

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



**STATISTICAL METHODS
FOR EVALUATION AND QUALITY:
TECHNIQUES, TECHNOLOGIES AND TRENDS (T³)**

**IES 2023 - Statistical Methods for Evaluation and Quality:
Techniques, Technologies and Trends (T³)**

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 (T³)

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



Contents

Plenary Session	1
Befitting Cross Validation with Three Case Studies	2
<i>Kenett R.S.</i>	
Solicited Session SS1 - <i>Statistical analysis and modeling of environmental pollution data</i>	3
Assessing environmental quality by clustering a structural equation model based index: An application to European cities air pollution	4
<i>Bottazzi Schenone M., Grimaccia E. and Vichi M.</i>	
Evaluating the nonlinear association between PM ₁₀ and emergency department visits	10
<i>Bucci A., Sanmarchi F., Santi L., Giostra F., Tubertini E., Rosa S., Nante N. and Golinelli D.</i>	
Estimating spatially varying Gaussian Graphical Models to unveil relationships among pollutants in the Red River Delta in Vietnam	16
<i>Pronello N., Cucco A., Ignaccolo R. and Ippoliti L.</i>	
Solicited Session SS2 - <i>Statistics in sports</i>	22
Clustering Athlete Performances in Track and Field Sports	23
<i>Argiento R., Colombi A., Modotti L. and Montagna S.</i>	
A Cross-Country Analysis of Engagement in Physical Activity and Sport Practice Learnt from Eurobarometer Survey Data	29
<i>Simone R.</i>	
Strong eras in male professional tennis	35
<i>Breznik K., Candila V., Milekhina A. and Restaino M.</i>	
NonParametric Combination method for data analytics in basketball matches	41
<i>Barzizza E., Biasetton N., Ceccato R., Disegna M. and Vezzosi G.</i>	
Solicited Session SS3 - <i>Statistical methods for the analysis of university student choices and academic performance</i>	47
The influence of labor market conditions on students' career disruption: first insights from Italy	48
<i>Usala C., Sulis I. and Porcu M.</i>	
Socio-economic aspects that may affect South-North students' mobility in Italy	53
<i>Genova V.G. and Boscaino G.</i>	
An analysis of student's performance in bachelor's degree	59

<i>La Rocca M., Niglio M. and Restaino M.</i>	
An exploratory strategy for analyzing students' mobility data	65
<i>Primerano I. and Giordano G.</i>	
Solicited Session SS4 - <i>Statistics in football</i>	71
Community Detection in Sport Market Networks: The Case of Italian Professional Football	72
<i>Rondinelli R. and Ievoli R.</i>	
An Original Application to Football of PLS-SEM for the xG Model	78
<i>Cefis M. and Carpita M.</i>	
Performance Assessment of Football Players and Combined Permutation Tests with application to Home-Field Advantage	84
<i>Bonnini S.</i>	
A First Proposal of the Triad Census for Weighted Networks: an Application to Football	90
<i>Rondinelli R. and Palazzo L.</i>	
Solicited Session SS5 - <i>Advances in statistical learning from high-dimensional data</i>	96
PCA approaches for vector functional time series	97
<i>Aguilera A.M., Alonso F.J. and Acal C.</i>	
Conformal Prediction for Functional Kriging Models	102
<i>Diana A., Romano E. and Adzic J.</i>	
Measuring Public-Private Connectedness in Financial Markets	108
<i>Sánchez García J. and Cruz Rambaud S.</i>	
An original approach to anomalies in intertemporal choices through Functional Data Analysis: Theory and application for the study of Hikikomori syndrome	112
<i>Martino R., Ventre V., Cruz Rambaud S. and Maturo F.</i>	
Solicited Session SS6 - <i>Labour market: trends, perspectives and new challenges</i>	118
Enriching Job Vacancy Official Information with Online Job Advertisements: Chances and Limits	119
<i>Lucarelli A. and Righi A.</i>	
Innovation in Management: towards the Open Manager	126
<i>Bruttini P., Gallo M., Mariani P. and Menini T.</i>	
Solicited Session SS7 - <i>Data analysis for web sources</i>	132
Enhancing SMEs default prediction with web-scraped data	133
<i>Crosato L., Domenech J. and Liberati C.</i>	
Web data as enabler for informed decisions in Labour Market	137
<i>Maggioni G.</i>	
The metaverse & luxury fashion brands: strategic communication exercise	141
<i>Forciniti A. and Zavarrone E.</i>	
Increasing the Geographical Granularity of Economic Indicators with Google Trends	147

Domenech J. and Marletta A.

Solicited Session SS8 - <i>Methodological and applicative contributions for evaluating sustainable development</i>	153
Evaluating sustainable development in EU countries through synthetic indicators	154
<i>Alaimo L.S. and Cucci M.</i>	
Naples and tourism sustainability: A survey of citizens' perceptions	160
<i>Aria M., Pagliara F., D'Aniello L. and Della Corte V.</i>	
Modelling inequalities for sustainable development in Italy countries	166
<i>Musella M., Borrata G., Camminatiello I. and Lombardo R.</i>	
Food Security and Sustainability: A Science Mapping Analysis	172
<i>Piscitelli A.</i>	
Solicited Session SS9 - <i>Inequalities in the labour market</i>	178
Skill similarities across Italian regions: an analysis based on the online job advertisements	179
<i>Kahlawi A., Buzzigoli L., Grassini L., Martelli C. and Giambona F.</i>	
Italian Labour Market reform and gender inequalities	185
<i>Marini C. and Nicolardi V.</i>	
Intergenerational transmission of disadvantages in the Italian labour market: evidence from AD-SILC data	191
<i>Busetta A., Fabrizi E., Ragozini G. and Sulis I.</i>	
Solicited Session SS10 - <i>Statistical methods and complexity for evaluation in finance</i>	197
Financial networks resilience and shocks propagation	198
<i>Cerqueti R., Cinelli M., Ferraro G. and Iovanella A.</i>	
How the choice of one parameter impacts the numerical stability of the efficient frontier	204
<i>Fassino C. and Uberti P.</i>	
Dynamic shrinkage for minimum variance combination of forecasts	210
<i>Mattera R.</i>	
Exploring the perception of the gender issue of Italian female entrepreneurs	217
<i>Castellano R., Riccioni J. and Rinaldi A.</i>	
Solicited Session SS11 - <i>Networks data analysis: new perspectives and applications</i>	223
Describing Italian mobility trajectories in higher education	224
<i>Genova V.G., Giordano G., Ragozini G. and Vitale M.P.</i>	
Collaboration networks: methodological issues and updated empirical evidence on Italian statisticians	230
<i>De Stefano D., Fabbrucci Barbagli A.G., Santelli F. and Zaccarin S.</i>	
Mapping Ashtma Complexity with Graph Theory: an Integrative Approach	236
<i>Cucco A., Simpson A., Murray C., Fontanella S. and Custovic A.</i>	
Investigating the patterns of Italian internal mobility: a network analysis at provincial level	242

Sarra A., D'Ingiullo D., Evangelista A., Nissi E., Quaglione D. and Di Battista T.

Solicited Session SS12 - <i>Innovations and challenges in official statistics</i>	248
Formal and informal networks of care for the elderly: regional profiles compared	249
<i>Sicuro L., Tucci D. and Coniglio R.</i>	
Gender Gap: a multidimensional approach	255
<i>Acampora C., Fusco D., Liguori M.A. and Pagliuca M.M.</i>	
Using Whatsapp in Official Statistics: a New tool for managing the Agriculture Census	259
<i>Fabi C.</i>	
Solicited Session SS13 - <i>Statistical methods and composite indicators for healthcare</i>	265
Longitudinal composite indicators to measure the quality of health services	266
<i>Crocetta C., Antonucci L., Cataldo R. and Mazza R.</i>	
Past and Future of Doctor-Patient Communication	272
<i>Tedesco N., Zavarrone E. and Forciniti A.</i>	
Network Analysis approach to customer satisfaction and service quality detection: an application to health-care services	277
<i>Crocetta C., Grassia M.G., Marino M., Mazza R., Simonacci V. and Stavolo A.</i>	
A project evaluation study on multiset Likert scale data	283
<i>Simonacci V., Marino M., Grassia M.G. and Gallo M.</i>	
Solicited Session SS14 - <i>Distance and depth-based statistical learning methods for robust data analysis</i>	289
Robust distance-based predictive models	290
<i>Boj E., Grané A. and Parron D.</i>	
Data depth for mixed-type data through multidimensional scaling. An application to biological age imputation	294
<i>Cascos I., Grané A. and Qian J.</i>	
A compared protocol to improve clustering procedures	298
<i>Grané A., Riani M. and Salini S.</i>	
Robust diagnostics for Linear Mixed Models with the Forward Search . . .	304
<i>Corbellini A., Grossi L. and Laurini F.</i>	
Solicited Session SS15 - <i>Advanced statistical methods for pattern recognition</i>	310
Unsupervised classification of NPLs recovery curves	311
<i>Carleo A. and Rocci R.</i>	
Living alone in Italian municipalities	317
<i>Vellucci P., Benassi F., Naccarato A. and Gallo G.</i>	
Supervised learning from high-dimensional data through dynamic updating of functional classification rules	323

<i>Maturo F., Fortuna F. and Di Battista T.</i>	
Assessing the effectiveness of coordination among public authorities in cohesion expenditure	329
<i>Coco G., Monturano G. and Resce G.</i>	
Solicited Session SS16 - <i>Recent advances in statistical learning and data analysis</i>	335
A Predictive Functional Principal Component Analysis of Well-Being Data	336
<i>Marcis L., Pagliarella M.C. and Salvatore R.</i>	
Detecting the partition in the extended hierarchy of a dendrogram: an application on biomedical data	343
<i>PolICASTRO V., Palazzo L. and VISTOCO D.</i>	
Concordance measure for rankings	350
<i>Bissiri P.G. and Nai Ruscone, M.</i>	
Quadratic discriminant scoring for selecting clustering solutions	355
<i>Coraggio L. and Coretto P.</i>	
Solicited Session SS17 - <i>Statistical methods for education and educational services</i>	361
Association between INVALSI scores and students' mobility in Italy: a preliminary analysis	362
<i>Bacci S., Bertaccini B., Lombardi G. and Tocchioni V.</i>	
Modelling Responses and Response Times: an application to Mathematics INVALSI data	368
<i>Bungaro L., Desiderati R. and Mignani S.</i>	
Latent potential outcomes: An analysis of the effects of programs aimed at improving students' non-cognitive skills	374
<i>Pennoni F., Bartolucci F. and Vittadini G.</i>	
Cognitive Skills and Non Cognitive Skills to Analyze School and Students Performances	380
<i>Vittadini G.</i>	
Solicited Session SS18 - <i>Statistical methods for the assessment of transport services and sustainable emissions</i>	386
Sustainability assessment of urban transport by an LCA comparison on different technologies vehicles	387
<i>Della Ragione L. and Meccariello G.</i>	
Passenger comfort prediction via time-series classification	393
<i>Vanacore A., Pellegrino M.S. and Ciardiello A.</i>	
A statistical model to analyse driving behavior: a case study	399
<i>Rodia G., Sarnacchiaro P. and Acciarino V.</i>	
Aggregating judgments in non negotiable group decisions in transport system	403
<i>Amenta P. and Lucadamo A.</i>	
Solicited Session SS19 - <i>New advanced statistical methods for data science</i>	408

A unified framework for two-dimensional clustering on preference-approvals: an analysis of Eurobarometer data	409
<i>Albano A., Sciandra M. and Plaia A.</i>	
Pandemic Data Quality Modelling: A Bayesian Approach	415
<i>Ferrari L., Manzi G., Micheletti A., Nicolussi F. and Salini S.</i>	
Explainable AI for Peer-to-Peer Credit Risk Management	421
<i>Babaei G., Pagnottoni P. and Do T. T.</i>	
Tackling misclassification in surveys about undeclared work via the EM algorithm	427
<i>Arezzo M.F., Guagnano G. and Vitale D.</i>	
Solicited Session SS20 - <i>Tourism, territory and data analysis</i>	432
Tourism, sustainability, and territorial impact: an input-output analysis . .	433
<i>Garau G. and El Meligi A.K.</i>	
The Impact of Big Data in Tourism	439
<i>Ciuffreda R., Choedon C. and Simonetti B.</i>	
Neural network-based prediction of domestic tourists' length of stay in Italy	443
<i>Antolini F. and Cesarini S.</i>	
The management of cultural heritage in contexts of undertourism: a model for assessing the economic sustainability of public-private partnerships	450
<i>Calabrò F.</i>	
Solicited Session SS21 - <i>Statistical methods and models for land monitoring with spatio-temporal data</i>	457
Geo-referenced data and complex networks for measuring road accident risk	458
<i>Cantaluppi G., Clemente C., Della Corte F. and Zappa D.</i>	
A comparison of geospatial models for car crash risk	464
<i>Cantaluppi G., Giardino G. and Zappa D.</i>	
Geostatistical modelling of livestock-related PM _{2.5} pollution and scenario analysis for policymakers - Work in progress	470
<i>Fassò A., Rodeschini J., Fusta Moro A. and Finazzi F.</i>	
Functional clustering methods for space-time big data from mobile phone networks	476
<i>Perazzini S., Metulini R. and Carpita M.</i>	
Solicited Session SS22 - <i>Methodological developments and applica- tions for the assessment of student competencies</i>	483
Modeling the main drivers of mathematical literacy of school-leaving stu- dents. Some evidence from the Invalsi tests	484
<i>Davino C., Palumbo F., Romano R. and Vistocco D.</i>	
Educational Data Mining: clustering students' performance over time . . .	490
<i>Taraborrelli G. and Farnè M.</i>	
The nexus of cultural capital with participation in early childhood education	496
<i>Ripamonti E.</i>	
High- and Low-Performing students and future career: a gender and social issue	502
<i>Falzetti P. and Ricci R.</i>	

Solicited Session SS23 - <i>Statistical methods for environmental monitoring and sustainability</i>	508
Clustering spatial data through optimal transports	509
<i>Balzanella A. and Verde R.</i>	
New interpretative insights for environmental air quality by means of FDA	514
<i>Terzi S., Naccarato A. and Fortuna F.</i>	
A Bayesian State-Space Model to Mitigate Unmeasured Confounding . . .	520
<i>Zaccardi C., Valentini P. and Ippoliti L.</i>	
Mining social media data for damage assessment in environmental disasters	526
<i>del Gobbo E., Cafarelli B., Ippoliti L. and Fontanella L.</i>	
Solicited Session SS24 - <i>Satisfaction and behavior in tourism</i>	531
The evaluation of the hotel stay through a new development of correspondence analysis coping with ordinal variables	532
<i>D'Ambra A. and Amenta P.</i>	
Assessing the role of knowledge and authenticity in the formation of attendee loyalty at cultural festivals	536
<i>Rivetti F., Lucadamo A. and Rossi C.</i>	
Residents' Opinions and Perceptions of Tourism Development in the Historic City of Matera	542
<i>Sarnacchiaro P., Di Gennaro R. and Di Taranto E.</i>	
Exploring tourism at religious sites: The case of Assisi	547
<i>Rivetti F., Dini M. and Splendiani S.</i>	
Session of free contributes SFC1 - <i>Education and labour</i>	552
Local concordance among the items of questionnaires on student's opinion (OPIS)	553
<i>Terzi S. and Petrarca F.</i>	
High School Proficiency of Future University Students: An Analysis based on INVALSI Data	560
<i>Santelli F., Di Credico G. and Di Caterina C.</i>	
Employment vulnerability of immigrants in the labour market – Does origin matter?	566
<i>Bittaye M.</i>	
The effect of pricing policies on students' use of university canteens	572
<i>Masserini L., Bini M. and Lorenzoni V.</i>	
Gig workers' identikit	576
<i>Zavarrone E. and Forciniti A.</i>	
Session of free contributes SFC2 - <i>Tourism and territory</i>	582
Multi-source approach for statistics in tourism sector	583
<i>Bianchino A., Fusco D., Giordano P., Liguori M.A. and Summa D.</i>	
Statistical analysis of tourism sustainability in Campania: post Covid-19 review	589
<i>Giacalone M., Basile V. and Bellucci M.</i>	
Investigating recent changes in dietary behavior	596
<i>D'Uggento A., d'Ovidio F.D., Toma E. and Onorati M.G.</i>	

Depopulation in the Abruzzo municipalities	603
<i>Carulli A.L., Di Spalatro D. and Valentini A.</i>	
The Productions System of Inland Areas	607
<i>Madia Carucci A.M. and Regano A.</i>	
Session of free contributes SFC3 - <i>Social issues</i>	613
Modelling the gender gap in youth mortality with an Age-Period-Cohort analysis	614
<i>Lanfiuti Baldi G. and Nigri A.</i>	
Random forest for classifying odor emission sources	620
<i>Distefano V., Palma M., De Iaco S. and Mazuruse G.</i>	
An Experimental Annotation Task Investigating Annotator Agreement Within a Misogynistic Dictionary and Corpus	626
<i>Tontodimamma A., Ignazzi E., Anzani S., Fontanella L. and Di Zio S.</i>	
Statistical analysis of COVID19 impact on Italian mortality	632
<i>Franchetti G. and Politano M.</i>	
Measuring multidimensional deprivation using objective and subjective data: an application of the Voronoi ranking method	638
<i>Ciommi M., Mariani F., Polinesi G. and Recchioni M.C.</i>	
Session of free contributes SFC4 - <i>Trends</i>	644
The role of big data analytics in circular supply chains: A bibliometric analysis	645
<i>Khan F. and Rapposelli A.</i>	
Estimation of the ranking of incentive policies for the adoption of 4.0 tech- nologies	652
<i>Bonnini S. and Borghesi M.</i>	
Risk Management and Future Scenarios. A proposal based on a mixed- method approach	658
<i>Di Zio S., Bolzan M., Marozzi M. and Scioni M.</i>	
Explainable artificial intelligence (XAI) through artificial intelligence from a human in the loop (HITL) perspective: an interview with ChatGPT	664
<i>Santarcangelo V., Lamacchia A., Vitullo S., Di Lecce M. and Giacalone M.</i>	
Relevance in official statistics: information needs, satisfaction with data quality, some results and future perspectives	670
<i>Loporcaro M.F.</i>	
Session of free contributes SFC5 - <i>Economic issues</i>	676
Evolutionary trends of start-ups in Italy: a case study	677
<i>Duttilo P., Caruso G., Iannone B. and Gattone S.A.</i>	
Permanent establishments and efficiency analysis with global enterprises .	682
<i>Frenda A. and Sepe E.</i>	
Techniques and constructs in some recent market and organizational research	685
<i>Sciascia I.A.</i>	
The value of buildings in the Italian general government balance sheet . . .	689
<i>Santoro P. and Regano A.</i>	

Linear and nonlinear factors affecting default risk in the peer-to-peer lending market	695
<i>Giordano F., Milito S. and Parrella M.L.</i>	
Session of free contributes SFC6 - <i>Methodological issues</i>	701
Macroeconomic Time Series Classification by Nonparametric Trend Estimation	702
<i>Feo G., Giordano F., Niglio M. and Parrella M.L.</i>	
A Normalization Method for Space-time Analysis of Evaluation and Quality Indicators	709
<i>Mazziotta M. and Pareto A.</i>	
Unveiling Latent Structures: exploring Multidimensional IRT Models using Dirichlet Process Mixtures	713
<i>Valentini P., Fontanella S. and Fontanella L.</i>	
On a technique to detect accounting data manipulation	718
<i>Passamonti C.</i>	
Session of free contributes SFC7 - <i>Economics and environment</i>	724
Determinants of Water Conservation Behaviour and Spatial Heterogeneity in their Coefficients	725
<i>Mammadli R. and Gigliarano C.</i>	
Modeling the economic burden of grass pollen allergoid immunotherapy . .	731
<i>Bilancia M. and Di Bona D.</i>	
Swine fever in Liguria: who does pay for economic losses? A causal analysis	735
<i>Baggetta C., Cavalletti B. and Corsi M.</i>	
Measuring Multidimensional Poverty of the Italian Regions in the era of COVID-19	741
<i>Chelli F.M.C., Ciommi M., Gigliarano C. and Polinesi G.</i>	
Drivers of inflation: relationships changing over time	747
<i>Sokolenko O., Palumbo A., Fortuna F., Naccarato A. and Marie J.</i>	
Environmental accounting and sustainable cities: an explorative bibliometric-based literature analysis	753
<i>Kaur A., Javed A., D'Andreamatteo A. and Rapposelli A.</i>	

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 ($m_{x,t}^M$) and females ($m_{x,t}^F$) over time:

$$SR_{x,t} = \frac{m_{x,t}^M}{m_{x,t}^F}. \quad (1)$$

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.

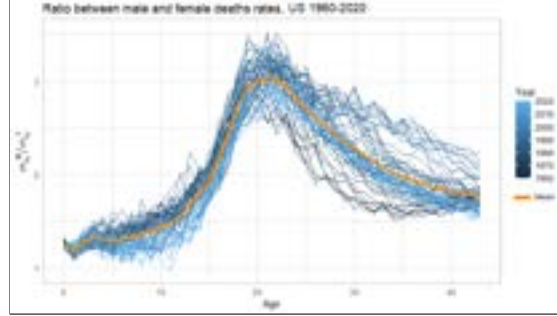


Fig. 1 Sex-Ratio of the Age-Specific Mortality Rates in US between 1960 and 2020 for young (under 45) population. Data source: HMD

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:

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

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:

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

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:

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

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

Specifically, let's consider $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$ as the vector of the response variable and $f(\mathbf{y}; \boldsymbol{\varphi})$, a density function with k parameters $\boldsymbol{\varphi}^T = (\varphi_1, \varphi_2, \dots, \varphi_n)$ 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 φ_i^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:

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

where: $\boldsymbol{\beta}^T = (\beta_{1k}, \beta_{2k}, \dots, \beta_{Jk})$ 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: $\boldsymbol{\varphi} = \boldsymbol{\mu}$ and $g(\cdot)$ is the identity function. So, we have the following model:

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

Thereby, the components of the model are: \mathbf{y} , is the response vector of length n ; $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)$ are the design matrices of the effects for the three dimensions in which we are interested (Age-Period-Cohort) and $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_p^T)$ 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 $\mathcal{A} = \{a_0, a_1, \dots, a_\omega\}$, $\mathcal{P} = \{p_0, p_1, \dots, p_n\}$ and $\mathcal{C} = \{c_0, c_1, \dots, c_m\}$ be the set of age, year and cohort categories, respectively. The APC model describes the gender ration of death rates at age $a \in \mathcal{A}$, time $p \in \mathcal{P}$, and cohort $c \in \mathcal{C}$. Using categorical coding for Age, Period and Cohort respectively we introduce the model:

$$\boldsymbol{\mu} = \boldsymbol{\beta}_{(a)} + \boldsymbol{\beta}_{(p)} + \boldsymbol{\beta}_{(c)}. \quad (4)$$

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:

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

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.

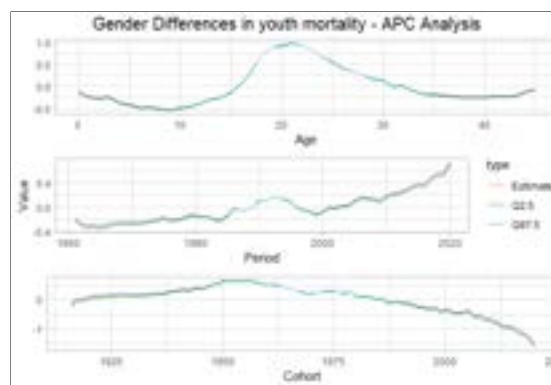


Fig. 2 APC model on the gender gap in youth mortality, US 1960-2021. Data source: HMD

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 (T⁸)*, 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: 979-12-803-3369-8

DOI: 10.60984/978-88-94593-36-5-IES2023

ISBN 9791280333698

