11th Scientific Meeting of the SIS Group
"Statistics for the Evaluation and Quality in Services"

BOOK OF SHORT PAPERS

Editors
Andrea Bucci
Alfredo Cartone
Adelia Evangelista
Andrea Marletta

IES PESCARA 2023

STATISTICAL METHODS FOR EVALUATION AND QUALITY: TECHNIQUES, TECHNOLOGIES AND TRENDS (T^3)

EDIZIONI Il Viandante

IES 2023 - Statistical Methods for Evaluation and Quality: Techniques, Technologies and Trends (T^3)

BOOK OF SHORT PAPERS

Editors: Andrea Bucci, Alfredo Cartone, Adelia Evangelista and Andrea Marletta

Book of Short papers
11th International Conference **IES 2023**
Statistical Methods for Evaluation and Quality: Techniques, Technologies and Trends (T^3)

University 'G. d'Annunzio' of Chieti-Pescara

Contents

Modelling the gender gap in youth mortality with an Age-Period-Cohort analysis

Modellazione delle differenze di genere nella mortalità giovanile con un'analisi Età-Periodo-Coorte

Giacomo Lanfiuti Baldi and Andrea Nigri

Abstract In this paper, we propose an Age-Period-Cohort (APC) model leveraging Skew-Normal distribution, aiming at modelling the gender gap in youth mortality. We noticed that gender differences in youth mortality are largest around the age of 20, but these differences are not symmetrical with respect to the peak. Following this evidence, we perform the APC analysis in which the response variable follows the Skew-Normal distribution. We adopt a sex ratio approach, using the ratio of age-specific mortality rates of men and women as the response variable.
Our research focuses on the population under age 45 in the United States between 1960 and 2020.
Abstract *In questo lavoro, proponiamo un modello Età-Periodo-Coorte (APC) che sfrutta la distribuzione Normale asimmetrica, con l'obiettivo di modellare il divario di genere nella mortalità giovanile.*
Abbiamo notato che le differenze di genere nella mortalità giovanile sono maggiori intorno ai 20 anni d'età, ma queste differenze non sono simmetriche rispetto al picco. In base a questa evidenza, eseguiamo l'analisi APC in cui la variabile di risposta segue la distribuzione Normale asimmetrica. Usiamo come variabile di risposta il rapporto tra i tassi di mortalità specifici per età di uomini e donne.
La nostra ricerca si concentra sulla popolazione di età inferiore ai 45 anni negli Stati Uniti tra il 1960 e il 2020.

Key words: Mortality modelling, Skew-Normal, Age-Period-Cohort Model

---

Giacomo Lanfiuti Baldi
Department of Statistics, Sapienza University of Rome, Rome, Italy e-mail: giacomo.lanfiutibaldi@uniroma1.it

Andrea Nigri
Department of Economics, Management and Territory, University of Foggia, Foggia, Italy e-mail: andrea.nigri@unifg.it

1 Introduction

Gender differences in mortality are increasingly discussed and studied in the social and demographic fields. Modelling and analysing gender differences in mortality can tell us a lot about the social context of a population or a country [8]. Thus, a better understanding of sex differences in mortality can guide public health efforts to reduce overall mortality rates and promote greater equity in health outcomes.

Gender differences are not constant at all ages and are driven by different causes of death. Many studies focus on gender differences in mortality at adult ages, but there is less literature on differences at younger ages. At the same time, gender differences in mortality are greater at younger ages [6] and this is mainly due to the different behaviour of the two groups. Men in particular run a much higher risk of accidental death around the age of 20 [7]. In addition to age, mortality differences change over time because the behaviour that generates them changes over the years [2].

We aim to study how the gender gap in mortality at younger ages (under 45) varies concerning individual age groups and over time. We want to define and interpret the effects of different ages and social, cultural and behavioural changes in society [11]. To do this, we work in an Age-Period-Cohort (APC) framework leveraging a model based on the Skew-Normal distribution. The Skew-Normal distribution is not widely used in the APC framework.

In particular, we are interested in studying in the United States (US), where more than the 30% of deaths at young ages are due to external causes (unintentional injuries) [3] and the issue of road accidents, and violent and risky behaviours among young people is often at the centre of public debate.

2 Data and Measure

We use a sex-ratio approach to study gender differences: we analyse the ratio of age-specific mortality rates between males ($m^M_{x,t}$) and females ($m^F_{x,t}$) over time:

$$SR_{x,t} = \frac{m^M_{x,t}}{m^F_{x,t}}. \tag{1}$$

This measure is useful for several reasons: it allows us to use a single variable to study the two sexes, it is less sensitive to the general level of mortality than the absolute difference in deaths [2], and finally, it has a well-defined and known shape [7]. We use age-specific mortality rates per sex and single year of age from the Human Mortality Database [10] of the United States in the study period.

Generally, this sex-ratio over the ages is characterized by a *peak* and a *hump*. The peak, which is the highest and most concentrated, coincides with youthful ages and is generally attributed to the highest male mortality due to riskier behaviours [7]. The hump corresponds to the adult ages and it was primarily caused by excess male mortality from cancer [2]. According to [7] we set the threshold age (between the peak and the hump) at 45 ages and we will focus only on the peak (Fig.1), in order to study the gender gap in mortality at young ages.

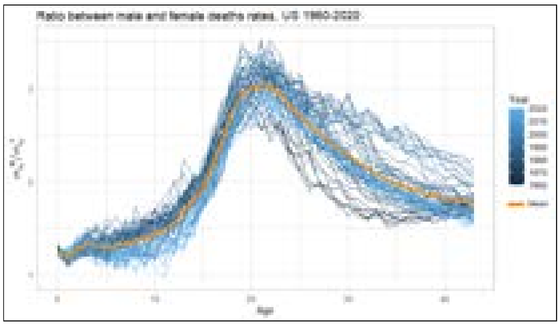

Fig. 1 Sex-Ratio of the Age-Specific Mortality Rates in US between 1960 and 2020 for young (under 45) population. Data source: HMD

Gender differences in infant mortality are very low in all periods considered, the differences start to increase in the years of adolescence. Male mortality at the peak comes to more than 3 times that of females in most years and there is no trend in the shift of the peak age over the years.
We have noticed that the sex-ratio at younger ages has a Gaussian shape, but in most of the years, it is not symmetrical around the peak.

3 Model specification

In order to study the age, period and cohort effects we suggest leveraging a model based on a *Skew-Normal* distribution. This distribution is useful for modelling both symmetric and skew data sets.

As defined by Azzalini in 1985 [1], Z has a skew-normal distribution with parameter $\lambda \in \mathbb{R}$ if:

$$f(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \tag{2}$$

Where $\phi(\cdot)$ and $\Phi(\cdot)$ are the $N(0,1)$ probability density function and cumulative distribution function, respectively.

If $Z \sim SN(\lambda)$, then the random variable $Y = \mu + \sigma^2 Z$ still has a skew-normal distribution with *location* parameter $\mu \in \mathbb{R}$, *scale* parameter $\sigma^2 \in \mathbb{R}^+$ and *skewness* parameter λ. The probability density function of $Y \sim SN(\mu, \sigma^2, \lambda)$ is given by:

$$f(y;\theta) = f(y;\mu,\sigma^2,\lambda) = \frac{2}{\sigma^2}\phi\left(\frac{y-\mu}{\sigma^2}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma^2}\right). \tag{3}$$

Following the framework proposed by Klein [5], we can generally set up the relationship between distribution parameters and the elements of the linear predictor as:

$$g(\varphi) = \sum_{j=1}^{J} f_j(\nu),$$

where f may comprise various forms, defined on basis of the covariate structure, in our case we consider a linear function $f_j(\nu) = \mathbf{X}\beta_j$, that represents the fixed effects.

Specifically, let's consider $y^{\mathrm{T}} = (y_1, y_2, \ldots, y_n)$ as the vector of the response variable and $f(y; \boldsymbol{\varphi})$, a density function with k parameters $\boldsymbol{\varphi}^{\mathrm{T}} = (\varphi_1, \varphi_2, \ldots, \varphi_n)$ modelled by linear additive models. We assume that observations y_i are independent conditional on $\boldsymbol{\varphi}$, with density function $f(y_i; \varphi_i)$, where ${\varphi_i}^{\mathrm{T}}$ is a vector of k parameters related to explanatory variables and random effects. Let $g(\cdot)$ be a known monotonic link function relating φ_k to explanatory variables through an additive model given by:

$$g(\varphi) = \eta = \mathbf{X}\beta$$

where: $\beta^{\mathrm{T}} = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{J'k})$ is a parameter vector of length J', $\mathbf{X}$ is the design matrix of order $n \times J'$. In our study, for the Skew-normal family distribution: $\varphi = \mu$ and $g(.)$ is the identity function. So, we have the following model:

$$\mu = \mathbf{X}\beta.$$

Thereby, the components of the model are: $\mathbf{y}$, is the response vector of length n; $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p)$ are the design matrices of the effects for the three dimensions in which we are interested (Age-Period-Cohort) and $\boldsymbol{\beta}^{\mathrm{T}} = \left(\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_p^{\mathrm{T}}\right)$ are the linear parameters.

Here, we provide the structure of the specific APC Skew-Normal model to be applied to the study of the gender gap while including the Age-Period-Cohort (A-P-C) structure as predictor.

Let $\mathscr{A} = \{a_0, a_1, \ldots, a_\omega\}$, $\mathscr{P} = \{p_0, p_1, \ldots, p_n\}$ and $\mathscr{C} = \{c_0, c_1, \ldots, c_m\}$ be the set of age, year and cohort categories, respectively. The APC model describes the gender ration of death rates at age $a \in \mathscr{A}$, time $p \in \mathscr{P}$, and cohort $c \in \mathscr{C}$.
Using categorical coding for Age, Period and Cohort respectively we introduce the model:

$$\mu = \beta_{(a)} + \beta_{(p)} + \beta_{(c)}. \tag{4}$$

As it is well known in the literature [9], since the cohort is obtained from a linear relationship of the other two variables (Age and Period), the model suffers from a lack of identifiability. To solve this it is necessary to impose constraints, which in this case are:

$$\sum_{\omega=1}^{\Omega} \beta_{(a_\omega)} = \sum_{n=1}^{N} \beta_{(p_n)} = \sum_{m=1}^{M} \beta_{(c_m)} = 0. \tag{5}$$

4 Results

In our analysis, we estimate the models described in the previous section (3) with a Bayesian approach. Samples from the posterior distributions of the parameters and effects were drawn by using Hamiltonian Monte Carlo sampling and specifically using the `stan` software package [4].

We specify a $\sigma^2 \sim \Gamma(0.01, 0.01)$ and a $\lambda \sim \Gamma(0.01, 0.01)$ as prior distribution.

Here we report the results (Fig.2) of the model with the constraints in which Age, Period and Cohort are treated as categorical variables, with the first category being the reference one.

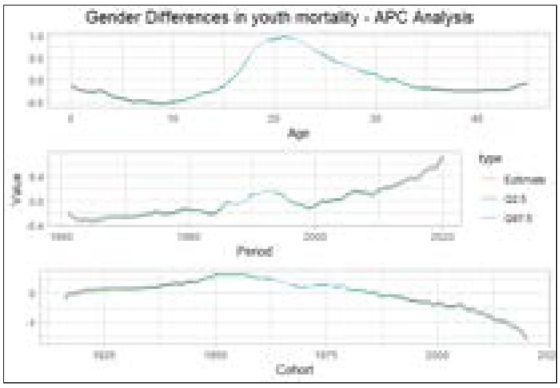

Fig. 2 APC model on the gender gap in youth mortality, US 1960-2021. Data source: HMD

The age parameters provide the structure of the gender gap in youth mortality, which can be found on average in all years of the period and for all cohorts. The value of the estimated parameters for age increases with adolescence. Researchers usually do not look at biological factors to explain the excess mortality among men at these ages, but rather at individual and social reasons. These reasons become even bigger and more important in the peak years of gender differences (21-22 years): car accidents, suicides and violence are by far the most important causes.

From the two graphs of period and cohort parameters, we can observe two different trajectories. Cohorts born after the mid-1950s gradually experienced smaller differences in mortality between the two sexes. In contrast, the period effect indicates an increase in the gender gap on average in the last decades.

5 Discussion

We observed the gender gap in youth mortality in the US between 1960 and 2020. For this aim, we used a sex-ratio approach in the Age-Period-Cohort framework, leveraging a model based on a Skew-Normal distribution.

The parameter estimates were performed adopting a Bayesian approach and using `stan` software.

The innovation of this work is to implement an Age-Period-Cohort analysis, assuming that the target variable has an asymmetric distribution: the Skew-Normal distribution is not widely used in the APC framework, which usually is based on the normal distribution in the demographic field.

Observing sex differences in mortality and how these vary across ages and over time is useful for understanding society and the behaviours that determine them. Moreover, the knowledge of the mortality dynamics and of the differences between the sexes can be an excellent tool in the hands of policymakers.

Preliminary results show that over the past 25 years in the US, we have observed that younger cohorts are benefiting from societal changes and the attention that the topic of youth mortality is receiving in the public debate.

References

1. Azzalini A.: A class of distributions which include the normal ones. Scandinavian Journal of Statistics. 171-178 (1985)
2. Bergeron-Boucher, M.-P., Canudas-Romo, V., Pascariu, M. and Lindahl-Jacobsen, R.: Modeling and forecasting sex differences in mortality: a sex-ratio approach. Genus, 74, 1-28 (2018) https://doi.org/10.1186/s41118-018-0044-8
3. Heron, M.: Deaths: Leading causes for 2019. National Vital Statistics Reports, 70(9). National Center for Health Statistics, Hyattsville, MD (2021) https://doi.org/10.15620/cdc:107021
4. Hilton, J., Dodd, E., Forster, J. J., and Smith, P. W. F.: Projecting UK mortality using Bayesian generalised additive models (2018) arXiv preprint arXiv:1802.03242. Retrieved from https://arxiv.org/abs/1802.03242
5. Klein, N., Kneib, T., Klasen, S. and Lang, S.: Bayesian structured additive distributional regression for multivariate responses. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64(4), 569–591 (2015) https://doi.org/10.1111/rssc.12090
6. Kung, H. C., Hoyert, D. L., Xu, J. and Murphy, S. L.: Deaths: final data for 2005. National Vital Statistics Reports, 56(10), 1-120 (2008)
7. Meslé, F.: Life expectancy: a female advantage under threat. Population and Societies, 402(4), 1-4 (2004)
8. Nathanson, C. A.: Sex differences in mortality. Annual Review of Sociology, 10(1), 191-213 (1984) https://doi.org/10.1146/annurev.so.10.080184.001203
9. O'Brien, R. M.: Mixed models, linear dependency, and identification in age-period-cohort models. Statistics in Medicine, 36(16), 2590-2600 (2017) https://doi.org/10.1002/sim.7305
10. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Human Mortality Database (2021) Retrieved from http://www.mortality.org
11. Yang, Y., Schulhofer-Wohl, S., Fu, W. J. and Land, K. C.: The intrinsic estimator for age-period-cohort analysis: what it is and how to use it. American Journal of Sociology, 113(6), 1697-1736 (2008) https://doi.org/10.1086/587154

This book collects the papers presented at the 11th Scientific Meeting of the SIS Group "Statistics for the Evaluation and Quality in Services" *Statistical Methods for Evaluation and Quality: Techniques, Technologies and Trends (T^3)*, which took place at the University of Chieti-Pescara, on 30th August-1st September, 2023.
The papers, which had been selected through a refereeing process, contain topics on statistical approaches and methodologies for the evaluation of public services in different contexts, and cover the areas of digital transition, e-commerce and digital marketing, enterprises, environment and territory, healthcare and wellness, finance, bank and FinTech, justice system, labour market, official statistics, public administration, food and wine, school, education and training, social, sports, sustainability, tourism, transport, university and research, well-being and welfare.

ISBN 9791280333698

ISBN: 979-12-803-3369-8
DOI: 10.60984/978-88-94593-36-5-IES2023

W UWOPVN QQQTWV \