



NOVA

IMS

Information
Management
School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

**Modelação das taxas de mortalidade associadas
a hábitos alimentares nos municípios
Portugueses**

Uma análise exploratória utilizando Geographically
Weighted Regression

Adriana Loução Guerreiro Ramos

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Estatística e Gestão de
Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**MODELAÇÃO DAS TAXAS DE MORTALIDADE ASSOCIADAS A
HÁBITOS ALIMENTARES NOS MUNICÍPIOS PORTUGUESES**
Uma análise exploratória utilizando Geographically Weighted Regression

por

Adriana Loução Guerreiro Ramos

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização em Análise e Gestão de Informação.

Orientadora: Professora Doutora Ana Cristina Costa

Coorientadora: Professora Doutora Sara Ribeiro

Agosto 2019

Dedico este trabalho à minha família.

AGRADECIMENTOS

Agradecer à Professora Doutora Ana Cristina Costa por todo o apoio, disponibilidade incondicional, por todos os ensinamentos, pelo rigor e pela paciência que teve ao longo deste trabalho.

Muito agradeço também à Professora Doutora Sara Ribeiro pela georreferenciação dos restaurantes de *fast food*, pelas sugestões e ideias, por todo o apoio e correções.

Sem o auxílio e motivação de ambas este trabalho certamente não estaria concluído. Foi uma aprendizagem constante e o meu agradecimento é pouco perante toda a ajuda que me concederam.

À minha irmã Daniela, que nos bons e maus momentos esteve sempre disponível para mim, ajudando-me a manter focada na reta final, auxiliando-me nas correções e a debater ideias.

Aos meus pais e avós, que mesmo distantes me apoiaram sempre e me incentivaram a não desistir dos meus sonhos, a fazer as minhas escolhas de forma independente e de acordo com os meus gostos.

Agradeço à Telma e à Andrea, as minhas eternas colegas de casa e amigas, por serem quem são e por me ajudarem em todos os momentos.

Às minhas amigas de sempre, Andreia S., Andreia V., Joana e Liza, que mesmo distantes se mantiveram presentes das mais diversas formas sempre com uma palavra de alento.

À tia Paula por ser quem é e por me aconselhar e ouvir.

À Maria Joana pelas sessões de pesquisa, companhia durante o processo de escrita e ainda todo o apoio.

Aos meus colegas (amigos) de trabalho (dos vários locais por onde passei durante este período) por todos os conselhos, incentivos e auxílio.

RESUMO

Uma das principais preocupações da investigação em saúde é o desenvolvimento e melhoria dos indicadores que permitem monitorizar a saúde e/ou os fatores de risco e conectá-los ao ambiente da população. É sempre difícil medir a informação da população relativa à saúde porque existem muitos fatores a ter em conta. A Organização Mundial de Saúde (OMS) refere as taxas de mortalidade como relevantes indicadores para caracterizar a saúde geral da população.

O principal objetivo deste estudo é identificar os indicadores socioeconómicos (ex. literacia, desemprego, poder de compra, consultas por habitante, restaurantes por habitante, etc.) e variáveis espaciais (ex. distância aos cuidados de saúde) que poderão ser associados com as taxas de mortalidade relacionadas com os hábitos alimentares, e os municípios onde são mais determinantes para essas taxas, com a finalidade de melhor compreender o estado de saúde nos municípios de Portugal continental.

A metodologia utilizada é composta por duas fases principais. Primeiro, o conjunto das variáveis socioeconómicas e espaciais são analisadas utilizando um processo iterativo que aplica o Ordinary Least Squares (OLS), para obter diferentes modelos de regressão linear alternativos, selecionando-se o melhor modelo possível tendo em consideração diversos testes estatísticos e medidas de diagnóstico. De forma a lidar com a não estacionaridade espacial e para investigar as relações locais, na segunda fase utiliza-se o modelo Geographically Weighted Regression (GWR) com as variáveis utilizadas no melhor modelo de OLS. O modelo GWR é também diagnosticado para a multicolinearidade local das variáveis explicativas. A estatística Global Moran's I é utilizada para diagnosticar a possível existência de autocorrelação espacial dos resíduos em todos os modelos testados. A precisão dos parâmetros estimados pelo GWR é avaliada através dos erros padrão locais.

As variáveis utilizadas no modelo GWR foram “Resposta hospitalar”, “Rácio entre supermercados e lojas de conveniência”, “Distância média a restaurantes *fast food* por habitante” e “Percentagem de população com educação de 2º ciclo”, as quais explicam entre 42% e 64% da variabilidade das taxas de mortalidade nos municípios. O modelo tem um maior poder explicativo em alguns dos municípios da região centro, destacando-se os municípios do distrito de Coimbra. A variável “Distância média a restaurantes *fast food* por habitante” tem os coeficientes positivos em todos os municípios, os quais são mais elevados no centro litoral e na

Área Metropolitana do Porto. As restantes variáveis têm coeficientes negativos e maior poder explicativo no interior do país.

Espera-se que esta análise exploratória possa contribuir para o conhecimento das conexões locais entre os padrões socioeconómicos da população e as taxas de mortalidade relacionadas com os hábitos alimentares.

PALAVRAS-CHAVE

Indicadores de saúde; Regressão espacial; Geographically Weighted Regression (GWR); Taxas de mortalidade; Portugal.

ABSTRACT

One of the main concerns of health research is the development and improvement of the indicators that allow monitoring the health and/or risk factors and connect them with the population environment. It is always difficult to measure health information from the population because there are many factors to consider. The World Health Organization (WHO) refers the indicators related to mortality rates as relevant for the characterization of the overall population health.

The main goal of this study is to identify socioeconomic indicators (e.g. illiteracy, unemployment, purchasing power, medical appointments by habitant, restaurants by habitant, etc.) and spatial variables (e.g. distance to health facilities) that might be associated with mortality rates caused by diseases associated to eating habits, and the municipalities where they are more determinant to these rates, in order to better understand the health status in the municipalities of mainland Portugal.

The methodological framework has two main stages. In the first one, a set of socioeconomic and spatial variables are analyzed using an iterative process that applies Ordinary Least Squares (OLS) to obtain different alternative linear regression models, choosing the best possible model according to multiple statistical tests and diagnostic measures. In order to deal with spatial nonstationary and to investigate local relationships, the second stage is based on a Geographically Weighted Regression (GWR) model with the variables that were included in the best OLS model. The GWR model is also diagnosed for local multicollinearity of the explanatory variables. The Global Moran's I statistic is used to diagnose the possible existence of spatial autocorrelation of the residuals in all tested models. The accuracy of the GWR parameter estimates is assessed through the local standard errors.

The selected variables in the GWR model were "Hospital response", "Grocery-to-convenience stores ratio", "Average distance to *fast food* restaurants per habitant" and "2nd cycle education", which explain between 42% and 64% of the mortality rates variability in the municipalities. The model has greater explanatory power in some municipalities of the center region, with more relevance in Coimbra district. The local coefficients of the variable "Average distance to *fast food* restaurants per habitant" are positive in all municipalities, and are higher in the coastal center and in the metropolitan area of Porto. The remaining variables have negative coefficients and higher explanatory power inland.

We hope that this exploratory spatial data analysis may contribute to the knowledge of the local connections and patterns of socioeconomic characteristics of the population and mortality rates caused by diseases related to eating habits in Portugal.

KEYWORDS

Health indicators; Spatial regression; Geographically Weighted Regression (GWR); Mortality rates; Portugal.

ÍNDICE

1. Introdução.....	1
1.1. Enquadramento e relevância do problema.....	1
1.2. Objetivos.....	2
1.3. Organização do documento.....	3
2. Revisão de Literatura	4
2.1. Saúde e Indicadores de saúde	4
2.2. Enquadramento metodológico.....	6
3. Metodologia.....	10
3.1. Região de estudo e dados.....	11
3.1.1. Variável dependente	13
3.1.2. Variáveis independentes	13
3.2. Análise exploratória	16
3.3. Métodos utilizados na modelação.....	17
3.3.1. Modelação Global (OLS)	17
3.3.2. Modelação Local (GWR)	18
4. Resultados e discussão	19
4.1. Análise exploratória.....	19
4.1.1. Variável dependente	19
4.1.2. Variáveis independentes	22
4.2. Modelação global (OLS).....	26
4.3. Modelo GWR.....	27
4.3.1. Coeficientes e respetivos erros padrão	31
4.3.2. Comparação dos modelos OLS vs. GWR.....	37
5. Conclusão.....	38
5.1. Limitações	39
5.2. Recomendações para trabalhos futuros.....	39
Referências	41
Anexos	46
Lista de municípios com valores em falta de Portugal	46
Modo de agrupamento das variáveis baseadas no CAE.....	47

ÍNDICE DE FIGURAS

Figura 2.1 Determinantes de saúde (Retirado de (Ministério da Saúde, 2018))	4
Figura 3.1 Processo de modelação com GWR.....	10
Figura 3.2 Enquadramento metodológico	11
Figura 3.3 Habitantes de Portugal continental por município.....	12
Figura 4.1. Projeção das taxas de mortalidade relacionadas com hábitos alimentares, em 2011.....	20
Figura 4.2 Histograma regional - Taxa de mortalidade.....	20
Figura 4.3 Output do método Local Anselin's Moran's.....	21
Figura 4.4. Projeção da percentagem de população com nível de educação de 2º Ciclo.	22
Figura 4.5 Projeção do rácio entre supermercados e lojas de conveniência.....	23
Figura 4.6 Projeção da resposta hospitalar.....	24
Figura 4.7 Projeção da distância média a restaurantes de fast food por habitante (Km/hab).....	25
Figura 4.8 Diagnóstico de multicolinearidade do modelo GWR.....	29
Figura 4.9 R ² Local do modelo GWR.....	30
Figura 4.10 Local Moran's I - Análise de resíduos do modelo GWR	30
Figura 4.11 GWR Educação de 2º ciclo	31
Figura 4.12 Erros padrão associados à educação de 2º ciclo.....	32
Figura 4.13 GWR Distância média a restaurantes fast food por habitante (Km/hab)...	33
Figura 4.14 Erros padrão associados à distância média a restaurantes fast food por habitante.	33
Figura 4.15 GWR Resposta hospitalar	34
Figura 4.16 Erros padrão associados à variável resposta hospitalar	35
Figura 4.17 GWR Rácio entre supermercados e lojas de conveniência.....	36
Figura 4.18 Erros padrão rácio entre supermercados e lojas de conveniência	36

ÍNDICE DE TABELAS

Tabela 3.1 Descrição das variáveis utilizadas e respetiva descrição e origem	13
Tabela 4.1 Percentagens de óbitos para cada uma das causas utilizadas na análise, para Portugal continental, no ano 2011.....	19
Tabela 4.2 Estatísticas descritivas da variável dependente [taxa de mortalidade (%)]	20
Tabela 4.3 Resultado do modelo OLS.....	27
Tabela 4.4 Síntese dos resultados do GWR.....	28
Tabela 4.5 Indicadores de comparação das regressões Global (OLS) e Local (GWR)	37
Tabela 0.1 Classificação das Atividades Económicas (CAE) de Restauração utilizadas no indicador.....	47
Tabela 0.2 Classificação das Atividades Económicas (CAE) de Comercio de bens alimentares utilizadas no indicado “Rácio entre o nº supermercados e o nº lojas de conveniência”	48

LISTA DE SIGLAS E ABREVIATURAS

AICc	Corrected Akaike's Information Criterion (Critério de Informação de Akaike Corrigido)
AVC	Acidente Vascular Cerebral
CAE	Classificação Portuguesa das Atividades Económicas
DGS	Direção Geral de Saúde
EUA	Estados Unidos da América
GWR	Geographically Weighted Regression
HIA	Health Impact Assessment
INE	Instituto Nacional de Estatística
OLS	Ordinary Least Squares (Método dos Mínimos Quadrados Ordinários)
OMS	Organização Mundial de Saúde
PNS	Plano Nacional de Saúde
SIG	Sistemas de Informação Geográfica
SNS	Serviço Nacional de Saúde
VIF	Variance Inflation Factor

1. INTRODUÇÃO

1.1. ENQUADRAMENTO E RELEVÂNCIA DO PROBLEMA

Melhorar o estado de saúde e o bem-estar da população é um dos principais objetivos da maioria dos governos. Portugal não é exceção, como se comprova na Lei de Bases da Saúde (Lei nº48/90 de 24 de Agosto da Assembleia da República, 1990), no artigo 64º da Constituição da República Portuguesa: “*Todos têm direito à proteção da saúde e o dever de a defender e promover*” (1976). Mais recentemente, a 16 de agosto de 2019, foi promulgada uma nova Lei de Bases da Saúde que complementa a anterior (Presidência da República, 2019).

Fazer investigação na área da saúde é de extrema importância, pois possibilita a identificação de problemas na população e os respetivos aspetos a melhorar em termos de políticas de saúde. Para avaliar o estado de saúde de uma determinada população, as ferramentas estatísticas facilitam o processo de gestão de informação e a tomada de decisão. Em particular, a estatística espacial e a utilização de sistemas de informação geográfica (SIG) permitem obter resultados que caracterizam a saúde da população em cada local e que podem posteriormente providenciar melhores políticas para cada uma delas (Castelli et al., 2013).

Os estilos de vida pouco saudáveis e os hábitos alimentares impactam a saúde física e psicossocial sendo ainda fatores de risco, bem conhecidos, para o desenvolvimento de doenças potencialmente fatais (Ruano-Rodríguez, Serra-Majem, & Dubois, 2015). São exemplos os acidentes vasculares cerebrais (AVC), doenças cardíacas e diabetes (Micha, Peñalvo, et al., 2017).

Neste sentido, de acordo com a Organização Mundial de Saúde (OMS), a mortalidade é adequada para avaliar o estado de saúde da população. O relatório anual das *Estatísticas Mundiais de Saúde 2018* tem presente a *Mortalidade devido a doenças não comunicáveis* (inclui doenças cardiovasculares, diabetes, entre outros fora da listagem das comunicáveis) (World Health Organization, 2018b) e no relatório **Core health indicators**, à semelhança das edições dos anos anteriores de ambos os relatórios, referem a mortalidade como um dos principais indicadores de saúde (World Health Organization, 2018a). Estas taxas variam de acordo com os objetivos dos estudos, podendo ir desde mortalidade infantil passando pelas taxas de mortalidade por causas específicas (World Health Organization, 2014b).

Na literatura encontram-se diversas investigações internacionais que utilizam as taxas de mortalidade relacionadas com a alimentação como indicador, com diferentes abordagens e metodologias. Como no Médio Oriente e norte de África (Afshin et al., 2015), na Coreia do Sul (Cho et al., 2017) ou ainda uma revisão sistemática com a meta análise de estudos que abordam os nutrientes e fazem ligação com as doenças crónicas (*NutriCoDE*) (Micha, Shulkin, et al., 2017).

Dos estudos referentes à população portuguesa, verificou-se que existe pouca investigação publicada no que diz respeito a modelos estatísticos, cujo objetivo seja identificar e analisar os fatores determinantes do estado de saúde da população.

1.2. OBJETIVOS

Com este estudo pretende-se fazer uma análise exploratória do estado de saúde da população, por municípios de Portugal continental, e averiguar quais os fatores que o influenciam. Para tal utiliza-se a taxa de mortalidade relacionada com hábitos alimentares como indicador do estado de saúde, a qual contabiliza apenas as doenças do aparelho digestivo, diabetes e doenças do aparelho circulatório.

Assim, o objetivo principal deste estudo consiste em identificar fatores determinantes da variabilidade espacial das taxas de mortalidade relacionadas com hábitos alimentares, bem como a relevância de cada fator em cada município. Em particular, serão investigados indicadores socioeconómicos (literacia, desemprego, poder de compra, consultas por habitante, restaurantes por habitante, entre outros) e variáveis espaciais (distância aos cuidados de saúde, distância a restaurantes *fast food*, etc.).

Em suma, os objetivos específicos encontram-se indicados de seguida:

- Analisar o padrão espacial das taxas de mortalidade relacionadas com hábitos alimentares (variável dependente) nos municípios de Portugal continental.
- Identificar os principais indicadores socioeconómicos que influenciam a variável dependente.
- Identificar os padrões geográficos para cada um dos indicadores encontrados.
- Estimar um modelo de regressão linear confiável, usando o método dos Mínimos Quadrados Ordinários (será abreviado pela sigla inglesa OLS – Ordinary Least Squares).
- Estimar um modelo de regressão espacial local para investigar a influência de cada indicador socioeconómico em cada município.

- Identificar as regiões/municípios com piores ou melhores resultados da modelação espacial local.

Portanto, neste estudo, tentaremos dar resposta às seguintes questões de investigação:

1. Quais são os fatores socioeconómicos que afetam as taxas de mortalidade relacionadas com hábitos alimentares?
2. Quais são as regiões/municípios em que cada fator exerce maior influência?

1.3. ORGANIZAÇÃO DO DOCUMENTO

Neste primeiro capítulo, apresentou-se o enquadramento do estudo, a sua relevância e os respetivos objetivos. O capítulo 2 é dedicado a uma breve revisão de literatura sobre indicadores de saúde, mortalidade e hábitos alimentares. Esta revisão de literatura contribui para a identificação de fatores socioeconómicos que poderão afetar as taxas de mortalidade relacionadas com hábitos alimentares. Aqui, são também revistos alguns estudos sobre este tema do ponto de vista metodológico. No capítulo 3 são apresentadas as principais metodologias utilizadas para modelar as taxas de mortalidade relacionadas com hábitos alimentares e são descritas as variáveis que foram utilizadas ao longo dos processos iterativos. No capítulo 4, são apresentados e discutidos os resultados da investigação, incluindo os resultados da modelação por OLS e por regressão espacial local (GWR). Por fim, no capítulo 5, são apresentadas as principais conclusões, bem como as limitações do estudo e sugestões para trabalhos futuros.

2. REVISÃO DE LITERATURA

Este capítulo subdivide-se em dois subcapítulos. Na primeira parte apresentam-se os conceitos e estudos científicos relacionados com a saúde e os fatores que a influenciam, enquanto a segunda parte é referente aos métodos utilizados e as suas aplicações na área da saúde.

2.1. SAÚDE E INDICADORES DE SAÚDE

De acordo com a OMS, a definição de saúde consiste “no estado de completo bem-estar físico, mental e social, não se restringido apenas à ausência de doenças ou enfermidades.” (World Health Organization, 1948).

A compreensão do estado de saúde de uma população é uma tarefa complexa, que reflete o nível de saúde. Este conceito não depende exclusivamente dos cuidados médicos prestados, engloba também uma dimensão sociocultural, estando interligada com o desenvolvimento e o bem-estar (Santana, 2014; World Health Organization, 2017).

Segundo a OMS, vários são os fatores que condicionam o estado de saúde, tanto a nível individual como na população em geral. De uma forma mais ampla, os determinantes de saúde refletem as circunstâncias socioeconómicas, culturais e ambientais. Desagregando-se em condições de vida e trabalho como a habitação, a educação, o desemprego, entre outros. Adicionalmente, o estilo de vida e as características individuais também têm influência (Figura 2.1) (World Health Organization, 2017).

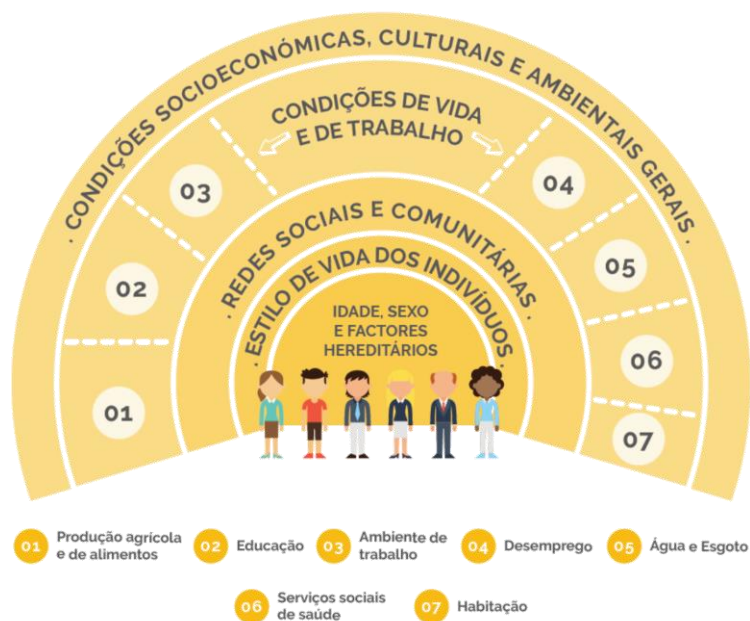


Figura 2.1 Determinantes de saúde (Retirado de (Ministério da Saúde, 2018))

Existem diversos estudos com a finalidade de conhecer e melhorar a saúde, através da análise dos fatores determinantes. A OMS desenvolveu o programa *Health Impact Assessment* (HIA) que pretende promover a saúde e auxiliar as tomadas de decisão que previnam doenças (World Health Organization, 2014a).

Por norma, os estudos desenvolvidos em Portugal apresentam uma abrangência regional reduzida, efetuando a decomposição por NUTS III ou por distritos (Barreto et al., 2016; Nogueira & Santana, 2005). O projeto *GeoHealthS* (<https://www.uc.pt/en/fluc/gigs/GeoHealthS/>; acessado em maio 2018), desenvolvido pela Universidade de Coimbra, tem como objetivo principal a “*avaliação da saúde da população portuguesa nos últimos 20 anos*”, através da elaboração de um *Índice de Saúde da População* ao nível dos municípios (Santana, 2015). Este índice incorpora duas grandes componentes com várias dimensões: para os “*Resultados em Saúde*” existem duas dimensões (mortalidade e morbilidade) e para “*Determinantes da Saúde*” existem quatro dimensões (económica e social, ambiente físico, cuidados de saúde e estilos de vida). As dimensões agregam 43 critérios de avaliação, aos quais estão associados 45 indicadores de desempenho, que descrevem os municípios em cada critério (Santana, 2015).

Uma vez que a análise do estado da saúde apresenta alguma complexidade e subjetividade, é importante encontrar indicadores. Estes são instrumentos de medida sumária que podem ser diretos (doenças, mortes, utilização de serviços de saúde, ...) ou indiretos (desenvolvimento social, educação, indicadores de pobreza, ...) (Larson & Mercer, 2004). Na área da saúde, os indicadores indiretos, também denominados por *proxies*, são os que mais se utilizam. Por norma, estes medem essencialmente a ausência de saúde (Santana, 2015).

A mortalidade é um dos principais indicadores de medidas do estado de saúde, tal como referido anteriormente na introdução. Este indicador é utilizado, partindo da premissa de que se o estado de saúde varia, este terá influência na mortalidade em causa. Por exemplo, se a população está mais afetada por determinada doença, a carga associada fragiliza o indivíduo levando a uma maior propensão de morte.

Para ir ao encontro dos objetivos deste trabalho, não serão analisados todos os tipos de mortalidade, mas apenas aqueles que correspondem a causas relacionadas com os hábitos alimentares (doenças cardiovasculares, diabetes e doenças do aparelho digestivo).

Estudos internacionais ao longo dos anos têm vindo a aumentar as análises que relacionam mortalidades e hábitos alimentares. Destaca-se uma investigação elaborada nos Estados Unidos

da América (EUA) com bastante detalhe na informação referente a consumos alimentares. O estudo foi conduzido em duas fases (1999-2002 e 2009-2012) e recorrendo a inquéritos, o seu objetivo principal era estimar a associação entre dez alimentos e a mortalidade (por doenças cardiovasculares, acidente vascular cerebral (AVC) e diabetes do tipo 2). Os resultados obtidos comprovam a importância de uma alimentação equilibrada (Micha, Peñalvo, et al., 2017).

As abordagens são as mais diversificadas, mas na sua grande maioria correspondem a mortalidade associada a doenças do aparelho circulatório (doenças cardiovasculares). Até ao momento já existem bastantes estudos que comprovam a relação de causalidade existente entre a mortalidade/doenças em causa e os consumos alimentares, exemplos disso são: alimentos prejudiciais à diabetes (fibras e hidratos de carbono em qualidade e quantidade) (Burger et al., 2012) ou ainda alimentos benéficos para as doenças cardiovasculares (fruta e vegetais) (Leenders et al., 2014; Verlangieri, Kapoghian, El-Dean, & Bush, 1985). O estudo de Leenders et al. (2014) também faz análise de outras causas de mortalidade específica como doenças do aparelho digestivo.

Em Portugal, como já foi referido existe pouca investigação sobre este assunto e a sua maioria corresponde a uma visão macro, com a limitação de não fornecer dados dos municípios em detalhe (Rodrigues, Trichopoulou, & De Almeida, 2008).

Mozaffarian (2016), numa revisão compreensiva relacionada com os hábitos alimentares e prioridades alimentares, apesar de ser apenas dirigida aos EUA, expõe aspetos importantes que podem ser transversais a outras populações. Salaria ainda a lacuna que existe na investigação das doenças cardiometabólicas e os fatores alimentares que possam estar associados (Mozaffarian, 2016).

Neste sentido os determinantes associados às doenças cardiovasculares seguem uma grande cadeia de acontecimentos, que por norma têm a sua origem no contexto socioeconómico em que cada um dos indivíduos se insere.

2.2. ENQUADRAMENTO METODOLÓGICO

O método GWR é uma técnica poderosa de modelação local concebida para lidar com a não estacionaridade dos dados, ou seja, quando estes não são constantes ao longo do espaço (C. Brunsdon, Fotheringham, & Charlton, 1996; D. C. Wheeler & Paez, 2002). A não estacionaridade

reflete-se geralmente em resíduos heterocedásticos nas regressões por Ordinary Least Squares (OLS). Fotheringham et al. (2002, p. 117) defendem que o método GWR permite também lidar com a autocorrelação espacial dos resíduos.

Este método é uma expansão da regressão linear clássica (OLS), sendo a principal diferença a forma como os dois métodos utilizam os dados. Enquanto o OLS é uma regressão global tratando os dados como um todo, o GWR é uma regressão local que trata cada observação (ponto ou polígono) de forma isolada. O GWR percorre todo o conjunto de dados centrando-se numa observação de cada vez e utilizando as observações vizinhas para estimar uma equação de regressão, na qual as observações vizinhas são ponderadas com base na distância à observação central (Bagheri, Holt, & Benwell, 2009; D. C. Wheeler & Paez, 2002).

Alguns especialistas duvidam da validade deste método argumentando que a multicolinearidade entre as variáveis independentes pode enviesar os resultados (Páez et al. 2011). Estudos por simulação mostraram que, além deste potencial problema, o GWR pode produzir um padrão nos coeficientes de regressão mesmo quando este não existe (D. Wheeler & Tiefelsdorf, 2005). Ou seja, mesmo quando o conjunto de dados é completamente aleatório, é possível que os coeficientes da regressão GWR apresentem algum padrão espacial (Fotheringham et al., 2002, p. 83). Em particular, Griffith (2008) refere que os coeficientes podem apresentar uma forte autocorrelação espacial positiva.

As técnicas de diagnóstico dos modelos GWR são limitadas porque os testes estatísticos clássicos não são válidos, dado que as observações não são independentes. A abordagem recomendada consiste então em obter em primeiro lugar um modelo OLS viável em termos dos seus pressupostos, exceto no que se refere à heterocedasticidade e autocorrelação espacial dos resíduos.

A maioria dos investigadores concorda que o método GWR pode ser utilizado com confiança como uma técnica exploratória para compreender de que forma a associação entre variáveis varia localmente (Bäumler, 2013; Leong & Yue, 2017; Matthews & Yang, 2012; Ogneva-Himmelberger, Rakshit, & Pearsall, 2013). De facto, o método tem sido amplamente usado nas mais diversas áreas, como por exemplo na modelação do desemprego (Lewandowska-Gwarda, 2018), na criminologia (Nezami & Khoramshahi, 2016), no planeamento urbano (Ye et al., 2017) e na modelação dos preços da habitação (Huang, Wu, & Barry, 2010).

A área da saúde não é exceção, contando com investigações nos mais variados tipos de doenças ou problemas de saúde, como por exemplo: cancro (Buck, 2016), HIV (Nakaya, Nakase, & Osaka, 2005), malária (Ndiath et al., 2015), doenças cardiovasculares e mortalidade por acidente vascular cerebral (Ford & Highfield, 2016; Odoi & Busingye, 2014a), asma (Bäumler, 2013), obesidade infantil (Chalkias et al., 2013; Shahid & Bertazzon, 2015), hábitos alimentares (Clary et al., 2016), e acessibilidade aos cuidados de saúde (Bascuñán & Quezada, 2016; Cosmber, Brunson, & Radburn, 2011). Em Portugal também se destaca uma investigação, ao nível dos municípios, relacionando fatores socioeconómicos com acidente vascular cerebral e que tem em parte da sua metodologia o modelo GWR Poisson (Oliveira, Cabral, Mendes, Martins, & Cabral, 2015). Mais recentemente foi publicado um estudo português sobre a gravidez na adolescência e os fatores socioeconómicos que a influenciam (David & Cabral, 2019).

Dos estudos que relacionam saúde e hábitos alimentares utilizando o método GWR, a grande maioria utiliza como variável dependente a percentagem ou taxas de obesidade, tanto infantil como na população em geral.

Na Grécia, foi desenvolvido um estudo com este método, no qual relacionam a obesidade infantil, de 18296 alunos de Atenas, com indicadores do estado socioeconómico (nível de escolaridade e rendimento familiar) e das condições de vida (densidade populacional, ocupação dos solos e áreas recreativas). No modelo global foi possível concluir que o nível de escolaridade é o indicador com mais influência no modelo. Por outro lado, na análise local os resultados são mais abrangentes, concluindo-se que zonas com baixo nível de escolaridade, grande densidade populacional, baixos rendimentos e com pouca abrangência de espaços verde e recreativos são propensas para a existência de crianças obesas (Chalkias et al., 2013).

Outro estudo, que também utiliza obesidade infantil como variável dependente, elaborado numa cidade do Canadá, considera como fatores explicativos o estado socioeconómico, a proximidade aos restaurantes de *fast food*, proximidade a parques e ainda como indicadores a acessibilidade pedonal e extensão dos seus percursos (Shahid & Bertazzon, 2015). Estes autores salientam a importância da utilização do GWR como ferramenta para identificar as regiões mais afetadas, pois com este método foi possível caracterizar os bairros de forma mais detalhada, permitindo assim identificar quais as medidas mais específicas para cada um deles.

No Reino Unido relacionam o ambiente alimentar com o consumo de fruta e vegetais, tendo como ideia principal medir a sua exposição sugerindo uma relação local não estacionária entre o e os hábitos alimentares/dieta (Clary et al., 2016).

Nos EUA (Califórnia) avaliam a relação entre a saúde dos jovens e o acesso a estabelecimentos de restauração, nomeadamente de *fast food*, e ainda lojas de conveniência. Os resultados indicam que tem um efeito moderado (West, 2014).

3. METODOLOGIA

A metodologia utilizada baseou-se em cinco fases, conforme esquematizado na **Figura 3.1**. Após a análise exploratória detalhada dos dados, a etapa seguinte consistiu na estimação de diferentes modelos de regressão linear múltipla pelo método dos Mínimos Quadrados Ordinários (OLS – *Ordinary Least Squares*), e no diagnóstico dos pressupostos do OLS de cada um deles. Estas etapas constituíram um processo iterativo que consistiu em remover ou adicionar variáveis explicativas aos modelos de regressão OLS, em função dos resultados do diagnóstico. Na quarta etapa, procedeu-se à estimação do modelo GWR com as variáveis explicativas encontradas no melhor modelo OLS, tendo em conta o R^2 ajustado e o Critério da Informação de Akaike Corrigido (AICc). A existência de multicolinearidade local e autocorrelação espacial dos resíduos do modelo GWR foram então investigados.

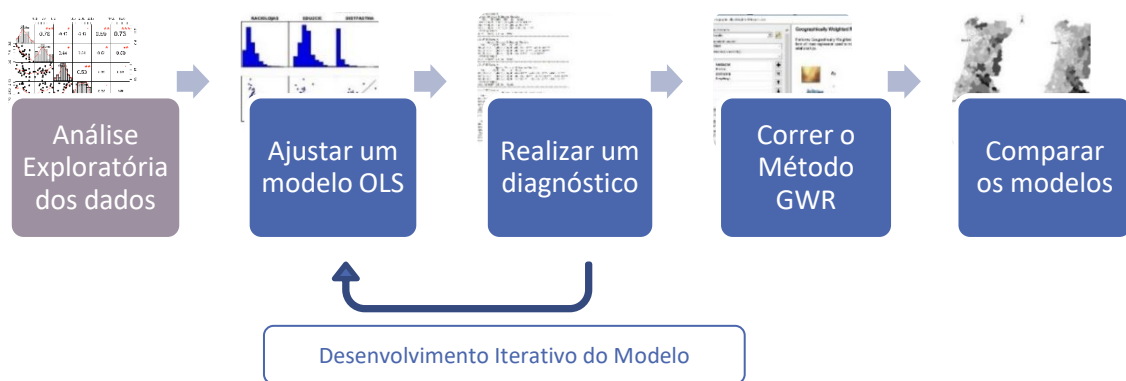


Figura 3.1 Processo de modelação com GWR

Os métodos e testes estatísticos utilizados na análise exploratória de dados e para diagnosticar os modelos OLS e GWR estão detalhados na Figura 3.2. Nas secções seguintes, descreve-se a região de estudo e variáveis analisadas. É também apresentada de forma resumida a metodologia associada a cada umas das três principais fases da análise de dados (Figura 3.2).

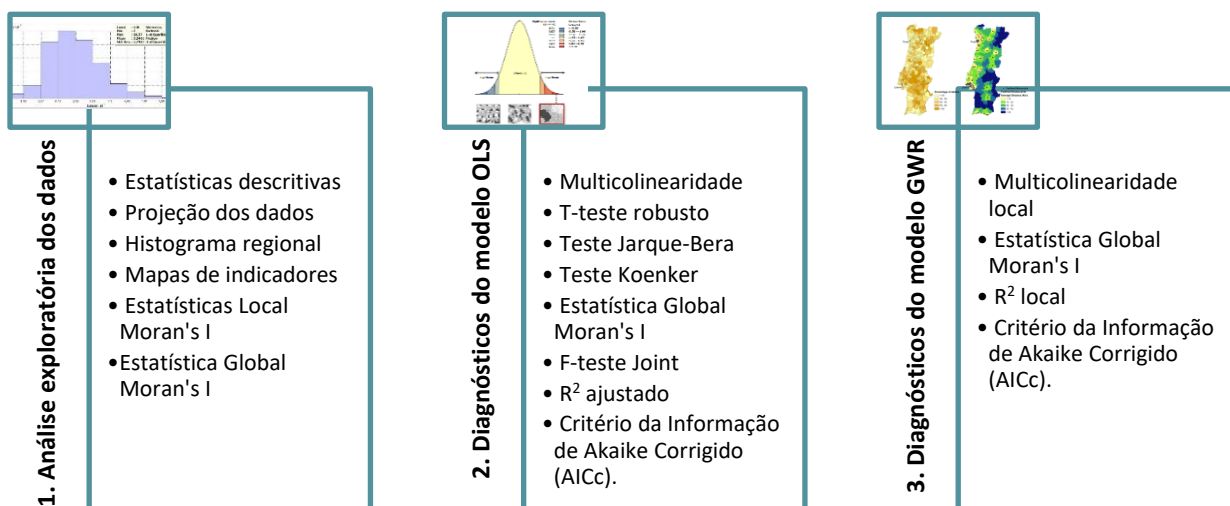


Figura 3.2 Enquadramento metodológico

3.1. REGIÃO DE ESTUDO E DADOS

A região de estudo deste trabalho corresponde ao território de Portugal continental que, à data dos Censos 2011, era composto por 10.047.621 habitantes (INE, 2011b). Sendo que as análises recaíram sobre as regiões administrativas correspondentes aos municípios do Continente, perfazendo um total de 278. Cada um dos municípios corresponde a uma área geográfica bem definida de tamanho variável (polígono irregular). Para o estudo utilizou-se a Carta Administrativa Oficial de Portugal – Versão de 2011, com a referência geográfica [Continente: ETRS89 / PT-TM06] (DGTerritório, 2011).

A população por município apresenta alguma heterogeneidade, variando entre 1822 habitantes em Barrancos e 542917 em Lisboa, com maior número nas Áreas Metropolitanas de Lisboa e do Porto e ainda no litoral a Norte de Lisboa (Figura 3.3) (INE, 2011b).

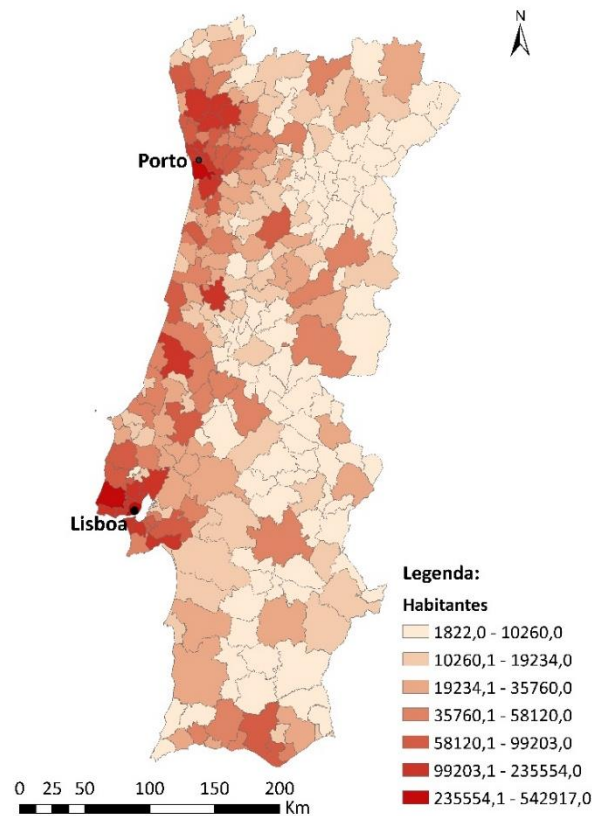


Figura 3.3 Habitantes de Portugal continental por município.

Com o objetivo de tornar o estudo mais abrangente e coerente, sempre que possível, utilizaram-se dados dos Censos 2011. Quanto aos dados não provenientes dos censos, teve-se o cuidado de serem referentes ao mesmo ano e disponíveis para todos os municípios de Portugal continental. Algumas das variáveis foram utilizadas na sua forma original, ou seja, retiradas diretamente da fonte sem tratamento adicional. Outras apresentaram-se em rácios já utilizados na literatura aplicados à situação em estudo. Destaca-se ainda que a variável “Distância aos restaurantes *fast food*” foi construída de raiz utilizando a georreferenciação dos mesmos e as distâncias médias dentro de cada município. Estas variáveis serão apresentadas com mais detalhe nos próximos pontos. Para tornar o estudo o mais completo possível foi feita uma pesquisa exaustiva de dados com o objetivo de encontrar o máximo número de potenciais variáveis explicativas (i.e. independentes).

3.1.1. Variável dependente

A variável de estudo escolhida foi a taxa de mortalidade utilizando causas de morte relacionadas com a alimentação (doenças do aparelho digestivo, diabetes e doenças do aparelho circulatório) (INE, 2011a). O valor da variável final corresponde à soma das três causas de óbitos para cada município, a dividir pelo número de população residente para cada município e multiplicado por 1000, para o ano 2011.

Em alguns municípios, por questões de confidencialidade, os dados não se encontravam disponíveis. Assim sendo, estes valores foram calculados com base nas diferenças dentro de cada região, apresentadas no Anexo “Lista de municípios com valores em falta de Portugal”.

3.1.2. Variáveis independentes

Na **Tabela 3.1** estão expostas as variáveis recolhidas de acordo com a literatura e com a sua disponibilidade nas bases de dados de livre acesso (indicadas na coluna mais à direita).

Tabela 3.1 Descrição das variáveis utilizadas e respetiva descrição e origem

Categoria	Nome da variável	Descrição	Unidades	Fonte/Origem
Socioeconómicos	SalDifer	Diferença entre o salário mínimo nacional e a remuneração base média mensal dos trabalhadores por conta de outrem	€	INE
	Desemprego	Taxa de desemprego segundo os Censos: total e por sexo	%	
	SemEdu	Sem nível de escolaridade	%	
	Edu1ciclo	Ensino básico 1º ciclo		
	Edu2ciclo	Ensino básico 2º ciclo		
	Edu3ciclo	Ensino básico 3º ciclo		
	EduSec	Ensino secundário		
	EduMedio	Ensino médio		
	EduSup	Ensino superior		
	ResidHab	Resíduos Urbanos indiferenciados por Habitante	%	
PropPCompra	Proporção de poder de compra	Rácio		

	DimFamilia	Dimensão média das famílias segundo os Censos	Nº médio	
	TxAtividade	Taxa de atividade segundo os Censos: total e por sexo	%	
Acesso à Alimentação	RestHab	Restaurantes por habitantes	Nº de res/habitantes	INE*
	RestKM	Restaurantes por Km ²	Nº restaurantes / Km ²	
	RacioLojas	Rácio entre o nº supermercados e o nº lojas de conveniência	Rácio	INE**
	RacioRestaurantes	Rácio entre nº restaurantes <i>fast food</i> e nº restaurantes com refeições completas	Rácio	
	FastFoodDist	Distância média de cada município a restaurantes <i>fast food</i>	m	-
	FastFHabit	Quociente entre a distância média e a nr. Habitantes	Km/hab.	-
	FastFDens	Quociente entre a distância média e a densidade	m/densidade	-
	WaterQual	Água com qualidade	%	INE
Cuidados de Saúde	SaudePrimaria	Utilização dos Cuidados de Saúde Primários	‰	GeoHealths
	ProxSPrimaria	Proximidade dos Cuidados de Saúde Primários	Minutos	
	RSaudePrimaria	Capacidade de resposta dos Cuidados de Saúde Primários	‰	
	ProxHospitalar	Proximidade dos Hospitais	Minutos	
	RespHospital	Capacidade de resposta dos Hospitais do SNS	‰	
	AcessoFarma	Acesso a cuidados farmacêuticos	‰	

INE* - Variáveis construídas com dados do INE disponíveis no *site*.

INE** - Variáveis construídas com dados do INE não disponíveis no *site* (pedidos por email).

De seguida estão detalhadas as variáveis calculadas para este trabalho, tal como o seu modo de interpretação, se aplicável.

Acesso à Alimentação

No que diz respeito a hábitos alimentares, a informação disponível ao nível dos municípios é bastante reduzida. Portanto surgiu a necessidade de encontrar alternativas para medir este fator. A existência de alimentos de *fast food* e alimentos pré-preparados que são por norma altamente calóricos e mais baratos, estando assim acessíveis à maioria das classes socioeconómicas, condicionam o estado de saúde das pessoas que os consomem (Schlosser 2001), esta foi a premissa para os indicadores seguintes.

- **Rácios:**

Criaram-se dois indicadores baseados num estudo desenvolvido nos EUA sobre obesidade (Chi et al., 2013), onde eram utilizadas variáveis semelhantes a estas para avaliar o acesso à alimentação utilizando estimativas do rácio entre as lojas de conveniência e os supermercados, e ainda o rácio entres os restaurantes de *fast food* e os restaurantes com refeições completas (Mehta & Chang, 2008).

Solicitou-se ao INE a informação do número de empresas pela CAE (Classificação das Atividade Económica) de restauração e retalho alimentar. Os dados enviados estavam organizados por freguesia. Para o tratamento destes dados procedeu-se ao agrupamento por município e por CAE, como indicado no Anexo (INE, 2007).

Rácio entre supermercados e lojas de conveniência

RacioLojas	Rácio entre o nº de supermercados e o nº de lojas de conveniência	>1	Supermercado > Conveniência
		=1	Supermercado = Conveniência
		<1	Supermercado < Conveniência

Nota: na categoria supermercados também estão englobadas as lojas de alimentos especializados (ex. lojas especializadas em frutas e legumes, etc.) (consultar Anexo).

Esta variável parte da premissa de que a quantidade de alimentos saudáveis é maior nos supermercados do que nas lojas de conveniência.

Rácio entre restaurantes *fast food* e restaurantes com refeições completas

À semelhança da variável anterior, agruparam-se os restaurantes de acordo com o tipo de serviço de restauração (Mehta & Chang, 2008).

RacioRest	Rácio entre o nº de restaurantes de <i>fast food</i> e o nº de restaurantes com refeições completas	>1	<i>fast food</i> > Restaurante
		=1	<i>fast food</i> = Restaurante
		<1	<i>fast food</i> < Restaurante

Nota: para mais detalhes consultar Anexo.

- **Distância média aos restaurantes *fast food* por habitante**

A georreferenciação dos restaurantes de *fast food* teve como base as coordenadas GPS dos restaurantes de uma das cadeias de *fast food* mais influentes no país. Com base nessa informação utilizou-se o ArcGIS®, para calcular as distâncias médias entre qualquer ponto do mapa e o restaurante mais próximo. Seguidamente calculou-se a média das distâncias por município (polígono), baseada na informação anterior. Por fim, para obter a variável final, procedeu-se ao quociente entre a distâncias média e o número de habitantes de cada município (km/hab.).

- **Densidade de restaurantes (número de restaurantes por km²):**

Quociente entre o número total de restaurantes de cada município pela área correspondente.

Cuidados de Saúde

A informação destas variáveis foi retirada do *site GeoHealths* (http://www.uc.pt/fluc/qigs/GeoHealthS/INES/cuidados_saude), nomeadamente do índice INES, onde estão presentes indicadores dos cuidados de saúde em três vertentes: **Acesso**, **Proximidade** e **Capacidade de Resposta** de entre essas categorias. A classificação destes indicadores em termos práticos varia de pior para melhor, sendo um valor mais baixo pior que um mais elevado (Santana, 2015).

3.2. ANÁLISE EXPLORATÓRIA

Com a utilização de estatísticas descritivas, matrizes de correlação, projeção dos dados e histogramas regionais é possível descrever e visualizar a distribuição espacial e analisar o comportamento das variáveis a escolher para o modelo. Posteriormente recorreram-se a

métodos gráficos e visuais, Estatísticas Local Moran's I e Estatística Global Moran's I para identificar os *outliers* espaciais e descobrir os padrões espaciais como, por exemplo, *Clusters*.

3.3. MÉTODOS UTILIZADOS NA MODELAÇÃO

3.3.1. Modelação Global (OLS)

A primeira fase da modelação tem como finalidade identificar as variáveis que estão relacionadas globalmente com a variável dependente do estudo. O modelo de regressão linear múltipla foi então estimado por OLS.

Para garantir que esta modelação está corretamente especificada é essencial que o modelo satisfaça os pressupostos que se seguem (Hayashi, 2000). Relativamente aos resíduos, a sua média deve ser zero, a sua variância deve ser constante (homocedasticidade) e ainda devem seguir a distribuição normal. Quanto às variáveis, as observações devem ser independentes entre si e não pode existir relação de multicolinearidade entre elas. Caso alguns destes pressupostos falhem, então o modelo não é BLUE (*Best Linear Unbiased Estimator*). Dependendo do pressuposto que falhe, existem outras consequências associadas à qualidade da modelação.

O diagnóstico desenvolvido para este método permite avaliar os três aspetos principais da modelação, nomeadamente as **estimativas dos parâmetros**, os **resíduos** e o **ajustamento global do modelo**.

Para que as estimativas dos parâmetros sejam adequadas devem ser garantidos os pressupostos do modelo. No que se refere à **multicolinearidade**, os coeficientes com valores de VIF (Variance Inflation Factor) superiores a 10 são geralmente indicados como indicadores de existência de multicolinearidade (O'Brien, 2007). As variáveis independentes com valor de VIF superior ou igual a 7,5 foram eliminadas dos modelos. Todas as variáveis independentes devem ser estatisticamente significativas, portanto, excluíram-se todas as variáveis cujo valor de p (*p-value*) do teste- t para o respetivo parâmetro era superior a 0,05. Os **resíduos** devem estar normalmente distribuídos e, para verificar esta situação, utiliza-se o teste de Jarque-Bera (JB), no qual o valor de p deve ser superior ao nível de significância (0,05). Foi também utilizado o teste *Koenker (BP)* para averiguar a homocedasticidade dos resíduos (caso em que o valor p é superior a 0,05). Os resíduos devem ser independentes, e, portanto, não devem estar agrupados (*clustered*) em termos de localização, nem apresentar autocorrelação espacial. A avaliação da

primeira situação (estacionaridade) é feita com base no teste Local Moran's I, e a segunda (autocorrelação espacial) com base no teste Global Moran's I. Nas situações em que os resíduos são não-estacionários o modelo local GWR é mais adequado.

O **ajustamento global do modelo** é avaliado utilizando o teste de *Wald* (valor de $p < 0,05$ indica evidência de um modelo bem ajustado aos dados), utilizando o R^2 ajustado que indica a variabilidade da variável dependente (Y) explicada pelo modelo, e utilizando o AICc que fornece informação sobre o desempenho do modelo (quanto menor o valor, melhor o desempenho do modelo).

3.3.2. Modelação Local (GWR)

Sendo o método GWR uma ferramenta de modelação local, para cada polígono há uma regressão associada (D. C. Wheeler & Paez, 2002). O método GWR foi utilizado para cada polígono dos municípios. As variáveis foram as pré-selecionadas pela regressão global. Para um bom funcionamento da modelação, para além da identificação das variáveis, é importante tomar decisões sobre alguns parâmetros da modelação. Em primeiro lugar, deve escolher-se o tipo de Kernel, o qual especifica a função usada para determinar os pesos associados às observações vizinhas usadas em cada regressão local. Neste modelo utilizou-se um Kernel adaptativo do tipo gaussiano, pois é mais adequado para dados poligonais. Em segundo lugar, é necessário especificar como é que o parâmetro Bandwidth do Kernel é determinado em cada regressão local. Utilizou-se então o AICc como critério, pois o software estima automaticamente diferentes modelos locais e determina a distância ótima ou o número de vizinhos ótimo. Para mais detalhes, veja-se por exemplo Fotheringham, Brunson, & Charlton (2002, pp. 44-45).

Os resultados da modelação local por GWR foram diagnosticados usando:

- Os R^2 locais avaliam o ajustamento do modelo a cada município.
- Os valores dos coeficientes exprimem a influência de cada variável para cada município, e os erros padrão avaliam a precisão dos respetivos coeficientes. Note-se que as estimativas com menores erros padrão correspondem a estimativas com maior confiança, enquanto os valores elevados podem indicar que os respetivos coeficientes não são significativos.
- O *Condition Number* é uma ferramenta de diagnóstico da multicolinearidade do modelo GWR (Chris Brunson, Charlton, & Harris, 2012), devendo tomar valores inferiores a 30, para garantir que não existe colinearidade nas regressões locais.

4. RESULTADOS E DISCUSSÃO

Este capítulo apresenta e discute os resultados obtidos nas fases de análise exploratória e de modelação. Na secção 4.1 são analisadas as taxas de mortalidade relacionadas com hábitos alimentares (variável dependente) nos municípios de Portugal continental, bem como as potenciais variáveis explicativas deste fenómeno. A secção 4.2 é dedicada aos resultados da modelação global (OLS), e a secção 4.3 apresenta e discute os resultados do melhor modelo local (GWR) obtido.

4.1. ANÁLISE EXPLORATÓRIA

4.1.1. Variável dependente

No ano de 2011, as causas de óbitos consideradas neste estudo correspondem a cerca de 39,4% dos óbitos em Portugal continental (**Tabela 4.1**), em que as doenças do aparelho circulatório têm a maior influência (INE, 2011a). Verifica-se que a grande maioria dos municípios junto ao litoral apresentam os valores mais baixos de mortalidade (**Figura 4.1**), não excedendo 5 óbitos por cada 1000 habitantes. Em contraste, nota-se um aumento gradual do valor da taxa à medida que se analisam os municípios mais interiores, este comportamento é mais nítido na região centro e no Alentejo.

Tabela 4.1 Percentagens de óbitos para cada uma das causas utilizadas na análise, para Portugal continental, no ano 2011.

Doenças do aparelho circulatório	Diabetes	Doenças do aparelho digestivo	Total
30,7%	4,3%	4,4%	39,4%

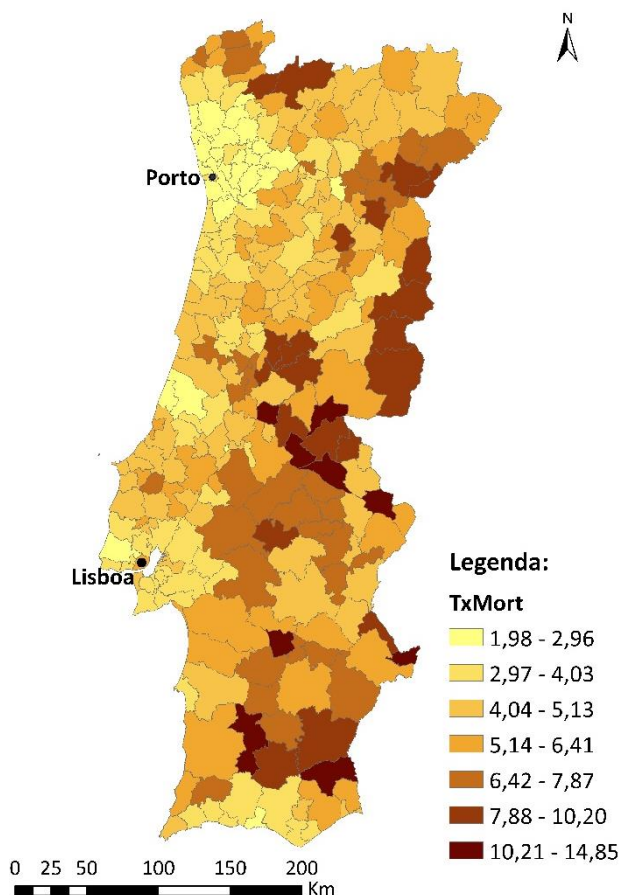


Figura 4.1. Projeção das taxas de mortalidade relacionadas com hábitos alimentares, em 2011.

A taxa de mortalidade média das causas em estudo corresponde a 5,21‰ com um desvio padrão de 2,17 (**Tabela 4.2**). Os valores máximos encontram-se no município de Arronches (14,85‰) e de Barrancos (13,09‰) por outro lado os valores mínimos estão presentes em Braga (1,98‰), Paredes (2,05‰) e Vizela (2,06‰). A distribuição dos valores é assimétrica positiva (**Figura 4.2**).

Tabela 4.2 Estatísticas descritivas da variável dependente [taxa de mortalidade (‰)]

Média	5,21
Erro Padrão	0,13
Mediana	4,77
Desvio Padrão	2,17
Variância Amostral	4,70
Amplitude	12,87
Mínimo	1,98
Máximo	14,85

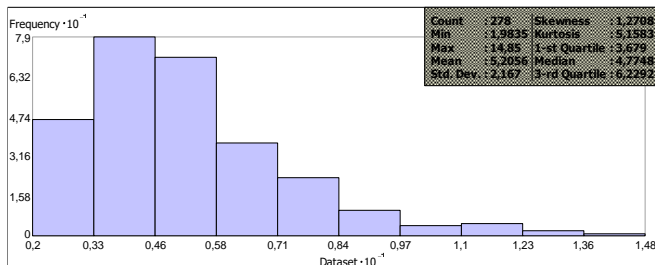


Figura 4.2 Histograma regional - Taxa de mortalidade

Para analisar o comportamento espacial da variável dependente, com base no comportamento nos polígonos vizinhos, recorreu-se à estatística *Local Moran's I* (**Figura 4.3**) para identificar os *outliers espaciais* (valores elevados rodeados por valores baixos – HL, ou a situação contrária – LH) estatisticamente significativos e os municípios com comportamentos de *cluster* (valores elevados rodeados por valores elevados – HH, ou valores baixos rodeados por valores baixos – LL).

Ao calcular esta estatística para cada município, é indicada como a autocorrelação espacial da mortalidade varia na região de estudo. Existem comportamentos de *cluster* com correlação positiva de valores baixos (■ LL) nas Áreas Metropolitanas de Lisboa e Porto e municípios adjacentes; e com correlação positiva de valores elevados (■ HH) no interior do país (do Baixo Alentejo até parte de Trás-os-Montes). *Outliers espaciais* com valores elevados de mortalidade (■ HL) em municípios próximos de *clusters* LL e com valores reduzidos de mortalidade (■ LH) aparecem maioritariamente em municípios do interior.

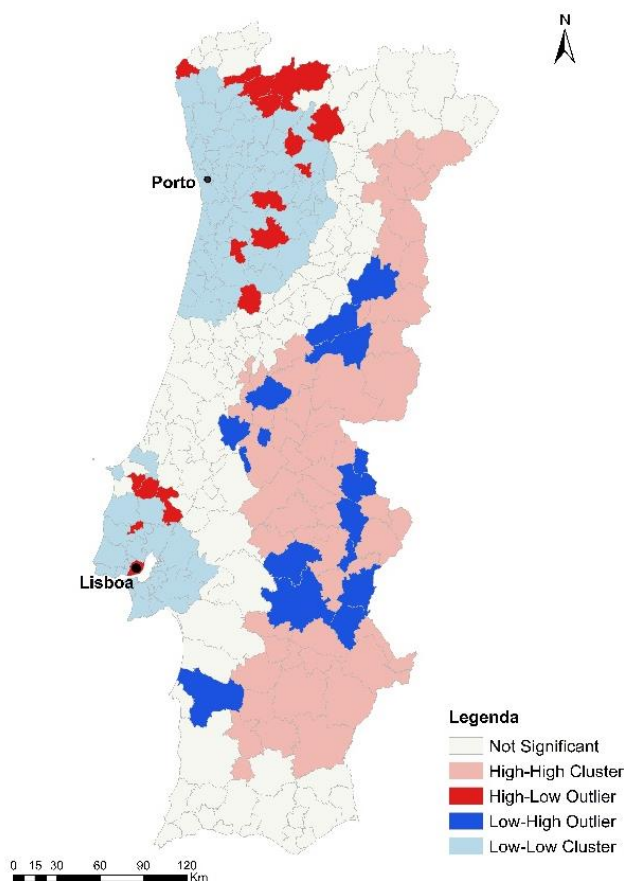


Figura 4.3 Output do método Local Anselin's Moran's

4.1.2. Variáveis independentes

De seguida será feita uma breve análise das projeções dos dados das variáveis seleccionadas, verificando os principais padrões.

A variável correspondente à “**Percentagem de população com educação de 2º Ciclo**” varia de 7,8%, em Lisboa, a 24%, em Barrancos. Observa-se maiores percentagens nos municípios das áreas circundantes ao Porto, por outro lado as menores percentagens estão em Lisboa, Oeiras, Cascais, no norte interior junto à fronteira com Espanha e no sul do país destacam-se, Almodôvar, Monchique e Faro (**Figura 4.4**).

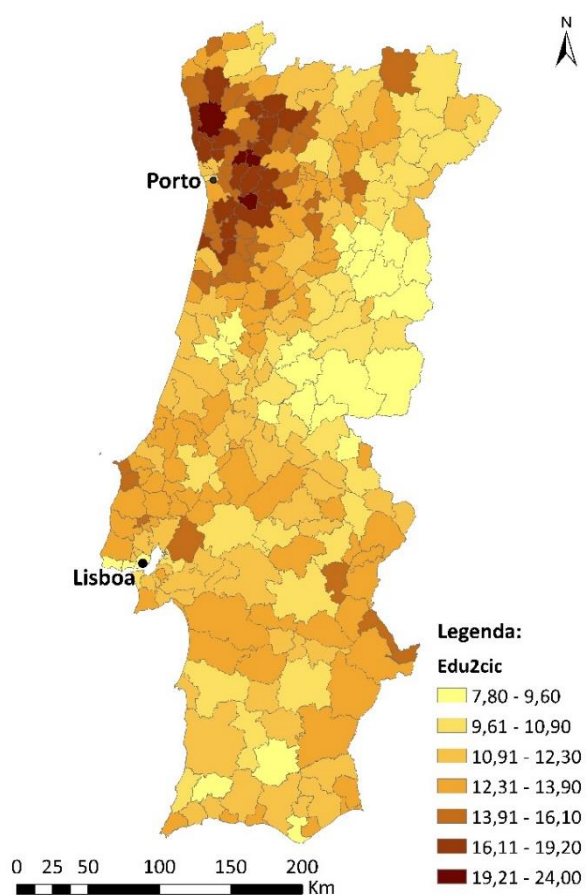


Figura 4.4. Projeção da percentagem de população com nível de educação de 2º Ciclo.

Para a variável “Rácio entre supermercados e lojas de conveniência” (Figura 4.5), a projeção dos dados apenas foi apresentada de acordo com a interpretação indicada na metodologia (<1; =1; >1). Verifica-se uma grande influência dos supermercados em quase todos os municípios do litoral e alguns do Norte junto à fronteira com Espanha, por outro lado os valores inferiores a um são mais predominantes no interior.

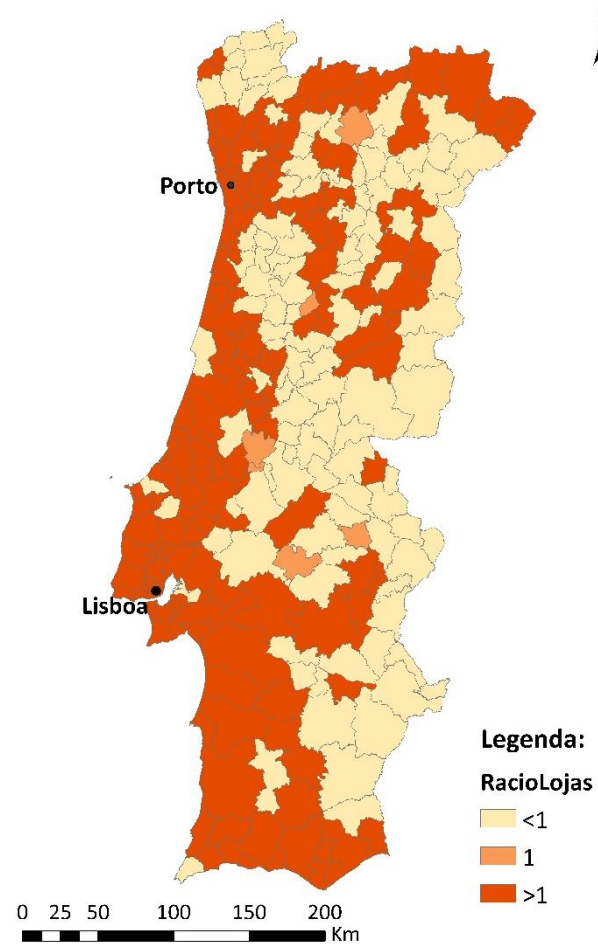


Figura 4.5 Projeção do rácio entre supermercados e lojas de conveniência

Quanto à “Resposta hospitalar”, os municípios vizinhos exibem valores muito semelhantes porque a sua área é abrangida pelos mesmos hospitais de referência [esta distribuição está implícita na construção da variável (Santana, 2015)]. Assim pode observar-se um melhor desempenho junto da Grande Lisboa bem como na região de Coimbra. Existe uma menor resposta hospitalar em regiões como o litoral alentejano e alguns municípios mais rurais pertencentes aos distritos do Porto e Lisboa (**Figura 4.6**).

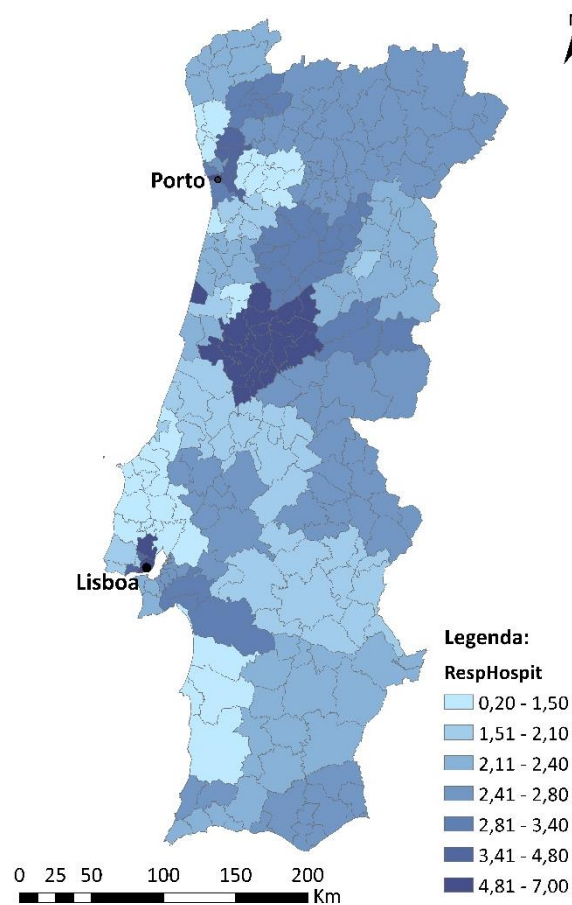


Figura 4.6 Projeção da resposta hospitalar

A variável “Distância média a restaurantes de *fast food* por habitante” apresenta valores entre 0,002 em Lisboa e 40,19 em Barrancos, sendo a distância média por habitante de 3,14 km. Na **Figura 4.7** pode observar-se uma maior distância por habitante no interior do país, nomeadamente no Alto Alentejo e no interior do distrito de Bragança. Por outro lado, as menores distâncias por habitante estão presentes maioritariamente junto à costa, a partir de Setúbal em direção ao norte, e ainda nos municípios da costa sul do Algarve.

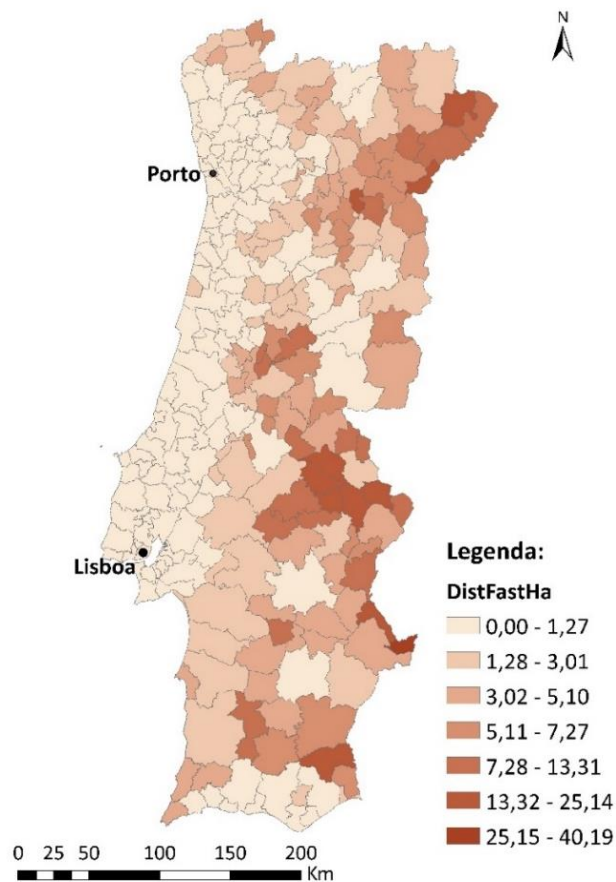


Figura 4.7 Projeção da distância média a restaurantes de *fast food* por habitante (Km/hab)

4.2. MODELAÇÃO GLOBAL (OLS)

O modelo foi alvo de diversas alterações ao longo do processo de investigação com o objetivo de melhorar a regressão. Ou seja, foram testados diversos modelos OLS com diferentes variáveis independentes, os quais foram alvo dos vários procedimentos de diagnóstico descritos no capítulo da Metodologia.

Nenhum dos modelos OLS construídos durante este trabalho passaram em todos os testes de diagnóstico, como seria de esperar. Assim, tomou-se como critério de seleção o modelo que apresentasse todos os parâmetros estatisticamente significativos (teste-*t*), sem multicolinearidade ($VIF < 7,5$), e globalmente bem ajustado (teste de Wald; maior valor de R^2 Ajustado; menor valor de AICc). Os resultados do melhor modelo obtido são apresentados na **Tabela 4.3** e discutidos em seguida.

As variáveis incluídas no modelo obtido foram “Educação 2º Ciclo (%)”, “Rácio supermercados e lojas de conveniência”, “Resposta hospitalar” e “Distância média a restaurantes *fast food* por habitante (km/hab.)”. Todas elas são estatisticamente significativas para o modelo (valor $p < 0,05$). O R^2 Ajustado indica que as variáveis independentes do modelo explicam 56,8% da variabilidade das taxas de mortalidade. O AICc do modelo corresponde ao valor 993,10.

Ao examinar os resíduos do modelo verificou-se que estes apresentam autocorrelação espacial (Estatística Global Moran's $I = 0,11$; estatística normalizada $Z(I) = 3,91$; valor $p = 0,000092$). Este diagnóstico demonstra a transgressão do pressuposto de independência dos resíduos do método OLS. Como seria de esperar, há também evidência de que os resíduos são heterocedásticos [teste Koenker (BP)], e, portanto, verifica-se a sua não estacionaridade. Por fim, o teste de Jarque-Bera rejeitou a hipótese nula da normalidade dos resíduos, o que implica que o modelo poderá não estar corretamente especificado (por faltarem variáveis explicativas relevantes, ou pelo facto de um modelo linear global não ser adequado). Segue-se a modelação com método GWR.

Tabela 4.3 Resultado do modelo OLS

Variável	Coefficiente	Estatística t	Valor p	VIF
Intercepção	9,45	15,82	0,000000	-
Educação 2º Ciclo (%)	-0,30	-8,40	0,000000	1,107969
Rácio supermercados e lojas de conveniência	-0,54	-4,83	0,000003	1,078835
Resposta hospitalar (‰)	-0,19	-2,37	0,018518	1,079621
Distância média a restaurantes <i>Fast food</i> / habitante (km/hab.)	0,24	13,00	0,000000	1,108712

Testes estatísticos diagnóstico	Estatística de teste	Valor p
R ² Múltiplo	0,574	-
R ² Ajustado	0,568	-
AICc	993,10	-
Joint F-statistic	91,90	0,003
Joint Wald Statistic	200,60	0,000
Koenker (BP) Statistic	57,95	0,000
Jarque-Bera Statistic	11,51	0,003
Global Moran's I	3,91	0,000092

4.3. MODELO GWR

Para a implementação da regressão GWR, tal como já foi referido anteriormente, utilizaram-se as mesmas variáveis do modelo OLS descrito na secção anterior. O desempenho geral do modelo está sistematizado na **Tabela 4.4**. As variáveis “Educação 2º Ciclo” e “Resposta hospitalar” apresentam apenas coeficientes locais negativos, enquanto os coeficientes de “Distância média a restaurantes *Fast food* por habitante” são todos negativos. A mediana dos coeficientes do “Rácio supermercados e lojas de conveniência” é negativa. Os resultados e discussão da distribuição espacial e da variabilidade das estimativas locais dos parâmetros para as variáveis em estudo são apresentadas e discutidas na subsecção seguinte.

No modelo GWR obtido, o valor do R² Ajustado é de 0,612 e o AICc igual a 974,91. Os valores de AICc obtidos por (Oliveira et al., 2015) para investigar o risco de acidentes vasculares cerebrais em Portugal não são comparáveis com os do modelo GWR deste estudo, porque aqueles autores usaram modelos GWR Poisson (e não modelos lineares). Num estudo sobre obesidade infantil na Grécia, foi obtido um valor de 0,27 para o R² Ajustado e um AICc igual a 528.49 no modelo GWR (Chalkias et al., 2013). Um modelo GWR para a prevalência de obesidade infantil numa cidade do Canadá apresentou valores de R² locais entre 0,09 e 0,62, não tendo sido apresentados os valores do R² Ajustado e do AICc (Shahid & Bertazzon, 2015).

Tabela 4.4 Síntese dos resultados do GWR

Variável	Estatísticas dos valores dos coeficientes locais			
	Min.	Mediana Q50	Max.	Desvio padrão
Interceção	5,314	8,241	14,178	2,159
Educação 2º Ciclo (%)	-0,696	-0,228	-0,005	0,151
Rácio supermercados e lojas de conveniência	-0,887	-0,571	0,025	0,195
Resposta hospitalar	-0,466	-0,167	-0,042	0,088
Distância média a restaurantes fast food / habitante (km/hab.)	0,194	0,312	0,566	0,089

R^2 Múltiplo	0,648
R^2 Ajustado	0,612
AICc	974,91

Como todos os municípios têm valores do *Condition Number* inferiores a 30 (**Figura 4.8**), pode afirmar-se que as variáveis do modelo não apresentam multicolinearidade, o que torna o estudo confiável para todos os municípios.

No entanto, os municípios do sul do país e da Área Metropolitana de Lisboa têm valores a variar entre 27,945 e 29,994, muito próximo do limite estipulado (a vermelho no mapa da **Figura 4.8**). Destacam-se os municípios alentejanos de Barrancos, Reguengos de Monsaraz, Mourão, Moura, Alandroal, Évora, Portel, Vidigueira, Serpa, Cuba, Viana do Alentejo, Alvito e Beja, apresentando valores na ordem de 29. Por outro lado, o município da Guarda tem o menor valor, com 15,91.

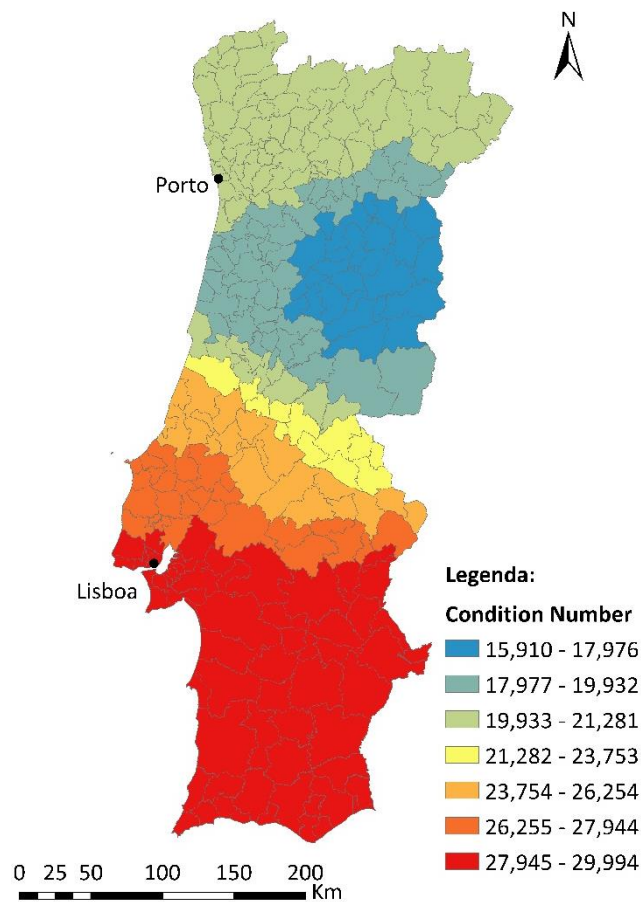


Figura 4.8 Diagnóstico de multicolinearidade do modelo GWR

R² Local

Os valores dos R² locais variam entre 42,2% e 64,1% (**Figura 4.9**), podendo considerar-se valores elevados o que demonstra um bom ajuste do modelo aos dados, conforme discutido anteriormente.

Os municípios com melhores valores (superiores a 61%) pertencem aos distritos de Aveiro e Coimbra. Contrariamente estão os municípios pertencentes ao distrito de Lisboa e alguns do Alto Alentejo (entre 42,3% e 45,5%), embora estes sejam valores relativamente altos.

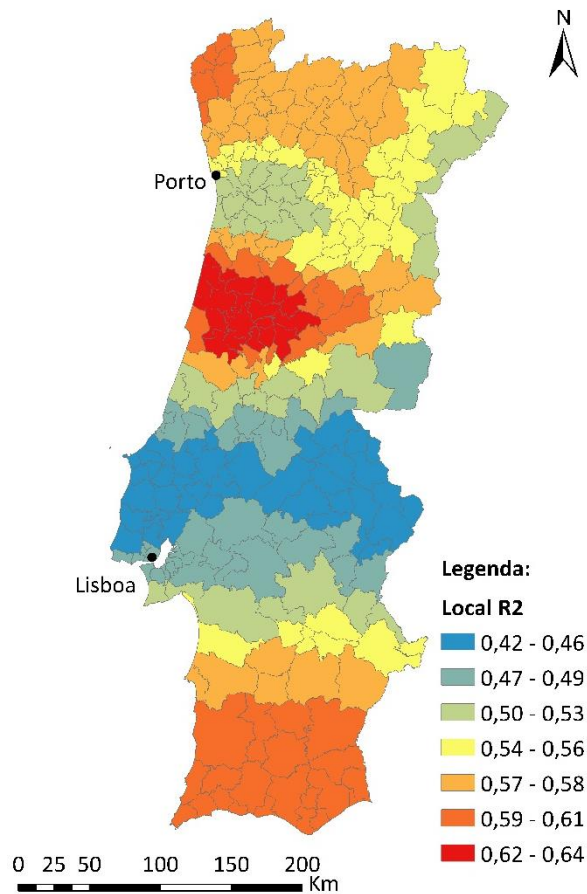


Figura 4.9 R² Local do modelo GWR

Resíduos do modelo

Para avaliar a autocorrelação espacial dos resíduos após a modelação GWR, utilizou-se novamente o Global Moran's I (Índice de Moran = 0,05; Z(I) = 1,88; p-value = 0,06), que revelou não haver evidência de autocorrelação espacial para o nível de significância de 0,05. Contudo, o Local Moran's I apresenta alguns *clusters* de valores elevados (■ HH) no interior e sul do país, especialmente em grande parte do Baixo Alentejo, e de valores baixos (■ LL) em Fronteira, Golegã, Portel, Santo Tirso, Sousel, Valongo, Vila Nova da Barquinha e Vila Nova de Famalicão (Figura 4.10). Estes resultados indicam que poderá haver problemas de especificação do modelo nesses municípios.

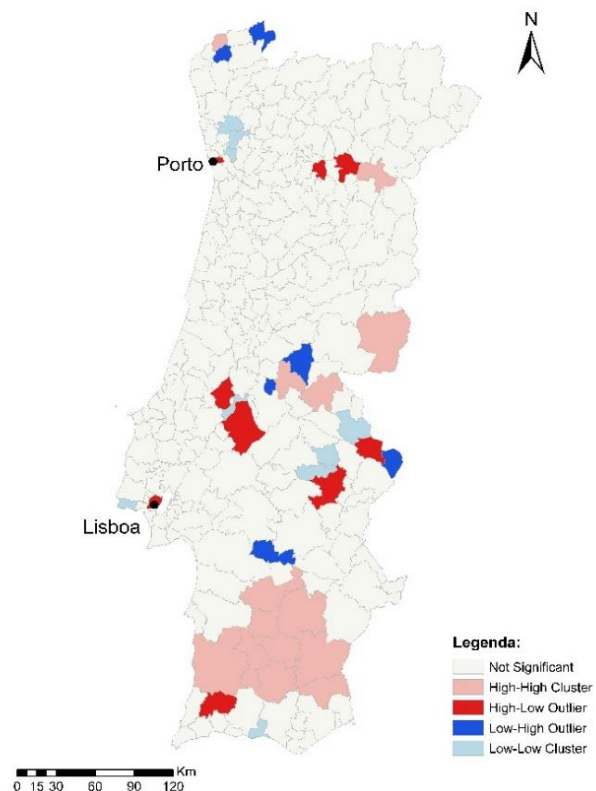


Figura 4.10 Local Moran's I - Análise de resíduos do modelo GWR

4.3.1. Coeficientes e respetivos erros padrão

Percentagem de população com educação de 2º ciclo

É de esperar que se a população tiver um maior nível de educação terá mais informação sobre as formas corretas de alimentação, e como tal terá uma menor propensão de sofrer doenças destes tipos e eventualmente menor mortalidade. A variável “População com educação de 2º ciclo” tem uma associação negativa com a taxa de mortalidade (**Figura 4.11**), com coeficientes mais negativos nos municípios do interior centro e sul (tons mais claros no mapa), e menos negativos nas Áreas Metropolitanas de Lisboa e Porto (tons mais escuros). Na zona norte apresentam-se valores negativos baixos, mas com erro padrão associado quase nulo. Os erros padrão variam entre 0,053 e 0,126 (**Figura 4.12**). Estes coeficientes negativos indicam que maior população com baixos níveis de educação (no máximo com o 6º ano de escolaridade), principalmente nos municípios do interior centro e sul, está associada a menores taxas de mortalidade relacionadas com hábitos alimentares. Estes resultados podem, eventualmente, ser explicados pelo facto de a educação de 2º ciclo poder ser considerada como um nível de escolaridade relativamente elevado, especialmente nos municípios do interior centro e sul. Em 2011, 74,5% da população residente com 15 e mais anos no Alentejo, e 69% no Algarve, não tinha ensino secundário (cf. dados do Pordata, <https://www.pordata.pt>, atualizados a 08/02/2019). A este nível de NUTS II, a Área Metropolitana de Lisboa era a que apresentava menor percentagem da população com 15 e mais anos sem ensino secundário do país (61,5%).

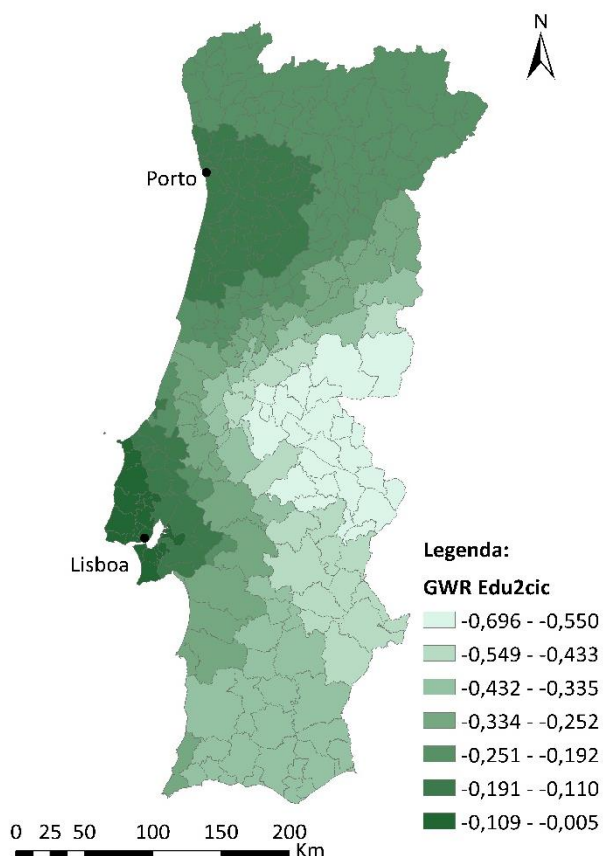


Figura 4.11 GWR Educação de 2º ciclo

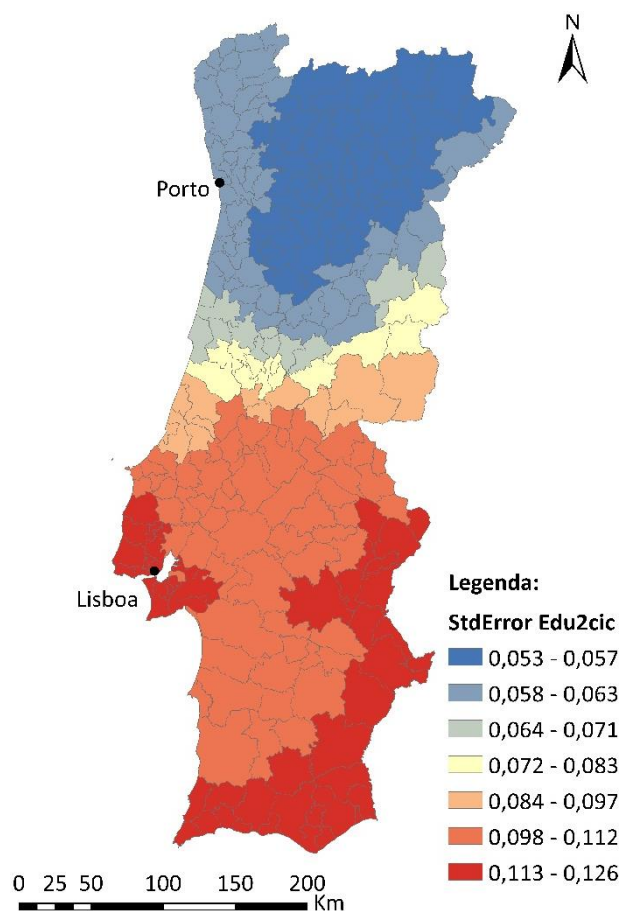


Figura 4.12 Erros padrão associados à educação de 2º ciclo

Distância média aos restaurantes *fast food* por habitante

Em todos os municípios existe uma associação positiva entre a variável “Distância média aos restaurantes *fast food* por habitante” e a variável de estudo (**Figura 4.13**), ao contrário do que seria de esperar. Tal poderá dever-se à forma como a variável foi construída e ao facto de se ter considerado restaurantes de uma única cadeia de *fast food*, não sendo, portanto, representativa do fator distância que se pretendia investigar. Note-se que, além da distância aos restaurantes, esta variável considera implicitamente o número de habitantes do município, que poderá ter um grande impacto nos resultados.

Observa-se uma maior influência desta variável nas zonas urbanas, com destaque para o centro litoral (região de Leiria e uma parte da região de Coimbra) e na Área Metropolitana do Porto. Reflete-se uma menor influência no interior, nomeadamente no Alto Alentejo e Alentejo Central (NUT III). Ao observar os erros padrão para esta variável são baixos em todos os municípios, variando de 0,025 a 0,092, o que indicia estimativas precisas para esta variável (**Figura 4.14**).

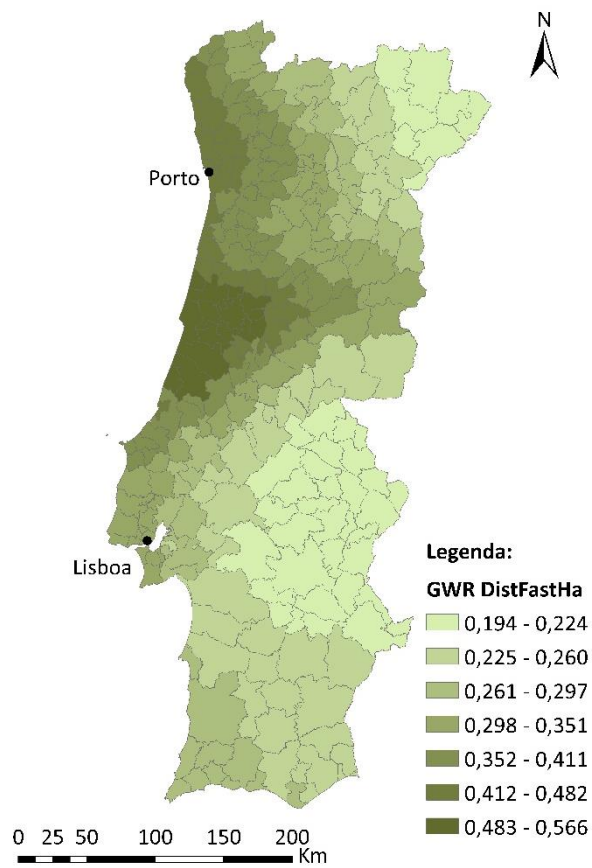


Figura 4.13 GWR Distância média a restaurantes fast food por habitante (Km/hab).

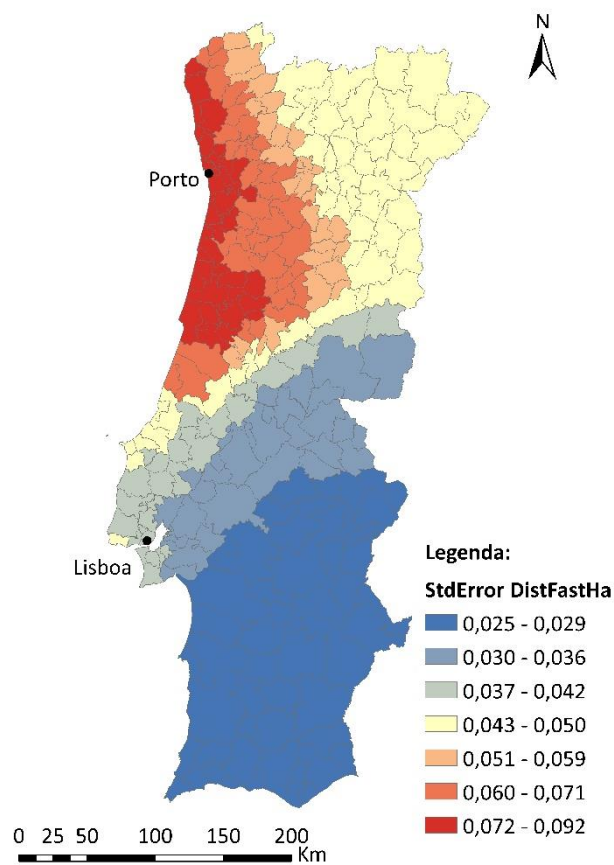


Figura 4.14 Erros padrão associados à distância média a restaurantes fast food por habitante.

Resposta hospitalar

Há uma associação negativa da “Resposta hospitalar” com a taxa de mortalidade em todos os municípios (**Figura 4.15**), como seria de esperar. Pois se a resposta hospitalar for boa, ou seja, existirem profissionais de saúde disponíveis para satisfazer as necessidades, será expectável um melhor estado de saúde da população. Esta variável tem menos influência nas Áreas Metropolitanas de Lisboa e Porto e ainda no centro interior e com maior influência nos municípios do interior do Alentejo e Algarve. Os erros padrão variam entre 0,098 a 0,232 (**Figura 4.16**), portanto a precisão dos coeficientes é elevada em todos os municípios. Com a região centro e parte do Alto Alentejo com os menores erros associados á variável e com valores mais elevados de erro no Alentejo interior junto à fronteira com Espanha.

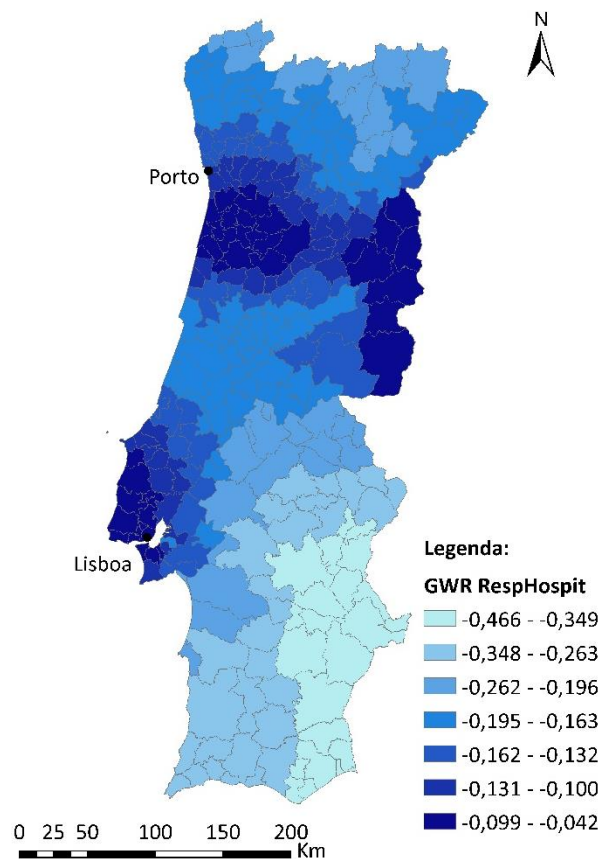


Figura 4.15 GWR Resposta hospitalar

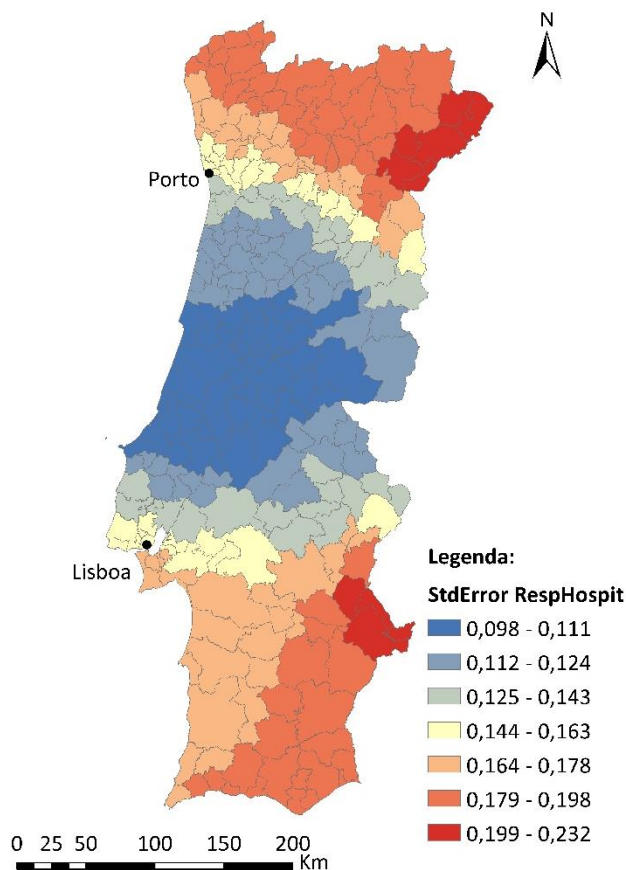


Figura 4.16 Erros padrão associados à variável resposta hospitalar

Rácio entre supermercados e lojas de conveniência

A variável “Rácio entre supermercados e lojas de conveniência” tem uma associação negativa com a variável de estudo na maioria dos municípios (**Figura 4.17**), com exceção de Leiria e Marinha Grande. Coeficientes inferiores localizam-se na região do Minho.

Os erros padrão variam de 0,166 a 0,322 (**Figura 4.18**), apresentando os maiores valores a Centro e Norte.

De acordo com o pressuposto desta variável, a quantidade de supermercados por lojas de conveniência ser superior (valores elevados do rácio) implica um melhor acesso a alimentos saudáveis, tendo um efeito de redução da mortalidade. Assim, os coeficientes têm um comportamento de acordo com o esperado.

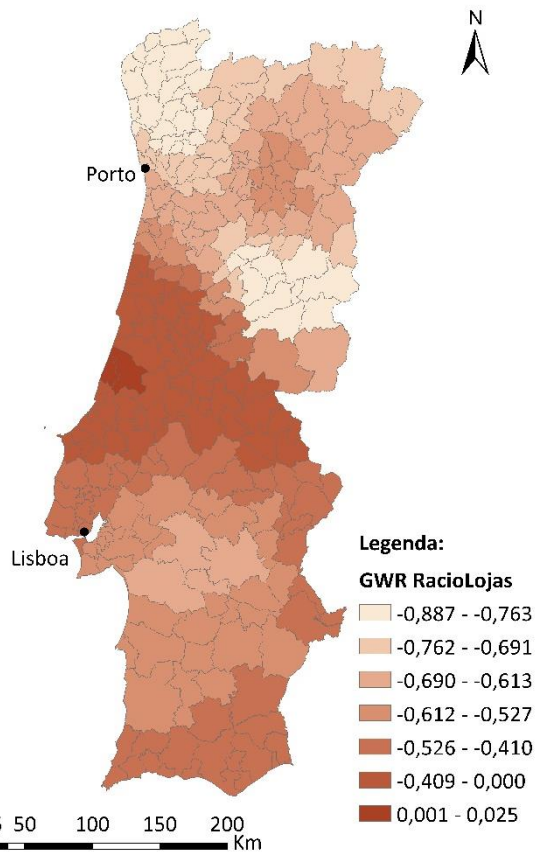


Figura 4.17 GWR Rácio entre supermercados e lojas de conveniência

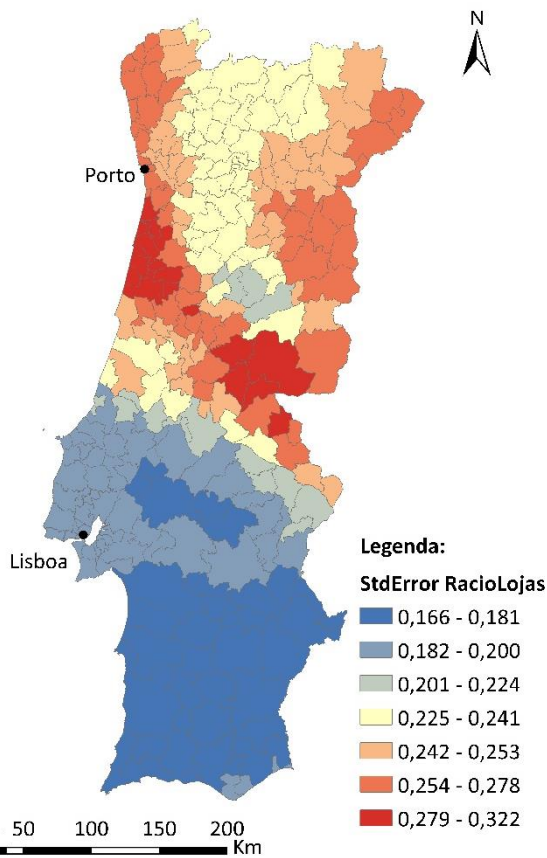


Figura 4.18 Erros padrão rácio entre supermercados e lojas de conveniência

4.3.2. Comparação dos modelos OLS vs. GWR

O R^2 Ajustado foi superior no modelo GWR (0,61) comparativamente com do modelo OLS (0,57). O Corrected Akaike Information Criterion (AICc) apresenta o valor 974,91 para o modelo GWR e 993,10 para o modelo OLS. Verificou-se, portanto, uma melhoria do ajustamento global do modelo espacial local em comparação com os resultados do modelo global (**Tabela 4.5**).

Em suma, de acordo com o esperado, o modelo GWR é melhor que o modelo OLS. Apesar de apresentar algumas limitações (que serão detalhadas posteriormente), o modelo espacial local permite analisar a variação da relação entre as taxas de mortalidade e as variáveis explicativas nos municípios portugueses do continente.

Tabela 4.5 Indicadores de comparação das regressões Global (OLS) e Local (GWR)

Modelo	R^2	R^2 ajustado	AICc	Spatial Moran's I	Autocorrelation Z-score	p-values
OLS	0,57	0,57	993,10	0,11	3,91	0,000092
GWR	0,64	0,61	974,91	0,05	1,88	0,059963

5. CONCLUSÃO

Tal como definido nos objetivos, pretendia-se identificar alguns fatores que pudessem influenciar a mortalidade relacionada com hábitos alimentares, que serviu como indicador do estado de saúde dos municípios. Para tal, primeiro realizou-se a modelação das taxas de mortalidade usando modelos de regressão linear múltipla para identificar as variáveis explicativas mais relevantes, e que permitissem obter um modelo OLS sem problemas de multicolinearidade e globalmente bem ajustado. A seleção das potenciais variáveis explicativas baseou-se na revisão de literatura, tendo a sua recolha sido limitada pela disponibilidade de dados. Depois, as variáveis explicativas identificadas nesse modelo global foram usadas na regressão espacial local (modelo GWR). Assim, tanto no modelo global como no local foram selecionadas as mesmas variáveis, relacionadas com educação, com o acesso à alimentação e com o acesso à saúde.

Mais concretamente, a “Educação de 2º ciclo” revelou ter uma associação negativa com a variável de estudo em ambos os modelos. Em termos locais, esta variável tem maior influência nos municípios do interior do país, dando a ideia de que à medida que este nível de educação aumenta a mortalidade diminui nestas regiões. Quanto à “Resposta hospitalar”, também com uma relação negativa em relação à variável dependente, apresenta nos municípios do Sul junto à fronteira com Espanha uma maior influência. Por outro lado, o “Rácio entre supermercados e lojas de conveniência”, apesar de manter a associação negativa como as variáveis anteriores, localmente tem mais relevância no norte de Portugal. A variável “Distância média a restaurantes *fast food* por habitante” apresenta uma associação positiva, e em termos locais tem coeficientes de valores mais elevados no litoral centro e no litoral norte.

O modelo local (GWR) ajusta-se melhor aos dados do que o modelo global (OLS), de acordo com as métricas utilizadas (R^2 e AICc). No entanto, as variáveis selecionadas não explicam a taxa de mortalidade na sua totalidade. Ambos os modelos estão sub-especificados devido à escassez de variáveis que ajudem a explicar a alimentação da população portuguesa. O modelo GWR apresentou maior poder explicativo (superior a 61%) nos municípios dos distritos de Aveiro e Coimbra.

Importa salientar que este estudo constitui uma análise exploratória com algumas limitações, que serão enumeradas na secção abaixo.

5.1. LIMITAÇÕES

Uma das limitações diz respeito ao facto da elaboração do estudo ao nível do município poder induzir em falácia ecológica, pois a análise foi elaborada ao nível da população do município (agrupado) e os fatores ao nível do individuo podem ser diferentes dos encontrados (Ford & Highfield, 2016).

Como o estudo foi elaborado ao nível do município, a disponibilidade da informação para algumas categorias de variáveis explicativas foi reduzida. Em especial, informação sobre a alimentação da população é praticamente inexistente, limitando assim as variáveis associadas aos hábitos alimentares. Por falta de informação, também não foram consideradas variáveis associadas a fatores biológicos ou genéticos.

Relativamente à variável dependente, surgem duas limitações principais. Em primeiro lugar, os óbitos por doenças cardiovasculares, apesar de serem influenciados pela alimentação, têm também outras causas. Em segundo lugar, alguns municípios (ver Anexo), têm valores em falta, devido à confidencialidade dos dados. Assim, para poder analisar todos os municípios, esses valores foram estimados, o que pode ter criado um ligeiro viés nos resultados.

5.2. RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Após a elaboração do presente trabalho surgiram algumas ideias e recomendações que poderão melhorar as futuras investigações na área.

Realizar um estudo semelhante, mas utilizando os óbitos por diabetes, por doenças do aparelho digestivo e acrescentando também os óbitos causados por tumores relacionados com o aparelho digestivo poderá originar um estudo mais completo. O estudo também beneficiaria da modelação do risco relativo da mortalidade associada a estas doenças, em vez da modelação direta das taxas de mortalidade (veja-se, por exemplo, o estudo sobre a mortalidade por cancro do pulmão em Portugal de Cardoso, Painho, & Roquette, 2019).

Uma vez que grande parte das variáveis obtidas para a análise exploratória não foram utilizadas no modelo final, por não serem adequadas, sugere-se a utilização de uma análise em componentes principais e a inclusão dos fatores obtidos na modelação GWR.

Referente aos fatores socioeconómicos, a utilização de índice/indicador de privação da população, similar ao *Townsend Deprivation Index* e tomar uma abordagem semelhante à do estudo de Ford & Highfield (2016), poderá enriquecer a análise. A modelação também poderia beneficiar da criação de uma variável que possibilite a identificação da Tipologia de Área Urbana, similar à que é utilizada no estudo de Odoi & Busingye (2014) sobre desigualdades geográficas em vizinhanças, a qual sugere a utilização da percentagem de população que habita em zonas predominantemente rurais/urbanas. Por outro lado, do ponto de vista clínico, uma informação pertinente poderia ser a utilização de dados sobre exames e rastreios realizados, pois a falta de diagnóstico poderá influenciar o tratamento, implicando que o tratamento necessário seja aplicado numa fase mais avançada das doenças. Com a ressalva de que estes dados são de difícil acesso, pois podem estar restritos a ambiente hospitalar.

Quanto à disponibilidade de dados referentes aos hábitos alimentares da população, apesar de já existir o IAN-AF (referido anteriormente, <https://ian-af.up.pt/>) que disponibiliza informação a nível regional (NUTII), seria importante ao nível do município efetuar um levantamento de algumas destas informações, através dos censos, com uma secção referente ao assunto, ou mesmo através da criação dos censos da alimentação. Apesar das limitações financeiras que poderiam advir da sua realização, certamente os resultados conseguiriam levar a uma implementação de medidas mais direcionadas e eficientes. O ideal seria a recolha da informação sobre os consumos de bebidas alcoólicas, consumo de alimentos processados e informação biométrica (peso e altura, etc.). Em complemento, seria benéfica a introdução de informação relacionada com a atividade física como o acesso aos ginásios, atividades desportivas e realização de ar livre.

REFERÊNCIAS

- Afshin, A., Micha, R., Khatibzadeh, S., Fahimi, S., Shi, P., Powles, J., ... Mozaffarian, D. (2015). The impact of dietary habits and metabolic risk factors on cardiovascular and diabetes mortality in countries of the Middle East and North Africa in 2010: A comparative risk assessment analysis. *BMJ Open*, 5(5). <https://doi.org/10.1136/bmjopen-2014-006385>
- Bagheri, N., Holt, A., & Benwell, G. L. (2009). Using geographically weighted regression to validate approaches for modelling accessibility to primary health care. *Applied Spatial Analysis and Policy*, 2(3), 177–194. <https://doi.org/10.1007/s12061-009-9021-0>
- Barreto, M., Gaio, V., Kislava, I., Antunes, L., Rodrigues, A. P., Silva, A. C., ... Matias Dias, C. (2016). 1º Inquérito Nacional de Saúde com Exame Físico (INSEF 2015): Estado de Saúde, 1–61. Retrieved from <http://repositorio.insa.pt/handle/10400.18/4115>
- Bascuñán, M. M., & Quezada, C. R. (2016). Geographically weighted regression for modelling the accessibility to the public hospital network in Concepción Metropolitan Area, Chile. *Geospatial Health*, 11(3), 263–273. <https://doi.org/10.4081/gh.2016.451>
- Bäumler, J. S. K. (2013). *The use of GIS to Analyze the Environmental and Social Influence to Medicated Asthma in New Zealand*. Dissertação de Mestrado, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa. Retrieved from <http://hdl.handle.net/10451/20513>
- Brunsdon, C., Fotheringham, S., & Charlton, M. (1996). Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431–443. <https://doi.org/10.1111/1467-9884.00145>
- Brunsdon, Chris, Charlton, M., & Harris, P. (2012). Living with Collinearity in Local Regression Models. *International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Sciences*, 67–72.
- Buck, K. D. (2016). Modelling of geographic cancer risk factor disparities in US counties. *Applied Geography*, 75, 28–35. <https://doi.org/10.1016/j.apgeog.2016.08.001>
- Burger, K. N. J., Beulens, J. W. J., van der Schouw, Y. T., Sluijs, I., Spijkerman, A. M. W., Sluik, D., ... Nöthlings, U. (2012). Dietary Fiber, Carbohydrate Quality and Quantity, and Mortality Risk of Individuals with Diabetes Mellitus. *PLoS ONE*, 7(8), e43127. <https://doi.org/10.1371/journal.pone.0043127>
- Cardoso, D., Painho, M., & Roquette, R. (2019). A geographically weighted regression approach to investigate air pollution effect on lung cancer: A case study in Portugal. *Geospatial Health*, 14(1), 35–45. <https://doi.org/10.4081/gh.2019.701>
- Castelli, A., Jacobs, R., Goddard, M., & Smith, P. C. (2013). Health, policy and geography: Insights from a multi-level modelling approach. *Social Science and Medicine*, 92, 61–73. <https://doi.org/10.1016/j.socscimed.2013.05.021>
- Chalkias, C., Papadopoulos, A. G., Kalogeropoulos, K., Tambalis, K., Psarra, G., & Sidossis, L. (2013). Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: Empirical evidence from Athens, Greece. *Applied Geography*, 37(1), 34–43. <https://doi.org/10.1016/j.apgeog.2012.10.007>

- Cho, Y., Cudhea, F., Park, J.-H., Mozaffarian, D., Singh, G., & Shin, M.-J. (2017). Burdens of Cardiometabolic Diseases Attributable to Dietary and Metabolic Risks in Korean Adults 2012–2013. *Yonsei Medical Journal*, 58(3), 540. <https://doi.org/10.3349/ymj.2017.58.3.540>
- Clary, C., Lewis, D. J., Flint, E., Smith, N. R., Kestens, Y., & Cummins, S. (2016). The local food environment and fruit and vegetable intake: A geographically weighted regression approach in the ORiEL study. *American Journal of Epidemiology*, 184(11), 837–846. <https://doi.org/10.1093/aje/kww073>
- Comber, A. J., Brunsdon, C., & Radburn, R. (2011). A spatial analysis of variations in health access: Linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics*, 10, 1–11. <https://doi.org/10.1186/1476-072X-10-44>
- Constituição da República Portuguesa. Artigo nº64 - Saúde, Pub. L. No. CAPÍTULO II Direitos e deveres sociais (1976). Portugal. Retrieved from <http://www.parlamento.pt/Legislacao/paginas/constituicaorepublikaportuguesa.aspx>
- David, J., & Cabral, P. (2019). Modelling youth pregnancy in continental Portugal through geographically weighted regression. *Geospatial Health*, 14(1). <https://doi.org/10.4081/gh.2019.680>
- DGTerritório. (2011). Carta Administrativa Oficial de Portugal - Versão 2011. Retrieved January 17, 2016, from http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal__caop_/caop__download_/carta_administrativa_oficial_de_portugal__versao_2011_2/
- Ford, M. M., & Highfield, L. D. (2016). Exploring the spatial association between social deprivation and cardiovascular disease mortality at the neighborhood level. *PLoS ONE*, 11(1), 1–17. <https://doi.org/10.1371/journal.pone.0146085>
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: the analysis of spatially varying relationships. Wiley.
- Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751–2769. <https://doi.org/10.1068/a38218>
- Hayashi, F. (2000). Finite-Sample Properties of OLS. In *Econometrics*. Princeton University Press.
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401. <https://doi.org/10.1080/13658810802672469>
- INE. (2011a). Óbitos (N.º) por Local de residência (NUTS - 2013). Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008206&contexto=bd&selTab=tab2
- INE. (2011b). População residente em lugares censitários (N.º) por Local de residência (NUTS - 2013) e Escalão de dimensão populacional; Decenal. Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0009085&contexto=bd&selTab=tab2

- Larson, C., & Mercer, A. (2004). Global health indicators: an overview. *CMAJ : Canadian Medical Association Journal*, *171*(10), 1199–1200. <https://doi.org/10.1503/cmaj.1021409>
- Leenders, M., Boshuizen, H. C., Ferrari, P., Siersema, P. D., Overvad, K., Tjønneland, A., ... Bueno-de-Mesquita, H. B. (2014). Fruit and vegetable intake and cause-specific mortality in the EPIC study. *European Journal of Epidemiology*, *29*(9), 639–652. <https://doi.org/10.1007/s10654-014-9945-9>
- Lei nº48/90 de 24 de Agosto da Assembleia da República, Pub. L. No. Diário da República: I Série, nº 195 (1990). Portugal: Diário da República Portuguesa. Retrieved from www.dre.pt
- Leong, Y. Y., & Yue, J. C. (2017). A modification to geographically weighted regression. *International Journal of Health Geographics*, *16*(1), 1–18. <https://doi.org/10.1186/s12942-017-0085-9>
- Lewandowska-Gwarda, K. (2018). Geographically Weighted Regression in the Analysis of Unemployment in Poland. *ISPRS International Journal of Geo-Information*, *7*(1), 17. <https://doi.org/10.3390/ijgi7010017>
- Matthews, S. A., & Yang, T. C. (2012). Mapping the results of local statistics: Using geographically weighted regression. *Demographic Research*, *26*, 151–166. <https://doi.org/10.4054/DemRes.2012.26.6>
- Mehta, N. K., & Chang, V. W. (2008). Weight Status and Restaurant Availability. A Multilevel Analysis. *American Journal of Preventive Medicine*, *34*(2), 127–133. <https://doi.org/10.1016/j.amepre.2007.09.031>
- Micha, R., Peñalvo, J. L., Cudhea, F., Imamura, F., Rehm, C. D., & Mozaffarian, D. (2017). Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States. *JAMA*, *317*(9), 912. <https://doi.org/10.1001/jama.2017.0947>
- Micha, R., Shulkin, M. L., Peñalvo, J. L., Khatibzadeh, S., Singh, G. M., Rao, M., ... Mozaffarian, D. (2017). Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE). *PLOS ONE*, *12*(4), e0175149. <https://doi.org/10.1371/journal.pone.0175149>
- Ministério da Saúde. (2018). *Retrato da Saúde*. Lisboa. Retrieved from https://www.sns.gov.pt/wp-content/uploads/2018/04/RETRATO-DA-SAUDE_2018_compressed.pdf
- Mozaffarian, D. (2016). Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. *Circulation*, *133*(2), 187–225. <https://doi.org/10.1161/CIRCULATIONAHA.115.018585>
- Nakaya, T., Nakase, K., & Osaka, K. (2005). Spatio-temporal modelling of the HIV epidemic in Japan based on the national HIV/AIDS surveillance. *Journal of Geographical Systems*, *7*(3–4), 313–336. <https://doi.org/10.1007/s10109-005-0008-3>
- Ndiath, M. M., Cisse, B., Ndiaye, J. L., Gomis, J. F., Bathiery, O., Dia, A. T., ... Faye, B. (2015). Application of geographically-weighted regression analysis to assess risk factors for malaria hotspots in Keur Soce health and demographic surveillance site. *Malaria Journal*, *14*(1), 1–

11. <https://doi.org/10.1186/s12936-015-0976-9>
- Nezami, S., & Khoramshahi, E. (2016). Spatial Modeling of Crime by Using of GWR Method. *Proceedings - 2016 Baltic Geodetic Congress (Geomatics), BGC Geomatics 2016*, 222–227. <https://doi.org/10.1109/BGC.Geomatics.2016.47>
- Nogueira, H., & Santana, P. (2005). Geographies of health and deprivation: relationship between them. In *Atti dell' VIII Seminario Internazionale do Geografia Medica* (pp. 539–546).
- O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Odoi, A., & Busingye, D. (2014). Neighborhood geographic disparities in heart attack and stroke mortality: Comparison of global and local modeling approaches. *Spatial and Spatio-Temporal Epidemiology*, 11, 109–123. <https://doi.org/10.1016/j.sste.2014.10.001>
- Ogneva-Himmelberger, Y., Rakshit, R., & Pearsall, H. (2013). Examining the Impact of Environmental Factors on Quality of Life Across Massachusetts. *The Professional Geographer*, 65(2), 187–204. <https://doi.org/10.1080/00330124.2011.639631>
- Oliveira, A., Cabral, A. J. R., Mendes, J. M., Martins, M. R. O., & Cabral, P. (2015). Spatiotemporal analysis of the relationship between socioeconomic factors and stroke in the Portuguese mainland population under 65 years old. *Geospatial Health*, 10(2), 171–184. <https://doi.org/10.4081/gh.2015.365>
- Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), 2992–3010. <https://doi.org/10.1068/a44111>
- Presidência da República. (2019). Presidente da República promulga sete diplomas da Assembleia da República. Retrieved August 26, 2019, from <http://www.presidencia.pt/?idc=10&idi=168194>
- Rodrigues, S. S. P., Trichopoulou, A., & De Almeida, M. D. V. (2008). Household diet quality in relation to mortality in Portuguese regions: An ecological study. *Journal of Public Health*, 16(1), 43–51. <https://doi.org/10.1007/s10389-007-0113-5>
- Ruano-Rodríguez, C., Serra-Majem, L., & Dubois, D. (2015). Assessing the impact of dietary habits on health-related quality of life requires contextual measurement tools. *Frontiers in Pharmacology*, 6, 101. <https://doi.org/10.3389/fphar.2015.00101>
- Santana, P. (2014). *Introdução à geografia da saúde território, saúde e bem-estar*. (Pombalina, Ed.) (Santana, P). Coimbra: Imprensa da Universidade. <https://doi.org/http://dx.doi.org/10.14195/978-989-26-0727-6>
- Santana, P. (2015). *A Geografia da Saúde da População. Evolução nos últimos 20 anos em Portugal Continental*. Centro de Estudos em Geografia e Ordenamento do Território (CEGOT). <https://doi.org/10.17127/cegot/2015.GS>
- Shahid, R., & Bertazzon, S. (2015). Local Spatial Analysis and Dynamic Simulation of Childhood Obesity and Neighbourhood Walkability in a Major Canadian City. *AIMS Public Health*, 2(4), 616–637. <https://doi.org/10.3934/publichealth.2015.4.616>

- Verlangieri, A. J., Kapeghian, J. C., El-Dean, S., & Bush, M. (1985). Fruit and vegetable consumption and cardiovascular mortality. *Medical Hypotheses*, 16(1), 7–15. [https://doi.org/10.1016/0306-9877\(85\)90035-0](https://doi.org/10.1016/0306-9877(85)90035-0)
- West, G. (2014). *The Socio-Spatial Relationship between Los Angeles County Youth Health and Food Establishments*. California State University, Northridge.
- Wheeler, D. C., & Paez, A. (2002). *Handbook of Applied Spatial Analysis*. (M. M. Fischer & A. Getis, Eds.), *Handbook of applied spatial analysis: Software tools, methods and applications*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-03647-7>
- Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161–187. <https://doi.org/10.1007/s10109-005-0155-6>
- World Health Organization. WHO | Frequently asked questions (1948). World Health Organization. Retrieved from <http://www.who.int/suggestions/faq/en/>
- World Health Organization. (2014a). *Health in Impact Assessments*. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0011/261929/Health-in-Impact-Assessments-final-version.pdf?ua=1
- World Health Organization. (2014b). Health status indicator: Mortality. Retrieved February 21, 2019, from <https://www.who.int/healthinfo/statistics/indadultmortality/en/>
- World Health Organization. (2017). The determinants of Health. Retrieved May 10, 2018, from <http://www.who.int/hia/evidence/doh/en/>
- World Health Organization. (2018a). *2018 Global reference list of 100 core health indicators (plus health-related SDGs)*. Retrieved from <https://apps.who.int/iris/bitstream/handle/10665/259951/WHO-HIS-IER-GPM-2018.1-eng.pdf?sequence=1>
- World Health Organization. (2018b). *World Health Statistics 2018- Monitoring Health for the SDG's (Sustainable development goals)*.
- Ye, Y., Zuo, Z., Yuan, X., Zhang, S., Zeng, X., An, Y., & Chen, B. (2017). Geographically weighted regression model for urban traffic black-spot analysis. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4866–4869). IEEE. <https://doi.org/10.1109/IGARSS.2017.8128093>

ANEXOS

LISTA DE MUNICÍPIOS COM VALORES EM FALTA DE PORTUGAL

Salienta-se que a informação a laranja foi calculada através da diferença entre o total da região e a junção valores dos municípios dessa região com informação.

Diabetes	
Região de Leiria	4
Alvaiázere	2
Pedrógão Grande	2

Ap. Digestivo	
Região de Coimbra	6
Pampilhosa da Serra	3
Penela	3

Médio Tejo	6
Constância	2
Vila Nova da Barquinha	4

Região de Leiria	6
Castanheira de Pêra	2
Pedrógão Grande	4

Beiras e Serra da Estrela	4
Belmonte	3
Manteigas	1

Médio Tejo	2
Constância	1
Sardoal	1

Alto Alentejo	2
Marvão	1
Monforte	1

Baixo Alentejo	3
Alvito	2
Barrancos	1

Alentejo Central	3
Alandroal	2
Mourão	1

Alto Alentejo	2
Castelo de Vide	1
Fronteira	1

Douro	3
Mesão Frio	2
Sernancelhe	1

Alentejo Central	2
Mourão	1
Vendas Novas	1

Soma 22

Algarve	3
Alcoutim	2
Vila do Bispo	1

Soma 24

MODO DE AGRUPAMENTO DAS VARIÁVEIS BASEADAS NO CAE.

Tabela 0.1 Classificação das Atividades Económicas (CAE) de Restauração utilizadas no indicador.

CAE-Rev.3	Descrição
56101	Restaurantes tipo tradicional.
56102	Restaurantes com lugares ao balcão.
56103	Restaurantes sem serviço de mesa.
56104	Restaurantes típicos.
56105	Restaurantes com espaço de dança.
56106	Confeção de refeições prontas a levar para casa (takeaway).
56107	Restaurantes, n. e. (inclui atividades de restauração em meios móveis).
56290	Outras atividades de serviço de refeições (Cantinas e similares).
Modo de agrupar:	
	Restaurantes de refeições completas
	Restaurantes de <i>fast food</i>

Tabela 0.2 Classificação das Atividades Económicas (CAE) de Comércio de bens alimentares utilizadas no indicado “Rácio entre o nº supermercados e o nº lojas de conveniência”.

CAE-Rev.3	Descrição
47111	Comércio a retalho em supermercados e hipermercados
47112	Comércio a retalho em outros estabelecimentos não especializados, com predominância de produtos alimentares, bebidas ou tabaco
47210	Comércio a retalho de frutas e produtos hortícolas, em estabelecimentos especializados
47220	Comércio a retalho de carne e produtos à base de carne, em estabelecimentos especializados
47230	Comércio a retalho de peixe, crustáceos e moluscos, em estabelecimentos especializados
47240	Comércio a retalho de pão, de produtos de pastelaria e de confeitaria, em estabelecimentos especializados
47250	Comércio a retalho de bebidas, em estabelecimentos especializados
47291	Comércio a retalho de leite e de derivados, em estabelecimentos especializados
47292	Comércio a retalho de produtos alimentares, naturais e dietéticos, em estabelecimentos especializados
47293	Outro comércio a retalho de produtos alimentares, em estabelecimentos especializados, n.e.
Modo de agrupar:	
	Lojas de conveniência
	Supermercados e lojas de venda de produtos especializados

