MASTER
APPLIED ECONOMETRICS AND FORECASTING

MASTER´S FINAL WORK
DISSERTATION

DOCUMENT SPECIFICALLY ELABORATED TO OBTAIN A MASTER'S DEGREE

THE HEART OF THE MATTER: AN ECONOMETRIC STUDY ON THE IMPACT OF SOCIOECONOMIC VARIABLES ON CIRCULATORY DISEASE MORTALITY IN PORTUGUESE MUNICIPALITIES

MATILDE COSTA FARTO VAZ ANTUNES

SUPERVISION:
ESMERALDA DE JESUS RATINHO LOPES ARRANHADO

JUNE - 2025

To António and Carla,
without whom I would not
have made it.

GLOSSARY

AIC – Akaike Information Criterion.

AMI – Acute Myocardial Infarction.

BIC – Bayesian Information Criterion.

CDF – Conditional Density Function.

CLOGLOG – Complementary Log-Log.

CVD – Cardiovascular Disease.

EU – European Union.

GDP – Gross Domestic Product.

HF – Heart Failure.

IHD – Ischemic Heart Disease.

INE – Instituto Nacional de Estatística.

LL – Log-Likelihood.

LMA – Lisbon Metropolitan Area.

ML – Maximum Likelihood.

OLS – Ordinary Least Squares.

PCI – Percutaneous Coronary Intervention.

PDF – Probability Density Function.

PE – Partial Effects.

PMA – Porto Metropolitan Area.

PPI – Purchasing Power Index.

QML – Quasi-Maximum Likelihood.

SES – Socioeconomic Status.

ABSTRACT, KEYWORDS AND JEL CODES

This dissertation investigates the determinants of circulatory disease mortality rates across all 308 Portuguese municipalities in 2021 through econometric analysis of socio-economic, demographic, geographic, and environmental variables. After identifying a research gap on the municipal health field in Portugal, the greatest aim is to provide essential insights to national health practicians and policymakers. By employing fractional regression models – logit, probit, and complementary log-log -, appropriate for bounded continuous outcomes, this study analyses the cross-sectional data recovered from two well-known Portuguese national databases, focusing essentially on the relevant socioeconomic indicators.

The results reveal significant associations between circulatory disease mortality and many factors: positive relationships with illiteracy rates, higher-aged populations, the presence of recognized cardiology reference centres, and the PPI; negative associations with available healthcare infrastructure (number of hospitals and doctors per 1,000 inhabitants), population density, and municipal expenditure in sports equipment.

These findings offer crucial insights for Portuguese healthcare decision-makers working to reduce regional differences in cardiac disease mortality rates. By understanding how socioeconomic factors like purchasing power, age demographics, and healthcare infrastructure influence cardiovascular outcomes, policymakers can develop more effective strategies to improve heart health across Portugal.

KEYWORDS: Circulatory Disease; Mortality Rate; Socioeconomic Factors; Healthcare Resources; Fractional Regression Models; Municipal-Level Analysis in Portugal.

JEL CODES: C21; H75; I14; I15; I18; R11.

Resumo, Palavras-Chave e Códigos JEL

Esta dissertação investiga os determinantes das taxas de mortalidade por doenças circulatórias em todos os 308 concelhos portugueses em 2021, através da análise econométrica de variáveis socioeconómicas, demográficas, geográficas e ambientais. Após a identificação de uma lacuna na investigação sobre a saúde concelhia em Portugal, o principal objetivo é fornecer informação essencial aos profissionais de saúde e decisores políticos nacionais. Recorrendo a modelos de regressão fracionária – *logit*, *probit* e *complementary log-log* –, apropriados para resultados contínuos limitados, este estudo analisa os dados transversais retirados de duas bases de dados nacionais portuguesas bem conhecidas, focando-se essencialmente nos indicadores socioeconómicos relevantes.

Os resultados revelam associações significativas entre a mortalidade por doenças circulatórias e diversos fatores: relações positivas com as taxas de analfabetismo, populações mais idosas, presença de centros de referência em cardiologia reconhecidos e índice de poder de compra; associações negativas com as infraestruturas de saúde disponíveis (número de hospitais e médicos por 1.000 habitantes), a densidade populacional e a despesa municipal em equipamento desportivo.

Estas descobertas oferecem *insights* cruciais para os decisores políticos portugueses na área da saúde que trabalham para reduzir as diferenças regionais nas taxas de mortalidade por doenças cardíacas. Ao compreender como fatores socioeconómicos como o poder de compra, a idade, a demografia e as infraestruturas de saúde influenciam os resultados cardiovasculares, os decisores políticos podem desenvolver estratégias mais eficazes para melhorar a saúde cardíaca em Portugal.

Palavras-Chave: Doença Circulatória; Taxa de Mortalidade; Fatores Socioeconómicos; Recursos de Saúde; Modelos de Regressão Fracionária; Análise Municipal em Portugal.

Códigos JEL: C21; H75; I14; I15; I18; R11.

TABLE OF CONTENTS

TABLE OF FIGURES

Acknowledgments

First, I wish to express my profound gratitude to Professor Esmeralda Arranhado for her encouragement and expert guidance throughout the development of this thesis. Her insightful feedback and continuous support have been essential in shaping this research.

I am also deeply grateful to Dr. Rui Anjos, the cardiologist who has been a fundamental part of my healthcare journey since birth. His extensive medical knowledge, coupled with his willingness to share valuable insights, provided me with a unique perspective for this thesis.

My sincere appreciation extends to my Master's colleagues, especially those who worked closely with me throughout this path. Our discussions, problem-solving sessions, and mutual support have been precious.

I am profoundly grateful to my work colleagues from two different teams over these past two years. Their extraordinary understanding and support were precious as I balanced professional responsibilities with academic studies. Despite changing work environments, both teams showed remarkable empathy and flexibility, allowing me to dedicate necessary time to my Master.

Profound thanks go to my dearest friends who have been a constant source of support throughout my journey(s). My long-standing pears from basic, secondary, and high-school – including Erasmus buddies – have been priceless in offering encouragement, while my fellow Scouts provided much-needed rest from research. I am also grateful to my partners from Muay Thai and other fitness activities, whose energy helped maintain my well-being during this challenging period.

Finally, my deepest gratitude goes to my family, particularly my parents, for their endless patience, emotional support, and constant encouragement while I pursued this project. Their belief in me, even during the most demanding times, has been a source of strength and motivation. This thesis is as much a product of their love and support as it is of my own efforts.

1. Introduction

Understanding the importance of the determinants of circulatory disease mortality can be crucial when trying to innovate into new ways of preventive health care and medical procedures. In fact, in 2021, *CVD* (cardiovascular disease) *was the leading cause of death worldwide*, according to Buscardini (2023) and the World Health Report from the same year. Professor Fausto Pinto, former President of the World Health Federation, mentioned that *good data can help drive good policy*, making a reference to the high value of having insightful data and the respective analysis. This is not a reality when aiming to compare several countries or regions. Very often, important questions such as: do *countries that invest more in healthcare as a percentage of gross domestic product (GDP) have lower CVD death rates?* – remain unresponded.

In Portugal, information and databases on the health sector are scarce. In reality, it tends to be extremely difficult for a researcher to have access to specific data or information, mostly due to the high bureaucracy, lack of resources, and most notably, confidentiality concerns. Thus, it is not a surprise that there are not many studies in this sector. Therefore, a cross-sectional study based on information publicly available for each Portuguese municipality will be beneficial to decision-making not only at a national level, but also in specific geographical areas.

The primary research question on the horizon with this research is: *What role do socioeconomic factors like unemployment rate or purchasing power index play in determining circulatory disease mortality rates across all 308 Portuguese municipalities?*. Despite the vast nature of variables and factors used for this study, the most recent topic of discussion when addressing health outcomes tends to be the influence of socioeconomic determinants. Hence, although there is also a great interest in understanding the relationship between cardiac diseases mortality rates and factors related to demography, geography, and environmental concerns, the primary focus will be to study how the economic conditions can impact the functioning of the health sector results.

This investigation aims to analyse the impact of several factors (especially socioeconomic) on circulatory disease mortality rates in Portuguese municipalities, by applying and comparing conditional mean models in the context of fractional data – using logit, probit, and complementary log-log models. The ultimate end is to provide insights that could aid medical practitioners and policymakers in addressing circulatory disease mortality.

The collection of data was made through several data sources, as the National Institute of Statistics (INE – Instituto Nacional de Estatística), Pordata, and the Portuguese Ministry of Health, which allowed for a more comprehensive set of variables. With data for all 308 Portuguese municipalities, a cross-sectional dataset was built, mostly using data regarding the 2021 Portuguese census. Among those variables that, due to their nature, arouse the most curiosity on the results of their relationship with the dependent variable – the mortality rate from diseases of the circulatory system -, there is the unemployment rate, purchasing power index (PPI) per capita, and the illiteracy rate.

To set the tone for this study, it is vital to create a strong literature review that helps understand what was already investigated in this field and what can still be done to add value to this thematic. This survey is presented in the following chapter. After doing so, and analysing what the existing studies did in econometric terms, it is feasible to advance to some model-specification and testing, in accordance with the type of data that is being used. The third chapter addresses the methodological background for the thesis. The fourth chapter describes the available dataset. The fifth chapter discusses the main results of the regression models. Finally, some concluding remarks are presented in chapter six.

2. LITERATURE REVIEW

The history of the study of cardiac disease and its determinants is rich, reflecting all the progress made in medical science, public health and even socioeconomic analysis. In former contributions, cardiac diseases were commonly attributed to lifestyle and individual behaviour, with very little thought given to wider relevant factors that emerged lately such as socioeconomic determinants. These include unemployment rates, educational attainment, income levels, and overall economic stability, which have been shown to significantly impact cardiovascular health outcomes. In fact, early influential studies, pioneered by the Framingham Heart Study (1948), conducted to about 5,200 individuals in the United States of America, under the direction of the National Heart Institute, emphasized the role of individual health-related risk factors, but subsequent work demonstrated the strong impact of higher-level economic factors on rates and trends in coronary heart disease and mortality.

The links between economic conditions and cardiac health have been the subject of investigation since the mid-20th century, when public health experts first noted disparities in disease outcomes in different socioeconomic groups. These disparities were initially blamed largely on differences in access to healthcare. Eventually, though, the rise of welfare economics and the development of econometric techniques in the mid- century allowed a more formal quantification of the impact of income, education and employment on health outcomes. Fast forward to today, and economic conditions continue to play an essential role in health outcomes, which today are an essential topic of concern for not just public health, but economic policy, as addressing structural inequalities is needed now more than ever to mitigate cardiac mortality.

As this field has advanced, two distinct approaches have been developed in empirical work: studies focusing on health outcomes and their determinants at individual level, and studies directed to the investigation of economic and social explanatory factors at a nation or regional level. The former approach is followed by studies such as Farahni et al. (2010) – that examined the effects of state-level public health spending on mortality probability in India – and Smits et al. (2004) – that analysed the effects of socioeconomic status (SES) on total and cause-specific mortality in the Netherlands in 1999, using data on all 15.8 million inhabitants – focused on individual-level data from national health surveys to assess the impact of public health expenditure on mortality. On the other hand, studies addressing the problem at a country or regional level include Granados & Ionides (2017) and Kawachi et al. (2023).

This thesis takes the latter approach. In the following sections, a comprehensive survey on the existing papers that may be relevant to the field is made available – these papers are presented in chronological order, within each section. On the Appendix, there is a summary table (table 5) of the approach and findings of several relevant documents that use econometric models.

2.1. Determinants of Cardiovascular Mortality: Socioeconomic Factors

This review provides a detailed analysis of the literature relating economic factors to cardiovascular health. They serve as a basis for better understanding how individual economic attributes influence heart conditions. The following papers highlight the findings in this area, demonstrating how factors such as personal income, employment status, and educational status correlate with cardiovascular health outcomes.

A first contribution in this area is due to Smits et al. (2004) that analysed the effects of SES on total and cause-specific mortality in the Netherlands in 1999, using data on all 15.8 million inhabitants. This paper focused on individual-level data from national health surveys to assess the impact of public health expenditure on mortality, and the main findings, despite the ones regarding the rural-urban disparities – studied further on this survey -, focused on the differences between regions' poverty levels. The authors found that *CVD death rates among high-poverty areas were higher than low-poverty areas, and that these findings reinforce the importance of prioritizing high-poverty* regions when developing and implementing cardiovascular health policies and programs.

Farahni et al. (2010) examined the effects of state-level public health spending on mortality probability in India – by aggregating individual-level data. They used data from the second National Family Health Survey (NFHS-2)[1], conducted in 1998/1999, covering 26 Indian states, and aim to address the fundamental question of whether public spending on health care improves health outcomes. To do so, they controlled for individual household, and state-level characteristics, and used a multilevel probit model to estimate mortality probability effects. During this research, the authors explored how public health spending interacted with other factors such as the ones listed above (gender, socioeconomic status, age). The key findings set the tone for this thesis: *a 10 percent increase in per capita public health spending (...) decreases the probability of death for an average individual by about 0.0004 (...) this is a reduction in mortality rates of about 2%; the effects of public health spending on mortality is predominately for women, the young*

[1] The National Family Health Survey-2 (NFHS-2) was a health, population and nutrition survey that involved around 91,000 women across 26 states of India. It offered national and state-level health and demographic data, as well as new measures on women's nutrition and anaemia. The survey was led by the International Institute for Population Sciences, and funded by the United Stated Agency for International Development, under the guidance of India's health ministry.

(under 18 years old) and the elderly (over 40 years old); and rural residence and household poverty were some of the major factors affecting mortality – mortality is about 50% higher among the poorest quintile.

Hagen et al. (2015) investigated whether the SES influences probability of death among cardiac patients. The study links microdata from Finnish and Norwegian national patient registers, combining treatment episodes SES indicators – income and education. The authors' econometric approach included a path analysis model, used to separate the direct effect of SES on mortality from the indirect effect through percutaneous coronary intervention (PCI) access; a panel data analysis, that accounts for background variables like age, gender, type of acute myocardial infarction (AMI), comorbidities, and travel distance to PCI facilities, and uses linear regression models to estimate the effects of SES on mortality and PCI provision. Regarding the findings, the study confirmed previous research showing a strong causality between higher mortality in AMI patients with lower income and education levels – *Of particular importance in this analysis is that most of the total effect on mortality are through the direct path for both income and education in both countries*. These results are largely associated with long-standing demographic divides in health and mortality that result from differences in diet, disease, clinical care and preventive care.

Recent studies like Birgisdóttir et al. (2020) analysed the effects of economic collapse and following crisis on the likelihood of ischemic heart disease (IHD) with data from Iceland following the 2008 crisis. The probability of a IHD event – death or hospital admission -, was estimated at an individual level, and analysed through regressors like the unemployment rate, economic collapse ad crisis periods' indicators, and individual-level mediators (such as age, gender, and real monthly income). This paper reached the conclusion that the probability of IHD increased in the long term for both males and females – *The results reveal that in the first twelve months the probability of an IHD event is 0.0052 percentage points higher than in the average twelve-month period in the sample, and for the latter twelve months the probability is 0.0080 percentage points higher*.

In the context of the region-level literature, researchers examined how wider economic conditions within regions or nations influence cardiovascular health outcomes across populations. By investigating regional-level economic potential determinants such as regional unemployment rates, average income levels, local educational status, and measures of income inequality, this set of research papers examines the complex relationship between geographically-specific economic circumstances and population cardiac health. The following papers demonstrate key contributions of how regional economic contexts can impact cardiovascular disease prevalence, incidence, and mortality rates.

The 2008 crisis and its impact on mortality trends has been a major topic of study for a few years now. Granados & Ionides (2017) investigated the waves caused by the Great Recession on mortality rates (age-specific and cause-specific) across all 27 European counties, using covariates like the unemployment rates, GDP per capita, and employment ratios. They concluded that an increase of 1 percentage point in the national unemployment rate is associated with a 0.5% reduction in age-adjusted mortality. This paper employs both linear and nonlinear panel regression models with effects and uses robust standard errors. In what concerns robustness tests, these results appeared to be consistent across different model specifications and business cycle indicators – like GDP per capita or employment-to-population ratios.

Kawachi et al. (2023), who provided an application for England and Wales, focused on the association of economic uncertainty with CVD mortality (evaluated according to weekly deaths). They have controlled for several economic indicators – like the unemployment rate, the GDP growth, and the consumer price index -, and considered country-level data when using models like OLS, probit, Poisson, or negative binomial. As anticipated, it was found that economic uncertainty is significantly correlated with short-term deaths from circulatory diseases – the coefficient between EPUI[2] and *Circulatory system diseases* was of 0.0111, and the one between EPUI and *Ischemic heart disease* 0.0117.

2.2. Regional Disparities and Environmental Considerations

From a different perspective, some very recent literature also focus on some demographic and spatial determinants, issues addressed in Atalay et al. (2023), that showed that coronary heart disease (CHD) is a leading cause of death for elder individuals, and that urban areas have being experiencing greater improvements in mortality outcomes

[2] Economic Policy Uncertainty Index.

compared to rural and remote areas – *health differentials exist not only between the rich and the poor (…) but also between urban and rural and remote areas* -, given that these latter face challenges such as higher disease burdens, fewer healthcare resources, and socioeconomic disadvantages – *policymakers must direct efforts to improve health inequalities (…) toward rural and regional areas*. This paper focused on the 3-year mortality rate in Australian regions and, to study it, used the percentile rank of each SA3 group[3], some year dummies, and interaction terms between the year and the rank. Sekkarie et al. (2024) investigated this a year later, when going through the rural-urban disparities in CVD mortality rates across different poverty levels and United States of America counties. When looking at the national findings, the results showed that the high-poverty rural areas had the highest CVD death rates (191 deaths per 100,000 inhabitants) and the largest disparity compared to low-poverty urban areas (where the rate ratio was of 1.76).

Environmental issues also influence CVD mortality rates, as Açiktepe et al. (2024) proved in their province-level analysis in Turkey – using data from 2010 to 2019. Factors like air pollution and temperature variations proved to be great contributors to the prevalence of these diseases, indicating that is reasonable to expect that a region with more environmental expenditure, less (rural) fires, and more protected areas has less propensity to display higher mortality rates by CVD. Included in the paper's discussion section, the authors even mentioned: *the study's findings are significant in understanding the complex interplay between environmental factors, lifestyle choices, and cardiovascular health*. During the same year, Khaltaev & Axelrod (2024) also investigated this matter, including all World Health Organization (WHO) member states when analysing the effects of factors like income level or type of air pollution (ambient or household) on age-standardized CVD mortality rates. The findings are in line to the ones mentioned previously and support the reasoning of studying this question in Portugal. In fact, in the section reserved for results, it is mentioned: *There is a gradual, statistically significant increase in CVD mortality (…) attributed to air pollution*; and: *IHD mortality associated with ambient air pollution is almost two to three times higher than mortality from stroke*. As most papers in this field, the authors close with a useful conclusion/contribution to policymakers: *Air*

[3] According to the Australian Bureau of Statistics, SA3s (Statistical Area Level 3) typically represent, in regional areas, the catchment area of larger cities with populations exceeding 20,000 inhabitants. In major metropolitan areas, SA3s generally correspond to the zones served by significant transportation and commercial centres.

pollution control should be an essential component of the preventive strategy along with lifestyle modifications and effective disease management.

2.3. The Portuguese Case

Over the past few years, several authors addressed issues related to mortality rates, cardiac and others, and the determinants of heart failure in Portugal. Gouveia et al. (2019) had the purpose of studying the burden of heart failure (HF) in Portugal's mainland in a 22-year horizon (2014-2036), focusing on the associated social and economic impacts. Their primary dependent variable was the burden of heart failure, measured in disability-adjusted life years, studied through the years of life lost due to premature death and to disability, and some population demographics – age, gender, population projections, and chronic heart failure class – and aging trends. The findings gave light to some relevant data: *the total number of deaths attributable to HF in 2014 was 4,688 (...), which corresponds to 4.7% of overall mortality in Portugal*; and *mortality due to HF is expected to increase by 73% in the next 20 years (with 8,112 deaths for HF in 2036) as a consequence of both HF mortality rate being higher in the older population and population ageing.* Heart failure mortality rates were also a subject of study for Marques-Alves et al. (2020) when trying to determine the characteristics of 1,024 individuals with acute decompensated heart failure (ADHF) admitted to Portugal's emergency rooms. Several HF mortality rates were considered and analysed via variables like age, gender, biomarkers, individual clinical diagnoses, and length of prior hospitalization. Using univariate and multivariate Cox proportional hazards models, and some statistical tests including Student's t-test, Chi-square or Fisher's exact tests, the authors concluded that the most affected by ADHF were older, male patients (*mean age was 78 [...], and 53% were male*); and that timely and appropriate care is critical in reducing hospital readmissions.

On the field of papers directly addressing (total) mortality rates, Costa & Santana (2021) aimed to analyse it in each Portuguese municipality for several time periods since 1990, having into account aspects like gender, age, and levels of socioeconomic deprivation. With the coefficients of the fitted regression, the main conclusion regarded the varying levels of mortality risk associated with deprivation: *this provides evidence that the* *association identified in previous studies between poverty and health outcomes in Portugal is weakening for younger age groups*. Education concerns also come to light with this paper, as they may *explain the increase in socioeconomic inequalities*, and there is *evidence that low health literacy is associated with (...) increased rates of hospitalization and longer hospitalization*. In fact, it is also mentioned that *one of the reasons why average incomes in Portugal are lower and poverty is high, affecting health outcomes* is Portugal's low education/training – especially when compared to European Union's (EU) values, and that *the turning point in socioeconomic inequalities in Portugal seems to be after compulsory education*.

Another approach is followed by Duarte et al. (2023), who analysed the relationship between fire-pollutant-atmospheric variables and health outcomes, specifically cardio-respiratory mortality rates, in Portugal, found differing associations between pollutants associated with fire exposure and various health conditions, calling for the consideration of socioeconomic and health behaviour variables. The study focused on regional-level analysis and used generalized additive models, linear regressions, and Principal Components Linear Regression. *Portugal is a highly fire-prone region and these release great quantities of pollutants to the atmosphere which are a risk factor for adverse cardiovascular outcomes*, which justify the use of data regarding the number of fires – rural or otherwise – in the study of cardiac mortality rates. Although this paper did not cover it, it was recommended that *including additional variables such as population density, socio-economic status, or other environmental factors that could potentially influence outcomes*, and that *factors such as age, gender, or other similar conditions may provide valuable insights.*

To conclude this section, it should be mentioned that, in 1982, was founded the Portuguese Journal of Cardiology, with the objective of ensuring that Portuguese cardiologists remained updated on various domains of cardiology. This year's publications included various articles of great interest on the topic in analysis in this thesis. January's edition comprised an article – Cabral et al. (2025) –, in which the aim was to analyse the economic impact of CVD in Portugal and suggest measures for improvement in cardiovascular care. By analysing patients with CVD, and taking into consideration factors such as healthcare expenditure, prevention measures, and patient demographics, the authors reached the conclusion that CVD has a significant contribution to Portugal's mortality figures. The findings also emphasize disparities in cardiovascular health outcomes based on demographic variables – age (*cardiovascular mortality reaches higher rates starting at age 75*), gender (*mortality from cardiovascular causes is higher in men),* and SES -, suggesting the need for interventions that address these inequalities. In February, Anjos (2025) included some "interesting, although not unexpected" ideas on how the patients' higher education correlates with better knowledge on their specific heart defect – *Patient education is an important and integral part of this concept and should be undertaken through a multidisciplinary approach, adapted to the patient's and family's educational level, culture, and age, among others.* -, and how empowering patients can lead to better outcomes. Given this, it becomes easy to deduct that, if a patient (or region) has a higher-level education, it is expected that the consequences for their health – cardiac and general – will be more positive, being even possible that the mortality rate will be lower.

3. METHODOLOGY

Consider a random sample, of $i = 1, \dots, N$ individuals, and let y_i be the fractional variable of interest, constrained in the interval $[0,1]$. The vector x_i represents a set of $1 \times K$ covariates and is denoted as $x_i = (x_1, x_2, \dots, x_k)$. $x_i'\beta$ is the linear predictor, where β be the $(K+1) \times 1$ vector of parameters to be estimated. The conditional density of y_i is denoted by $f(y_i|x_i,\beta)$, which may or may not be specified *a priori*. This framework allows to analyse the relationship between various characteristics, represented by the co-variates, and the fractional outcome variable of interest across the sampled individuals.

3.1. Linear Regression Model

Linear regression is a commonly used technique that quantifies the associations between predictor and outcome variables. The standard specification is denoted as:

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u_i \tag{1}$$

, where u_i is the error term. Assuming that $E[u_i|x_1, \dots, x_k] = 0$, it is possible to derive:

$$E[y_i|x_1, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \tag{2}$$

Linear regression models are estimated by the ordinary least squares (OLS) method. In fact, OLS estimators hold several crucial properties as they are consistent, unbiased, and efficient, exhibiting the smallest variance among all unbiased linear estimators.

3.2. Nonlinear Conditional Mean Regression Models

In the study of fractional response variables (considering *heartmort* as a proportion obtained by dividing the original values by 1,000), nonlinear conditional mean models are the prevalent methodological approach. These models are typically introduced as:

$$E[y_i|x_i] = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(x_i'\beta) \quad (3)$$

, where $G(\cdot)$ is a known cumulative distribution function, bounded by $[0,1]$.

This approach was pioneered by Papke and Wooldridge (1996), who proposed that *typically, $G(\cdot)$ is chosen to be a cumulative distribution function*. This selection ensures that predicted values fall within the $[0,1]$ range, aligning with the nature of fractional responses. Papke and Wooldridge (1996)'s framework has become the standard convention for handling ordinary fractional responses, offering a method that both accommodates the bounded nature of the dependent variable, in contrast to the linear regression model, and provides a useful approach for modelling proportions and rates across various data scenarios.

In the context of fractional response models, where the dependent variable, y_i, is bounded by $[0,1]$, it can be relevant to also retrieve the partial effects (PE) of the covariates. These values offer vital insights into the marginal impact of explanatory variables on the response. The derivation of these effects is based on the following assumption:

$$\Delta x_{ij} = k \Rightarrow \Delta E[y_i|x_i] = k \times \beta_j g(x_i'\beta) \quad (4)$$

, where $g(\cdot)$ represents the probability density function of each model. (4) describes the changes of the conditional mean relative to changes in a predictor. Within this framework, the sign of PE consistently aligns with the corresponding estimate for $\hat{\beta}_j$. However, it is important to note that the precise formulation and interpretation of PE depends on the nature of the explanatory variables. The subsequent analysis will explain these different approaches, providing a comprehensive understanding of how different types of covariates influence the fractional response.

3.2.1. Alternative Fractional Regression Models

Given the fractional nature of the data, this research explores a range of nonlinear conditional mean models, focusing specifically on three: logit, probit, and complementary log-log (cloglog) models. These models are widely recognized for their effectiveness in handling bounded continuous outcomes. Drawing from the work of Ramalho et al. (2011), the following table defines and details the distribution functions for these models. This approach allows for a comprehensive examination of how these different models can be applied to fractional response data, providing insights into their respective strengths and characteristics in modelling bounded continuous variables.

Model	Distribution Function	Distribution Symmetry	Tail Behaviour
Logit	Standard Normal	Symmetric	Heavy tails
Probit	Extreme maximum	Symmetric	Thin tails
Cloglog	Extreme minimum	Asymmetric	Right-skewed

Table 1: Nonlinear Conditional Mean Models for Fractional Data. Self-elaborated.

In what regards the fractional logit model, originally formulated using quasi-maximum likelihood (QML) estimation, it is designed to model the conditional probability of a positive outcome given a set of regressors. The fractional probit model offers an alternative symmetric specification. Finally, the complementary log-log model offers an alternative asymmetric functional form for the conditional mean.

Across all three models, the sign of the PE aligns with the one of the corresponding coefficient estimates. However, their specific formulation varies depending on whether the explanatory variable is continuous or binary. For continuous variables, PEs represent the marginal impact in the expected outcome, while for dummy variables, they quantify the change in the expected outcome when the binary predictor shifts from 0 to 1, holding all other variables constant (in a *ceteris paribus* framework).

The following table presents the conditional density function (CDF), probability density function (PDF), and the PE of each alternative fractional regression model. For subsequent analysis, let us consider $z = x_i'\beta$.

Model	$G(\cdot)$	$g(\cdot)$	$\widehat{PE}_{ij}$ Cont.	$\widehat{PE}_{ij}$ Dummy
Fractional Logit	$\Lambda(z) = \frac{e^z}{1+e^z}$	$\lambda(z) =$ $= \Lambda(z)[1-\Lambda(z)]$ $= \frac{e^z}{(1+e^z)^2}$	$\hat{\beta}_j \frac{\partial\Lambda(x_i'\hat{\beta})}{\partial x_{ij}} =$ $= \hat{\beta}_j \frac{e^{x_i'\hat{\beta}}}{(1+e^{x_i'\hat{\beta}})^2}$	$\Lambda\left(x'_{i,x_j=1}\hat{\beta}\right)$ $-\Lambda\left(x'_{i,x_j=0}\hat{\beta}\right)$
Fractional Probit	$\Phi(z) =$ $= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}} dz$	$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}$	$\hat{\beta}_j \frac{\partial\Phi(x_i'\hat{\beta})}{\partial x_{ij}} =$ $= \hat{\beta}_j \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i'\hat{\beta})^2}{2}}$	$\Phi(x'_{i,x_j=1}\hat{\beta})$ $-\Phi(x'_{i,x_j=0}\hat{\beta})$
Fractional Cloglog	$C(z) = 1 - e^{-e^z}$	$c(z) = e^{z-e^z}$	$\hat{\beta}_j \frac{\partial C(x_i'\hat{\beta})}{\partial x_{ij}} =$ $= \hat{\beta}_j e^{x_i'\hat{\beta}-e^{x_i'\hat{\beta}}}$	$C(x'_{i,x_j=1}\hat{\beta})$ $-C(x'_{i,x_j=0}\hat{\beta})$

Table 2: CDF, PDF, and Partial Effects ─ for continuous and dummy explanatory variables, respectively ─ of the Alternative Fractional Regression Models. Self-elaborated.

3.2.2. Estimation Method

Fractional response models share similarities with binary regression models in their analytical framework. Both model types employ identical expressions for calculating PE, highlighting their methodological similarity. Moreover, the estimation process for both is based on the Bernoulli log-likelihood (LL) function, further emphasizing these structural parallels.

However, a key distinction lies in the estimation technique employed for fractional models. These models utilize QML estimation, which enables robust estimation without imposing strict distributional assumptions on the dependent variable, accommodating the unique characteristics of fractional data. The Bernoulli log-likelihood function is expressed as:

$$LL_i(\beta) = \sum_{i=1}^{N} \{y_i \ln[G(x_i'\beta)] + (1-y_i)\ln[1-G(x_i'\beta)]\} \tag{5}$$

The QML Bernoulli estimator of the parameter vector β, denoted by $\hat{\beta}$, is obtained by the maximization:

$$\hat{\beta} = arg \max_{\beta} \sum_{i=1}^{N} LL_i(\beta) \quad (6)$$

As proved by Papke and Wooldridge (1996), $\hat{\beta}$ is *consistent and* $\sqrt{N}$*-asymptotically normal* regardless *of the distribution of* y_i *conditional on* x_i. This property ensures the estimator's reliability across various data distributions.

3.3. Beta Fractional Regression Model

To achieve more efficient estimators, it can be advantageous to specify not only the conditional mean $E[y_i|x_i]$ as expressed in (1), but also the conditional distribution $f(y_i|x_i)$. This approach leads to the use of beta regression, which assumes $y_i \sim Beta$ and has a precision parameter ϕ, besides the parameters included in the conditional mean. The estimation in this case is performed exclusively through maximum likelihood (ML), a method that, while less robust, because requires the specification of the data distribution, relative to QML, delivers efficient estimators. However, the beta distributional assumption comes with certain limitations: firstly, it is only defined for the open interval]0,1[, making it impossible to accommodate extreme values of 0 or 1; secondly, as it does not belong to the linear exponential family, the resulting estimators lack robustness when faced with deviations from the assumed distribution. These considerations highlight the trade-offs between efficiency and flexibility in modelling fractional response data.

The beta density function is given by:

$$f(y_i; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \, y_i^{p-1} \, (1-y_i)^{q-1} \quad (7)$$

, where $\Gamma(\cdot)$ denotes the gamma function, $y_i \in]0,1[$, and $p > 0$ and $q > 0$ are the shape parameters exponentially derived from the covariates. With this specification:

$$E[y_i] = \frac{p}{p+q} = \mu \quad (8)$$

$$Var(y_i) = \frac{pq}{(p+q)^2 \, (p+q+1)} = \frac{\mu \, (1-\mu)}{\phi + 1} \quad (9)$$

, where μ is the mean of y_i and ϕ represents a precision parameter.

3.4. Specification Testing

Specification tests are crucial in econometric analysis to document the validity and reliability of model estimates. These tests help identify potential misspecifications, such as omitted variables and incorrect functional form which could lead to biased estimates.

In the case of the RESET test, it evaluates model misspecification by testing whether nonlinear combinations of fitted values significantly improve the model's explanatory power. This test starts with the estimation of the original model and obtention of the respective fitted values of the index, followed by the creation of new variables – by raising those fitted values to higher powers (usually $(X\hat{\beta})^2$ and $(X\hat{\beta})^3$) – and addition of the mentioned to the original model as new regressors to create an auxiliary model. The null and alternative hypothesis, tested by a Wald statistic, are:

$$H_0: \gamma_1 = \gamma_2 = \gamma_3 = \cdots = \gamma_p = 0 \rightarrow Suitable\ model\ functional\ form \tag{10}$$

$$H_1: \exists_j \gamma_j \neq 0, j = 1, 2, \ldots, p \rightarrow Unsuitable\ model\ functional\ form \tag{11}$$

Under H_0, this test follows a $\chi^2_{(p)}$ distribution, where p corresponds to the number of new regressors added to the original model (usually 2). This test provides valuable insights into the adequacy of the model's functional form, helping researchers identify potential improvements or necessary adjustments to their econometric specifications.

4. DATA DESCRIPTION

This section provides a comprehensive overview of the dataset used in this thesis. First, summary statistics, including measures of central tendency and dispersion to give an overall picture of data characteristics, are analysed. Second, histograms visually represent the distribution of important variables, allowing for an intuitive understand of their shapes and potential outliers. Finally, Pearson correlation coefficients examine the strength and direction of linear relationships between variables. Together, these tools provide a solid foundation for the empirical analysis that follows, ensuring a clear understanding of the data's structure and relationships.

4.1. Dataset and Variables

The data for this analysis were primarily sourced from INE and Pordata, Portugal's database of contemporary information. Additionally, the variables denoted as *metro* and *coastline*, concerning information on the municipalities located in a metropolitan area (either Lisbon's – LMA – or Porto's – PMA) and on Portugal's coastline (municipalities bathed by the Atlantic Ocean or the Mediterranean Sea, with the exceptions of 10 municipalities that are located in either Sado's or Tagus' estuaries – Alcácer do Sal, Alcochete, Barreiro, Benavente, Lisbon, Loures, Moita, Montijo, Seixal, and Vila Franca de Xira) respectively, were created by the author based on geographical knowledge of Portugal's municipalities. Moreover, the variable designated as *refcentre* was constructed by the author using information published in March 2024 by Direção Geral da Saúde, the government's health authority managed by the Portuguese Ministry of Health, regarding the location of type A[4] reference centres for cardiology and paediatric cardiology (in total, 6 and 4, respectively, according to Serviço Nacional de Saúde (2024)). Further details on each variable can be found in table 6 of the Appendix.

To facilitate a more structured analysis, the explanatory variables in this study were aggregated into categories based on their nature. Economic variables were grouped to include unemployment rate, PPI, environmental expenditure, and expenses in sports activities. Demographic variables encompassed illiteracy rate, population aged 65 and over, female population percentage, and population density. Health infrastructure variables consisted of hospitals per 1,000 inhabitants, doctors per 1,000 inhabitants, and the existence of recognized cardiology reference centres. Geographic and environmental variables included the natural protected areas, number of rural fires, and coastline location. Finally, urban and transportation variables were represented by long commute times (60 minutes or more) and metropolitan area status. This labelling allows a clearer examination of how different characteristics may influence cardiovascular mortality rates.

Notice that while variables such as unemployment rate (*unem*), PPI (*purpower*), population aged 65 and over (*pop65*), female population percentage (*fem*), hospitals per 1,000

4 The Type A reference centres are the ones that deal with structural intervention.

inhabitants (*hosp*), and metropolitan area status (*metro*) are standard in the literature explaining the mortality rate from diseases of the circulatory system, variables such as environmental expenditure (*environ*), population density (*popdens*), or number of rural fires (*fire*) have been rarely used. Some new potential determinants considered in this thesis are the expenses in *sports* activities, *illiteracy* rate, *doctors* per 1,000 inhabitants, existence of recognized cardiology reference centres (*refcentre*), natural protected areas (*natura*), *coastline* location, and long commute times (*commute60*).

4.2. Variable Description

To gain deeper insights into the available variables, an examination of their descriptive statistics[5] is essential. Using STATA, key measures were calculated, including the mean, standard deviation, minimum, and maximum values for each variable. Visual representation of our variables through histograms also provides relevant insights into the core data distributions, complementing and expanding the summary statistics. This analysis allows a more complete understanding of patterns, potential skewness, and irregularities.

Starting with the dependent variable, the mortality rate from diseases of the circulatory system per thousand, *heartmort*, the analysis reveals a mean mortality rate of 4.16‰, with a standard deviation of 0.0172. These statistics, obtained from INE, indicate that Portugal exhibits a slightly higher average mortality rate from circulatory system diseases than the EU. According to a statistics' highlight made by the European Union (2024), *there were 1.71 million deaths in the EU from diseases of the circulatory system*, and this was the leading cause of death in 2021. Considering the 443.2 million EU residents during the year in analysis – European Union (2023) data -, this mortality rate was of approximately 3.86‰. Portugal displays a high variability in the data. This dispersion is further evidenced by the wide range between the minimum observed value of 0‰ and the maximum of 12.4‰ (in Corvo and Santa Cruz das Flores – both in Azores archipelago -, and Gavião, respectively). Analysis of the *heartmort* histogram, as depicted in Figure 1, reveals a right-skewed distribution. The majority of data points cluster between 2 and 7 per

[5] All summary statistics can be found in table 7 of the Appendix.

thousand inhabitants. However, there are some outliers between 10 and 15 per thousand, which indicated the existence of specific regions with higher mortality rates from cardiac diseases. This asymmetry suggests heterogeneity in circulatory disease outcomes across different municipalities of Portugal, indicating areas that require focus on prevention and treatment plans.

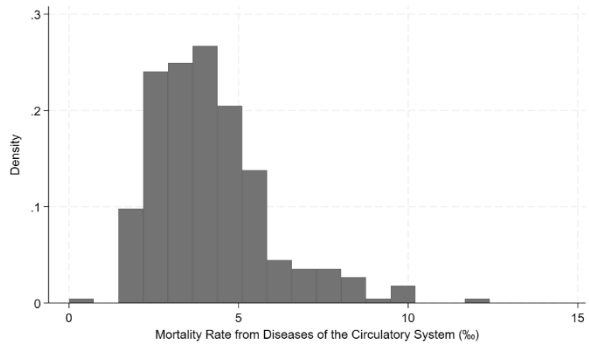

Figure 1: Dependent variable's (*heartmort*) histogram. Self-elaborated in STATA.

Examination of economic variables across Portuguese municipalities reveals significant disparities. Unemployment rates average around 7%, aligning with EU trends – according to O'Neill (2024) – but showing regional variations. PPI per capita averages 82.26%, with substantial differences between municipalities. Environmental and sports-related spending exhibit high variability, reflecting diverse local priorities and resources. Data distributions – represented in figure 2 – for these variables are generally right-skewed, indicating concentrations at lower values with extended right tails. These patterns suggest economic inequalities and varying resource allocations among municipalities, potentially impacting health outcomes, particularly cardiovascular disease mortality rates. Analysis underscores the need for targeted interventions to address regional economic imbalances and promote more equitable development across Portugal.

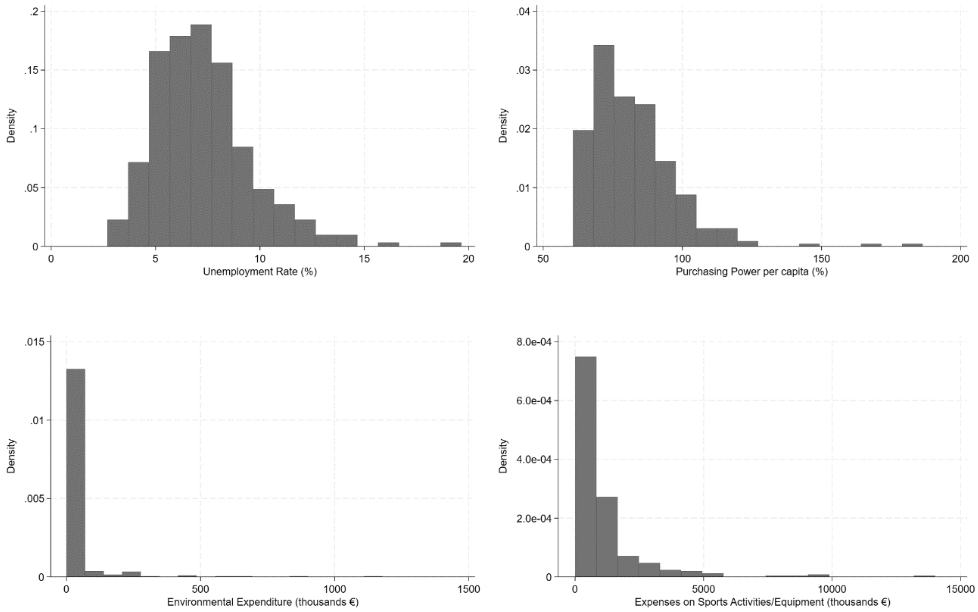

Figure 2: Economic variables' (*unem*, *purpower*, *environ*, and *sports*) histograms. Self-elaborated in STATA.

Demographic analysis of Portuguese municipalities reveals heterogeneity across several indicators. Illiteracy rates average 4.85%, ranging from 1.41% to 12.26%. The population aged 65 and older is 11% to 48%, with a mean approaching 30%. Females constitute an average of 52% of the population, with minimal variation across regions. Population density shows extreme disparities, averaging 295.28 inhabitants per square kilometre but ranging from 4.4 to over 7,200. Data distributions for these variables are generally right-skewed, as figure 3 displays, particularly for *illiteracy* and population density, indicating concentrations at lower values with extended right tails. The aging population and female percentage distributions are more symmetrical. These patterns reflect ongoing demographic transitions, including rural exodus and population aging, which contribute to regional disparities. Such demographic variations across municipalities highlight the need for tailored policies addressing education, healthcare, and urban planning to manage diverse population needs and potential impacts on health outcomes, including cardiovascular disease mortality rates.

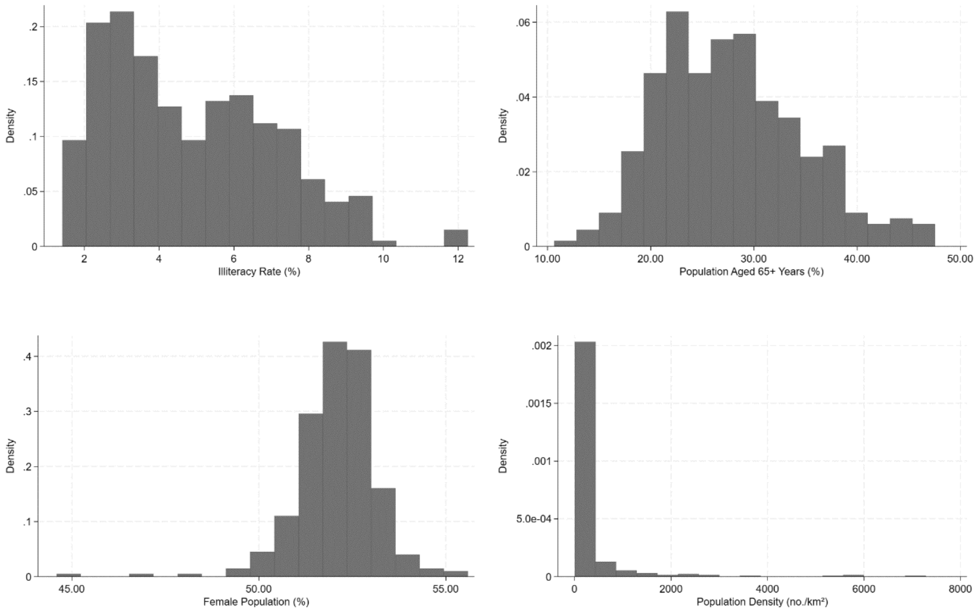

Figure 3: Demographic variables' (*illiteracy*, *pop65*, *fem*, and *popdens*) histograms. Self-elaborated in STATA.

Analysis of health infrastructure variables across Portuguese municipalities reveals disparities in resource distribution. Hospital density averages 0.015 per 1,000 inhabitants, aligning with European norms – as data from Degenhard (2024) and European Union (2023) prove -, but shows high variability (with the standard deviation exceeding the mean). Doctor-to-population (1,000 inhabitants) ratios average 2.940, with substantial variation across municipalities (ranging from 0.490 to 34.300). Only 1.3% of municipalities (4 out of 308 – Coimbra, Lisbon, Porto, and Vila Nova de Gaia) possess specialized cardiology reference centres, as the country only has 10 centres of such area. Distributions (found in figure 4) for hospitals and doctors per 1,000 inhabitants are markedly right-skewed, indicating concentrations at lower values with extended right tails. This pattern suggests considerable inequalities in healthcare access and resource allocation among municipalities. These disparities in health infrastructure across Portugal highlight potential impacts on healthcare access, quality, and outcomes, particularly concerning cardiovascular disease mortality rates, and underscore the need for targeted interventions to address regional imbalances.

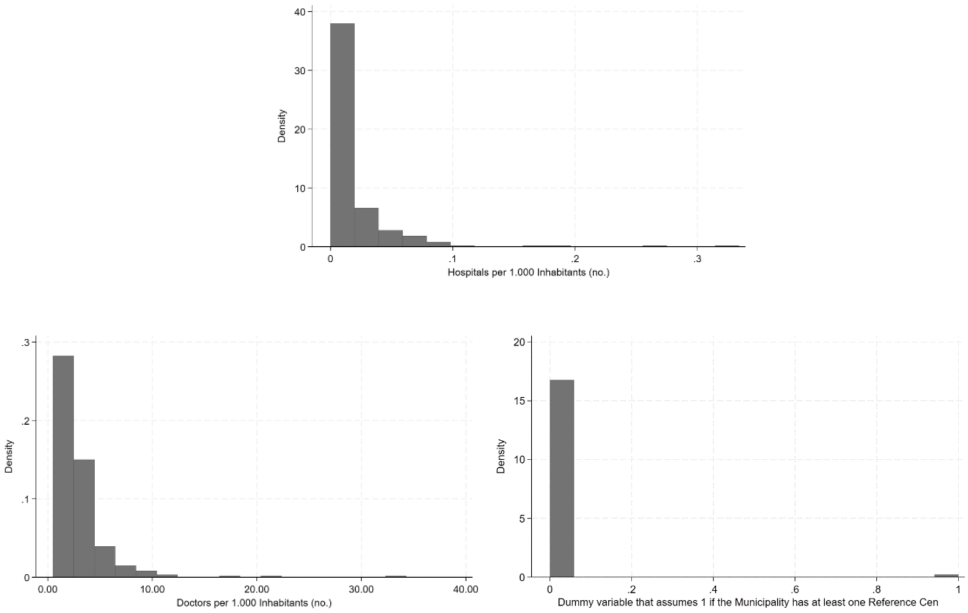

Figure 4: Health infrastructure variables' (*hosp*, *doctors*, and *refcentre*) histograms. Self-elaborated in STATA.

Analysis of geographical and environmental variables across Portuguese municipalities shows that protected natural habitats (Natura 2000 Network) range from 0 to 78,161 hectares, with a mean of 6,291 hectares. Rural fire incidence in 2021 varied widely, from 0 to 202 fires per municipality, averaging 27 fires. Approximately 30% of municipalities are coastal. Distributions for protected areas and rural fires are highly right-skewed, as figure 5 displays, indicating concentrations at lower values with extended right tails. The coastline variable, being binary, shows a clear asymmetry with most municipalities situated inland. These patterns highlight concentrated nature conservation efforts, geographically focused fire risks, and an uneven coastal-inland distribution. Such geographical and environmental disparities across municipalities may significantly impact local health conditions, development policies, and resource allocation, potentially influencing health outcomes including cardiovascular disease mortality rates.

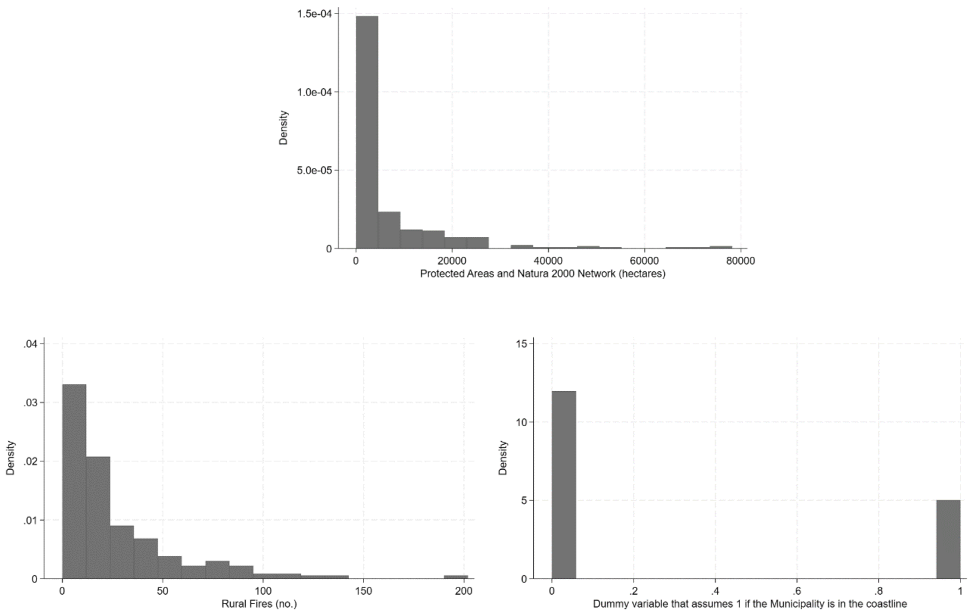

Figure 5: Geographic/environmental variables' (*natura*, *fire*, and *coastline*) histograms. Self-elaborated in STATA.

Analysis of urban and transportation variables across Portuguese municipalities reveals significant disparities. On average, 3.13% of the population experiences commute times of 60 minutes or more, ranging from 0% to 16%. This positions Portugal as the 5th country with the shortest average commute times in the EU, according to Yanatma (2024). Only 11.36% of municipalities (35 out of 308) are classified as part of metropolitan areas, split between LMA (18) and PMA (17). The distribution of long commute times is right-skewed – as concluded from figure 6's analysis -, with most municipalities having less than 10% of residents facing extended travel times. The metropolitan status variable is highly skewed, reflecting the concentrated nature of Portugal's major urban centres. These patterns highlight regional variations in accessibility, transportation infrastructure, and urban development across the country. Such disparities in urban and transportation factors may significantly impact socioeconomic outcomes, including health and quality of life, and underscore the need for targeted policies to address regional imbalances in urban planning and resource allocation.

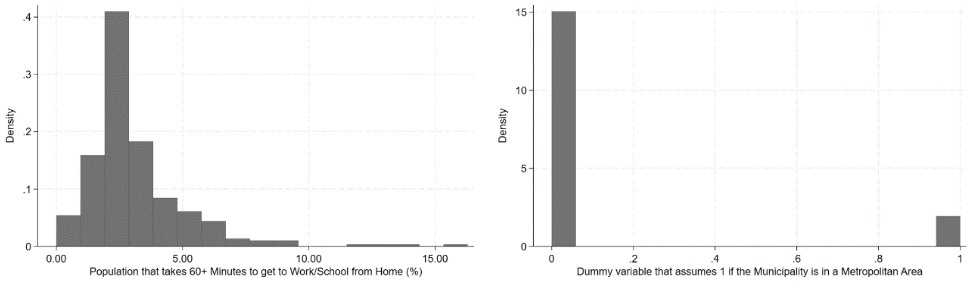

Figure 6: Urban/transportation variables' (*commute60* and *metro*) histograms. Self-elaborated in STATA.

4.3. Relationships Between Variables

Having examined individual variables, the next step is to look for relationships between them. This examination is crucial to understand the patterns and interactions that cannot be observed when considering each variable by itself. By analysing these correlations, it is possible to achieve a better understanding of the contributions of each variable related to the differences in heart disease mortality rates in Portuguese municipalities. So, this analysis provides a strong basis for the econometric modelling approach.

Starting with Pearson correlation coefficients[6], that measure linear relationships among variables, the analysis of the results involving the dependent variable reveals several significant relationships. While *fem* and *commute60* show no statistically significant correlations, all other variables demonstrate significant associations. *Illiteracy*, *natura*, and *pop65* exhibit positive correlations with *heartmort*, while variables such as *hosp*, *environ*, *popdens*, *unem*, *natura*, *fire*, and *doctors* display weak correlations but need further analysis due to potential nonlinear relationships. Strong correlations are also observed between pairs of independent variables, as between *popdens* and *purpower*, *illiteracy* and *pop65*, and *purpower* with both *doctors* and *sports*. The only strong negative correlation is between *illiteracy* and *purpower*. This analysis provides insights into the complex relationships between variables influencing heart disease mortality rates in Portuguese municipalities, setting the stage for more in-depth statistical modelling.

[6] Results for Pearson's correlation coefficient can be found in table 8 of the Appendix.

To further examine the relationships between key variables and heart disease mortality rates, boxplots were created – and displayed in figure 7 of the Appendix. These visualizations illustrate how various factors distribute across different mortality rate groups (“low”, “medium”, and “high”). On the economic variables’ category, the boxplots show a decreasing trend in the variability of *unem* and *purpower*’s values, as the categories progress from “low” to “high”, with some outliers presented in each group, meaning that, as the unemployment rate and PPI increase, the mortality rate from heart diseases in each municipality tends to decrease. The figures reveal a positive relationship between the demographic variables *pop65* and *fem*, and heart diseases mortality rates, given that, as mortality groups progress from “low” to “high”, both the median and variability of these variables increase, suggesting higher elderly populations correspond to higher mortality rates, and so do female populations. The number of hospitals per 1,000 inhabitants’ plot indicates a negative relationship, meaning that, as *hosp* values increase, the *heartmort* values tend to be lower. The *metro* graph shows that areas in the “low” category are predominantly metropolitan, while “medium” and “high” classes are mostly non-metropolitan, with a single outlier metropolitan area in the centre category.

5. EMPIRICAL RESULTS

5.1. Regression Results and Model Selection

This chapter reports the results of the regression models selected from several experiments involving different combinations of explanatory variables. Table 3 includes the estimates of the models’ coefficients as well as some selection criteria and tests.

Having the literature review and relevant descriptive statistics into account, the starting point of the estimation process was the linear regression model by OLS. This model, despite validated by the RESET test, revealed that only six of the initial sixteen explanatory variables were statistically significant at the 10% significance level. After performing a Wald test for significance of the individually insignificant variables, and concluding these were not jointly significant, the decision was to remove those variables from the model. After doing so, the restricted model only presented a statistical insignificant variable (*hosp*), but continued presenting a valid functional form, according to the RESET test. Even though the model presented a suitable functional form, it is known that the linear regression imposes an inappropriate structure on rate data, as it ignores its boundaries and proportional relationships that can lead to meaningless negative and larger than 1 outcomes.

All alternative fractional regression models considered – logit, probit, and cloglog – were validated by the RESET test. Nevertheless, there were seven independent variables that did not display any statistical significance (neither individual nor jointly) in all three models. After the removal of such variables from the regression, not only the RESET test kept displaying encouraging results, but also the selection criteria (Akaike Information Criterion – AIC – and Bayesian Information Criterion – BIC) revealed significant decreases, indicating that the restricted models would be a better fit for the data.

Finally, a beta model was estimated using logit as the default link function. The RESET test, in this case, indicated an unsuitable model functional form. These results are not so surprising, as this regression uses ML estimation, that requires the exact specification of the data distribution, making it more vulnerable to distributional misspecification errors and leading to inconsistent estimators.

Given so, the chosen and analysed models will be the ones treated as "alternative": fractional logit, probit and complementary log-log. The following sections will regard and dive deeper into these.

Variables	Linear				Frac. Logit			
	Unrestricted		Restricted		Unrestricted		Restricted	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
metro	0.0000	0.0003	-	-	-0.0493	0.0434	-	-
hosp	-0.0038 *	0.0020	-0.0032	0.0019	-0.8300 **	0.3763	-0.6919 *	0.3860
environ	0.0000	0.0000	-	-	-0.0001	0.0001	-	-
popdens	0.0000	0.0000	-	-	0.0000 **	0.0000	-0.0001 ***	0.0000
unem	0.0000	0.0000	-	-	-0.0090	0.0073	-	-
illiteracy	0.0003 ***	0.0000	0.0002 ***	0.0000	0.0566 ***	0.0119	0.0507 ***	0.0112
natura	0.0000	0.0000	-	-	0.0000	0.0000	-	-
fire	0.0000 *	0.0000	0.0000 **	0.0000	-0.0013 ***	0.0004	-0.0011 ***	0.0004
purpower	0.0000 ***	0.0000	0.0000 ***	0.0000	0.0070 ***	0.0016	0.0065 ***	0.0014
pop65	0.0001 ***	0.0000	0.0001 ***	0.0000	0.0266 ***	0.0042	0.0303 ***	0.0038
fem	0.0001	0.0001	-	-	0.0223	0.0173	-	-
doctors	-0.0001 *	0.0000	-0.0001 *	0.0000	-0.0240 ***	0.0090	-0.0211 **	0.0083

coastline	0.0000	0.0002	-	-	-0.0234	0.0334	-	-
sports	0.0000	0.0000	-	-	0.0000 **	0.0000	0.0000 ***	0.0000
refcentre	0.0012	0.0008	-	-	0.4471 ***	0.1618	0.3862 **	0.1533
commute60	0.0000	0.0000	-	-	0.0088	0.0065	-	-
Constant	-0.0044	0.0035	-0.0017 ***	0.0006	-8.0755 ***	0.9431	-7.0015 ***	0.1567
N	308				308			
R²	0.6239		0.6112		-			
Pseudo- R²	0.6032		0.6035		-			
AIC	-				0.1556		0.1101	
BIC	-				-1,667.388		-1,707.497	
RESET Test								
Test Statistic	1.84		0.97		2.55		3.34	
p-value	0.1396		0.4074		0.2799		0.1882	

Variables	Frac. Probit				Frac. Cloglog				Unrestricted Beta (with Logit)	
	Unrestricted		Restricted		Unrestricted		Restricted			
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
metro	-0.0144	0.0141	-	-	-0.0494	0.0434	-	-	-0.0547	0.0732
hosp	-0.2827 **	0.1259	-0.2382 *	0.1289	-0.8281 **	0.3756	-0.6901 *	0.3850	-0.7299 *	0.4315
environ	0.0000	0.0000	-	-	-0.0001	0.0001	-	-	-0.0001	0.0002
popdens	0.0000 **	0.0000	0.0000 ***	0.0000	0.0000 **	0.0000	-0.0001 ***	0.0000	-0.0001	0.0000
unem	-0.0033	0.0024	-	-	-0.0090	0.0073	-	-	-0.0083	0.0071
illiteracy	0.0194 ***	0.0041	0.0173 ***	0.0038	0.0565 ***	0.0119	0.0506 ***	0.0112	0.0559 ***	0.0103
natura	0.0000	0.0000	-	-	0.0000	0.0000	-	-	0.0000	0.0000
fire	-0.0004 ***	0.0001	-0.0004 ***	0.0001	-0.0013 ***	0.0004	-0.0011 ***	0.0004	-0.0010 *	0.0006
purpower	0.0023 ***	0.0005	0.0021 ***	0.0005	0.0070 ***	0.0016	0.0065 ***	0.0014	0.0078 ***	0.0017
pop65	0.0092 ***	0.0014	0.0104 ***	0.0013	0.0266 ***	0.0041	0.0302 ***	0.0038	0.0262 ***	0.0033
fem	0.0072	0.0060	-	-	0.0223	0.0173	-	-	0.0410 ***	0.0149
doctors	-0.0078 ***	0.0030	-0.0069 **	0.0027	-0.0240 ***	0.0090	-0.0211 **	0.0083	-0.0287 ***	0.0094
coastline	-0.0065	0.0111	-	-	-0.0235	0.0333	-	-	-0.0206	0.0404
sports	0.0000 **	0.0000	0.0000 ***	0.0000	0.0000 **	0.0000	0.0000 ***	0.0000	0.0000	0.0000
refcentre	0.1431 ***	0.0569	0.1232 **	0.0496	0.4466 ***	0.1616	0.3858 **	0.1531	0.4557 **	0.1911
commute60	0.0030	0.0021	-	-	0.0088	0.0065	-	-	0.0115	0.0080
Constant	-3.4965 ***	0.3241	-3.1546 ***	0.0523	-8.0736 ***	0.9408	-7.0001 ***	0.1565	-9.1120 ***	0.7582
Scale Cons.	-								8.3230 ***	0.0810
N	308				308				308	
AIC	0.1556		0.1101		0.1556		0.1101		-	
BIC	-1,667.388		-1,707.497		-1,667.387		-1,707.497		-	
RESET Test										
Test Statistic	2.19		2.73		2.56		3.36		12.01	

p-value	0.3347	0.2553	0.2786	0.1867	0.0025

Legend: * p-value<0.1 ; ** p-value<0.05 ; *** p-value<0.01

Table 3: Regression Results: Linear, Fractional Logit, Fractional Probit, Fractional Cloglog, and Beta Models. Self-elaborated using STATA outputs.

5.2. Selected Models Interpretation

As noticed in table 3, for all restricted fractional logit, probit, and complementary log-log models, most variables are statistically significant at 1% level. Only *hosp*, and *doctors* and *refcentre* are significant at 10% and 5% levels, respectively.

The equations for the fractional logit, probit, and cloglog, that make clear that all variables display the same sign for the models under study, are, respectively:

$$\Lambda(z) = \Lambda(-7.0015 - 0.6919\, hosp_i - 0.0001\, popdens_i + 0.0507\, illiteracy_i - 0.0011\, fire_i + 0.0065\, purpower_i + \\ + 0.0303\, pop65_i - 0.0211\, doctors_i - 0.00004\, sports_i + 0.3862\, refcenter_i) \quad (12)$$

$$\Phi(\text{z}) = \Phi(-3.1546 - 0.2382\, hosp_i - 0.00002\, popdens_i + 0.0173\, illiteracy_i - 0.0004\, fire_i + 0.0021\, purpower_i + \\ + 0.0104\, pop65_i - 0.0069\, doctors_i - 0.00001\, sports_i + 0.1232\, refcenter_i) \quad (13)$$

$$C(z) = C(-7.0001 - 0.6901\, hosp_i - 0.0001\, popdens_i + 0.0506\, illiteracy_i - 0.0011\, fire_i + \\ + 0.0065\, purpower_i + 0.0302\, pop65_i - 0.0211\, doctors_i - 0.00003\, sports_i + 0.3858\, refcenter_i) \quad (14)$$

Having more hospitals per 1,000 inhabitants creates better conditions to treat any eventual heart matter that may arise and may avoid serious health complications or even prevent the death of the patients. Hence, not surprisingly, this covariate has a negative impact (*ceteris paribus*) in the mortality rate from diseases of the circulatory system, being this is the independent variable with the higher absolute impact on the response variable.

Municipalities with higher population density in Portugal are usually integrated in urban areas, and tend to have better access to healthcare, less response times from emergency services, and more developed socioeconomic factors. These considerations justify the negative sign displayed by *popdens*, that indicates that the mortality rate from diseases of the circulatory system tends to decrease as the number of individuals per squared kilometre increases (*ceteris paribus*).

Considering “illiteracy” as an issue that regards individuals with 10 or more years old that cannot read nor write (is not capable of reading and understanding a written sentence not of writing one), this problem has great impact in several other areas and situations. The ability to comprehend medical instructions, recognize warning signs of (heart) diseases, or follow complex treatment plans is highly harmed. Because of that, and because there tends to be a reduced preventive care or poorer health management among this niche of the population, the sign of *illiteracy* was the first appearing positive, meaning that the mortality rate from diseases of the circulatory system tends to increase as the illiteracy rate of the municipalities increases (*ceteris paribus*).

Surprisingly, the sign of *fire* showed negative, indicating that, as the number of rural fires increases, in a *ceteris paribus* framework, the mortality rate from diseases of the circulatory system tends to be lower.

Another unexpected relationship emerged from the sign of *purpower*. According to the data, as the PPI per capita increased (*ceteris paribus*), the mortality rate from diseases of the circulatory system is likely to also increase. This complex bond can be (remotely) explained by some lifestyle changes or work-related stress that may occur from having higher affluence, indicating that prosperity alone does not guarantee better cardiovascular health without appropriate lifestyle and policy interventions.

Returning to the expected relationships, the increase of population aged 65 or more years old (*ceteris paribus*) also increases the mortality rate from diseases of the circulatory system. Elder people tend to be more susceptible to cardiovascular pathologies, as they have higher cumulative exposure, faster disease progression, comorbidities, and reduced physiological reserves. These factors create a straightforward explanation for the positive sign of *pop65*’s coefficients.

Similarly to *hosp*’s reasoning, having more doctors per 1,000 inhabitants improves healthcare timing and intervention. A negative sign in *doctors* underlines the anticipated association: *ceteris paribus*, as the number of doctors per 1,000 inhabitants increases, the mortality rate from diseases of the circulatory system is also expected to decrease.

Despite being the variable with lower absolute impact on the conditional mean of the dependent variable, if a municipality invests more in *sports* activities and equipment, the mortality rate from diseases of the circulatory system has the tendency to be lower (*ceteris paribus*). The logic behind this output is that a higher municipal spending on such activities and equipment likely indicates increased physical activity and may imply better secondary lifestyle effects, which improve cardiovascular health.

Finally, the positive sign of *refcentre* also carries an unexpected conclusion: if the municipality has at least one reference centre recognized by the Ministry of Health (regarding cardiology or paediatric cardiology), in a *ceteris paribus* context, the mortality rate from diseases of the circulatory system appears to be higher than when it does not. However, this outcome can be explained by the fact that, having such centre attracts more patients in need, that unfortunately may end up dying in that municipality.

5.3. Partial Effects

When working with nonlinear models like the logit, probit and complementary log-log, it becomes relevant to derive the average PE. These effects measure the marginal change in the probability of the outcome (*heartmort*) when a predictor changes by one unit. This analysis facilitates the comparison across models, as the PE are remarkably similar – unlike the raw coefficients.

Table 4 presents the PE of the explanatory variables on circulatory disease mortality probability across three selected models. These effects, calculated at the mean values of all variables, quantify the marginal impact of each predictor for a representative municipality while holding other factors constant.

Variables	**Frac. Logit**	**Frac. Probit**	**Frac. Cloglog**
	dy/dx	**dy/dx**	**dy/dx**
hosp	-0.0028627 *	-0.0029041 *	-0.0028619 *
popdens	-0.0000002 ***	-0.0000002 ***	-0.0000002 ***
illiteracy	0.0002099 ***	0.0002105 ***	0.0002099 ***
fire	-0.0000047 ***	-0.0000046 ***	-0.0000046 ***
purpower	0.0000268 ***	0.000026 ***	0.0000268 ***
pop65	0.0001254 ***	0.0001269 ***	0.0001254 ***

doctors	-0.0000874 **	-0.0000843 **	-0.000875 **
sports	-0.0000001 ***	-0.0000001 ***	-0.0000001 ***
refcentre	0.0015979 **	0.0015019 **	0.0016 **

Legend: * p-value<0.1 ; ** p-value<0.05 ; *** p-value<0.01

Table 4: Average Partial Effects for the Selected Models: Fractional Logit, Fractional Probit, and Fractional Cloglog. Self-elaborated using STATA outputs.

The similar values across all three models, including the same direction, magnitude and significance, suggest that this approach produces robust estimates regardless of the link function chosen. As anticipated, and mentioned in the Methodology chapter, the signs and individual significance of PE are aligned with those of the coefficients. Hence, the reasoning behind the signs, presented in the previous chapter, can also be applied to this analysis. Note that the small magnitude of the PE could be anticipated in a framework where the proportion represents a permillage.

For the variables that present a negative sign, the interpretation can be made as: an increase of one hospital per 1,000 inhabitants; or an increase of one doctor per 1,000 inhabitants is associated with a reduction in the mortality rate from diseases of the circulatory system by approximately 0.0029; or 0.0001, respectively, holding all other variables constant.

When the sign is positive but there are still small divergencies across models, a one percentage point increase in the PPI per capita; or the presence of at least one reference centre recognized by the Ministry of Health for cardiology/paediatric cardiology tends to increase the mortality rate from diseases of the circulatory system by around 0.00002 to 0.00003; or 0.0015 to 0.0016, respectively, holding all other variables constant.

Moreover, an increase of one individual per km² in population density; an additional rural fire; or an additional thousand euros in expenses for sports activities/equipment is expected to negatively impact the mortality rate from diseases of the circulatory system by, respectively, 0.0000002; 0.000005; or 0.0000001.

Lastly, a one percentage point increase in the illiteracy rate; or a one percentage point increase in the population aged 65 years and over is predicted to increase the mortality rate from diseases of the circulatory system by about 0.0002; or 0.0001.

6. CONCLUSION

Existing research on economic determinants of cardiac disease mortality highlights how macroeconomic performance, healthcare access, and SES influence health outcomes. As noted in papers like Farahani et al. (2010) and Granados & Ionides (2017), macroeconomic conditions modulate mortality incidence, while Hagen et al. (2015) demonstrates how disadvantaged economic groups experience higher cardiac mortality. Cardiovascular health disparities are further influenced by geographic and environmental factors according to Atalay et al. (2023) and Açiktepe et al. (2024). However, municipal-level studies in Portugal remain uncommon despite comprehensive data availability from institutions like INE and Pordata.

This study represents a pioneering approach in Portuguese health research by applying advanced econometric techniques to analyse circulatory disease mortality rates across all 308 municipalities. The analysis revealed several significant determinants: positive impacts from illiteracy rate, proportion of population aged 65 and over, presence of cardiology reference centres, and PPI; negative relationships with hospitals and doctors per 1,000 inhabitants, population density, expenses in sporting equipment, and rural fires, suggesting complex wealth-health dynamics.

Based on the findings, several concrete policy measures are recommended: targeted literacy programs in municipalities with high illiteracy rates; strategic redistribution of healthcare resources; specialized cardiovascular prevention programs for elderly populations; stress-reduction initiatives and lifestyle interventions in high-PPI areas; and innovative healthcare delivery models like mobile clinics and telemedicine in low-density municipalities to improve access to cardiac care.

This study, therefore, not only contributes to academic literature but also provides a valuable tool for public health decision-making, potentially facilitating a more effective resource allocation and improving cardiovascular health outcomes across Portugal. Future research could build upon this foundation by examining longitudinal changes (spatial analysis) in these relationships, evaluating specific policy interventions, and incorporating additional variables that may influence cardiovascular health outcomes at the municipal level throughout Portugal.

REFERENCES

Açiktepe, B., Esirgun, S. N., & Kocak, M. (2024, September 23). Association of environmental and behavioural factors with cardiovascular disease mortality. *ESC Heart Failure*, 12(1), 401-407. doi:10.1002/ehf2.14976

Anjos, R. (2025). Knowledge and perception of disease - Time to empower the needy and change the world! *Portuguese Journal of Cardiology*, 44(2), 105-106. Retrieved from https://doi.org/10.1016/j.repc.2024.11.007

Atalay, K., Edwards, R., & Georgiakakis, F. (2023). Mortality inequality, spatial differences and health care access. *Health Economics*, 2632-2654. doi:10.1002/hec.4746

Birgisdóttir, K. H., Hauksdóttir, A., Ruhm, C. J., Valdimarsdóttir, U. A., & Ásgeirsdóttir, T. L. (2020, May). The effect of the economic collapse in Iceland on the probability of cardiovascular events. *Economics & Human Biology*, 37. Retrieved from https://doi.org/10.1016/j.ehb.2020.100861

Buscardini, L. (2023, May 20). Clippings. *Deaths From Cardiovascular Disease Surged 60% Globally Over the Last 30 Years: Report.* Retrieved from https://faustopinto.com/2023/05/deaths-from-cardiovascular-disease-surged-60-globally-over-the-last-30-years-report/

Cabral, S., Gavina, C., Almeida, M., Sousa, A., Francisco, A. R., Oliveira, E. I., . . . Pereira, H. (2025). Strategic Plan for Cardiovascular Health in Portugal - Portuguese Society of Cardiology (PESCP-SPC). *Portuguese Journal of Cardiology*, 44(1), 41-56. Retrieved from https://doi.org/10.1016/j.repc.2024.11.006

Costa, C., & Santana, P. (2021). Gender and Age Differences in Socio-economic Inequalities in Total and Avoidable Mortality in Portugal: A Trend Analysis. *Fiscal Studies*, 42(1), 123-145.

Degenhard, J. (2024, January 30). *Number of hospitals in Europe 2014-2029*. Retrieved from Statista: https://www.statista.com/forecasts/1166223/hospital-count-forecast-in-europe

Duarte, E. d., Salgueiro, V., Costa, M. J., Lucio, P. S., Potes, M., Bortoli, D., & Salgado, R. (2023). Fire-Pollutant-Atmosphere Components and Its Impact on Mortality in Portugal During Wildfire Seasons. *GeoHealth*. doi:10.1029/2023GH000802

European Union. (2023, March). *Population and housing census 2021 - population grids*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_housing_census_2021_-_population_grids

European Union. (2023, March 30). *Population and housing census 2021: first EU results*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230330-2

European Union. (2024, July). *Cardiovascular diseases statistics*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cardiovascular_diseases_statistics

Farahani, M., Subramanian, S. V., & Canning, D. (2010, November). Effects of State-level Public Spending on Health on the mortality Probability in India. *Health Economics*, 19(11), 1361-1376. doi:10.1002/hec.1557

Framingham Heart Study. (1948). *History of the Framingham Heart Study*. Retrieved from Framingham Heart Study - Three Generations of Dedication: https://www.framinghamheartstudy.org/fhs-about/history/

Gouveia, M., Ascenção, R., Fiorentino, F., Costa, J., Caldeira, D., Broeiro-Gonçalves, P., Borges, M. (2019, January 8). The current and future burden of heart failure in Portugal. *ESC Heart Failure*, 6, 254-261. doi:10.1002/ehf2.12399

Granados, J. A., & Ionides, E. L. (2017, December). Population health and the economy: Mortality and the Great Recession in Europe. *Health Economics*, 26(12), e219-e235. Retrieved from https://doi.org/10.1002/hec.3495

Hagen, T. P., Häkkinen, U., Iversen, T., Klitkou, S. T., & Moger, T. A. (2015, December 3). Socio-economic Inequality in the Use of Procedures and Mortality Among AMI Patients: Quantifying the Effects Along Different Paths. *Health Economics*, 24(S2), 102-115. doi:10.1002/hec.3269

Kawachi, I., Kyriopoulos, I., & Vandoros, S. (2023). Economic uncertainty and cardiovascular disease mortality. *Health Economics*, 32(7), 1550-1560. doi:10.1002/hec.4678

Khaltaev, N., & Axelrod, S. (2024, February 7). Cardiovascular disease mortality and air pollution in countries with different socioeconomic status. *Chronic Diseases and Translational Medicine*, 10(3), 247-255. doi:10.1002/cdt3.116

Marques-Alves, P., Marinho, A. V., Almeida, J. P., Gonçalves, T., Costa, M., Ferreira, M., Gonçalves, L. (2020, February 5). Real-world analysis of acute decompensated heart failure outcomes in Portugal. *ESC Heart Failure*, 7, 551-558. doi:10.1002/ehf2.12599

O'Neill, A. (2024, September 27). *Unemployment rate in the European Union and the Euro area from 2013 to 2023*. Retrieved from Statista: https://www.statista.com/statistics/267906/unemployment-rate-in-eu-and-euro-area/

Papke, L. E., & Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics*, 11, 616-632.

Ramalho, E. A., Ramalho, J. J., & Murteira, J. M. (2011). Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models. *Journal of Economic Surveys*, 25(1), 19-68.

Sekkarie, A., Woodruff, R. C., Casper, M., Thompson-Paul, A. M., & Vaughan, A. S. (2024, August 16). Rural–urban disparities in cardiovascular disease mortality vary by poverty level and region. *The Journal of Rural Health*, 41(1), e12874.

Serviço Nacional de Saúde. (2024, March 6). *SNS Cardiologia. Rede de Referenciação Hospitalar de Cardiologia 2023*, p. 15.

Serviço Nacional de Saúde. (2024, March 6). *SNS Cardiologia Pediátrica. Rede de Referenciação Hospitalar 2023*, pp. 44-48.

Smits, J., Keij-Deerenberg, I., & Westert, G. (2004). Effects of socio-economic status on mortality: separating the nearby from the farther away. *Health Economics*, 14(6), 595-608. Retrieved from https://doi.org/10.1002/hec.950

Yanatma, S. (2024, September 25). *Workers in this country endure Europe's longest commutes. How does your country compare?* Retrieved from Euronews: https://www.euronews.com/next/2024/09/25/these-are-europes-longest-and-shortest-commutes-to-work-how-does-your-country-compare

APPENDICES

Author	Year	Object	Region	Dep. Var.	Regressors	Models
Smits et al.	2004	Individuals Regions	Netherlands	Total mortality Cause-specific mortality	SES indicators	Multilevel
Farahani et al.	2010	Individuals Households Regions	India	Probability of death	Public health spending Av. household spending on health Per Capita Net State Domestic Product Rural/urban residence Access to toilet facilities Access to water sources Religion Caste Asset index quintiles	Multilevel probit Instrumental variable probit Probit with interaction terms Probit for morbidity outcomes
Sekkarie et al.	2014	Regions	USA	Age-standardized CVD mortality rate	Rural/urban classification Poverty level USA Census region	-
Hagen et al.	2015	Individuals	Finland Norway	Probability of death after AMI admission Provision of PCI	SES variables (income and education) Age Gender AMI type Associated health conditions Distance to the nearest PCI	Linear regression Path analysis
Granados & Ionides	2017	Regions	EU-27	Mortality rate	Unemployment rate GDP per capita Employment-to-population ratios	Fixed effects Nonlinear detrending Differenced regressions
Gouveia et al.	2019	Regions	Portugal	Burden of heart failure	Years of life lost due to premature death and disability Age Gender Population projections Chronic heart failure class	Shift-share analysis DISMOD II
Birgisdóttir et al.	2020	Individuals	Iceland	Probability of an IHD event	Unemployment rate Indicators of economic collapse and crisis Age Gender Marital status Number of children Real monthly income Real equity	Linear probability with fixed effects
Marques-Alves et al.	2020	Individuals	Portugal	In-hospital HF mortality rate 30-day HF mortality rate Follow-up HF mortality rate Readmission rates (at 30 days and overall follow-up)	Age Gender Biomarkers Clinical diagnoses Length of prior hospitalization	Univariate and multivariate Cox proportional hazards
Costa & Santana	2021	Regions	Portugal	Age-specific mortality rate	Gender Age	Linear regression

					Level of socioeconomic deprivation	
Kawachi et al.	2023	Regions	England Wales	Weekly deaths from diseases of the circulatory system, IHD, and cerebrovascular disease	EPUI Unemployment rate Consumer price index GDP growth Population size	OLS Probit with interaction terms Poisson and negative binomial Lagged dependent variables Time trends
Atalay et al.	2023	Regions	Australia	3-year mortality rate	Percentile rank of each SA3 group Year dummies Interaction terms between year and rank	Linear regression DiD
Duarte et al.	2023	Regions	Portugal	Cardio-respiratory mortality rates	Fire-pollutant-atmospheric variables	Generalized additive PCRL Linear regression
Açiktepe et al.	2024	Regions	Turkey	CVD mortality rate	Environmental factors Lifestyle factors (smoking and alcohol consumption)	SAS Ordinal logistic regression Univariable and multivariable FDR
Khaltaev & Axelrod	2024	Regions	WHO member states	Age-standardized CVD mortality rate	Income level Type of air pollution Annual mean concentration of fine particulate matter Proportion of population using clean fuels and technologies for cooking	-

Table 5: Literature Review – Econometric Studies Overview. Self-elaborated.

Variable	Description
heartmort	Mortality rate from diseases of the circulatory system (‰)
metro	Dummy variable that assumes 1 if the municipality is in a metropolitan area
hosp	Number of hospitals per 1,000 inhabitants
environ	Environmental expenditure (thousands €)
popdens	Population density (individuals/km²)
unem	Unemployment rate (%)
illiteracy	Illiteracy rate (%)
natura	Protected areas and Natura 2000 Network (hectares)
fire	Number of rural fires
purpower	Purchasing power index per capita (%)
pop65	Population aged 65 years and over (%)
fem	Female population (%)
doctors	Number of doctors per 1,000 inhabitants
coastline	Dummy variable that assumes 1 if the municipality is in the coastline
sports	Expenses in sports activities/equipment (thousand €)
refcentre	Dummy variable that assumes 1 if the municipality has at least one reference centre recognized by the Ministry of Health, regarding cardiology or paediatric cardiology

commute60	Population that takes at least 60 minutes to get to work/school from home (%)

Table 6: Variable Description. Self-elaborated.

Variable	Mean	Std. Dev.	Minimum	Maximum
heartmort	4.156818	1.717201	0	12.4
metro	0.1136364	0.3178856	0	1
hosp	0.0147532	0.0338328	0	0.334
environ	25.80519	111.1071	0	1,178
popdens	295.2792	809.4718	4.4	7,281.6
unem	7.272273	2.313841	2.69	19.65
illiteracy	4.853149	2.234932	1.41	12.26
natura	6,290.763	11,750.22	0	78,161
fire	26.72078	30.60806	0	202
purpower	82.25974	15.63603	60.7	186.3
pop65	27.77523	6.960037	10.6442	47.52279
fem	52.08585	1.101367	44.59002	55.58986
doctors	2.939819	2.914924	0.4899559	34.30008
coastline	0.2954545	0.4569893	0	1
sports	1,049.829	1,561.463	8.5	13,998.33
refcentre	0.012987	0.1134024	0	1
commute60	3.137258	2.053839	0	16.28693

Table 7: Descriptive Statistics. Self-elaborated using STATA outputs.

	heartmort	*metro*	*hosp*	*environ*	*popdens*	*unem*	*illiteracy*	*natura*
heartmort	**1**							
metro	-0.301 (0.000)	**1**						
hosp	-0.171 (0.003)	0.048 (0.398)	**1**					
environ	-0.159 (0.005)	0.355 (0.000)	0.058 (0.309)	**1**				
popdens	-0.270 (0.000)	0.622 (0.000)	0.110 (0.053)	0.306 (0.000)	**1**			
unem	-0.198 (0.001)	0.213 (0.000)	-0.017 (0.768)	0.064 (0.262)	0.243 (0.000)	**1**		
illiteracy	0.666 (0.000)	-0.415 (0.000)	-0.157 (0.006)	-0.209 (0.000)	-0.374 (0.000)	-0.020 (0.728)	**1**	
natura	0.169 (0.003)	-0.122 (0.032)	-0.040 (0.481)	0.002 (0.967)	-0.156 (0.006)	0.024 (0.673)	0.284 (0.000)	**1**
fire	-0.245 (0.000)	0.298 (0.000)	-0.025 (0.668)	0.096 (0.093)	0.097 (0.089)	0.112 (0.049)	-0.161 (0.005)	0.0612 (0.279)

purpower	-0.385 (0.000)	0.483 (0.000)	0.269 (0.000)	0.360 (0.000)	0.576 (0.000)	0.106 (0.062)	-0.629 (0.000)	-0.051 (0.372)
pop65	0.743 (0.000)	-0.310 (0.000)	-0.129 (0.023)	-0.178 (0.002)	-0.266 (0.000)	-0.234 (0.000)	0.711 (0.000)	0.213 (0.000)
fem	0.128 (0.025)	0.151 (0.008)	0.141 (0.013)	0.043 (0.456)	0.235 (0.000)	0.101 (0.078)	0.009 (0.882)	-0.236 (0.000)
doctors	-0.268 (0.000)	0.278 (0.000)	0.328 (0.000)	0.191 (0.001)	0.455 (0.000)	0.101 (0.076)	-0.381 (0.000)	-0.065 (0.255)
coastline	-0.351 (0.000)	0.239 (0.000)	0.020 (0.728)	0.196 (0.001)	0.195 (0.000)	0.263 (0.000)	-0.401 (0.000)	-0.050 (0.385)
sports	-0.337 (0.000)	0.446 (0.000)	0.147 (0.010)	0.345 (0.000)	0.520 (0.000)	0.203 (0.000)	-0.425 (0.000)	-0.029 (0.618)
refcentre	-0.074 (0.195)	0.230 (0.000)	0.170 (0.003)	0.278 (0.000)	0.434 (0.000)	0.105 (0.065)	-0.149 (0.009)	-0.061 (0.288)
commute60	-0.028 (0.629)	0.316 (0.000)	-0.064 (0.266)	0.082 (0.149)	0.166 (0.004)	0.124 (0.030)	-0.056 (0.326)	-0.078 (0.174)

(cont.)	*fire*	*purpower*	*pop65*	*fem*	*doctors*	*coastline*	*sports*	*refcentre*	*commute60*
fire	**1**								
purpower	0.109 (0.056)	**1**							
pop65	-0.218 (0.000)	-0.461 (0.000)	**1**						
fem	0.043 (0.458)	0.076 (0.182)	0.260 (0.000)	**1**					
doctors	0.090 (0.117)	0.651 (0.000)	-0.227 (0.000)	0.240 (0.000)	**1**				
coastline	-0.126 (0.027)	0.379 (0.000)	-0.474 (0.000)	-0.143 (0.012)	0.167 (0.003)	**1**			
sports	0.271 (0.000)	0.615 (0.000)	-0.312 (0.000)	0.133 (0.019)	0.529 (0.000)	0.264 (0.000)	**1**		
refcentre	0.053 (0.357)	0.415 (0.000)	-0.062 (0.276)	0.140 (0.014)	0.681 (0.000)	0.114 (0.045)	0.391 (0.000)	**1**	
commute60	0.280 (0.000)	0.042 (0.459)	-0.040 (0.485)	0.151 (0.008)	-0.080 (0.162)	-0.061 (0.289)	0.057 (0.317)	-0.021 (0.715)	**1**

Table 8: Pearson's Correlation (p-values in parentheses). Self-elaborated using STATA outputs.

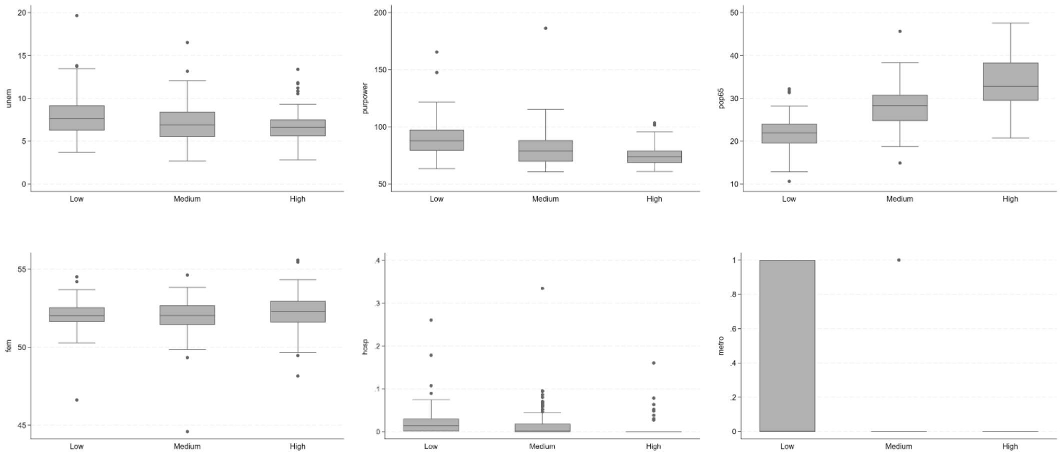

Figure 7: Boxplots of *unem*, *purpower*, *pop65*, *fem*, *hosp*, and *metro* variables' distributions across low, medium, and high heart disease mortality rate groups. Self-elaborated in STATA.