Evaluating the Effectiveness of Statistical Models in Predicting Mortality RatesAn Analysis of Negative Binomial Distribution Model in Canada*

An Analysis of Negative Binomial Distribution Model

Xiyou Wang, Yetao Guo

2024-03-13

The study examines the correlation between gender imbalance, specifically the excess of males in a population, and the consequent competition for mates, as well as the attitudes towards hate crimes against refugees. Utilizing data from various sources, it aims to illuminate the complex interplay between mate competition and the endorsement of hate crimes, addressing gaps in current research. The findings suggest that higher male ratios may contribute to increased mate competition and potentially to greater support for hate-motivated acts against refugees. This research underscores the need for policies that promote gender balance and societal tolerance, contributing to the prevention of hate crimes and enhancing social cohesion.

Contents

1 Introduction **3**

2 Measurement/Methodology **4**

3 Data **5**

3.1 raw data 5

3.2 data cleaning 5

4 Result **7**

*Code and data are available at: https://github.com/wxywxy666/Mortality-in-Canada.

5 Discussion **8**
5.1 Finding . . . 8
5.2 Economic Impact Insigths . . . 8
5.3 Societal and Technological Influences . . . 8
5.4 Weakness and Future Research Directions . . . 9

6 Conclusion **10**

1 Introduction

Understanding the field of public health mortality rate is important for policymakers, healthcare professionals, and society in general. The barometers of the mortality rate serve the efficiency of the well-being of society and the health system. This provides more knowledge on the challenges of general health, thus guiding resource allocation. A key case study emerges for Canada with the reflection of the morality rate landscape through the more overall societal trends and the dynamics of healthcare. In this research study, the morality complicated rate is solved within the population of Canada through the age-specific dissection of mortality rate and the lead on the causes of death illustration in the various cohorts demographically; thus, our main aim is to give more insights on the factors underlying shaping Canada's mortality trends.

Due to COVID-19, The number of casualties in this research is the key critical revolution. At the same time, it acts as the subsequent analysis and the strategies for intervention. While the research findings are highlighted in the abstract, the introduction explores the comprehensive understanding of the research, providing an overview that is extensive while on an exhaustive component, ensuring there is refrainment. This research study sets a foundation for strategic address and decision-making based on evidence. Additionally, we seek a better understanding of the efforts guiding an intervention that is subsequent through the research study. To ensure the research findings efforts for the future have a structured analysis are presented in the section following an analysis that is more structured exploring the importance of Canada's results on the employed methodologies. Through the research structure description, the main aim is to offer the readers an informed road map for the multifaceted landscape navigation concerning the Canadian mortality rate.

2 Measurement/Methodology

The dataset comes from Statistics Canada and focuses on mortality trends in Canada from 2000 to 2022, categorized by age, sex, and cause of death. This information is crucial for understanding public health trends and guiding policy decisions. Therefore, this dataset was selected due to its comprehensive coverage, high data quality and reliability. Data is updated annually, and the specific data used in this article is the latest available as of January 27, 2023. Raw data set presents data on the total number of deaths in Canada and ranks the leading causes of death, such as salmonella infections, shigellosis and amoebiasis, and tuberculosis, etc.

Also included are maps of the age at time of death, the distribution of both genders, and partial places of residence. It is worth noting that The category “Age at time of death, all ages” includes the number of deaths for children aged under one year old,The deaths for which age is not stated are included in the “Age at the time of death, all ages” category but not distributed among age groups. All of the data is possessed and cleaned through R studio,a programming language for statistical computing and graphics.

3 Data

This section seeks to provide a comprehensive understanding of the dataset used in our analysis. The dataset captures the mortality rates across various demographics in Canada from 2000 to 2022. The data offer a wider perspective allowing trends to be analyzed over time, including periods of COVID-19 outbreaks.

3.1 raw data

First, the original dataset file of “Leading causes of death”, named “1310039401-eng”, was downloaded from Statistics Canada(https://www.statcan.gc.ca). Five key variables are included. “Leading causes of death” tabulated is the underlying cause of death. This is defined as the disease or injury which initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury. is selected from the conditions listed on the medical certificate of cause of death. “Characteristics” contains the “Rank of leading causes of death” which is based on the “Number of deaths”. “Age at time of death” is attained at the last birthday preceding death. “Reference period” contains specific numbers, from 2000 to 2022. This article mainly studies the statistical model of death toll, so only the data of all ages and both sexes has been downloaded.

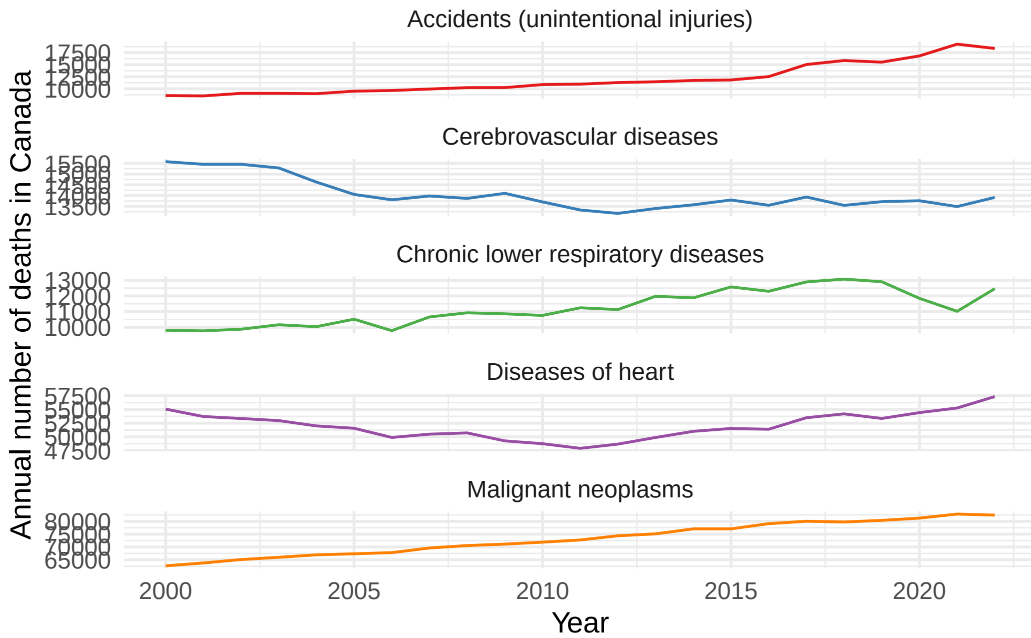
3.2 data cleaning

Year	Cause	Deaths	Ranking	Years
2022	Malignant neoplasms	82,412	1	23
2022	Diseases of heart	57,357	2	23
2022	COVID-19	19,716	3	3
2022	Accidents (unintentional injuries)	18,365	4	23
2022	Cerebrovascular diseases	13,915	5	23
2022	Chronic lower respiratory diseases	12,462	6	23
2022	Diabetes mellitus	7,557	7	23
2022	Influenza and pneumonia	5,985	8	23
2022	Alzheimer’s disease	5,413	9	23
2022	Chronic liver disease and cirrhosis	4,530	10	23

In order to understand the data more intuitively, data cleansing is necessary. A detailed description of each variable explains how these variables are critical for understanding public health trends and policy implications.

table1 showcases the first ten rows of Leading Causes of Death in Canada for the Year 2022.To further understand the table,“cause” refers to the event that led to the deaths recorded in the dataset. Each cause is identified by its medical name."deaths" represents the total number of people who passed away due to each listed cause. Years represent……….It should be noted that since the COVID-19 outbreak began in 2020, the data for 'years' is limited to only three years

4 Result



Accidents (unintentional injuries)
Cerebrovascular diseases
Chronic lower respiratory diseases
Diseases of heart
Malignant neoplasms
Annual number of deaths in Canada
Year
2000
2005
2010
2015
2020
17500
10000
15500
13500
13000
12000
11000
10000
57500
55000
52500
50000
47500
80000
65000

5 Discussion

Plots of time-series changes in the number of deaths in Canada due to different factors are provided in the visualisation results of this study. The excellence of two forecasting models for mortality is analysed in comparison.

5.1 Finding

The main focus of this document revolves around a comparison of the performance of two statistical models - the negative binomial distribution model and the Poisson distribution model - in predicting mortality in Canada. Through detailed data analysis, we find that the negative binomial distribution model outperforms the Poisson distribution model in terms of both the expected log-point-by-point predicted density difference (elpd_diff) and the k-fold cross-validated ELPD value (elpd_kfold). In addition, the negative binomial distribution model has a smaller standard deviation, suggesting that its predictions are more stable. This finding provides a new perspective that negative binomial distribution models may be more appropriate for describing and predicting mortality in Canada.

5.2 Economic Impact Insigths

From the perspective of economic impact, accurate mortality prediction is an important reference value for a number of industries, including insurance, pension, and healthcare. The superiority of the negative binomial distribution model means that we can predict future mortality rates more accurately, thus helping these industries to make more rational economic decisions, such as setting insurance premium rates and planning medical resources. In addition, for governments, accurate mortality prediction can also help to formulate more effective social security policies and reduce financial pressure. In terms of insight, the reason why the negative binomial distribution model can achieve better prediction results may be related to its ability to better handle the discrete and over-discrete nature of the data. This suggests that we need to pay more attention to the characteristics of the data and choose more appropriate statistical models when dealing with similar problems.

5.3 Societal and Technological Influences

In terms of social impact, accurate mortality forecasts help the public to better understand the trends in demographic and health conditions, and thus make more informed life decisions. For example, the public can adjust their health management and retirement planning based on the prediction results. In terms of technical implications, the findings in this paper provide new ideas for the application of statistical models in the field of mortality prediction. In the future, we can further explore and optimise the negative binomial distribution model, or find other more suitable models to improve the prediction accuracy and stability.

5.4 Weakness and Future Research Directions

Despite the superiority of the negative binomial distribution model in predicting mortality in Alberta, there are still some shortcomings. For example, only two models were compared in this paper and other possible models were not considered; furthermore, the parameters of the models were not analysed in detail to identify the key factors affecting the prediction results. Future research directions can be developed in the following aspects: firstly, other possible statistical models can be further explored to find a more suitable model for predicting mortality in the Canada; secondly, an in-depth analysis of the model's parameters can be carried out to reveal the specific factors affecting the prediction effect; and lastly, attempts can be made to incorporate more factors into the model in order to improve the prediction accuracy and the scope of application. In summary, the findings of this paper are of great guiding significance for the modelling of mortality in the Canada, and also provide us with an opportunity to explore in depth the application of statistical models in the field of mortality prediction.

6 Conclusion