

# Book of Short Papers

## SIS 2021



Editors: **Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo**



**TStat**



Distribuzione Software | Formazione Professionale  
Statistica | Economia | Finanza | Biostatistica | Epidemiologia  
Sanità Pubblica | Scienze Sociali  
[www.tstat.it](http://www.tstat.it) | [www.tstattraining.eu](http://www.tstattraining.eu)

Copyright © 2021

PUBLISHED BY PEARSON

WWW.PEARSON.COM

*ISBN 9788891927361*

# Contents

<b>Preface</b>	<b>XIX</b>
<b>1 Plenary Sessions</b>	<b>1</b>
1.1 Citizen data, and citizen science: a challenge for official statistics. <i>Monica Pratesi</i>	2
<b>2 Specialized Sessions</b>	<b>8</b>
2.1 A glimpse of new data and methods for analysing a rapidly changing population	9
2.1.1 The diffusion of new family patterns in Italy: An update. <i>Arnstein Aassve, Letizia Mencarini, Elena Pirani and Daniele Vignoli</i>	10
2.1.2 Causes of death patterns and life expectancy: looking for warning signals. <i>Stefano Mazzucco, Emanuele Aliverti, Daniele Durante and Stefano Campostrini</i>	16
2.2 Advances in ecological modelling	22
2.2.1 A Bayesian joint model for exploring zero-inflated bivariate marine litter data. <i>Sara Martino, Crescenza Calculli and Porzia Maiorano</i>	23
2.3 Advances in environmental statistics	29
2.3.1 Bayesian small area models for investigating spatial heterogeneity and factors affecting the amount of solid waste in Italy. <i>Crescenza Calculli and Serena Arima</i>	30
2.3.2 A spatial regression model for for predicting abundance of lichen functional groups. <i>Pasquale Valentini, Francesca Fortuna, Tonio Di Battista and Paolo Giordani</i>	36
2.4 Advances in preference and ordinal data theoretical improvements and applications	42
2.4.1 Boosting for ranking data: an extension to item weighting. <i>Alessandro Albano, Mariangela Sciandra and Antonella Plaia</i>	43
2.4.2 An Extended Bradley-Terry Model For The Analysis Of Financial Data. <i>Alessio Baldassarre, Elise Dusseldorp and Mark De Rooij</i>	49

<b>2.5</b>	<b>Business system innovation, competitiveness, productivity and internationalization</b>	<b>55</b>
2.5.1	An analysis of the dynamics of the competitiveness for some European Countries. <i>Andrea Marletta, Mauro Mussini and Mariangela Zenga</i>	56
2.5.2	National innovation system and economic performance in EU. An analysis using composite indicators. <i>Alessandro Zeli</i>	62
<b>2.6</b>	<b>Challenges for observational studies in modern biomedicine</b>	<b>68</b>
2.6.1	Data integration: a Statistical view. <i>Pier Luigi Conti</i>	69
2.6.2	Exploring patients' profile from COVID-19 case series data: beyond standard statistical approaches. <i>Chiara Brombin, Federica Cugnata, Pietro E. Cippà, Alessandro Ceschi, Paolo Ferrari and Clelia di Serio</i>	75
2.6.3	On the statistics for some pivotal anti-COVID-19 vaccine trials. <i>Mauro Gasparini</i>	81
<b>2.7</b>	<b>Data Science for Industry 4.0 (ENBIS)</b>	<b>87</b>
2.7.1	Sample selection from a given dataset to validate machine learning models. <i>Bertrand Iooss</i>	88
2.7.2	Reliable data-drive modelling and optimisation of a batch reactor using bootstrap aggregated deep belief networks. <i>Changhao Zhu and Jie Zhang</i>	94
<b>2.8</b>	<b>Integration of survey with alternative sources of data</b>	<b>100</b>
2.8.1	A parametric empirical likelihood approach to data matching under nonignorable sampling and nonresponse. <i>Daniela Marella and Danny Pfeffermann</i>	101
2.8.2	Survey data integration for regression analysis using model calibration.	107
2.8.3	Latent Mixed Markov Models for the Production of Population Census Data on Employment. <i>Daniela Filippini, Ugo Guarnera and Roberta Varriale</i>	112
<b>2.9</b>	<b>Media, social media and demographic behaviours</b>	<b>118</b>
2.9.1	Monitoring the Numbers of European Migrants in the United Kingdom using Facebook Data. <i>Francesco Rampazzo, Jakub Bijak, Agnese Vitali, Ingmar Weber and Emilio Zagheni</i>	119
<b>2.10</b>	<b>New developments in ensemble methods for classification</b>	<b>125</b>
2.10.1	An alternative approach for nowcasting economic activity during COVID-19 times. <i>Alessandro Spelta and Paolo Pagnottoni</i>	126
2.10.2	Assessing the number of groups in consensus clustering by pivotal methods. <i>Roberta Pappadà, Francesco Pauli and Nicola Torelli</i>	132
2.10.3	Clustering of data recorded by Distributed Acoustic Sensors to identify vehicle passage and typology. <i>Antonio Balzanella and Stefania Nacchia</i>	138
<b>2.11</b>	<b>New developments in latent variable models</b>	<b>144</b>
2.11.1	A Hidden Markov Model for Variable Selection with Missing Values. <i>Fulvia Pennoni, Francesco Bartolucci, and Silvia Pandolfi</i>	145
2.11.2	Comparison between Different Likelihood Based Estimation Methods in Latent Variable Models for Categorical Data. <i>Silvia Bianconcini and Silvia Cagnone</i>	151
2.11.3	A Comparison of Estimation Methods for the Rasch Model. <i>Alexander Robitzsch</i>	157

2.12	<b>New issues on multivariate and univariate quantile regression</b>	<b>163</b>
2.12.1	Directional M-quantile regression for multivariate dependent outcomes. <i>Luca Merlo, Lea Petrella and Nikos Tzavidis</i>	164
2.13	<b>Semi-parametric and non-parametric latent class analysis</b>	<b>170</b>
2.13.1	Stepwise Estimation of Multilevel Latent Class Models. <i>Zsuzsa Bakk, Roberto di Mari, Jennifer Oser and Jouni Kuha</i>	171
2.13.2	Distance learning, stress and career-related anxiety during the Covid-19 pandemic: a students perspective analysis. <i>Alfonso Iodice D'Enza, Maria Iannario, Rosaria Romano</i>	177
2.13.3	A Tempered Expectation-Maximization Algorithm for Latent Class Model Estimation. <i>Luca Brusa, Francesco Bartolucci and Fulvia Pennoni</i>	183
2.14	<b>Statistics for finance high frequency data, large dimension and networks</b>	<b>189</b>
2.14.1	The Italian debt not-so-flash crash. <i>Maria Flora and Roberto Reno'</i>	190
<b>3</b>	<b>Solicited Sessions</b>	<b>197</b>
3.1	<b>Advances in social indicators research and latent variables modelling in social sciences</b>	<b>198</b>
3.1.1	A composite indicator to measure frailty using administrative healthcare data. <i>Margherita Silan, Rachele Brocco and Giovanna Boccuzzo</i>	199
3.1.2	Clusters of contracting authorities over time: an analysis of their behaviour based on procurement red flags. <i>Simone Del Sarto, Paolo Coppola and Matteo Troia</i>	205
3.1.3	An Application of Temporal Poset on Human Development Index Data. <i>Leonardo Salvatore Alaimo, Filomena Maggino and Emiliano Seri</i>	211
3.1.4	The SDGs System: a longitudinal analysis through PLS-PM. <i>Rosanna Cataldo, Maria Gabriella Grassia and Laura Antonucci</i>	217
3.2	<b>Changes in the life course and social inequality</b>	<b>223</b>
3.2.1	Heterogeneous Income Dynamics: Unemployment Consequences in Germany and the US. <i>Raffaele Grotti</i>	224
3.2.2	In-work poverty in Germany and in the US: The role of parity progression. <i>Emanuela Struffolino and Zachary Van Winkle Z.</i>	230
3.2.3	Parenthood, education and social stratification. An analysis of female occupational careers in Italy. <i>Gabriele Ballarino and Stefano Cantalini</i>	236
3.3	<b>Composition in the Data Science Era</b>	<b>242</b>
3.3.1	Can we Ignore the Compositional Nature of Compositional Data by using Deep Learning Approaches? <i>Matthias Templ</i>	243
3.3.2	Principal balances for three-way compositions. <i>Violetta Simonacci</i>	249
3.3.3	Robust Regression for Compositional Data and its Application in the Context of SDG. <i>Valentin Todorov and Fatemah Alqallaf</i>	255

<b>3.4</b>	<b>Evaluation of undercoverage for censuses and administrative data</b>	<b>261</b>
3.4.1	Spatially balanced indirect sampling to estimate the coverage of the agricultural census. <i>Federica Piersimoni, Francesco Pantalone and Roberto Benedetti</i>	262
3.4.2	Next Census in Israel: Strategy, Estimation and Evaluation. <i>Danny Pfeffermann</i>	268
3.4.3	Administrative data for population counts estimations in Italian Population Census. <i>Antonella Bernadini, Angela Chieppa, Nicola Cibella and Fabrizio Solari</i>	274
3.4.4	LFS non response indicators for population register overcoverage estimation. <i>Lorella Di Consiglio, Stefano Falorsi</i>	279
<b>3.5</b>	<b>Excesses and rare events in complex systems</b>	<b>285</b>
3.5.1	Space-time extreme rainfall simulation under a geostatistical approach. <i>Gianmarco Callegher, Carlo Gaetan, Noemie Le Carrer and Ilaria Prosdocimi</i>	286
<b>3.6</b>	<b>Hierarchical forecasting and forecast combination</b>	<b>292</b>
3.6.1	Density calibration with consistent scoring functions. <i>Roberto Casarin and Francesco Ravazzolo</i>	293
3.6.2	Forecasting combination of hierarchical time series: a novel method with an application to CoVid-19. <i>Livio Fenga</i>	298
<b>3.7</b>	<b>Household surveys for policy analysis</b>	<b>304</b>
3.7.1	Did the policy responses to COVID-19 protect Italian households' incomes? Evidence from survey and administrative data. <i>Maria Teresa Monteduro, Dalila De Rosa and Chiara Subrizi</i>	305
<b>3.8</b>	<b>Learning analytics methods and applications</b>	<b>311</b>
3.8.1	Open-Source Automated Test Assembly: the Challenges of Large-Sized Models. <i>Giada Spaccapanico Proietti</i>	312
3.8.2	How Much Tutoring Activities May Improve Academic Careers of At-Risk Students? An Evaluation Study. <i>Marta Cannistra, Tommaso Agasisti, Anna Maria Paganoni and Chiara Masci</i>	318
3.8.3	Composite-based Segmentation Trees to Model Learners' performance. <i>Cristina Davino and Giuseppe Lamberti</i>	324
3.8.4	Test-taking Effort in INVALSI Assessments. <i>Chiara Sacco</i>	330
<b>3.9</b>	<b>Light methods for hard problems</b>	<b>336</b>
3.9.1	Fast Divide-and-Conquer Strategies to Solve Spatial Big Data Problems. <i>Michele Peruzzi</i>	337
3.9.2	Application of hierarchical matrices in spatial statistics. <i>Anastasiia Gorshechnikova and Carlo Gaetan</i>	343
<b>3.10</b>	<b>Management and statistics in search for a common ground (AIDEA)</b>	<b>349</b>
3.10.1	Customer Segmentation: it's time to make a change. <i>Fabrizio Laurini, Beatrice Luceri and Sabrina Latusi</i>	350
3.10.2	Multivariate prediction models: Altman's ZScore and CNDCEC's sectoral indicators. <i>Alessandro Danovi, Alberto Falini and Massimo Postiglione</i>	356
3.10.3	Comparing Entrepreneurship and Perceived Quality of Life in the European Smart Cities: a "Posetic" Approach. <i>Lara Penco, Enrico Ivaldi and Andrea Ciacci</i>	362

3.10.4	The Relationship between Business Economics and Statistics: Taking Stock and Ways Forward. <i>Amedeo Pugliese</i>	368
<b>3.11</b>	<b>Mathematical methods and tools for finance and insurance (AMASES)</b>	<b>373</b>
3.11.1	On the valuation of the initiation option in a GLWB variable annuity. <i>Anna Rita Bacinello and Pietro Millosovich</i>	374
3.11.2	Modern design of life annuities in view of longevity and pandemics. <i>Annamaria Olivieri</i>	380
3.11.3	Risk Management from Finance to Production Planning: An Assembly-to-Order Case Study. <i>Paolo Brandimarte, Edoardo Fadda and Alberto Gennaro</i>	386
3.11.4	Some probability distortion functions in behavioral portfolio selection. <i>Diana Barro, Marco Corazza and Martina Nardonthors</i>	392
<b>3.12</b>	<b>Multiple system estimation</b>	<b>398</b>
3.12.1	Multiple Systems Estimation in the Presence of Censored Cells. <i>Ruth King, Oscar Rodriguez de Rivera Ortega and Rachel McCrea</i>	399
3.12.2	Bayesian population size estimation by repeated identifications of units. A semi-parametric mixture model approach. <i>Tiziana Tuoto, Davide Di Cecco and Andrea Tancredi</i>	405
<b>3.13</b>	<b>Network sampling and estimation</b>	<b>411</b>
3.13.1	Targeted random walk sampling. <i>Li-Chun Zhang</i>	412
3.13.2	Estimation of poverty measures in Respondent-driven sampling. <i>María del Mar Rueda, Ismael Sánchez-Borrego and Héctor Mullo</i>	418
3.13.3	Sampling Networked Data for Semi-Supervised Learning Algorithms. <i>Simone Di Zio, Lara Fontanella, Francesco Pantalone and Federica Piersimoni</i>	423
3.13.4	A sequential adaptive sampling scheme for rare populations with a network structure. <i>Emilia Rocco</i>	429
<b>3.14</b>	<b>New perspectives on multidimensional child poverty</b>	<b>435</b>
3.14.1	Estimating uncertainty for child poverty indicators: The Case of Mediterranean Countries. <i>Ilaria Benedetti, Federico Crescenzi and Riccardo De Santis</i>	436
3.14.2	Child poverty and government social spending in the European Union during the economic crisis. <i>Angeles Sánchez and María Navarro</i>	442
3.14.3	The Children's Worlds Study: New perspectives on children's deprivation research. <i>Caterina Giusti and Antoanneta Potsi</i>	448
3.14.4	The impact of different definition of "households with children" on deprivation measures: the case of Italy. <i>Laura Neri and Francesca Gagliardi</i>	454
<b>3.15</b>	<b>Perspectives in social network analysis applications</b>	<b>460</b>
3.15.1	A comparison of student mobility flows in Erasmus and Erasmus+ among countries. <i>Kristijan Breznik, Giancarlo Ragozini and Marialuisa Restaino</i>	461
3.15.2	Network-based approach for the analysis of LexisNexis news database. <i>Carla Galluccio and Alessandra Petrucci</i>	467
3.15.3	A multiplex network approach to study Italian Students' Mobility. <i>Ilaria Primerano, Francesco Santelli and Cristian Usala</i>	473
3.15.4	Ego-centered Support Networks:a Cross-national European Comparison. <i>Emanuela Furfaro, Elvira Pelle, Giulia Rivellini and Susanna Zaccarin</i>	479

<b>3.16</b>	<b>Statistical analysis of energy data</b>	<b>485</b>
3.16.1	Machine learning models for electricity price forecasting. <i>Silvia Golia, Luigi Grossi, Matteo Pelagatti</i>	486
3.16.2	The impact of hydroelectric storage in the Italian power market. <i>Filippo Beltrami</i>	492
3.16.3	Jumps and cojumps in electricity price forecasting. <i>Peru Muniain, Aitor Ciarreta and Ainhoa Zarraga</i>	498
<b>3.17</b>	<b>Statistical methods and models for the analysis of sports data</b>	<b>507</b>
3.17.1	Football analytics: a Higher-Order PLS-SEM approach to evaluate players' performance. <i>Mattia Cefis and Maurizio Carpita</i>	508
3.17.2	Bayesian regularized regression of football tracking data through structured factor models. <i>Lorenzo Schiavon and Antonio Canale</i>	514
3.17.3	A dynamic matrix-variate model for clustering time series with multiple sources of variation. <i>Mattia Stival</i>	520
3.17.4	Evaluating football players' performances using on-the-ball data. <i>David Dandolo</i>	526
<b>3.18</b>	<b>The social and demographic consequences of international migration in Western societies</b>	<b>532</b>
3.18.1	Employment and job satisfaction of immigrants: the case of Campania (Italy). <i>Alessio Buonomo, Stefania Capecchi, Francesca Di Iorio and Salvatore Strozza</i>	533
3.18.2	Social stratification of migrants in Italy: class reproduction and social mobility from origin to destination. <i>Giorgio Piccitto, Maurizio Avola and Nazareno Panichella</i>	539
<b>3.19</b>	<b>Well-being, healthcare, integration measurements and indicators (SIEDS)</b>	<b>545</b>
3.19.1	A Composite Index of Economic Well-being for the European Union Countries. <i>Andrea Cutillo, Matteo Mazziotta and Adriano Pareto</i>	546
3.19.2	Poverty orderings and TIP curves: an application to the Italian regions. <i>Francesco M. Chelli, Mariateresa Ciommi and Chiara Gagliarano</i>	552
<b>4</b>	<b>Contributed Sessions</b>	<b>558</b>
<b>4.1</b>	<b>Advances in clinical trials</b>	<b>559</b>
4.1.1	Quantitative depth-based [18F]FMCH-avid lesion profiling in prostate cancer treatment. <i>Lara Cavinato, Alessandra Ragni, Francesca Ieva, Martina Sollini, Francesco Bartoli and Paola A. Erba</i>	560
4.1.2	Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patients. <i>Marta Spreafico, Francesca Ieva and Marta Fiocco</i>	566
4.1.3	Information borrowing in phase II basket trials: a comparison of different designs. <i>Marco Novelli</i>	572
4.1.4	Q-learning Estimation Techniques for Dynamic Treatment Regime. <i>Simone Bogni, Debora Slanzi and Matteo Borrotti</i>	578
4.1.5	Sample Size Computation for Competing Risks Survival Data in GS-Design. <i>Mohammad Anamul Haque and Giuliana Cortese</i>	584



<b>4.2</b>	<b>Advances in neural networks</b>	<b>590</b>
4.2.1	Linear models vs Neural Network: predicting Italian SMEs default. <i>Lisa Crosato, Caterina Liberati and Marco Repetto</i>	591
4.2.2	Network estimation via elastic net penalty for heavy-tailed data. <i>Davide Bernardini, Sandra Paterlini and Emanuele Taufer</i>	596
4.2.3	Neural Network for statistical process control of a multiple stream process with an application to HVAC systems in passenger rail vehicles. <i>Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini</i>	602
4.2.4	Forecasting air quality by using ANNs. <i>Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Francesco Bucci</i>	608
<b>4.3</b>	<b>Advances in statistical methods</b>	<b>614</b>
4.3.1	Robustness of Fractional Factorial Designs through Circuits. <i>Roberto Fontana and Fabio Rapallo</i>	615
4.3.2	Multi-objective optimal allocations for experimental studies with binary outcome. <i>Alessandro Baldi Antognini, Rosamarie Frieri, Marco Novelli and Maroussa Zagoraiou</i>	621
4.3.3	Analysis of three-way data: an extension of the STATIS method. <i>Laura Bocci and Donatella Vicari</i>	627
4.3.4	KL-optimum designs to discriminate models with different variance function. <i>Alessandro Lanteri, Samantha Leorato and Chiara Tommasi</i>	633
4.3.5	Riemannian optimization on the space of covariance matrices. <i>Jacopo Schiavon, Mauro Bernardi and Antonio Canale</i>	639
<b>4.4</b>	<b>Advances in statistical methods and inference</b>	<b>645</b>
4.4.1	Estimation of Dirichlet Distribution Parameters with Modified Score Functions. <i>Vincenzo Gioia and Euloge Clovis Kenne Pagui</i>	646
4.4.2	Confidence distributions for predictive tail probabilities. <i>Giovanni Fonseca, Federica Giummolè and Paolo Vidoni</i>	652
4.4.3	Impact of sample size on stochastic ordering tests: a simulation study. <i>Rosa Arboretti, Riccardo Ceccato, Luca Pegoraro and Luigi Salmaso</i>	658
4.4.4	On testing the significance of a mode. <i>Federico Ferraccioli and Giovanna Menardi</i>	664
4.4.5	Hommel BH: an adaptive Benjamini-Hochberg procedure using Hommel's estimator for the number of true hypotheses. <i>Chiara G. Magnani and Aldo Solari</i>	670
<b>4.5</b>	<b>Advances in statistical models</b>	<b>676</b>
4.5.1	Specification Curve Analysis: Visualising the risk of model misspecification in COVID-19 data. <i>Venera Tomaselli, Giulio Giacomo Cantone and Vincenzo Miracula</i>	677
4.5.2	Semiparametric Variational Inference for Bayesian Quantile Regression. <i>Cristian Castiglione and Mauro Bernardi</i>	683
4.5.3	Searching for a source of difference in undirected graphical models for count data – an empirical study. <i>Federico Agostinis, Monica Chiogna, Vera Djordjilovic, Luna Pianesi and Chiara Romualdi</i>	689
4.5.4	Snipped robust inference in mixed linear models. <i>Antonio Lucadamo, Luca Greco, Pietro Amenta and Anna Crisci</i>	695

<b>4.6</b>	<b>Advances in time series</b>	<b>701</b>
4.6.1	A spatio-temporal model for events on road networks: an application to ambulance interventions in Milan. <i>Andrea Gilardi and Riccardo Borgoni and Jorge Mateu</i>	702
4.6.2	Forecasting electricity demand of individual customers via additive stacking. <i>Christian Capezza, Biagio Palumbo, Yannig Goude, Simon N. Wood and Matteo Fasiolo</i>	708
4.6.3	Hierarchical Forecast Reconciliation on Italian Covid-19 data. <i>Andrea Marcocchia, Serena Arima and Pierpaolo Brutti</i>	714
4.6.4	Link between Threshold ARMA and tdARMA models. <i>Guy M�elard and Marcella Niglio</i>	720
<b>4.7</b>	<b>Bayesian nonparametrics</b>	<b>726</b>
4.7.1	Bayesian nonparametric prediction: from species to features. <i>Lorenzo Masoero, Federico Camerlenghi, Stefano Favaro and Tamara Broderick</i>	727
4.7.2	A framework for filtering in hidden Markov models with normalized random measures. <i>Filippo Ascolani, Antonio Lijoi, Igor Pr�unster and Matteo Ruggiero</i>	733
4.7.3	On the convex combination of a Dirichlet process with a diffuse probability measure. <i>Federico Camerlenghi, Riccardo Corradin and Andrea Ongaro</i>	739
4.7.4	Detection of neural activity in calcium imaging data via Bayesian mixture models. <i>Laura D'Angelo, Antonio Canale, Zhaoxia Yu and Michele Guindani</i>	745
<b>4.8</b>	<b>Clustering for complex data</b>	<b>751</b>
4.8.1	Clustering categorical data via Hamming distance. <i>Edoardo Filippi-Mazzola, Raffaele Argiento and Lucia Paci</i>	752
4.8.2	Penalized model-based clustering for three-way data structures. <i>Andrea Cappelozzo, Alessandro Casa, and Michael Fop</i>	758
4.8.3	Does Milan have a smart mobility? A clustering analysis approach. <i>Nicola Cornali, Matteo Seminati, Paolo Maranzano and Paola M. Chiadini</i>	764
4.8.4	A Fuzzy clustering approach for textual data. <i>Irene Cozzolino, Maria Brigida Ferraro and Peter Winker</i>	770
4.8.5	Valid Double-Dipping via Permutation-Based Closed Testing. <i>Anna Vesely, Livio Finos, Jelle J. Goeman and Angela Andreella</i>	776
<b>4.9</b>	<b>Data science for complex data</b>	<b>782</b>
4.9.1	Text mining on large corpora using Taltac4: An explorative analysis of the USPTO patents database. <i>Pasquale Pavone, Arianna Martinelli and Federico Tamagni</i>	783
4.9.2	Emotion pattern detection on facial videos using functional statistics. <i>Rongjiao Ji, Alessandra Micheletti, Natasa Krklec Jerinkic and Zoranka Desnica</i>	789
4.9.3	The spread of contagion on Twitter: identification of communities analysing data from the first wave of the COVID-19 epidemic. <i>Gianni Andreozzi, Salvatore Pirri, Giuseppe Turchetti and Valentina Lorenzoni</i>	795
4.9.4	Composition-on-Function Regression Model for the Remote Analysis of Near-Earth Asteroids. <i>Mara S. Bernardi, Matteo Fontana, Alessandra Menafoglio, Alessandro Pisello, Massimiliano Porreca, Diego Perugini and Simone Vantini</i>	801
4.9.5	Determinants of football coach dismissal in Italian League Serie A. <i>Francesco Porro, Marialisa Restaino, Juan Eloy Ruiz-Castro and Mariangela Zenga</i>	805
<b>4.10</b>	<b>Data science for unstructured data</b>	<b>810</b>
4.10.1	Identification and modeling of stop activities at the destination from GPS tracking data. <i>Nicoletta D'Angelo, Giada Adelfio, Antonino Abbruzzo and Mauro Ferrante</i>	811

4.10.2	A generalization of derangement. <i>Maurizio Maravalle and Ciro Marziliano</i>	817
4.10.3	Analysis of clickstream data with mixture hidden markov models. <i>Furio Urso, Antonino Abbruzzo and Maria Francesca Cracolici</i>	823
4.10.4	Using Google Scholar to measure the credibility of preprints in the COVID-19 Open Research Dataset (CORD-19). <i>Manlio Migliorati, Maurizio Carpita, Eugenio Brentari</i>	829
4.10.5	Mobile phone use while driving: a Structural Equation Model to analyze the Behavior behind the wheel. <i>Carlo Cavicchia and Pasquale Samacchiaro</i>	835
<b>4.11</b>	<b>Demographic analysis</b>	<b>841</b>
4.11.1	Life expectancy in the districts of Taranto. <i>Stefano Cervellera, Carlo Cusatelli and Massimiliano Giacalone</i>	842
4.11.2	Family size and Human Capital in Italy: a micro-territorial analysis. <i>Gabriele Ruii, Marco Breschi and Alessio Fornasin</i>	848
4.11.3	Estimate age-specific fertility rates from summary demographic measures. An Indirect Model Levering on Deep Neural Network. <i>Andrea Nigri</i>	854
4.11.4	Patterns in the relation between causes of death and gross domestic product. <i>Andrea Nigri and Federico Crescenzi</i>	860
4.11.5	Locally sparse functional regression with an application to mortality data. <i>Mauro Bernardi, Antonio Canale, Marco Stefanucci</i>	866
<b>4.12</b>	<b>Environmental statistics</b>	<b>871</b>
4.12.1	A Distribution-Free Approach for Detecting Radioxenon Anomalous Concentrations. <i>Michele Scagliarini, Rosanna Gualdi, Giuseppe Ottaviano, Antonietta Rizzo and Franca Padoani</i>	872
4.12.2	Ecosud Car, a novel approach for the predictive control of the territory. <i>Giacomo Iula, Massimo Dimo, Saverio Gianluca Crisafulli, Marco Vito Calciano, Vito Santarcangelo and Massimiliano Giacalone</i>	878
4.12.3	Effect of ties on the empirical copula methods for weather forecasting. <i>Elisa Perrone, Fabrizio Durante and Irene Schicker</i>	884
4.12.4	Spatio-temporal regression with differential penalization for the reconstruction of partially observed signals. <i>Eleonora Amone and Laura M. Sangalli</i>	890
4.12.5	Sea Surface Temperature Effects on the Mediterranean Marine Ecosystem: a Semiparametric Model Approach. <i>Claudio Rubino, Giacomo Milisenda, Antonino Abbruzzo, Giada Adelfio, Mar Bosch-Belmar, Francesco Colloca, Manfredi Di Lorenzo and Vita Gancitano</i>	895
<b>4.13</b>	<b>Functional data analysis</b>	<b>901</b>
4.13.1	Remote Analysis of Chapas Stops in Maputo from GPS data: a Functional Data Analysis Approach. <i>Agostino Torti, Davide Ranieri and Simone Vantini</i>	902
4.13.2	A Conformal approach for functional data prediction. <i>Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini</i>	907
4.13.3	Block testing in covariance and precision matrices for functional data analysis. <i>Marie Morvan, Alessia Pini, Madison Giacomci and Valerie Monbet</i>	911
4.13.4	Analysing contributions of ages and causes of death to gender gap in life expectancy using functional data analysis. <i>Alessandro Feraldi, Virginia Zarulli, Stefano Mazzuco and Cristina Giudici</i>	917
4.13.5	Supervised classification of ECG curves via a combined use of functional data analysis and random forest to identify patients affected by heart disease. <i>Fabrizio Maturo and Rosanna Verde</i>	923

<b>4.14</b>	<b>Mixture models</b>	<b>929</b>
4.14.1	Alternative parameterizations for regression models with constrained multivariate responses. <i>Roberto Ascari, Agnese Maria Di Brisco, Sonia Migliorati and Andrea Ongaro</i>	930
4.14.2	Spatially dependent mixture models with a random number of components. <i>Matteo Gianella, Mario Beraha and Alessandra Guglielmi</i>	936
4.14.3	Finite mixtures of regression models for longitudinal data. <i>Marco Alfò and Roberto Rocci</i>	942
4.14.4	Mixtures of regressions for size estimation of heterogeneous populations. <i>Gianmarco Caruso</i>	948
4.14.5	Finite mixtures of regressions with random covariates using multivariate skewed distributions. <i>Salvatore D. Tomarchio, Michael P.B. Gallagher, Antonio Punzo and Paul D. McNicholas</i>	954
<b>4.15</b>	<b>New applications of regression models</b>	<b>960</b>
4.15.1	The Shapley-Lorenz decomposition approach to mitigate cyber risks. <i>Paolo Giudici and Emanuela Raffinetti</i>	960
4.15.2	A spatially adaptive estimator for the function-on-function linear regression model with application to the Swedish Mortality dataset. <i>Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i>	967
4.15.3	POSetR: a new computationally efficient R package for partially ordered data. <i>Alberto Arcagni, Alessandro Avellone and Marco Fattore</i>	972
4.15.4	Multi Split Conformal Prediction. <i>Aldo Solari and Vera Djordjilović</i>	978
4.15.5	Changes in the consumption of fruits and vegetables among university students during master courses: an analysis of data automatically collected from cashier transactions. <i>Valentina Lorenzoni, Giuseppe Turchetti and Lucio Masserini</i>	984
<b>4.16</b>	<b>New challenges in clustering and classification techniques</b>	<b>990</b>
4.16.1	A Dynamic Stochastic Block Model with infinite communities. <i>Roberto Casarin and Ovielt Baltodano López</i>	991
4.16.2	Cross-Subject EEG Channel Selection for the Detection of Predisposition to Alcoholism. <i>Michela Carlotta Massi and Francesca Ieva</i>	997
4.16.3	Some Issues on the Parameter Selection in the Spectral Methods for Clustering. <i>Cinzia Di Nuzzo and Salvatore Ingrassia</i>	1003
4.16.4	The link-match tale: new microdata from unit level association. <i>Riccardo D'Alberto, Meri Raggi and Daniela Cocchi</i>	1009
<b>4.17</b>	<b>New developments in Bayesian methods</b>	<b>1015</b>
4.17.1	Spatio-temporal analysis of the Covid-19 spread in Italy by Bayesian hierarchical models. <i>Nicoletta D'Angelo, Giada Adelfio and Antonino Abbruzzo</i>	1016
4.17.2	Modelling of accumulation curves through Weibull survival functions. <i>Alessandro Zito, Tommaso Rigon and David B. Dunson</i>	1021
4.17.3	Model fitting and Bayesian inference via power expectation propagation. <i>Emanuele Degani, Luca Maestrini and Mauro Bernardi</i>	1026
4.17.4	Bayesian quantile estimation in deconvolution. <i>Catia Scricciolo</i>	1032
4.17.5	Bayesian inference for discretely observed non-homogeneous Markov processes. <i>Rosario Barone and Andrea Tancredi</i>	1038

<b>4.18</b>	<b>New developments in composite indicators applications</b>	<b>1044</b>
4.18.1	Building composite indicators in the functional domain: a suggestion for an evolutionary HDI. <i>Francesca Fortuna, Alessia Naccarato and Silvia Terzi</i>	1045
4.18.2	Small Area Estimation of Inequality Measures via Simplex Regression. <i>Silvia De Nicolò, Maria Rosaria Ferrante and Silvia Pacei</i>	1051
4.18.3	Relational Well-Being and Poverty in Italy Benessere relazionale e povertà in Italia. <i>Elena Dalla Chiara and Federico Perali</i>	1057
4.18.4	A composite indicator to assess sustainability of agriculture in European Union countries. <i>Alessandro Magrini and Francesca Giambona</i>	1063
4.18.5	Interval-Based Composite Indicators with a Triplex Representation: A Measure of the Potential Demand for the “Ristori” Decree in Italy. <i>Carlo Drago</i>	1069
<b>4.19</b>	<b>New developments in GLM theory and applications</b>	<b>1075</b>
4.19.1	Variational inference for the smoothing distribution in dynamic probit models. <i>Augusto Fasano and Giovanni Rebaudo</i>	1076
4.19.2	Interpretability and interaction learning for logistic regression models. <i>Nicola Rares Franco, Michela Carliotta Massi, Francesca Ieva and Anna Maria Paganoni</i>	1082
4.19.3	Entropy estimation for binary data with dependence structures. <i>Linda Altieri and Daniela Cocchi</i>	1088
4.19.4	A Comparison of Some Estimation Methods for the Three-Parameter Logistic Model. <i>Michela Battauz and Ruggero Bellio</i>	1094
4.19.5	A statistical model to identify the price determinations: the case of Airbnb. <i>Giulia Contu, Luca Frigau, Gian Paolo Zammarchi and Francesco Mola</i>	1100
<b>4.20</b>	<b>New developments in social statistics analysis</b>	<b>1106</b>
4.20.1	Data-based Evaluation of Political Agents Against Goals Scheduling. <i>Giulio D'Epifanio</i>	1107
4.20.2	Local heterogeneities in population growth and decline. A spatial analysis for Italian municipalities. <i>Federico Benassi, Annalisa Busetta, Gerardo Gallo and Manuela Stranges</i>	1113
4.20.3	The assessment of environmental and income inequalities. <i>Michele Costa</i>	1119
4.20.4	Household financial fragility across Europe. <i>Marianna Brunetti, Elena Giarda and Costanza Torricelli</i>	1125
4.20.5	Refugees' perception of their new life in Germany. <i>Daria Mendola and Anna Maria Parroco</i>	1131
<b>4.21</b>	<b>New perspectives in clinical trials</b>	<b>1137</b>
4.21.1	Improved maximum likelihood estimator in relative risk regression. <i>Euloge C. Kenne Pagui, Francesco Pozza and Alessandra Salvan</i>	1138
4.21.2	Development and validation of a clinical risk score to predict the risk of SARS-CoV-2 infection. <i>Laura Savaré, Valentina Orlando and Giovanni Corrao</i>	1144
4.21.3	Functional representation of potassium trajectories for dynamic monitoring of Heart Failure patients. <i>Caterina Gregorio, Giulia Barbati<sup>1</sup> and Francesca Ieva</i>	1150
4.21.4	Effect of lung transplantation on the survival of patients with cystic fibrosis: IMaCh contribution to registry data. <i>Cristina Giudici, Nicolas Brouard and Gil Bellis</i>	1156
4.21.5	Categories and Clusters to investigate Similarities in Diabetic Kidney Disease Patients. <i>Veronica Distefano, Maria Mannone, Claudio Silvestri and Irene Poli</i>	1162

<b>4.22</b>	<b>New perspectives in models for multivariate dependency</b>	<b>1168</b>
4.22.1	Parsimonious modelling of spectroscopy data via a Bayesian latent variables approach. <i>Alessandro Casa, Tom F. O'Callaghan and Thomas Brendan Mur</i>	1169
4.22.2	Bias reduction in the equicorrelated multivariate normal. <i>Elena Bortolato and Euloge Clovis Kenne Pagui</i>	1175
4.22.3	Some results on identifiable parameters that cannot be identified from data. <i>Christian Hennig</i>	1181
<b>4.23</b>	<b>Novel approaches for official statistics</b>	<b>1187</b>
4.23.1	Web data collection: profiles of respondents to the Italian Population Census. <i>Elena Grimaccia, Gerardo Gallo, Alessia Naccarato, Novella Cecconi and Alessandro Fratoni</i>	1188
4.23.2	Trusted Smart Surveys: architectural and methodological challenges at a glance. <i>Mauro Bruno, Francesca Inglese and Giuseppina Ruocco</i>	1194
4.23.3	On bias correction in small area estimation: An M-quantile approach. <i>Gaia Bertarelli, Francesco Schirripa Spagnolo, Raymond Chambers and David Haziza</i>	1200
4.23.4	The address component of the Statistical Base Register of Territorial Entities. <i>Davide Fardelli, Enrico Orsini and Andrea Pagano</i>	1206
4.23.5	A well-being municipal indicator using census data: first results. <i>Massimo Esposito</i>	1212
<b>4.24</b>	<b>Prior distribution for Bayesian analysis</b>	<b>1218</b>
4.24.1	On the dependence structure in Bayesian nonparametric priors. <i>Filippo Ascolani, Beatrice Franzolini, Antonio Lijoi, and Igor Prünster</i>	1219
4.24.2	Anisotropic determinantal point processes and their application in Bayesian mixtures. <i>Lorenzo Ghilotti, Mario Beraha and Alessandra Guglielmi</i>	1226
4.24.3	Bayesian Screening of Covariates in Linear Regression Models Using Correlation Thresholds. <i>Ioannis Ntzoufras and Roberta Paroli</i>	1232
<b>4.25</b>	<b>Recent advances in clustering methods</b>	<b>1238</b>
4.25.1	Biclustering longitudinal trajectories through a model-based approach. <i>Francesca Martella, Marco Alfò and Maria Francesca Marino</i>	1239
4.25.2	Monitoring tools for robust estimation of Cluster Weighted models. <i>Andrea Cappozzo and Francesca Greselin</i>	1245
4.25.3	Co-clustering Models for Spatial Transcriptomics: Analysis of a Human Brain Tissue Sample. <i>Andrea Sottosanti and Davide Risso</i>	1251
4.25.4	Graph nodes clustering: a comparison between algorithms. <i>Ilaria Bombelli</i>	1257
<b>4.26</b>	<b>Social demography</b>	<b>1263</b>
4.26.1	Childcare among migrants: a comparison between Italy and France. <i>Eleonora Trappolini, Elisa Barbiano di Belgiojoso, Stefania Maria Lorenza Rimoldi and Laura Terzera</i>	1264
4.26.2	Employment Uncertainty and Fertility in Italy: The Role of Union Formation. <i>Giammarco Alderotti, Valentina Tocchioni and Alessandra De Rose</i>	1270
4.26.3	Determinants of union dissolution in Italy: Do children matter? <i>Valentina Tocchioni, Daniele Vignoli, Eleonora Meli and Bruno Arpino</i>	1276
4.26.4	Working schedules and fathers' time with children: A Sequence Analysis. <i>Annalisa Donno and Maria Letizia Tanturri</i>	1282
4.26.5	Correlates of the non-use of contraception among female university students in Italy. <i>Annalisa Busetta, Alessandra De Rose and Daniele Vignoli</i>	1288

<b>4.27</b>	<b>Social indicators applications and methods</b>	<b>1294</b>
4.27.1	A logistic regression model for predicting child language performance. <i>Andrea Briglia, Massimo Mucciardi and Giovanni Pirrotta</i>	1295
4.27.2	Subject-specific measures of interrater agreement for ordinal scales. <i>Giuseppe Bove</i>	1301
4.27.3	A Tucker3 method application on adjusted-PMRs for the study of work-related mortality. <i>Vittoria Carolina Malpassuti, Vittoria La Serra and Stefania Massari</i>	1307
4.27.4	Two case-mix adjusted indices for nursing home performance evaluation. <i>Giorgio E. Montanari and Marco Doretti</i>	1313
4.27.5	The ultrametric covariance model for modelling teachers' job satisfaction. <i>Carlo Cavicchia, Maurizio Vichi and Giorgia Zaccaria</i>	1319
<b>4.28</b>	<b>Some recent developments in compositional data analysis</b>	<b>1325</b>
4.28.1	A Robust Approach to Microbiome-Based Classification Problems. <i>Gianna Serafina Monti and Peter Filzmoser</i>	1326
4.28.2	What is a convex set in compositional data analysis? <i>Jordi Saperas i Riera, Josep Antoni Martín Fernández</i>	1332
4.28.3	Compositional Analysis on the Functional Distribution of Extended Income. <i>Elena Dalla Chiara and Federico Perali</i>	1338
4.28.4	Evaluating seasonal-induced changes in river chemistry using Principal Balances. <i>Caterina Gozzi and Antonella Buccianti</i>	1344
4.28.5	Compositional Data Techniques for the Analysis of the Ragweed Allergy. <i>Gianna S. Monti, Maira Bonini, Valentina Ceriotti, Matteo Pelagatti and Claudio M. Ortolani</i>	1350
<b>4.29</b>	<b>Spatial data analysis</b>	<b>1356</b>
4.29.1	Spatial multilevel mixed effects modeling for earthquake insurance losses in New Zealand. <i>F. Marta L. Di Lascio and Selene Perazzini</i>	1357
4.29.2	Weighted distances for spatially dependent functional data. <i>Andrea Diana, Elvira Romano, Claire Miller and Ruth O'Donnell</i>	1363
4.29.3	Spatial modeling of childcare services in Lombardia. <i>Emanuele Aliverti, Stefano Campostrini, Federico Caldura and Lucia Zanotto</i>	1369
4.29.4	On the use of a composite attractiveness index for the development of sustainable tourist routes. <i>Claudia Cappello, Sandra De Iaco, Sabrina Maggio and Monica Palma</i>	1375
<b>4.30</b>	<b>Statistical applications in education</b>	<b>1381</b>
4.30.1	Does self-efficacy influence academic results? A separable-effect mediation analysis. <i>Chiara Di Maria</i>	1382
4.30.2	Statistics Knowledge assessment: an archetypal analysis approach. <i>Bruno Adabbo, Rosa Fabbricatore, Alfonso Iodice D'Enza and Francesco Palumbo</i>	1388
4.30.3	Exploring drivers for Italian university students' mobility: first evidence from AlmaLaurea data. <i>Giovanni Boscaino and Vincenzo Giuseppe Genova</i>	1394
4.30.4	Can Grading Policies influence the competition among Universities of different sizes? <i>Gabriele Lombardi and Antonio Pio Distaso</i>	1400
4.30.5	The class A journals and the Italian academic research outcomes in Statistical Sciences. <i>Maria Maddalena Barbieri, Francesca Bassi, Antonio Irpino and Rosanna Verde</i>	1406
<b>4.31</b>	<b>Statistical methods for finance</b>	<b>1412</b>
4.31.1	Hypotheses testing in mixed-frequency volatility models: a bootstrap approach. <i>Vincenzo Candila and Lea Petrella</i>	1413



4.31.2	Quantile Regression Forest with mixed frequency Data. <i>Mila Andreani, Vincenzo Candila and Lea Petrella</i>	1419
4.31.3	Higher order moments in Capital Asset Pricing Model betas. <i>Giuseppe Arbia, Riccardo Bramante and Silvia Facchinetti</i>	1425
4.31.4	When Does Sentiment Matter in Predicting Cryptocurrency Bubbles? <i>Arianna Agosto and Paolo Pagnottoni</i>	1431
<b>4.32</b>	<b>Statistical methods for high dimensional data</b>	<b>1437</b>
4.32.1	Virtual biopsy in action: a radiomic-based model for CALI prediction. <i>Francesca Ieva, Giulia Baroni, Lara Cavinato, Chiara Masci, Guido Costa, Francesco Fiz, Arturo Chiti and Luca Viganò</i>	1438
4.32.2	Functional alignment by the “light” approach of the von Mises-Fisher-Procrustes model. <i>Angela Andreella and Livio Finos</i>	1444
4.32.3	A screening procedure for high-dimensional autologistic models. <i>Rodolfo Metulini and Francesco Giordano</i>	1450
4.32.4	Covariate adjusted censored gaussian lasso estimator. <i>Luigi Augugliaro, Gianluca Sottile and Veronica Vinciotti</i>	1456
4.32.5	Ranking-Based Variable Selection for ultra-high dimensional data in GLM framework. <i>Francesco Giordano, Marcella Niglio and Marialuisa Restaino</i>	1462
<b>4.33</b>	<b>Statistical methods in higher education</b>	<b>1468</b>
4.33.1	Effects of remote teaching on students' motivation and engagement: the case of the University of Modena & Reggio Emilia. <i>Isabella Morlini and Laura Sartori</i>	1469
4.33.2	A random effects model for the impact of remote teaching on university students' performance. <i>Silvia Bacci, Bruno Bertaccini, Simone Del Sarto, Leonardo Grilli and Carla Rampichini</i>	1475
4.33.3	Multinomial semiparametric mixed-effects model for profiling engineering university students. <i>Chiara Masci, Francesca Ieva and Anna Maria Paganoni</i>	1481
4.33.4	Evaluating Italian universities: ANVUR periodic accreditation judgment versus international rankings. <i>Angela Maria D'Uggento, Nunziata Ribecco and Vito Ricci</i>	1487
4.33.5	Women's career discrimination in the Italian Academia in the last 20. <i>Daniele Cuntrera, Vincenzo Falco and Massimo Attanasio</i>	1493
<b>4.34</b>	<b>Statistical methods with Bayesian networks</b>	<b>1499</b>
4.34.1	Statistical Micro Matching Using Bayesian Networks. <i>Pier Luigi Conti, Daniela Marella, Paola Vicard and Vincenzina Vitale</i>	1500
4.34.2	Modeling school managers challenges in the pandemic era with Bayesian networks. <i>Maria Chiara De Angelis and Flaminia Musella and Paola Vicard</i>	1506
4.34.3	Structural learning of mixed directed acyclic graphs: a copula-based approach. <i>Federico Castelletti</i>	1512
4.34.4	Inference on Markov chains parameters via Large Deviations ABC. <i>Cecilia Viscardi, Fabio Corradi, Michele Boreale and Antonietta Mira</i>	1518
4.34.5	A propensity score approach for treatment evaluation based on Bayesian Networks. <i>Federica Cugnata, Paola M.V. Rancoita, Pier Luigi Conti, Alberto Briganti, Clelia Di Serio, Fulvia Mecatti and Paola Vicard</i>	1524
<b>4.35</b>	<b>Statistical modelling for the analysis of contemporary societies</b>	<b>1530</b>
4.35.1	Social Network Analysis to analyse the relationship between 'victim-author' and 'motivation' of violence against women in Italy. <i>Alessia Forciniti</i>	1531
4.35.2	Satisfaction and sustainability propensity among elderly bike-sharing users. <i>Paolo Maranzano, Roberto Ascari, Paola Maddalena Chiodini and Giancarlo Manzi</i>	1537



4.35.3	Media and Investors' Attention. Estimating analysts' ratings and sentiment of a financial column to predict abnormal returns. <i>Riccardo Ferretti and Andrea Sciandra</i>	1543
4.35.4	Predictions of regional HCE: spatial and time patterns in an ageing population framework. <i>Laura Rizzi, Luca Grassetti, Divya Brundavanam, Alvisa Palese and Alessio Fornasin</i>	1549
<b>4.36</b>	<b>Surveillance methods and statistical models in the Covid-19 crisis</b>	<b>1555</b>
4.36.1	The Italian Social Mood on Economy Index during the Covid-19 Crisis. <i>Alessandra Righi and Diego Zardetto</i>	1556
4.36.2	Modeling the first wave of the COVID-19 pandemic in the Lombardy region, Italy, by using the daily number of swabs. <i>Claudia Furlan and Cinzia Mortarino</i>	1562
4.36.3	Analysing the Covid-19 pandemic in Italy with the SIPRO model. <i>Martina Amongero, Enrico Bibbona and Gianluca Mastrantonio</i>	1568
4.36.4	Intentions of union formation and dissolution during the COVID-19 pandemic. <i>Bruno Arpino and Daniela Bellani</i>	1574
<b>4.37</b>	<b>Time series methods</b>	<b>1580</b>
4.37.1	Bootstrap-based score test for INAR effect. <i>Riccardo Ievoli and Lucio Palazzo</i>	1581
4.37.2	Evaluating the performance of a new picking algorithm based on the variance piecewise constant models. <i>Nicoletta D'Angelo, Giada Adelfio, Antonino D'Alessandro and Marcello Chiodi</i>	1587
4.37.3	Conditional moments based time series cluster analysis. <i>Raffaele Mattera and Germana Scepi</i>	1593
4.37.4	On the asymptotic mean-squared prediction error for multivariate time series. <i>Gery Andrés Díaz Rubio, Simone Giannerini, and Greta Goracci</i>	1599
4.37.5	Spherical autoregressive change-point detection with applications. <i>Federica Spoto, Alessia Caponera and Pierpaolo Brutti</i>	1605
<b>5</b>	<b>Posters</b>	<b>1611</b>
5.1	A method for incorporating historical information in non-inferiority trials. <i>Fulvio De Santis and Stefania Gubbiotti</i>	1612
5.2	Optimal credible intervals under alternative loss functions. <i>Fulvio De Santis and Stefania Gubbiotti</i>	1618
5.3	Statistical learning for credit risk modelling. <i>Veronica Bacino, Alessio Zoccarato, Caterina Liberati and Matteo Borrotti</i>	1624
5.4	Evaluating heterogeneity of agreement with strong prior information. <i>Federico M. Stefanini</i>	1630
5.5	Analysis of the spatial interdependence of the size of endoreduplicated nuclei observed in confocal microscopy. <i>Ivan Sciascia, Andrea Crosino, Gennaro Carotenuto and Andrea Genre</i>	1636
5.6	A Density-Peak Approach to Clustering Graph-Structured Data. <i>Riccardo Giubilei</i>	1642
5.7	The employment situation of people with disabilities in Tuscany, A Survey on the workplace. <i>Paolo Addis, Alessandra Coli and Gianfranco Francese</i>	1648
5.8	Robustness of statistical methods for modeling paired count data using bivariate discrete distributions with general dependence structures. <i>Marta Nai Ruscone and Dimitris Karlis</i>	1654

<b>6</b>	<b>Satellite events</b>	<b>1660</b>
6.1	Measuring uncertainty in key official economic statistics	1661
6.1.1	Uncertainty in production and communication of statistics: challenges in the new data ecosystem. <i>Giorgio Alleva and Piero Demetrio Falorsi</i>	1662
6.1.2	Uncertainty and variance estimation techniques for poverty and inequality measures from complex surveys: a simulation study. <i>Riccardo De Santis, Lucio Barabesi and Gianni Betti</i>	1668
6.1.3	Pandemics and uncertainty in business cycle analysis. <i>Jacques Anas, Monica Billio, Leonardo Carati, Gian Luigi Mazzi and Hionia Vlachou</i>	1674
6.2	<b>Covid-19: the urgent call for a unified statistical and demographic challenge</b>	<b>1680</b>
6.2.1	Environmental epidemiology and the Covid-19 pandemics	1681
6.2.1.1	The Covid-19 outbreaks and their environment: The Valencian human behaviour. <i>Xavier Barber, Elisa Espín, Lucía Guevara, Aurora Mula, Kristina Polotskaya and Alejandro Rabasa</i>	1682
6.2.2	Estimation of Covid 19 prevalence	1686
6.2.2.1	Optimal spatial sampling for estimating the SARS-Cov-2 crucial parameters. <i>Piero Demetrio Falorsi and Vincenzo Nardelli</i>	1687
6.2.2.2	Survey aimed to estimate the seroprevalence of SARS-CoV-2 infection in Italian population at national and regional level. <i>Stefano Falorsi, Michele D'Alò, Claudia De Vitiis, Andrea Fasulo, Danila Filippini, Alessio Guandalini, Francesca Inglese, Orietta Luzi, Enrico Orsini and Roberta Radini</i>	1693
6.2.3	Measuring and modeling inequalities following the Covid-19 crisis	1699
6.2.3.1	COVID-19 impacts on young people's life courses: first results. <i>Antonietta Bisceglia, Concetta Scolorato and Giancarlo Ragozini</i>	1700
6.2.3.2	Exploring Students' Profile and Performance Before and After Covid-19 Lock-down. <i>Cristina Davino and Marco Gherghi</i>	1705
6.2.4	Nowcasting the Covid-19 outbreaks methods and applications	1711
6.2.4.1	Modeling subsequent waves of COVID-19 outbreak: A change point growth model. <i>Luca Greco, Paolo Girardi and Laura Ventura</i>	1712
6.2.4.2	The second wave of SARS-CoV-2 epidemic in Italy through a SIRD model. <i>Michela Baccini and Giulia Cereda</i>	1718
6.2.5	The impact of Covid-19 on survey methods	1724
6.2.5.1	Collecting cross-national survey data during the COVID-19 pandemic: Challenges and implications of data collection for the 50+ population in the Survey of Health, Ageing and Retirement in Europe (SHARE). <i>Michael Bergmann, Arne Bethmann, Yuri Pettinicchi and Borsch-Supan</i>	1725
6.2.5.2	Adapting a Long-Term Panel Survey to Pandemic Conditions. <i>Peter Lynn</i>	1731
6.2.6	Young contributions in Covid-19 statistical modelling	1737
6.2.6.1	Statistical communication of COVID-19 epidemic using widely accessible interactive tools. <i>Marco Mingione and Pierfrancesco Alaimo Di Loro</i>	1738
6.2.6.2	Modelling COVID-19 evolution in Italy with an augmented SIRD model using open data. <i>Vincenzo Nardelli, Giuseppe Arbia, Andrea Palladino and Luigi Giuseppe Atzeni</i>	1744

# **Analysing contributions of ages and causes of death to gender gap in life expectancy using functional data analysis**

## ***Analisi dei contributi per età e cause di morte alla differenza di genere nell'aspettativa di vita attraverso l'analisi dei dati funzionali***

Alessandro Feraldi, Virginia Zarulli, Stefano Mazzuco and Cristina Giudici

**Abstract** The work consists of application of functional data analysis (FDA) to demographic data: it analyses the contribution of ages and causes of death to gender gap in life expectancy in 14 European and non-European countries between 1998 and 2016. Causes-of-death data and life tables were retrieved from the Human Causes-of-Death Database (HCD) and from the Human Mortality Database (HMD). Our analysis allows to identify two main components that capture most of the variability and which captures the extent of the cause-specific gender differences and the age pattern, respectively. Over time, an increase in the most relevant contributions is observed, especially around the modal age and a shift of the contributions towards older age.

**Abstract** Il lavoro consiste in un'applicazione dell'analisi dei dati funzionali (FDA) a dati demografici: si analizza il contributo delle età e cause di morte alle differenze di genere nella speranza di vita in 14 paesi europei ed extraeuropei nel periodo compreso tra il 1998 e il 2016. I dati sui decessi per causa provengono dallo Human Cause-of-Death Database (HCD), mentre le tavole di mortalità sono tratte dallo Human Mortality Database (HMD). L'analisi consente di individuare due componenti principali che colgono gran parte della variabilità e che descrivono rispettivamente l'entità delle differenze di genere specifiche per causa e i contributi età-specifici. Nel tempo, si osserva un aumento dei contributi più rilevanti soprattutto intorno all'età modale ed uno spostamento degli stessi verso l'età avanzata.

**Key words:** gender gap, life expectancy, causes of death, functional data analysis

---

Alessandro Feraldi, Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5 (00185) Rome, e-mail: [alessandro.feraldi@uniroma1.it](mailto:alessandro.feraldi@uniroma1.it)

Virginia Zarulli, Interdisciplinary Centre on Population Dynamics, CPOP, University of Southern Denmark. Campusvej 55. Odense M DK-5230, e-mail: [vzarulli@sdu.dk](mailto:vzarulli@sdu.dk)

Stefano Mazzuco, Department of Statistical Sciences, University of Padova, Via Cesare Battisti 241, 35121 Padova, Italy, email: [stefano.mazzuco@unipd.it](mailto:stefano.mazzuco@unipd.it)

Cristina Giudici, MEMOTEF Department, Sapienza University of Rome, Via del Castro Laurenziano, 9, 00161 Rome, Italy, e-mail: [cristina.giudici@uniroma1.it](mailto:cristina.giudici@uniroma1.it)

## 1 Introduction

On average and worldwide, women live longer than men and the absolute difference between male and female mortality risk reaches its maximum at old ages. From the beginning of the 1920s, in most industrialized countries the gap in life expectancy between the two sexes widened until the 1970s, when the difference started to narrow (Austad (2006); Zarulli et al. 2020)). The literature shows that where some convergence has taken place, men have experienced more rapid gains in survival than women (Meslé & Vallin (2011)).

To explain the recent narrowing in the sex gap in LE, several studies have focused on the causes of death that contributed to the gender gap in mortality rates and, thereby, either narrowed or widened this gap (Klenk et al. (2016)). In most of the studies, the topic of the variability in the gender gap in mortality has usually been tackled comparing age trajectories of cause-specific death rates between men and women by fitting specific parametric models on cause-specific life table death rate for women and men separately (Horiuchi et al. (2013)). Other studies used life table or aggregated mortality indicators to provide summary measures of mortality levels (e.g. life expectancy) and dispersion (e.g. lifespan variation) both separately for men and women and with the decomposition of the difference between sexes according to age and causes of death (Trias-Llimós & Janssen (2018)).

Although numerous studies have decomposed the sex gap in life expectancy according to age and causes of death, they did not study the principal components of the contributions of age and causes of death from a functional perspective, which has been shown to be more informative approach (Ramsay & Silverman (2002); Léger & Mazzuco (2020)). To fill this gap, we study absolute and relative contributions of age and causes of death to the gender gap in life expectancy (GGLE) for several countries, using the Functional Data Analysis (FDA) (Ramsay & Silverman (2002)). Following this approach, we consider age- and cause-specific contributions as functions, and therefore we analyse curves rather than scalar data. More specifically we propose a Functional Principal Component Analysis (FPCA) of the contribution profiles of several countries, in order to identify the main components of the distribution of age-specific contributions according to causes of death. To the best of our knowledge, this is the first study analysing age- and cause-specific contributions to the gender gap in life expectancy with a functional data analysis approach.

## 2 Data and Method

### 2.1 Data

Cause-specific mortality data were retrieved by gender, 5-year age interval and year from the Human Cause-of-Death Database (HCD (2019)) and life tables were taken from the Human Mortality Database (HMD (2019)). HMD and HCD are open-source projects: the former offers harmonized data on constructed series of mortality rates,

## Analysing contributions of age and cause of death to the gender gap in life expectancy using functional data analysis

life tables, death counts and population exposures; the latter contains reconstructed long-term trends in cause-specific mortality for sixteen countries over time. Since this study aims at analysing patterns of causes of death across several countries, we focussed on the last 15 years available for each country, within the period 1998-2016. Countries were grouped according to geographical areas: Eastern Europe (EE) (i.e. Russia, Ukraine, Poland, the Czech Republic, Estonia, Latvia, Lithuania and Belarus); Western Europe (WE) (i.e. France, Spain, Germany and the United Kingdom (UK)), and extra-European countries (i.e. Japan and the United States (US)). Romania and Moldova were excluded from the analysis because no data were available in the HMD. Within this time frame all the causes of death are coded in each country according to the 10th version of International Classification of Disease (ICD-10). This allows us to avoid problems related to differences in the classification and to obtain comparable information for all the countries under study. We further restricted our focus on the short list of the ICD-10, in which all the causes are grouped into sixteen major categories, each including a set of similar diseases (e.g. heart diseases, neoplasms, external causes, respiratory diseases etc.) (HCD (2019)). Finally, cause-specific mortality data for all the ages above 85 were grouped in the open-end age interval 85+, to avoid problems related to the data quality, which are particularly common at very old ages.

### 2.2 Analysis

Age- and cause-specific contributions to the GGLE (female - male) were obtained for each country and over time applying Arriaga's age- and cause-specific decomposition technique combining life tables from the HMD and cause-specific mortality data from the HCD (Arriaga (1984)). FDA was applied to the age-specific relative contributions to the GGLE, separately for each cause of death. Discrete age-specific relative contribution data  $x(t_1), \dots, x(t_N)$  were assumed to be independent realizations drawn from the same continuous stochastic process  $X(t)$  (Ramsay & Silverman (2002)). To obtain the functional representation, each  $X(t)$  was approximated by using a basis expansion of cubic B-splines functions (1) and the B-splines basis coefficients were estimated by the ordinary least squares method minimizing the sum of squared residuals (Léger & Mazzuco (2020)). Therefore,

$$X(t) = \sum_{j=1}^p \gamma_j \psi_j(t) \quad (1)$$

where  $\psi_j$  are  $p$  known basis functions and  $\gamma_j$  are the corresponding coefficients to be estimated. In order to maintain the data structure, we used a sequence of  $p = 19$  equally distributed knots (i.e. one for 5-years age interval). Afterwards, we performed FPCA separately for each cause of death in order to synthesize the variability of the curves and to identify the main components of the distributions of age-specific contributions according to causes of death, across countries. FPCA is the extension of the more classical multivariate PCA to functional data: for a generic curve  $x_i(t)$  we can obtain the approximation (2)

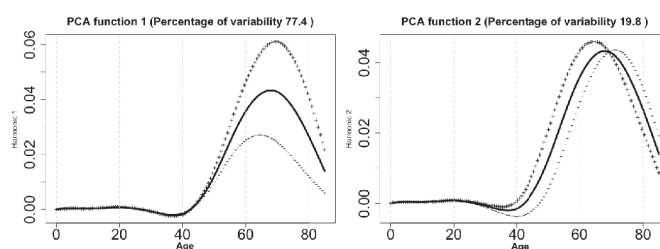
$$x_i(t) = \sum_{k=1}^{\infty} c_{i,k} \phi_k(t) \quad (2)$$

where  $c_{i,k}$  are the principal component scores and  $\phi_k(t)$  are the eigenfunctions or harmonics. Therefore, the information on the curve  $x_i(t)$  were then synthesized by the first  $q$  terms. All the analyses were conducted using the R package *fda* (Ramsay et al. (2011)).

### 3 Results

Decomposition results confirmed that neoplasms, heart diseases and external causes of death made the largest contributions to the GGLE in all the countries, explaining together more than two third of the overall gap. Additionally, the largest contributions to the GGLE were given by old ages for most of the causes of death over the entire period (Meslé & Vallin (2011); Trias-Llimós & Janssen (2018)). Results of FPCA focus on the three most relevant causes of death.

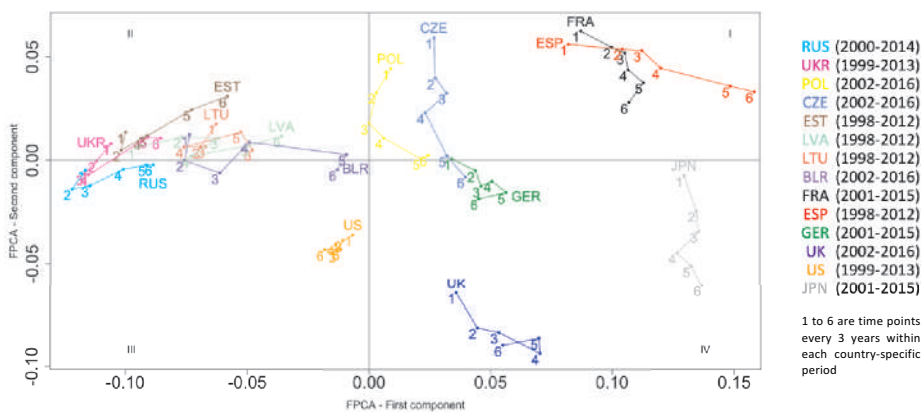
Most of the variability in the age-specific contributions is explained by the first two principal components for each cause of death (e.g. 97%, 94% and 95% for neoplasm, heart diseases and external causes, respectively). The first FPC mainly captures the extent of the cause-specific gender differences, while the second FPC captures the age pattern. A classical way to interpret the FPCs is to plot the group mean function (solid curve in Figure 1) as well as the functions obtained by adding (+ curve) and subtracting (- curve) to the mean function twice the square root of the principal component variance (Ramsay & Silverman (2002); Léger & Mazzuco (2020)). Regarding neoplasm (Figure 1), for the first FPC, the variability is concentrated at 40 years' age and older and especially around the modal age. A high score on this component suggests an above-average contribution. The second FPC corresponds to a shift of the curves with respect to the overall contribution mean towards older ages. The (+) curve has a higher contribution than the (-) curve with respect to the mean curve before 70 years, lower afterwards. A low score on this component suggests an above-average shift of the distribution towards older ages.



**Figure 1.** Effect of the first two FPCs on age-specific contributions to GGLE for neoplasm: overall mean and mean  $\pm$  a suitable multiple of the principal component weight function.

**Analysing contributions of age and cause of death to the gender gap in life expectancy using functional data analysis**

In order to study the evolutions of age-cause contributions over time, the scores of the two first components for each country are plotted at every 3 years. With regards to neoplasm (Figure 2), the first axis indicates that, throughout the whole period, relative contributions to the GGLE were higher in France, Spain (I quarter) and in Japan (IV quarter) than in the other countries, followed by UK, Germany, Poland, Czech Republic and US. The lowest contributions were shown in Ukraine, Estonia, Latvia, Lithuania, Belarus and in Russia (II quarter). The second axis denotes that the distribution of age-specific contributions to the GGLE was more concentrated at older ages in UK than in the other countries. Furthermore, distributions were more concentrated at older ages in US and in Japan (scores of the second FPC < 0) than in the overall mean distribution (i.e. for all the countries).



**Figure 2.** Principal component subspace - neoplasm.

On the contrary, in the remaining countries distributions were more concentrated at younger ages than in the overall mean distribution. Moreover, the increasing trends in the first FPC scores, especially in Spain, together with the decreasing trends in second FPC scores showed also in France and in UK, indicate increasing relative contribution of neoplasm over time as well as a shift of the distribution towards older ages. Similarly, in Poland, Czech Republic and in Japan, although the relative contributions of neoplasm to the GGLE stagnated over time (small variations in first FPC scores), the decreasing trends of the second FPC scores denote that the distributions of age-specific contributions shifted towards older ages over time. A small increase in the score of the first component in Estonia and in Belarus (especially at the second half of the period) suggests a slight increase in the relative contribution of neoplasm to the GGLE. Finally, stagnation in both FPC scores was reported in Germany, Russia, Ukraine and in US over time. Therefore, independent from trends in relative contributions of neoplasm to the GGLE, in most of the countries the distributions shifted towards older ages over time.

For the sake of brevity, we do not display the results of FPCA for the other two main causes, however the analysis shows similar patterns across countries and over time.



## Conclusion

This work gives a deeper insight of the main contributing factors to the gender differences in life expectancy, analysing components of the relative age-specific contributions according to the most relevant causes of death in determining the gap. The study also aims at illustrating the demographic application of FDA, a new method for demographic analysis but which could prove useful to deepen our understanding of complex demographic phenomena. Our results allow to identify two main components which capture most of the variability. The first component mainly captures the extent of the cause-specific gender differences, while the second FPC captures the age patterns. Over time, an increase in the most relevant contributions is observed, especially around the modal age and a shift of the contributions towards older age. The analysis confirms that FDA allows to highlight country-specific patterns in the context of the epidemiological transition which need to be further analysed. Further analyses include functional cluster analysis to group countries according to age-cause contributions and study the evolution of the contributions in each cluster over time. Finally, following the FDA approach we also apply other statistical analyses (i.e. regression and hypothesis tests) to the same data and suggest to increase the use of such approach in population studies.

## References

1. Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. *Demography*, 21(1), 83–96.
2. Austad, S. N. (2006). Why women live longer than men: Sex differences in longevity. *Gender Medicine. Official Journal of the Partnership for Gender-Specific Medicine at Columbia University*, 3(2), 79–92.
3. Horiuchi, S., Ouellette, N., Cheung, S. L. K., & Robine, J.-M. (2013). Modal age at death: Lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, 37–69.
4. Human Cause-of-Death Database. (2019). French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany) (downloaded on 2019).
5. Human Mortality Database. (2019). Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (downloaded on 2019).
6. Klenk, J., Keil, U., Jaensch, A., Christiansen, M. C., & Nagel, G. (2016). Changes in life expectancy 1950–2010: Contributions from age- and disease-specific mortality in selected countries. *Population Health Metrics*, 14, 20.
7. Léger, A. E., & Mazzuco, S. (2020). What can we learn from functional clustering of mortality data? An application to HMD data. *arXiv preprint arXiv:2003.05780*.
8. Meslé, F., & Vallin, J. (2011). Historical trends in mortality. In *International handbook of adult mortality* (pp. 9–47). Springer.
9. Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies*. Springer Series in Statistics, Springer-Verlag.
10. Ramsay, J., Wickham, H., Graves, S., and Hooker, G. (2011). *fda: Functional Data Analysis*. R package.
11. Trias-Llimós, S., & Janssen, F. (2018). Alcohol and gender gaps in life expectancy in eight Central and Eastern European countries. *European Journal of Public Health*, 28(4), 687–692.
12. Zarulli, V., Lindahl-Jacobsen, R., & Vaupel, J. W. (2020). Onset of the old-age gender gap in survival. *Demographic Research*, 42, 727–740.