*Nishchay Pandya*

*12/10/2024*

*Mortality of Chilean pensioners*

Table of Contents

[Descriptive Analysis of the Profile of Annuitants 2](#_Toc163914454)

[Profile of Annuitants 2](#_Toc163914455)

[Survival Analysis 3](#_Toc163914456)

[Overview and Key Assumptions 3](#_Toc163914457)

[Kaplan-Meier Estimate 3](#_Toc163914458)

[Cox Proportional Hazard Model 4](#_Toc163914459)

[5](#_Toc163914460)

[Conclusion 5](#_Toc163914461)

[Graduation of Unisex Life Table 6](#_Toc163914462)

[Rationale 6](#_Toc163914463)

[Comparison with Chilean Life Tables 6](#_Toc163914464)

[Ethical Implications of Unisex Annuity Pricing 7](#_Toc163914465)

[Technical Appendix 8](#_Toc163914466)

[Survival Analysis 8](#_Toc163914467)

[Kaplan-Meier Curve 8](#_Toc163914468)

[8](#_Toc163914469)

[Interpretation of Coefficients 8](#_Toc163914470)

[Graduation 9](#_Toc163914471)

[Methodology 9](#_Toc163914472)

[Chi-square Test 9](#_Toc163914473)

[Standardized Deviations Test 10](#_Toc163914474)

[Serial Correlations Test 10](#_Toc163914475)

[Grouping of Signs Test 11](#_Toc163914476)

[Signs Test 11](#_Toc163914477)

[Cumulative Deviations Test 12](#_Toc163914478)

[Graphs of Other Graduation Methods 12](#_Toc163914479)

# **Descriptive Analysis of the Profile of Annuitants**

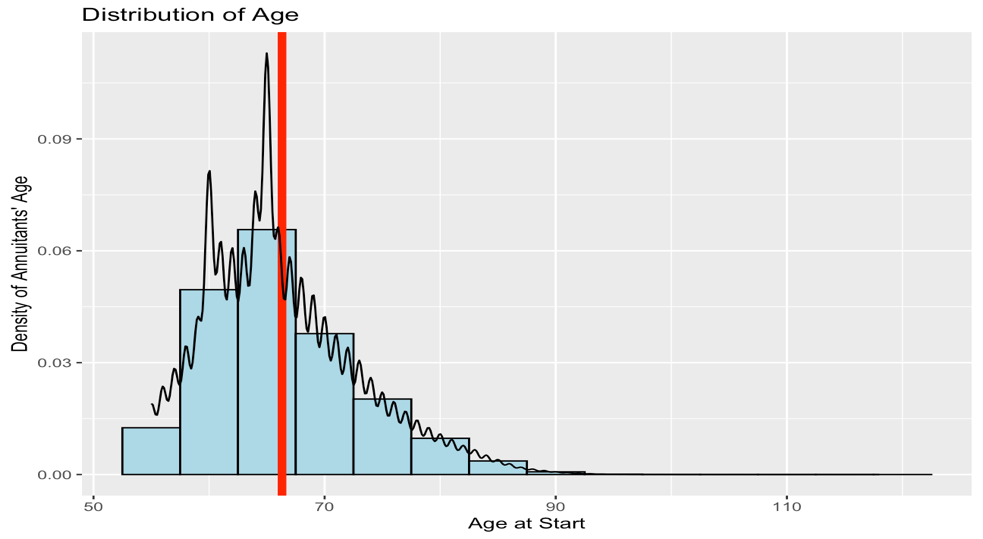
## A pie chart with numbers and a number of people Description automatically generatedA pie chart with numbers and a number of deaths Description automatically generated**Profile of Annuitants**

***Figure 1.1***

***Figure 1.2***

Figures 1.1 and 1.2 reflects the demographics of the population in the dataset, whereby the pie graph in figure 1.1 illustrates that most of the annuitants or beneficiaries are females, as they account for 57% of the population, hence males account for 43% of the pensioners. Despite this, the pie chart in figure 1.2 reveals that males comprise of approximately two-thirds of the total number deaths. In fact, males exhibit a mortality rate of 4.6% of the total population, whereas females have a lower mortality rate of 3.44%. This trend corresponds with the knowledge that on average, males tend to have higher mortality rates compared to females, which may be attributed to several biological and/or behavioural factors.

A graph with a red and green rectangle

Description automatically generated

***Figure 1.4***

***Figure 1.3***

Figure 1.3 illustrates the distribution and density of the populations’ age at the start of the registration. The histogram is positively skewed, which suggests that the majority of the pensioners are relatively early in their retirement age. This is further reinforced by the mean and mode age at the start of registration, which is 66.31 and 65 respectively. This statistic is in line with Chile’s official retirement age; 65 for males, and 60 for females. Therefore, it may be reasoned that the histogram is positively skewed since many retirees will enrol in the pension system as soon as possible to maximise their payoffs.

The boxplot in figure 1.4 portrays the distribution of the populations’ age, dependent on whether the subject has died or not. An observable trend from the boxplot is that older-aged pensioners experience a higher mortality rate compared to their younger counterparts, which is expected as their survival probability will decrease with age. Additionally, for pensioners that do not die, their mean age is 65.83, whereas for the pensioners that pass away, their mean age is 71.87. Hence, the pattern suggested by the boxplot indicates that age is a significant factor is determining mortality rates, however, further survival analysis must be conducted to reach a valid conclusion.

***Figure 1.5***

A graph of health status and status

Description automatically generatedThe bar chart in figure 1.5 indicates that healthy annuitants and beneficiaries form a majority of the population (91.20%), whereby the disabled pensioners make up 8.80% of the population. Additionally, deaths are evenly distributed amongst the heathy males and females, which is quite dissimilar in the case of the disabled. For the disabled pensioners, male mortality significantly exceeds that of females, as the mortality rate for males is 10.75%, whereas it is 0.06% for females. It is difficult to decipher the exact factors that may cause this disparity, however, as previously suggested in the report, generally, we observe a higher rate of mortality for males compared to that of females. Moreover, access to healthcare and lifestyle choices may potentially be other factors that influence this trend. Furthermore, it is worth noting that the number of disabled males in the pension systems is 73989, while the number of disabled females is 39676. The high degree of variability between those two values may account for the differences in the number of deaths amongst the two genders to some degree.

# **Survival Analysis**

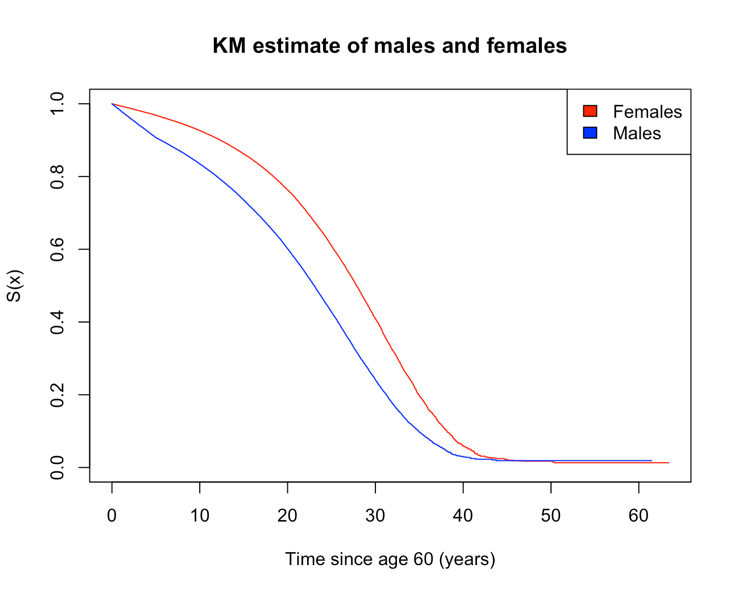
## **Overview and Key Assumptions**

Computing the Kaplan-Meier estimate and regressing the cox proportional hazard model are the methods undertaken to perform survival analysis on the given dataset. The Kaplan-Meier estimate was computed in R by accounting for left truncation and right censoring. The results of the analysis will be discussed under the assumptions of non-informative censoring, independent lives, proportional hazards, and no multicollinearity. Regarding the Kaplan-Meier survival analysis, three estimates have been produced that will assist in identifying any differences amongst the survival probabilities of the individuals differ by gender, health status, or type of annuitant. Furthermore, three Cox proportional hazard models have been constructed to test if health status or person type are significant predictors of mortality amongst males and females respectively, and whether gender has a significant association with the outcome of death.

## **Kaplan-Meier Estimate**

As observed by figure 1.6, the survival function for males decreases at a faster rate than that of females, which supports the preliminary analysis conducted on gender-wise mortality. The difference amongst the survival curves leads to the insight that the usage of different life tables based on genders is valid. However, to determine whether the association between gender and mortality is significant, a cox proportional hazard model will need to be fitted.

A graph of a person with a disability

Description automatically generated with medium confidence

***Figure 1.7***

***Figure 1.6***

A graph of an average annuitant

Description automatically generated

***Figure 1.8***

Figure 1.7 presents the disparity between the survival functions for healthy and disabled males. The Kaplan-Meier curve for disabled males declines more rapidly compared to that of males, which can be attributed to their underlying health issues that may significantly increase their risk of mortality. In contrast with the previous pairs of survival functions, the Kaplan-Meier curves for females that are either the main annuitant or beneficiary in figure 1.8 do not differ by much, as their survival functions exhibit a similar rate of decrease. Through visual inspection, it is justifiable to proceed with the argument that using different life tables for males based on health status is valid, however, such is not the case for females that are split into the two annuitant types.

## **Cox Proportional Hazard Model**

**A number of numbers on a white background

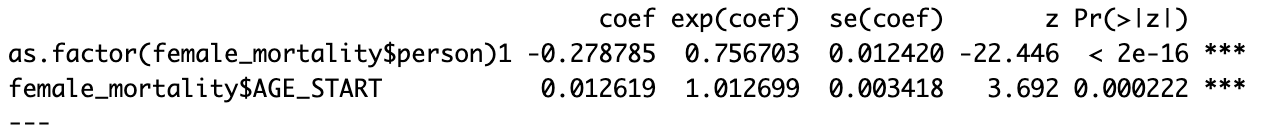
Description automatically generated**

***Figure 1.9***

The result of the full Cox proportional hazard model’s regression is shown in figure 1.9. Where ‘gender’, ‘health’, and ‘person’ are encoded as binary variables, with the dummy variables being the case of ‘male’, ‘unhealthy’, and ‘beneficiary’ respectively. The p-value constructed from a 95% confidence level for ‘gender’ is less than 0.05, indicating that the association between gender and death is significant. This reinforces our results from the descriptive analysis and Kaplan-Meier estimate. The coefficient of the ‘gender’ variable implies that holding all other factors constant, a female has a 57.12% less chance of fatality compared to a male.

## **A number and numbers on a white background Description automatically generated**

From the results of the Cox regression model for males, we observe that the p-values for ‘health’ and ‘person’ are less than 0.05, indicating their significant association with male mortality. The equation for the Cox regression model in figure 2.1 is:

The coefficient of ‘health’ indicates that healthy males have a 37.02% lower probability of dying compared to disabled males, ceteris paribus. This reaffirms the findings from the Kaplan-Meier estimate females that are healthy or disabled.

The results from the Cox proportional hazard model for females suggests that the type of annuitant is a distinct predictor of mortality, as the p-value of the test statistic is less than 0.05. This does contrast with the Kaplan-Meier estimate, as the survival curves for female main annuitants and beneficiaries are almost identical. The equation for the model is as follows:

Here, the coefficient of ‘person’ implies that a female main annuitant 75.67% lesser chance of dying, compared to a female beneficiary, ceteris paribus.

## **Conclusion**

The results from the Kaplan-Meier estimate and Cox proportional hazard model suggests that the survival function for individuals will vary based on their gender, health status, and annuitant type. The disparity between the survival curves in figure 1.7 strongly suggest that separating males based on their health status is reasonable and valid, hence the usage of life tables “CB-H-2020” and “MI-H-2020” is justified as disabled males are at a much higher risk of mortality compared to healthy males. This is further reinforced by the p-value of ‘health’ in the male Cox regression model, as it confirms that health status does have a significant association with male mortality. Furthermore, although the Kaplan-Meier survival curves do not justify using a different life table for females based on their annuitant status, the Cox regression’s results suggest that person type is a significant variable to model female mortality. Hence, the use of different life tables for the type of female annuitant is not supported by the Kaplan-Meier estimate, however, it is justified by the Cox regression model. Finally, the separation of male and female life tables is supported by both the Kaplan-Meier estimate, and Cox regression model.

# **Graduation of Unisex Life Table**

## **Rationale**

A graph of age and age

Description automatically generated

The cubic spline graduation approach is chosen as the most appropriate method given that it has passed five out of the six statistical tests for adherence. On the other hand, the Gompertz and Makeham mortality laws have passed four out of six tests, whereas the smoothing spline has also passed five out of six tests. The reason behind choosing the cubic spline over the smoothing spline is because the cubic spline performs better than the smoothing spline in the standardised deviations test, as suggested by their respective Q-Q plots. Additionally, the cubic spline is an extremely flexible method to employ as we may stich polynomials at the chosen knots to accurately model to mortality rate. The knots chosen in this case closely reflects the knots chosen for graduating the male mortality rates from the “ALT2010-12” dataset, however, certain changes were made such as only vales from 60 to 100 were included, and a couple of knots were placed in the early to mid 60s and late 90s to account for the unisex Chilean mortality rate. The cubic spline passes the Chi-square test, which suggests that the graduation approach is a good fit for the actual mortality rates, as there will be minimal discrepancies between the graduated rates and observed rates. Additionally, passing the standardised deviation test implies that the cubic spline’s distribution of its standardised deviations aligns with the standard normal distribution, which does not indicate under graduation or over graduation. Moreover, passing the signs test reinforces the fact that the positive and negative deviations of the cubic splines are binomially distributed and are balanced. Furthermore, passing the cumulative deviations test reflects the fact that the overall number of deaths conforms to the model’s graduated mortality rates. Finally, the results of the serial correlation test imply that in the cubic spline graduation approach, the deviations are independent of each other at consecutive ages. Additionally, since the curve fluctuates from the positive to negative auto correlations, and does not pass through the crude rates often, we may remove the possibility of over graduation and under graduation, meaning that the cubic spline is a ‘good’ graduation method. For the results of the statistical tests, please refer to technical appendix.

***Figure 2.1***

## **Comparison with Chilean Life Tables**

Figure 2.2 depicts the graduated death probabilities alongside with the death probabilities of healthy individuals from the Chilean life tables. As the graduated rates are for the healthy unisex population, it will slightly differ from the gender and annuitant type specific mortality rates, however, the nature of the graduated rates is very similar to the actual rates, where the graphs exhibit the characteristics of a convex function. A difference between the graduated and actual rates is that at age 60, the probability of death in the cubic spline model is higher than the actual rates, in fact the graduated rates temporarily decrease in the subsequent years, which may have to do with the high degree of volatility at age 60 as numerous individuals register for the pension system at that age. Additionally, the graduated rates rapidly increase at approximately age 95, which again may be a consequence of high volatility in the mortality data set as the number of observations are quite low.

A graph of different colored lines

Description automatically generatedA graph of a number of people

Description automatically generated with medium confidence

***Figure 2.3***

***Figure 2.2***

# **Ethical Implications of Unisex Annuity Pricing**

From an ethical perspective, using gender as a rating criterion for setting annuity prices is an intricate situation, as insurers justifiably will attempt to maximise their payoffs by pricing premiums that accurately reflect the mortality rates for the insured. The Kaplan-Meier and Cox regression survival analysis shows that the gender of an individual does have a significant association with the outcome of death, thus from a risk management perspective, it is warranted that insurance companies charge a higher premium to males as they are higher risk individuals. Nevertheless, this approach does raise an ethical dilemma as charging a different premium based on gender can be a discriminatory practice and perpetuate gender inequality, especially if the actuarial data does not show reasonable proof of gender being a determining factor of mortality.

By considering the gender of the insured, insurance companies can effectively accommodate premiums that reflect the expected lifespan of the policy holder, which assists the insurers to manage their financial capital. However, the inherent nature of this practice can be considered as discriminatory, as it tends to fixate on a certain characteristic of individuals which is beyond their control and only forms a minor segment of their individual risk profile. Furthermore, insurance companies may potentially overlook other lifestyle factors that may affect mortality, which include but are not limited to occupation, education, income, health status, and family medical history.

Most of the stakeholders in this case are the insurance companies and female policy holders, and using gender as a factor for pricing annuities will lead to financial benefits for those stakeholders. Hence, from a utilitarianism point of view, this action is right as it is advantageous for the majority. However, from a deontological perspective, every individual has the fundamental right to be treated equally, and consequences of an action does not justify the action itself. Hence, this practice can be considered unethical as it reinforces gender inequality.

Overall, the decision to use gender as a rating factor to price life annuities will need to consider the accuracy and reliability of the actuarial estimates, as well as the societal repercussions of enacting such a practice. In the case of the Chilean mortality dataset, and considering a utilitarianism point of view, it is recommended to use gender as a rating factor for setting life annuity prices. However, it is important to note that this recommendation is subject to change based on the mortality datasets and ethical framework employed.

# **Technical Appendix**

# **Survival Analysis**

## **Kaplan-Meier Curve**

## A graph with a line Description automatically generated

## **Interpretation of Coefficients**

The result of the Cox proportional hazard model’s regression is shown in figure 1.9. The p-values constructed from a 95% confidence level for each variable are less than 0.05, indicating that the association between the regressed variables and the outcome variable are significant. Hence, we may confirm our preliminary findings in the descriptive and survival analysis section and conclude that each variable is a distinct factor in determining and/or predicting the outcome of death for the pensioners. Building upon our findings, we may deduce that the usage of life tables that are separated by gender, health status, and annuitant type is justified.

**A number of numbers on a white background

Description automatically generated**

The full cox proportional hazard model’s equation is as follows:

As observed from figure 1.9, the coefficient for the binary variables is negative, meaning that the hazard rate will decrease in the case of them being equal to one (female, healthy, and main annuitant). The continuous variable ‘age’ has a positive coefficient, meaning that the hazard rate increases with age, supporting our findings in the previous section(s) of the report.

With the equation, the relative risk for certain groups may be computed, as in this case to investigate the use of five different life tables in Chile, we will calculate and compare the relative risk of an individual with risk factors ‘main annuitant’ and ‘beneficiary, also ‘healthy’ and ‘disabled’, and ‘female’ and ‘male’:

# **Graduation**

## **Methodology**

Using age last birthday as the standard definition for the calculations, the graduation of the unisex life table was performed by computing the central exposed to risk. Hence, the mortality rate is calculated by using the formula , for ages 60 to 100. Thus, the survival and death probabilities are extrapolated using the identity:

## **Chi-square Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **P-value from Chi-square Test** | | | |
| Gompertz | Makeham | Smoothing Spline | Cubic Spline |
| 0 | 0 | 0.06 | 0.65 |

The Chi-square test statistics for the fitted models were computed in R with their respective degrees of freedom, and in this case, the p-value for both the Gompertz and Makeham models were less than 0.05, which leads to the rejection of the null hypothesis that the actual data is consistent with the results predicted by the graduated rates. Hence, the Gompertz and Makeham graduation are an unsatisfactory fit to the actual data. Conversely, the p-values for Smoothing Spline and Cubic Spline are greater than 0, thus we fail to reject the null hypothesis, implying that the graduated rates are not different than the actual rates.

## **Standardized Deviations Test**

A graph of a normal q-q plot

Description automatically generatedA graph of a normal q-q plot

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **P-value from Standardized Deviations Test** | | | |
| Gompertz | Makeham | Smoothing Spline | Cubic Spline |
| 0.02 | 0.01 | 0.18 | 0.41 |

The standardized deviations test was conducted to determine whether the distribution of the residuals was consistent with the normal standard distribution, with the test statistic being equal to . The Gompertz, and Makeham model have failed the standardized deviation test, as the distribution of their standard deviations are not consistent with a standard normal distribution. The smoothing and cubic spline do indeed have residuals that follow a standard normal distribution, and to further investigate the conjecture that the values followed a standard normal distribution, a normal Q-Q plot was constructed for both graduation approaches, and as seen from both plots, the Q-Q plot is a better fit for the cubic spline as it deviates less on both ends of the line.

## **Serial Correlations Test**

**A graph with lines and numbers

Description automatically generated**

A graph with lines and numbers

Description automatically generated

A graph of a line

Description automatically generated**A graph with lines and numbers

Description automatically generated**

To test for over- and under-graduation, the auto correlation functions of the test statistics were plotted in R, and as observed through the plots of the Gompertz, Makeham and splines graduation, the lags are not significantly serially correlated. Additionally, since the lags do not cross the 95% confidence interval, we fail to reject the null hypothesis that the individualised standard deviations at consecutive ages are independent.

## **Grouping of Signs Test**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Results from Grouping of Signs Test** | | | |
| Gompertz | Makeham | Smoothing Spline | Cubic Spline |
| Runs | 6 | 4 | 14 | 10 |
| P-value | 0.004 | 8.020e-05 | 0.994 | 0.436 |
| Critical value | 8.100 | 7.810 | 7.943 | 8.125 |

The grouping of signs test is conducted to determine whether the positive and negative deviations are in random order, and if over-graduation is present in the mortality law. Groups of deviations of the same sign are tested as we can observe 6, 4, 14, and 10 runs for the Gompertz, Makeham, Smoothing, and Cubic spline approaches respectively. Since the number of groups for the Gompertz and Makeham graduation are less than their respective critical values, we can reject the null hypothesis that the positive deviations and negative deviations are in random order. On the contrary, we fail to reject the null hypothesis on the smoothing and cubic splines as their runs exceed their critical value.

## **Signs Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **P-value from Signs Test** | | | |
| Gompertz | Makeham | Smoothing Spline | Cubic Spline |
| 1.00 | 0.35 | 0.53 | 1.00 |

If the graduated rates have a binomial distribution, the balance of their positive and negative deviations can be tested using the sign tests. Using a binomial probability density function, the calculates values of the test statistic and probability return p-values that are greater than 0.05 for all the graduation approaches. Hence, the null hypothesis cannot be rejected, which implies that the graduation approaches has passed this test, since the number of their positive and negative deviations are equal.

## **Cumulative Deviations Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **P-value from Cumulative Deviations Test** | | | |
| Gompertz | Makeham | Smoothing Spline | Cubic Spline |
| 0.055 | 0.099 | 0.968 | 0.919 |

The cumulative deviations test is used to determine whether the total number of deaths are in alignment with the graduated mortality rates. Using a significance level of 5%, we cannot reject the null hypothesis that the graduation methods fail the test. However, based on the outcomes of the prior and previous tests, it is evident that the smoothing, and cubic splines outperform the Gompertz and Makeham laws. This is quite intriguing as theoretically, the Gompertz and Makeham mortality laws are accurate for large datasets.

## Graphs of Other Graduation Methods

A graph with red dots

Description automatically generated

A graph with green and black dots

Description automatically generatedA graph with dots and lines

Description automatically generated