
Edade_maior65	Pearson Correlation	,751**	,218**	,953**	,872**	1	,884**	,992**	,983**	,838**	,963**	,951**	,942**	,913**	,852**	,882**	,061	,308**
	Sig. (2-tailed)	,000	,007	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,455	,000
Edade_entre_15e24	Pearson Correlation	,912**	,483**	,979**	,994**	,884**	1	,927**	,952**	,939**	,973**	,952**	,984**	,929**	,861**	,888**	,066	,296**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,421	,000
Edade_maior50	Pearson Correlation	,801**	,288**	,980**	,917**	,992**	,927**	1	,997**	,877**	,987**	,966**	,972**	,946**	,879**	,915**	,078	,328**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,340	,000
Idade_15e24_maior50	Pearson Correlation	,832**	,330**	,991**	,943**	,983**	,952**	,997**	1	,900**	,995**	,975**	,985**	,953**	,886**	,920**	,077	,326**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,349	,000
Desempregados	Pearson Correlation	,813**	,387**	,929**	,941**	,838**	,939**	,877**	,900**	1	,919**	,908**	,923**	,904**	,802**	,866**	,111	,370**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,173	,000
Edade_mulheremaior15	Pearson Correlation	,863**	,389**	,999**	,968**	,963**	,973**	,987**	,995**	,919**	1	,976**	,994**	,963**	,896**	,928**	,084	,339**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,307	,000
Sem_nivel_de_escolaridade_completo	Pearson Correlation	,850**	,333**	,976**	,948**	,951**	,952**	,966**	,975**	,908**	,976**	1	,968**	,918**	,853**	,857**	,048	,287**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,562	,000
Ensino_básico	Pearson Correlation	,894**	,426**	,996**	,980**	,942**	,984**	,972**	,985**	,923**	,994**	,968**	1	,946**	,882**	,904**	,076	,305**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,353	,000
Ensino_secundario	Pearson Correlation	,781**	,361**	,966**	,936**	,913**	,929**	,946**	,953**	,904**	,963**	,918**	,946**	1	,930**	,962**	,165	,480**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,043	,000
Ensino_póssecundario	Pearson Correlation	,719**	,341**	,898**	,871**	,852**	,861**	,879**	,886**	,802**	,896**	,853**	,882**	,930**	1	,872**	,144	,429**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,078	,000
Ensino_superior	Pearson Correlation	,737**	,338**	,927**	,891**	,882**	,888**	,915**	,920**	,866**	,928**	,857**	,904**	,962**	,872**	1	,171**	,501**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,036	,000
Ganho_medio_mensual	Pearson Correlation	,000	-,031	,092	,088	,061	,066	,078	,077	,111	,084	,048	,076	,165**	,144	,171*	1	,584**
	Sig. (2-tailed)	,998	,703	,263	,281	,455	,421	,340	,349	,178	,307	,562	,353	,043	,036	,036		,000
Poder_de_compra	Pearson Correlation	,095	,011	,348**	,335**	,308**	,296**	,326**	,326**	,370**	,339**	,287**	,305**	,480**	,429**	,501**	,584**	1
	Sig. (2-tailed)	,247	,897	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N=151

Correlations^c

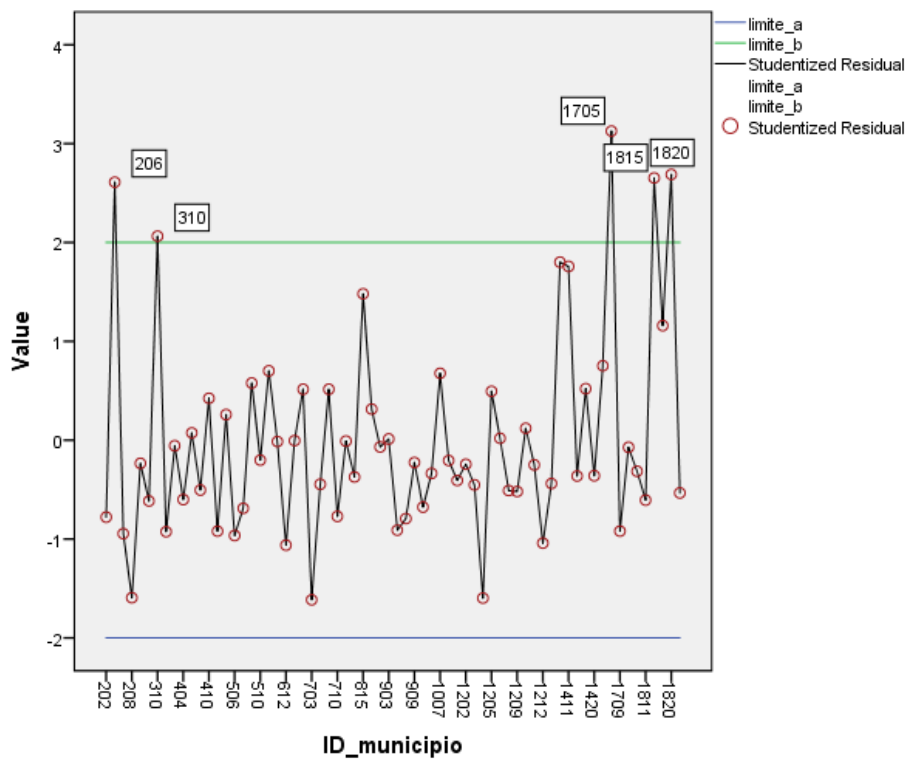
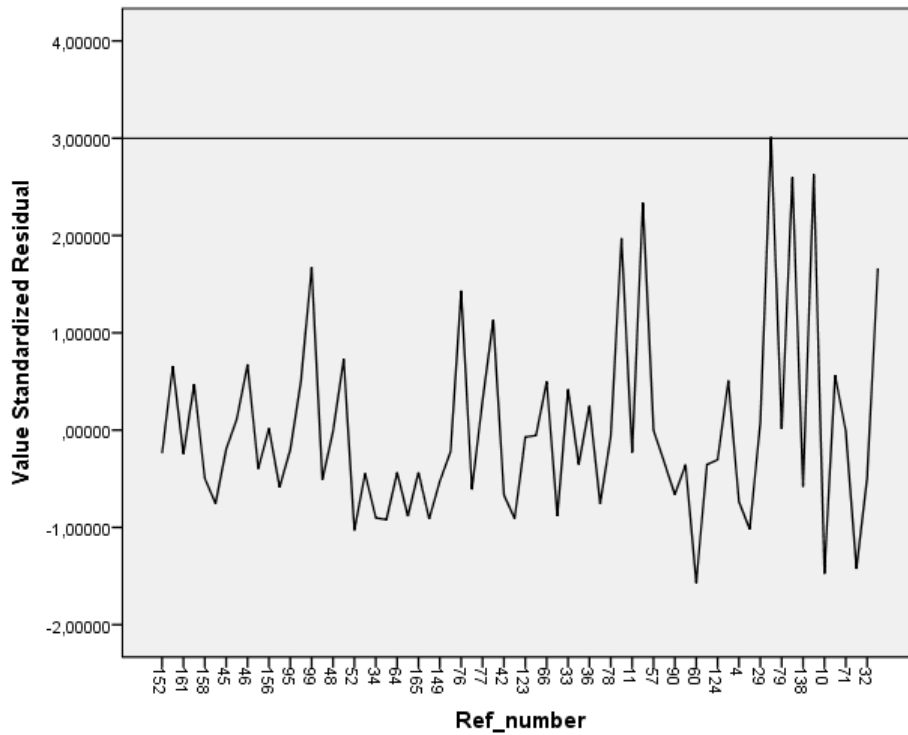
			Movimentos_ pend_autocar ro	Densidade	População	Edade_meno r21	Edade_mayor 65	Edade_entre _15e24	Edade_mayor 50	Idade_15e24 _mayor50	Desemprega dos	Edade_mulh eremayor15	Sem_nivel_d e_escolarida de_completo	Ensino_básic o	Ensiono_sec undario	Ensiono_pós secundario	Ensiono_sup erior	Ganho_medi o_mensual	Poder_de_co mpra
Spearman's rho	Movimentos_ pend_autocar ro	Correlation Coefficient	1,000	,579**	,898**	,919**	,782**	,922**	,836**	,863**	,857**	,887**	,875**	,906**	,849**	,762**	,813**	,110	,073
		Sig. (2-tailed)	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,178
	Densidade	Correlation Coefficient	,579**	1,000	,496**	,590**	,308**	,576**	,387**	,426**	,484**	,485**	,414**	,512**	,507**	,534**	,488**	,031	-,030
		Sig. (2-tailed)	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,701
	População	Correlation Coefficient	,898**	,496**	1,000	,972**	,942**	,979**	,975**	,987**	,923**	,997**	,981**	,997**	,965**	,981**	,952**	,220**	,266**
		Sig. (2-tailed)	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,007	,001
	Edade_menor21	Correlation Coefficient	,919**	,590**	,972**	1,000	,849**	,993**	,903**	,930**	,939**	,960**	,940**	,971**	,956**	,873**	,930**	,248**	,276**
		Sig. (2-tailed)	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,002	,001
	Edade_mayor65	Correlation Coefficient	,782**	,308**	,942**	,849**	1,000	,872**	,990**	,979**	,819**	,956**	,955**	,938**	,884**	,808**	,885**	,140	,222**
		Sig. (2-tailed)	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,087	,006
	Edade_entre_15e24	Correlation Coefficient	,922**	,576**	,979**	,993**	,872**	1,000	,922**	,947**	,936**	,971**	,952**	,980**	,950**	,870**	,930**	,219**	,235**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,007	,004
	Edade_mayor50	Correlation Coefficient	,836**	,387**	,975**	,903**	,990**	,922**	1,000	,996**	,865**	,984**	,975**	,972**	,923**	,843**	,923**	,176	,233**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000	,031	,004
	Idade_15e24_mayor50	Correlation Coefficient	,863**	,428**	,987**	,930**	,979**	,947**	,996**	1,000	,887**	,994**	,983**	,986**	,939**	,858**	,933**	,180	,236**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,027	,004
	Desempregados	Correlation Coefficient	,857**	,484**	,923**	,939**	,819**	,936**	,865**	,887**	1,000	,911**	,901**	,916**	,922**	,820**	,893**	,297**	,327**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,000	,000
	Edade_mulheremayor15	Correlation Coefficient	,887**	,485**	,997**	,960**	,956**	,971**	,984**	,994**	,911**	1,000	,983**	,996**	,957**	,879**	,948**	,203	,257**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,013	,001
	Sem_nivel_de_escolarid ade_completo	Correlation Coefficient	,875**	,414**	,981**	,940**	,955**	,952**	,975**	,983**	,901**	,983**	1,000	,975**	,922**	,829**	,903**	,158	,212**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,052	,009
	Ensino_básico	Correlation Coefficient	,906**	,512**	,997**	,971**	,938**	,980**	,972**	,986**	,916**	,996**	,975**	1,000	,954**	,871**	,941**	,202	,241**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,013	,003
	Ensiono_secundario	Correlation Coefficient	,849**	,507**	,965**	,956**	,884**	,950**	,923**	,939**	,922**	,957**	,922**	,954**	1,000	,917**	,974**	,327**	,390**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000
	Ensiono_póssecundario	Correlation Coefficient	,762**	,534**	,881**	,873**	,808**	,870**	,843**	,858**	,820**	,879**	,829**	,871**	,917**	1,000	,911**	,289**	,362**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,000
	Ensiono_superior	Correlation Coefficient	,813**	,488**	,952**	,930**	,885**	,930**	,923**	,933**	,893**	,948**	,903**	,941**	,974**	,911**	1,000	,344**	,405**
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000
	Ganho_medio_mensual	Correlation Coefficient	,110	,031	,220**	,248**	,140	,219**	,176	,180	,297**	,203	,158	,202	,327**	,289**	,344**	1,000	,613**
		Sig. (2-tailed)	,178	,701	,007	,002	,087	,007	,031	,027	,000	,013	,052	,013	,000	,000	,000	.	,000
	Poder_de_compra	Correlation Coefficient	,073	-,030	,266**	,276**	,222**	,235**	,233**	,236**	,327**	,257**	,212**	,241**	,390**	,362**	,405**	,613**	1,000
		Sig. (2-tailed)	,376	,713	,001	,001	,006	,004	,004	,004	,000	,001	,009	,003	,000	,000	,000	,000	.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 151

ANEXO C



Anexo A –

	Condição	GRUPO 1		GRUPO 2		GRUPO 3	
		k = 2	n = 68	k = 2	n = 36	k = 2	n = 18
		Valor/Intervalo	Estado	Valor	Estado	Valor	Estado
Linearidade	As variáveis independentes apresentam relação linear com a variável dependente	SnE: R=0,75 ; Dsgd:R=0,66	Duas variáveis apresentam relação linear	IM: R=0,56 ; PdC: R= 0,22	Só uma variável apresenta relação moderada	Dsgd: R=0,81 ; PdC: R=0,14	Só uma variável apresenta relação linear
Homocedasticidade	Os gráficos: Studentized residual Vs. Standardized Predicted; Standardized Residual Vs. Unstandardized Predicted não apresentam tendências crescentes ou decrescentes	-	Cumpe condição	-	Cumpe condição	-	Cumpe condição
Independência	O valor do teste de Darwin-Watson tem que estar dentro da região de não rejeição da hipótese nula	[1,668;2,332]	D-W = 1,878 cumpe	[1,587;2,413]	D-W = 1,800 cumpe	[1,535;2,465]	D-W = 2,341 cumpe
Normalidade	Os resíduos apresentam um comportamento normal, isto é que a significância do teste de Kolmogorov-Smirnov ou Shapiro-Wilk seja superior a 0,05	K-S: sig > 0,05	sig = 0,041 Não cumpe	S-W: sig > 0,05	sig = 0,104 cumpe	S-W: sig > 0,05	sig = 0,2 cumpe
Multicolinearidade	As variáveis independentes não estão correlacionadas entre si. Analisa-se o valor do VIF(Variance Inflation Factor)	VIF < 5	VIF -SnE = 1,5 ; VIF - Dsgd = 1,5	VIF < 5	VIF -IM = 1,048 ; VIF -PdC = 1,048	VIF < 5	VIF - Dsg = 1,045 ; VIF - PdC = 1,045

ANEXO D

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
1215	Sousel	5074	191	1325	18	69,45	192	0
0904	Figueira de Castelo Rodrigo	6260	246	1598	12	66,35	243	1
1509	Santiago do Cacém	29749	989	6388	28	94,98	997	1
1816	São Pedro do Sul	16851	739	3737	48	68,28	747	1
1210	Marvão	3512	143	882	23	66,29	145	2
0407	Mirandela	23850	877	5143	36	84,51	861	2
0912	Seia	24702	916	4898	57	77,25	932	2
0405	Macedo de Cavaleiros	15776	592	3503	23	74,67	576	3
1806	Mangualde	19880	832	4030	91	81,94	853	3
0501	Belmonte	6859	334	1501	58	74,92	321	4
1007	Castanheira de Pêra	3191	171	731	48	69,78	178	4
0606	Góis	4260	172	1126	16	64,86	178	4
1706	Montalegre	10537	464	2776	13	63,65	482	4
0710	Redondo	7031	254	1677	19	71,97	243	4
1814	Santa Comba Dão	11597	651	2461	104	71,12	676	4
0713	Viana do Alentejo	5743	157	1500	15	78,88	151	4
0910	Pinhel	9627	398	2267	20	62,39	418	5
0213	Serpa	15623	640	4092	14	72,56	671	5
0903	Celorico da Beira	7693	345	1939	31	67,05	364	6

Anexo A –

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1411	Ferreira do Zêzere	8619	477	2176	45	67,91	450	6
0906	Gouveia	14046	586	3061	47	67,6	621	6
0913	Trancoso	9878	399	2328	27	66,97	424	6
1813	Resende	11364	911	3113	92	58,16	844	7
1409	Coruche	19944	804	5296	18	78,39	869	8
1606	Ponte da Barca	12061	635	2975	66	64,78	688	8
1817	Sátão	12444	641	2980	62	61,71	694	8
1820	Tarouca	8048	585	1821	80	62,67	535	8
0803	Aljezur	5884	230	1449	18	64,1	250	9
0601	Arganil	12145	595	2852	37	68,46	542	9
0711	Reguengos de Monsaraz	10828	365	2681	23	89,8	327	10
0410	Vila Flor	6697	328	1537	25	63,07	296	10
1302	Baião	20522	1528	5441	118	57,14	1367	11
0403	Carraceda de Ansiães	6373	281	1616	23	61,66	312	11
0208	Ferreira do Alentejo	8255	310	2096	13	75,43	277	11
1505	Grândola	14826	454	3680	18	88,15	505	11
0209	Mértola	7274	318	1945	6	66,78	282	11
0117	Sever do Vouga	12356	531	2260	95	74,08	591	11
0901	Aguiar da Beira	5473	263	1605	27	66,05	294	12
0202	Almodôvar	7449	284	2200	10	80,83	251	12
1204	Campo Maior	8456	244	2063	34	95,16	214	12
0305	Celorico de Basto	20098	1374	4695	111	56,08	1214	12
0809	Monchique	6045	238	1490	15	61,26	266	12
1821	Tondela	28946	1396	6013	78	74,07	1230	12
1801	Armamar	6297	413	1410	54	64,1	361	13

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0504	Fundão	29213	1081	6772	42	77,93	1225	13
1212	Nisa	7450	213	1871	13	74,7	241	13
0508	Proença-a-Nova	8314	363	1938	21	69,33	316	13
1804	Cinfães	20427	1398	5132	85	56,32	1208	14
1417	Sardoal	3939	160	752	43	73,92	138	14
1413	Mação	7338	239	1767	18	69,26	276	15
1807	Moimenta da Beira	10212	574	2315	46	66,75	487	15
0409	Torre de Moncorvo	8572	321	2119	16	63,83	369	15
0714	Vila Viçosa	8319	300	1775	43	84,86	256	15
0406	Miranda do Douro	7482	269	1778	15	75,79	225	16
0611	Oliveira do Hospital	20855	1159	4438	89	74,86	969	16
0212	Ourique	5389	180	1457	8	74,14	152	16
0509	Sertã	15880	817	3874	36	75,12	685	16
0311	Vieira do Minho	13001	754	2948	60	69,03	633	16
1803	Castro Daire	15339	975	4069	41	64,39	808	17
1209	Gavião	4132	138	1048	14	71,16	115	17
1013	Pedrógão Grande	3915	150	979	30	68,24	176	17
0614	Penela	5983	353	1384	44	69,22	292	17
1501	Alcácer do Sal	13046	550	3340	9	81,75	453	18
0607	Lousã	17604	1007	3234	127	81,7	829	18
1213	Ponte de Sor	16722	509	4070	20	85,92	599	18
0616	Tábua	12071	720	2665	60	68,54	587	18
1712	Valpaços	16882	758	4550	31	60,13	893	18
1707	Murça	5952	268	1490	31	61,42	319	19
1710	Sabrosa	6361	294	1585	41	63,79	351	20

Anexo A –

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0909	Meda	5202	202	1348	18	62,1	244	21
1818	Sernancelhe	5671	352	1312	25	59,23	278	21
1610	Vila Nova de Cerveira	9253	376	1998	85	82,74	453	21
0902	Almeida	7242	221	1504	14	75,42	172	22
0615	Soure	19245	807	4686	73	71,75	981	22
1713	Vila Pouca de Aguiar	13187	750	3011	30	63,13	586	22
0804	Castro Marim	6747	312	1523	22	69,69	241	23
1701	Alijó	11942	444	2744	40	66,02	550	24
1802	Carregal do Sal	9835	461	2206	84	70,1	571	24
0706	Montemor-o-Novo	17437	503	4355	14	87,26	623	24
1702	Boticas	5750	422	1659	18	59,91	315	25
0304	Cabeceiras de Basto	16710	1186	4028	69	65,16	891	25
1407	Chamusca	10120	474	2384	14	73	350	26
0704	Estremoz	14318	343	3355	28	95,22	432	26
0408	Mogadouro	9542	304	2398	13	67,65	383	26
1605	Paredes de Coura	9198	474	2483	67	65,01	597	26
0104	Arouca	22359	1403	4774	68	69,49	997	29
0505	Idanha-a-Nova	9716	375	3077	7	68,61	484	29
0510	Vila de Rei	3452	110	916	18	65,39	142	29
1705	Mondim de Basto	7493	656	1939	44	58,76	460	30
0914	Vila Nova de Foz Côa	7312	203	1697	18	69,1	264	30
0401	Alfândega da Fé	5104	152	1242	16	64,56	201	32
1708	Peso da Régua	17131	808	3645	181	84,52	1065	32
1709	Ribeira de Pena	6544	325	1971	30	57,33	430	32
0815	Vila do Bispo	5258	347	1054	29	60,94	235	32

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0701	<i>Alandroal</i>	5843	345	1499	11	64,61	231	33
1815	<i>São João da Pesqueira</i>	7874	551	1915	30	65	368	33
1808	<i>Mortágua</i>	9607	282	2104	38	73,23	378	34
1810	<i>Oliveira de Frades</i>	10261	654	2072	71	80,54	432	34
0507	<i>Penamacor</i>	5682	222	1705	10	59,94	297	34
0207	<i>Cuba</i>	4878	152	1101	28	66,36	204	35
0709	<i>Portel</i>	6428	387	1616	11	64,96	250	35
0702	<i>Arraiolos</i>	7363	341	1723	11	72,67	220	36
1809	<i>Nelas</i>	14037	550	2887	112	76,25	749	36
0214	<i>Vidigueira</i>	5932	308	1459	19	72,84	196	36
1822	<i>Vila Nova de Paiva</i>	5176	211	1347	30	61,25	287	36
1208	<i>Fronteira</i>	3410	96	894	14	75,32	58	39
0210	<i>Moura</i>	15167	466	4154	16	77,8	654	40
0310	<i>Terras de Bouro</i>	7253	534	1647	26	63,97	314	41
1002	<i>Alvaiázere</i>	7287	279	1860	45	66,9	398	43
1203	<i>Avis</i>	4571	167	1208	8	75,36	96	43
1201	<i>Alter do Chão</i>	3562	94	898	10	74,39	52	44
0404	<i>Freixo de Espada à Cinta</i>	3780	117	1052	16	63,64	170	45
1206	<i>Crato</i>	3708	135	931	9	71,85	73	46
1202	<i>Arronches</i>	3165	106	853	10	72,8	55	48
0707	<i>Mora</i>	4978	150	1286	11	82,42	76	49
0201	<i>Aljustrel</i>	9257	422	2024	20	88,92	201	52
1211	<i>Monforte</i>	3329	120	1035	8	76,51	57	52
1819	<i>Tabuaço</i>	6350	267	1474	47	55,85	406	52
1811	<i>Penalva do Castelo</i>	7956	357	2136	59	58,59	550	54

Anexo A –

ID	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0412	Vinhais	9066	266	2210	13	58,25	412	55
0411	Vimioso	4669	123	1240	10	62,6	192	56
1003	Ansião	13092	438	3145	75	72,84	695	59
1008	Figueiró dos Vinhos	6169	179	1452	36	67,24	288	61
0911	Sabugal	12544	367	3435	15	63,67	611	67
0703	Borba	7333	209	1811	51	74,96	354	69
0617	Vila Nova de Poiares	7281	264	1496	86	69,52	450	70
0905	Fornos de Algodres	4989	177	1217	38	59,56	302	71
0511	Vila Velha de Ródão	3521	135	787	11	73,69	39	71
1408	Constância	4056	215	851	51	93,21	56	74
1711	Santa Marta de Penaguião	7356	369	1973	106	60,23	666	80
0908	Manteigas	3430	71	645	28	63,91	135	90
0506	Oleiros	5721	139	1624	12	63,52	266	91
0612	Pampilhosa da Serra	4481	111	1417	11	65,3	213	92
0712	Vendas Novas	11846	192	2646	53	95,96	380	98
0206	Castro Verde	7276	391	1578	13	102,41	5	99
1205	Castelo de Vide	3407	107	793	13	82,58	-10	110
1603	Melgaço	9213	217	2443	39	64,17	502	131

ANEXO E

REGIÃO NORTE								
ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
0407	<i>Mirandela</i>	23850	877	5143	36	84,51	861	2
0405	<i>Macedo de Cavaleiros</i>	15776	592	3503	23	74,67	576	3
1706	<i>Montalegre</i>	10537	464	2776	13	63,65	482	4
1813	<i>Resende</i>	11364	911	3113	92	58,16	844	7
1606	<i>Ponte da Barca</i>	12061	635	2975	66	64,78	688	8
1820	<i>Tarouca</i>	8048	585	1821	80	62,67	535	8
0410	<i>Vila Flor</i>	6697	328	1537	25	63,07	296	10
1302	<i>Baião</i>	20522	1528	5441	118	57,14	1367	11
0403	<i>Carraceda de Ansiães</i>	6373	281	1616	23	61,66	312	11
0305	<i>Celorico de Basto</i>	20098	1374	4695	111	56,08	1214	12
1801	<i>Armamar</i>	6297	413	1410	54	64,1	361	13
1804	<i>Cinfães</i>	20427	1398	5132	85	56,32	1208	14
1807	<i>Moimenta da Beira</i>	10212	574	2315	46	66,75	487	15
0409	<i>Torre de Moncorvo</i>	8572	321	2119	16	63,83	369	15
0406	<i>Miranda do Douro</i>	7482	269	1778	15	75,79	225	16
0311	<i>Vieira do Minho</i>	13001	754	2948	60	69,03	633	16
1712	<i>Valpaços</i>	16882	758	4550	31	60,13	893	18
1707	<i>Murça</i>	5952	268	1490	31	61,42	319	19
1710	<i>Sabrosa</i>	6361	294	1585	41	63,79	351	20
1818	<i>Sernancelhe</i>	5671	352	1312	25	59,23	278	21

Anexo A –

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1610	Vila Nova de Cerveira	9253	376	1998	85	82,74	453	21
1713	Vila Pouca de Aguiar	13187	750	3011	30	63,13	586	22
1701	Alijó	11942	444	2744	40	66,02	550	24
1702	Botícas	5750	422	1659	18	59,91	315	25
0304	Cabeceiras de Basto	16710	1186	4028	69	65,16	891	25
0408	Mogadouro	9542	304	2398	13	67,65	383	26
1605	Paredes de Coura	9198	474	2483	67	65,01	597	26
0104	Arouca	22359	1403	4774	68	69,49	997	29
1705	Mondim de Basto	7493	656	1939	44	58,76	460	30
0914	Vila Nova de Foz Côa	7312	203	1697	18	69,1	264	30
0401	Alfândega da Fé	5104	152	1242	16	64,56	201	32
1708	Peso da Régua	17131	808	3645	181	84,52	1065	32
1709	Ribeira de Pena	6544	325	1971	30	57,33	430	32
1815	São João da Pesqueira	7874	551	1915	30	65	368	33
0310	Terras de Bouro	7253	534	1647	26	63,97	314	41
0404	Freixo de Espada à Cinta	3780	117	1052	16	63,64	170	45
1819	Tabuaço	6350	267	1474	47	55,85	406	52
0412	Vinhais	9066	266	2210	13	58,25	412	55
0411	Vimioso	4669	123	1240	10	62,6	192	56
1711	Santa Marta de Penaguião	7356	369	1973	106	60,23	666	80
1603	Melgaço	9213	217	2443	39	64,17	502	131

REGIÃO CENTRO

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
0904	<i>Figueira de Castelo Rodrigo</i>	6260	246	1598	12	66,35	243	1
1816	<i>São Pedro do Sul</i>	16851	739	3737	48	68,28	747	1
0912	<i>Seia</i>	24702	916	4898	57	77,25	932	2
1806	<i>Mangualde</i>	19880	832	4030	91	81,94	853	3
0501	<i>Belmonte</i>	6859	334	1501	58	74,92	321	4
1007	<i>Castanheira de Pêra</i>	3191	171	731	48	69,78	178	4
0606	<i>Góis</i>	4260	172	1126	16	64,86	178	4
1814	<i>Santa Comba Dão</i>	11597	651	2461	104	71,12	676	4
0910	<i>Pinhel</i>	9627	398	2267	20	62,39	418	5
0903	<i>Celorico da Beira</i>	7693	345	1939	31	67,05	364	6
1411	<i>Ferreira do Zêzere</i>	8619	477	2176	45	67,91	450	6
0906	<i>Gouveia</i>	14046	586	3061	47	67,6	621	6
0913	<i>Trancoso</i>	9878	399	2328	27	66,97	424	6
1817	<i>Sátão</i>	12444	641	2980	62	61,71	694	8
0601	<i>Arganil</i>	12145	595	2852	37	68,46	542	9
0117	<i>Sever do Vouga</i>	12356	531	2260	95	74,08	591	11
0901	<i>Aguiar da Beira</i>	5473	263	1605	27	66,05	294	12
1821	<i>Tondela</i>	28946	1396	6013	78	74,07	1230	12

Anexo A –

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentage de erro [%e]
0504	Fundão	29213	1081	6772	42	77,93	1225	13
0508	Proença-a-Nova	8314	363	1938	21	69,33	316	13
1417	Sardoal	3939	160	752	43	73,92	138	14
1413	Mação	7338	239	1767	18	69,26	276	15
0611	Oliveira do Hospital	20855	1159	4438	89	74,86	969	16
0509	Sertã	15880	817	3874	36	75,12	685	16
1803	Castro Daire	15339	975	4069	41	64,39	808	17
1013	Pedrógão Grande	3915	150	979	30	68,24	176	17
0614	Penela	5983	353	1384	44	69,22	292	17
0607	Lousã	17604	1007	3234	127	81,7	829	18
0616	Tábua	12071	720	2665	60	68,54	587	18
0909	Meda	5202	202	1348	18	62,1	244	21
0902	Almeida	7242	221	1504	14	75,42	172	22
0615	Soure	19245	807	4686	73	71,75	981	22
1802	Carregal do Sal	9835	461	2206	84	70,1	571	24
0505	Idanha-a-Nova	9716	375	3077	7	68,61	484	29
0510	Vila de Rei	3452	110	916	18	65,39	142	29
1808	Mortágua	9607	282	2104	38	73,23	378	34
1810	Oliveira de Frades	10261	654	2072	71	80,54	432	34
0507	Penamacor	5682	222	1705	10	59,94	297	34

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1809	Nelas	14037	550	2887	112	76,25	749	36
1822	Vila Nova de Paiva	5176	211	1347	30	61,25	287	36
1002	Alvaiázere	7287	279	1860	45	66,9	398	43
1811	Penalva do Castelo	7956	357	2136	59	58,59	550	54
1003	Ansião	13092	438	3145	75	72,84	695	59
1008	Figueiró dos Vinhos	6169	179	1452	36	67,24	288	61
0911	Sabugal	12544	367	3435	15	63,67	611	67
0617	Vila Nova de Poiares	7281	264	1496	86	69,52	450	70
0905	Fornos de Algodres	4989	177	1217	38	59,56	302	71
0511	Vila Velha de Ródão	3521	135	787	11	73,69	39	71
1408	Constância	4056	215	851	51	93,21	56	74
0908	Manteigas	3430	71	645	28	63,91	135	90
0506	Oleiros	5721	139	1624	12	63,52	266	91
0612	Pampilhosa da Serra	4481	111	1417	11	65,3	213	92

REGIÃO SUL								
ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1215	Sousel	5074	191	1325	18	69,45	192	0
1509	Santiago do Cacém	29749	989	6388	28	94,98	997	1
1210	Marvão	3512	143	882	23	66,29	145	2
0710	Redondo	7031	254	1677	19	71,97	243	4
0713	Viana do Alentejo	5743	157	1500	15	78,88	151	4
0213	Serpa	15623	640	4092	14	72,56	671	5
1409	Coruche	19944	804	5296	18	78,39	869	8
0803	Aljezur	5884	230	1449	18	64,1	250	9
0711	Reguengos de Monsaraz	10828	365	2681	23	89,8	327	10
0208	Ferreira do Alentejo	8255	310	2096	13	75,43	277	11
1505	Grândola	14826	454	3680	18	88,15	505	11
0209	Mértola	7274	318	1945	6	66,78	282	11
0202	Almodôvar	7449	284	2200	10	80,83	251	12
1204	Campo Maior	8456	244	2063	34	95,16	214	12
0809	Monchique	6045	238	1490	15	61,26	266	12
1212	Nisa	7450	213	1871	13	74,7	241	13
0714	Vila Viçosa	8319	300	1775	43	84,86	256	15
0212	Ourique	5389	180	1457	8	74,14	152	16
1209	Gavião	4132	138	1048	14	71,16	115	17
1501	Alcácer do Sal	13046	550	3340	9	81,75	453	18
1213	Ponte de Sor	16722	509	4070	20	85,92	599	18
0804	Castro Marim	6747	312	1523	22	69,69	241	23
0706	Montemor-o-Novo	17437	503	4355	14	87,26	623	24

ID_MUNICIPIO	Nome_Municipio	População residente	VGA - real	Sem escolaridade	Densidade populacional	Poder de compra	VGA - estimada	Porcentagem de erro [%e]
1407	Chamusca	10120	474	2384	14	73	350	26
0704	Estremoz	14318	343	3355	28	95,22	432	26
0815	Vila do Bispo	5258	347	1054	29	60,94	235	32
0701	Alandroal	5843	345	1499	11	64,61	231	33
0207	Cuba	4878	152	1101	28	66,36	204	35
0709	Portel	6428	387	1616	11	64,96	250	35
0702	Arraiolos	7363	341	1723	11	72,67	220	36
0214	Vidigueira	5932	308	1459	19	72,84	196	36
1208	Fronteira	3410	96	894	14	75,32	58	39
0210	Moura	15167	466	4154	16	77,8	654	40
1203	Avis	4571	167	1208	8	75,36	96	43
1201	Alter do Chão	3562	94	898	10	74,39	52	44
1206	Crato	3708	135	931	9	71,85	73	46
1202	Arronches	3165	106	853	10	72,8	55	48
0707	Mora	4978	150	1286	11	82,42	76	49
0201	Aljustrel	9257	422	2024	20	88,92	201	52
1211	Monforte	3329	120	1035	8	76,51	57	52
0703	Borba	7333	209	1811	51	74,96	354	69

0712	Vendas Novas	11846	192	2646	53	95,96	380	98
0206	Castro Verde	7276	391	1578	13	102,41	5	99
1205	Castelo de Vide	3407	107	793	13	82,58	-10	110