Case Study – Age-specific Mortality Model indexed by Child Mortality or Both Child and Adult Mortality

Kaan Gumrah

20/12/2020

Abstract

This report will reproduce the findings of Samuel J. Clark (2019) and discuss the effects of the findings along the replication process.Clark 2019 is an article that develops a model which relates child mortality to full age schedules of mortality by looking at the relationships and patterns in the Human Mortality Database life tables. Such relationships are significant for predictions involving populations, and some poor countries lack statistical tools to conducts such forecasts. Clark's 'simple linear structure' proves to be a viable function of covariates that not only achieves accurate representation of the mortality figures in the HMD, but it is also slightly advantageous to the already existing models due to its ability to 'produce mortality schedules by single year of age.'

Keywords

Mortality - Model - Africa - Forecasting - Population - HMD - UN - Regression

Introduction

Many various demographic and epidemiological methods utilize complete age-specific mortality schedules as one of the most significant inputs. Policymakers and researchers associated with many fields, including but not limited to public health, sociology, economics, business and anthropology, rely on such data to plan in the long run; therefore, the quality and the accuracy of such data is fundamental for optimized forecasting. The World Population Prospects (WFF), prepared by the United Nations Population Division, is an example database for such models; it includes time, age, and sex-specific mortality, fertility, and population size information from 1950 to the present and forecasts of respective parameters until the year 2100 for every country. Hence, age-specific mortality schedules are present for the period between 1950 and 2100. Some of the countries in the developing world do not have the technological and statistical infrastructure to report on neither fertility nor mortality accurately, and a majority of these countries are located in Africa. The majority of the countries that do not have any information available on either child mortality or adult mortality are located in Africa. Out of 50 countries in the world that do not have any information on either child or adult mortality rates in, 33 of them are in Africa, which accounts for more than half of the continent's population, at 666 million people. In contrast, the entire world has at least some information regarding child mortality rates due to the extensive coverage of household surveys that investigate fertility and maternal/child health. (Li 2015) Clark's paper "A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality" develops a model that utilizes data on child mortality to create mortality schedules for other age groups, eliminating the need for data on mortality at other ages.

Methodology

The analysis aims to investigate factors that affect child mortality rates. For the purposes of this case study, these factors were age, weeks, and area. Age interval, access and type were also associated with the deaths recorded from the children.

The data was collected from github in accordance with the 'Software and Reproducibility' section of the original paper by Clark. Each of the variables are either categorical and continuous in nature, and the main variables of the study were age, gender, weeks, deaths, and area. Prior to the analysis, the collected data was saved as a CSV file.

Firsly, exploratory data analysis was conducted for the distribution of the variables. Secondly, regression analysis was conducted to test for the significance of age, area, and weeks on the number of deaths. All analysis was conducted in RStudio.

Results

```
##
## ===================================================================
## Statistic   N     Mean    St. Dev.  Min  Pctl(25) Pctl(75)  Max
## -------------------------------------------------------------------
## Area      5,382   1.000    0.000     1      1        1        1
## Year      5,382 2,017.341  1.637   2,015  2,016    2,019    2,020
## Week      5,382  25.920    15.082    1      13       39      53
## Deaths    5,382  591.953  662.741    1      154      760    3,348
## -------------------------------------------------------------------
```

```
##
## Results
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                               Deaths
## -----------------------------------------------
## Week                           0.021
##                               (0.290)
##
## as.factor(Age)45            151.992***
##                              (15.171)
##
## as.factor(Age)65            243.178***
##                              (15.171)
##
## as.factor(Age)75            445.377***
##                              (15.171)
##
## as.factor(Age)85            746.712***
##                              (15.171)
##
## as.factor(Age)TOT          1,738.153***
##                              (15.171)
##
## Constant                     37.181***
##                              (13.105)
##
## -----------------------------------------------
```

```
## Observations                      5,382
## R2                                0.765
## Adjusted R2                       0.765
## Residual Std. Error     321.294 (df = 5375)
## F Statistic         2,920.040*** (df = 6; 5375)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Discussion & Summary

On the model, the R-squared value of 0.7652 indicates that 76.52% of number of deaths was explained by the model. In addition, the model proved statistically adequate ($F(6,5375)=2920, P<0.05$).At 5% level of significance, week and ages proved to be statistically significant.

Conclusion

During the analysis, there were many deaths recorded on average as compared to age, weeks, and area. It was observed that area and age had a significant impact on the number of deaths recorded, impliying the idea that both age and weeks were desired for the prediction of the number of deaths.

Taking all these statistically significant results into consideration, the model developed by Samuel J. Clark is an adequate model for forecasting mortality. The model should be beneficial for countries that lack statistical framework to calculate and predict mortality, as the model is able to calculate mortality by taking only child mortality into consideration, which is a type of information at hand (Li 2015) unlike adult mortality. Countries can utilize this model make more accurate forecasts and direct policies and research according to their results. Fields like public health, economics, business, and sociology can benefit from these forecasts, both on an international and on a domestic scale.

Weaknesses & Next Steps

The statistical research and the model is limited to age, area, and week as the factors affecting deaths. Although the study showed the model is able to adequately explain about three quarters of the number of deaths, other factors that might cause child deaths are not captured by the model. This summons the need to include other variables to be implemented into the model to further enhance the explanatory power of the model.

Code Repository

The link below contains the rmd and csv files that were used to conduct the reproduction of the paper. Github: https://github.com/KaenG49/reproductionpaper

References

Clark, S.J. A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality. Demography 56, 1131–1159 (2019). https://doi.org/10.1007/s13524-019-00785-3

Li, N. (2015). Estimating life tables for developing countries (Technical Paper No. 2014/4). New York, NY: United Nations, Department of Economic and Social Affairs, Population Division. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/technical/TP2014-4.pdf