Lisbon School of Economics & Management
Universidade de Lisboa

MASTER
ACTUARIAL SCIENCE

MASTER´S FINAL WORK
PROJECT

ADJUSTMENT OF MORTALITY TABLES IN WORKERS' COMPENSATION INSURANCE
(A CASE STUDY)

ALAN GUILHERME PINHEIRO

SUPERVISION:
CARLOS EDUARDO DA ROSA
ONOFRE ALVES SIMÕES

OCTOBER - 2023

Abstract

The present study focuses on the selection of an appropriate mortality table for calculating the mathematical provision for workers' compensation pensions, specifically permanent disability pensions. The research examines the experience of Fidelidade, a Portuguese insurance company, to determine a mortality table that is accurate and reliable enough to be used in calculating these types of pension provisions. By analyzing both Fidelidade's experience and broader market trends, this study aims to provide a comprehensive and well-supported recommendation for the utilization of a specific mortality table in this context.

The primary argument of this work is to assess the adequacy of existing mortality tables in accurately describing the mortality patterns of disabled individuals covered by workers' compensation regulations. Additionaly, the study explores the possibility of adjusting an existing mortality table to align more closely with the observed mortality patterns within Fidelidade experience.

Keywords - Mortality Table; Workers' Compensation; Pensions; Reserve; Regression Tree; GLM

Resumo

Este estudo se concentra na seleção de uma tabela de mortalidade apropriada para calcular a provisão matemática destinada a pensões por acidente de trabalho. A pesquisa examina a experiência da Fidelidade, uma companhia de seguros em Portugal, com o objetivo de determinar qual tabela de mortalidade é a mais precisa e confiável para utilizar no cálculo dessas provisões de pensão. Ao analisar tanto a experiência da Fidelidade quanto as tendências gerais do mercado, nosso estudo visa fornecer uma recomendação abrangente e bem fundamentada sobre a tabela de mortalidade mais apropriada a ser empregada nesse contexto.

O ponto central do trabalho consiste em avaliar a adequação das tabelas de mortalidade existentes para descrever com precisão os padrões de mortalidade de indivíduos com incapacidades cobertos pelas apólices de acidente de trabalho. Além disso, exploramos a possibilidade de ajustar uma tabela de mortalidade que esteja disponível, de forma a refletir os padrões de mortalidade observados na experiência da Fidelidade.

Palavras-chave - Tábua de Mortalidade; Acidente de Trabalho; Pensões; Reservas; Árvore de Regressão; GLM.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The task of selecting an appropriate mortality table that best models the mortality of a particular population poses a significant challenge, as there is a multitude of tables available for the general population that we might consider. However, the complexity deepens when we turn our attention to the disabled population covered by the workers compensation line of business. Questions arise, such as: does mortality change for them? What types of disabilities can impact mortality? What is the life expectancy for individuals living with one leg or one arm? Such inquiries present a formidable challenge when it comes to deciding which mortality table accurately represents the mortality of this specific population. While there are mortality tables designed for disabled individuals, it is important to note that these tables often reflect the realities of specific countries. This raises the question: Do disabled individuals in one country show similar mortality patterns to those in other countries?

The present study is focusing on a particular group of individuals: pensioners receiving compensation due to work-related accidents through Fidelidade, a Portuguese insurance company that offers workers' compensation coverage. Unlike broader mortality studies, the reach of our study is limited by the availability of data. Workers' compensation studies inherently deal with smaller datasets, as not the entire population is engaged in work that exposes them to occupational hazards. Consequently, our study population is limited to individuals who are both old enough to work and not yet retired. In contrast, general mortality studies have access to larger and more diverse populations, encompassing various age groups and characteristics.

The selected methodology drew inspiration from the works of Guo (2008) in the United States[11]. In his paper, Guo employs the CART model and the generalized linear model to investigate advanced age mortality. Additionally, Benjamin's study[2] on mortality table adjustments with a credibility approach also served as a reference for our methodology.

1.1 Literature Review

Prior research in the field of worker's compensation reserving has delved into key discussions and considerations, which we explore in the following sections. Earlier studies, like those conducted by Ferguson (1971)[8] and Steeneck (1996)[21], underscored the significance of accounting for mortality improvement over time. More recent papers, exemplified by Jones et al. (2013)[13], have taken a deeper dive into this aspect, offering comprehensive guidance on constructing mortality models.

The literature review predominantly revolves around three critical issues. Firstly, it examines the selection of an appropriate mortality table, a subject that Benjamin (2008)[2] addresses. Secondly, it considers mortality improvement trends over time, a theme emphasized by earlier work such as that of Ferguson (1971)[8]. Lastly, it explores the implications of dynamic mortality tables on future liability and reserves, as explored by Jones et al. (2013)[13]. Throughout these discussions, various models and methodologies have been proposed to effectively tackle these complex challenges.

In his work, Benjamin (2008)[2], underscores the crucial role of judgment in actuarial work, particularly when it comes to adjusting mortality assumptions. This author elaborates on the credibility approach, which assumes that the shape of a standard mortality table is appropriate for the plan being assessed, necessitating only a proportional adjustment. Nevertheless, if an actuary strongly believes that the plan's mortality curve significantly deviates from available standard tables, Benjamin suggests that actuaries have the option to construct a bespoke table using plan-specific experience data, even when such data is limited. Benjamin's insights emphasize that, with the exercise of informed judgment, credibility theory remains a potent tool for actuaries when establishing mortality assumptions for pension plans.

Ferguson (1971)[8] highlighted the necessity of considering mortality in long-term pension-type worker's compensation awards. Special care is required when excess loss reinsurance coverage is involved. Reserves are calculated by breaking down gross reserves into net and ceded reserves, with the net reserve based on a temporary life annuity, accounting for both mortality and interest discounting. Steeneck (1996)[21] expanded on Ferguson's work, incorporating factors like escalation of indemnity benefits and medical inflation into mortality-based forecasts. This included modeling indemnity and medical expenses using annuities and arguing for the inclusion of these factors for greater accuracy in reserve estimation.

Another crucial aspect was addressed by Snader (1987)[20], focusing on life contingency concepts in establishing reserves for claimants requiring lifetime medical care. This approach involves a comprehensive evaluation process encompassing claim assessment, medical evaluation, and actuarial evaluation, with thorough discussions on the selection of key assumptions.

Gillam (1993)[10] delved into mortality assumptions in response to the National Council on Compensation Insurance (NCCI) Special Call for Injured Worker Mortality Data in

1987 and 1988, in the United States. His research uncovered intriguing findings. Notably, he observed that for individuals below the age of 60, the reported injured worker mortality rates exceeded what was found in standard U.S. life mortality tables. However, when examining the age group between 60 and 74, the injured worker mortality rates did not significantly deviate from the standard mortality figures. Importantly, Gillam's conclusion emphasized that disparities in mortality rates did not necessarily indicate redundancy or inadequacy in reserves at that particular time.

Blumsohn (1997)[4], on the other hand, embarked on a comparison between deterministic and stochastic approaches. He highlighted the errors that could arise from assuming deterministic values for key parameters such as medical usage, inflation, cost of living adjustments (COLAs), and investment income. His work demonstrated that eliminating these deterministic assumptions led to enhanced accuracy in calculations.

Shifting the focus to Kahn (2002)[14], this author engaged in discussions surrounding "ultimate" loss reserves within the context of runoff operations. He considered various factors, including medical escalation and the longevity of claimants, in the exploration of this topic. Jones et al. (2013)[13] provided a practical framework for modeling lifetime worker's compensation claims based on mortality, emphasizing the importance of considering the variability of mortality and its impact on reserve projections.

In summary, the research in worker's compensation reserving has evolved to incorporate more sophisticated mortality modeling techniques and considerations, leading to improved accuracy in estimating reserves and liabilities. These insights are valuable for insurers, actuaries, and policymakers involved in managing worker's compensation programs.

1.2 Structure of the text

Our study follows a structured approach, beginning with an overview of workers' compensation regulations and specifications. We will then proceed into an examination of the dataset, exploring the available variables that can help us understand mortality patterns within the specific population at Fidelidade. Moving on to the methodology section, we employ two distinct yet complementary approaches for forecasting and analyzing the relationships between variables and the response variable. In the following sections, we will present and discuss our findings, with the goal of identifying the most suitable mortality table for Fidelidade's context.

Chapter 2

Overview

2.1 Workers' Compensation: a Brief History

Since the late 1800s, European countries have implemented laws to protect employees from work-related accidents. Germany, led by Chancellor Otto von Bismarck, was a pioneer in this regard and served as a model for industrialized nations. The "Workman's Compensation Act," introduced in Germany in 1884, was later adopted by the UK in 1897. Similarly, in 1913, Portugal made it mandatory for employers to cover the costs of work-related accidents.

As stated by Ladou[16], European countries have different approaches for their workers' compensation systems. The German system involves self-governed insurance associations funded by employers, offering comprehensive services for prevention, rehabilitation, and compensation. The second approach is where the state manages compensation for occupational injuries and diseases, often as part of broader social security, with funding collected from employers. Portugal, along with other European countries, have a combination of both state and private insurance systems.

Today, in Portugal, the Workers' Compensation (WC) program provides coverage for two types of incidents: work-related injuries and occupational illnesses resulting from repeated exposure. WC is mandatory for employers, but the liabilities are divided between insurance companies, responsible for work-related injuries, and the national social security system, responsible for occupational illnesses. The program is regulated by Law 98/2009[1], which ensures the protection of beneficiaries through detailed specifications of benefits and responsibilities.

WC aims to cover medical expenses for workers injured during their professional activities, including travel to and from work. It also provides compensation for lost wages and death benefits for dependents of workers who die in work-related accidents. These benefits can be provided in the form of cash or in kind, and they may result in temporary or permanent obligations for the employers.

Under the Solvency II framework, WC consists of two components: the workers' compensation insurance Line of Business (LoB) and annuities related to health insurance obligations LoB. Non-similar to life techniques (NSLT) are used for workers' compensation insurance, covering temporary medical expenses, pharmaceutical costs, nursing care, and temporary salary compensation. On the other hand, similar to life techniques (SLT) are employed for annuities, which include disability pensions categorized based on the degree of disability and pensions for dependents (spouse, ascendants, and descendants) in the event of a worker's death. Additionally, SLT encompasses long-term assistance such as lifelong medical support and the replacement/maintenance of prostheses for injured workers.

Despite having liabilities similar to life insurance, WC is present in the Portuguese market as a non-life business, representing 19,9% of the total[1], with an stable amount of 60,1% loss ratio[2] in 2022.

2.2 Workers' Compensation Coverage

As previously mentioned, WC SLT liabilities include pensions and lifetime assistance payments. Lifetime assistance comprises permanent medical assistance payments and prostheses replacement for the injured worker's lifetime. Pensions are further divided into disability pensions and pensions for dependants in case of the insured's death. The pensions, regulated by Law 98/2009[1], can be redeemable under the following conditions:

1. Both the annual lifelong pension for a person with a permanent partial disability below 30% and the annual lifelong pension for a legal beneficiary can be redeemed. However, the redeemed amount cannot exceed six times the minimum monthly wage as of the day after discharge or death.

2. The annual lifelong pension corresponding to a disability of 30% or more, or the annual lifelong pension of a legal beneficiary, may be partially redeemed, at the request of the injured person or the legal beneficiary, provided that the following limits are met cumulatively:

    a) The remaining annual pension cannot be less than six times the minimum monthly guaranteed wage in force on the date of the redemption authorization;

    b) The redemption capital cannot exceed what would result from a pension calculated based on a 30% disability.

3. In the case of a work-related accident suffered by a foreign worker, resulting in permanent disability or death, the annual lifelong pension can be redeemed in capital, by agreement between the responsible entity and the pension beneficiary, if the beneficiary chooses to permanently leave Portugal.

[1]This information is available on www.asf.com.pt

[2]Loss ratio information is available in the annual report of the APS.

4. The provisions of the preceding numbers do not apply to the legal beneficiary of an annual lifelong pension who suffers from a disability or chronic illness that permanently reduces their overall earning capacity by more than 75%.
5. In the event that the injured person suffers multiple accidents, the pension to be redeemed is the total pension.

On the side of the insurer, the calculation of WC SLT liabilities needs to be done on a case-by-case basis for the entire portfolio, considering the diversity and different risk exposures. It is important to analyze non-redeemable pensions, redeemable pensions, and lifetime assistance separately, as they contain different risks. Non-redeemable pensions are calculated as annuities and paid monthly, with additional allowances for holidays and Christmas. The pension value can be revised once per civil year without time limit, based on the policyholder's or undertaking's request. If the worker's incapacity changes during the year, the pension value is adjusted accordingly.
Disability pensions have three stages: provisional, defined, and definitive. The provisional stage occurs when there is not yet a decision on the degree of incapacity. In the defined stage, the degree of incapacity is determined, and in the final stage, the pension value is legally defined.

The categories for pension calculations are as follows:

- Fully disabled to perform any kind of work: This category receives a lifelong pension of 80% of the worker's wage, with an additional 10% for each dependant (limited to the worker's wage).
- Fully disabled for the usual work: Workers who are fully disabled for their usual work receive a lifelong pension ranging from 50% to 70% of their wage, depending on their capability to perform other types of jobs.
- Partial permanent incapacity: Those with partial permanent incapacity receive a lifelong pension equal to 70% of the reduction in their wage.
- Pensions for worker's dependants in case of death: The main beneficiaries of these pensions are spouses or equivalent, descendants, and ascendants. The specific individuals recognized in each category are outlined in Law 98/2009[1].

The main pensions paid in case of death include:

- Spouse or equivalent: 30% of the worker's annual wage until the normal retirement age, increasing to 40% afterward.
- Descendants: 20% of the worker's annual wage if there is one descendant, 40% if there are two descendants, and 50% if there are three or more descendants. These pensions are paid until the descendants reach 18 years old, but the payment period can be extended if they continue their education beyond that age. Descendants with significant permanent disability or chronic disease may receive the pension for their whole life. If both parents die, these pensions can double (up to a maximum of 80% of the worker's annual wage).

- Ascendants: Each ascendant receives 10% of the worker's annual wage until a maximum of 30%. If there are no other beneficiaries, they receive 15% each until retirement age or the onset of a chronic disease, and 20% afterward.

- "Fundo de Acidente de Trabalho" (FAT): If the deceased worker has no dependants, the insurer should pay three times the worker's annual wage to this Workers' Compensation Fund.

In any case, the sum of all pensions mentioned above cannot exceed 80% of the annual wage of the deceased worker. If the total pensions exceed this limit, they need to be proportionally revised to ensure compliance.

2.3 Reserving for Similar to Life Techniques

Solvency II's best estimate represents the expected present value of all future cash flows related to existing contractual obligations, encompassing past, present, and future exposures. These cash flows consist of claim payments, allocated expenses associated with claims, unallocated expenses, and expected future premiums from policies in force.

Regarding the best estimate for Similar to Life Techniques (SLT), it aims to respond to the longevity risk and relies on two main pillars: the adjustment of mortality experience and the variability of expected lifetime.

It is important to note that longevity risk primarily impacts non-compulsory recoverable pensions and Lifetime Assistance. In cases involving individuals who have suffered accidents resulting in disabilities, their mortality patterns often deviate from those of the general Portuguese population. Typically, life insurance mortality tables are based on a population that is wealthier, better educated, and more likely to be married. These specific characteristics are associated with lower mortality rates and more significant improvements in mortality, as elaborated by Jones et al. (2013)[13], when compared to the broader population.

As a result, the mortality tables published by the National Institute of Statistics (Instituto Nacional de Estatística - INE) may not be the most cost-effective or suitable choice. However, they still serve a valuable purpose in capturing mortality trends within this context.

Next we will go through some the technical aspects of workers' compensation reserving, as described by Rosa (2012)[19]. Consider a random variable Y_x representing the present value of unit monthly payments made in advance for a pensioner aged x. The expected value of Y_x depends on the type of beneficiary:

$$\textit{Injured Person:} \quad E[Y_x] = \ddot{a}_x^{(12)} \quad \textit{(Whole life annuity due).} \tag{2.1}$$

$$\textit{Orphans:} \quad E[Y_x] = \ddot{a}_{x:25-x}^{(12)} \quad \textit{(Term life annuity due).} \tag{2.2}$$

$$\textit{Husband/Wife:} \quad E[Y_x] = \ddot{a}^{(12)}_{x:65-x} + \frac{4}{3} \times {}_{65-x}E_x \times \ddot{a}^{(12)}_{65} \quad \textit{(Whole life annuity due).} \tag{2.3}$$

$$\textit{Ascendants:} \quad E[Y_x] = \ddot{a}^{(12)}_{x:65-x} + \frac{4}{3} \times {}_{65-x}E_x \times \ddot{a}^{(12)}_{65} \quad \textit{(Whole life annuity due).} \tag{2.4}$$

Where,

$$\ddot{a}^{(12)}_{x} = \sum_{k \in \{0, \frac{1}{12}, \frac{2}{12}, \ldots\}} v^k {}_kp_x \quad \textit{(Whole life annuity payable 12 times a year).} \tag{2.5}$$

$$\ddot{a}^{(12)}_{x:n} = \sum_{k \in \{0, \frac{1}{12}, \frac{2}{12}, \ldots, n-\frac{1}{12}\}} v^k {}_kp_x \quad \textit{(Term annuities payable 12 times a year).} \tag{2.6}$$

$$v = (1+i)^{-1} \quad \textit{(Discounting factor).} \tag{2.7}$$

$${}_tp_x = Pr[T_x > t] = S_x(t) \quad \textit{(Prob. a life aged x survives for at least t years).} \tag{2.8}$$

$${}_tq_x = Pr[T_x \leq t] = 1 - S_x(t) \quad \textit{(Prob. a life aged x does not survive beyond age x+t).} \tag{2.9}$$

$${}_tE_x = v^n {}_np_x \quad \textit{(Actuarial discount factor).} \tag{2.10}$$

In Equation (2.3), it is important to note that we have not accounted for the possibility of either the husband or the wife getting remarried. In such cases, the husband or wife would lose their entitlement to the pension, but the company would be obligated to pay a lump sum equivalent to three times the annual pension amount. Interestingly, according to the benchmark study conducted by the Portuguese Association of Insurers, this scenario has been rarely utilized since 2006. For insurance companies, the most conservative approach is to assume a remarriage rate of zero, even though it incurs a cost.

Now, let us denote R^I as the amount of reserves for a portfolio of annuities at time I:

$$R^I = \sum_{w=1}^{W} R^I_w \quad . \tag{2.11}$$

where W is the number of pensioners and R^I_w is the reserve for beneficiary w at time I (depending on the annual amount and the expected present value $E[Y_x]$).

It is expected that some pensioners will die during accounting year $(I, I+1]$ and release reserve at time $I+1$. When it does not occur the reserve will be recalculated at time $I+1$ with the pensioner one year older. Let P^{I+1} be the payments that will occur during accounting year $(I, I+1]$,

$$P^{I+1} = \sum_{w=1}^{W} \frac{P_w}{12} \times [\ddot{a}_{\overline{12}|j} \times I_w + \ddot{a}_{\overline{6}|j} \times (1 - I_w)] \quad . \tag{2.12}$$

P_w is the annual amount paid to pensioner w,

$$I_w = \begin{cases} \text{1, pensionser } w \text{ does not die during accounting year (I,I+1]} \\ \text{0, pensionser } w \text{ dies during accounting year (I,I+1]} \end{cases}$$

and the factor $\ddot{a}_{\overline{n}|i}$ corresponds to the present value of the n ($n = 12$,$n = 6$) certain monthly payments of one monetary unit in advance (not depending on human life) and j is the effective monthly rate of interest.

Theoretically, the relation between reserves at time I (already observed), expected reserves at time $I+1$ and expected payments occurring in accounting year $(I, I+1]$ is

$$R^I = E[R^{I+1}] \times v + E[P^{I+1}] \quad . \tag{2.13}$$

It is important to recognize that the expected reserves at time $I+1$ and the expected payments within the accounting year $(I, I+1]$ are subject to uncertainty, contingent on whether pensioners pass away during that period or not. To assess solvency, companies should conduct simulations to model the behavior described in the RHS of Equation (2.13). For a specific replica m, the capital requirement CR^I can be calculated as follows:

$$CR^I(m) = R^{I+1}(m) \times v + P^{I+1}(m) - R^I \quad . \tag{2.14}$$

Here, $R^{I+1}(m)$ represents the observed reserves, and $P^{I+1}(m)$ represents the observed payments. Notably, in each simulation replica, we determine whether each pensioner survives or passes away, using a distribution based on $Bi(1, q_x)$, where x denotes the age of the pensioner.

When $CR^I(m)$ yields a positive value, it indicates that the reserves are insufficient for that particular replica. To calculate the solvency capital requirement for longevity risk, we consider the 99.5% confidence level of the distribution of CR^I in accordance with Solvency II regulations.

Chapter 3

Data

Before selecting an appropriate mortality table for calculating provisions for WC pensioners, a comprehensive understanding of the available data is essential. The insurance company Fidelidade has recorded pertinent information about pensioners, including their date of birth, sex, level of incapacity, and annual salary. For example, they have records of a 21-year-old male pensioner earning 1.2 times the minimum wage annually, with a 70% disability level, still alive in 2015. Another insured individual is a 46-year-old female earning 0.9 times the minimum wage annually, with a disability level of 16.5% that died in 2019. Moreover, for pensioners receiving life assistance, additional data points are available, for instance the average cost of the life assistance.
In our analysis of the effects of these variables on mortality, we will not consider dependents like spouses, ascendants, and descendants. This omission is because they are not recipients of the pension because they suffered work-related accidents, and standard mortality tables can be more appropriate to describe their mortality patterns, as in general they do not have any work-related disabilities.

3.1 Pensioners over Time

In our analysis of pensioners over time, with the exclusion of dependents as previously mentioned, we have examined data range from 2015 to 2022, as illustrated in Table 3.1. This table provides insights into the annual count of pensioners at the beginning of each year, along with the corresponding total number of deaths in each year. Notably, there was a substantial surge in the number of pensioners from 2015 to 2016, with an increase of 322 individuals. Subsequently, there was a more consistent upward trend in the following years, averaging 29 new pensioners per year. It is worth highlighting that 2022 marked the highest number of registered deaths, totaling 208.

Furthermore, in Table 3.2, we have presented the total number of deaths of the Portuguese population, sourced from Human Mortality Database (HMD)[1]. We can see that, from 2019 to 2021, there was a increase in the number of deaths, which diverges from the trend observed since 2015. This raise the question whether the significant increase in deaths in

[1]Table 3.2 is available on mortality.org

Table 3.1 and Table 3.2 can be attributed to the repercussions of the 2020 pandemic.

Year	Number of Pensioners	Number of Deaths	%
2015	5788	152	2.63%
2016	6110	114	1.87%
2017	6156	130	2.11%
2018	6230	131	2.10%
2019	6300	126	2.00%
2020	6332	117	1.85%
2021	6351	135	2.13%

Table 3.1: Pensioners and Deaths (2015-2021)

Year	Population	Number of Deaths	%
2015	10,399,507	108,539	1.04%
2016	10,365,966	110,573	1.07%
2017	10,334,170	109,586	1.06%
2018	10,315,591	113,051	1.10%
2019	10,298,681	111,843	1.09%
2020	10,315,540	123,396	1.20%
2021	10,315,101	124,802	1.21%

Table 3.2: Population and Deaths (2015-2021) in Portugal

In a study publish by Mosher (2023)[18], it is highlighted that "COVID-19 has had a significant impact on mortality and life expectancy in the short term, but the impact on mortality assumptions used going forward should be much lower. Many of the mortality shocks will be temporary, and mortality rates can be expected to return to their prior trajectories. Longer-term impacts are significantly less certain, however, thereby increasing the potential risk that experience will deviate from best estimate assumptions. This calls for the ongoing monitoring of longevity experience. Impacts will also vary widely from one country to the next, depending on factors such as the level of development, baseline health, and conditions during lockdown. These differences will need to be considered when assessing the impact of the COVID-19 pandemic on mortality assumptions."(p.44).

3.2 Pensioners by Sex

Upon examining the distribution by sex described in Figures 3.1 and 3.2 bellow, it becomes evident that on average 87% of the pensioners covered by workers' compensation, with either partial or total disability, are males, and 90% of the recorded deaths are also of males.

In a study conducted by Islam (2001)[12], it was highlighted that the overall injury/illness rate demonstrated a significant gender disparity, with females exhibiting a lower rate than males. This trend was observed across all major industrial classes except for the service

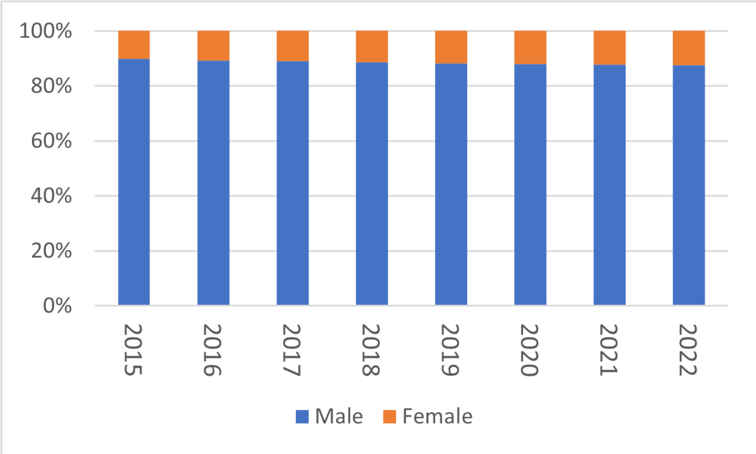

Figure 3.1: Pensioners by Sex (2015-2022)

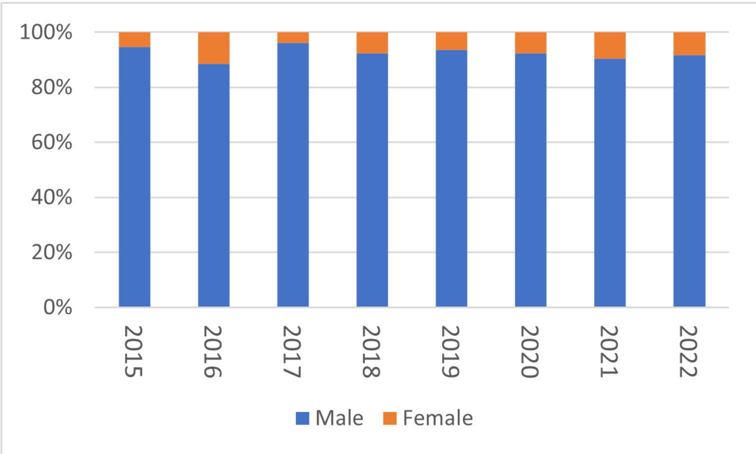

Figure 3.2: Pensioners' Mortality by Sex (2015-2022)

and agricultural sectors. The author also concluded that the distribution of injury/illness types varied based on gender, occupation, and industry. For instance, females showed a notably higher risk of carpal tunnel syndrome, burns, sprains, and fractures in comparison to their male counterparts.

In a recent study conducted by Biswas (2022) [3], the author highlights significant differences in injury risks between men and women across various occupational sectors. The research underscores that men faced a notably higher risk of injury in specific job roles within the primary and secondary industry sectors. These high-risk occupations frequently involved substantial physical exposures and, in certain instances, exposure to particular chemical or biological factors.

3.3 Pensioners by Age

To handle the variable age, an adjustment is necessary. For actuarial purposes, age is calculated by considering the nearest birthday to the observed date. This birthday can either be earlier, matching the person's actual age, or later, corresponding to their upcoming birthday. In the latter scenario, we add one more year for the purpose of our study. An example will illustrate: assume an individual born on February 18, 1977, who enrolls in a workers' compensation pension scheme on November 20, 2022. As of today, this person is 45 years old. However, in the context of our study, their age would be considered as 46 years, because their next birthday is closer to the date they joined the pension scheme than their previous birthday. Thus, the actuarial age for this person is 46 years.

Year	Total	Male	Female
2015	59.97	59.91	60.44
2016	60.24	60.16	60.89
2017	60.54	60.48	61.04
2018	61.22	61.14	61.82
2019	61.65	61.57	62.21
2020	62.14	62.06	62.74
2021	62.53	62.44	63.21
2022	63.14	63.04	63.87

Table 3.3: Average Age of Pensioners in 2015-2022 (years)

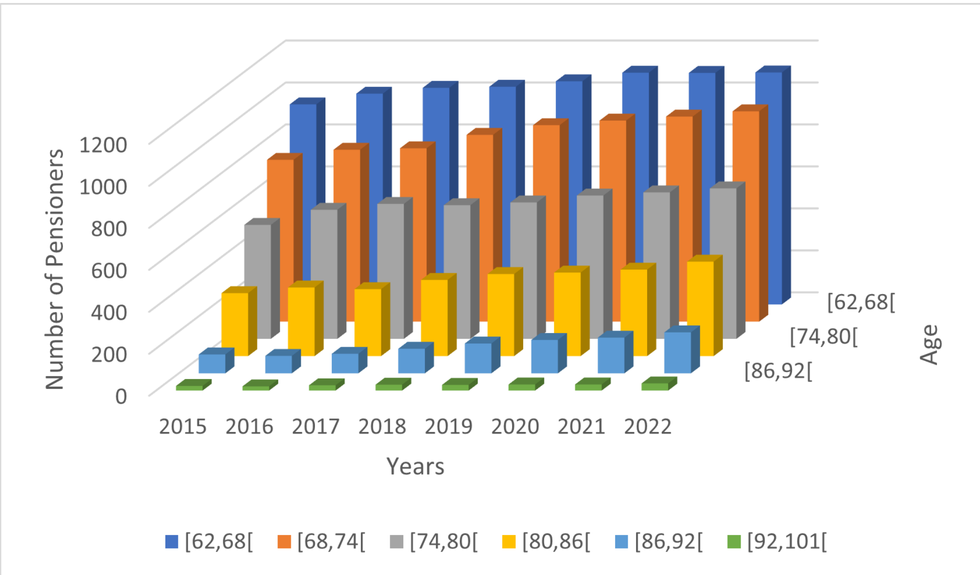

Figure 3.3: Pensioners by Age (62 to 100 Years) in 2015 - 2022

When we analyze the period from 2015 to 2022, we observe a consistent upward trend in the average age of pensioners, both for males and females, as shown in Table 3.3. Interestingly, the average age of female pensioners consistently exceeds that of their male

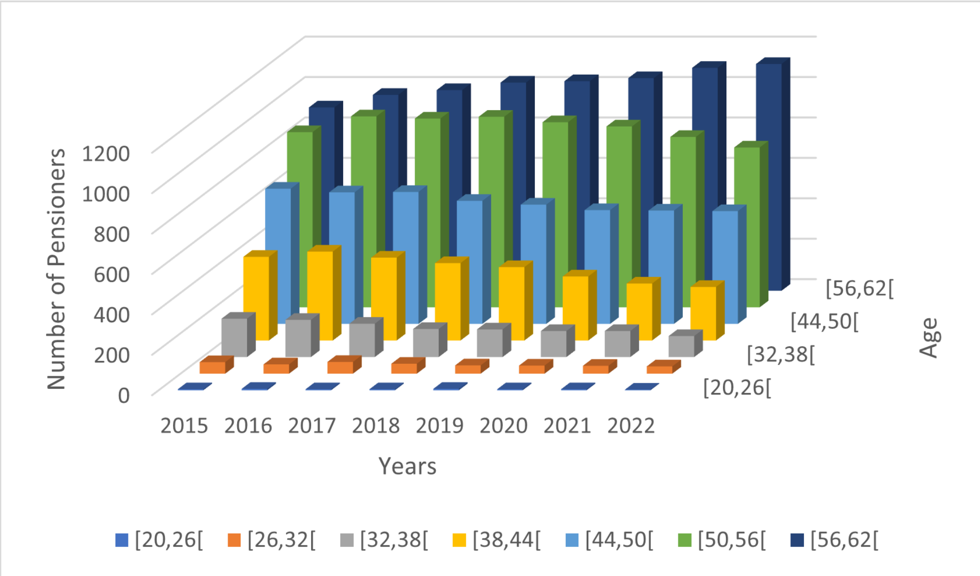

Figure 3.4: Pensioners by Age (20 to 61 Years) in 2015 - 2022

counterparts.
The rate of increase in the average age varies from year to year but generally shows a steady progression, with more significant changes observed in the later years of the dataset.

Figure 3.3 illustrates the age distribution of pensioners from 2015 to 2022, offering a clear representation of data within the age range 62 to 100. Additionally, Figure 3.4 further breaks down the distribution of pensioners between the ages of 20 to 61. This breakdown simplifies the interpretation of the figures.

These visualizations emphasize a noticeable trend: there is an increase in the number of pensioners aged 56 and above, accompanied by a corresponding decrease in the number of pensioners under the age of 56.

Regarding mortality, Figure 3.5 presents the distribution of deaths by age in the same years. However, there is limited information available for individuals below the age of 55. Consequently, constructing a comprehensive mortality table covering all age groups is not feasible.

3.4 Pensioners by Level of Incapacity

An essential component of our study on mortality is the Level of Incapacity, which is regulated by Decree-Law No. 352/2007 [7]. This decree-law addresses the intricate matter of medical-legal evaluation of bodily injuries and disabilities. The law recognizes the importance of safeguarding workers and treats labor-related disabilities differently from the general civil liabilities. To ensure fairness and accuracy, the liability introduces two tables for evaluating disabilities: one for labor liability and another for civil liability. These

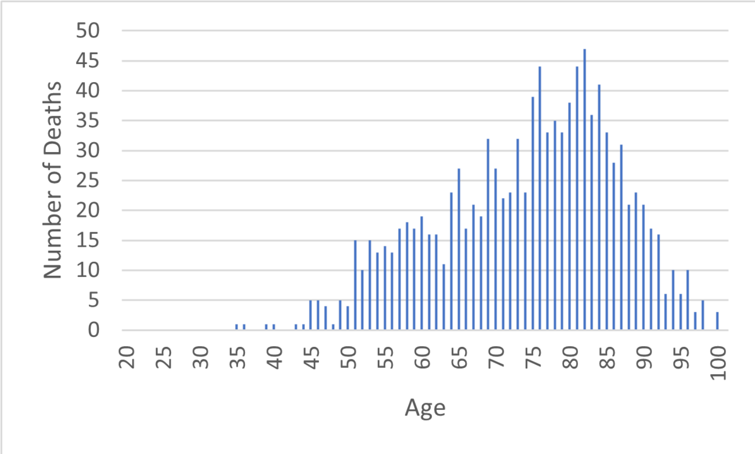

Figure 3.5: Pensioners' Mortality by Age in 2015 - 2022

tables are employed by medical specialists with expertise in legal medicine, or related fields, as appropriate tools for assessing disabilities.

In our data analysis, as depicted in Figure 3.6, we observe a notable increase of 40% in the number of pensioners with an incapacity level ranging from 5% to 30%. Additionally, there is a 15% increase in the number of pensioners with an incapacity level ranging from 30% to 50%, and a corresponding decrease of 5% in the number of pensioners with an incapacity level exceeding 50%. On average, 42% of pensioners exhibit an incapacity level exceeding 50%, while 38% of pensioners fall within the range of 30% to 50% incapacity.

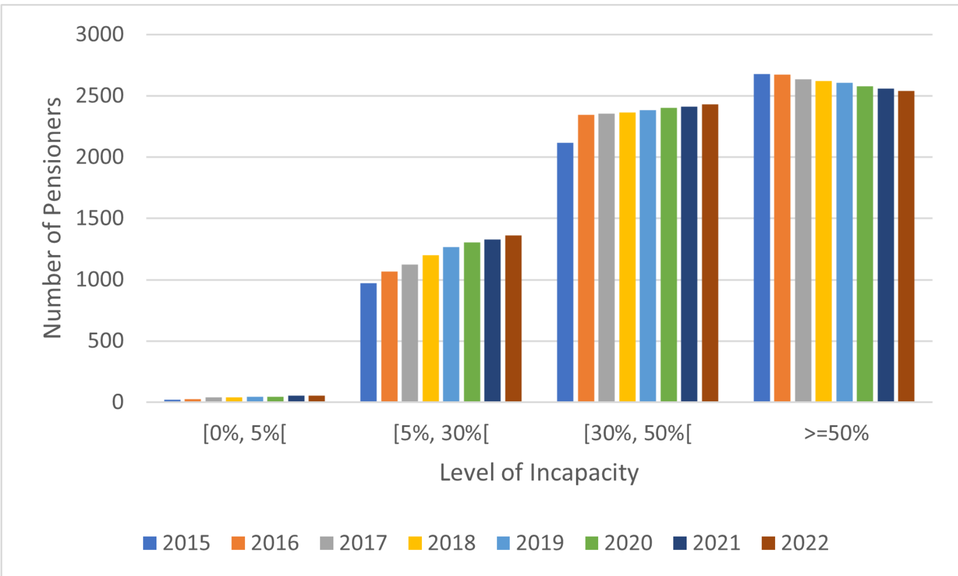

Figure 3.6: Pensioner by Level of Incapacity in 2015 - 2022

Figure 3.7 displays the number of deaths over time categorized by the level of incapacity. While a distinct mortality trend is not easily observable, it is noteworthy that, on average, 48% of deaths occur within the group exhibiting more than 50% incapacity, and 39% of deaths are observed in the group with incapacity levels ranging from 30% to 50%.

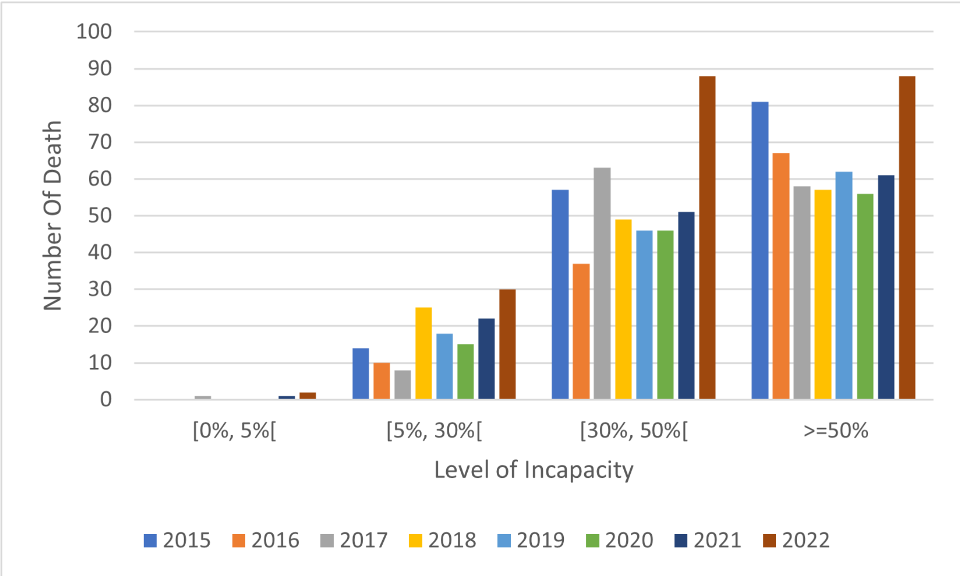

Figure 3.7: Mortality of Pensioner by Level of Incapacity in 2015 - 2022

When analyzing the average duration for which deceased pensioners received their pensions, we find interesting patterns. Among pensioners with more than 50% incapacity, the average duration of pension receipt was 28 years. In contrast, pensioners with an incapacity level ranging from 30% to 50% received pensions for an average of 26 years, while those with less than 30% incapacity received pensions for an average of 20 years.
The lower average duration for pensioners with incapacity levels less than 30% can be attributed to cases where the pensions are redeemable, leading to shorter periods of pension receipt.

3.5 Pensioners by Annual Wage

Incorporating annual wages as an explanatory variable requires us to normalize these wages with respect to Portugal's national minimum wage. We have access to data on the initial annual wage, when the pensioner began receiving their pension. Using this information, our goal is to determine the additional amount these pensioners received compared to the prevailing minimum wage during that period. Historical data on Portugal's minimum wages can be found on pordata[2].

First, let the proportional wage (PW) be defined as:

$$PW = \frac{\textit{Pensioner's Annual Wage}}{\textit{National Minimum Wage}}$$

[2]www.pordata.pt

Within our dataset, PW exhibits an average value of 2.65 and a median value of 1.7. Further insight into the distribution of PW can be gained by examining the Table 3.4, which allows us to classify PW into five distinct groups.

Level	Quantile
90%	4.0500
75%	2.6300
50%	1.7400
25%	1.3100
10%	1.0000

Table 3.4: Quantiles of the Proportional Wage

Table 3.5 provides details on the distribution of the number of pensioners and their corresponding mortality, based on the proportional wage (PW). Among pensioners, the group with the most deaths tends to have an annual salary less than 1.5 times the national minimum wage. Interestingly, those who earn between 3.5 and 4.5 times the minimum wage have a higher proportion of deaths relative to their income and most of those pensioners have a level of incapacity higher than 30%.

PW	Number of Pensioners	Number of Deaths	%
]0,1.5]	2321	80	3.45%
]1.5,2.5]	2165	74	3.42%
]2.5,3.5]	824	17	2.06%
]3.5,4.5]	415	18	4.34%
> 4.5	657	19	2.89%

Table 3.5: Mortality by Proportional Wage in 2022

Our goal is to investigate whether wages help elucidate mortality trends. As highlighted by Wolfson (2000)[22], there is accumulating evidence that residing in a society characterized by greater income inequalities predisposes to higher mortality rates, in general. However, widespread evidence also indicates that, on an individual level, higher income acts as a protective factor.

Chapter 4

Methodology section

The main objective of this study is to establish a meaningful link between the mortality rates observed in the Worker's Compensation portfolio under investigation and a well-established mortality table. To accomplish this, we will assess two distinct approaches that will guide us in identifying the most appropriate mortality table. Firstly, we will employ a classification and regression tree (CART), as explained by Friedman (2017)[9], enabling us to explore the interaction among variables that contribute significantly to estimating the mortality adjustment outlined by Benjamin (2008)[2].
Secondly, we will leverage a Generalized Linear Model (GLM), which takes into consideration factors such as age, gender, and level of incapacitation to search into their collective influence on mortality adjustments within the target population. This methodological approach, as described by McCullagh (1989) [17], will yield valuable insights into the underlying determinants of mortality rates.

4.1 CART Model Approach

Classification and regression tree models provide a flexible method for specifying the conditional distribution of a variable Y, given a vector of predictor variables X. Such models use a binary tree to recursively partition the predictor space into subsets in which the distribution of Y is successively more homogeneous. The terminal nodes of the tree correspond to the distinct regions of the partition, and the partition is determined by splitting rules associated with each of the internal nodes. By moving from the root node through to the terminal nodes of the tree, each observation is then assigned to a unique terminal node where the conditional distribution of Y is determined. CART models were popularized in the statistical community by the seminal work of Breiman, Friedman, Olshen, and Stone (1984)[5].

As described by Friedman (2017)[9], which will be our main reference in this section, let us consider a regression problem involving a continuous response variable Y and input variables X_1 and X_2, each taking values within the unit interval. The initial step involves partitioning the space into two regions and modeling the response as the mean of Y within each region. The choice of variable and split-point is made to achieve the best fit. Subsequently, one or both of these regions can be further divided into additional regions. This iterative process continues until a specified stopping rule is fulfilled.

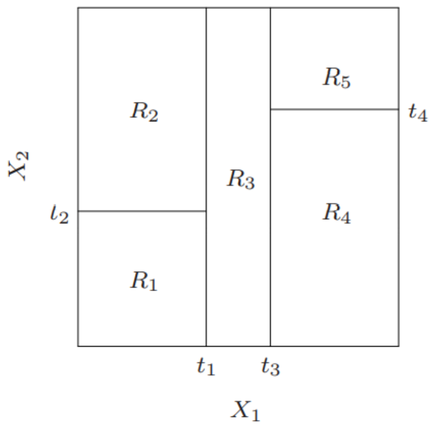

Figure 4.1: Example of a partition of a two-dimensional feature space performed by CART method (Friedman,2017,p.306)

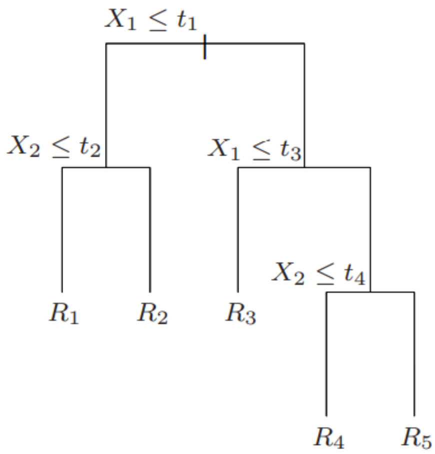

Figure 4.2: Tree corresponding to the partition in Figure 4.1 (Friedman,2017,p.306)

For example, in Figure 4.1, we first split the region at $X_1 = t_1$. Then the region $X_1 \leq t_1$ is split at $X_2 = t_2$ and the region $X_1 > t_1$ is split at $X_1 = t_3$. Finally, the region $X_1 > t_3$ is split at $X_2 = t_4$. The result of this process is a partition of the space (X_1,X_2) into the five regions R_1, R_2, . . . , R_5 shown in Figure 4.1. This partition of the space can also be represented through a tree graph, as in Figure 4.2.
The corresponding regression model predicts Y with a constant c_m in region R_m, that is:

$$Y = \hat{f}(X_1, X_2) = \sum_{m=1}^{5} c_m I\{(X_1, X_2) \in R_m\}. \tag{4.1}$$

Next, we must explore how to construct a regression tree. The goal here is to have the algorithm automatically determining the variables for splitting, the points at which to split, and the overall structure or shape of the tree. Consider p inputs and a response, for each of N observations: that is, (x_i,y_i) for i = 1,2,....,N, with x_i = (x_{i1},x_{i2},....,x_{ip}). Suppose first that we have a partition into M regions R_1, R_2, . . . , R_M, and we model the response as a constant c_m in each region:

$$f(x) = \sum_{m=1}^{M} c_m I\{x \in R_{\mathrm{m}}\}. \tag{4.2}$$

If we adopt the sum of squares $\sum(y_i - f(x_i))^2$ as the minimization criterion, it is easy to see that the optimal $\hat{c}_m$ is simply the average of y_i in region R_m:

$$\hat{c}_m = ave(y_i|x_i \in R_{\mathrm{m}}). \tag{4.3}$$

Now, identifying the optimal binary partition with the least sum of squares is typically a computationally challenging task. Therefore, we employ a greedy algorithm for this purpose. We begin with the entire dataset and then evaluate a splitting variable *j* and a split point *s*. Next, we define two half-planes using this split:

$$R_1(j,s) = \{X|X_{\rm j} \leq {\rm s}\} \text{ and } R_2(j,s) = \{X|X_{\rm j} > {\rm s}\}. \tag{4.4}$$

Then, we seek the splitting variable *j* and split point *s* that

$$\min_{j,s} = [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2]. \tag{4.5}$$

For any choice *j* and *s*, the inner minimization is solved by

$$\hat{c}_1 = ave(y_i|x_i \in R_1(j,s)) \text{ and } \hat{c}_2 = ave(y_i|x_i \in R_2(j,s)). \tag{4.6}$$

When it comes to each splitting variable *j*, we can efficiently identify the split point *s*. By scanning through all the inputs, we can feasibly determine the best pair (*j*, *s*).
Having found the best split, we partition the data into the two resulting regions and repeat the same splitting process on each of them, and so on.
The question that arises is: how large should we grow the tree?. Clearly a very large tree might over fit the data, while a small tree might not capture the important structure. The preferred strategy to obtain the optimal tree size is to grow a large tree T_0, stopping the splitting process only when some minimum node size (say 5) is reached. A node consists of the elements from *Y* that belong to a specific subset of the feature space, which is defined by the corresponding splits of the input variables. Then this large tree is pruned using cost-complexity pruning, which we describe in the following.
We define a subtree $T \subset T_0$ to be any tree that can be obtained by pruning T_0, that is by collapsing any number of its internal (non-terminal) nodes. We index terminal nodes by *m*, with node *m* representing region R_m. Let $|T|$ denote the number of terminal nodes in T. Letting

$$N_m = \#\{x_i \in R_{\rm m}\}, \tag{4.7}$$

$$\hat{c}_{\rm m} = \frac{1}{N_m} \sum_{x_i \in R_m} y_i, \tag{4.8}$$

$$Q_m(T) = \frac{1}{N_m} \sum_{x_i \in R_m} (y_i - \hat{c}_{\rm m})^2, \tag{4.9}$$

we define the cost complexity criterion

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_{\rm m} Q_m(T) + \alpha|T|. \tag{4.10}$$

The idea is to find, for each α, the sub-tree $T_\alpha \subseteq T_0$ to minimize $C_\alpha(T)$. The tuning parameter, α ($\alpha \geq 0$), plays a crucial role in balancing the size of the tree and how well it fits the data. When α is large, it leads to smaller trees (T_α), and conversely, smaller α values result in larger trees. When α equals 0, we get the full tree T_0.
To find the unique smallest subtree T_α that minimizes $C_{\alpha(T)}$ for a given α, we employ a method called weakest link pruning. This involves iteratively collapsing the internal node

that contributes the smallest per-node increase in the $\sum_m N_m Q_m(T)$. We continue this process until we arrive at the single-node (root) tree. This procedure generates a finite sequence of subtrees, and it is guaranteed that this sequence contains T_α. More details can be found in works like Breiman et al. (1984)[6].
To estimate the optimal α value, we employ five or tenfold cross-validation. We choose the value denoted as $\hat{\alpha}$ that minimizes the sum of squared errors in cross-validation. The final tree we use is $T_{\hat{\alpha}}$, where $\hat{\alpha}$ is the estimated α value.

If the target is a classification outcome taking values 1, 2, . . . , K, the only changes needed in the tree algorithm pertain to the criteria for splitting nodes and pruning the tree. For the regression we consider the squared-error node impurity measure $Q_m(T)$ defined in Equation (4.9). However this is not suitable for classification. For this reason, let us consider, in a node m representing a region R_m with N_m observations, the next quantity

$$\hat{p}_{mk} = \frac{1}{N_{\mathrm{m}}} \sum_{x_i \in R_m} I(y_i = k), \tag{4.11}$$

which represents the proportion of class k observations in node m. We classify the observations in node m to class $k(m) = \arg\max_k \hat{p}_{mk}$, which is the class that has most observations in node m.
In the CART algorithm, we calculate the impurity of a node using an alternative metric, known as the Gini index (GI), expressed as:

$$GI = \sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'}. \tag{4.12}$$

Gini index measures how often a randomly chosen element of a set would be incorrectly labeled if it were labeled randomly and independently, according to the distribution of labels in the set.

In summary, CART models provide a powerful tool for partitioning data into regions and modeling the conditional distribution of a response variable. By iteratively splitting data based on predictor variables, they can capture complex relationships within the data while also being capable of pruning to avoid overfitting.

4.2 Generalized Linear Model approach

A Generalized Linear Model (GLM) is a versatile statistical tool used to model relationships between variables. It is especially valuable in situations where the response variable does not follow a normal distribution or has a varying variance. This modeling framework, as outlined by authors like McCullagh (1989) [17], which is the main reference in this section, serves as a robust option when traditional linear regression assumptions are not met. GLMs expand upon the classical linear regression model, allowing it to handle a broader range of data types and relationships.

GLM Framework:

In a GLM, we work with a dataset comprising response values y_i and corresponding covariates x_i. It is crucial to note that these responses are observed independently for fixed covariate values $(x_{i1}, x_{i2}, \ldots, x_{ip})$. The influence of covariate variables on the distribution of y_i is captured through a linear function represented as:

$$\eta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}. \tag{4.13}$$

The expected response mean, denoted as $\mu = E(y_i)$, is linked to the linear predictor η_i through a smooth and invertible link function:

$$h(\mu_i) = \eta. \tag{4.14}$$

The inverse of this function, often referred to as $g(t) = h^{-1}(t)$, is known as the inverse link function.

Furthermore, the distribution of the response variable y_i belongs to the exponential family, characterized by the density function:

$$f(y_i|\beta, \phi) = \exp\left[\frac{A_i}{\phi} y_i \vartheta_i - \gamma(\theta_i) + \tau\left(y_i, \frac{\phi}{A_i}\right)\right]. \tag{4.15}$$

Here, ϕ represents the scale parameter, also referred to as the dispersion parameter. A_i is a known constant, and $\theta(.) = \theta(\phi_i)$ is a function of the linear predictor ϕ_i.

The exponential family comprises a wide range of probability distributions, such as the Normal, Poisson, inverse Gaussian, gamma, binomial, and exponential distributions. Within a GLM, the choice of both the link function h and the form of the response distribution, represented by the function γ, play pivotal roles and together determine the model's characteristics. In GLM, we can interpret the slope β_p as representing the expected change in $h(E(y))$ when the p^{th} covariate experiences a unit increase. This interpretation provides insight into how the covariates influence the response variable within the model.

Distribution Examples:

Normal Distribution: $\phi = \sigma^2$ and $V(\theta_i) = 1$, leading to

$$\theta_i = \mu_i = \eta_i, \textit{ and } \gamma(\theta_i) = \frac{\theta_i^2}{2}. \tag{4.16}$$

Poisson Distribution: Setting $\phi = 1$ and $V(\theta_i) = \theta$ results in

$$\theta = \eta_i, \textit{ and } \gamma(\theta_i) = e^{\theta_i}. \tag{4.17}$$

Binomial Distribution: For the binomial distribution, we set $\phi = 1$ and $V(\theta_i) = \theta_i(1-\theta_i)$, giving us

$$\theta = \eta_i, \textit{ and } \gamma(\theta_i) = \log\left(\binom{n_1}{n_i y}\right). \tag{4.18}$$

Gamma Distribution: Here, with $\phi = \frac{1}{\alpha}$ and $V(\theta_i) = \theta_i^2$, we have

$$\theta = \eta_i, \textit{ and } \gamma(\theta_i) = \log(\theta_i). \tag{4.19}$$

Link Functions:

The link function serves as the bridge connecting the expected value of the response variable to the linear predictor. It transforms the linear combination of predictor variables into an appropriate range for the response variable's distribution. Commonly used link functions include:

$$\textit{Logit Link: } h(t) = \log\left(\frac{t}{1-t}\right). \tag{4.20}$$

$$\textit{Probit Link: } h(t) = \phi^{-1}(t), \textit{ where } \phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{\frac{z^2}{2}} dz. \tag{4.21}$$

$$\textit{Log-Link: } h(t) = \log(t). \tag{4.22}$$

$$\textit{Square Root Link: } h(t) = \sqrt{t}. \tag{4.23}$$

$$\textit{Inverse Link: } h(t) = \frac{1}{t}. \tag{4.24}$$

In contrast to traditional linear regression, which assumes additivity concerning covariates for y_i, GLM only requires that some transformation of y_i, represented as $g(y_i)$, exhibits additivity concerning the covariates. This flexibility empowers GLM to excel in a wide range of modeling scenarios.

Generalized Linear Models (GLMs) offer several advantages, including the ability to handle diverse data types, model non-linear relationships, and incorporate multiple predictor variables simultaneously. They provide a versatile framework for statistical analysis and are widely applied in various fields, such as biology, economics, epidemiology, and social sciences.

Chapter 5

Results and Discussion

We used R to compute both the CART and GLM models. In particular, for the CART model, we relied on the 'caret' package developed by Kuhn (2008) [15]. As for our data, we collected information on 74,359 pensioners between 2015 and 2022, noting a total of 1,553 observed deaths.

5.1 CART Results

The decision to use the CART model was based on its suitability for non-linearly distributed data. It allows for a straightforward analysis and offers an insightful means of interpreting relationships between variables. Furthermore, the CART model proves valuable for forecasting the desired response variables, specifically in our case, the ratio of observed to expected mortality.

To calculate the expected mortality, we utilized the most recent Portuguese mortality table available for the respective year of reference. Specifically, for pensioners exposed in 2015, we employed the mortality table $INE12\_14$. For those exposed in 2016, we used the $INE13\_15$ table, and so on, until the year 2022, where we calculated it using the $INE19\_21$ mortality table. These specific tables were chosen because they accurately reflect the mortality patterns observed in the studied population.

Our objective is to examine how the variables of Age, Sex, *Incapacity Level*, and PW can assist in explaining the necessary mortality adjustment. This relationship can be expressed as follows:

$$\frac{\textit{Number Of Deaths}}{\textit{Expected Deaths}} \sim Age + Sex + \textit{Incapacity Level} + PW$$

We began by partitioning randomly the data into two subsets: training data and test data. With the training data, we implemented a crucial process known as cross-validation, using the R package "Caret" [15]. This procedure allowed us to evaluate the performance of our regression tree and make informed decisions regarding the best model.

In the cross-validation process, we first determined the minimum number of observations required in each terminal node, represented by the parameter "$minbucket$." We tested with four different values: 500, 1000, 2000, and 4000. For each $minbucket$ value, we trained the regression tree while adjusting the tuning parameter α ranging from 0.00001 to 0.01 in steps of one unit.

Table 5.1 presents the optimal outcomes achieved during each cross-validation iteration. The model selection criterion used was the root mean squared error (RMSE), with a preference for the model displaying the smallest RMSE. This rigorous process allowed us to identify the regression tree model that best fit our data and its underlying patterns.

α	minbucket	RMSE
0.01	500	14.38521
0.00036	1000	14.26767
0.00035	2000	14.34649
0.00017	4000	14.3711

Table 5.1: Cross-validation

As illustrated in Table 5.1, we opted for a α value of 0.00036 and set the $minbucket$ to 1000. This selection was made to ensure that the trees achieved their maximum complexity.

Figure 5.1 reports the regression tree obtained for the mortality adjustment against the variables Age, $Gender$, *Incapacity Level*, and *PW*. In this case, the CART algorithm has divided the feature space in 5 regions, and the values predicted for each region, together with the percentage of observable data, are reported in the squares of Figures 5.1 which represent the final nodes of the trees. The final outcome of this segmentation prioritizes the *Incapacity Level*, Age and PW influential factors, while not assigning significance to the variable $Gender$.

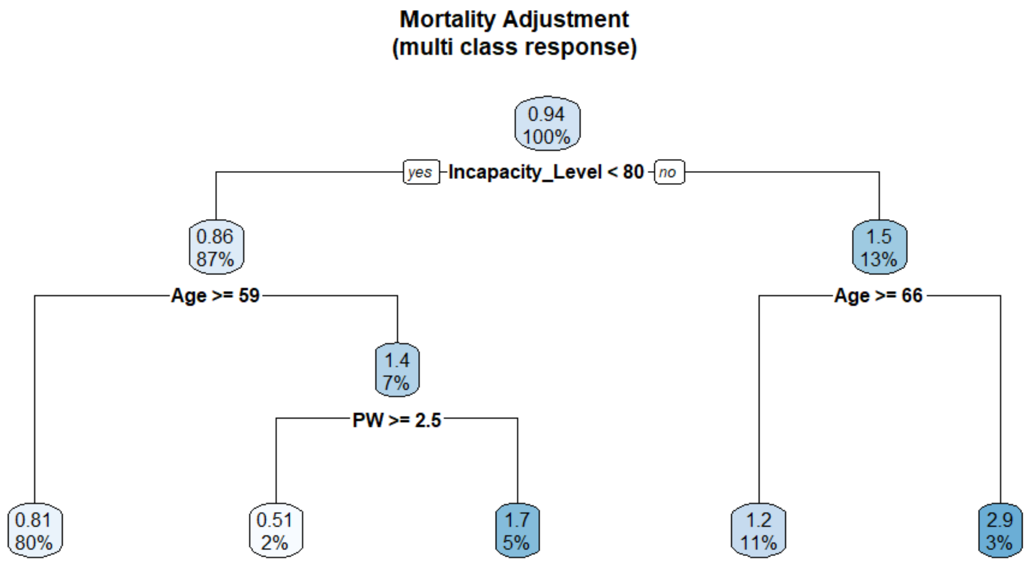

Figure 5.1: Regression Tree for Mortality Ratio Adjustment

We can interpret the tree as follows:

- If the pensioner's *Incapacity Level* is below 80% and they are over 59 years old, we estimate a mortality adjustment of 0.81 (Node 1).
- For pensioners with *Incapacity Level* below 80%, aged 59 or younger, and with a PW greater than 2.5, we estimate a mortality adjustment of 0.51 (Node 2).
- For pensioners with *Incapacity Level* below 80%, aged 59 or younger, and PW equal to or less than 2.5, the estimated mortality adjustment is 1.7 (Node 3).
- On the other side of the tree, for pensioners with an *Incapacity Level* above 80% and aged over 66, the estimated adjustment is 1.2 (Node 4).
- For the remaining cases, we estimate a mortality adjustment of 2.9 (Node 5).

The results from the regression tree highlight intriguing patterns. In Nodes 2 and 3, we see that pensioners with greater financial resources tend to have lower mortality adjustments. Many of these wealthier pensioners can afford better living conditions. In this group, we also find football players who, on average, receive higher salaries than the general population. Despite the potential for work-related injuries that could fully disable them in their sport, they are often able to engage in other activities. On the flip side, pensioners with fewer financial resources tend to experience higher mortality adjustments.

Now, when we look at Node 5, we notice that it has a higher mortality adjustment compared to Node 4. This is intriguing because Node 4 represents older pensioners with a high level of disability. One possible interpretation is that for pensioners with a high incapacity level, the likelihood of an earlier death due to the limitations of their disability is greater than for those who have already lived with the disability for several years.

Employing the estimated adjustments derived from the tree depicted in the final nodes, as shown in Figure 5.1, we applied them to the test data, resulting in the outcomes presented in the Table5.2.

Tree Node	Adjustment Prediction	Expected Mortality (1)	Mortality Adjusted (2)	Observed Deaths (3)	(3) - (1)	(3) - (2)
Node 1	0.811	484	392	456	-28	64
Node 2	0.505	11.2	5.67	17	5.8	11.33
Node 3	1.74	30.4	52.8	44	13.6	-8.8
Node 4	1.15	65.5	75.6	85	19.5	9.4
Node 5	2.86	15.9	45.7	29	13.1	-16.7

Table 5.2: Adjustment Prediction Table

The *Mortality Adjusted* value is the result of multiplying the *Adjustment Prediction* by the expected mortality. When we consider Node 3 and Node 4, the adjustments appear to yield more accurate estimates compared to the expected mortality alone. However, for the remaining nodes, the expected mortality without any adjustment tended to be closer to the observed mortality.

5.2 GLM results

To predict the number of deaths, we will use Logistic Regression since we can categorize our pensioners' status as either alive or dead. We chose Logistic Regression because it offers simple and easy-to-understand results, especially for modeling binary variables. Similar to what we did for the CART model, we randomly divided the data into two subsets: training data and a test set. For estimating expected mortality, we applied the same tables as described in the CART section.
Additionally, we can utilize the classifications obtained from the CART model to group the *Incapacity Level* into two categories: above 80% and below 80%, and PW into two categories: above 2.5 and below 2.5. By incorporating these categories into our model, we can then compare the following models:

$$\textit{Logistic 1:} \quad \textit{Death} \sim Age + Sex + \textit{Incapacity Level} + PW$$
$$\textit{Logistic 2:} \quad \textit{Death} \sim Age + Sex + \textit{Incapacity Level} \geq \textit{80} + \textit{PW} \geq \textit{2.5}$$

Upon comparing the two models, we can draw the following conclusions: *Logistic 2* exhibits lower AIC and BIC values than *Logistic 1*. Additionally, the bayes factor, which indicates the strength of evidence in favor of a specific model, supports *Logistic 2*, as illustrated in Table 5.3.

Model	*AIC*	*BIC*	*Bayes Factor*
Logistic 1	4990.438	5031.634	0.235
Logistic 2	4987.543	5028.739	4.251

Table 5.3: Model Comparison Statistics (*Logistic 1* vs *Logistic 2*)

Table 5.4 presents the coefficients of the *Logistic 2* model. It is worth noting that the variable *PW* $\geq$ *2.5* has a p-value greater than 5%. To assess the model's adequacy, we compared it with a model that excludes this variable, denoted as *Logistic 3*, as described in Table 5.5. Interestingly, we observed that the *Logistic 3* model performs better than the *Logistic 2* model.

Variable	*Estimate*	*Std. Error*	*z value*	*Pr(> |z|)*
(Intercept)	-10.29438	0.32991	-31.203	$< 2 \times 10^{-16}$
Age	0.08785	0.00390	22.525	$< 2 \times 10^{-16}$ ***
SexM	0.46648	0.15610	2.988	0.002804
Incapacity_Level>= 80	0.38967	0.11059	3.524	0.000426
PW1>= 2.5	-0.16660	0.09683	-1.721	0.085338

Table 5.4: *Logistic 2* Coefficients

Model	*AIC*	*BIC*	*Bayes Factor*
logistic2	4987.543	5028.739	0.027
logistic3	4988.567	5021.524	36.877

Table 5.5: Model Comparison Statistics (*Logistic 2* vs *Logistic 3*)

We can see the coefficients of the model *Logistic 3* on Table 5.6.

Variable	*Estimate*	*Std. Error*	*z value*	*Pr(*$>|z|$*)*
(Intercept)	-10.333152	0.329790	-31.332	$< 2 \times 10^{-16}$
Age	0.087950	0.003909	22.497	$< 2 \times 10^{-16}$
SexM	0.443609	0.155540	2.852	0.00434
Incapacity_Level$>= 80$	0.410291	0.109954	3.731	0.00019

Table 5.6: *Logistic 3* Coefficients

Table 5.7 displays the results of our model, *Logistic 3*, with the test dataset. Notably, when we compare the predicted number of deaths to the observed values overall, the expected number of deaths based on the mortality tables provides a better explanation for mortality behavior. However, when we break down the projected number of deaths by *Incapacity Level*, some interesting patterns emerge. Specifically, for individuals with an *Incapacity Level* greater than or equal to 80%, the model appears to provide a better explanation for mortality.

Incapacity Level	*Observed Deaths (1)*	*Logistic 3 (2)*	*Expect Deaths (3)*	*(2) - (1)*	*(3) - (1)*
< 80	506	461	523	-45	17
$>= 80$	124	105	78.1	-19	-45.9
Total	630	566	601.1	-64	-28.9

Table 5.7: Mortality Projection (*Logistic 3* Model)

In conclusion, our model cannot be used to construct a complete mortality table for our pensioners. However, the results of our model have highlighted the significance of adjusting mortality for pensioners with a high level of incapacity. In such cases, the expected deaths based on the mortality tables consistently underestimate the observed deaths.

5.3 Mortality Deviation

Before we proceed, the following definitions are necessary for our analysis:

- m_x,$x = 0$ to 100, represents the number of members and former members in a mortality experience study with ages from 0 to 100. Integral ages are assumed to be used for the purpose of the experience study, a member's actual age at the time of the study rounded to the nearest integral age.

- q_x^E, $x = 0$ to 100, represents the mortality rate at age x based on the mortality table currently employed for valuation purposes.
- d_{xj}, $x = 0$ to 100 and $j = 1$ to m_x, is equal to 1 if the j^th member or former member at age x dies during the year, and zero otherwise.

To illustrate the required adjustment in mortality rates, we can compare the actual number of deaths with the total expected deaths. To be more specific, we will define $\hat{f}$ as the ratio of actual to total expected death counts across all ages, as:

$$\hat{f} = \frac{\sum_{x=0}^{100} \sum_{j=1}^{m_x} d_{xj}}{\sum_{x=0}^{100} \sum_{1}^{m_x} q_x^E} \tag{5.1}$$

We will apply this calculation to the following mortality tables: $INE10\_12$, $INE12\_14$, $INE13\_15$, $INE14\_16$, $INE15\_17$, $INE16\_18$, $INE17\_19$, $INE18\_19$, and $INE19\_21$. If the mortality table accurately represents the true underlying mortality rates of plan members and former members, the ratios of actual to expected deaths should closely approach 1, as indicated by the black line in Figure 5.2. In the chart, it becomes evident that the ratio of actual to expected deaths is predominantly zero for individuals below the age of 50, primarily due to the limited availability of data in this age range. Beyond the age of 50, the actual deaths closely align with the expected deaths.

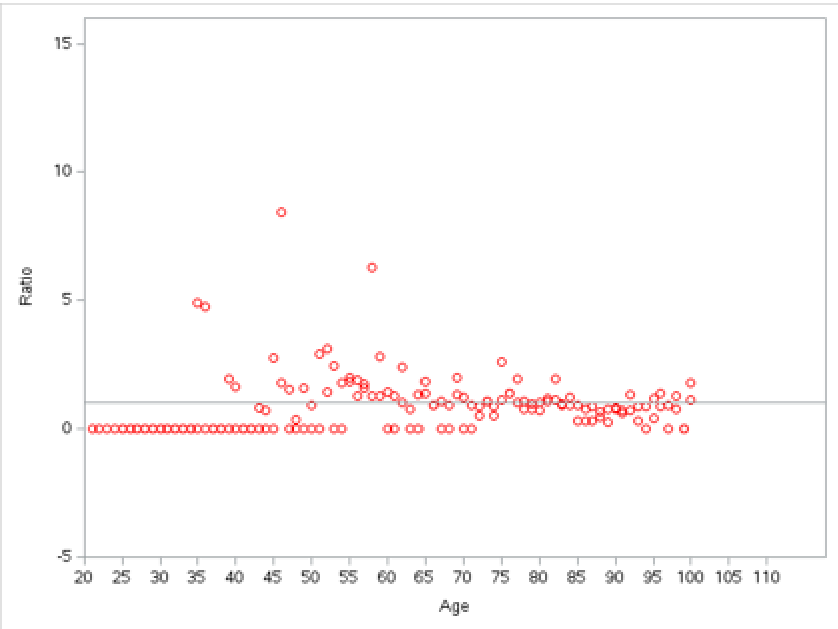

Figure 5.2: Actual-to-Expected Deaths Ratio for Pensioners by Age

Table 5.8 presents necessary mortality adjustments for various years concerning different mortality tables. For instance, the mortality table $INE14\_16$ predicted a total of 122 deaths for the year 2017, while the observed deaths for that year were 130. If this table had been implemented in that year, it would have required a 6% adjustment to accurately

Year	Number of Deaths	Mortality Table	Expected Mortality	$\hat{f}$
2015	152	INE12_14	116.08	1.31
2016	114	INE13_15	120.05	0.95
2017	130	INE14_16	122.14	1.06
2018	131	INE15_17	129.80	1.01
2019	126	INE16_18	136.60	0.92
2020	117	INE17_19	138.70	0.84
2021	135	INE18_20	140.52	0.96
2022	208	INE19_21	156.72	1.33

Table 5.8: Observed Mortality to the Expected Mortality

reflect the mortality data.

The data in Table 5.8 provides valuable insights into how well the INEs (Instituto Nacional de Estatística) mortality tables can explain observed mortality patterns. Overall, these mortality tables exhibit a strong fit for explaining mortality among WC pensioners. However, there are exceptions, notably in the years 2022 and 2015, where higher adjustments are required. If we exclude these two years with higher adjustments, the average adjustment needed would be a decrease by 4% in the mortality rate.

We can leverage the insights from both the CART model and the GLM to refine the necessary adjustment for the mortality table. Given that the average adjustment implies a reduction in the mortality rate, we can apply this reduction to pensioners with an incapacity level below 80%. This approach is justified by our earlier findings, which indicated that pensioners with a higher level of incapacity exhibit different mortality patterns compared to the general population.

Year	**Number Of Deaths (1)**	**Mortality Table**	**Expected No. of Deaths (2)**	**$\hat{\mathbf{f}}$ (2)**	**Adjusted Mortality (3)**	**$\hat{\mathbf{f}}$ (3)**
2016	114	INE13_15	120.05	0.95	114.75	0.99
2017	130	INE14_16	122.14	1.06	116.75	1.11
2018	131	INE15_17	129.80	1.01	124.06	1.06
2019	126	INE16_18	136.60	0.92	130.56	0.97
2020	117	INE17_19	138.70	0.84	132.57	0.88
2021	135	INE18_20	140.52	0.96	134.31	1.01
Total	753		787.80	0.96	753	1.00

Table 5.9: Adjusted Mortality

Applying the average adjustment of 0.96, calculated by considering the years from 2016 to 2021, results in an overall improvement, as demonstrated in Table 5.9. It is crucial to exclude outliers from this calculation because when determining the required reserves, we are essentially forecasting how long our pensioners will survive. Overestimating mortality can lead to higher future costs due to expected premature death, resulting in additional pension payouts beyond our calculations. Conversely, underestimating mortality can lead to excessive costs as we anticipate longer payment periods.

Chapter 6

Conclusion

To determine the most fitting mortality table for pension reserves in Workers' Compensation, we explored the potential of more advanced methods like the CART model and GLM to predict the necessary adjustments or even construct our own mortality experience. However, we encountered limitations due to the amount of data available, which posed a significant challenge throughout this study. Despite these limitations, these models offered valuable insights into how variables like age, sex, level of incapacity, and proportional wage can aid in explaining mortality patterns. It is worth noting that with more extensive data on Workers' Compensation pensioners in the future, we may be able to provide more definitive results.

Utilizing a practical approach outlined by Benjamin (2008) [2], we estimate a adjustment of 0.96 on the expected morality. When we excluded the periods with elevated death rates, some of which were attributed to the COVID-19 pandemic, we arrived at an unexpected finding. Contrary to our initial expectations, the adjustment needed is not an increase in the mortality rate; instead, it indicates a decrease. Our CART model also highlighted that pensioners with a high level of incapacity would necessitate a more substantial adjustment. As a result, we recommend not applying this adjustment to those pensioners.

A valuable enhancement for this study would involve improving the GLM model by attempting to model the level of adjustment using a Gamma distribution that is widely used in modeling continuous, non-negative and positive-skewed data, such as insurance claims and survival data. Furthermore, the continuation of this study is essential to ensure the company's calculations for reserves are based on the most precise assumptions. This accuracy in representing mortality is vital to provide pensioners with the best possible protection for their pensions.

Bibliography

[1] Assembleia da República. Lei n.º 98/2009 de 4 de setembro, Setembro 2009.

[2] Gavin Benjamin. Selecting mortality tables: A credibility approach. *Research paper*, 2008.

[3] Aviroop Biswas, Shireen Harbin, Emma Irvin, Heather Johnston, Momtaz Begum, Maggie Tiong, Dorothy Apedaile, Mieke Koehoorn, and Peter Smith. Differences between men and women in their risk of work injury and disability: A systematic review. *American journal of industrial medicine*, 65(7):576–588, 2022.

[4] Gary Blumsohn. Levels of determinism in workers compensation reinsurance commutations. In *AdS*, page 55. Citeseer, 1997.

[5] Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. Cart. *Classification and regression trees*, 1984.

[6] Leo Breiman and Ross Ihaka. *Nonlinear discriminant analysis via scaling and ACE*. Department of Statistics, University of California Davis One Shields Avenue ..., 1984.

[7] Direcção-Geral da Segurança Social. Ministério do trabalho e da solidariedade social. *Aviso (extracto)*, 29:830, 2007.

[8] Ronald E Ferguson. Actuarial note on workmen's compensation loss reserves. *PCAS LVIII*, 1971.

[9] Jerome H Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. springer open, 2017.

[10] William R Gillam. Injured worker mortality. In *Proceedings of the Casualty Actuarial Society*, volume 80, pages 34–54, 1993.

[11] Lijia Guo. Predictive modeling for advanced age mortality. *Living to*, 100, 2008.

[12] Syed S Islam, Angela M Velilla, Edward J Doyle, and Alan M Ducatman. Gender differences in work-related injury/illness: analysis of workers compensation claims. *American journal of industrial medicine*, 39(1):84–91, 2001.

[13] Brian A Jones, Craig J Scukas, Kathryn S Frerman, Melissa S Holt, and Vicki A Fendley. A mortality-based approach to reserving for lifetime workers' compensation claims. In *Casualty Actuarial Society E-Forum. Retrieved December*, volume 20, page 2019, 2013.

[14] James Kahn. Reserving for runoff operations a real life claims specific methodology for reserving a workers compensation runoff entity. In *Casualty Actuarial Society Forum*, pages 139–210. Citeseer, 2002.

[15] Max Kuhn. Building predictive models in `R` using the `caret` package. *Journal of Statistical Software*, 28(5):1–26, 2008.

[16] Joseph LaDou. The european influence on workers' compensation reform in the united states. *Environmental health*, 10(1):1–10, 2011.

[17] Peter McCullagh and John A Nelder. Generalized linear models 2nd edition chapman and hall. *London, UK*, 1989.

[18] Jessica Mosher. Mortality and the provision of retirement income: The impact of the covid-19 pandemic on mortality. 2023.

[19] Carlos Eduardo Barrenho da Rosa. *An internal model for workers compensation*. PhD thesis, Instituto Superior de Economia e Gestão, 2012.

[20] Richard H Snader. Reserving long term medical claims. In *Proceedings of the Casualty Actuarial Society*, volume 74, pages 322–353, 1987.

[21] Lee R Steeneck. Actuarial note on workmen's compensation loss reserves-25 years later. In *Casualty Actuarial Society Forum*, pages 245–271, 1996.

[22] Michael C Wolfson, George Kaplan, John Lynch, Nancy Ross, and Eric Backlund. Relation between income inequality and mortality: empirical demonstration. *Western Journal of Medicine*, 172(1):22, 2000.