

## **MESTRADO**

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

## **TRABALHO FINAL DE MESTRADO**

### **PROJETO**

**“SOBREVOANDO” AS VENDAS DO DISTRITO DE LISBOA: UMA ANÁLISE  
DE ECONOMETRIA ESPACIAL**

**ANA RITA GERALDO CAGUERELHO**

**SETEMBRO - 2019**

## **MESTRADO**

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

## **TRABALHO FINAL DE MESTRADO PROJETO**

**“SOBREVOANDO” AS VENDAS DO DISTRITO DE LISBOA: UMA ANÁLISE  
DE ECONOMETRIA ESPACIAL**

**ANA RITA GERALDO CAGUERELHO**

**SUPERVISÃO:**  
**ISABEL PROENÇA**

**SETEMBRO - 2019**

*A todos os que me apoiaram  
nas fases boas, mas  
sobretudo aos que nunca me  
deixaram desistir.*

## ABSTRACT

The increasing and widespread use of technology, coupled with the increasing evolution of data storage capacity and the subsequent ease of computation, fosters the interest and the possibility to perform analytically more sophisticated studies using complex data structures.

This paper aims at analysing a spatial dataset on a company's food retail sales. Through both exploratory analysis and econometric models suitable for this purpose, it is intended to understand which factors influence the variation of sales volume in the various parishes of the Lisbon district and the importance of the spatial dimension in explaining this phenomenon.

From the panoply of available information, a set of variables considered relevant was defined to explain the variation in the sales volume of each parish. Through this data set, we carry out analyses supported by statistical software and software with graphical interface capability, through which it is possible to perform spatial exploratory analyses.

After several exploratory analyses of the data, it was concluded the need to include factors that consider spatial dependence. The presence of spatial dependence in the data under analysis was also confirmed empirically in the econometric modelling. This conclusion allows us to understand that the sales volume of a particular parish in the district of Lisbon is influenced by characteristics of that parish, but also by sales volume values registered in neighbouring parishes.

**KEYWORDS:** spatial econometrics, spatial autocorrelation, exploratory analysis, spatial data

## RESUMO

O incremento e a generalização do uso da tecnologia, aliados à crescente evolução da capacidade de armazenamento de dados e à sucessiva maior facilidade de computação, fomentam o interesse e a possibilidade em realizar estudos analiticamente mais sofisticados utilizando estruturas de dados complexas.

Este trabalho tem como objetivo analisar um conjunto de dados espaciais sobre vendas a retalho de uma empresa na área da alimentação. Através da análise exploratória e de modelos econométricos adequados para o efeito, pretende-se perceber quais os fatores que influenciam a variação do volume de vendas nas diversas freguesias do distrito de Lisboa e a importância da dimensão espacial na explicação deste fenómeno.

Da panóplia de informação disponível, foi definido um conjunto de variáveis consideradas relevantes para explicar a variação do volume de vendas de cada freguesia. Através deste conjunto de dados, procedemos a análises suportadas por softwares estatísticos e softwares com capacidade de interface gráfico, através dos quais é possível realizar análises exploratórias espaciais.

Após várias análises exploratórias dos dados, chegou-se à conclusão da necessidade da inclusão de fatores que ponderem a dependência espacial. A presença de dependência espacial nos dados em análise foi também confirmada empiricamente na modelação econométrica. Esta conclusão permite-nos perceber que o volume de vendas de determinada freguesia do distrito de Lisboa é influenciado por características dessa mesma freguesia, mas também por valores do volume de vendas registados em freguesias vizinhas.

**PALAVRAS-CHAVE:** Econometria espacial, dependência espacial, análise exploratória, dados espaciais

## AGRADECIMENTOS

Primeiramente gostaria de agradecer à professora Isabel Proença, pelas suas importantes sugestões, conhecimentos e disponibilidade ao longo desta jornada.

Ao Dr. Nuno Santos e à Produtos e Serviços de Estatística (PSE) pela disponibilização dos dados analisados.

A todos os professores e professoras que fizeram parte do meu percurso académico e que de alguma forma contribuíram para o meu desenvolvimento académico, que conduziu ao presente trabalho. Um particular agradecimento a todos os docentes do mestrado de Métodos Quantitativos para a Decisão Económica e Empresarial, pela forma como direcionam os seus esforços em prol dos alunos.

Aos meus pais e avós, pelos princípios e educação que me deram, que ainda hoje me guiam e que são responsáveis por todas as minhas conquistas.

Ao Daniel por todo o apoio, paciência e carinho, que em muito contribui para o meu sucesso.

Por último, gostaria de agradecer ao Instituto Superior de Economia e Gestão e a todos os seus colaboradores, cuja influência é crucial na formação de profissionais de excelência.

TABELA DE CONTEÚDOS

Abstract.....	i
Resumo .....	ii
Agradecimentos .....	iii
Tabela de Conteúdos .....	iv
Glossário .....	vii
1. Introdução .....	8
2. Revisão da Literatura.....	9
3. Metodologia .....	12
3.1. Abordagem Espacial.....	12
3.2. Matriz W de ponderação espacial.....	13
3.3 Modelos econométricos espaciais .....	15
3.3.1 Modelos SAR .....	16
3.3.2 Modelos SEM.....	17
3.3.3 Modelos SAC .....	18
3.3.4 Modelos SDM.....	18
3.4. Índices de Dependência Espacial .....	18
3.2.1. Índice de Moran (I) .....	18
3.4.2. Indicador Local de Dependência Espacial .....	20
4. Análise Exploratória dos Dados.....	21
4.1 Caracterização dos Dados .....	21
4.2 Caracterização das Variáveis .....	21
4.3 Análise Exploratória de Dados Espaciais .....	24
4.4 Dependência Espacial Global.....	25
4.5 Dependência Espacial Local .....	26

4.6 Dependência Espacial Bivariada .....	26
5. Modelação e resultados .....	27
5.1 Modelo Econométrico Espacial.....	27
6. Conclusões e Possíveis Desenvolvimentos Futuros.....	32
Referências .....	34
Apêndice.....	38
<b>Tabelas</b> .....	38
<b>TABELA A1</b> – Estatística Descritiva das variáveis incluídas no modelo final.....	38
<b>TABELA A2</b> – Matriz de Correlação .....	38
<b>FIGURA 1</b> – Diagrama de Dispersão de Moran para a variável vendas – Ano 2018 .....	39
<b>FIGURA 2</b> – Estatística I de Moran Local referente à variável vendas – Ano 2018 .....	39
<b>Figura 3</b> – Estatística I de Moran Bivariada Global.....	40
<b>FIGURA 4</b> – Estatística I de Moran Bivariada Local – Ano 2018.....	40
<b>Output 1</b> – Resultados da estimação por MQO.....	41
<b>Output 2</b> – Resultados da estimação do modelo SAR com todas as variáveis por MV .....	41
<b>Output 3</b> – Resultados da estimação do modelo SAR restrito por MV (após extração das variáveis irrelevantes) .....	42
<b>Output 4</b> – Resultados da estimação do modelo SEM com todas as variáveis por MV .....	42
<b>Output 5</b> – Resultados da estimação do modelo SEM restrito por MV (após extração das variáveis irrelevantes) .....	43
<b>Output 6</b> – Resultados da estimação do modelo SAC com todas as variáveis por MV .....	43



<b>Output 7</b> – Resultados da estimação do modelo SAC restrito por MV (após extração das variáveis irrelevantes) .....	44
<b>Output 8</b> – Resultados da estimação do modelo SAR irrestrito por Pseudo Máxima Verosimilhança .....	44
<b>Output 9</b> – Resultados da estimação do modelo SAR restrito por Pseudo Máxima Verosimilhança .....	45
<b>Output 10</b> – Resultados da estimação do modelo SEM irrestrito por Pseudo Máxima Verosimilhança .....	45
<b>Output 11</b> – Resultados da estimação do modelo SEM restrito por Pseudo Máxima Verosimilhança .....	46
<b>Output 12</b> – Resultados da estimação do modelo SAC irrestrito por Pseudo Máxima Verosimilhança .....	46
<b>Output 13</b> – Resultados da estimação do modelo SAC restrito por Pseudo Máxima Verosimilhança .....	47

GLOSSÁRIO

LISA – Indicador espacial de associação local (*Local Indicators of Spatial Association*).

MQO – Mínimos Quadrados Ordinários

SAC – *Spatial Lag Combined*

SAR – *Spatial Autoregressive Model*

SDM – *Spatial Durbin Model*

SEM – *Spatial Error Model*

AEDE – Análise Espacial de Dados Espaciais

MV – Máxima Verosimilhança

PMV – Pseudo Máxima Verosimilhança

PSE – Produtos e Serviços de Estatística

TFM – Trabalho Final de Mestrado

## 1. INTRODUÇÃO

Primordialmente, como retratado em Fujita *et al.* (1999), era considerada a inexistência de economias externas, sendo os estudos e modelos centrados apenas em cidades isoladas. Contudo esta teoria foi dando lugar a conjeturas que afirmam a possibilidade de existência de interações no espaço, denominadas em Krugman (1979) como economias de escala.

Neste sentido verifica-se um incremento da importância dada ao fenómeno designado por dependência espacial, ou seja, ao modo como valores assumidos por determinadas variáveis podem ser influenciados pela sua localização e por valores assumidos em regiões vizinhas. Este fator, segundo Anselin & Bera (1998), é definido como similaridade de valores em localidades próximas.

Várias são as análises que retratam o aumento do número de estudos que incorporam efeitos espaciais. Estas técnicas começam a fazer parte da econometria aplicada, facilitando o estudo de dados, como por exemplo, ao nível da agricultura, recursos económicos ou, até mesmo, ao nível do bem habitação.

Aliado a este acréscimo surgem as necessidades de melhorias nas capacidades de especificação e interpretação de modelos de dados de natureza espacial. No que concerne à presença de dependência espacial no modelo, podemos fazer distinção entre dois tipos: modelos de defasamento espacial e modelos de erro espacial.

O presente Trabalho Final de Mestrado tem como foco a especificação e estimação de modelos que permitam perceber, tendo em conta os dados em estudo, quais os fatores que influenciam o volume de vendas da empresa de retalho em análise. Perante a situação anteriormente exposta, o propósito geral deste estudo é identificar as causas que possam estar na origem das variações existentes na quantidade de vendas de determinadas regiões.

Um dos objetivos deste estudo é identificar a relevância da dependência espacial entre os locais onde os indivíduos vivem para determinar o volume de vendas da empresa em questão.

Este trabalho encontra-se estruturado em seis capítulos. No primeiro capítulo encontra-se a Introdução. O segundo capítulo inclui uma breve revisão da literatura sobre estudos empíricos que tratam a dependência espacial na modelação de fenómenos próximos das vendas a retalho, variável em estudo neste trabalho. O terceiro capítulo apresenta a

metodologia que fundamenta o estudo em causa, definindo as principais medidas e modelos a adotar na presença de dependência espacial nos dados. O quarto capítulo descreve os dados, as variáveis consideradas relevantes e é elaborada uma análise exploratória dos dados com vista a analisar a importância da componente espacial. O quinto capítulo contém a modelação econométrica e análise dos resultados obtidos após definição dos pressupostos e modelos mais adequados aos dados em estudo. No sexto capítulo são apresentadas as principais conclusões da análise efetuada e um conjunto de limitações do estudo com sugestões para trabalhos futuros.

## 2. REVISÃO DA LITERATURA

O tema da dependência espacial tem presenciado um incremento da sua importância nas últimas décadas. A sua crescente utilização esboçou o desenvolvimento de mecanismos que se encontram diretamente relacionados com o estudo de fenómenos espaciais.

O objetivo desta seção é referenciar, de forma sucinta, um conjunto de trabalhos empíricos que estudam problemas próximos daquele que é objeto desta tese, tendo em consideração a dependência espacial.

No seu estudo, Besner (2002), considera a oscilação do preço das residências, atendendo a alterações verificadas quer em termos de localização, quer em termos do preço de residências na vizinhança. Considera que as decisões dos agentes económicos estão subjacentes às características partilhadas entre os bens com localização comum. De forma semelhante, em 2009, Rogers & Winter aplicam o conceito de dependência espacial a dados de habitações onde é retratada a questão das execuções hipotecárias.

No seu estudo, Besner (2002), incorpora a componente espacial através de um modelo *spatial lag*. o autor procede a uma análise do comportamento dos agentes económicos e da estrutura de preços, recorrendo a testes de duas amostras: uma delas composta exclusivamente por condomínios (próximos do CBD<sup>1</sup>) e a outra composta exclusivamente por moradias individuais unifamiliares. Cada uma das amostras deu origem a um modelo com as respetivas variáveis explicativas. Através dos resultados obtidos, Besner (2002) concluiu que existem variáveis com poder explicativo importante numa das amostras que desempenham um papel secundário na outra amostra (como, por exemplo, o caso da

---

<sup>1</sup> Abreviatura para “*Central Business District*”.

variável “área de convivência” que assume um papel dominante na amostra de condomínios). É, ainda, afirmado que a variável “idade do imóvel” constitui uma medida de similaridade para a amostra constituída por casas individuais, conduzindo a dependência espacial de preços, isto porque se considera que residências construídas durante um determinado período partilham características tecnológicas e arquitetónicas específicas.

Também Kalnins (2003) estuda a possível existência de dependência espacial, introduzindo, na sua investigação, um conceito particularmente interessante: o *franchising*, relação que permite às empresas maximizar os seus lucros. Porém, e tal como referido no estudo, é importante ter em conta que nem sempre os consumidores consideram os produtos (neste caso, hambúrgueres) das diferentes cadeias, como produtos substitutos. Assim, só se assume dependência espacial em postos de venda considerados competidores ativos uma vez que, só neste caso concreto, se pode falar nos conceitos de substituibilidade entre produtos e relação entre localizações de empresas dispersas no espaço. Para explicação dos seus dados, o autor recorreu a um modelo de regressão que permite incorporar a presença de dois tipos de correlação espacial: a relação de causalidade entre as variáveis dependentes de regiões vizinhas sobre o valor assumido pela região de referência e a correlação entre os erros do modelo.

Lee & Pace (2005), introduzem o conceito de dependência espacial, ao analisarem a importância da localização para determinar a dependência entre consumidores e vendedores. Assim, estudam possíveis ligações existentes entre o comércio a retalho e características dos consumidores, construindo modelos que permitem realçar a influência da dimensão das lojas e da distância existente entre postos de venda. Os autores formulam quatro modelos cujo intuito é considerar diferentes componentes de dependência espacial entre consumidores e vendedores. O objetivo central é explicar o modo como se processa o comércio de retalho entre duas regiões. Para tal e, através do modelo *spatial lag*, são introduzidas variáveis caracterizadoras do produto em estudo, da procura e da oferta, como por exemplo, facilidade na acessibilidade, distância a pontos de interesse, entre outros.

A ideia de correlação entre as características do mercado e o espaço físico é reforçada por Campos (2017), que reflete sobre a importância na ponderação de fatores que influenciam

o mercado e o mecanismo de preços. Neste tipo de estudos é importante analisar os vários condicionantes da procura e da oferta, tendo em conta questões como as características da população como, por exemplo, a faixa etária, o nível de rendimento, o género, a nacionalidade, entre outros. Isto porque, o nível de vendas pode estar sob influência de fatores como o crime/violência, acessibilidade a serviços, nível de interação social ou, até mesmo, a perspectiva de emprego (Lipton *et al.*, 1995).

Na sua análise Campos (2017) recorre a modelos que permitem comparação de multimercados, procedendo a uma segmentação tendo em conta a área de localização do respetivo posto de venda. No estudo de Campos, à semelhança do que já tinha sido feito por Kalnins (2003), Martens *et al.* (2005), Rogers & Winter (2009), entre outros, são aplicados modelos que incorporam a dependência espacial. Os preços hedónicos das casas são definidos em função do perfil do consumidor, sendo este perfil explicado pela distância entre a habitação e o centro de negócios do distrito a que esta pertence, a sensação de bem-estar do indivíduo e a vizinhança definida para esse mesmo bem. Através dos resultados obtidos foi possível perceber que o consumo do bem em questão (imóveis) é influenciado negativamente pelo seu preço quando o imóvel se localiza longe do centro de negócios (havendo concentração de imóveis mais caros na envolvente do CBD). Por outro lado, também se concluiu que a quantidade de imóveis lançados em distritos com maior insegurança é menor, sendo que nestas regiões o preço dos imóveis é tendencialmente mais baixo.

No caso de Rogers & Winter (2009), dada a dimensão da amostra e assimetria de ponderações espaciais, o estudo recorreu ao estimador do método dos momentos tendo em conta a existência de dependência espacial, procedendo-se à estimativa entre os anos de 2000 e 2007. Esta análise aponta para a importância da realização de estudos baseados em amostras observadas ao longo de um período temporal.

Outro tipo de formulação possível para o estudo da presença de dependência espacial, é o uso de modelos de otimização não-linear (verificado no estudo efetuado por Besner, 2002). Neste tipo de modelos é recomendada a logaritmação das variáveis analisadas com distribuições assimétricas.

No estudo desenvolvido, Lipton *et al.* (1995) recorreram a uma amostra de pontos de venda aleatoriamente selecionados de uma zona alvo, previamente, definida. Para

proceder à análise foi aplicado um modelo de estimação econométrico espacial linear, através do qual foi possível concluir que, do conjunto de variáveis integradas, as que influenciam positivamente as vendas de tabaco são: o rendimento familiar médio, a percentagem de pessoas nascidas no estrangeiro e a percentagem de população entre os 15 e os 24 anos.

### 3. METODOLOGIA

#### 3.1. *Abordagem Espacial*

O incremento e a generalização do uso da tecnologia, a crescente evolução da capacidade de armazenamento de dados, a sucessiva maior facilidade de computação, o baixo custo de sistemas de informação geográfica (SIG), conduzem a uma maior disponibilidade de dados onde é incorporada a localização das unidades observadas. Por outro lado, o reconhecimento de falhas nas aplicações econométricas tradicionais, fomentam o interesse e a necessidade de utilização de técnicas mais sofisticadas. Assim e, de modo a evitar incorreções nas inferências estatísticas provenientes do uso da Econometria Clássica, manifesta-se o interesse por modelos que questionam a independência espacial no conjunto de dados em análise, recorrendo-se a metodologias da designada Econometria Espacial. Esta tem em conta, na especificação e estimação dos modelos, a eventual dependência entre observações e regiões vizinhas.

Segundo LeSage (1999), a econometria espacial é caracterizada tendo por base dois fatores: dependência espacial e heterogeneidade espacial. Dependência espacial aponta para a existência de influência entre observações com localização diferente, ou seja,

$$y_i = f(y_j) \quad i, j = 1, \dots, n; i \neq j \quad (1)$$

em que,  $i$  representa a observação para a localização  $i$  e  $n$  o número de diferentes localizações consideradas.

Por outro lado, a heterogeneidade espacial ocorre quando se verifica uma alteração da relação de causalidade entre  $y_i$  e  $y_j$  tendo em conta a sua localização e, é dada por,

$$y_i = X_i \beta_i + \varepsilon_i \quad (2)$$

em que,  $X_i$  representa um vetor ( $1 \times k$ ) composto por variáveis explicativas,  $\beta_i$  representa um vetor de parâmetros ( $k \times 1$ ) associados a  $X_i$ ,  $y_i$  é a variável dependente observada em

cada localização  $i$  e  $\varepsilon_i$  representa o erro estocástico na relação linear. Abordagem à qual optámos por não recorrer ao longo do estudo em causa.

O estudo desenvolvido por Câmara *et al.* (2009) afirma existir necessidade de caracterizar os fenómenos ponderando os fatores de localização. Sandberg (2004) acrescenta que “*Location is everything*”, corroborando a ideia de um incremento de características comuns entre regiões próximas. Também Almeida (2004) defende esta conjetura, à qual se refere como 1ª Lei da Geografia de Tobler, através da qual é introduzido o fundamento da dependência espacial e que postula, em termos gerais, que quanto mais próximos no espaço os fenómenos estiverem mais relacionados estão.

O conceito de dependência espacial apresenta semelhanças quando comparado com o conceito de dependência temporal, uma vez que a sua presença tem influência sobre os estimadores de variância e conduz à inadequação da estimação por Mínimos Quadrados Ordinários (MQO). Porém é um fenómeno muito mais complexo, uma vez que a dependência temporal possui apenas um sentido, enquanto a dependência espacial é multidirecional. Por outro lado, a dependência espacial pode ser modelada de várias formas: através da introdução nas variáveis explicativas dos valores da variável dependente e/ou independente verificados em regiões vizinhas (dando origem a modelos que incorporam os designados “desfasamentos espaciais”), de mecanismo de dependência espacial no erro ou de uma combinação entre as situações anteriores.

### 3.2. Matriz $W$ de ponderação espacial

A matriz de ponderação espacial, doravante designada por  $W$ , permite captar a dependência espacial entre regiões (LeSage & Pace, 2009). É uma matriz exógena (dado ser definida *a priori*) e com estrutura não estocástica, comum a diversos modelos.

Esta matriz pode ser definida através de diferentes estruturas de ponderação, expressas por Tyszler (2006) como atributos geográficos ou socioeconómicos. Tal como afirmado por Almeida (2012), para que a matriz  $W$  não constitua uma fonte de má especificação do modelo econométrico, deverá ser apropriada às características da análise em execução. Porém, no processo de obtenção desta matriz, encontra-se muitas vezes implícito o processo de “tentativa erro”, tendo sido frisado por Anselin (1988) que poderá ser possível a obtenção de valores próximos dos considerados ideais permitindo, para a

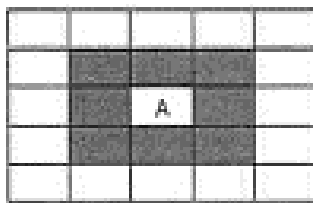


análise efetuada, a incorporação de grande parte da dependência espacial presente nos dados.

Para definir a matriz  $W$  podemos recorrer a várias regras, destacando-se os critérios baseados no conceito de contiguidade entre regiões e no conceito de distância, seja ela a distância Euclidiana ou económica entre as regiões.

O conceito de contiguidade define vizinhança da região de referência tendo em conta as regiões que estabelecem, com ela, fronteira física. No entanto, existem vários critérios para definir essa fronteira como, por exemplo, o critério Rainha ou Torre. Tais denominações são resultantes da semelhança existente entre o movimento da peça (designadas de rainha e torre) num jogo de xadrez e a fronteira estabelecida entre regiões consideradas contíguas. Segundo o critério de contiguidade rainha, são consideradas vizinhas as regiões que apresentam fronteiras físicas com extensão ou que apresentam vértices comuns. Note-se que, no caso da convenção denominada de Torre, as regiões são consideradas contíguas apenas se existirem fronteiras físicas com extensão entre elas.

**Figura 1:** Exemplo de distribuição espacial de regiões



A figura 1, presente em Almeida (2012), permite caracterizar o critério de contiguidade segundo o critério Rainha, em que a região sombreada define a vizinhança da região A. Por outro lado, a título de exemplo, considera-se para um conjunto de três regiões, a construção de uma matriz  $W$  (abaixo representada) do tipo binário. Quando  $w_{ij}$  assume valor unitário manifesta a existência de relação de vizinhança entre as regiões  $i$  e  $j$  (situação das regiões 1 e 2 e 1 e 3). Por outro lado, se  $w_{ij}$  apresenta valor nulo então as regiões  $i$  e  $j$  são não contíguas ou não vizinhas (situação das regiões 2 e 3).

$$W = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (3)$$

Outra forma de especificação da matriz de ponderação espacial é suportada pelo critério da distância Euclidiana inversa, baseada na distância medida a partir dos centroides de cada unidade de estudo (no caso em análise, freguesias). Este critério de definição de matrizes, descrito por Tyszler (2006), baseia-se no pressuposto de que quanto mais afastadas forem as regiões, menor será a dependência espacial que estabelecem entre si. Neste sentido são definidos os  $k$  vizinhos mais próximos e os valores assumidos pela matriz de pesos espaciais terão em consideração a distância geográfica entre essas referências (de entre as  $k$  anteriormente definidas).

É, ainda, importante ter em conta que, sempre que essa matriz assumir valor nulo, pode concluir-se que as observações em análise são não correlacionadas espacialmente. Por outro lado, quanto maior for o valor assumido por  $w_{ij}$ , maior será o potencial efeito colateral da observação  $j$  sobre a observação  $i$ . É relevante destacar que, por convenção, é assumido que a diagonal principal de  $W$  é composta por zeros, dado que não se considera que a região seja vizinha de si mesma (Almeida, 2012). Como tal,

$$w_{ij} = 0, \text{ para } i = j \quad (4)$$

Tendo em conta a descrição anterior, é perceptível que a matriz caracterizada tem um papel preponderante em modelos que pretendem assinalar relações entre regiões vizinhas, pelo que se torna crucial a escolha adequada de uma matriz de ponderação espacial (decisão esta abordada ao longo do procedimento experimental). Segundo Vieira (2009), a matriz formada pelo inverso da distância geográfica com definição de um número fixo de vizinhos, possui a importante capacidade de permitir que regiões isoladas (que não possuem fronteiras físicas com nenhuma outra região), possam estar associadas a regiões próximas.

### 3.3 Modelos econométricos espaciais

Para estimar os parâmetros de um modelo linear é frequente utilizar o método dos MQO. Porém, segundo LeSage (1999), caso se prove a presença de dependência espacial associada à variável dependente e/ou ao termo erro, verificar-se-á violação dos pressupostos básicos de Gauss-Markov<sup>2</sup> (devido à presença de endogeneidade nos regressores). Assim, poderemos estar perante enviesamento e inconsistência ou

---

<sup>2</sup> Para mais detalhes consultar LeSage (1999).

ineficiência do estimador e invalidade da inferência habitual, efeito este que será dependente do tipo de modelo especificado para incorporar a dependência espacial. Deste modo, na presença de dependência espacial nos dados, tal como afirmado por Tyszler (2006), é frequente recorrer-se a métodos de estimação alternativos como: estimação por Máxima Verosimilhança, Variáveis Instrumentais ou Método dos Momentos Generalizado.

Consequentemente será necessário perceber se o mecanismo de dependência espacial presente nos dados pode ser explicado com a introdução de defasamentos espaciais da variável dependente nos regressores ou com a consideração de um mecanismo espacial autorregressivo do erro do modelo. Para o efeito podem ser usados testes específicos como: teste de Wald, Razão de Verosimilhança ou testes fundamentados na técnica dos Multiplicadores de Lagrange e, em caso de necessidade, recorre-se às respetivas versões robustas (Tyszler, 2006; Vieira, 2009; Silva, Borges & Parré, 2013). Porém, Anselin (2003) realça que os testes de Multiplicador de Lagrange (LM) são os que propiciam uma escolha mais adequada do modelo que deve ser tido em consideração.

Como referido por Anselin (1988), LeSage (1999), LeSage & Pace (2009), Golgher (2012), entre outros, vários são os modelos paramétricos utilizados em econometria espacial, aos quais são atribuídas diferentes justificações teóricas para o respetivo uso. Para o presente estudo, limitar-nos-emos a caracterizar alguns dos modelos mais usados no tratamento de dados *cross-section*, visto serem deste tipo os dados de que dispomos.

### 3.3.1 Modelos SAR

Segundo Ywata & Albuquerque (2011), o Modelo Autorregressivo Espacial (SAR) é um dos mais utilizados para modelação de dependência espacial.

LeSage (1999) afirma que este é um modelo adequado para descrever o comportamento de determinada variável tendo em conta os valores assumidos por defasamentos espaciais da variável dependente (valores da variável verificados em regiões vizinhas), valores assumidos por variáveis independentes (que representam diferentes características da respetiva área) e os valores assumidos pelo termo erro.

O valor assumido pela variável dependente na localidade  $i$  (dados *cross-section*),  $y_i$ , é calculado em função dos valores espacialmente defasados dessa mesma variável, ou seja,

valores verificados em regiões vizinhas e por valores assumidos por outras características da própria região (variáveis independentes).

Este tipo de modelos, frequentemente denominados por *spatial autoregressive models* (LeSage & Pace, 2009), assumem uma forma simplificada:

$$y_i = \rho \sum_{j=1}^n W_{ij} y_j + \varepsilon_i \quad i = 1, \dots, n \quad (5)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

em que o termo,  $\sum_{j=1}^n W_{ij} y_j$  denominado por desfasamento espacial (*spatial lag*), reproduz a combinação linear de valores assumidos pela variável dependente nas outras regiões;  $\rho$  representa a intensidade da dependência espacial existente entre as observações em análise;  $\varepsilon$  vetor  $n \times 1$  é um erro estocástico não observado com distribuição normal (média e covariância zero e variância  $\sigma^2$ ).

Porém, o modelo anteriormente proposto (5) não considera a possível existência de efeitos por parte de outras variáveis no valor assumido pela variável dependente. Para colmatar este aspeto, LeSage (1999) propõe a seguinte especificação,

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \rho \sum_{j=1}^n W_{ij} y_j + \varepsilon_i \quad i = 1, \dots, n \quad (6)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

onde  $x_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, k$ ) são variáveis explicativas observadas na região  $i$  e  $\beta_j$  ( $j = 0, \dots, k$ ) são parâmetros desconhecidos.

### 3.3.2 Modelos SEM

O Modelo de Erro Espacial (SEM), caracterizado por Anselin & Bera (1998), LeSage (1999), Vieira (2009), entre outros, permite estudar a dependência espacial presente no termo de erro. Este modelo, na sua forma generalizada, pode ser definido por:

$$y = X\beta + u \quad \text{com} \quad u = \lambda Wu + \varepsilon \quad (7)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

com  $y$  vetor  $n \times 1$ ,  $X$  matriz  $n \times (k + 1)$ ,  $\beta$  vetor  $(k + 1) \times 1$ ,  $u$  vetor  $n \times 1$  de variáveis estocásticas não observadas.

Tal como pode ser verificado através da formulação anterior, o modelo de erro espacial apresenta uma estrutura autorregressiva ao nível dos erros, que permite quantificar o efeito médio dos erros dos vizinhos em relação ao erro da região em questão.

### 3.3.3 Modelos SAC

Também referido por Anselin & Bera (1998), LeSage (1999), Vieira (2009), entre outros, o Modelo Espacial Autorregressivo (SAC) permite combinar a dependência espacial no desfasamento e no termo de erro. Na forma generalizada, este tipo de modelos é dado por:

$$y = \rho W_1 y + X\beta + u \quad \text{com} \quad u = \lambda W_2 u + \varepsilon \quad (8)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

em que,  $W_1$  e  $W_2$  representam matrizes de ponderação espacial não necessariamente iguais (Vieira, 2009).

### 3.3.4 Modelos SDM

O Modelo Espacial de Durbin (SDM) é, segundo Vieira (2009), indicado para regressões que possuem variáveis dependentes e variáveis explicativas espacialmente autocorrelacionadas. Este modelo considera que não se verifica a presença de dependência espacial ao nível do termo erro.

Este tipo de modelos define-se, de forma genérica, através da seguinte expressão:

$$y = \rho W_1 y + X\beta + \theta W_3 X\beta + \varepsilon \quad (9)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

em que,  $W_1$  e  $W_3$  representam matrizes de ponderação espacial, associadas às variáveis dependente e explicativas, não necessariamente iguais. (Vieira, 2009).

## 3.4. Índices de Dependência Espacial

### 3.2.1. Índice de Moran (I)

O Índice de Moran permite testar a presença de dependência espacial global numa dada variável. Possui distribuição assintótica normal e, como caracterizado por Almeida (2012), fornece informação acerca da tendência de convergência ou divergência desses mesmos valores. A estatística de Moran é sensível à ausência de normalidade nos erros do modelo, bem como à escolha da matriz de ponderação espacial.

Como evidenciado por Almeida (2012), é dado por:

$$I = \frac{n}{S_0} \frac{z'Wz}{z'z} \quad (10)$$

em que  $n$  é o número de regiões presentes no modelo,  $z$  representa os valores da variável de interesse padronizada (vetor de observações),  $Wz$  representa os valores médios da variável de interesse padronizada na vizinhança, tendo em conta a matriz de ponderação espacial  $W$  (onde cada elemento referente às regiões  $i$  e  $j$  são inscritos como  $w_{ij}$ ).  $S_0$  representa a soma dos elementos da matriz  $W$ . Da equação anteriormente apresentada, é importante realçar que no termo do numerador ( $z'Wz$ ) é apresentada a expressão da autocovariância espacial (a menos de uma constante) e no denominador ( $z'z$ ) a variância dos dados (a menos de uma constante).

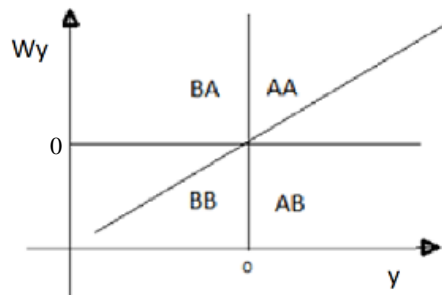
Este índice assume valores num intervalo entre  $-1$  e  $+1$ . Valores que se encontrem significativamente abaixo de  $-1/(N-1)$  manifestam a presença de dependência espacial negativa. Por outro lado, valores significativamente acima de  $-1/(N-1)$  manifestam a presença de dependência espacial positiva. É importante salientar que, para a realização de testes de hipóteses estatísticas, a estatística do Índice de Moran é standardizada de forma a assumir distribuição assintótica normal standardizada.

A execução do teste considera, sob hipótese nula, ausência de dependência espacial no erro dos dados em estudo (Almeida, 2004). Porém, e tal como enunciado por Cardoso (2002), duas das desvantagens da análise estatística em termos globais é não permitir identificar o tipo de dependência espacial presente nos dados e a possibilidade de omissão de padrões locais do comportamento da variável observada havendo, por isso, necessidade de recorrer a Índices Locais de Associação Espacial (LISA).

Associado a esta estatística surge o diagrama de dispersão de Moran (presente na figura 2) que, tal como referido por Tyszler (2006), Vieira (2009), Marconato *et al.* (2010), entre outros, permite visualizar a dependência espacial presente num determinado fenómeno e identificar diferentes padrões de associação entre regiões que estabelecem, entre si, relação de vizinhança. A variável de interesse é expressa através da abcissa (presente no eixo horizontal) do gráfico e o seu desfasamento espacial expresso no eixo vertical do respetivo gráfico. O coeficiente  $I$  de Moran é evidenciado através da inclinação da curva

de regressão estabelecida entre os valores do vetor da variável de interesse e o vetor do seu desfasamento espacial.

**Figura 2:** Diagrama de Dispersão de Moran



A interpretação do diagrama de dispersão é feita segundo a presença de observações de um de entre os quatro quadrantes: Alto-Alto (AA), Baixo-Baixo (BB), Alto-Baixo (AB) e Baixo-Alto (BA). O quadrante AA contém as unidades que apresentam elevado valor para a variável de interesse e para o respetivo desfasamento espacial, a aglomeração BB refere-se a unidades com baixo valor assumido pela variável em análise e pelo respetivo desfasamento espacial. O *cluster* BA representa as unidades com baixo valor para a variável de interesse e elevado valor para o seu desfasamento espacial e, caso contrário, para o *cluster* AB.

#### 3.4.2. Indicador Local de Dependência Espacial

Segundo Cardoso (2002), este tipo de indicador permite detetar e quantificar o grau de dependência espacial, evitando distorções de resultados evidenciados aquando da realização de uma análise global (como a análise global através do Índice de Moran). Silva *et al.* (2013) destacam a sua importante capacidade de facilitar a identificação de *clusters* (regiões que verifiquem comportamento muito semelhante no que diz respeito à variável em análise). Assim, através destes indicadores, é possível proceder a análises de peculiaridades (ou características próprias) confinadas a determinados locais do território em estudo.

No entanto e, tal como referido por Lopes *et al.* (2007), Marques *et al.* (2010), Marconato *et al.* (2010), entre outros, para aplicação dos indicadores locais de dependência espacial, dois pressupostos necessitam de ser satisfeitos: na envolvente de cada observação, devem existir padrões significativos de dependência espacial entre regiões e deve verificar-se proporcionalidade entre indicadores espaciais locais e globais.

#### 4. ANÁLISE EXPLORATÓRIA DOS DADOS

Através desta análise pretende-se perceber, por um lado, quais as características das freguesias do distrito de Lisboa que podem estar associadas a variações entre localizações no valor assumido pelas vendas a retalho e, por outro lado, a formação de *clusters*. Para tal são utilizados métodos caracterizados em capítulos anteriores.

##### 4.1 Caracterização dos Dados

A base de dados que compila toda a informação utilizada no presente estudo contém informação estatística selecionada pela PSE<sup>3</sup>. Esta informação permite caracterizar cada uma das presentes secções (municípios do distrito de Lisboa) e as suas respetivas subsecções (freguesias do mesmo distrito), onde se verifica a presença de pontos de vendas da loja de retalho de referência, no ano de 2018. A unidade de observação é a freguesia, sendo a amostra constituída por 123 observações (freguesias e uniões de freguesias).

Para tratar os dados da amostra optámos por recorrer a *softwares* estatísticos (como o Stata) e a *softwares* com capacidade de interface gráfico (como o QGIS e o GeoDa), programas que permitem ao utilizador a visualização de diferentes tipos de associação, bem como editar e efetuar análises exploratórias espaciais. Através destes *softwares* procedemos a cálculos de natureza estatística, estimação de modelos espaciais e análises exploratórias espaciais.

Os dados fornecidos, à semelhança do estudo de Almeida (2004), são de natureza *cross-section* pelo que, todos os valores assumidos por cada uma das variáveis, correspondem a um mesmo momento temporal.

##### 4.2 Caracterização das Variáveis

A base de dados de que dispomos é formada por um elevado número de variáveis caracterizadoras da oferta, da procura e das condições do mercado.

Todas as subdivisões territoriais são portadoras de um código, doravante denominado “DICOFRE” (designação presente na base de dados utilizada). Este, dada a diferenciação

---

<sup>3</sup> Produtos e Serviços de Estatística (PSE) é uma empresa de consultoria de natureza analítica com a qual estabelecemos parceria. Esta empresa incorpora a inteligência preditiva nas organizações e nos processos de negócio, aumentando a eficácia, certeza e valor na tomada de decisão dos seus clientes e parceiros.



feita pela presença de combinações de Algarismos, permite identificar univocamente cada uma das freguesias ou união de freguesias (por exemplo, “DICOFRE = 110301” diz respeito à freguesia de Alcoentre).

O processo de eleição das variáveis mais adequadas ao estudo passa por dois critérios: por um lado é tida em conta a matriz de correlação de Pearson<sup>4</sup> (tabela A2 no Apêndice), que auxilia na escolha das variáveis a incorporar no modelo (critério adotado por Vieira, 2009) e, por outro, procede-se a uma interpretação económica das relações existentes entre variáveis, tendo por base as características do problema em causa. Para conferir um carácter de maior qualidade ao modelo, é de extrema importância integrar variáveis que permitam caracterizar as condições de procura, oferta e da concorrência com as quais a empresa se depara.

**Tabela 1:** Caracterização das variáveis usadas no estudo

Variável	Descrição	Efeito Esperado
Vendas	Volume de vendas (milhões de euros) na freguesia	-
Perc_IND_RESID_65	Percentagem de indivíduos com mais de 65 anos residentes na freguesia	Negativo
Perc_IND_ENS_SUP	Percentagem de indivíduos com o ensino superior existentes na freguesia	Positivo
Perc_IND_DESEMP	Percentagem de indivíduos desempregados na freguesia	Negativo
TMD	Tráfego Médio Diário ( <i>proxy</i> da mobilidade)	Positivo
AEV/AEV_C	Rácio entre a área de vendas da loja de referência e a área de vendas da concorrência na freguesia	Positivo
RESID_HA	Residentes por hectare na freguesia	Positivo
NLojas/NLC	Rácio entre o número de lojas de referência e o número de lojas da concorrência na freguesia	Positivo
N_EDIF_EXC_RES	Número de edifícios exclusivos de residentes da freguesia	Positivo
Perc_Mulheres_RES	Percentagem de mulheres existentes na freguesia	Positivo
AEV_NLojT	Rácio entre a área de vendas da loja de referência e o número total de lojas na freguesia	Positivo
AEV_perC	Área de vendas da loja de referência <i>per capita</i> da freguesia	Positivo

A partir dos dados fornecidos, foram formuladas e integradas as variáveis presentes na tabela 1. Através deste conjunto de variáveis pretendemos caracterizar da melhor forma as condições de mercado para o negócio em análise presentes na região de Lisboa.

Para tal, e como verificado no estudo de Valduga & Zancan (2018), recorremos à inclusão de variáveis *proxy* (utilizadas para substituir outras de difícil quantificação e que estabelecem com elas relações pertinentes) e de uma classificação em categorias tendo em conta as variáveis: idade, grau de escolaridade, situação de emprego e género.

As variáveis Perc\_IND\_RESID\_65 e Perc\_Mulheres\_RES traduzem, respetivamente, a percentagem de indivíduos residentes na freguesia com mais de 65 anos de idade (variável

<sup>4</sup> Pela análise não se detetam problemas de multicolinearidade entre as variáveis.

caracterizadora das classes etárias presentes em cada freguesia<sup>5</sup>) e a percentagem de indivíduos do sexo feminino residentes na freguesia. Perc\_IND\_ENS\_SUP e Perc\_IND\_DESEMP são variáveis consideradas como *proxies* do nível de rendimento e definem, respetivamente, a percentagem de indivíduos residentes na freguesia que possuem ensino superior e a percentagem de indivíduos desempregados residentes na freguesia. A variável tráfego médio diário (TMD) tem propósito de captar o efeito da mobilidade para as várias freguesias. AEV/AEV\_C, NLojas/NLC e AEV\_NLojT são variáveis que captam a influência do mercado concorrente nas vendas da empresa em estudo, e podem ser definidas como: rácio entre a área de vendas da loja de referência e a área de vendas das lojas da concorrência, rácio entre o número de lojas de referência e o número de lojas da concorrência e rácio entre a área de vendas da loja de referência e o número total de lojas para determinada freguesia. Por outro lado, recorremos a variáveis como RESID\_HA, N\_EDIF\_EXC\_RES para caracterizar a população residente na freguesia em análise. Estas definem-se, respetivamente, como o número de residentes por hectare na freguesia e o número de edifícios exclusivos a residentes na freguesia. A variável AEV\_perC define a área de vendas da loja de referência *per capita* existentes na freguesia.

Através da estatística descritiva das variáveis presente tabela A1 no Apêndice, é possível a visualização dos valores médios e extremos assumidos quer pela variável em análise, quer pelas variáveis eleitas para modelação das vendas a retalho no ramo alimentar.

As freguesias do distrito de Lisboa apresentam uma média de vendas de, aproximadamente, 12,59 milhões de euros. Porém existem freguesias cujo volume médio de vendas é de, aproximadamente, zero milhões de euros. É, ainda, de salientar que a média de residentes por hectare é de, aproximadamente, 35 indivíduos. Em média 18% dos residentes têm mais de 65 anos, 20% têm o ensino superior, cerca de 5% estão desempregados e 52% são mulheres. Verifica-se ainda que, em média, o número de lojas da marca em estudo constitui 20% do número de lojas da concorrência e a sua área corresponde a 35% da área do conjunto das lojas da concorrência.

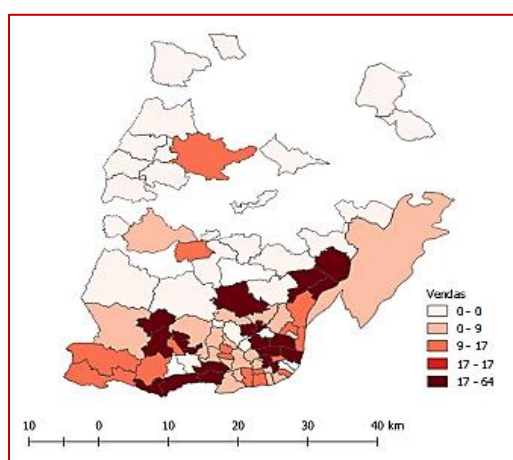
---

<sup>5</sup> Para estimar a regressão foram incluídos outros escalões de idades. No entanto, não foram considerados estatisticamente significativos.

### 4.3 Análise Exploratória de Dados Espaciais

À semelhança do que se verifica em Vieira (2009), através da AEDE do caso em estudo, é possível verificar a existência de valores que evidenciam a concentração geográfica do volume de vendas da empresa de retalho nas zonas sul e sudoeste do distrito de Lisboa. Tal verifica-se através da graduação de cores presente na imagem (em que cores mais claras correspondem a menores valores assumidos pelas vendas e cores mais escuras são atribuídas a maiores valores de vendas).

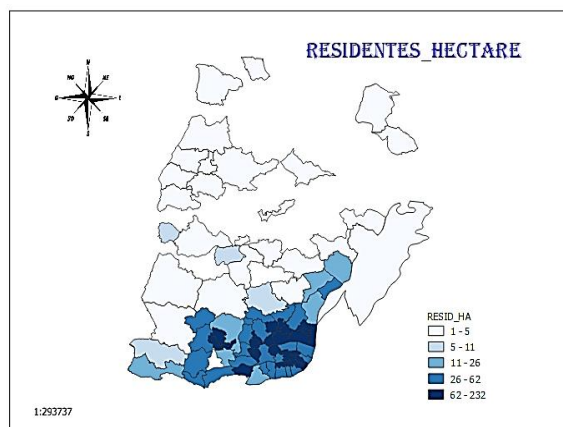
**Figura 3:** Dispersão geográfica da variável vendas (milhões de euros) – Ano 2018



Fonte: elaboração própria através do QGIS.

Tendo em conta todos estes fatores, não parece existir aleatoriedade na distribuição do valor da variável vendas, podendo suspeitar-se da existência de dependência espacial entre freguesias.

**Figura 4:** Dispersão geográfica da do número de residentes por hectare – Ano 2018



Fonte: elaboração própria através do QGIS.

Por outro lado (a título de exemplo) ao analisar o diagrama que representa uma graduação de cores, tendo por base o número de indivíduos residentes em cada uma das freguesias, parece fazer sentido aferir que as variáveis estão positivamente associadas. Regiões onde o número de indivíduos por hectare é mais elevado (mancha azul mais escura), coincidem com regiões onde se verificam maiores volumes de vendas.

Contudo, para comprovar as anteriores suposições, surge a necessidade de calcular a dependência espacial presente nos dados (Silva *et al.*, 2013).

#### 4.4 Dependência Espacial Global

Através do Diagrama de Moran, tendo por base as conclusões dos estudos de Vieira (2009) e Marques *et al.* (2010), entre outros, é possível determinar se as observações evidenciam dependência espacial. Porém, como visto anteriormente, antes de calcular o Índice de Moran propriamente dito, existe a necessidade de eleger a matriz de ponderação espacial que melhor se ajusta aos dados em questão.

Para o presente estudo recorreremos à matriz de ponderação espacial definida segundo o critério de contiguidade Rainha. Assim sendo, todos os testes e estimações de modelos econométricos verificados em diante terão em conta esta matriz  $W$ .

Os testes realizados com base nesta especificação para a matriz  $W$ , indicam a existência de dependência espacial positiva. Tal pode ser concluído a partir dos valores positivos e estatisticamente significativos que constam na tabela 2 (presente em baixo), que contém os resultados obtidos para o teste de Moran, cujo objetivo é testar a presença de dependência espacial para o modelo estimado. O valor obtido para a estatística de Moran, estandardizada conduz à rejeição da hipótese nula, para um nível de significância de 5%, da inexistência de dependência espacial. Tal resultado leva à necessidade de controlar a inclusão de dependência espacial nos modelos econométricos. Este tipo de resultados evidencia a existência de similaridade, em termos de localização, nos valores da variável vendas (Silva *et al.*, 2013).

**Tabela 2:** Resultado obtido para o teste de Moran.

Estatística de Moran	Valor-p
5,07	0,0243

A mesma conclusão pode ser retirada da figura 1 presente no Apêndice, que exhibe o diagrama de dispersão de Moran, cujo valor obtido para estatística I de Moran<sup>6</sup> é de, aproximadamente, 0,178. Assim e, através da inclinação positiva da reta de regressão, é possível verificar a existência de dependência espacial positiva nos dados.

Nos dados em análise verifica-se a presença de freguesias em cada um dos quatro quadrantes (AA, BB, AB, BA). Através da sua observação é possível verificar a existência de um aglomerado de freguesias que assumem elevado valor das vendas e cujo seu desfasamento espacial também apresenta valor elevado. Por outro lado, também é evidente a concentração de freguesias onde se verifica valor baixo quer para o volume de vendas quer para o respetivo desfasamento espacial.

#### *4.5 Dependência Espacial Local*

Os LISA permitem identificar agrupamentos similares ou atípicos, designados por *clusters* cuja relação no espaço é significativa.

A figura 2 no Apêndice facilita a visualização da localização dos agrupamentos estatisticamente significativos (Marconato *et al.*, 2010). Através destes indicadores é possível identificar dois principais *clusters* para os dados analisados. O primeiro envolve quatro freguesias destacadas a vermelho (localizadas a Sul do distrito de Lisboa), que apresentam elevado volume de vendas (AA). O segundo *cluster* é referente a sete freguesias destacadas a azul escuro, que possuem baixo volume de vendas quer na própria região quer em regiões vizinhas (BB). Freguesias cujo padrão apresentado é baixo-alto (BA) são destacadas a azul mais claro, e as que apresentam padrão AB evidenciam cor rosa.

#### *4.6 Dependência Espacial Bivariada*

Na análise multivariada, a variável vendas é relacionada com as diversas variáveis explicativas consideradas relevantes: AEV/AEV\_C, RESID\_HA, NLojas/NLC, N\_EDIF\_EXC\_RES, AEV\_perC.

Na figura 3 (no Apêndice) encontra-se a representação do diagrama de dispersão de Moran onde é estabelecida a relação de dependência entre as vendas (cujos valores são

---

<sup>6</sup> Não estandardizada.

representados no eixo das abcissas) e os desfasamentos espaciais de cada uma das variáveis explicativas (valores presentes no eixo das ordenadas).

No que concerne ao Índice de Moran bivariado (por observação da figura 3) podemos salientar que, à exceção da variável AEV/AEV\_C, dada a inclinação positiva das retas de regressão, existe evidência da presença de dependência espacial positiva entre as vendas e variáveis independentes em análise.

Os mapas de *clusters* bivariados, presentes na figura 4, permitem visualizar os agrupamentos estatisticamente significativos entre o volume de vendas e os valores assumidos pelas diferentes variáveis explicativas (análise espacial local). Nestes mapas as associações espaciais alto-alto (AA) são destacadas a vermelho e os agrupamentos baixo-baixo (BB) a azul mais escuro.

Nesta análise verifica-se que as associações espaciais alto-alto (AA) se encontram, predominantemente, localizadas na região Sudoeste do distrito de Lisboa.

## 5. MODELAÇÃO E RESULTADOS

### 5.1 Modelo Económico Espacial

Para iniciar o estudo foi estimado um modelo das vendas com as variáveis explicativas anteriormente definidas. O resultado estimado do modelo por MQO é apresentado no *output 1* no Apêndice. Através do valor evidenciado pelo  $R^2$  podemos afirmar que, aproximadamente, 77% da variação que ocorre no volume de vendas é explicado pelas variáveis explicativas presentes no modelo.

Porém, como referido por Almeida (2004), Arbia (2014), entre outros, a presença de dependência espacial torna inválidas as estimações obtidas através do método de MQO (Anselin, 1988; Bera & Anselin, 1998; LeSage, 1999; Elhorst, 2000; Vieira, 2009).

De modo a especificar o modelo, para que seja possível captar e quantificar este tipo de influência, recorreremos à econometria espacial (Silva *et al.*, 2013). Dada a inadequação do MQO e, como a distribuição do erro associado aos modelos que descrevem os dados em estudo é desconhecida, procede-se à estimação dos parâmetros recorrendo ao método de *pseudo* máxima verosimilhança.

Os resultados das estimações encontram-se resumidos nas tabelas 3 e 4 e podem ainda ser visíveis através dos *outputs 2 a 13* presentes no Apêndice.

Os resultados das estimações presentes na tabela 3 são obtidos através da estimação pelo método da máxima verosimilhança e dizem respeito, respetivamente, às versões não restritas e restritas dos modelos SAR, SEM e SAC. O mesmo raciocínio pode ser tido para a tabela 4, onde procedemos à estimação recorrendo ao método da *pseudo* máxima verosimilhança.

As estimações dos modelos iniciais (não restritos) SAR, SEM e SAC (modelos (1), (3) e (5) presentes na tabela 1), tiveram em conta o conjunto de variáveis explicativas presentes na tabela 1. Porém, dada a inexistência de significância estatística inerente a um conjunto de variáveis para os três modelos econométricos, foram retiradas as variáveis estatisticamente irrelevantes para explicar possíveis variações do volume de vendas. Tal conclusão deriva da análise do teste de significância conjunta dos regressores omitidos, cujo valor-p assume sempre valor superior a 5%.

Após estimar as novas regressões espaciais, às quais foram excluídas as variáveis estatisticamente irrelevantes, concluímos que, para um nível de confiança de 5%, todas as variáveis presentes (AEV/AEV\_C, RESID\_HA, NLojas/NLC, N\_EDIF\_EXC\_RES, AEV\_perC) possuem significância estatística (modelos (2), (4) e (6) da tabela 3).

Considerando a estimação do modelo SAR restrito, verifica-se que a estimativa do coeficiente que capta a dependência espacial, com valor-p de 0,054 (de acordo com o *output* 3 em Apêndice), é estatisticamente significativa a 10%, evidenciando que um aumento das vendas nas freguesias vizinhas vai provocar um aumento das vendas na freguesia em análise.

De forma semelhante, também o modelo SEM restrito, demonstra a presença de dependência espacial ao nível do erro, dada a significância estatística a 1% da estimativa do coeficiente que lhe está associado.

Relativamente aos resultados obtidos para a estimação dos modelos SAR e SEM robustos (modelos presentes na tabela 4), é possível concluir que, também nestes casos, há evidência da presença de dependência espacial. Estas conclusões induzem à necessidade de uso de um modelo de estimação que considere a presença de dependência espacial ao nível da variável dependente e do erro do modelo (SAC) (modelos (11) e (12) da tabela 4).

**Tabela 3:** Modelos SAR, SEM e SAC com estimação por máxima verosimilhança para o volume de vendas

	(1)	(2)	(3)	(4)	(5)	(6)
Perc_IND_RESID_65	-0,121 (0,140)	-	-0,028 (0,161)	-	-0,063 (0,181)	-
Perc_IND_ENS_SUP	-0,058 (0,100)	-	-0,058 (0,102)	-	-0,061 (0,102)	-
Perc_IND_DESEMP	-0,425 (0,820)	-	-0,341 (0,842)	-	-0,370 (0,840)	-
TMD	-0,003 (0,015)	-	-0,006 (0,015)	-	0,002 (0,019)	-
AEV/AEV_C	-1,410** (0,560)	-1,411*** (0,517)	-1,430*** (0,541)	-1,382*** (0,501)	-1,438*** (0,549)	-1,382*** (0,500)
RESID_HA	0,067*** (0,016)	0,052*** (0,013)	0,072*** (0,016)	0,064*** (0,014)	0,071*** (0,017)	0,065*** (0,015)
NLojas/NLC	9,427** (3,725)	8,644** (3,406)	8,847** (3,539)	9,112*** (3,340)	9,014** (3,643)	9,112*** (3,336)
N_EDIF_EXC_RES	0,00097*** (0,0002)	0,0009*** (0,0001)	0,0009*** (0,0002)	0,0008*** (0,0001)	0,0009*** (0,0002)	0,0008*** (0,0002)
Perc_Mulheres_RES	0,086 (0,391)	-	0,074 (0,389)	-	0,082 (0,391)	-
Area/NLojT	0,0006 (0,001)	-	0,0005 (0,001)	-	0,0006 (0,001)	-
AEV_perC	191,8965*** (15,268)	189,905*** (14,026)	194,193*** (14,506)	190,916*** (13,545)	194,479*** (14,840)	190,879*** (13,550)
Constante	-5,1867 (20,235)	-4,443 (1,251)	-5,787 (19,777)	-3,831 (1,304)	-5,703 (19,981)	-3,793 (1,440)
W*Vendas	0,187** (0,088)	0,136* (0,070)	-	-	0,076 (0,223)	-0,008 (0,134)
$\lambda$			0,431** (0,178)	0,424*** (0,158)	0,296 (0,447)	0,437* (0,256)
Estatística Wald	443,79	427,90	423,67	418,26	425,26	418,52
Valor-p	0,000	0,000	0,000	0,000	0,000	0,000
Dimensão da Amostra	123	123	123	123	123	123
Sig. Conj. Termos Espaciais - $\chi^2$	4,49	3,73	5,85	7,23	5,12	7,37
Valor-p	0,034	0,054	0,016	0,007	0,077	0,025
Sig. Conj. Var. Omitidas		3,39		1,11		1,24
Valor-p		0,758		0,981		0,975

Notas:

Erros padrão encontram-se entre parênteses; (1) e (2) modelos SAR, (3) e (4) modelos SEM, (5) e (6) modelos SAC; Níveis de Significância: \*10%, \*\*5%, \*\*\*1%.



**Tabela 4:** Modelos SAR, SEM e SAC com estimação por *pseudo* máxima verosimilhança para o volume de vendas

Variável	(7)	(8)	(9)	(10)	(11)	(12)
Perc_IND_RESID_65	-0,121 (0,138)	-	-0,028 (0,146)	-	-0,063 (0,143)	-
Perc_IND_ENS_SUP	-0,058 (0,100)	-	-0,059 (0,101)	-	-0,061 (0,102)	-
Perc_IND_DESEMP	-0,425 (0,818)	-	-0,341 (0,835)	-	-0,370 (0,839)	-
TMD	-0,003 (0,015)	-	-0,006 (0,015)	-	0,002 (0,016)	-
AEV/AEV_C	-1,410** (0,559)	-1,411*** (0,517)	-1,430*** (0,539)	-1,382*** (0,500)	-1,438*** (0,548)	-1,382*** (0,500)
RESID_HA	0,067*** (0,016)	0,052*** (0,013)	0,072*** (0,016)	0,064*** (0,013)	0,071*** (0,016)	0,065*** (0,014)
NLojas/NLC	9,427** (3,717)	8,644** (3,404)	8,847** (3,488)	9,112*** (3,331)	9,014** (3,629)	9,112*** (3,332)
N_EDIF_EXC_RES	0,001*** (0,0002)	0,0009*** (0,0001)	0,0009*** (0,0002)	0,0008*** (0,0001)	0,0009*** (0,0002)	0,0008*** (0,0002)
Perc_Mulheres_RES	0,086 (0,391)	-	0,074 (0,388)	-	0,082 (0,390)	-
Area/NLojT	0,0006 (0,001)	-	0,0005 (0,001)	-	0,0006 (0,001)	-
AEV_perC	191,897*** (15,281)	189,905*** (14,043)	194,193*** (14,495)	190,916*** (13,525)	194,479*** (14,783)	190,879*** (13,510)
Constante	-5,1867 (20,233)	-4,443 (1,256)	-5,787 (19,771)	-3,831 (1,292)	-5,703 (19,981)	-3,793 (1,412)
W*Vendas	0,187** (0,094)	0,136* (0,076)	-	-	0,076 (0,119)	-0,008 (0,095)
$\lambda$			0,431** (0,175)	0,424** (0,175)	0,296 (0,217)	0,437** (0,186)
Estatística Wald	445,10	429,87	423,80	418,83	425,48	418,99
Valor-p	0,000	0,000	0,000	0,000	0,000	0,000
Dimensão da Amostra	123	123	123	123	123	123
Sig. Conj. Termos Espaciais - $\chi^2$	3,98	3,21	6,09	5,84	4,02	6,16
Valor-p	0,046	0,073	0,014	0,016	0,134	0,046
Sig. Conj. Var. Omitidas		3,30		1,11		1,43
Valor-p		0,770		0,981		0,964

Notas:

Erros padrão encontram-se entre parênteses; (7) e (8) modelos SAR, (9) e (10) modelos SEM, (11) e (12) modelos SAC;

Níveis de Significância: \*10%, \*\*5%, \*\*\*1%.

À semelhança da estimação por máxima verosimilhança, também na estimação por *pseudo* máxima verosimilhança se verificou a inexistência de significância estatística para um conjunto de regressores. Assim, também nesta fase e, após recorrermos ao teste de significância conjunta, foram estimados os modelos restritos (modelos (8), (10) e (12)). Após analisar os modelos SAR e SEM restritos, verifica-se a presença de dependência espacial ao nível da variável de interesse e do erro (respetivamente) e, como tal, decidimos recorrer à estimação do modelo SAC. Através dos resultados obtidos para este modelo é possível afirmar que o *spatial lag* não apresenta significância estatística para um nível de confiança de 5%, pelo que apenas o *error lag* deve ser considerado para o caso (dada a sua significância estatística).

Assim e, tal como afirmado por Vieira (2009), existe evidência de dependência espacial no erro ou nas suas variáveis omitidas, originando falhas na sua especificação. Conforme enunciado em Almeida *et al.* (2008), esta dependência poderá ser resultante de efeitos não-modelados não aleatoriamente distribuídos através do espaço. Vieira (2009) acrescenta que, este tipo de resultados, se podem dever à omissão de variáveis importantes na caracterização dos dados ou transformações de variáveis incluídas que não foram devidamente consideradas.

Todas as estimações conduzem a resultados semelhantes relativamente à significância estatística das variáveis. Deste modo pode concluir-se que o volume de vendas de determinada freguesia é influenciado positivamente pelo número de residentes por hectare da respetiva freguesia, pelo rácio entre o número de postos de venda da marca em estudo e do número de postos de venda da concorrência da freguesia, pelo número de edifícios exclusivos a residentes na freguesia e pela área de vendas da loja de referência *per capita* da respetiva freguesia. Verificamos que o rácio entre a área de vendas da loja de referência e a área de vendas da concorrência na freguesia influencia negativamente o volume de vendas da freguesia. É, ainda, importante salientar que a percentagem de indivíduos residentes com idade superior 65 anos, a percentagem de indivíduos com ensino superior, a percentagem de indivíduos desempregados, o tráfego médio diário, a percentagem de mulheres residentes e o rácio entre a área de vendas da loja de referência e o número total de lojas de cada freguesia são variáveis não relevantes para explicar as vendas da respetiva freguesia.

Um ponto bastante interessante e que apoia os resultados obtidos ao longo do presente TFM, é o facto de se encontrarem alinhados com as conclusões obtidas pela PSE em estudos similares.

## 6. CONCLUSÕES E POSSÍVEIS DESENVOLVIMENTOS FUTUROS

Vários são os fatores que podem influenciar a procura dos consumidores. Através do presente trabalho e, tendo em conta os dados de que dispomos e a sua natureza espacial, pretendemos analisar quais os fatores que influenciam o volume de vendas a retalho no setor alimentar de uma dada marca.

Definimos variáveis consideradas relevantes, procedemos a análises empíricas e exploratórias dos dados, modelações econométricas e interpretações das estimações e *outputs* obtidos. Para tal, recorremos à matriz W de ponderação espacial (definida através do critério de contiguidade Rainha), através da qual é possível captar a dependência espacial entre regiões do distrito de Lisboa.

Os valores obtidos para o teste de Moran evidenciam a presença de dependência espacial nos dados. Tal conclusão é corroborada pela análise dos gráficos que ilustram a dispersão geográfica das variáveis vendas e número de residentes por hectare na freguesia, que apontam para a existência de maior volume de vendas nas regiões com maior número de habitantes por metro quadrado, fator que fomenta a suspeita de inexistência de independência espacial. Esta conclusão é reforçada aquando da estimação dos diferentes modelos econométricos. Ambos os coeficientes estimados associados ao termo espacial, presentes nos modelos SAR e SEM, apontam para a presença de dependência espacial nos dados.

É ainda possível concluir que o número de residentes por hectare na freguesia, o rácio entre o número de postos de venda da marca em estudo e o número de postos de venda da concorrência na freguesia, o número de edifícios exclusivos a residentes na freguesia e a área de vendas da loja de referência *per capita* da freguesia, possuem um impacto positivo sobre a variável vendas da respetiva freguesia enquanto que o rácio entre a área de vendas da loja de referência e a área de vendas da concorrência na freguesia possui um impacto negativo. Deste modo, parece vantajoso para a empresa proceder à instalação de postos de venda de maior dimensão em localidades com elevado número de residentes e em regiões próximas do Centro de Negócios da cidade. Através do estudo parece-nos

adequado afirmar que existem vantagens no aumento do número de postos de vendas nas regiões a Sul e Sudoeste do distrito de Lisboa. É, ainda, importante salientar que, para uma dada área de vendas de uma freguesia, quanto maior for o mercado concorrencial, maior será o volume de vendas da loja em análise.

Para proceder ao estudo da presença de dependência espacial podemos ainda recorrer modelo de Durbin (sugerido pela análise da estatística de Moran bivariada). Este modelo é baseado na estimação do modelo com desfasamento espacial nas variáveis independentes. Porém este método não é considerado para o presente estudo por não se encontrar no âmbito deste trabalho dada a reduzida dimensão da amostra e a complexidade que é inerente a esta modelação.

Para a realização deste trabalho foram assumidos alguns pressupostos que devem ser tidos em conta e contornados em análises futuras. O estudo em causa partiu de uma amostra de dados e variáveis caracterizadores do distrito de Lisboa. Porém, para que ocorra a agregação dos dados para determinada área/freguesia, é apresentada a hipótese de presença de regiões homogéneas. Tal conjectura é vista como um inconveniente subjacente ao modelo dado que, segundo Golgher (2012), tendencialmente, as regiões não apresentam características de total homogeneidade.

Outro aspeto que deve ser tido em consideração é o facto de, para definição da matriz de pesos espaciais, se considerar que influências entre regiões vizinhas são similares independentemente da direção de referência. Porém, seria mais realista a construção de uma matriz que permita tomar em consideração a posição da região em análise face às regiões consideradas suas vizinhas. Por outras palavras e, como denominado por Dantas (2003) proceder à construção de uma matriz que capte os “efeitos de anisotropia”.

A estes fatores acresce o facto de, apesar da sua grande utilização, o índice de Moran apresentar alguns problemas no tratamento de uma amostra de tamanho reduzido. Assim, análises futuras poderão recorrer a métodos e modelos alternativos.

A definição do modelo mais adequado aos dados poderá ser obtida através do Método de Simulação de Monte Carlo, que permite testar as propriedades de diferentes matrizes em diferentes modelos.

Outro possível futuro desenvolvimento seria a elaboração de uma análise a uma escala geográfica mais detalhada, através da qual seria possível desenvolver outro tipo de

análises, como por exemplo, outros efeitos da concorrência, o impacto da mobilidade e dos movimentos pendulares, infraestruturas e acessibilidade ou, até mesmo, de pontos de interesse (efeitos estes não são visíveis com a presente escala geográfica).

#### REFERÊNCIAS

- Almeida, E. (2004). *Curso de Econometria Espacial Aplicada*, 1ªEd. São Paulo: Departamento de Economia, Administração e Sociologia, da ESALQ-USP.
- Almeida, E. (2012). *Econometria Espacial Aplicada*, 1ª Ed. Guanabara: Editora Alínea.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, 1ªEd. Berlin: Springer Netherlands.
- Anselin, L. (2001). *Spatial Econometrics*, 1ª Ed. Nova Iorque: Badi H. Baltagi.
- Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27, pp. 247-267.
- Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International regional science review* 26, 2: 153–166.
- Anselin, L. & Bera, A. (1998) Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics, *Handbook of Applied Economic Statistics*, pp. 237–289.
- Aranha, F. (1999). Autocorrelação espacial na área de loja de supermercados nos municípios paulistas: mensuração por meio do Índice de Geary, *Revista de Administração de Empresas*, pp. 38 – 45.
- Arbia, G., (2014). *A primer for spatial econometrics: with applications in R*, 1ªEd. UK: Palgrave Macmillan UK.
- Besner, C. (2002) A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme. *Journal of Real Estate Research*. Vol.24, pp. 193-211.

- Câmara, G., Monteiro, A., Fucks, S. & Carvalho, M. (2009). *Análise Espacial e Geoprocessamento*. Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, SP, Brazil.
- Campos, R. (2017) O mercado imobiliário residencial no município de São Paulo: uma abordagem de preços hedônicos espacial. *Nova Economia* v.27 n.1 pp.303-337.
- Cardoso, C. (2002). *Dependência Espacial*. Paraná: Universidade Estadual de Maringá.
- Dantas, R. (2003). *Modelos espaciais aplicados ao mercado habitacional: um estudo de caso para a cidade do Recife*, 1ªEd. Pernambuco: Universidade Federal de Pernambuco.
- Elhorst, J. (2001). *Spatial Panel Data Models*, 2ªEd. Ohio: Geographical Analysis. pp. 119–140.
- Fingleton, B. & McCombie, J. (1998). Increasing returns and economic growth: some evidence for manufacturing from the European Union regions, *Oxford Economic Papers* 50 (1998), pp. 89-105.
- Fotheringham, A., Brunsdon, C. & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *International Journal of Geographical Geographically Information Science* 00(00), pp. 1–22.
- Fujita, M., Krugman, P. & Venables, A. (1999). *The Spatial Economy Cities, Regions, and International Trade*, 1ªEd. London: Massachusetts Institute of Technology.
- Golgher, A. (2012) Introduction to Spatial Econometrics, *Associação Brasileira de Estudos Populacionais*, pp. 1 – 36.
- Hoeckel, P., Casagrande, D., Santos, C. *Análise Espacial da Pobreza no Rio Grande do Sul*. Pernambuco.
- Kalnins, A. (2003) Hamburger Prices and Spatial Econometrics. *The Scholarly Commons*. University of Southern California, California.

- Kelejian, H. & Prucha, I. (1997). A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive. *Journal of Real Estate Finance and Economics*, Vol. 17:1, pp. 99-121.
- Krugman, P. (1979). Increasing Returns, Monopolistic Competition and International Trade, *Journal of International Economics* 9 (1979), pp. 469-479.
- Lee, M., Pace, R. (2005) Spatial Distribution of Retail Sales. *The Journal of Real Estate Finance and Economics*, 31:1, pp. 53-69.
- LeSage, J. (1999). *The Theory and Practice of Spatial Econometrics*, 1ªEd. Toledo: Universidade de Toledo.
- LeSage & Pace (2009). *Introduction to Spatial Econometrics*, 1ªEd. U.S.A.: Taylor and Francis Group.
- Lipton, R., Banerjee, A., Levy, D., Manzanilla, N. & Cochrane, M. (1995). The Spatial Distribution of Underage Tobacco Sales in Los Angeles. *Substance Use & Misuse*, 43: pp. 1597-1617.
- Lopes, S., Brondino, N. & Silva, A. (2007). *Análise do Desempenho de Modelos de Regressão Espacial na Previsão de Demanda por Transportes*. Brasil: UNESP.
- Marconato, M., Moro, O. & Parré, J. (2010). *Análise Espacial sobre Saúde nos Municípios Brasileiros em 2010*, 1ª Ed. Paraná: Universidade Estadual de Maringá.
- Marques, A., Holzschuh, M., Tachibana, V. & Imai, N. (2010). Análise exploratória de dados de área para índices de furto, Recife - PE, 27-30, pp. 1 – 8.
- Martens, B., Florax, R., Dooley, F. (2005). *The Effect of Entry by Supercenter and Warehouse Club Retailers on Grocery Sales and Small Supermarkets: A Spatial Analysis*. Purdue University, EUA.

- Martinho, V. (2005). *Aplicação da econometria espacial, com métodos “cross- section” e em painel, na análise dos efeitos “spillovers” na produtividade dos sectores económicos das regiões portuguesas*, Portugal: Millenium, pp. 190 – 221.
- Rogers, W., Winter, W. (2009). The Impact of Foreclosures on Neighboring Housing Sales. *Journal of Real Estate Research*. Vol.31, No.4.
- Sandberg, K. (2004). *Hedonic Prices, Economic Growth, and Spatial Dependence*, 1ªEd. Suécia.
- Santos, J. (2014). *Estratégias de Renovação do Retalho Alimentar Tradicional*, 1ªEd. Lisboa: Instituto Superior Técnico.
- Silva, L., Borges, M. & Parré, J. (2013), *Distribuição Espacial da Pobreza no Paraná*. Paraná: Universidade Estadual de Maringá.
- Tyszler, M. (2006). *Econometria espacial: discutindo medidas para a matriz de ponderação espacial*, 1ªEd. São Paulo: Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas pp. 1 – 155.
- Valduga, L., Zancan, E. (2018). *Utilização de modelo de regressão espacial para avaliação de terrenos na cidade de criciúma, sc*. UNESC: Brasil.
- Vieira, R. (2009). *Crescimento econômico no estado de São Paulo: uma análise espacial*. São Paulo: Cultura Acadêmica Editora.
- Ywata, A. & Albuquerque, P. (2011). Métodos e modelos em econometria espacial: uma revisão, *Revista Brasileira de Biometria*, 29 (2), pp. 273 – 306.



## APÊNDICE

## Tabelas

*TABELA A1 – Estatística Descritiva das variáveis incluídas no modelo final*

Variável	Unidades	Média	Desvio-Padrão	Mínimo	Máximo
Vendas	Milhões de euros	12,59142	10,95368	0,000000	64,30463
Perc_IND_RESID_65	Porcentagem	18,19942	4,618062	8,506583	29,52387
Perc_IND_ENS_SUP	Porcentagem	19,97323	9,54863	5,597448	50,14212
Perc_IND_DESEMP	Porcentagem	4,821081	0,879722	2,767528	7,226054
TMD	Número médio de veículos (unidades)	74,39161	53,44548	0,132000	313,5530
AEV/AEV_C	Rácio	0,354398	1,123921	0,000000	11,33700
RESID_HA	Número (unidades)	34,82853	41,51622	0,733000	231,9190
NLojas/NLC	Rácio	0,193504	0,170362	0,000000	1,000000
N_EDIF_EXC_RES	Número (unidades)	5515,000	4139,506	875,0000	13717,00
Perc_Mulheres_RES	Porcentagem	52,30439	1,958725	37,00696	56,34375
AEV_NLojT	Rácio	505,9830	768,3287	6,265000	4705,000
AEV_perC	Metros quadrados	0,043346	0,045419	0,000000	0,262427

Fonte: Elaboração Própria.

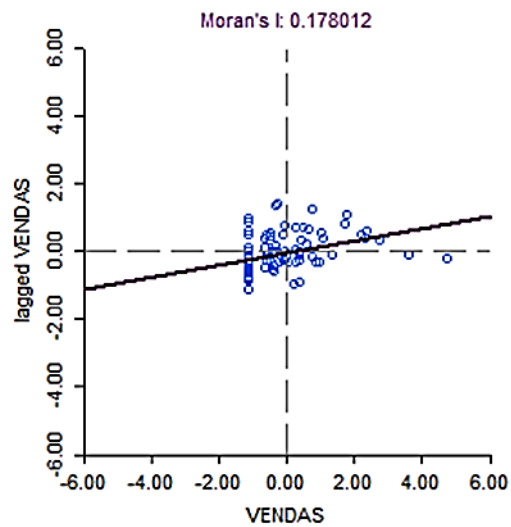
*TABELA A2 – Matriz de Correlação*

	VENDAS	Perc_~65	Perc_~UP	Perc_~MP	TMD	AEV_AE~C	RESID_HA	NLojas~C	N_EDIF~S	Perc_M~S	Area_N~T	AEV_perC
VENDAS	1.0000											
Perc_IND_~65	-0.1029	1.0000										
Perc_IND_E~P	0.2974	0.4664	1.0000									
Perc_IND_D~P	0.1179	-0.2755	-0.3752	1.0000								
TMD	0.3406	0.3619	0.6521	-0.0572	1.0000							
AEV_AEV_C	0.3380	0.1164	-0.0245	0.2033	0.1514	1.0000						
RESID_HA	0.2246	0.3079	0.4062	0.1999	0.3908	0.1547	1.0000					
NLojas_NLC	0.5531	-0.0965	0.1175	0.1552	0.3605	0.2778	-0.0188	1.0000				
N_EDIF_EXC~S	0.2009	-0.2780	-0.0520	0.1013	-0.0214	-0.1745	-0.4278	0.3199	1.0000			
Perc_Mulhe~S	0.2210	0.4631	0.6238	0.0087	0.4635	0.1118	0.4318	0.0900	-0.1089	1.0000		
Area_NLojT	-0.3767	-0.0706	-0.4451	-0.3680	-0.4482	-0.1371	-0.4317	-0.2620	-0.0971	-0.6465	1.0000	
AEV_perC	0.7923	-0.0120	0.2722	0.0598	0.2811	0.5798	0.1955	0.4606	-0.0949	0.2290	-0.3179	1.0000

Fonte: Elaboração Própria (com recurso ao STATA).

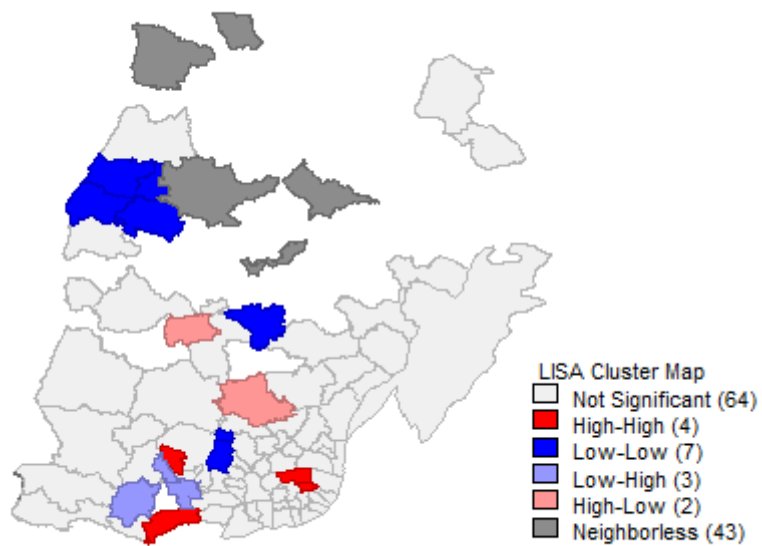
## Figuras

*FIGURA 1 – Diagrama de Dispersão de Moran para a variável vendas – Ano 2018*



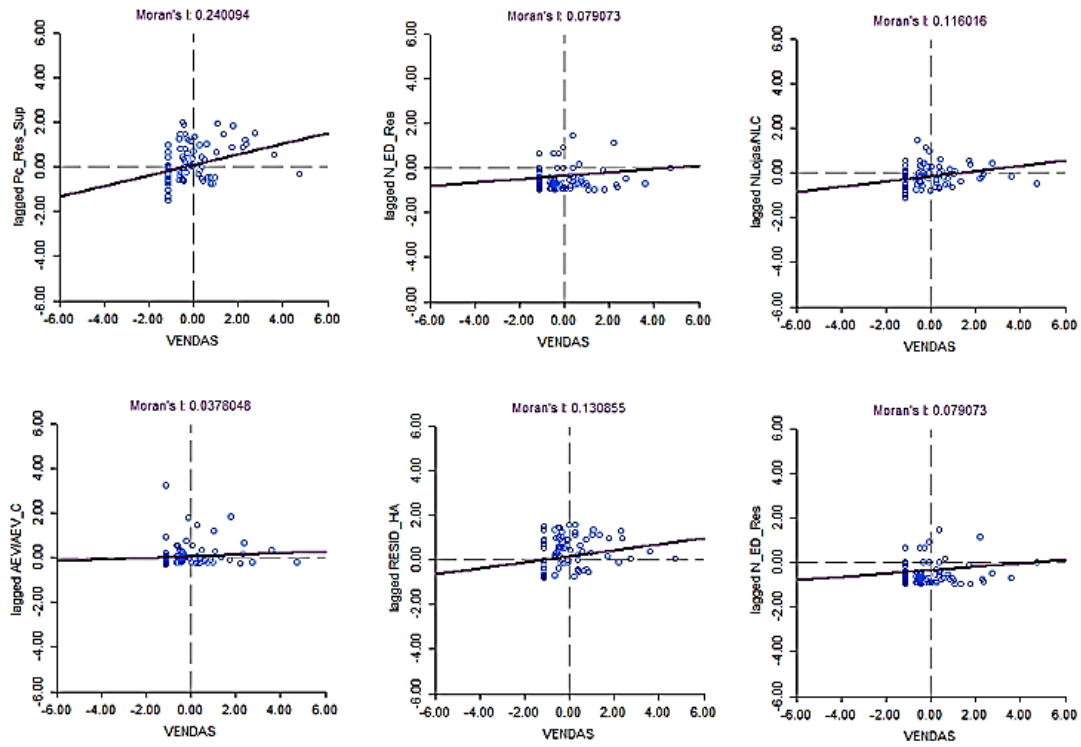
Fonte: Elaboração Própria (com recurso ao GeoDa).

*FIGURA 2 – Estatística I de Moran Local referente à variável vendas – Ano 2018*



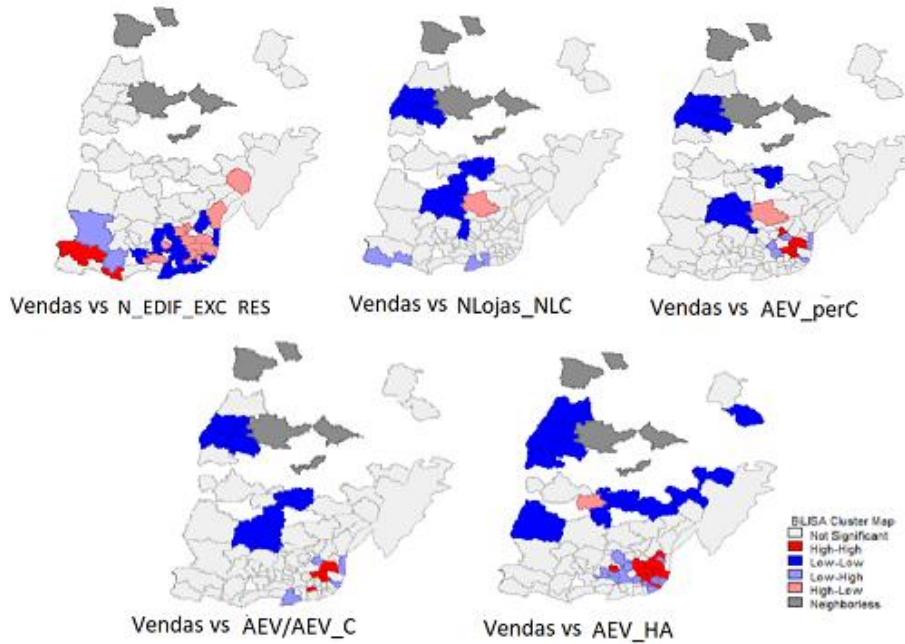
Fonte: Elaboração Própria (com recurso ao GeoDa).

**Figura 3 – Estatística I de Moran Bivariada Global**



Fonte: Elaboração Própria (com recurso ao GeoDa).

**FIGURA 4 – Estatística I de Moran Bivariada Local – Ano 2018**



Fonte: Elaboração Própria (com recurso ao GeoDa).

*Output 1 – Resultados da estimação por MQO*

Source	SS	df	MS	Number of obs	=	123
Model	11337.9476	11	1030.72251	F(11, 111)	=	34.67
Residual	3300.00312	111	29.7297578	Prob > F	=	0.0000
Total	14637.9507	122	119.983203	R-squared	=	0.7746
				Adj R-squared	=	0.7522
				Root MSE	=	5.4525

VENDAS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Perc_IND_RESID_65	-.1868486	.1462994	-1.28	0.204	-.4767506 .1030534
Perc_IND_ENS_SUP	-.0129971	.1053076	-0.12	0.902	-.2216711 .195677
Perc_IND_DESEMP	-.0664204	.8616343	-0.08	0.939	-1.773806 1.640965
TMD	.0111169	.0139622	0.80	0.428	-.0165502 .038784
AEV_AEV_C	-1.302926	.5980518	-2.18	0.031	-2.488006 -.1178468
RESID_HA	.064259	.0172166	3.73	0.000	.0301431 .0983749
NLojas_NLC	7.474738	3.873324	1.93	0.056	-.2005119 15.14999
N_EDIF_EXC_RES	.0008107	.0001583	5.12	0.000	.0004971 .0011243
Perc_Mulheres_RES	.1200869	.4196621	0.29	0.775	-.7115015 .9516754
Area_NLojT	.0006958	.0011259	0.62	0.538	-.0015352 .0029268
AEV_perC	191.8496	16.38617	11.71	0.000	159.3793 224.3199
_cons	-6.898295	21.69987	-0.32	0.751	-49.89803 36.10144

Fonte: Elaboração Própria (com recurso ao STATA).

*Output 2 – Resultados da estimação do modelo SAR com todas as variáveis por MV*

Spatial autoregressive model	Number of obs	=	123
Maximum likelihood estimates	Wald chi2(12)	=	443.79
	Prob > chi2	=	0.0000
Log likelihood = -374.64699	Pseudo R2	=	0.7781

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
<b>VENDAS</b>					
Perc_IND_RESID_65	-.1212859	.1397807	-0.87	0.386	-.395251 .1526791
Perc_IND_ENS_SUP	-.0575332	.1003451	-0.57	0.566	-.2542061 .1391397
Perc_IND_DESEMP	-.4248454	.820445	-0.52	0.605	-2.032888 1.183197
TMD	-.0029362	.0146027	-0.20	0.841	-.0315571 .0256846
AEV_AEV_C	-1.410096	.5595142	-2.52	0.012	-2.506724 -.3134681
RESID_HA	.0668341	.0160872	4.15	0.000	.0353038 .0983645
NLojas_NLC	9.426584	3.724644	2.53	0.011	2.126415 16.72675
N_EDIF_EXC_RES	.0009706	.0001656	5.86	0.000	.0006459 .0012952
Perc_Mulheres_RES	.0858624	.3913473	0.22	0.826	-.6811643 .8528891
Area_NLojT	.00064	.0010493	0.61	0.542	-.0014167 .0026966
AEV_perC	191.8965	15.2675	12.57	0.000	161.9728 221.8203
_cons	-5.1867	20.23465	-0.26	0.798	-44.84588 34.47248
<b>W_C</b>					
VENDAS	.1873699	.0884461	2.12	0.034	.0140187 .3607211
var(e.VENDAS)	25.809	3.292027			20.10006 33.13941

Wald test of spatial terms: chi2(1) = 4.49 Prob > chi2 = 0.0341

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 3 – Resultados da estimação do modelo SAR restrito por MV (após extração das variáveis irrelevantes)**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(6)    =     427.90
                                  Prob > chi2      =     0.0000
Log likelihood = -376.3183        Pseudo R2      =     0.7717
```

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
AEV_AEV_C	-1.411385	.5170699	-2.73	0.006	-2.424823	-.3979465
RESID_HA	.0518818	.0131679	3.94	0.000	.0260732	.0776904
NLojas_NLC	8.643817	3.405611	2.54	0.011	1.968942	15.31869
N_EDIF_EXC_RES	.0008898	.000144	6.18	0.000	.0006075	.0011721
AEV_perC	189.9049	14.02612	13.54	0.000	162.4142	217.3956
_cons	-4.442617	1.251176	-3.55	0.000	-6.894877	-1.990358
<b>W_C</b>						
VENDAS	.1357785	.0703243	1.93	0.054	-.0020545	.2736115
var(e.VENDAS)	26.5612	3.387281			20.68693	34.10352

Wald test of spatial terms:                      chi2(1) = 3.73                      Prob > chi2 = 0.0535

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 4 – Resultados da estimação do modelo SEM com todas as variáveis por MV**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(11)   =     423.67
                                  Prob > chi2      =     0.0000
Log likelihood = -374.40206      Pseudo R2      =     0.7710
```

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
Perc_IND_RESID_65	-.0278038	.1607005	-0.17	0.863	-.342771	.2871635
Perc_IND_ENS_SUP	-.0578711	.1022098	-0.57	0.571	-.2581986	.1424564
Perc_IND_DESEMP	-.3410474	.8415336	-0.41	0.685	-1.990423	1.308328
TMD	.0062254	.0148196	0.42	0.674	-.0228205	.0352713
AEV_AEV_C	-1.430083	.5411754	-2.64	0.008	-2.490767	-.369399
RESID_HA	.0722183	.0162814	4.44	0.000	.0403073	.1041293
NLojas_NLC	8.846519	3.539344	2.50	0.012	1.909533	15.78351
N_EDIF_EXC_RES	.0008514	.0001613	5.28	0.000	.0005353	.0011675
Perc_Mulheres_RES	.0743296	.3887135	0.19	0.848	-.6875349	.836194
Area_NLojT	.0005295	.0010271	0.52	0.606	-.0014836	.0025426
AEV_perC	194.1929	14.50593	13.39	0.000	165.7618	222.624
_cons	-5.787395	19.77731	-0.29	0.770	-44.55021	32.97542
<b>W_C</b>						
e.VENDAS	.4309413	.1782388	2.42	0.016	.0815997	.780283
var(e.VENDAS)	25.31637	3.255725			19.67594	32.57371

Wald test of spatial terms:                      chi2(1) = 5.85                      Prob > chi2 = 0.0156

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 5 – Resultados da estimação do modelo SEM restrito por MV (após extração das variáveis irrelevantes)**

```

Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates     Wald chi2(5)    =     418.26
                                  Prob > chi2      =     0.0000
Log likelihood = -374.95287      Pseudo R2      =     0.7692

```

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
AEV_AEV_C	-1.382187	.5006389	-2.76	0.006	-2.363421	-.4009524
RESID_HA	.0644901	.0136982	4.71	0.000	.0376421	.0913381
NLojas_NLC	9.111971	3.339892	2.73	0.006	2.565903	15.65804
N_EDIF_EXC_RES	.0008023	.0001468	5.46	0.000	.0005145	.0010901
AEV_perC	190.9155	13.54526	14.09	0.000	164.3672	217.4637
_cons	-3.830937	1.303545	-2.94	0.003	-6.385839	-1.276035
<b>W_C</b>						
e.VENDAS	.4241683	.1577478	2.69	0.007	.1149884	.7333483
var(e.VENDAS)	25.56015	3.280192			19.8759	32.87003

Wald test of spatial terms:                      chi2(1) = 7.23                      Prob > chi2 = 0.0072

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 6 – Resultados da estimação do modelo SAC com todas as variáveis por MV**

```

Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates     Wald chi2(12)   =     425.26
                                  Prob > chi2      =     0.0000
Log likelihood = -374.36        Pseudo R2      =     0.7751

```

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
Perc_IND_RESID_65	-.0629064	.1814697	-0.35	0.729	-.4185806	.2927677
Perc_IND_ENS_SUP	-.0609248	.1021309	-0.60	0.551	-.2610978	.1392481
Perc_IND_DESEMP	-.3701316	.8403924	-0.44	0.660	-2.017271	1.277007
TMD	.0021794	.0186818	0.12	0.907	-.0344363	.038795
AEV_AEV_C	-1.438166	.548782	-2.62	0.009	-2.513759	-.3625729
RESID_HA	.070743	.0170218	4.16	0.000	.0373809	.1041051
NLojas_NLC	9.013682	3.64314	2.47	0.013	1.873259	16.15411
N_EDIF_EXC_RES	.000898	.0002143	4.19	0.000	.000478	.0013181
Perc_Mulheres_RES	.0823518	.3905562	0.21	0.833	-.6831244	.847828
Area_NLojT	.0005701	.0010428	0.55	0.585	-.0014737	.0026139
AEV_perC	194.4791	14.84047	13.10	0.000	165.3923	223.5658
_cons	-5.702635	19.98133	-0.29	0.775	-44.86533	33.46006
<b>W_C</b>						
VENDAS	.0762596	.2228025	0.34	0.732	-.3604252	.5129444
e.VENDAS	.2960006	.4472916	0.66	0.508	-.5806747	1.172676
var(e.VENDAS)	25.54526	3.312805			19.81177	32.93801

Wald test of spatial terms:                      chi2(2) = 5.12                      Prob > chi2 = 0.0772

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 7 – Resultados da estimação do modelo SAC restrito por MV (após extração das variáveis irrelevantes)**

```
Spatial autoregressive model      Number of obs      =      123
Maximum likelihood estimates      Wald chi2(6)       =      418.52
                                   Prob > chi2         =      0.0000
Log likelihood = -374.95094       Pseudo R2         =      0.7688
```

VENDAS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
AEV_AEV_C	-1.381762	.5000783	-2.76	0.006	-2.361898	-.4016265
RESID_HA	.0648991	.015169	4.28	0.000	.0351684	.0946298
NLojas_NLC	9.111537	3.336138	2.73	0.006	2.572828	15.65025
N_EDIF_EXC_RES	.0007987	.0001577	5.07	0.000	.0004897	.0011077
AEV_perC	190.8789	13.55025	14.09	0.000	164.3209	217.4368
_cons	-3.793408	1.44014	-2.63	0.008	-6.616031	-.9707855
<b>W_C</b>						
VENDAS	-.0082556	.1338771	-0.06	0.951	-.2706499	.2541387
e.VENDAS	.4366535	.255537	1.71	0.087	-.0641899	.9374969
var(e.VENDAS)	25.5294	3.314876			19.7932	32.92799

Wald test of spatial terms:                      chi2(2) = 7.37                      Prob > chi2 = 0.0251

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 8 – Resultados da estimação do modelo SAR irrestrito por Pseudo Máxima Verosimilhança**

```
Spatial autoregressive model      Number of obs      =      123
Maximum likelihood estimates      Wald chi2(12)     =      445.10
                                   Prob > chi2         =      0.0000
Log pseudolikelihood = -374.64699   Pseudo R2         =      0.7781
```

VENDAS	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
Perc_IND_RESID_65	-.1212859	.1376274	-0.88	0.378	-.3910306	.1484588
Perc_IND_ENS_SUP	-.0575332	.099761	-0.58	0.564	-.2530612	.1379948
Perc_IND_DESEMP	-.4248454	.8178946	-0.52	0.603	-2.027889	1.178198
TMD	-.0029362	.0145031	-0.20	0.840	-.0313617	.0254892
AEV_AEV_C	-1.410096	.5587074	-2.52	0.012	-2.505142	-.3150495
RESID_HA	.0668341	.0160482	4.16	0.000	.0353802	.0982881
NLojas_NLC	9.426584	3.716537	2.54	0.011	2.142306	16.71086
N_EDIF_EXC_RES	.0009706	.000166	5.85	0.000	.0006453	.0012958
Perc_Mulheres_RES	.0858624	.3912461	0.22	0.826	-.680966	.8526907
Area_NLojT	.00064	.0010491	0.61	0.542	-.0014162	.0026961
AEV_perC	191.8965	15.28111	12.56	0.000	161.9461	221.847
_cons	-5.1867	20.23283	-0.26	0.798	-44.84232	34.46892
<b>W_C</b>						
VENDAS	.1873699	.0939287	1.99	0.046	.0032731	.3714667
var(e.VENDAS)	25.809	7.930191			14.13276	47.13193

Wald test of spatial terms:                      chi2(1) = 3.98                      Prob > chi2 = 0.0461

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 9 – Resultados da estimação do modelo SAR restrito por Pseudo Máxima Verosimilhança**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(6)    =     429.87
                                  Prob > chi2      =     0.0000
Log pseudolikelihood = -376.3183    Pseudo R2      =     0.7717
```

VENDAS	Robust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
<b>VENDAS</b>						
AEV_AEV_C	-1.411385	.5171395	-2.73	0.006	-2.42496	-.3978101
RESID_HA	.0518818	.0134442	3.86	0.000	.0255316	.078232
NLojas_NLC	8.643817	3.403704	2.54	0.011	1.97268	15.31495
N_EDIF_EXC_RES	.0008898	.0001455	6.11	0.000	.0006045	.0011751
AEV_perC	189.9049	14.04296	13.52	0.000	162.3812	217.4286
_cons	-4.442617	1.255966	-3.54	0.000	-6.904266	-1.980968
<b>W_C</b>						
VENDAS	.1357785	.0758325	1.79	0.073	-.0128504	.2844074
var(e.VENDAS)	26.5612	7.982156			14.73822	47.86855

Wald test of spatial terms:                    chi2(1) = 3.21                    Prob > chi2 = 0.0734

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 10 – Resultados da estimação do modelo SEM irrestrito por Pseudo Máxima Verosimilhança**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(11)   =     423.80
                                  Prob > chi2      =     0.0000
Log pseudolikelihood = -374.40206    Pseudo R2      =     0.7710
```

VENDAS	Robust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
<b>VENDAS</b>						
Perc_IND_RESID_65	-.0278038	.1458826	-0.19	0.849	-.3137285	.2581209
Perc_IND_ENS_SUP	-.0578711	.1007442	-0.57	0.566	-.255326	.1395838
Perc_IND_DESEMP	-.3410474	.8350003	-0.41	0.683	-1.977618	1.295523
TMD	.0062254	.0146278	0.43	0.670	-.0224446	.0348954
AEV_AEV_C	-1.430083	.5393482	-2.65	0.008	-2.487186	-.3729802
RESID_HA	.0722183	.0159259	4.53	0.000	.0410041	.1034324
NLojas_NLC	8.846519	3.487801	2.54	0.011	2.010555	15.68248
N_EDIF_EXC_RES	.0008514	.0001596	5.34	0.000	.0005386	.0011642
Perc_Mulheres_RES	.0743296	.3880642	0.19	0.848	-.6862623	.8349215
Area_NLojT	.0005295	.0010239	0.52	0.605	-.0014773	.0025363
AEV_perC	194.1929	14.49465	13.40	0.000	165.7839	222.6019
_cons	-5.787395	19.77058	-0.29	0.770	-44.53703	32.96224
<b>W_C</b>						
e.VENDAS	.4309413	.1746666	2.47	0.014	.088601	.7732817
var(e.VENDAS)	25.31637	7.725749			13.92008	46.04272

Wald test of spatial terms:                    chi2(1) = 6.09                    Prob > chi2 = 0.0136

Fonte: Elaboração Própria (com recurso ao STATA).



**Output 11 – Resultados da estimação do modelo SEM restrito por Pseudo Máxima Verosimilhança**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(5)    =     418.83
                                  Prob > chi2      =      0.0000
Log pseudolikelihood = -374.95287  Pseudo R2      =      0.7692
```

VENDAS	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
AEV_AEV_C	-1.382187	.5006342	-2.76	0.006	-2.363412	-.4009617
RESID_HA	.0644901	.0134967	4.78	0.000	.038037	.0909432
NLojas_NLC	9.111971	3.330788	2.74	0.006	2.583746	15.64019
N_EDIF_EXC_RES	.0008023	.0001468	5.46	0.000	.0005145	.00109
AEV_perC	190.9155	13.52523	14.12	0.000	164.4065	217.4244
_cons	-3.830937	1.291503	-2.97	0.003	-6.362237	-1.299637
<b>W_C</b>						
e.VENDAS	.4241683	.1754728	2.42	0.016	.080248	.7680887
var(e.VENDAS)	25.56015	7.744406			14.11432	46.28783

Wald test of spatial terms:                      chi2(1) = 5.84                      Prob > chi2 = 0.0156

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 12 – Resultados da estimação do modelo SAC irrestrito por Pseudo Máxima Verosimilhança**

```
Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(12)   =     425.48
                                  Prob > chi2      =      0.0000
Log pseudolikelihood = -374.36                      Pseudo R2      =      0.7751
```

VENDAS	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
Perc_IND_RESID_65	-.0629064	.1431364	-0.44	0.660	-.3434486	.2176357
Perc_IND_ENS_SUP	-.0609248	.1018831	-0.60	0.550	-.2606121	.1387624
Perc_IND_DESEMP	-.3701316	.8388488	-0.44	0.659	-2.014245	1.273982
TMD	.0021794	.0164262	0.13	0.894	-.0300154	.0343741
AEV_AEV_C	-1.438166	.5479704	-2.62	0.009	-2.512168	-.3641636
RESID_HA	.070743	.0160095	4.42	0.000	.039365	.1021211
NLojas_NLC	9.013682	3.628731	2.48	0.013	1.901501	16.12586
N_EDIF_EXC_RES	.000898	.0001812	4.95	0.000	.0005428	.0012533
Perc_Mulheres_RES	.0823518	.3900218	0.21	0.833	-.6820768	.8467804
Area_NLojT	.0005701	.0010351	0.55	0.582	-.0014587	.0025989
AEV_perC	194.4791	14.78281	13.16	0.000	165.5053	223.4528
_cons	-5.702635	19.98086	-0.29	0.775	-44.86439	33.45912
<b>W_C</b>						
VENDAS	.0762596	.1192698	0.64	0.523	-.157505	.3100241
e.VENDAS	.2960006	.2167649	1.37	0.172	-.1288508	.720852
var(e.VENDAS)	25.54526	7.834063			14.00455	46.59631

Wald test of spatial terms:                      chi2(2) = 4.02                      Prob > chi2 = 0.1341

Fonte: Elaboração Própria (com recurso ao STATA).

**Output 13 – Resultados da estimação do modelo SAC restrito por Pseudo Máxima Verosimilhança**

```

Spatial autoregressive model      Number of obs   =      123
Maximum likelihood estimates      Wald chi2(6)    =     418.99
                                  Prob > chi2      =     0.0000
Log pseudolikelihood = -374.95094  Pseudo R2      =     0.7688

```

VENDAS	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>VENDAS</b>						
AEV_AEV_C	-1.381762	.500033	-2.76	0.006	-2.361809	-.4017154
RESID_HA	.0648991	.0139342	4.66	0.000	.0375887	.0922096
NLojas_NLC	9.111537	3.331698	2.73	0.006	2.581529	15.64155
N_EDIF_EXC_RES	.0007987	.0001536	5.20	0.000	.0004977	.0010997
AEV_perC	190.8789	13.51022	14.13	0.000	164.3993	217.3584
_cons	-3.793408	1.412284	-2.69	0.007	-6.561434	-1.025382
<b>W_C</b>						
VENDAS	-.0082556	.095128	-0.09	0.931	-.1947031	.1781919
e.VENDAS	.4366535	.1863726	2.34	0.019	.0713699	.8019371
var(e.VENDAS)	25.5294	7.747732			14.08366	46.27704

```

Wald test of spatial terms:      chi2(2) = 6.16      Prob > chi2 = 0.0461

```

Fonte: Elaboração Própria (com recurso ao STATA).