

## **MESTRADO**

# **MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL**

## **TRABALHO FINAL DE MESTRADO**

### **TRABALHO DE PROJETO**

#### **ANÁLISE DOS VALORES TRANSACIONADOS DOS ALOJAMENTOS FAMILIARES NA ÁREA METROPOLITANA DE LISBOA**

**CÍNTIA CARLICE LOPES PEREIRA DOS REIS**

**OUTUBRO-2018**

**MESTRADO EM**  
**MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA**  
**E EMPRESARIAL**

**TRABALHO FINAL DE MESTRADO**

TRABALHO DE PROJETO

ANÁLISE DOS VALORES TRANSACIONADOS DOS ALOJAMENTOS  
FAMILIARES NA ÁREA METROPOLITANA DE LISBOA

CÍNTIA CARLICE LOPES PEREIRA DOS REIS

**ORIENTAÇÃO:**

PROFESSORA DOUTORA AMÉLIA BASTOS  
DR. FRANCISCO VALA

OUTUBRO-2018

## **Agradecimentos**

A elaboração deste trabalho não seria possível sem a fé em Deus e o apoio e compreensão de algumas pessoas que sempre estiveram ao meu lado.

Agradeço de um modo especial aos meus orientadores Professora Amélia Bastos e Dr. Francisco Vala, pela enorme disponibilidade que demonstraram para que este trabalho fosse o melhor possível e para que eu conseguisse chegar ao fim com muito agrado e satisfação.

Um muito obrigado a todos os familiares e amigos que de uma forma ou outra me apoiaram, em especial à minha prima Neusa Vieira e à minha irmã Tânia Carlice dos Reis que foram um grande suporte desde o início até o fim desta caminhada.

Por fim, agradeço aos meus Pais que sempre me serviram de modelo e inspiração durante toda a minha vida e o meu percurso académico e que apesar da distância nunca me deixaram sentir só, dando sempre o seu máximo para que eu chegasse onde cheguei hoje.

## Resumo

O presente trabalho final de mestrado tem como principal objetivo analisar e compreender a partir de um modelo econométrico, quais as variáveis que determinam os preços dos alojamentos familiares clássicos de residência habitual na AML (Área metropolitana de Lisboa).

Devido à grande heterogeneidade desta área e para uma melhor compreensão sobre quais as variáveis que influenciam o preço, procedeu-se ao agrupamento da AML obtendo-se três *clusters*.

Foram estimados e analisados quatro modelos, um modelo que engloba todas as freguesias da AML e outros três modelos respeitantes aos grupos obtidos, de modo a perceber se as variáveis explicativas variam de grupo para grupo e relativamente ao modelo global.

Concluiu-se que os melhores modelos incluem variáveis diferenciadas, observando-se que a única variável comum a todos os modelos obtidos é a duração média dos movimentos pendulares da população empregada ou estudante (DURACAOMEDIAMOVPEND), cuja relação com a variável dependente é negativa.

Concluiu-se que o recente fenómeno de alojamento local apresenta uma elevada significância estatística como variável explicativa da variável dependente no modelo global, constatando a sua importância enquanto determinante do preço dos alojamentos.

Palavras chaves: preços das casas; análise de *cluster*; regressão linear múltipla

## **Abstract**

The main objective of this master's thesis is to analyze and understand, from an econometric model, the variables that determine the prices of classic family dwellings in AML.

Due to great heterogeneity of this area and to better understand which variables influence the price, a cluster analysis was conducted, and it showed three clusters.

In order to understand whether the explanatory variables differ from group to group and whether this difference occurs when specific cluster models are compared to a global one, four models were estimated and analyzed. One that encompasses all AML parishes and three other models related to the three acquired clusters.

It was concluded that the best models are all different, noticing that the average commute time of students and employees is the only variable included in all obtained models, and its result is negative compared to dependent variable.

It was also concluded that the recent cases of local housing presents a high statistical significance as an explanatory variable of the dependent variable in the global model.

Keywords: House price; cluster analysis; multiple linear regression

## Índice

Agradecimentos.....	I
Resumo .....	II
Abstract.....	III
Índice .....	IV
Índice de Tabelas .....	V
Índice de Figuras.....	VI
Lista de Siglas e Acrónimos.....	VII
1. Introdução.....	1
2. Revisão da literatura.....	3
2.1. <i>Preços das casas em Portugal</i> .....	3
2.2. <i>Fatores determinantes do valor da transação de um imóvel</i> .....	4
2.2.1. <i>Fatores de acessibilidade, amenidade e de localização</i> .....	4
2.2.2. <i>Fatores socioeconómicos</i> .....	6
2.3. <i>Variável dependente</i> .....	7
3. Descrição dos dados e da metodologia .....	8
3.1. <i>Agrupamento de dados- Análise de Cluster</i> .....	9
3.2. <i>Modelo de Regressão linear</i> .....	13
4. Análise empírica.....	17
4.1. <i>Apresentação dos dados e análise de cluster</i> .....	17
4.2. <i>Modelo global da AML</i> .....	21
4.3. <i>Modelo cluster 1</i> .....	23
4.4. <i>Modelo cluster 2</i> .....	24
4.5. <i>Modelo cluster 3</i> .....	25
5. Conclusão .....	28
6. Bibliografia .....	31
7. ANEXO.....	34
Anexo A - Apresentação dos Dados .....	34
Anexo B - Modelo AML .....	36
Anexo C - Modelo <i>cluster 1</i> .....	39
Anexo D - Modelo <i>cluster 2</i> .....	42
Anexo E - Modelo <i>cluster 3</i> .....	44

## Índice de Tabelas

Tabela I-Indicadores em estudo.....	17
Tabela A- I-Indicadores Socioeconómicos.....	34
Tabela A- II-Estatísticas Descritivas .....	35
Tabela A- III-Constituição dos <i>clusters</i> .....	36
Tabela B- I- Output (modelo AML) .....	36
Tabela B- II- Teste Breusch-Pagan-Godfrey(modelo AML).....	37
Tabela B- III- Teste Breusch-Godfrey(modelo AML) .....	37
Tabela B- IV- Teste RESET(modelo AML) .....	38
Tabela B- V- Teste de multicolinearidade (modelo AML) .....	39
Tabela C- I- Output (modelo <i>cluster 1</i> ) .....	39
Tabela C- II- Teste de Breusch-Pagan-Godfrey(modelo <i>cluster 1</i> ).....	40
Tabela C- III- Teste de Breusch-Godfrey(modelo <i>cluster 1</i> ).....	40
Tabela C- IV- Teste RESET(modelo <i>cluster 1</i> ) .....	41
Tabela C- V- Teste de multicolinearidade (modelo <i>cluster 1</i> ) .....	42
Tabela D- I- Output (modelo <i>cluster 2</i> ).....	42
Tabela D- II- Teste de Breusch-Pagan-Godfrey (modelo <i>cluster 2</i> ).....	42
Tabela D- III- Teste de Breusch-Godfrey(modelo <i>cluster 2</i> ) .....	43
Tabela D- IV- Teste RESET(modelo <i>cluster 2</i> ) .....	43
Tabela D- V- Teste de multicolinearidade(modelo <i>cluster 2</i> ) .....	44
Tabela E- I-Output(modelo <i>cluster 3</i> ) .....	44
Tabela E- II- Teste de Breusch-Pagan-Godfrey (modelo <i>cluster 3</i> ) .....	45
Tabela E- III- Teste de Breusch-Godfrey (modelo <i>cluster 3</i> ).....	45
Tabela E- IV- Teste RESET (modelo <i>cluster 3</i> ).....	46
Tabela E- V- Teste de multicolinearidade (modelo <i>cluster 3</i> ).....	46

## Índice de Figuras

Figura 1 – Gráfico dos Coeficientes de aglomeração .....	19
Figura 2 - Mapa de agrupamento.....	19
Figura 3 - Mapa dos Municípios da AML .....	20
Figura B- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo AML).....	39
Figura C- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo <i>cluster</i> 1).....	41
Figura D- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo <i>cluster</i> 2).....	44
Figura E- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo <i>cluster</i> 3) .....	46



## **Lista de Siglas e Acrónimos**

**AML-** Área Metropolitana de Lisboa

**AMP -** Área Metropolitana do Porto

**PHMS-** *Portuguese Housing Market Survey*

**SIR-** Sistema de Informação Residencial

**INE-** Instituto Nacional de Estatística

**GET-** Gabinete para a Coordenação das Estatísticas Territoriais

**SIOU-** Sistema de Informação das Operações Urbanísticas

**PIB-** Produto Interno Bruto

**EU-** União Europeia

**VIF-** *Variance Inflation factor*

**SSR-** Soma dos Quadrados da Regressão

**SSE-** Soma dos Quadrados dos Resíduos

**RICS-** *Royal Institute of chartered Surveyors*

## 1. Introdução

A habitação é um bem prioritário na vida de qualquer indivíduo. A sua aquisição tem em conta as necessidades e preferências inerentes a cada indivíduo no que diz respeito às características do imóvel, do contexto onde está inserido e a sua disponibilidade de recursos.

O mercado imobiliário é bastante amplo e diversificado, uma vez que o imóvel é um bem heterogéneo. A heterogeneidade é definida pelas características únicas do bem em causa, mas este é avaliado considerando também o contexto onde está inserido, a sua localização e as condições de acessibilidade no que diz respeito às infraestruturas e serviços. Atualmente com o aumento contínuo e exponencial dos preços dos imóveis nos centros urbanos em particular na Área Metropolitana de Lisboa (AML), torna-se cada vez mais importante a disponibilização de informação clara, concisa e credível aos agentes económicos, sobre os determinantes do valor de transação de um imóvel.

Na AML a procura por habitações tem seguido uma trajetória ascendente desde o último trimestre de 2013, segundo os dados do PHMS <sup>1</sup>(*Portuguese Housing Market Survey*). Os dados reportados ao SIR (sistema de informação residencial), apontam também para um acréscimo da procura no mercado de compra e venda da AML, tendo o volume de fogos transacionados nesta região pelo grupo de empresas que participam nesses sistemas mais do que duplicado entre 2013 e o 2º trimestre de 2017, passando das 4,4 mil unidades para as 10,3 mil unidades (Confidencial Imobiliário, 2018).

O presente estudo tem por objetivo analisar e compreender a partir de um modelo econométrico, quais as variáveis que determinam os preços dos alojamentos na AML. Devido à grande heterogeneidade desta zona realizou-se uma análise de *clusters*, com vista a construir grupos homogéneos de freguesias nesta área. Estimam-se modelos para cada grupo identificado, com o objetivo de perceber quais as variáveis explicativas que são estatisticamente significativas em cada um dos *clusters* obtidos. Estima-se também o

---

<sup>1</sup> Inquérito de confiança e sentimento de mercado produzido pela Confidencial Imobiliário em parceria com o RICS- *Royal Institute of chartered Surveyors*. Este inquérito permite um acompanhamento mensal de um painel de aproximadamente 150 agentes de mercado, produzindo indicadores assentes em metodologias de saldos de respostas extremas (% de inquiridos com expectativas positivas- % inquiridos com expectativas negativas)

modelo global para a AML que visa explicar o preço dos imóveis nesta zona em termos globais.

Em parceria com o INE (Instituto Nacional de Estatística) procedeu-se à construção de indicadores socioeconómicos (Carvalho (2012), Cruz (2016), Censos-2011, Preparação, Metodologia e Conceitos (2013), Estatísticas de Preços da Habitação ao nível local-2º Trimestre de 2017, Retrato Territorial de Portugal. Estatísticas oficiais (2017), tipologia Socioeconómica das Áreas Metropolitanas de Lisboa e Porto 2011(2014), O Corvo (2018), Pereira(2018), Resolução do conselho de Ministros nº4/2015, Soldado(2018) ), ao nível da freguesia, associados aos preços da habitação mais concretamente aos preços dos alojamentos familiares clássicos de residência habitual - variável e unidade estatística de análise no presente trabalho. O processo da construção de indicadores ocorreu no Gabinete para a Coordenação das Estatísticas Territoriais (GET) e dividiu-se em três fases: discussão e levantamento dos indicadores, cálculo dos indicadores e construção da base de dados de trabalho.

Foram consideradas algumas fontes para a obtenção dos dados que permitiram calcular os indicadores à escala pretendida tais como: o Censos de 2011, o SIOU (Sistema de Informação das Operações Urbanísticas), o Turismo de Portugal e a análise espacial recorrendo ao *OpenStreetMap* e *softwares* geográficos (Ex: Arcgis)

Após o presente capítulo, será realizada uma breve revisão da literatura sobre os estudos existentes acerca do tema em análise, de seguida proceder-se-á à descrição dos dados e da metodologia aplicada, prosseguindo-se com a análise empírica dos modelos obtidos. Por fim serão apresentadas as conclusões do estudo, referindo-se alguns contributos e limitações do mesmo e sugestões para futuras investigações.

## 2. Revisão da literatura

### 2.1. Preços das casas em Portugal

A habitação é um bem complexo, quando se considera o cruzamento das dimensões económica, social, urbana e política. Para além de ser reconhecida como um direito fundamental, a habitação é destacada como um motor económico muito importante no crescimento da União Europeia (Dijol et al, (2015)).

Os preços das casas na União Europeia e em particular em Portugal têm vindo a apresentar aumentos exponenciais e contínuos, conduzindo a uma situação de *bolha imobiliária* - indiciando instabilidade financeira (Claeys et al, (2017)). Após um período caracterizado por uma fraca evolução (durante os anos 90 e início dos anos 2000), o preço da habitação em Portugal tem apresentado um aumento significativo no passado recente (Lourenço & Rodrigues, 2017).

Em 2015, Portugal foi juntamente com a Alemanha considerado um dos únicos países da Zona Euro cujos preços dos imóveis nunca experimentaram um *boom* desde os anos 80. Dijol et al, (2015) mostraram que em 2015 Portugal inverteu a queda significativa dos preços dos imóveis após à crise financeira global em 2007. Desde o último trimestre de 2015 e ao longo de 2016 registou-se um aumento do índice de preços residenciais em Portugal.

De acordo com os dados apurados pela Confidencial Imobiliária até setembro de 2017, o sector imobiliário atravessava um bom momento pois o índice de preços residenciais registou uma variação positiva (subida acumulada de 9% até o último mês apurado) durante este período. Segundo o que foi captado pelo PHMS (*Portuguese Housing Market Survey*), as perspetivas dos agentes relativamente ao sector imobiliário são positivas no sentido em que existe uma tendência de aceleração do índice de preços residenciais.

A partir do ano 2007 até aos dias de hoje tem-se verificado uma fase de recuperação do mercado imobiliário em Portugal.

## **2.2. Fatores determinantes do valor da transação de um imóvel**

*A habitação não é um elemento que possa ser pensado isolado de uma política de cidade, de um ambiente externo de equipamentos e de espaços coletivos, que faz com que seja simultaneamente um produto para as famílias resolverem as suas necessidades, mas também um património económico de uma sociedade e um meio de ligação entre espaços, pessoas e organizações com vidas e objetivos, diferentes e complexos. É preciso não esquecer que a habitação é também um bem económico essencial, quer para a iniciativa privada, quer para as exigências do erário público, sendo um sector que reflete o dinamismo económico e financeiro de uma sociedade.*

In Guerra (2008), p. 47-61.

Foram realizados vários estudos com o intuito de avaliar quais são os fatores que influenciam a decisão dos agentes na determinação do preço de imóveis, fatores estes que podem ser variados dependendo das necessidades e preferências de cada indivíduo.

A tendência é de pensarmos que na escolha de uma casa consideram-se essencialmente as suas características estruturais, variáveis que caracterizam estruturalmente os imóveis como por exemplo se é novo ou usado, a sua tipologia, se tem estacionamento, garagem, varandas, a área útil, a idade, o número de casas de banho, etc. (Tarré, 2009), mas essa não é a realidade. É claro que estas características têm uma grande importância, mas existem outros fatores de grande impacto nessa escolha. Podemos dividir estes fatores em dois grandes grupos: os fatores de acessibilidade, amenidade e de localização e os fatores socioeconómicos.

### **2.2.1. Fatores de acessibilidade, amenidade e de localização**

A escolha do local de residência é uma decisão muito importante para um indivíduo pois tende a influenciar o seu grau de satisfação no que diz respeito ao acesso a oportunidades de trabalho, comércio, escolas, hospitais, locais de culto, entretenimento e proximidade de parentes e amigos (Goodall, 1972). Assim a acessibilidade é considerada como um fator crucial na decisão de compra de uma casa.

Os fatores de amenidade e de localização exercem uma grande influência na decisão de compra das casas. Uma entidade (famílias ou empresas) valoriza muito a proximidade das estações de metro, dos espaços verdes, de zonas industriais, de praias, a boa vista, o ar puro, etc.

De acordo com Bazyl (2009) a proximidade dos espaços verdes apresenta um impacto positivo de 4% no valor do imóvel. Gaspar (2011), no seu estudo revelou que o valor dos imóveis se reduz em cerca de 135 euros por cada metro de distância dos espaços verdes mais próximos.

No que diz respeito às estações do metro, um aumento da distância relativamente a este equipamento tem um impacto negativo no valor do imóvel (Estrella, 2008). De acordo com Bazyl (2009), a existência de uma estação de metro a 1 km de um apartamento implica que o preço deste aumente em 15% relativamente aos outros apartamentos situados a uma maior distância.

No seu estudo Bazyl (2009) mostra também que a proximidade das zonas industriais afeta negativamente (7%) os preços dos alojamentos.

A localização apresenta-se como fator de elevada importância na determinação do preço de um imóvel. Vieira (2011) analisou o impacto da localização no valor das transações a partir do coeficiente de localização<sup>2</sup> que se apresentou como um bom estimador confirmando que a localização exerce um impacto positivo no valor de transação de um imóvel.

Claeys et al, (2017) sugerem a divisão entre as capitais e as outras regiões quando se pretende analisar o preço das casas de um determinado país, revelando que o crescimento dos preços nas capitais geralmente tende a ser mais volátil e mais pronunciado em relação ao resto do país. De acordo com a análise realizada por estes autores, a diferença

---

<sup>2</sup> Coeficiente aplicado em cada zona homogénea de um determinado município que varia conforme se trate de edifícios destinados à habitação, comércio, indústrias ou serviço e o seu valor está compreendido entre 0,4 e 3,5. Na fixação deste coeficiente tem-se em consideração algumas características, tais como: as acessibilidades (nível de qualidade e variedade das vias rodoviárias, ferroviárias, fluviais e marítimas); a proximidade de equipamentos sociais (escolas, serviços públicos e comércio); serviços de transportes públicos; localização em zonas de elevado valor de mercado imobiliário.

no crescimento de preços entre as capitais e as demais regiões de cada país consideradas no estudo varia entre 0,6 e 3,5 pontos percentuais.

Teixeira (2011), considerou como variável explicativa do preço de um imóvel, a sua localização relativa dentro de uma cidade.

### **2.2.2. Fatores socioeconómicos**

Alguns fatores de cariz socioeconómico têm sido considerados como determinantes dos preços, como por exemplo a acessibilidade no que diz respeito ao poder de compra, a situação económica do país (período de crise ou pós crise), etc.

A acessibilidade dos preços das casas em termos do poder de compra dos compradores. Claeys et al, (2017) analisaram o índice de preços para renda<sup>3</sup> para avaliar a acessibilidade da habitação em seis países da União Europeia. Segundo estes autores, um comportamento crescente deste indicador significa uma habitação menos acessível, com os preços de imóveis residenciais crescendo mais rapidamente do que o rendimento disponível das famílias. De acordo com os resultados obtidos no estudo, a diminuição persistente da acessibilidade nas capitais é de extrema importância para a estabilidade financeira destas, uma vez que põe em causa a capacidade das famílias de liquidar as dívidas em caso de acontecimentos inesperados (como por exemplo no caso de mudanças nos níveis das taxas de juros ou renda, quebra de rendimento disponível por motivos de doença ou desemprego).

Atualmente há que considerar a separação dos períodos de crise e pós crise na análise dos fatores determinantes dos preços das casas, pois com a crise foram várias as alterações verificadas em todos os setores, inclusive no sector imobiliário em que se verificou um forte impacto no curto prazo (Dijol et al, 2015).

O impacto dos determinantes do preço de um imóvel difere consoante se consideram os períodos de crise e pós crise (Lourenço & Rodrigues, 2017). Na sua análise, Lourenço & Rodrigues (2017) revelam que durante a crise, de entre as variáveis socioeconómicas introduzidas no modelo, o logaritmo natural do PIB real *per capita*, a taxa de desemprego, a taxa de hipoteca real, e o investimento direto estrangeiro em imóveis num

---

<sup>3</sup> preços das casas em relação ao rendimento disponível no domicílio

determinado período, todas se apresentaram como significativas no modelo exceto a taxa de hipoteca real. No entanto no período pós crise apenas o PIB real *per capita* e a taxa de hipoteca real se apresentaram como estatisticamente significativas.

Outros fatores socioeconómicos foram considerados em estudos sobre os preços das casas como o índice de preço da habitação e o índice de preço para renda. Fatores estes que foram classificados como alguns dos indicadores de alerta precoce mais confiáveis de bolhas imobiliárias insustentáveis num cenário a nível da União Europeia (Claeys et al, 2017).

### ***2.3. Variável dependente***

Em vários estudos realizados sobre os preços das casas, considerou-se o logaritmo do preço como variável dependente, ou seja, o modelo apresenta-se na forma semi-logarítmica isto porque nenhuma das variáveis explicativas inseridas no modelo está na sua forma logarítmica (Rodrigues, 2008).

Alguns autores apresentam a forma semi-logarítmica como uma solução para o problema de heterocedasticidade (*Follain & Malpezzi (1980) e Fletcher et al 2000*). Segundo Vieira (2001), esta forma pode conduzir a modelos homocedásticos, ou pelo menos com uma menor variância do erro. Estas questões metodológicas serão desenvolvidas nos capítulos seguintes.



### 3. Descrição dos dados e da metodologia

O presente estudo tem por objetivo analisar e compreender a partir do modelo de regressão linear, quais as variáveis que determinam os preços medianos dos alojamentos familiares clássicos de residência habitual, na AML.

Para a realização da análise, foi construída uma base de dados constituída por indicadores socioeconómicos agregados ao nível da freguesia (*Tabela A-I*), recorrendo a algumas fontes como o Censos, o Sistema de Informação das operações urbanísticas (SIOU), o Turismo de Portugal e a análise espacial através do *OpenStreetMap* e de *softwares* estatísticos (Ex: Arcgis). Durante o processo de construção da base de dados, procedeu-se à discussão e levantamento dos indicadores, ao cálculo dos indicadores e por fim à construção da base de dados de trabalho.

Foi considerada a mediana dos preços uma vez que esta vem expurgada do efeito da presença dos *outliers*<sup>4</sup> observados no mercado de transações de habitação e dos possíveis efeitos sazonais no comportamento dos preços, evitando assim o risco de uma estimação enviesada. Estes preços dizem respeito ao 4º trimestre de 2016.

É de salientar que na base de dados de preços disponibilizada, das 118 freguesias da AML, 5 (Fanhões, União das freguesias de Gaio-Rosário e Sarilhos Pequenos, Canha, Sarilhos Grandes, União das freguesias de Pegões) não apresentam informação sobre o preço mediano dos alojamentos, por este motivo a amostra reduziu-se para 113 freguesias.

Dada a heterogeneidade da área de estudo procedeu-se à análise de *clusters*, recorrendo ao *software* estatístico *SPSS*, obtendo-se 3 *clusters*. Após a formação dos *clusters* aplicou-se o modelo de regressão linear múltipla a cada grupo e à área metropolitana no seu todo com recurso ao *software* estatístico *Eviews*.

---

<sup>4</sup> Valores extremos

### 3.1. Agrupamento de dados- Análise de Cluster

A análise de *clusters* é uma forma de produzir grupos, atuando sobre um conjunto de dados e utilizando métodos matemáticos normalmente de forma automática (Branco, 2004). O primeiro passo para a formação de grupos de elementos pertencentes a um universo é conhecer os seus atributos, as suas propriedades, isto é, as variáveis que caracterizam estes elementos.

Esta análise é realizada com alguns objetivos específicos tais como: a exploração de dados (permite perceber se os dados estão bem estruturados ou não); a redução dos dados a partir da obtenção de grupos homogêneos e a geração de hipóteses (podem resultar classificações não esperadas o que leva a criação de conjeturas e formação de hipóteses para explicar a estrutura obtida).

Na análise de *cluster* utilizam-se medidas de semelhança ou de dissemelhança aplicadas inicialmente entre dois sujeitos e posteriormente entre dois *Clusters* de objetos utilizando técnicas hierárquicas ou não-hierárquicas de agrupamento. Existem várias medidas de semelhança (proximidade) e de dissemelhança (distância) mas normalmente na análise de *clusters* as mais utilizadas são: a distância euclidiana, a distância de *Minkowski*, a distância de *Mahalanobis*, a medida de semelhança do coseno, os coeficientes de *Jaccard*, de *Russel & Rão* e medidas de associação binárias e por último, medidas de semelhanças para as variáveis- Marôco (2014)

A distância Euclidiana é uma medida que consiste em avaliar o comprimento do segmento de reta que une duas observações num espaço  $p$ -dimensional.

Segundo *Johnson & Wichern* (2002), se considerarmos  $p$  variáveis a distância euclidiana entre os sujeitos  $i$  e  $j$  é dada por:

$$(4) \quad D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

em que  $x_{ik}$  é o valor da variável  $k$  no sujeito  $i$  e  $x_{jk}$  é o valor da variável  $k$  para o sujeito  $j$ . Do mesmo modo tem-se a representação do método na forma matricial

$$(5) D_{ij} = \sqrt{(x_i - x_j)'(x_i - x_j)}$$

onde  $x_i$  e  $x_j$  representam os vetores  $p$ -dimensionais das observações que correspondem aos indivíduos  $i$  e  $j$  respetivamente.

No cálculo da distância euclidiana podem ser consideradas as variáveis originais, em que variáveis com maior amplitude apresentam maiores distâncias, ou as standardizadas em que a distância euclidiana não é influenciada pela amplitude das variáveis. Independentemente do tipo de variáveis o *SPSS Statistics- software* utilizado na construção dos *clusters*- utiliza, por defeito, o quadrado da distância euclidiana como medida de dissimilaridade entre sujeitos.

A Distância de *Minkowski*, pode ser considerada uma generalização da distância euclidiana representada por:

$$(6) D_{ij} = \sqrt[m]{\sum_{k=1}^p |x_{ik} - x_{jk}|^m},$$

para  $m=1$ ,  $D_{ij}$  é o módulo da distância absoluta entre os sujeitos  $i$  e  $j$  relativamente às  $p$ -variáveis medidas; para  $m=2$  obtêm-se a distância Euclidiana.

A Distância de Mahalanobis mede a distância estatística entre dois sujeitos  $i$  e  $j$  (Johnson, 1998):

$$(7) D_{ij} = \sqrt{(x_i - x_j)' S^{-1} (x_i - x_j)}$$

Onde  $S$  é uma estimativa amostral da matriz de variâncias-covariâncias  $\Sigma$  dentro dos grupos ou *Clusters*. Esta medida tem em conta a correlação entre as variáveis e é particularmente útil quando a matriz  $\Sigma$  é diferente da matriz identidade (se  $\Sigma=I$  a distância de *Mahalanobis* reduz-se à distância euclidiana). O uso desta distância na análise de clusters tem uma limitação que resulta da necessidade de conhecer à partida os grupos sobre os quais se vai calcular  $S$ . É claro que se os grupos são definidos à partida, então não será necessário utilizar a análise de clusters. Este método não se encontra disponível no **SPSS**.

A Medida de Semelhança do Coseno, mede a proximidade entre dois sujeitos para  $p$ -variáveis pelo menos intervalares, dada por:

$$(8) \text{ CoSIN}(i, j) = \frac{\sum_{k=1}^p x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^p x_{ik}^2 \sum_{k=1}^p x_{jk}^2}}$$

Coefficiente de *Jaccard*, de *Russel & Rão* e Medidas de Associação Binárias definidas como medidas de associação apropriadas para tabelas de contingência, são utilizadas quando os sujeitos de uma amostra multivariada são caracterizados por variáveis nominais.

As Medidas de semelhança para variáveis, como o próprio nome diz são utilizadas quando a análise de Clusters tem como objetivo o agrupamento de variáveis (e não de sujeitos). Neste caso as medidas de semelhança apropriadas são os coeficientes de correlação amostrais. No caso de variáveis contínuas normalmente o mais utilizado é o coeficiente de correlação de *Pearson* (ou, de um modo geral, o seu valor absoluto). Para variáveis ordinais deve utilizar-se o coeficiente de *Spearman*, e finalmente no caso de variáveis nominais deve utilizar-se o coeficiente *Phi*. Estes coeficientes podem também ser utilizados como medidas de semelhança entre sujeitos, mas nestes casos, o seu uso é por vezes desprovido do significado claro das medidas de correlação.

Neste estudo utilizou-se o quadrado da distância Euclidiana que é a utilizada por defeito no SPSS.

Uma vez que no presente estudo recorreu-se à técnica hierárquica considerando o método de Ward para a determinação das distâncias entre os sujeitos do primeiro *cluster* (método aglomerante), vai-se realizar um estudo mais pormenorizado desta técnica e do método referido.

### **Técnicas hierárquicas- Método de Ward**

Nas técnicas hierárquicas existem dois tipos de métodos de agrupamentos, os aglomerantes e os divisivos. Nos métodos aglomerantes recorre-se a passos sucessivos de agregação dos sujeitos considerados individualmente, ou seja, considera-se que cada um dos sujeitos é um *cluster*, posteriormente estes serão agrupados considerando o seu grau de proximidade. Quando os sujeitos são inicialmente agrupados num único *cluster* e posteriormente são divididos em subgrupos de acordo com as suas medidas de distância, está-se perante o método divisivo. Normalmente os métodos aglomerantes são os mais utilizados.

No presente estudo utilizou-se o método de *Ward*, uma vez que do ponto de vista estatístico de entre todos os métodos existentes no *software* utilizado (SPSS), este é o que melhor se adequa ao objetivo da análise de *clusters* – a produção de grupos mais homogêneos e melhor separados entre si (Centro de Estudos de Gestão(CEGE), Análise de dados com IBM SPSS).

Em 1963, *Ward* definiu o seu método em que o critério de fusão de dois grupos A e B baseia-se no incremento da soma dos quadrados resultantes da aglutinação entre os dois *clusters*. Esse incremento é definido por:

$$(9) \quad SSW_C - (SSW_A + SSW_B),$$

Onde (10)  $SSW_A = \sum_{i \in A} \sum_{j=1}^p (x_{ijA} - \bar{x}_{jA})^2$  é a soma dos quadrados dentro do grupo A e (11)  $SSW_B = \sum_{i \in B} \sum_{j=1}^p (x_{ijB} - \bar{x}_{jB})^2$  é a soma dos quadrados dentro do grupo B e (12)  $SSW_C = \sum_{i \in C} \sum_{j=1}^p (x_{ijC} - \bar{x}_{jC})^2$  é a soma dos quadrados dentro do grupo C=A U B, resultante da aglutinação do grupo A com o grupo B;  $x_{ijA}$  ( $x_{ijB}$ ) é a observação do objeto i do grupo A (B) na variável j,  $\bar{x}_{jA}$  e  $\bar{x}_{jB}$  são as médias da variável j nos grupos A e B, respetivamente.

Na implementação do método de *Ward* as distâncias não são calculadas, mas sim os *clusters* são construídos de forma a minimizar a soma dos quadrados dos erros, retendo em cada passo do algoritmo os *clusters* com a menor soma de quadrados dos erros possíveis, ou seja, aqueles a que corresponde a menor perda de informação derivada da aglutinação. Este método é positivamente caracterizado pela boa performance com os dados de grande dispersão e pela tendência em produzir *clusters* com dimensões semelhantes.

Para além do método de *Ward* existem outros métodos hierárquicos tais como: o método de menor distância, o método de maior distância, a distância média entre *clusters*, a distância média dentro dos *clusters*, a distância mediana e o método do centróide.

Um dos grandes problemas da análise de *clusters* é a determinação do número de *clusters* a reter quando este não está definido. Neste estudo foi utilizado o método da distância entre *clusters*, mas existe um outro método também muito utilizado que é o critério do R quadrado. O método da distância entre *clusters* permite sugerir o número de *clusters* “naturais” a partir da análise do gráfico construído em função dos coeficientes de

aglomeração dos *clusters*, quando se verifica um ponto de inflexão na curva (Marôco, 2014).

### 3.2. Modelo de Regressão linear

Para analisar quais são os fatores que afetam os preços dos imóveis, recorreu-se ao modelo de regressão linear múltipla.

O modelo de regressão linear múltipla geral pode ser apresentado da seguinte forma:

$$(13) \quad Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u,$$

Onde  $\beta_0$  é o intercepto,  $\beta_1$  é o parâmetro associado a  $x_1$ ,  $\beta_2$  é o parâmetro associado a  $x_2$ , assim sucessivamente e  $u$  é o termo de erro que contém outros fatores, para além do  $x_1, x_2, \dots, x_k$ , que afetam o  $y$ . Independentemente do número de variáveis explicativas incluídas no modelo, existirão sempre variáveis que não foram incluídas e que estão todas contidas na variável aleatória residual  $u$ . Desde que existam  $k$  variáveis independentes e o intercepto, a equação acima definida contém  $k+1$  parâmetros (desconhecidos) da população em estudo.

De modo a avaliar a qualidade dos modelos obtidos, será realizado um conjunto de testes de modo a perceber se os modelos obedecem as hipóteses do modelo de regressão linear múltipla, cujo detalhe poderá ser visto em Wooldridge (2000). Neste contexto foram realizados os seguintes testes de hipótese:

- Teste de homocedasticidade:

O pressuposto da homocedasticidade das variáveis aleatórias residuais é testado recorrendo ao teste *Breusch-Pagan-Godfrey* cuja hipótese nula é a homocedasticidade (variância constante das variáveis aleatórias residuais) contra a hipótese alternativa de heterocedasticidade

$$(14) \quad \begin{cases} H_0: \text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2 \\ H_1: \text{Var}(u|x_1, x_2, \dots, x_k) = \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \dots + \delta_k x_k \end{cases}$$

e a estatística de teste é dada por:

$$(15) \quad LM = n \cdot R_{\hat{\epsilon}^2}^2 \sim \chi_{k-1}^2 \quad ^5$$

- Teste de autocorrelação:

O pressuposto da ausência da autocorrelação exige que as variáveis aleatórias residuais sejam independentes, ou seja não estejam correlacionadas. A hipótese nula da ausência de autocorrelação é testada a partir do teste de *Breusch-Godfrey*, cujas hipóteses são:

$$(16) \quad \begin{cases} H_0: Cov(u_t, u_s | \mathbf{X}) = Cov(u_t, u_s) = 0 \\ H_1: Cov(u_t, u_s | \mathbf{X}) \neq 0 \end{cases},$$

e a estatística de teste é:

$$(17) \quad LM = n \cdot R_{\hat{u}}^2 \sim \chi_q^2.$$

- Teste de especificação ou ajustamento do modelo:

Permite analisar se o modelo está bem especificado ou não, declarando problemas de omissão de variáveis importantes e ou de inclusão de variáveis não relevantes para o modelo. A hipótese nula de que o modelo está bem especificado será testada a partir do teste *RESET* cuja estatística de teste é:

$$(18) \quad F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} \sim F_{n-k-1}^k.$$

Em que *SSR* é a soma dos quadrados da regressão e *SSE* é a soma dos quadrados dos resíduos.

- A normalidade dos resíduos é testada pelo teste de *Jarque-Bera*, cuja hipótese nula é a existência de normalidade dos resíduos. A estatística do teste de *Jarque-Bera* é dada por:

$$(19) \quad JB = \frac{n-k+1}{6} \left( S^2 + \frac{1}{4} (C - 3)^2 \right).$$

Onde *n* é o número de observações (ou normalmente graus de liberdade), *S* é o *skewness* da amostra, *C* é a *Kurtose* da amostra e *K* é o número de regressores.

- O pressuposto da ausência de multicolinearidade será analisado considerando o valor do *VIF* (*Variance Inflation factor*), dado por:

$$(20) \quad VIF_j = \frac{1}{1-R_j^2},$$

onde  $R_j$  é o coeficiente de determinação da regressão de  $x_j$  sobre as outras variáveis explicativas e mede o quanto a variância do coeficiente  $\beta_j$  é inflacionada por sua colinearidade. Se  $VIF > 10$ , está-se na presença de multicolinearidade.

Para além destes testes ainda são efetuados os testes de significância individual e global das variáveis.

- Teste de significância individual

A significância individual consiste em analisar se o impacto de uma variável inserida no modelo é estatisticamente significativo ou não, ou seja, permite avaliar o poder explicativo de cada variável isoladamente. As hipóteses a serem testadas neste teste são:

$$(21) \begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}, j=0, 1, \dots, k \text{ e a estatística do teste em causa é:}$$

$$(22) T = \frac{\hat{\beta}_j}{S\sqrt{c_{jj}}} \sim t_{n-k-1},$$

em que (23)  $S = \sqrt{\frac{SSE}{n-k-1}}$  é o estimador de  $\sigma$ .

Se  $H_0: \beta_j = 0$ , não for rejeitada pode-se retirar  $x_j$  do modelo uma vez que esta não influencia significativamente o valor da variável dependente

- Teste de significância global

Este permite-nos testar o modelo de regressão, fornecendo informação sobre a sua validade. Fornece-nos informação sobre a existência ou não de uma relação linear entre a variável explicada e as variáveis explicativas. As hipóteses do teste são:

$$(24) \begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1: \text{pelo menos um } \beta_j \neq 0 \end{cases}$$

E a estatística de teste é definida por:

$$(25) F = \frac{\frac{SSR}{k}}{S^2} \sim F_{n-k-1}^k$$



Rejeita-se  $H_0$  para valores muito grandes de F, estatisticamente tabulados. Caso não se rejeite a hipótese nula, conclui-se que o conjunto de variáveis explicativas contribui pouco para a explicação da variável dependente.

## 4. Análise empírica

### 4.1. Apresentação dos dados e análise de cluster

Na tabela que se segue estão apresentadas as variáveis incluídas nos modelos e suas correspondentes designações no decorrer do estudo:

*Tabela I-Indicadores em estudo*

Logaritmo da mediana do valor das transações	LVALORMEDIANO
Proporção do alojamento local <sup>5</sup> (nos alojamentos clássicos)	AL
Proporção da população com 23 anos ou mais e com ensino superior	PROPPOPCOM23OUMAI-COMES
Duração média dos movimentos pendulares da população empregada ou estudante	DURACAOMEDIAMOV-PEND
Proporção da população residente pertencente aos grupos socioeconómicos 11,20 e 23 <sup>6</sup>	GSE_11_20_23
Proporção de alojamentos familiares clássicos de residência habitual com aquecimento central	PROPALOJCOMAQUEC
Proporção de alojamentos familiares clássicos de residência habitual sobrelotados	PROPALOJSOBRELOT
Cobertura populacional dos estabelecimentos de ensino superior	COBPOPES
Cobertura populacional dos estabelecimentos de ensino básico	COBPOPEB
Proporção de edifícios com necessidades de grandes reparações ou muito degradadas	PROPEDIFNECESSREPARACO
Idade média dos Edifícios	IDADEMEDIAEDIF
Cobertura territorial dos transportes (autocarro, elétrico e metro)	COBTERTRANSP
Proporção de edifícios não exclusivamente residenciais	PROPEDIFNAOEXCLUSRESIDE

Fonte: INE

A cobertura populacional dos estabelecimentos de ensino básico e do ensino superior, indica a proporção da população residente numa freguesia servida pelo serviço do ensino básico e do ensino superior considerando um *buffer*<sup>7</sup> de 1500 e 2000 metros respetivamente. Na mesma ótica a cobertura territorial dos transportes indica a proporção de superfície servida pelos transportes considerando um *buffer* de 500 metros.

<sup>5</sup> Nova tipologia dos alojamentos cuja finalidade é o arrendamento de curta duração para fins turísticos

<sup>6</sup> O grupo socioeconómico 11 é designado como o grupo dos Trabalhadores industriais e artesanais independentes; o grupo socioeconómico 20 é designado como o grupo dos Trabalhadores administrativos do comércio e serviços, não qualificados e o grupo socioeconómico 23 é designado como o grupo dos Operários não qualificados- Censos 2011-Preparação, Metodologia e Conceitos, 2013.

<sup>7</sup> Área de influência medida em raios.

Para uma melhor perceção do comportamento das variáveis consideradas nos modelos, recorreu-se à análise das estatísticas descritivas apresentadas na *Tabela A-II*.

Relativamente à mediana dos preços dos alojamentos clássicos de residência habitual, o valor máximo é atingido na freguesia de Santa Maria Maior (município de Lisboa) e o valor mínimo é atingido na União das freguesias de Baixa da banheira e Vale da Amoreira, (município de Moita).

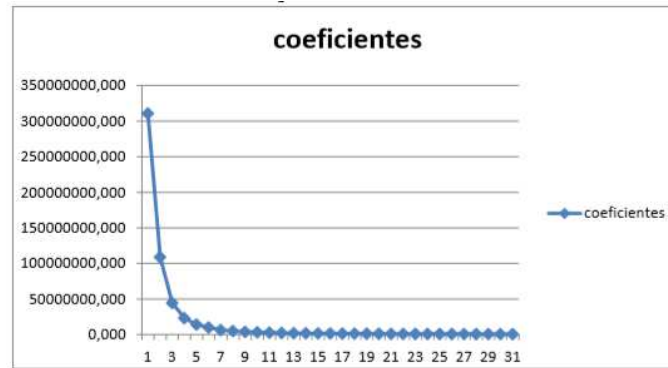
Após a análise exploratória dos dados procedeu-se ao agrupamento da AML. Como referido anteriormente, no processo de agrupamento recorreu-se ao método de *Ward* e para a determinação do número de *clusters* a reter utilizou-se o método da distância entre *clusters*.

No processo de formação de clusters foram considerados os 32 indicadores socioeconómicos construídos, apresentados na *Tabela A-I* do Anexo. Considerou-se todos os indicadores uma vez que o objetivo é perceber quais destes influenciam os preços em cada grupo.

As freguesias foram então agrupadas de acordo com a semelhança apresentada pelos indicadores. Ou seja, se duas ou mais freguesias pertencem ao mesmo *Cluster* significa que os indicadores apresentam valores próximos nelas e ou os indicadores com maior peso são os mesmos para as duas freguesias.

Após a aplicação do método de Ward, recorreu-se à análise do gráfico apresentado abaixo, onde estão representados os 30 maiores coeficientes de aglomeração<sup>8</sup> por ordem decrescente:

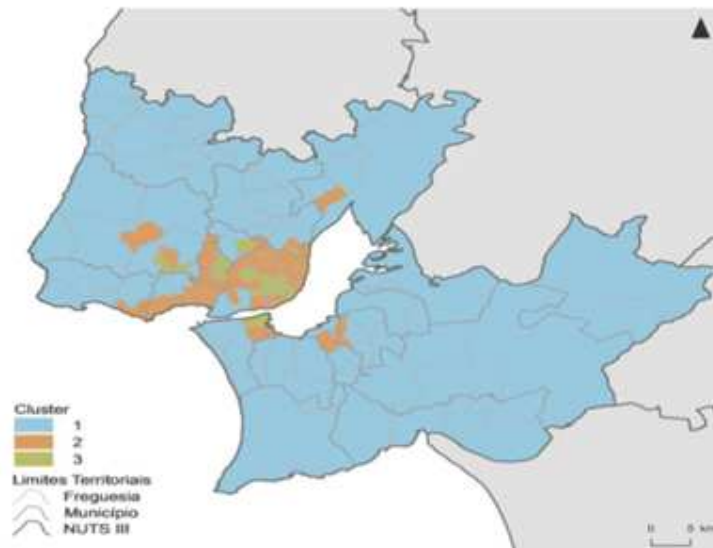
**Figura 1 – Gráfico dos Coeficientes de aglomeração**



Fonte: cálculos efetuados pela autora a partir dos dados fornecidos pelo INE relativos aos indicadores socioeconómicos que constam no anexo A, Tabela A-I.

Constatando evidências para a construção de 3 *clusters*. Procedeu-se então à construção dos *clusters* cuja constituição é apresentada na figura 2:

**Figura 2 - Mapa de agrupamento**



Fonte: Output Arcgis

<sup>8</sup>Coefficientes disponibilizados pelo SPSS no método hierárquico (*Agglomeration Schedule*) e consiste na “medida” da distância entre as unidades de observação permitindo a formação dos grupos, uma vez que valores idênticos revelam semelhanças entre as unidades observadas

**Figura 3 - Mapa dos Municípios da AML**

Fonte: Output Arcgis

Analisando o mapa temático acima e a *Tabela A-III*, podemos verificar que as freguesias dos municípios de Alcochete, Loures, Mafra, Montijo, Palmela, Seixal, Sesimbra e Setúbal pertencem, na sua totalidade, ao *cluster 1*. A maior parte das freguesias dos municípios de Vila Franca de Xira, Sintra, Odivelas, Moita, Barreiro e Cascais pertence ao *cluster 1*. Este é constituído por 67 freguesias.

O *cluster 2* é constituído por 30 freguesias, principalmente por freguesias dos municípios de Lisboa e Oeiras. Fazem parte deste grupo as freguesias pertencentes à região Sul, Sudeste, Oeste e noroeste do município de Amadora.

O *cluster 3* é constituído por 16 freguesias, na sua maioria por freguesias da região central do município de Lisboa e pelas freguesias do município de Amadora pertencentes às regiões Nordeste, Este e lés-sudeste.

Da constituição dos *Clusters* podemos observar que existem freguesias que não pertencem aos mesmos Municípios ou não pertencem a Municípios “próximos”, mas pertencem ao mesmo *Cluster*, o que significa que os valores dos preços transacionados destas freguesias são influenciadas pelos mesmos indicadores, ou os indicadores de maior peso são iguais para ambas. Ou seja, neste estudo há indicadores que revelam um peso elevado pondo em causa a ideia de que nas freguesias mais próximas os indicadores dos preços

são os mesmos. E este resultado “não muito esperado” pode ser explicado pelo desfaseamento do tempo entre os períodos a que se referem os dados utilizados no cálculo dos diferentes indicadores- informação apresentada na *Tabela A-I*.

Após a construção e definição dos *clusters*, procedeu-se à estimação dos modelos. Durante o processo de estimação, optou-se por um modelo não linear nos preços pelos motivos apresentados no capítulo 2. Assim, a variável dependente é o logaritmo da mediana do valor das transações (LVALORMEDIANO).

Primeiramente analisar-se-á o modelo global que explica o preço dos alojamentos pertencentes à AML e depois proceder-se-á à análise dos modelos respeitantes a cada um dos *clusters* obtidos, considerando um nível de significância de 5% ( $\alpha=0,05$ ) e o desfaseamento do tempo entre os preços medianos e os indicadores calculados.

#### **4.2. Modelo global da AML**

O modelo global da AML relaciona o preço mediano dos alojamentos (LVALORMEDIANO) com a percentagem do alojamento local nos alojamentos clássicos (AL), a proporção da população com 23 anos ou mais com ensino superior (PROPPOPCOM23OUMAISSCOMES), a duração média dos movimentos pendulares da população empregada ou estudante (DURACAOMEDIAMOVPEND), a proporção da população residente pertencente aos grupos socioeconómicos 11, 20 e 23 (GSE\_11\_20\_23), a proporção de alojamentos com aquecimento central (PROPALOJCOMAQUEC), a proporção de alojamentos sobrelotados (PROPALOJSOBRELOT) e a cobertura populacional do ensino superior (COBPOPES), representando-se da seguinte forma:

$$(26) \text{LVALOMEDIANO} = \beta_0 + \beta_1 \text{AL} + \beta_2 \text{PROPPOPCOM23OUMAISSCOMES} + \\ \beta_3 \text{DURACAOMEDIAMOVPEND} + \beta_4 \text{GSE\_11\_20\_23} + \\ \beta_5 \text{PROPALOJCOMAQUEC} + \beta_6 \text{PROPALOJSOBRELOT} + \beta_7 \text{COBPOPES} + u.$$

Considerando a definição de cada uma das variáveis, a relação entre a variável dependente e as variáveis AL, PROPPOPCOM23OUMAISSCOMES, PROPALOJCOMAQUEC, COBPOPES, PROPALOJSOBRELOT é positiva. E é negativa relativamente às variáveis DURACAOMEDIAMOVPEND e GSE\_11\_20\_23.

A relação positiva entre os preços e a proporção de alojamentos sobrelotados reflete o problema de acesso às casas devido à incapacidade dos indivíduos em aceder aos preços de mercado, implicando que os filhos permaneçam em casa dos pais ou adquiram casas mais pequenas.

Analisando a *Tabela B-I*, pode observar-se que as variáveis AL ( $p\text{-value}=0,000$ ), PROPOPCOM23OUMAISSCOMES ( $p\text{-value}=0,007$ ), DURACAOMEDIAMOVPEND ( $p\text{-value}=0,000$ ), GSE\_11\_20\_23 ( $p\text{-value}=0,001$ ), PROPALOJSOBRELOT ( $p\text{-value}=0,004$ ) apresentam-se como estatisticamente significativas e as variáveis PROPALOJCOMAQUEC ( $p\text{-value}=0,863$ ) e COBPOPES ( $p\text{-value}=0,698$ ) não se apresentam como estatisticamente significativas, para explicar a variável dependente, uma vez que apresentam um  $p\text{-value}$  superior ao nível de significância (5%). As variáveis do modelo explicam cerca de 77,77% ( $R^2 = 0,7777$ ) da variação da variável dependente.

Relativamente ao teste F, teste de significância global das variáveis, obteve-se uma estatística F de 52,506 com um  $p\text{-value}=0<0,05$ , rejeitando-se a hipótese nula e confirmando a adequabilidade do modelo, na sua globalidade, para explicar a variável dependente.

Analisando a *Tabela B-II- Teste Breusch-Pagan-Godfrey*, verifica-se que  $p\text{-value} = 0,254>0,05$ . Não se rejeita a hipótese nula, evidenciando ausência de heterocedasticidade. De acordo com o teste de *Breusch-Godfrey (Tabela B-III)*,  $p\text{-value}=0,229>0,05$ , não se rejeita a hipótese nula constatando ausência de autocorrelação.

O teste *RESET (Tabela B-IV)* permite-nos concluir que o modelo não apresenta evidências de má especificação devido a um  $p\text{-value}=0,830>0,05$ , o que impede a rejeição da hipótese nula (boa qualidade de ajustamento).

Da análise do histograma dos resíduos (*Figura B-1*), verifica-se que estes seguem uma distribuição normal, resultado confirmado pela estatística de teste de JB=5,858 e um  $p\text{-value}=0,053>0,05$ .

Analisando a *Tabela B-V*, comprova-se ausência de multicolinearidade uma vez que todas as variáveis apresentam  $VIF < 10$ .

### 4.3. Modelo Cluster 1

Este modelo relaciona o preço mediano (LVALORMEDIANO) com as variáveis proporção da população pertencente aos grupos socioeconómicos 11, 20 e 23 (GSE\_11\_20\_23), duração média dos movimentos pendulares da população empregada ou estudante (DURACAOMEDIAMOVPEND), cobertura populacional dos estabelecimentos de Ensino Básico (COBPOPEB), percentagem do alojamento local nos alojamentos clássicos (AL), proporção de alojamentos com aquecimento central (PROPALOJCOMAQUEC) e proporção de alojamentos sobrelotados (PROPALOJSOBRELOT) sendo representado pela equação a seguir apresentada:

$$(27) \text{LVALORMEDIANO} = \beta_0 + \beta_1 \text{GSE}_{11_20_23} + \beta_2 \text{DURACAOMEDIAMOVPEND} + \beta_3 \text{COBPOPEB} + \beta_4 \text{AL} + \beta_5 \text{PROPALOJCOMAQUEC} + \beta_6 \text{PROPALOJSOBRELOT} + u.$$

O Output do modelo do *cluster* 1 é apresentado na *Tabela C-I*, revelando que as variáveis GSE\_11\_20\_23 e DURACAOMEDIAMOVPEND têm um impacto negativo na variável dependente, por outro lado as variáveis COBPOPEB, AL, PROPALOJCOMAQUEC e PROPALOJSOBRELOT têm um impacto positivo.

As variáveis COBPOPEB ( $p\text{-value}=0,415$ ) e PROPALOJCOMAQUEC ( $p\text{-value}=0,061$ ) não têm significância estatística para explicarem o comportamento da variável dependente apresentando um  $p\text{-value}$  superior a 0.05. Por outro lado, as variáveis GSE\_11\_20\_23 ( $p\text{-value}=0,000$ ), DURACAOMEDIAMOVPEND ( $p\text{-value}=0,0062$ ), AL ( $p\text{-value}=0,0187$ ), PROPALOJSOBRELOT ( $p\text{-value}=0,002$ ), são estatisticamente significativas. Com um  $R^2 = 0,599976$ , constata-se que o modelo tem uma capacidade explicativa de aproximadamente 60%.

As variáveis apresentam significância estatística, conjunta, significativa com uma estatística  $F=14,999$  e um  $p\text{-value}= 0,000$ .

De acordo com o teste de heterocedasticidade (*Tabela C-II*) o modelo apresenta um  $p\text{-value}= 0,371 > 0,05$  o que implica a não rejeição da hipótese nula da presença de homocedasticidade. O modelo não apresenta vestígios de autocorrelação, isto porque analisando a *Tabela C-III -Teste Breusch-Godfrey*, verifica-se que este teste apresenta um  $p\text{-value}$



$value=0,901>0,05$ , sugerindo a não rejeição da hipótese nula da ausência de autocorrelação.

Uma vez que o teste *RESET* (*Tabela C-IV*) apresenta um  $p-value=0,127>0,05$ , não há evidências de má qualidade no ajustamento do modelo.

O teste de *Jarque-Bera* (*Figura C-1*) apresenta um  $p-value=0,224>0,05$  rejeitando a hipótese nula e evidenciando a normalidade dos resíduos.

As variáveis apresentam um  $VIF<10$  (*Tabela C-V*), constatando que o modelo respeita o pressuposto de ausência de multicolinearidade entre as variáveis.

#### 4.4. Modelo Cluster 2

As variáveis que melhor explicam o preço mediano (*LVALORMEDIANO*) dos alojamentos pertencentes ao *cluster 2*, são a proporção da população pertencente aos grupos socioeconómicos 11, 20 e 23 (*GSE\_11\_20\_23*), a proporção de edifícios com necessidade de reparação (*PROPEDIFNECESSREPARACO*), a duração média dos movimentos pendulares da população empregada ou estudante (*DURACAOMEDIAMOVPEND*) e a percentagem do alojamento local nos alojamentos clássicos (*AL*), formando assim o melhor modelo para este grupo:

$$(28) \quad LVALORMEDIANO = \beta_0 + \beta_1 GSE_{11_20_23} + \beta_2 PROPEDIFNECESSREPARACO + \beta_3 DURACAOMEDIAMOVPEND + \beta_4 AL + u$$

Analisando o output da regressão (*Tabela D-I*), conclui-se que as variáveis *GSE\_11\_20\_23* e *DURACAOMEDIAMOVPEND* têm um impacto negativo no valor das transações, enquanto que *PROPEDIFNECESSREPARACO* e *AL* apresentam um impacto positivo.

O impacto positivo da proporção de edifícios com necessidades de grandes reparações ou muito degradadas nos preços dos alojamentos (*PROPEDIFNECESSREPARACO*), pode ser consequência de obras de reestruturação ou reabilitação de edifícios ocorridas com o tempo, implicando a valorização dos alojamentos neles situados.

De entre as variáveis do modelo, a proporção de alojamento local nos alojamentos clássicos (*AL*) é a única que apresenta um  $p-value$  acima do nível de significância ( $p$ -

$value=0,233>0,05$ ) o que significa que não tem significância estatística individual no modelo. Por outro lado, as variáveis GSE\_11\_20\_23 ( $p-value=0,000$ ), PROPEDIFNECES-SREPARACO ( $p-value=0,005$ ) e DURACAOMEDIAMOVPEND ( $p-value=0,010$ ), apresentam um valor  $p$  inferior a 5% revelando a sua significância estatística individual no modelo. A capacidade explicativa é revelada pelo coeficiente de determinação  $R^2 = 0,837$ , querendo com isto dizer que as variáveis independentes explicam cerca de 83,7% da variável dependente.

Quanto à significância conjunta conclui-se que as variáveis, em conjunto, explicam muito bem a variável dependente apresentando uma estatística  $F=31,978$  e  $p-value=0$ .

Dos testes de *Breusch-Pagan-Godfrey* (Tabela D-II) e *Breusch-Godfrey* (Tabela D-III) resultaram  $p-value$  de 0,356 e 0,477 respetivamente, sugerindo ausência de heterocedasticidade e de autocorrelação no modelo.

O teste *RESET* apresenta um  $p-value=0,667>0,05$  (Tabela D-IV), valor que indica que o modelo não apresenta evidências de má qualidade de ajustamento.

Quanto à normalidade dos resíduos, do teste de *Jarque-Bera* (Figura D-1) resultou um  $p-value=0,597>0,05$ , sugerindo que os resíduos do modelo seguem uma distribuição normal.

Verifica-se  $VIF<10$  para todas as variáveis explicativas (Tabela D-V) evidenciando ausência de multicolinearidade.

#### **4.5. Modelo Cluster 3**

Neste modelo relaciona-se o preço mediano dos alojamentos (LVALORMEDIANO) com a duração média dos movimentos pendulares da população empregada ou estudante (DURACAOMEDIAMOVPEND), a idade média dos edifícios (IDADEMEDIAMEDIF), a cobertura territorial dos transportes (COBTERRANSP), a proporção da população com 23 anos ou mais e com ensino superior (PROPPOPCOM23OUMAISSCOMES) e a proporção de edifícios não exclusivamente residenciais (PROPEDIFNAOEXCLUSRESIDE) e é representado da seguinte forma:

$$(29) LVALORMEDIANO = \beta_0 + \beta_1 DURACAOMEDIAMOVEND + \beta_2 IDADEMEDIAEDIF + \beta_3 COBTERTRANSP + \beta_4 PROPOPCOM23OUMAISSCOMES + \beta_5 PROPEDIFNAOEXCLUSRESIDE + u$$

Analisando a *Tabela E-I*, conclui-se que as variáveis DURACAOMEDIAMOVEND, COBTERTRANSP, PROPEDIFNAOEXCLUSRESIDE têm um impacto negativo na variável dependente e as variáveis IDADEMEDIAEDIF, PROPOPCOM23OUMAISSCOMES revelam um impacto positivo sobre o preço mediano dos alojamentos.

Relativamente à variável idade média dos edifícios (IDADEMEDIAEDIF) o impacto positivo nos preços poderá ser explicado pelo facto de edifícios antigos estarem mais propensos à realização de obras de reestruturação e reabilitação tendo como consequência a sua valorização monetária.

O impacto negativo da variável cobertura territorial dos Transportes (COBTERTRANSP) nos preços pode ser explicado pelo facto do *buffer* considerado ser “pequeno”, o que faz com que a proximidade dos transportes seja considerada uma desvantagem por causa dos ruídos causados.

As variáveis DURACAOMEDIAMOVEND ( $p\text{-value}=0,002$ ), IDADEMEDIAEDIF ( $p\text{-value}=0,000$ ), COBTERTRANSP ( $p\text{-value}=0,001$ ), PROPOPCOM23OUMAISSCOMES ( $p\text{-value}=0,000$ ) são estatisticamente significativas para explicar o preço dos alojamentos e a variável PROPEDIFNAOEXCLUSRESIDE não é estatisticamente significativa apresentando um  $p\text{-value}=0,210 > 0,05$ . Nestas condições estas variáveis explicam 98,7% ( $R^2 = 0,987$ ) dos preços medianos.

A estatística  $F=154,494$  e o  $p\text{-value}=0,000$  revelam significância estatística conjunta das variáveis, confirmando a qualidade do modelo.

O teste de *Breusch-Pagan-Godfrey* (*Tabela E-II*) constata ausência de heterocedasticidade no modelo, pois apresenta um  $p\text{-value}=0,240 > 0,05$ . Quanto ao teste de *Breusch-Godfrey* (*Tabela E-III*) não há evidências de autocorrelação ( $p\text{-value}=0,586$ ).

Analisando o output do teste *RESET* o modelo está bem ajustado pois este apresenta um  $p\text{-value} = 0,161 > 0,05$  (*Tabela E-IV*) não rejeitando a hipótese nula da boa qualidade de ajustamento do modelo.

O teste de *Jarque-Bera* (*Figura E-1*) apresenta um  $p\text{-value} = 0,807$ , declarando a normalidade dos resíduos.

As variáveis não apresentam multicolinearidade uma vez que os respetivos valores de VIF não excedem o valor 10 (*Tabela E-V*).

## 5. Conclusão

A heterogeneidade da AML implicou a sua divisão em 3 grupos para uma melhor compreensão sobre que variáveis têm impacto no valor de transação de um alojamento, localizado numa determinada zona desta área. O presente estudo permitiu-nos perceber que o melhor modelo para a AML quando esta é considerada no seu todo, diferencia dos melhores modelos quando esta é dividida em grupos. O facto de um alojamento situar-se no centro de uma área ou mais próximo da periferia influencia o seu preço uma vez que cada freguesia tem as suas características, que podem ser definidas ou não pelo fator proximidade.

Após a análise dos modelos criados para cada um dos *clusters* e para a AML no seu todo, pode-se afirmar que nem todas as variáveis são significativas, quando consideradas individualmente, na explicação da variável dependente. Entretanto, quando consideradas conjuntamente satisfazem todos os pressupostos de regressão linear múltipla revelando um bom poder explicativo apresentando assim um coeficiente de determinação superior a 0,5 ( $R^2 > 0,5$ ).

Na AML conclui-se o modelo que melhor explica os preços, inclui as variáveis: proporção do alojamento local nos alojamentos clássicos (AL), proporção da população com 23 anos ou mais com ensino superior (PROPPOPCOM23OUMAISSCOMES), duração média dos movimentos pendulares da população empregada e estudante (DURACAOMEDIAMOVPEND), proporção da população pertencente aos grupos socioeconómicos 11, 20 e 23 (GSE\_11\_20\_23), proporção de alojamentos com aquecimento central (PROPALOJCOMAQUEC), proporção de alojamentos sobrelotados (PROPALOJSOBRELOT) e cobertura populacional do ensino superior (COBPOPES).

No *Cluster* 1, o modelo que melhor explica os preços inclui as variáveis: proporção da população pertencente aos grupos socioeconómicos 11, 20 e 23 (GSE\_11\_20\_23), duração média dos movimentos pendulares da população empregada e estudante (DURACAOMEDIAMOVPEND), cobertura populacional do ensino básico (COBPOPEB), proporção do alojamento local nos alojamentos clássicos (AL), proporção de alojamentos com aquecimento central (PROPALOJCOMAQUEC) e proporção de alojamentos sobrelotados (PROPALOJSOBRELOT).

Conclui-se que no *Cluster 2* o modelo que melhor explica os preços inclui as variáveis: proporção da população pertencente aos grupos socioeconómicos 11,20 e 23 (GSE\_11\_20\_23), proporção de edifícios com necessidades de grandes reparações ou muito degradadas (PROPEDIFNECESSREPARACOES), duração média dos movimentos pendulares da população empregada e estudante (DURACAOMEDIAMOVPEND) e proporção do alojamento local nos alojamentos clássicos (AL).

O modelo que melhor explica o *Cluster 3* inclui as seguintes variáveis: duração média dos movimentos pendulares da população empregada e estudante (DURACAOMEDIAMOVPEND), idade média dos edifícios (IDADEMEDIAEDIF), cobertura territorial dos transportes(COBTERTRANSP), proporção da população com 23 anos ou mais com ensino superior (PROPOPCOM23OUMAISSCOMES) e proporção de edifícios com necessidades de grandes reparações ou muito degradadas (PROPEDIFNECESSREPARACOES).

A duração média dos movimentos pendulares da população empregada ou estudante (DURACAOMEDIAMOVPEND) é a única variável comum a todos os modelos obtidos. Revela-se como uma variável de extrema importância pois funciona como um indicador de acessibilidade numa determinada zona, isto é, quanto maior for a duração média dos movimentos pendulares menor será o nível de acessibilidade de um alojamento em termos dos serviços de transportes, estabelecimentos de ensino e de outros serviços dentro da área em estudo.

A variável GSE\_11\_20\_23 (que representa as profissões menos qualificadas) está presente no modelo global e nos modelos *cluster 1* e *cluster 2* revelando um impacto negativo na variável dependente o que pode significar que o mercado imobiliário cresce essencialmente para as classes médias e superiores.

O fenómeno alojamento local (fenómeno recente<sup>9</sup>), tem tido um grande impacto nos preços das casas, impacto confirmado pelo presente estudo revelando-se assim como uma variável explicativa dos preços medianos na AML, com elevada

---

<sup>9</sup> Estudos apontam que os primeiros registos de alojamento local ocorreram em 2010, mas o *boom* deste fenómeno ocorreu em 2014 continuando até o presente momento. Soldado, C. (2018).

significância estatística ( $p\text{-value}=0.000$ ). Uma das grandes consequências deste fenómeno são as alterações significativas nos alojamentos e edifícios verificadas, resultantes de obras de reestruturação e de reabilitação realizadas.

Com a evolução do tempo ocorrem algumas transformações, umas mais significativas que outras. No decorrer do estudo, apercebeu-se que o desfasamento do tempo entre o período a que se referem os dados utilizados na construção dos indicadores (*Tabela A-I*) e o período a que se referem os preços explicados (2016) constitui uma limitação pois, observaram-se algumas relações entre variáveis independentes e a variável dependente que só podem ser explicadas pelas transformações entretanto ocorridas (transformações nos edifícios, alojamentos e no espaço geográfico).

O presente estudo pode servir como modelo para outros estudos sobre os preços das casas, que permitirão aos agentes estar mais informados sobre os fatores que influenciam o valor da transação dos alojamentos durante um processo de avaliação imobiliária e num processo de transação de um imóvel.

Será interessante a realização de um estudo semelhante para a Área Metropolitana do Porto (AMP), que é também uma área onde tem-se verificado um aumento significativo dos preços das casas. Este estudo permitirá perceber se os valores de transação de alojamentos das duas áreas metropolitanas são influenciados pelas mesmas variáveis e, no caso afirmativo deve-se comparar o seu nível de impacto.

Será pertinente a inclusão, em estudos futuros, de mais indicadores relacionados com as características estruturais dos alojamentos, como por exemplo indicadores relacionados com as tipologias, com o estado do imóvel (novo ou usado), etc.

## 6. Bibliografia

- Bazyl, M. (2009). Hedonic price model for Warsaw housing market. Warsaw School of Economics, Institute of Econometrics, Department of Applied Econometrics.
- Branco, J. (2004). Uma Introdução à Análise de *Clusters* - Congresso anual da sociedade portuguesa de estatística
- Carvalho, S. (2012). Mais de 450 famílias vivem em casa sobrelotadas. Diário de notícias 26 de novembro de 2012 [Em linha]. Disponível em: <https://www.dn.pt/portugal/interior/mais-de-450-mil-familias-vivem-em-casas-sobrelotadas-2909498>.
- Claeys, G., Efstathiou, K. e Schoemaker, D. Spotting excessive regional house price growth and what to do about it. Policy contribution Issue nº26, October 2017 [Em linha]. Disponível em: <http://bruegel.org/wp-content/uploads/2017/10/PC-26-2017.pdf>
- Centro de Estudos de Gestão (CEGE), Análise de dados com IBM SPSS- Mix essencial para relatórios profissionais e teses académicas.
- Cruz, T. (2016). O impacto do alojamento local na reabilitação urbana em Lisboa. Disponível em: <https://fenix.tecnico.ulisboa.pt/download-File/281870113703300/O%20impacto%20do%20alojamento%20local%20na%20reabilitacao%20urbana%20em%20Lisboa.pdf>
- Dijol, J., Ghekière, L., Kiss, I., Pittini, A., (2015). The State of Housing in the EU 2015. A Housing Europe Review [Em linha]. Disponível em: <http://www.housingeurope.eu/resource-468/the-state-of-housing-in-the-eu-2015>
- Estrella, D. (2008). Analysis of the impact of London underground on the supply value of residential real estate.
- Fletcher, M., Gallimore, P. e Mangan, J. (2000). Heteroscedasticity in hedonic house price models - Journal of Property Research, 17(2), p. 93-108
- Follain, J.R. and Stephen Malpezzi (1980). Dissecting Housing Value and Rent. Washington, DC: The Urban Institute.



- Gaspar, A. (2011). Valorização dos Espaços Verdes Urbanos Através do Valor dos Imóveis Utilizando o Método do Preço Hedónico. Aplicado em Cinco Freguesias da Cidade de Lisboa-Dissertação de Mestrado.
- Goodall, B. (1972). The economic of urban areas. Urban and Regional Planning series, V (3), 84-88.
- Guerra, Isabel. (2008). Europa e Políticas habitacionais. Mudanças em curso. Cidades-Comunidades e Territórios Dez. 2008, n. °17, pp. 47-61
- INE (2013). Censos 2011-Preparação, Metodologia e Conceitos. Edição 2013.pp 166.ISSN: 2182-9594
- INE (2017). Destaque. Estatísticas de Preços da Habitação ao nível local- 2º Trimestre de 2017, pp 12. Disponível em : [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_destaques&DESTAQUES-dest\\_boui=282086493&DESTAQUEStema=00&DESTAQUESmodo=2](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUES-dest_boui=282086493&DESTAQUEStema=00&DESTAQUESmodo=2)
- INE (2017). Retrato Territorial de Portugal. Estatísticas oficiais. Edição 2017. pp.218. ISSN: 1646-0448
- INE (2014). Tipologia Socioeconómica das Áreas metropolitanas de Lisboa e Porto 2011. Estatísticas oficiais. Edição 2014. pp.104.ISSN 2183-1645
- Johnson, R. A., Wichern, D.W. (2002). Applied Multivariate Statistical Analysis, 4th Ed., Prentice- Hall, New Jersey
- Lourenço, R. e Rodrigues P. (2017). House prices in Portugal- what happened since the crisis
- Marôco, J. (2014). Análise Estatística com o *SPSS Statistics*, 6ª edição
- O Corvo (2018). Reabilitação Urbana em Alfama é quase sempre “selectiva” e “virada para o turismo” [Em linha]. Disponível em: <https://ocorvo.pt/reabilitacao-urbana-em-alfama-e-quase-sempre-selectiva-e-virada-para-o-turismo/> [Acesso em: 04/2018]

- Pereira, S. (2018). O problema da habitação em modo global. Expresso, 21-04-2018 semanal, p.32.
- Portugal lidera recuperação pós-crise. (2018, janeiro). Confidencial imobiliário, newsletter imobiliária Portuguesa, p.12-15. [Em linha]. Disponível em: <https://confidencialimobiliario.com/editorial/revista/2018-01/>
- Resolução do Conselho de Ministros nº 4/2015, 15/07/2015 - Regula o processo da habitação e a aprovação e implementação de uma estratégia para a habitação. Diário da República, 1ª série- nº 136
- Rodrigues, J. (2008). Especificação de um modelo Hedónico de preços de habitação para Portugal. (Tese de Mestrado)
- Soldado, C. (2018). Alojamento local em Lisboa disparou mais de 3000% desde 2010 [Em linha]. Disponível em: <https://www.publico.pt/2018/07/24/local/noticia/alojamento-local-em-lisboa-disparou-mais-de-3000-desde-2010-1838922>
- Tarré, A. (2009). Análise de Valores de avaliação de apartamentos no âmbito do Crédito a Habitação, para duas zonas distintas do concelho de Lisboa-recurso a Modelos Hedónicos. (Projeto de Mestrado)
- Teixeira, M. (2011). Modelos econométricos para el precio de los inmuebles: Un caso de estudio en Portugal. (Dissertação de Doutoramento).
- Vieira, R. (2011). Desenvolvimento de um modelo hedónico de avaliação de apartamentos aplicando coeficientes de localização, no concelho de Lisboa. (Tese de Mestrado)
- Wooldridge, M. (2000). Introductory Econometrics: a modern approach. South Western College.

## 7. ANEXO

### Anexo A - Apresentação dos Dados

*Tabela A- I-Indicadores Socioeconómicos*

Indicador	Unidade de medida	Fonte	Período de referência
Área média útil habitável dos alojamentos clássicos de residência habitual	m2	Censos	2011
Proporção de alojamentos familiares clássicos de residência habitual vagos	%	Censos	2011
Densidade de alojamentos familiares clássicos de residência habitual	Nº/km2	Censos	2011
Proporção de alojamentos familiares clássicos de residência habitual com aquecimento central	%	Censos	2011
Proporção de alojamentos familiares clássicos de residência habitual sobrelotados	%	Censos	2011
Proporção de alojamentos familiares clássicos de residência habitual com lugar de estacionamento ou garagem	%	Censos	2011
Proporção de edifícios com necessidades de grandes reparações ou muito degradadas	%	Censos	2011
Proporção de edifícios não exclusivamente residenciais	%	Censos	2011
Proporção de edifícios com 3 pisos ou mais e servidos por elevador	%	Censos	2011
Idade média dos Edifícios	anos	Censos	2011
Proporção da população residente de nacionalidade estrangeira	%	Censos	2011
Duração média dos movimentos pendulares da população empregada ou estudante	minutos	Censos	2011
Proporção da população com 23 anos ou mais e com ensino superior	%	Censos	2011
Proporção da população residente pertencente aos grupos socioeconómicos 1, 2, 4, 5, 6 e 7	%	Censos	2011
Proporção da população residente pertencente aos grupos socioeconómicos 11, 20 e 23	%	Censos	2011
Proporção da superfície da superfície destinada à construção nova	%	SIOU	2011-2014
Proporção da superfície destinada às obras de reabilitação	%	SIOU	2011-2015
Proporção do alojamento local (nos alojamentos clássicos)	%	Turismo de Portugal	Antes e depois de 2013
Cobertura territorial dos Espaços verdes	%	COS 2010	2010
Cobertura territorial dos estabelecimentos de ensino pré-escolar	%	DGEEC e DREM	2014-2017
Cobertura territorial dos estabelecimentos de ensino básico	%	DGEEC e DREM	2014-2014
Cobertura territorial dos estabelecimentos de ensino superior	%	DGEEC e DREM	2015-2016
Cobertura territorial dos transportes (autocarro, elétrico e metro)	%	OpenStreetMap	2018
Cobertura territorial dos comboios	%	IMOB	2017
Cobertura territorial das indústrias	%	COS 2010	2010
Cobertura populacional dos Espaços verdes	%	COS 2010	2010
Cobertura populacional dos estabelecimentos de ensino pré-escolar	%	DGEEC e DREM	2014-2017
Cobertura populacional dos estabelecimentos de ensino básico	%	DGEEC e DREM	2014-2017
Cobertura populacional dos estabelecimentos de ensino superior	%	DGEEC e DREM	2015-2016
Cobertura populacional dos transportes (autocarro, elétrico e metro)	%	OpenStreetMap	2018
Cobertura populacional dos comboios	%	IMOB	2017
Cobertura populacional das indústrias	%	COS 2010	2010

Fonte: INE

*Tabela A- II-Estatísticas Descritivas*

VARIÁVEIS	MÉDIA	MEDIANA	DESVIO PADRÃO	MÍNIMO	MÁXIMO
PREÇO MEDIANO	1199,17	989	555,6984328	477	3022
AL	1,25783	0,093639	3,81425	0	28,75516
PROPOPCOM23OUMAISSCOMES	19,739	17,1549	10,94754	4,86622	55,21152
DURACAOMEDIAMOVPEND	25,5928	25,48853	3,137666	18,7965	33,64603
GSE_11_20_23	8,33901	8,442292	3,397268	1,71225	15,6839
PROPALOJCOMAQUEC	84,9961	85,62348	4,767143	66,5394	95,1952
PROPALOJSOBRELOT	12,5378	11,783	3,341381	6,62309	23,21708
COBPOPES	37,3459	8,945233	42,9125	0	100
COBPOPEB	93,0967	100	15,39777	24,156	100
PROPEDIFNECESSREPARACO	4,69686	3,51822	3,35631	0,199	17,98501
IDADEMEDIAEDIF	40,7635	37,43416	14,03601	18,502	93,57315
COBTERTRANSP	37,6696	26,5193	35,62889	0	100
PROPEDIFNAOEXCLUSRESIDE	11,5029	9,69901	9,39282	0,39063	49,00162

Fonte: cálculos efetuados pela autora a partir dos dados fornecidos pelo INE relativos aos indicadores socioeconómicos que constam na *Tabela A-I*.

**Tabela A- III-Constituição dos clusters**

<i>Clusters</i>	<b>Constituição</b>
<i>cluster 1</i>	Alcabideche, São Domingos de Rana, União das freguesias de Cascais e Estoril, Alcântara, Belém, Bucelas, Fanhões, Loures, Lousa, União das freguesias de Santa Iria de Azoia, São João da Talha e Bobadela, União das freguesias de Santo Antão e São Julião do Tojal, União das freguesias de Santo António dos Cavaleiros e Frielas, União das freguesias de Camarate, Unhos e Apelação, Carvoeira, Encarnação, Ericeira, Mafra, Milharado, Santo Isidoro, União das freguesias de Azueira e Sobral da Abelheira, União das freguesias de Enxara do Bispo, Gradil e Vila Franca do Rosário, União das freguesias de Igreja Nova e Cheleiros, União das freguesias de Malveira e São Miguel de Alcniça, União das freguesias de Venda do Pinheiro e Santo Estevão das Galés, Barcarena, Porto Salvo, Colares, Rio de Mouro, União das freguesias de Almargem do Bispo, Pêro Pinheiro e Montelavar, União das freguesias de Queluz e Belas, União das freguesias de São João das Lampas e Terrugem, União das freguesias de Sintra (Santa Maria e São Miguel, São Martinho e São Pedro de Penaferrim), Vialonga, Vila Franca de Xira, União das freguesias de Alhandra São João dos Montes e Calhandriz, União das freguesias de Alverca do Ribatejo e Sobralinho, União das freguesias de Castanheira do Ribatejo e Cachoeiras, União das freguesias de Pontinha e Famões, União das freguesias de Ramada e Caneças, Alcochete, Samouco, São Francisco, Costa da Caparica, União das freguesias de Caparica e Trafaria, União das freguesias de Charneca de Caparica e Sobreda, Santo António da Charneca, União das freguesias de barreiro e lavradio, União das freguesias de Palhais e Coia, Alhos Vedros, Moita, União das freguesias de Gaio-Rosário e Sarilhos Pequenos, Canha, Sarilhos grandes, União das freguesias de Atalaia e Alto-Estanqueiro-Jardia, União das freguesias de Montijo e Afonsoeiro, União das freguesias de Pegões, Palmela, Pinhal Novo, Quinta do Anjo, União das freguesias de Poceirão e Marateca, Amora, Corroios, Fernão Ferro, União das freguesias do Seixal, Arrentela e Aldeia de Paio Pires, Sesimbra (Castelo), Sesimbra (Santiago), Quinta do Conde, Setúbal (São Sebastião), Gâmbia-Pontes-Alto da Guerra, Sado, União das freguesias de Azeitão (São Lourenço e São Simão), União das freguesias de Setúbal (São Julião, Nossa Senhora da Anunciada e Santa Maria da Graça)
<i>cluster 2</i>	União das freguesias de Carcavelos e Parede, Ajuda, Beato, Benfca, Campolide, Carnide, Lumiar, Marvila, Olivais, Alvalade, Estrela, Misericórdia, Parque das Nações, Santa Clara, Santa Maria Maior, União das freguesias de Sacavém e Prior Velho, União das freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo, União das freguesias de Carnaxide e Queijas, União das freguesias de Oeiras e São Julião da Barra, Paço de Arcos e Caxias, Algueirão-Mem Martins, Casal de Cambra, União das freguesias de Aqualva e Mira-Sintra, União das freguesias de Póvoa de Santa Iria e Forte da Casa, Alfragide, Mina de Água, Venteira, União das freguesias de Póvoa de Santo Adrião e Olival Basto, União das freguesias de Laranjeiro e Feijó, União das freguesias de Alto de Seixalinho, Santo André e Verderena e União das freguesias de baixa da banheira e Vale da Amoreira
<i>cluster 3</i>	São Domingos de Benfca, Areiro, Arroios, Avenidas Novas, Campo de Ourique, Penha de França, Santo António, São Vicente, União das freguesias de Moscavide e Portela, União das freguesias de Cacém e São Marcos, União das freguesias de Massamá e Monte Abraão, Águas Livres, Encosta do Sol, Falagueira-venda Nova, Odivelas e União das freguesias de Almada, Cova da Piedade, Pragal e Cacilhas

Fonte: INE

**Anexo B - Modelo AML****Tabela B- I- Output (modelo AML)**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.378430	0.637454	11.57485	0.0000
AL	0.025522	0.005751	4.437863	0.0000
PROPPOPCOM23OUMISCO	0.013134	0.004801	2.735795	0.0073
DURACAOMEDIAMOVPEND	-0.028364	0.007463	-3.800763	0.0002
GSE_11_20_23	-0.046633	0.013727	-3.397106	0.0010
PROPALOJCOMAQUEC	0.001195	0.006926	0.172512	0.8634
PROPALOJSOBRELOT	0.025451	0.008524	2.985790	0.0035
COBPOPES	0.000267	0.000686	0.388814	0.6982
R-squared	0.777797	Mean dependent var	6.998059	
Adjusted R-squared	0.762984	S.D. dependent var	0.416867	
S.E. of regression	0.202949	Akaike info criterion	-0.283558	
Sum squared resid	4.324774	Schwarz criterion	-0.090468	
Log likelihood	24.02102	Hannan-Quinn criter.	-0.205204	
F-statistic	52.50592	Durbin-Watson stat	1.620333	
Prob(F-statistic)	0.000000			

**Tabela B- II- Teste Breusch-Pagan-Godfrey(modelo AML)**

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	1.306854	Prob. F(7,105)	0.2543	
Obs*R-squared	9.055975	Prob. Chi-Square(7)	0.2487	
Scaled explained SS	10.92083	Prob. Chi-Square(7)	0.1421	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/12/18 Time: 19:14				
Sample: 1 118				
Included observations: 113				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.019183	0.199902	0.095964	0.9237
AL	-0.001659	0.001803	-0.919608	0.3599
PROPPOPCOM23OUMAISCO	-0.000793	0.001505	-0.526809	0.5994
DURACAOMEDIAMOVPEND	-0.001396	0.002340	-0.596445	0.5522
GSE_11_20_23	0.002535	0.004305	0.588962	0.5572
PROPALOJCOMAQUEC	0.000199	0.002172	0.091805	0.9270
PROPALOJSOBRELOT	0.002263	0.002673	0.846411	0.3992
COBPOPES	0.000180	0.000215	0.836075	0.4050
R-squared	0.080141	Mean dependent var	0.038272	
Adjusted R-squared	0.018817	S.D. dependent var	0.064251	
S.E. of regression	0.063644	Akaike info criterion	-2.602872	
Sum squared resid	0.425303	Schwarz criterion	-2.409782	
Log likelihood	155.0622	Hannan-Quinn criter.	-2.524518	
F-statistic	1.306854	Durbin-Watson stat	1.917679	
Prob(F-statistic)	0.254302			

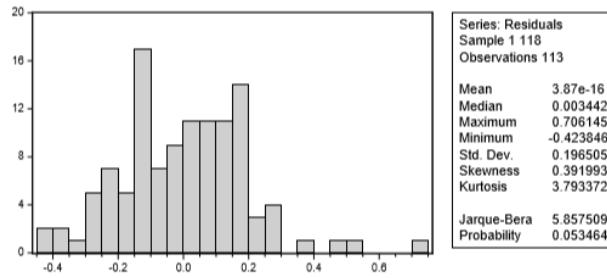
**Tabela B- III- Teste Breusch-Godfrey(modelo AML)**

Breusch-Godfrey Serial Correlation LM Test				
Null hypothesis: No serial correlation at up to 2 lags				
F-statistic	1.493528	Prob. F(2,103)	0.2294	
Obs*R-squared	3.184704	Prob. Chi-Square(2)	0.2034	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 06/12/18 Time: 19:14				
Sample: 1 118				
Included observations: 113				
Presample and interior missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.104554	0.637881	-0.163908	0.8701
AL	-0.000311	0.005728	-0.054222	0.9569
PROPPOPCOM23OUMAISCO	-0.000525	0.004889	-0.107364	0.9147
DURACAOMEDIAMOVPEND	0.000696	0.007441	0.093583	0.9256
GSE_11_20_23	0.000397	0.013816	0.028729	0.9771
PROPALOJCOMAQUEC	0.001001	0.006955	0.143988	0.8858
PROPALOJSOBRELOT	0.000347	0.008509	0.040807	0.9675
COBPOPES	9.79E-05	0.000686	0.142708	0.8868
RESID(-1)	0.141265	0.106358	1.328202	0.1870
RESID(-2)	0.085923	0.106536	0.806515	0.4218
R-squared	0.028183	Mean dependent var	3.87E-16	
Adjusted R-squared	-0.056733	S.D. dependent var	0.196505	
S.E. of regression	0.202002	Akaike info criterion	-0.276748	
Sum squared resid	4.202888	Schwarz criterion	-0.035386	
Log likelihood	25.63624	Hannan-Quinn criter.	-0.178806	
F-statistic	0.331895	Durbin-Watson stat	1.864149	
Prob(F-statistic)	0.962544			

**Tabela B- IV- Teste RESET(modelo AML)**

Ramsey RESET Test				
Equation: UNTITLED				
Specification: LVALORMEDIANO C AL PROPOPCOM23OUMAISSCOMES DURACAOMEDIAMOVPEND GSE_11_20_23 PROPALOJCOMAQUEC PROPALOJSOBRELOT COBPOPES				
Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	0.215792	104	0.8296	
F-statistic	0.046566	(1, 104)	0.8296	
Likelihood ratio	0.050585	1	0.8220	
F-test summary:				
	Sum of Sq	df	Mean Squares	
Test SSR	0.001936	1	0.001936	
Restricted SSR	4.324774	105	0.041188	
Unrestricted SSR	4.322839	104	0.041566	
LR test summary:				
	Value			
Restricted LogL	24.02102			
Unrestricted LogL	24.04631			
Unrestricted Test Equation:				
Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 19:15				
Sample: 1 118				
Included observations: 113				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.852788	11.48425	0.857939	0.3929
AL	0.043630	0.084109	0.518726	0.6051
PROPOPCOM23OUMAISSCO	0.022422	0.043309	0.517709	0.6058
DURACAOMEDIAMOVPEND	-0.046547	0.084597	-0.550223	0.5833
GSE_11_20_23	-0.074387	0.129354	-0.575068	0.5665
PROPALOJCOMAQUEC	0.001832	0.007559	0.242408	0.8089
PROPALOJSOBRELOT	0.041770	0.076106	0.548839	0.5843
COBPOPES	0.000493	0.001256	0.392656	0.6954
FITTED^2	-0.046006	0.213197	-0.215792	0.8296
R-squared	0.777897	Mean dependent var	6.998059	
Adjusted R-squared	0.760812	S. D. dependent var	0.416867	
S. E. of regression	0.203877	Akaike info criterion	-0.266306	
Sum squared resid	4.322839	Schwarz criterion	-0.049081	
Log likelihood	24.04631	Hannan-Quinn criter.	-0.178159	
F-statistic	45.53133	Durbin-Watson stat	1.614734	
Prob(F-statistic)	0.000000			

**Figura B- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo AML)**



**Tabela B- V- Teste de multicolinearidade (modelo AML)**

Variance Inflation Factors			
Date: 06/12/18 Time: 19:20			
Sample: 1 118			
Included observations: 113			
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.406347	1114.812	NA
AL	3.31E-05	1.516744	1.360872
PROPOPCOM23O	2.30E-05	33.34527	7.474677
DURACAOMEDIAMO	5.57E-05	102.1808	1.492626
GSE_11_20_23	0.000188	40.69073	5.884040
PROPALOJCOMAQU	4.80E-05	950.8696	2.813185
PROPALOJSOBREL	7.27E-05	33.69588	2.286938
COBPOPES	4.70E-07	4.336679	2.375531

**Anexo C - Modelo cluster 1**

**Tabela C- I- Output (modelo cluster 1)**

Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 17:00				
Sample: 1 72				
Included observations: 67				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.261221	0.813642	7.695305	0.0000
GSE_11_20_23	-0.060465	0.012607	-4.796059	0.0000
DURACAOMEDIAMOVP	-0.026546	0.009361	-2.835812	0.0062
COBPOPEB	0.001482	0.001805	0.821290	0.4147
AL	0.062797	0.025980	2.417137	0.0187
PROPALOJCOMAQUEC	0.014510	0.007593	1.910915	0.0608
PROPALOJSOBRELOT	0.037363	0.011760	3.177191	0.0024
R-squared	0.599976	Mean dependent var	6.845909	
Adjusted R-squared	0.559974	S.D. dependent var	0.303316	
S.E. of regression	0.201203	Akaike info criterion	-0.270398	
Sum squared resid	2.428958	Schwarz criterion	-0.040057	
Log likelihood	16.05834	Hannan-Quinn criter.	-0.179252	
F-statistic	14.99852	Durbin-Watson stat	1.880177	
Prob(F-statistic)	0.000000			



**Tabela C- II- Teste de Breusch-Pagan-Godfrey(modelo cluster 1)**

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	1.104308	Prob. F(6,60)	0.3706	
Obs*R-squared	6.663057	Prob. Chi-Square(6)	0.3531	
Scaled explained SS	6.550516	Prob. Chi-Square(6)	0.3644	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/12/18 Time: 17:03				
Sample: 1 72				
Included observations: 67				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.168534	0.230197	-0.732129	0.4669
GSE_11_20_23	0.005392	0.003567	1.511649	0.1359
DURACAOMEDIAMOVP	-0.002176	0.002648	-0.821589	0.4146
COBPOPEB	0.000537	0.000511	1.051900	0.2971
AL	2.08E-05	0.007350	0.002826	0.9978
PROPALOJCOMAQUEC	0.001386	0.002148	0.645202	0.5213
PROPALOJSOBRELOT	0.003576	0.003327	1.074690	0.2868
R-squared	0.099449	Mean dependent var	0.036253	
Adjusted R-squared	0.009393	S.D. dependent var	0.057194	
S.E. of regression	0.056925	Akaike info criterion	-2.795568	
Sum squared resid	0.194425	Schwarz criterion	-2.565227	
Log likelihood	100.6515	Hannan-Quinn criter.	-2.704421	
F-statistic	1.104308	Durbin-Watson stat	1.878909	
Prob(F-statistic)	0.370561			

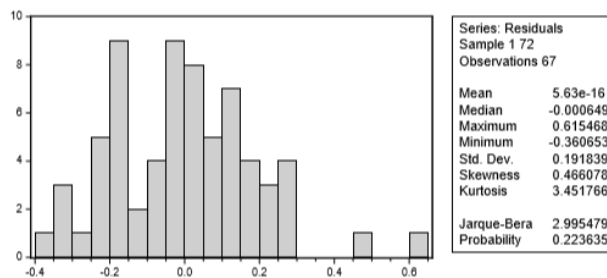
**Tabela C- III- Teste de Breusch-Godfrey(modelo cluster 1)**

Breusch-Godfrey Serial Correlation LM Test:				
Null hypothesis: No serial correlation at up to 1 lag				
F-statistic	0.015508	Prob. F(1,59)	0.9013	
Obs*R-squared	0.017606	Prob. Chi-Square(1)	0.8944	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 06/12/18 Time: 17:02				
Sample: 1 72				
Included observations: 67				
Presample and interior missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.006093	0.821858	-0.007414	0.9941
GSE_11_20_23	-2.89E-05	0.012714	-0.002270	0.9982
DURACAOMEDIAMOVP	-3.57E-05	0.009443	-0.003775	0.9970
COBPOPEB	8.42E-06	0.001821	0.004624	0.9963
AL	0.000390	0.026383	0.014788	0.9883
PROPALOJCOMAQUEC	5.88E-05	0.007671	0.007664	0.9939
PROPALOJSOBRELOT	0.000109	0.011890	0.009198	0.9927
RESID(-1)	-0.018065	0.145061	-0.124531	0.9013
R-squared	0.000263	Mean dependent var	5.63E-16	
Adjusted R-squared	-0.118350	S.D. dependent var	0.191839	
S.E. of regression	0.202874	Akaike info criterion	-0.240810	
Sum squared resid	2.428319	Schwarz criterion	0.022437	
Log likelihood	16.06714	Hannan-Quinn criter.	-0.136643	
F-statistic	0.002215	Durbin-Watson stat	1.854188	
Prob(F-statistic)	1.000000			

**Tabela C- IV- Teste RESET(modelo cluster 1)**

Ramsey RESET Test				
Equation: UNTITLED				
Specification: LVALORMEDIANO C GSE_11_20_23 DURACAOMEDIAMOV PEND COBPOPEB AL PROPALOJCOMAQUEC PROPALOJSOBRELO T				
Omitted Variables: Powers of fitted values from 2 to 3				
	Value	df	Probability	
F-statistic	2.135901	(2, 58)	0.1273	
Likelihood ratio	4.761402	2	0.0925	
F-test summary:				
	Sum of Sq	df	Mean Squares	
Test SSR	0.166626	2	0.083312	
Restricted SSR	2.428958	60	0.040483	
Unrestricted SSR	2.262333	58	0.039006	
LR test summary:				
	Value			
Restricted LogL	16.05834			
Unrestricted LogL	18.43904			
Unrestricted Test Equation:				
Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 17:04				
Sample: 1 72				
Included observations: 67				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-389.8543	783.2103	-0.497765	0.6205
GSE_11_20_23	6.180195	11.98609	0.515614	0.6081
DURACAOMEDIAMOV	2.714949	5.261708	0.515983	0.6078
COBPOPEB	-0.150822	0.293241	-0.514329	0.6090
AL	-6.426541	12.44830	-0.516259	0.6076
PROPALOJCOMAQUEC	-1.477768	2.875283	-0.513956	0.6092
PROPALOJSOBRELOT	-3.813522	7.406339	-0.514900	0.6086
FITTED^2	14.10790	28.56648	0.493862	0.6233
FITTED^3	-0.640144	1.371051	-0.466900	0.6423
R-squared	0.627418	Mean dependent var	6.845909	
Adjusted R-squared	0.576027	S.D. dependent var	0.303316	
S.E. of regression	0.197499	Akaike info criterion	-0.281762	
Sum squared resid	2.262333	Schwarz criterion	0.014390	
Log likelihood	18.43904	Hannan-Quinn criter.	-0.164574	
F-statistic	12.20878	Durbin-Watson stat	2.055244	
Prob(F-statistic)	0.000000			

**Figura C- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo cluster 1)**



**Tabela C- V- Teste de multicolinearidade (modelo cluster 1)**

Variance Inflation Factors			
Date: 06/12/18 Time: 17:05			
Sample: 1 72			
Included observations: 67			
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.662013	1095.652	NA
GSE_11_20_23	0.000159	26.87643	2.098836
DURACAOMEDIAMO	8.76E-05	97.92439	1.216042
COBPOPEB	3.26E-06	44.73773	1.795403
AL	0.000675	2.492128	1.960958
PROPALOJCOMAQU	5.77E-05	706.3859	1.874376
PROPALOJSOBREL	0.000138	32.79840	1.820266

**Anexo D - Modelo cluster 2****Tabela D- I- Output (modelo cluster 2)**

Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 18:15				
Sample: 1 30				
Included observations: 30				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.772456	0.385015	22.78470	0.0000
GSE_11_20_23	-0.082368	0.019585	-4.205618	0.0003
PROPEDIFNECESSREPARACO	0.030988	0.010087	3.072141	0.0051
DURACAOMEDIAMOVPEND	-0.048298	0.017375	-2.779753	0.0102
AL	0.008753	0.007165	1.221748	0.2332
R-squared	0.836507	Mean dependent var	7.204942	
Adjusted R-squared	0.810348	S.D. dependent var	0.477606	
S.E. of regression	0.207993	Akaike info criterion	-0.151615	
Sum squared resid	1.081524	Schwarz criterion	0.081917	
Log likelihood	7.274232	Hannan-Quinn criter.	-0.076906	
F-statistic	31.97790	Durbin-Watson stat	1.613532	
Prob(F-statistic)	0.000000			

**Tabela D- II- Teste de Breusch-Pagan-Godfrey (modelo cluster 2)**

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	1.151237	Prob. F(4,25)	0.3558	
Obs*R-squared	4.666396	Prob. Chi-Square(4)	0.3233	
Scaled explained SS	3.201153	Prob. Chi-Square(4)	0.5247	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/12/18 Time: 18:18				
Sample: 1 30				
Included observations: 30				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.018298	0.094424	0.193782	0.8479
GSE_11_20_23	0.006715	0.004803	1.398012	0.1744
PROPEDIFNECESSREPARACO	-0.000786	0.002474	-0.317878	0.7532
DURACAOMEDIAMOVPEND	-0.000710	0.004261	-0.166603	0.8690
AL	-0.000691	0.001757	-0.393308	0.6974
R-squared	0.155547	Mean dependent var	0.036051	
Adjusted R-squared	0.020434	S.D. dependent var	0.051539	
S.E. of regression	0.051010	Akaike info criterion	-2.962593	
Sum squared resid	0.065050	Schwarz criterion	-2.729060	
Log likelihood	49.43889	Hannan-Quinn criter.	-2.887883	
F-statistic	1.151237	Durbin-Watson stat	2.565416	
Prob(F-statistic)	0.355781			

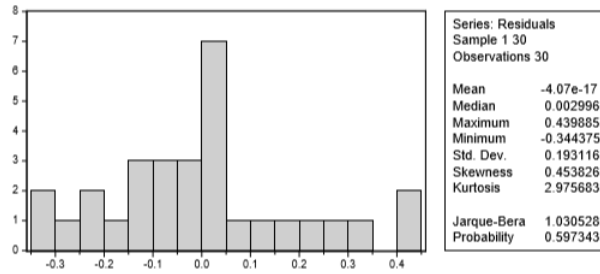
**Tabela D- III- Teste de Breusch-Godfrey(modelo cluster 2)**

Breusch-Godfrey Serial Correlation LM Test				
Null hypothesis: No serial correlation at up to 2 lags				
F-statistic	0.765504	Prob. F(2,23)	0.4766	
Obs*R-squared	1.872334	Prob. Chi-Square(2)	0.3921	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 06/12/18 Time: 18:18				
Sample: 1 30				
Included observations: 30				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.105086	0.398639	-0.263612	0.7944
GSE_11_20_23	-0.004085	0.020122	-0.203017	0.8409
PROPEDIFNECESSREPARACO	-0.002496	0.010505	-0.237554	0.8143
DURACAOMEDIAMOVPEND	0.005506	0.018175	0.302963	0.7646
AL	-0.001318	0.007353	-0.179217	0.8593
RESID(-1)	0.148530	0.218400	0.680080	0.5032
RESID(-2)	0.252998	0.244471	1.034879	0.3115
R-squared	0.062411	Mean dependent var	-4.07E-17	
Adjusted R-squared	-0.182177	S.D. dependent var	0.193116	
S.E. of regression	0.209972	Akaike info criterion	-0.082726	
Sum squared resid	1.014025	Schwarz criterion	0.244220	
Log likelihood	8.240888	Hannan-Quinn criter.	0.021867	
F-statistic	0.255168	Durbin-Watson stat	1.734857	
Prob(F-statistic)	0.952076			

**Tabela D- IV- Teste RESET(modelo cluster 2)**

Ramsey RESET Test				
Equation: UNTITLED				
Specification: LVALORMEDIANO C GSE_11_20_23 PROPEDIFNECESSR EPARACOES DURACAOMEDIAMOVPEND AL				
Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	0.436268	24	0.6665	
F-statistic	0.190330	(1, 24)	0.6665	
Likelihood ratio	0.236974	1	0.6264	
F-test summary:				
	Sum of Sq	df	Mean Squares	
Test SSR	0.008509	1	0.008509	
Restricted SSR	1.081524	25	0.043261	
Unrestricted SSR	1.073015	24	0.044709	
LR test summary:				
	Value			
Restricted LogL	7.274232			
Unrestricted LogL	7.392719			
Unrestricted Test Equation:				
Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 18:19				
Sample: 1 30				
Included observations: 30				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.417107	23.35947	-0.060665	0.9521
GSE_11_20_23	0.076583	0.364840	0.209853	0.8356
PROPEDIFNECESSREPARACO	-0.029756	0.139612	-0.213131	0.8330
DURACAOMEDIAMOVPEND	0.046284	0.217515	0.212784	0.8333
AL	-0.011592	0.047200	-0.245591	0.8081
FITTED^2	0.137102	0.314262	0.436268	0.6665
R-squared	0.837793	Mean dependent var	7.204942	
Adjusted R-squared	0.804000	S.D. dependent var	0.477606	
S.E. of regression	0.211445	Akaike info criterion	-0.092848	
Sum squared resid	1.073015	Schwarz criterion	0.187392	
Log likelihood	7.392719	Hannan-Quinn criter.	-0.003197	
F-statistic	24.79188	Durbin-Watson stat	1.598123	
Prob(F-statistic)	0.000000			

**Figura D- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo cluster 2)**



**Tabela D- V- Teste de multicolinearidade(modelo cluster 2)**

Variance Inflation Factors			
Date: 06/12/18 Time: 18:20			
Sample: 1 30			
Included observations: 30			
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.148237	102.7971	NA
GSE_11_20_23	0.000384	13.00912	2.301977
PROPEDEFNECESSR	0.000102	4.167873	1.477194
DURACAOMEDIAMO	0.000302	143.8281	2.308438
AL	5.13E-05	1.722911	1.547136

### Anexo E - Modelo cluster 3

**Tabela E- I-Output(modelo cluster 3)**

Dependent Variable: LVALORMEDIANO				
Method: Least Squares				
Date: 06/12/18 Time: 18:43				
Sample: 1 16				
Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.330411	0.521679	15.96846	0.0000
DURACAOMEDIAMOVPEND	-0.056371	0.013613	-4.140935	0.0020
IDADEMEDIAEDIF	0.014703	0.001317	11.16212	0.0000
COBTERTRANSP	-0.009466	0.001877	-5.042480	0.0005
PROPOPCOM230UMAISSCOMES	0.015996	0.003105	5.151058	0.0004
PROPEDEFNAOEXCLUSRESIDE	-0.004168	0.003109	-1.340471	0.2097
R-squared	0.987220	Mean dependent var	7.247281	
Adjusted R-squared	0.980830	S.D. dependent var	0.443263	
S.E. of regression	0.061372	Akaike info criterion	-2.463714	
Sum squared resid	0.037666	Schwarz criterion	-2.173993	
Log likelihood	25.70971	Hannan-Quinn criter.	-2.448878	
F-statistic	154.4937	Durbin-Watson stat	1.403781	
Prob(F-statistic)	0.000000			

**Tabela E- II- Teste de Breusch-Pagan-Godfrey (modelo cluster 3)**

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	1.623545	Prob. F(5, 10)	0.2403	
Obs*R-squared	7.168868	Prob. Chi-Square(5)	0.2084	
Scaled explained SS	2.058899	Prob. Chi-Square(5)	0.8409	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/12/18 Time: 18:48				
Sample: 1 16				
Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014494	0.022803	0.635631	0.5393
DURACAOMEDIAMOVPEND	-6.24E-05	0.000595	-0.104883	0.9185
IDAEMEDIAEDIF	-5.81E-05	5.76E-05	-1.009417	0.3366
COBTERTRANSP	-0.000111	8.21E-05	-1.356802	0.2047
PROPPOPCOM23OUMAISSCOMES	7.90E-06	0.000136	0.058167	0.9548
PROPEDIFNAOEXCLUSRESIDE	7.54E-05	0.000136	0.555119	0.5910
R-squared	0.448054	Mean dependent var	0.002354	
Adjusted R-squared	0.172081	S.D. dependent var	0.002948	
S.E. of regression	0.002683	Akaike info criterion	-8.724032	
Sum squared resid	7.20E-05	Schwarz criterion	-8.434312	
Log likelihood	75.79226	Hannan-Quinn criter.	-8.709196	
F-statistic	1.623545	Durbin-Watson stat	2.218726	
Prob(F-statistic)	0.240284			

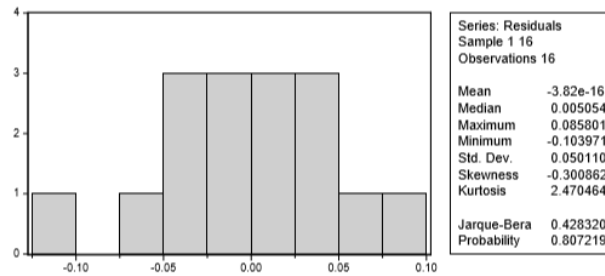
**Tabela E- III- Teste de Breusch-Godfrey (modelo cluster 3)**

Breusch-Godfrey Serial Correlation LM Test:				
Null hypothesis: No serial correlation at up to 2 lags				
F-statistic	0.571253	Prob. F(2,8)	0.5863	
Obs*R-squared	1.999463	Prob. Chi-Square(2)	0.3680	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 06/12/18 Time: 18:47				
Sample: 1 16				
Included observations: 16				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.526353	0.765213	0.687852	0.5110
DURACAOMEDIAMOVPEND	-0.013052	0.019529	-0.668335	0.5227
IDAEMEDIAEDIF	-0.000163	0.001544	-0.105346	0.9187
COBTERTRANSP	-0.001481	0.002416	-0.613037	0.5569
PROPPOPCOM23OUMAISSCOMES	-0.002821	0.004239	-0.665410	0.5245
PROPEDIFNAOEXCLUSRESIDE	0.000355	0.003269	0.108606	0.9162
RESID(-1)	0.453844	0.439278	1.033158	0.3317
RESID(-2)	0.152307	0.472339	0.322452	0.7554
R-squared	0.124966	Mean dependent var	-3.82E-16	
Adjusted R-squared	-0.640688	S.D. dependent var	0.050110	
S.E. of regression	0.064186	Akaike info criterion	-2.347207	
Sum squared resid	0.032959	Schwarz criterion	-1.960913	
Log likelihood	26.77766	Hannan-Quinn criter.	-2.327425	
F-statistic	0.163215	Durbin-Watson stat	1.973001	
Prob(F-statistic)	0.986364			

**Tabela E- IV- Teste RESET (modelo cluster 3)**

Ramsey RESET Test Equation: UNTITLED Specification: LVALORMEDIANO C DURACAOMEDIAMOVPEND IDADEMEDIAEDIF COBTERTRANSP PROPOPCOM23OUMAISSCOM ES PROPEDIFNAOEXCLUSRESIDENC Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	1.525951	9	0.1614	
F-statistic	2.328527	(1, 9)	0.1614	
Likelihood ratio	3.681592	1	0.0550	
F-test summary:				
	Sum of Sq	df	Mean Squares	
Test SSR	0.007742	1	0.007742	
Restricted SSR	0.037666	10	0.003767	
Unrestricted SSR	0.029924	9	0.003325	
LR test summary:				
	Value			
Restricted LogL	25.70971			
Unrestricted LogL	27.55051			
Unrestricted Test Equation: Dependent Variable: LVALORMEDIANO Method: Least Squares Date: 06/12/18 Time: 18:50 Sample: 1 16 Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.26397	10.45322	2.321196	0.0454
DURACAOMEDIAMOVPEND	-0.248963	0.126857	-1.982545	0.0813
IDADEMEDIAEDIF	0.063667	0.032111	1.982689	0.0787
COBTERTRANSP	-0.040875	0.020659	-1.978810	0.0792
PROPOPCOM23OUMAISSCOMES	0.068294	0.034396	1.985506	0.0784
PROPEDIFNAOEXCLUSRESIDE	-0.017212	0.009034	-1.905330	0.0891
FITTED^2	-0.231688	0.151832	-1.525951	0.1614
R-squared	0.989847	Mean dependent var	7.247281	
Adjusted R-squared	0.983078	S.D. dependent var	0.443263	
S.E. of regression	0.057662	Akaike info criterion	-2.568813	
Sum squared resid	0.029924	Schwarz criterion	-2.230806	
Log likelihood	27.55051	Hannan-Quinn criter.	-2.551505	
F-statistic	145.2370	Durbin-Watson stat	1.445718	
Prob(F-statistic)	0.000000			

**Figura E- 1- Histograma dos resíduos- Teste de Jarque-Bera (modelo cluster 3)**



**Tabela E- V- Teste de multicolinearidade (modelo cluster 3)**

Variance Inflation Factors Date: 06/12/18 Time: 18:52 Sample: 1 16 Included observations: 16			
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.272149	1156.057	NA
DURACAOMEDIAMO	0.000185	478.3404	8.928697
IDADEMEDIAEDIF	1.74E-06	22.04316	1.889836
COBTERTRANSP	3.52E-06	114.8035	2.944445
PROPOPCOM23O	9.64E-06	37.97893	6.444530
PROPEDIFNAOEXCL	9.67E-06	27.96754	3.863197