Scientific Conference on

Statistics for Health and Well-being

University of Brescia
Department of Economics and Management
25 – 27 September 2019

SUMMARY SCHEME

CONFERENCE PROGRAM

BOOK OF ABSTRACTS
INTRODUCTION

The ASA – Association for Applied Statistics and the DMS StatLab - Data Methods and Systems Statistical Laboratory of the Department of Economics and Management of the University of Brescia organized in Brescia the Scientific Conference "Statistics for Health and Well-being", aimed at promoting applications that use statistical techniques and models suitable for health and well-being analyses.

The conference is three days long, from 25 to 27 September, 2019. Scientific reports at the Conference are partly invited by the Scientific Committee and partly contributed by participants. Concerned scholars have submitted to the Scientific Committee an abstract of one page and, if accepted, also an optional short paper of four pages.

Moreover, other scientific events, round-table and sensory games on food and beverages are held. At some invited session two basic speeches will open the discussion, one held by an academic and the other by an expert from a company or an institution currently using statistics. This juxtaposition is aimed at favouring a better understanding between academia and other worlds.

The final program is composed by 25 parallel sessions with a total of 82 contributions, 4 plenary sessions (ISTAT-USCI, SIS-ASA, AICQCN-AISS, DBSPORTS) and 4 special events (ISTAT-ASA Open Conference with the President of ISTAT, IASA Sensory Experiment, Visit to Capitolium, Kick-off meeting ISI-SPG in Sports Statistics).

We are grateful to the members of the Scientific and Local Program Committees, to the chair and to the speakers of the ASA Conference 2019.

Brescia, September 25, 2019

Co-chair Scientific Program Committee
Maurizio Carpita
University of Brescia

Chair Local Program Committee
Luigi Fabbri
University of Padova

Eugenio Brentari
University of Brescia
Conference session topics include, but are not limited to, the following areas of special interest:

- Health and healthcare
- Education and health
- Health Psychology
- Work and life balance
- Economic well-being
- Social relationships and social health
- Welfare and well-being
- Safety and security
- Subjective well-being
- Environment and pollution
- Innovation, research and creativity
- Quality of health services
- Equitable and sustainable well-being
- Resilience and vulnerability
- Sport, Health and wellbeing
- Sport analytics
- Health and fitness
- Sport psychology
- Statistics and tourism
- Food and beverage, health, well-being and life quality
- Qualitative and quantitative methods for sensory analysis
- Psychology and food
- Food and beverage industries and markets
- Methods and models for health and well-being analysis
- Technology for health analysis

**Scientific Program Committee:**

- Luigi Fabbris (University of Padua, co-chair)
- Maurizio Carpita (University of Brescia, co-chair)
- Giuseppe Arbia (SIS - Università Cattolica di Milano)
- Rosella Berni (University of Florence)
- Matilde Bini (SIS - European University of Rome)
- Giovanna Bocuzzi (University of Padova)
- Eugenio Brentari (University of Brescia)
- Vittoria Buratta (ISTAT)
- Giulia Cavrini (University of Bolzano-Bozen)
- Alessandro Celegato (AICQ-AISS, PSV Project Service and Value)
- Giuliana Coccia (ISTAT)
- Adriano Decarli (University of Milan)
- Tonio Di Battista (‘G. D’Annunzio’ University of Chieti and Pescara)
- Simone Di Zio (‘G. D’Annunzio’ University of Chieti and Pescara)
- Benito Vittorio Frosini (Sacred Heart Catholic University of Milan)
- Antonio Giusti (University of Florence)
- Silvia Golia (University of Brescia)
- Maria Gabriella Grassia (Federico II University of Naples)
- Maria Iannario (Federico II University of Naples)
- Domenica Fioredistella Iezzi (Tor Vergata University of Rome)
- Michele Lalla (University of Modena and Reggio Emilia)
- Fabio Lucidi (SIPS - La Sapienza University of Rome)
- Marica Manisera (University of Brescia)
- Paolo Mariani (University of Milan-Bicocca)
- Francesco Mola (University of Cagliari)
- Antonio Mussino (La Sapienza University of Rome)
- Luigi Odelo (International Academy of Sensory Analysis)
- Francesco Palumbo (Federico II University of Naples)
- Maurizio Pessato (Assirm)
- Alessandra Petrucci (University of Florence)
- Alfonso Piscitelli (Federico II University of Naples)
- Marco Trentini (Unione Statistica Comuni Italiani)
- Fabio Vernau (Federico II University of Naples)
- Domenico Vistocco (Federico II University of Naples)
- Paola Zuccolotto (University of Brescia)

**Local Program Committee:**

- Eugenio Brentari (University of Brescia, chair)
- Maurizio Carpita (University of Brescia)
- Silvia Golia (University of Brescia)
- Marica Manisera (University of Brescia)
- Manlio Miglieri (University of Brescia)
- Anna Simonetto (University of Brescia)
- Marika Vezzoli (University of Brescia)
- Mariangela Zenga (University of Milano-Bicocca)
- Paola Zola (University of Brescia)
- Paola Zuccolotto (University of Brescia)
INDEX OF ASA CONFERENCE 2019

• Summary Scheme pag. I

• Conference Program pag. II

• Special Events pag. 1

• Plenary Sessions pag. 3

• Book of Abstracts pag. 12
# SUMMARY SCHEME

<table>
<thead>
<tr>
<th>Wednesday - Sept. 25, 2019</th>
<th>Thursday - Sept. 26, 2019</th>
<th>Friday - Sept. 27, 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UniBS</strong></td>
<td><strong>UniBS</strong></td>
<td></td>
</tr>
<tr>
<td>C.da S. Chiara, 50</td>
<td>C.da S. Chiara, 50</td>
<td></td>
</tr>
<tr>
<td>08:30 - 09:15</td>
<td>Registration of partecipants</td>
<td></td>
</tr>
<tr>
<td>09:15 - 09:30</td>
<td>Welcome Addresses</td>
<td></td>
</tr>
<tr>
<td>09:30 - 10:30</td>
<td>Plenary Session 1 ISTAT &amp; USCI</td>
<td>09:30 - 10:30 Plenary Session 1 AICQ-CN &amp; AISS</td>
</tr>
<tr>
<td>10:30 - 11:00</td>
<td>Coffee Break</td>
<td>10:30 - 11:00 Coffee Break</td>
</tr>
<tr>
<td>11:00 - 12:00</td>
<td>1st parallel sessions</td>
<td>11:00 - 12:00 1st parallel sessions</td>
</tr>
<tr>
<td>12:00 - 13:00</td>
<td>Plenary Session 2 SIS &amp; ASA</td>
<td>12:00 - 13:00 Plenary Session 2 DBsports</td>
</tr>
<tr>
<td>13:00 - 14:00</td>
<td>Lunch</td>
<td>13:00 - 14:30 Lunch</td>
</tr>
</tbody>
</table>

**MO.CA**<br>Via Moretto, 78

| 14:00 - 15:00              | Special Event IASA experiment Chocolate Sensory Maps | 14:30 -15:30 2nd parallel sessions |
| 15:00 - 15:15              | Institutional greetings |                         |
| 15:00 - 16:00              | 2nd parallel sessions  | 15:30 - 16:30 Special Event Kick-off meeting ISI-SPG in Sports Statistics |
| 15:15 - 17:00              | Special Event ISTAT & ASA | 16:00 - 17:00 3rd parallel sessions |
| 17:00 - 17:30              | Coffee Break             | 16:30 - 17:00 Conference closing and drink to the health |
| 17:30 - 18:30              | ASA General Assembly     | Special Event Visit to the Capitolium |
| 20:30                      | ASA Dinner               | ASA Social Dinner     |
**ASA 2019 SCIENTIFIC CONFERENCE ON STATISTICS FOR HEALTH AND WELL-BEING CONFERENCE PROGRAM**

**DEPARTMENT OF ECONOMICS AND MANAGEMENT**

**UNIVERSITY OF BRESCIA**

**SEPTEMBER 25 – 27, 2019**

---

**Wednesday - September 25, 2019**

**MO.CA - Via Moretto, 78**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.00 – 15.15</td>
<td><strong>INSTITUTIONAL GREETINGS</strong></td>
</tr>
<tr>
<td>15.15 – 17.00</td>
<td><strong>SPECIAL EVENT ORGANIZED BY ISTAT &amp; ASA &amp; UniBS</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SCENARI STATISTICI SU POPOLAZIONE SALUTE E BENESSERE IN ITALIA</strong></td>
</tr>
<tr>
<td></td>
<td>Relazione pubblica del Presidente ISTAT Prof. Gian Carlo Blangiardo</td>
</tr>
<tr>
<td></td>
<td>Seguirà dibattito moderato da Anna Della Moretta del Giornale di</td>
</tr>
<tr>
<td></td>
<td>Brescia con gli interventi di:</td>
</tr>
<tr>
<td></td>
<td>• Emilio Del Bono, <strong>Sindaco del Comune di Brescia</strong></td>
</tr>
<tr>
<td></td>
<td>• Filomena Maggino, <strong>Università di Roma La Sapienza, Consigliere e</strong></td>
</tr>
<tr>
<td></td>
<td><strong>rappresentante del Presidente del Consiglio dei Ministri per la</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Cabina di regia &quot;Benessere Italia&quot;</strong></td>
</tr>
<tr>
<td></td>
<td>• Clelia di Serio, <strong>Università Vita-Salute San Raffaele e Presidente</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Società Italiana di Biometria</strong></td>
</tr>
<tr>
<td></td>
<td>• Maurizio Tira, <strong>Magnifico Rettore dell'Università degli Studi di</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Brescia</strong></td>
</tr>
<tr>
<td>17.00 – 17.30</td>
<td><strong>COFFEE BREAK</strong></td>
</tr>
<tr>
<td>17.30 – 18.30</td>
<td><strong>ASA GENERAL ASSEMBLY</strong></td>
</tr>
<tr>
<td>20.30 – 23.00</td>
<td><strong>ASA DINNER</strong></td>
</tr>
</tbody>
</table>
Thursday - September 26, 2019

Department of Economics and Management - C.da S. Chiara, 50

08.30–09.15  REGISTRATION OF PARTECIPANTS

09.15–09.30  WELCOME ADDRESSES

09.30 – 10.30  PLENARY SESSION ORGANIZED BY ISTAT & USCI

I NUOVI SISTEMI DI PRODUZIONE ISTAT: CENSIMENTO, INDAGINI E REGISTRI

Auditorium  Pag. 3

Chairman: Marco Trentini (USCI)  Speaker: Vittoria Buratta (ISTAT)

• Il nuovo modello di produzione delle statistiche ufficiali

10.30 – 11.00  COFFEE BREAK

11.00 – 12.00  IASA SESSION ON STATISTICS APPLIED TO SENSORY ANALYSIS (INVITED)  Room B1  Pag. 12

Chairman: Bonizzi

• A pre-post sensory experiment on the effect of a seminar on olive oil preferences of Italian consumers
  Pomarici & Piscitelli & Fabbri & Sacchi

• The sensory sustainability of wine
  Odello & Brentari & Braceschi

• The sensory classification of coffee using Coffee Sensory Style
  Odello & Braceschi

11.00 – 12.00  STATISTICS APPLIED TO STUDY AND WORK HEALTH (CONTRIBUTED)  Room B2  Pag. 15

Chairman: Bini

• The effects of attitude toward statistics and math knowledge on statistical anxiety
  A path model approach
  Fabbricatore & Galluccio & Davino & Pacella & Vistocco & Palumbo

• The studyholism comprehensive model: a Bayesian reanalysis
  Stefanini & Loscalzo

• Investigating well-being at work via composite indicators
  Capecchi & Cappelli & Curtarelli & Di Iorio

11.00 – 12.00  ADVANCED METHODS IN BIOSTATISTICS (INVITED)  Room B3  Pag. 18

Chairman: Calza

• Statistical analysis of copy number alteration from next-generation sequence data
  Gusnanto

• From a black box of regression trees ensemble to a reliable Representative Tree
  Vezzoli & Sandri & Renzetti & Calza

• Modeling the joint effect of intensity and duration of alcohol drinking with bivariate spline models
  Di Credico & Polesel & Dal Maso & La Vecchia & Pauli & Torelli & Edefonti

• Potential impact fraction for a continuous risk factor: assessing the burden of oral and pharyngeal cancer according to the adherence to the healthy eating index
  Di Maso & Tomaino & Ferraroni & La Vecchia & Edefonti & Bravi
11.00 – 12.00  STATISTICS APPLIED TO CULTURAL PLACES OF THE BRESCIA TERRITORY (INVITED)  ROOM B4  PG. 22

**Chairman:** Karadjov  
- The information of the Web for local administrations  
  *Chiesa & Palamenghi*  
- The mobile phone big data tell the story of the impact of Christo’s The Floating Piers on the Lake Iseo  
  *Carpita*  
- Evaluating museum visitor experience in Brescia: the Capitolium Archaeological Area and the Santa Giulia museum complex  
  *Biggi & Comune & Manisera & Scassa & Zuccolotto*  
- Tangible and multi sensory museum visiting experience: the Tosio Martinengo Gallery in Brescia  
  *Cirillo & Manisera & Palamenghi & Zanoletti & Zuccolotto*

11.00 – 12.00  STATISTICS APPLIED TO EDUCATION (CONTRIBUTED)  ROOM A1  PG. 26

**Chairman:** Petrucci  
- Issues in prior achievement adjustment for value added analysis: an application to Invalsi tests in Italian schools  
  *Arpino & Bacci & Grilli & Guetto & Rampichini*  
- Strengths, constraints and issue of progress test  
  *Antonucci & Cataldo & Crocetta & Grassia*  
- Measuring Schein’s career anchors on University students to promote self-realization and organizational well-being  
  *Martini & Galli & Tedeschi*  
- Internationality as a psychological and social trigger to participants to Erasmus+ VET international mobility  
  *Bortolato & Fabbris & Scioni*

12.00 – 13.00  PLENARY SESSION 2 ORGANIZED BY SIS & ASA  AUDITORIUM  PG. 4

**Chairman:** Matilde Bini (SIS)  
**Speakers:** Robert Alexander (IBM)  
Andrea Mannini (BioRobotics Institute)  
- La guerra tra cervello e pensiero non ha senso. Verso un futuro più intelligente  
- Potenzialità dell’intelligenza artificiale come strumento di pre-screening di massa

13.00 – 14.00  LUNCH

14.00 – 15.00  SPECIAL EVENT ORGANIZED BY IASA: EXPERIMENT WITH CHOCOLATE SENSORY MAPS

15.00 – 16.00  STATISTICS APPLIED TO SPORTS (INVITED)  ROOM B1  PG. 30

**Chairman:** Manisera  
- Three is a magic number: Evidence on the three-point rule application to Italian soccer league  
  *Gallo & Alfano & Gaeta & Rotondo*  
- Evaluating karate performances using a wearable wireless inertial measurement unit  
  *Di Zio & Fontanella & Ippoliti*  
- Application of multivariate statistics in sports: Exploration of recall and recognition of UEFA Champions league sponsors  
  *Maricic*
15.00 – 16.00  MACHINE LEARNING AND DATA MINING IN SENSORY ANALYSIS (INVITED)  ROOM B2  PAG. 33

Chairman: Palumbo

- RERT: a novel approach for extracting a synthetic representation from an ensemble. The case study of the Italian wine quality
  *Vezzoli & Brentari*
- Prediction of wine taste preferences based on physicochemical properties and data mining models
  *Cortez*
- Automatic identification of wines via supervised text classification
  *Aria & Misuraca & Sorrentino & Spano*

15.00 – 16.00  DECISION SUPPORT DATA: THE ROLE OF DATA IN POLICY DEVELOPMENT (INVITED)  ROOM B3  PAG. 36

Chairman: Simonetto

- Data and indicators to support social policies
  *Scalvini*
- Impact assessment and wellbeing indexes in social enterprises: some real cases
  *Depedri*
- Using Data for improvement. The case of Cooperativa di Bessimo, social enterprise in addiction treatment
  *Chief*

15.00 – 16.00  STATISTICS APPLIED TO FOOD CHOICES AND BEHAVIOURS (CONTRIBUTED)  ROOM B4  PAG. 39

Chairman: Piscitelli

- Impact of changes in the price of meals at university canteens on students’ eating behaviours
  *Lorenzoni & Masserini & Bini*
- The blockchain for the certification of the dairy supply chain, the “Lucanum” basket and the bakery products for well-being
  *Notarnicola & Santarcangelo & Martulli*
- Family lifestyle habits: what is passed down from adults to children?
  *Borgia & Castellari & Sckokai*

15.00 – 16.00  INTELLIGENT SYSTEMS AND STATISTICS APPLIED TO HEALTH (CONTRIBUTED)  ROOM A1  PAG. 42

Chairman: Di Battista

- Intelligent systems to support patients
  *Santarcangelo & Massa & Sinitò & Scavone*
- Intelligent device and method for a better electroneurographic examination
  *Gallicchio*
- The determinants of vaccination behaviour of general practitioners in South Tyrol: differences and similarities between Italian and German respondents
  *Cavrini & Lazzerini*

16.00 – 17.00  STATISTICAL MODELS APPLIED TO SOCCER PREDICTIONS (INVITED)  ROOM B1  PAG. 45

Chairman: Carpita

- Balancing multi-class response variable’s classes into a training dataset through SCUT function
  *Lucas & Bueno*
- Exploring the statistical structure of soccer team performance variables using the PCovR
  *Carpita & Ciavolino & Pasca*
- Comparing statistical models and machine learning algorithms in predicting football outcomes
  *Egidi & Torelli*
### 16.00 – 17.00  STATISTICS APPLIED TO FOODS RESEARCH (INVITED)  
**Room B2  Pag. 48**

**Chairman:** Vezzoli  
- Segmenting consumers through the dietary identity questionnaire - the case of Modern Food Habits  
  *Verneau & Amato & Coppola*  
- Validation of a food insecurity scale through structural equation models  
  *Grimaccia & Naccarato*  
- Perception of the food quality in children: a compared approach between Bayesian Network and Structural Equation Model  
  *Simonetto & Golia & Malo & Gilioli*

### 16.00 – 17.00  STATISTICS APPLIED TO SOCIAL WELL-BEING (CONTRIBUTED)  
**Room B3  Pag. 51**

**Chairman:** Brentari  
- City Prosperity Index: A comparative analysis of Latin American and Mediterranean cities based on well-being and social inclusion features  
  *Surian & Sciandra*  
- Assessing mental health therapeutic communities functioning  
  *Parroco & Mancuso & Genova & Giannone*  
- Cyberbullying: A threat for relationships and social health  
  *D'Uggento & Grattagliano & Ribeco & Toma*

### 16.00 – 17.00  TEXT MINING AND SENTIMENT ANALYSES (CONTRIBUTED)  
**Room B4  Pag. 54**

**Chairman:** Zenga  
- A functional data analysis of Google Trends on health and wellness  
  *Fortuna & Caruso & Di Battista*  
- Emotional text mining and health psychology: the culture of organ donation in Spain  
  *Greco & Monaco & Di Trani & Cordella*  
- Overhearing Italian subjective well-being on Twitter  
  *Iacus & Porro & Salini & Siletti*  
- Quantity and mood of final open-ended comments on an Erasmus+ VET mobility questionnaire  
  *Vivian & Bortolato & Fabbris*

### 16.00 – 17.00  STATISTICS AND TOURISM (INVITED)  
**Room A1  Pag. 58**

**Chairman:** Giusti  
- Museum preferences analysis: an item response model applied to network data  
  *Bacci & Bertaccini & Petrucci*  
- A multi-inflated hurdle regression model for the total number of overnight stays of Italian tourists in the years of the economic recession  
  *Bocci & Grassini & Rocco*  
- Ecotourism and food geographic areas  
  *Antolini & Truglia*  
- Computing ordinal consistency thresholds for pairwise comparison matrices  
  *Amenta & Lucadamo & Marcarelli*

### 17.00 – 17.30  COFFEE BREAK

### 18.00 – 19.30  SPECIAL EVENT ORGANIZED BY DMS STATLAB: VISIT TO THE CAPITOLIUM

### 20.30 – 23.00  ASA SOCIAL DINNER  
**Trattoria urbana il Mangiafuoco**
Friday – September 27, 2019
Department of Economics and Management - C.da S. Chiara, 50

09.30 – 10.30  PLENARY SESSION 1 ORGANIZED BY AICQ-CN & AISS  AUDITORIUM  PAG. 6

SIX SIGMA AS A METHOD FOR IMPROVEMENT IN HEALTH FACILITIES AND IN THE MEDICAL AND PHARMACEUTICAL INDUSTRY

Chairman: Alessandro Celegato  Speakers: Alessandro Celegato, Luigi Mosca, Michela Palmas

- Six Sigma as a method for improvement in health facilities and in the medical and pharmaceutical industry  
  Celegato
- Lean six sigma as a method for improving the healthcare and pharmaceutical processes  
  Mosca & Nutta & Palmas & Celegato
- The FMEA as a patient risk management tool  
  Nutta & Palmas & Mosca & Celegato
- Communication of the AICQ-CN announcement for best paper/project defined by AICQ-CN  
  (In memory of Eng. Egidio Cascini)  
  Celegato

10.30 – 11.00  COFFEE BREAK

11.00 – 12.00  STATISTICAL METHODS FOR PERFORMANCE EVALUATION IN SPORT FRAMEWORK (INVITED)  ROOM B1  PAG. 62

Chairman: D’Ambra

- Three-way log-ratio analysis for assessing sport performance  
  Lombardo & Camminatiello & D’Ambra
- A composite indicator via hierarchical disjoint factor analysis for measuring the Italian football teams’ performances  
  Cavicchia & Sarnacchiaro & Vichi
- Analysis of the financial performance in Italian football championship clubs via longitudinal count data and diagnostic test  
  Crisci & D’Ambra

11.00 – 12.00  METHODS AND MODELS FOR ORDINAL CATEGORICAL DATA APPLIED TO HEALTH & WELL-BEING (INVITED)  ROOM B2  PAG. 65

Chairman: Iannario

- Another look at the relationship between perceived well-being and income satisfaction  
  Paccagnella & Zanin
- Profile patterns of Italians NEET by nonlinear PCA  
  Parola & Palumbo
- A mixture model with discrete variables for depression diagnosis in infertile couples  
  Iannario & Vistocco & Zurlo

11.00 – 12.00  STATISTICS AND ERRITORIAL INEQUALITIES OF WELL-BEING AND SUSTAINABILITY (INVITED)  ROOM B3  PAG. 68

Chairman: Davino

- Quantile Composite-based path modelling to handle differences in territorial well-being  
  Davino & Dolce & Taralli & Vistocco
- Composite indicators of gender equality in Italy at regional level  
  Di Bella
- A composite indicator to stratify old people resident in Piedmont according to their frailty level  
  Silan & Bocuzzo & Signorin
### DEVELOPMENTS IN DATA SCIENCE (INVITED)

**Chairman:** Mariani  
- Missing values in social media: an application on Twitter data  
  Mariani & Marletta & Missineo  
- Harmonised administrative databases: a new approach in the era of big data  
  Nicolardi & Marini  
- Inferring Twitter users home location based on trend topics  
  Zola & Rago & Cortez  
- Invariance in the structural topic models  
  Zavarrone & Grassia & Mazza

### STATISTICAL METHODS AND MODELS APPLIED TO HEALTH (CONTRIBUTED)

**Chairman:** Golia  
- Elderly with and without children: Do they report different health conditions?  
  Bocci & Salvini  
- Short-run and long-run persistence of bad health among elderly  
  Mendola & Li Donni  
- Selecting features for machine learning in Alzheimer’s diagnostics  
  Lucas & Bueno & Lucas

### STATISTICS MEETS SPORTS

**Chairman:** Marica Manisera  
**Speaker:** Christophe Ley  
- Statistics meets Sports: When figures are more than numbers

### HEALTH AND WELFARE MEASURES (INVITED)

**Chairman:** Marletta  
- Dynamic of efficiency data: an analysis of Italian geriatric wards  
  Mazzoleni & Zenga  
- Spatial inequality and urban poverty: Evidence from U.S. metropolitan areas  
  Mussini & Andreoli & Prete  
- Socioeconomic inequalities and cancer risk: the challenges and opportunities of worldwide epidemiological data consortia  
  Galeone & Bonzia & Pelucchi & Turati & La Vecchia  
- A data analytics framework: medical prescription pattern dynamics  
  Giordani & Arosio & Battiston & Archetti
14.30 – 15.30  STATISTICS APPLIED TO WELL-BEING (CONTRIBUTED)  ROOM B3  PAG. 85

**Chairman:** Fabbris  
- Applying network modelling to uncover the relationships among well-being dimensions  
  Giuntoli & Vidotto  
- Evaluating health performance and inequalities in Marche region of Italy  
  Franci & Renzi  
- The determinants of physical inactivity throughout the analysis of well-being and sustainability indicators: the Italian BES framework  
  Murianni & De Carli & Bologna & Iannucci & Tinto

14.30 – 15.30  STATISTICS APPLIED TO SOCIAL ISSUES (CONTRIBUTED)  ROOM B4  PAG. 88

**Chairman:** Frosini  
- Household wealth and income in Italy: Analysis by locally weighted quantile regression  
  Amerise & Tarsitano  
- Understanding local administrations policies effects on well-being in Italian inner areas  
  Romagnoli & Mastronardi  
- Family and personal determinants of gambling risk among Italian adolescent students  
  Fabbris & Andreotti & Genetti & Vian & Mortali & Mastrobattista & Minutillo & Pacifici

14.30 – 15.30  STATISTICS APPLIED TO HEALTH CARE AND SUSTAINABILITY (CONTRIBUTED)  ROOM A1  PAG. 91

**Chairman:** Vistocco  
- Does industry change affect strategic, governance, and financial configurations of private hospital providers. A survey of Italian private healthcare organizations  
  Aria & Cuccurullo  
- Demand vs estimated burden of health care: A comparative evaluation based on spatial analysis  
  Alfonzetti & Rizzi & Grassetti & Gobbato  
- Partial Least Squares - Path Modeling approach for sustainability using qualitative information  
  Cataldo & Grassia & Marino

15.30 – 16.30  SPECIAL EVENT ORGANIZED BY BDSPORTS & ASA  AUDITORIUM  PAG. 2

**Kick-off meeting Special Interest Group ISI in Sports Statistics**  
**Promoters:** Marica Manisera & Paola Zuccolotto  
- After a brief presentation of the new ISI-SIG in Sport Statistics, the members (also in international videoconference) will introduce their main research interests and ideas about possible research cooperation as well as initiatives and events to organize in the future.

16.30 – 17.00  CONFERENCE CLOSING AND DRINK TO THE HEALTH
incontro pubblico sul tema

Scenari statistici su popolazione salute e benessere in Italia

RELAZIONE DEL PRESIDENTE ISTAT

Prof. Gian Carlo Blangiardo

Brescia, 25 Settembre 2019 ore 15

Spazio MO.CA - Centro per le nuove Culture, Via Moretto 78

Nell’ambito delle iniziative promosse per il Convegno Nazionale “Statistica per la Salute e il Benessere” (www.sa-ijas.org/statistics-for-health-and-well-being/), l’Associazione per la Statistica Applicata e l’Istituto Nazionale di Statistica, con il supporto della Società Italiana di Statistica e dell’Università degli Studi di Brescia, e con il patrocinio del Comune di Brescia, organizzano presso lo spazio MO.CA (www.spaziomoca.com) l’incontro pubblico sul tema “Scenari statistici su popolazione, salute e benessere in Italia”:

Come emerge con chiarezza dall’ultimo rapporto annuale dell’ISTAT, il quadro demografico e sociale italiano è caratterizzato da un forte invecchiamento della popolazione, risultato di una marcata crescita della sopravvivenza e da un altrettanto marcato calo della natalità, ma anche da un miglioramento delle condizioni di salute e delle condizioni di vita nelle età anziane mentre si fanno più critiche le condizioni delle famiglie giovani e in generale le condizioni socio-economiche dei giovani.

L’intervento del Presidente dell’ISTAT fornirà un contributo per la lettura dello scenario attuale, delle prevedibili tendenze nel medio periodo e della sostenibilità per il sistema Paese.

Seguirà il dibattito moderato da Anna Della Moretta del Giornale di Brescia con gli interventi di:

• Emilio Del Bono, Sindaco del Comune di Brescia
• Filomena Maggino, Università di Roma La Sapienza, Consigliere e rappresentante del Presidente del Consiglio dei Ministri per la Cabina di regia "Benessere Italia"
• Clelia di Serio, Università Vita-Salute San Raffaele e Presidente Società Italiana di Biometria
• Maurizio Tira, Magnifico Rettore dell’Università degli Studi di Brescia

La partecipazione è libera e ci si può prenotare inviando un’email a: dms.statlab@unibs.it
Kick-off meeting Special Interest Group ISI in Sports Statistics

Marica Manisera, Paola Zuccolotto (promoters)
Department of Economics and Management, University of Brescia, Italy
BDsports - bodai.unibs.it/bdsports/, Italy

The Special Interest Group in Sports Statistics of the International Statistical Institute (isi-web.org/index.php/news-from-isii/127-isii-sports-statistics-committee, SIG on Sport) has been recently revitalized. The aim of this Special Interest Group is to promote the understanding, development and good practice of sports statistics worldwide. The kick-off meeting of the SIG on Sports Statistics will take place at the Department of Economics and Management of the University of Brescia, with the participation of members from several countries even through video conference.

It is a Satellite Event of the ASA 2019 conference on "Statistics for Health and Well-being" (www.sa-ijas.org/statistics-for-health-and-well-being/), organized in collaboration with the “Big Data analytics in sports” network BDsports (bdsports.unibs.it), a research project developed within the activities of the Big&Open Data Innovation Laboratory at the University of Brescia, Italy.

After a brief presentation of the SIG and its activities, the SIG members will introduce their main research interests and ideas about possible research cooperation as well as initiatives and events to organize in the future within the activities of the SIG.
Il nuovo modello di produzione delle statistiche ufficiali

Marco Trentini (chair)
Comune di Brescia; Unione Statistica Comuni Italiani - www.usci.it.

Vittoria Buratta (speaker)
Istituto Nazionale di Statistica, Roma, Italia

Il modello di produzione delle statistiche ufficiali dell’Istat è in profonda trasformazione, muovendo da un assetto basato sostanzialmente sulle singole fonti ad uno basato sulla integrazione dei dati di indagine e di quelli degli archivi amministrativi e statistici.
Il nostro Istituto ha fatto negli anni recenti importanti progressi verso le nuove esigenze di informazione statistica cercando al tempo stesso di ridurre l’onere sui rispondenti e ottimizzando i sistemi di produzione.

Alcuni di questi miglioramenti sono stati la conseguenza di nuove modalità di raccolta dati, con minore invasività e “molestia” da parte della statistica ufficiale, altri sono scaturiti dall’integrazioni di diverse modalità di raccolta. Tuttavia il progresso è avvenuto in maniera disomogenea tra i vari settori produttivi e con inefficienze di varia natura: dalla duplicazione delle iniziative a diversità di approccio e di realizzazione. L’organizzazione strutturale e quella infrastrutturale non avevano ancora tratto i benefici possibili dalle molte innovazioni introdotte.

Il recente processo di modernizzazione ha portato invece ad una reale e pervasiva innovazione di sistema. Cosa ha significato concretamente questo? L’adozione di un linguaggio comune (classificazioni, definizioni,… metadati) tra tutti i processi, l’istituzione di un comune e unitario sistemi di raccolta dati, l’impostazione di comuni e condivisi metodi di trattamento dati, un comune ambiente di rilascio e diffusione. Questo percorso di modernizzazione è principalmente sostenuto da fattori interni, come la necessità di ottimizzare e modificare i sistemi di produzione, e da fattori esterni in primo luogo la rivoluzione delle nuove tecnologie che ha riscritto le regole della produzione e della comunicazione.

I tradizionali modelli di rilevazione basati sull’acquisizione diretta dei dati dalle fonti (cittadini e imprese) attraverso le rilevazioni e i censimenti sono stati messi in discussione anche per l’impatto, a volte troppo invasivo, sui rispondenti, che influenza l’abbassamento dei tassi di risposta. È emersa, pertanto, la spinta a disegnare e a utilizzare registri statistici, essenzialmente derivati dalla combinazione delle fonti (amministrative e statistiche) con meccanismi di alimentazione nel continuo basati quasi esclusivamente su flussi telematici.

L’introduzione del sistema dei registri che integra e gestisce in un unico ambiente di produzione dati provenienti da una molteplicità di fonti ha configurato, di fatto, un nuovo frame di riferimento per l’informazione statistica. In questo nuovo contesto il censimento permanente della popolazione rappresenta una straordinaria innovazione ed è pienamente integrato con le principali indagini sociali sia sotto il profilo del disegno d’indagine sia sotto il profilo delle definizioni e classificazioni. Il nuovo assetto è coerente con il regolamento europeo sulle statistiche sociali IESS appena adottato.

Keywords: Modernizzazione, Registri statistici, Censimento, Indagini sociali

References
We are now at the peak of the technological age and we can no longer do anything without automated or robotic devices. We use technology - referring to robotics, synthetic biology, computational science, nanotechnology, quantum computing, 3D and 4D printing, Internet of Things, cognitive science, semantic web, human intelligence augmentation, blockchain, self-driving vehicles, conscious-technology, and synergies - as a tool of connection around the world and as a maker of the well-being of human nature. So artificial intelligence helps us in the medical field, just think about the laboratories composed of machines that replace the work of man or tools that help him to survive; or in the industry where it is used to speed up processing times, but nowadays, never as before, it became a fundamental tool for the progress in many fields (engineering, physics, biology, agriculture, human sciences and so on). This means that it is became a common denominator of all scientific disciplines. All this leads to an improvement in the quality of life, in fact in daily life it leads towards a perspective of development of the standard of living. Now artificial intelligence, although sometimes is very expensive, contributes to our well-being and health, nevertheless, we cannot completely replace it to men, but rather we can find a right balance between its capabilities and those of human mind.

**The war between brain and thought doesn't make sense.**

*Towards a smarter future*

Robert J. Alexander
Health & Life Science Integrated Industry Unit, IBM, Milan, Italy.

The perception around use and benefit of Artificial Intelligence for research, health and other applications is often driven by sensationalistic bipolar assessments: on one side it is seen as a savior on the other as a harbinger of a dystopian future. Some of this misinformation derives from both AI and biological and medical practices to be very complex and wide and that subject matters on one side often misunderstand the other. We will try to clarify the main topics around AI and health and give a few indications on attention points and useful applications.

**References**


Six Sigma as a method for improvement in health facilities and in the medical and pharmaceutical industry

Celegato Alessandro (chair)
Associazione Italiana Cultura Qualità Centronord - Milano
Accademia Italiana del Sei Sigma - Prato

Six Sigma, through the culture of “Statistical Thinking” and in conjunction with the study and analysis of business processes, makes it possible to achieve reduction and control in the variability of processes and product performance. The works that will be presented will provide the reference model and an application case.

Sei sigma come metodo per il miglioramento nelle strutture sanitarie e nella industria medicale e farmaceutica

Celegato Alessandro (chair)
Associazione Italiana Cultura Qualità Centronord - Milano
Accademia Italiana del Sei Sigma - Prato

Lean six sigma as a method for improving the healthcare and pharmaceutical processes

Luigi Mosca\textsuperscript{a}, Ilaria Nutta\textsuperscript{a}, Michela Palmas\textsuperscript{a}, Celegato Alessandro\textsuperscript{b}

\textsuperscript{a} Associazione Italiana Cultura Qualità Centronord - Milano
\textsuperscript{b} Associazione Italiana Cultura Qualità Centronord - Milano
Accademia Italiana del Sei Sigma – Prato

The healthcare sector, the medical and pharmaceutical industries are facing since some time the challenge to implement cost reduction without compromising the quality and access to healthcare and, at the same time, assuring a continuous improvement of the processes, products and services. The quickly changing context in which these organizations operate paired with the constant reference they keep to the stakeholders, compel the same organizations to carry out sudden and effective adjustments to internal and external quality. The common target is to supply a “product” with an excellent qualitative level, within the expected timing and without losing or underestimate the involved risk factors.

The present work presents Lean Six Sigma methodology which stems from the combination of two separate methods: Six Sigma and Lean Thinking approach. These methods have been developed and applied initially in the industrial environment and later they found application in different contexts, especially in healthcare, medical and pharmaceutical industries.

Six Sigma is a rigorous approach where the study is based on data analysis, developed with DMAIC methodology (Define-Measure-Analyze-Improve-Control). Lean Thinking approach focuses on processes organization, with the aim to streamline them by removing the waste upon a thorough evaluation of the workflow.

Lean Six Sigma incorporates the strengths of both techniques (Six Sigma and Lean Thinking), representing one of the most advanced methods aimed to problem solving in a framework of quality improvement with special care to efficiency and efficacy of processes, products development and services offering.

The work, by describing Lean Six Sigma methodology, shows same applications within healthcare and pharmaceutical processes.

Keywords: Sei sigma, Six sigma, Lean six sigma, Dmaic, Problem solving

References:
Norma UNI ISO 13053-1:2014, Metodi quantitativi per il miglioramento dei processi - Sei Sigma - Parte 1: metodologia DMAIC.
Lean six sigma come metodo per il miglioramento dei processi sanitari e farmaceutici

Luigi Mosca\textsuperscript{a}, Ilaria Nutta\textsuperscript{a}, Michela Palmas\textsuperscript{a}, Celegato Alessandro\textsuperscript{b}

\textsuperscript{a} Associazione Italiana Cultura Qualità Centronord - Milano
\textsuperscript{b} Associazione Italiana Cultura Qualità Centronord - Milano
Accademia Italiana del Sei Sigma – Prato

Il settore sanitario, l'industria nell'ambito medicale e farmaceutico, si trovano da diverso tempo ad affrontare la sfida della riduzione dei costi garantendo il diritto alla salute, e nello stesso tempo, assicurando un miglioramento continuo dei processi, dei prodotti e dei servizi offerti.

Queste organizzazioni devono affrontare costantemente uno scenario in cui il presente che le circonda muta velocemente, senza mai dimenticare gli stakeholder, obbligando le stesse organizzazioni ad un repentino adattamento nella qualità sia interna che esterna all'organizzazione stessa.

Per tutti l’obiettivo primario è quello di fornire un “prodotto” con un livello qualitativo eccellente, in tempo senza mai dimenticare o sottovalutare i fattori di rischio connessi.

Il lavoro presenta la metodologia Lean Six Sigma, che nasce dalla sinergia tra due metodi: il Six Sigma e l’approccio Lean Thinking.

Tali metodi sono stati sviluppati e applicati in ambito industriale, e successivamente hanno trovato una diffusione in diversi contesti; in particolare nel campo sanitario e dell’industria medica e farmaceutica.

Il Six Sigma è un approccio rigoroso in cui lo studio si basa sulla analisi dei dati, che viene sviluppata con la metodologia DMAIC (Define-Measure-Analyze-Improve-Control).

La metodologia Lean Thinking si focalizza sull’organizzazione dei processi, con l’obiettivo di renderli snelli eliminando gli sprechi e valutandone il flusso di lavoro.

Il metodo Lean Six Sigma recepisce i punti di forza delle due tecniche (Six Sigma e Lean) e rappresenta uno dei metodi più evoluti per il problem solving orientato al miglioramento della qualità in ottica di efficienza e di efficacia, dei processi, dei prodotti e dei servizi.

Il lavoro nel descrivere la metodologia Lean Six Sigma, ne presenta delle applicazioni nell’ambito del healthcare e pharmaceutical processes.

**Keywords:** Sei sigma, Six sigma, Lean six sigma, Dmaic, Problem solving

**Riferimenti bibliografici:**


Norma UNI ISO 13053-1:2014, Metodi quantitativi per il miglioramento dei processi - Sei Sigma - Parte 1: metodologia DMAIC.

The FMEA as a patient risk management tool

Ilaria Nutta\textsuperscript{a} e Michela Palmas\textsuperscript{a}, Luigi Mosca\textsuperscript{a}, Celegato Alessandro\textsuperscript{b}

\textsuperscript{a} Associazione Italiana Cultura Qualità Centronord - Milano
\textsuperscript{b} Associazione Italiana Cultura Qualità Centronord - Milano

Accademia Italiana del Sei Sigma – Prato

Developed at first in aerospace industry, particularly in the project development of Apollo missions, this method has been later adopted by military industry, followed by automotive and then by industry at large. At last it has been adopted by Service business and particularly healthcare context.

FMEA (Failure Mode and Effect Analysis) is employed in Lean Six Sigma methodology to identify defects (more broadly unexpected outputs) and to prevent the potential causes which could intervene in the different process or design phases, through the implementation of improvement/mitigation.

The study is composed by a set of activities aimed to detect a potential failure and to highlight the effects which could create, finally resulting in an associated risk index for the patient. This can be achieved by mean of a qualitative or quantitative approach, by mean of characterizing data and variables analysis. The work presents the method and an application within medical industry.

**Keywords:** Sei sigma, Six sigma, Lean six sigma, Fmea, Fmeca, Risk management

**References:**


norma CEI EN IEC 60812:2018, Analisi dei modi e degli effetti di guasto (FMEA e FMECA).
La fmea come strumento per la gestione del rischio per il paziente

Ilaria Nutta\textsuperscript{a} e Michela Palmas\textsuperscript{a}, Luigi Mosca\textsuperscript{a}, Celestino Alessandro\textsuperscript{b}

\textsuperscript{a} Associazione Italiana Cultura Qualità Centronord - Milano
\textsuperscript{b} Associazione Italiana Cultura Qualità Centronord - Milano

Accademia Italiana del Sei Sigma – Prato

Sviluppata ed utilizzata per la prima volta dall’industria aerospaziale nello sviluppo del progetto per le missioni Apollo, è stata successivamente adottata dall’industria militare poi dall’industria dell’automobile e in seguito dall’industria in generale per raggiungere infine i servizi in particolare in ambito della sanità.

La FMEA (Failure Mode and Effect Analysis) è utilizzata nella metodologia Lean Six Sigma per identificare difetti (o più in generale output non attesi) e prevenirne le potenziali cause che possono intervenire nelle varie fasi del processo o di una progettazione, attraverso l’implementazione di azioni di miglioramento / mitigazione.

Lo studio si articola in un insieme di attività orientate ad individuare un guasto potenziale e gli effetti che può generare con risultante un indice di rischio associato per il paziente.

Lo studio può essere effettuato in modo qualitativo oppure quantitativo attraverso l’analisi dei dati e delle variabili che lo caratterizzano.

Il lavoro presenta sia il metodo che una applicazione nell’ambito dell’industria medicale.

Keywords: Sei sigma, Six sigma, Lean six sigma, Fmea, Fmeca, Risk management

Riferimenti bibliografici:
norma CEI EN IEC 60812:2018, Analisi dei modi
In the past few decades, the application of statistics to sports has rapidly gained interest, as documented by the publication of a wide variety of scientific articles and some insightful collections of statistical analyses applied to data from several sports. The plenary talk by Christophe Ley will summarize interesting contributions and give an overview of possible directions of future research in this field.

**Statistics meets sports:**
*When figures are more than numbers*

Christophe Ley  
Department of Applied Mathematics, Computer Science and Statistics  
Ghent University, Gent, Belgium

The combination of advanced statistical methods with sports, be it under the name Sport Statistics, Sport Analytics, Machine Learning or Artificial Intelligence methods in Sports, has been very fruitful over the past decade. More and more sport clubs have recognised this importance and hired data scientists, with concrete success. In this talk I will give an overview of recent advances in this fertile research field (with principal focus on soccer), present modern challenges and conclude with an outlook on the future of sport statistics.

**Keywords:** Injury Prevention, Machine Learning, Match Prediction, Statistical Modeling, Tournament evaluation
A pre-post sensory experiment on the effect of a seminar on olive oil preferences of Italian consumers

Eugenio Pomarici\textsuperscript{a}, Alfonso Piscitelli\textsuperscript{b}, Luigi Fabbris\textsuperscript{c} and Raffaele Sacchi\textsuperscript{d}

\textsuperscript{a} Department of Land, Environment, Agriculture and Forestry (TeSAF), University of Padua, Italy;
\textsuperscript{b} Department of Political Sciences, University of Naples “Federico II”, Naples, Italy;
\textsuperscript{c} Department of Statistical Sciences, University of Padua, Padua, Italy;
\textsuperscript{d} Department of Agricultural Sciences, University of Naples “Federico II”, Naples, Italy;

The relative importance for consumers of different attributes of olive oil helps them to define a hypothetical product more preferred (Van der Lans \textit{et al.}, 2001, Krystallis and Ness 2005; Mtimet \textit{et al.}, 2013). Physical characteristics such as colour, taste, and flavour play an important role in consumers’ perception (Grunert, 1997).

This study attempts to measure the effect of the extra information on extra-virgin olive oil sensory and taste aspects and the effect of the components of this information on consumers’ perceived quality of extra-virgin olive oil (EVOO).

To this end, a two-time sensory evaluation experiment was conducted on five different EVOO. Two-hundred forty-six assessors were randomly assigned to the experimental group (n=117) and control group (n=129); for each tasted EVOO, the assessors had to express a preferential judgment by stating a value in a nine-point scale in which one means “Awful” oil; nine means “Excellent” oil. The first and second tasting experiments were conducted by following a double-blind control procedure. After the first session of taste, the experimental group was given a special training course on smells and tastes in extra-virgin olive oil.

The sensory evaluation of EVOO is used in order to define a procedure to draw a profile of consumers, their discriminatory capability and to understand the impact of the training course on consumers’ preference. Furthermore, the tasting experiment conducted draws the conclusion that consumers chose EVOO according to easy-to-perceive (that is, non-technical), general-type attributes.

\textbf{Keywords:} Consumers’ preference, EVOO attributes, Sensory evaluation experiment

\textbf{References:}


The sensory sustainability of wine

Luigi Odello\textsuperscript{a}, Eugenio Brentari\textsuperscript{b}, Gian Paolo Braceschi\textsuperscript{a}

\textsuperscript{a} Centro Studi Assaggiatori - www.assaggiatori.com, Italy;
\textsuperscript{b} Department of Economics and Management, University of Brescia, Brescia, Italy;

Is the sensory model that we are currently following in the making of our wines sustainable? Or should we think about redefining it to ensure that we have a successful future?

In our culture, there is no one that does not know what wine is, but this is no longer true when we talk about sensory sustainability. Sustainability is the characteristic of a process or a state that can be maintained indefinitely - or in any case for a long time - and sensoriality is the complex mechanism that allows us to perceive any reality. A wine is therefore sensory sustainable:

- when it gives us pleasure;
- when the pleasure can be prolonged because you can drink it in adequate volumes, without suffering from alcohol alteration;
- when the pleasure is increased by the characteristics of the land in the glass.

Fashions are against sensory sustainability, because they generate models that are often unsustainable and that make the sensory profiles of wines try to homologate with the consequent loss of the characteristics of the land.

\textbf{Keywords:} Sensory sustainability, Wine.
The sensory classification of coffee using
Coffee Sensory Style

Luigi Odello\textsuperscript{a}, Gian Paolo braceschi\textsuperscript{a}

\textsuperscript{a} Centro Studi Assaggiatori - www.assaggiatori.com, Italy;

In all sectors, defining the quality of the product becomes difficult, especially in the food industry where the quality is strongly linked to pleasure. Quality is difficult to measure because it is closely linked to expectations and personal habits. Therefore there is nothing better than generating narrative models understandable by consumers so that the final user can easily orient himself in the rich offer on the market.

The world of coffee, particularly espresso, is very varied; there are many types of products with very different characteristics. We have often tried to define quality models based on sensory attributes or instrumental analysis with the help of statistical tools, but the results weren’t always what we expected.

For this reason, a different method has been chosen. It started with a database of about 600 coffee samples from around 20 countries. The International Coffee Tasters Institute has created a sensory profile for each coffee.

Considering the objective descriptors alone, an algorithm has been generated that can position the samples according to three variables: structure, aroma and body.

Coffee is then classified according to precise sensory styles easily recognizable by the consumer. These sensory styles are usable by the coffee roasters for product development and product communication.

\textbf{Keywords}: Sensory analysis, Coffee.
The effects of attitude toward Statistics and Math knowledge on Statistical anxiety: A path model approach

Rosa Fabbriacatore\textsuperscript{a}, Carla Galluccio\textsuperscript{a}, Cristina Davino\textsuperscript{b}, Daniela Pacella\textsuperscript{a}, Domenico Vistocco\textsuperscript{a}, Francesco Palumbo\textsuperscript{a}

\textsuperscript{a} Department of Political Science, Università di Napoli Federico II;
\textsuperscript{b} Department of Economics and Statistics, Università di Napoli Federico II.

Academic well-being and performance are important tasks to achieve in all school grades, and are negatively affected by stress and anxiety. Actually, students feel discomfort with regard to specific subjects, like Math and Statistics. For this reason, mathematical and statistical anxiety have been widely studied (among the others, see Primi et al., 2016). Statistics is a mandatory course in the curriculum in most of humanities programs. Many authors showed that these students consider Statistics as a burden and exhibit higher levels of statistical anxiety. Statistical anxiety can be defined as “the feeling of anxiety encountered when attending a Statistics course or doing statistical analyses” (Cruise et al., 1985, p.92). These students are made weary by anything related to Mathematics and believe that Statistics is not important for their degree programs and careers (Primi et al., 2016). Moreover, Statistics is viewed as an unpleasant and difficult subject, making students feel uncomfortable and leading them to believe that they are not able to achieve the task that is being requested from them. Several studies on this topic classified the statistical anxiety antecedents. They are typically divided in situational factors (e.g. math skills, previous statistical experience), dispositional factors (e.g. attitude toward Statistics, self-concept and self-efficacy) and demographic factors (e.g. gender, age). The most common and widely used psychometrics tools to assess statistical anxiety are STARS (Cruise et al., 1985) and SAS (Vigil-Colet et al., 2008).

In this work we present a preliminary analysis based on data collected within the ALEAS (Adaptive LEArning in Statistics) ERASMUS+ project (https://aleas-project.eu/wordpress/) about statistical anxiety in undergraduate students enrolled in the Psychology course at Federico II University of Naples. Path analysis (Duncan, 1966) was carried out to study the interplay between statistical anxiety and a set of considered variables. Our results show that math background affects the attitude towards Statistics and the statistical anxiety. Moreover, statistical anxiety also depends on other variables such as math comprehension, gender, high school final mark, and past experience in Statistics.

Keywords: Academic well-being, Statistical anxiety, Path analysis.

References:


The Studyholism Comprehensive Model:  
a Bayesian reanalysis

Federico M. Stefanini a, Yura Loscalzo b

a Department of Statistics, Computer Science, Applications, University of Florence, Italy;  
b Department of Health Sciences, University of Florence, Italy;

The construct of Studyholism has been defined in the seminal work of Loscalzo and Gian- 
nini (20171) to characterize the obsession toward study in terms of its potential antecedents and 
outcomes. In a recent work, a path analysis model has been proposed (Loscalzo and Giannini, 
2019) to include some antecedents and outcomes of both studyholism and study engagement. 
The above model was fitted to 1958 Italian college students aged between 18 and 60 years and 
quite heterogeneous about their year and major of study, as well as concerning the city in which 
they attended their courses (details in cited works).

The above studies bear important implications not only because Studyholism is theoretically 
declared as an OCD-related disorder, but also for potential insights connected to preventive and 
clinical practice. For example, in a clinical setting it is important to distinguish between dis- 
engaged studyholics and engaged studyholics: although they both have functional impairment 
this is located in different areas.

Factor and path analysis models have been questioned from a statistical point of view: (i) 
the multivariate normal framework could be a coarse approximation, even after summing related 
ordinal variables; (ii) statistical tests based on the multivariate normal family of distributions do 
not closely match the nature of underlying variables.

In this work, we took the ordinal nature of most qualitative variables seriously while imple- 
menting a Bayesian Structural Equation Model. Model fitting was performed by Markov Chain 
Monte Carlo simulation. We closed the discussion of results from the Bayesian reanalysis by 
emphasizing the distinctive achievements as compared to previous work.

Keywords: Study addiction; Workaholism; Bayesian SEM; Ordinal manifest variables;

References

for a possible new clinical condition, in Advances in Psychology Research, A. M. Columbus 
Loscalzo, Y., Giannini, M. (2019). Heavy Study Investment in Italian College Students. An 
Analysis of Loscalzo and Giannini’s (2017) Studyholism Comprehensive Model. Frontiers 

1The dataset considered in this work has been discussed in the following work (in press): 
Investigating well-being at work via composite indicators

Stefania Capecchi\textsuperscript{a}, Carmela Cappelli\textsuperscript{a}, Maurizio Curtarelli\textsuperscript{b}, Francesca Di Iorio\textsuperscript{a}

\textsuperscript{a} Department of Political Sciences, University of Naples Federico II, Naples, Italy;
\textsuperscript{b} Prevention and Research Unit, EU-OSHA (European Agency for Safety and Health at Work), Bilbao, Spain

The paper investigates the effect on workers’ well-being of selected risk factors existing in the workplace. Following the framework proposed by EU-OSHA (2013), the two categories of risk factors under investigation will be on the one hand the physical risk factors and on the other hand the psychosocial ones. More in particular, our objective is to display how psychosocial risk factors can affect workers’ health conditions as much as the physical risk factors, using some of the evidence of the European Working Conditions Survey (Eurofound, 2017). As a proxy of workers’ well-being, the variable of interest, that is the self-assessed health (SAH), stems from question Q75: “How is your health in general? Would you say it is (…)” measured on a 5-point Likert scale (1 Very good; 2 Good; 3 Fair; 4 Bad; 5 Very bad).

Current literature, as summarized in OECD handbook (2008), emphasizes the stages to achieve effective and consistent composite indicator. In order to build a model-based composite indicator for the SAH, the methodology chosen for this paper involved a number of steps. First, due to the ordinal variable of interest, Ordered Probit analyses were run based on explanatory variables describing the physical risks at work, the psychosocial risks and some individual characteristics. Then, a Principal Components Analysis was carried out to build a composite indicator summarising the selected variables, and the resulting Principal Components (PCs) have subsequently been used as explanatory variables in further Ordered Probit analyses. Moreover, results from the latter models are compared to measure the intensity of the relationship between SAH and variables identifying physical and psychosocial risks, highlighting those more relevant in influencing SAH.

The results display that both types of risk factors do exert a significant impact on workers’ health, and in both cases the synthetic indicator (i.e. the first PC) accounts for most of the variance providing and effective synthesis of the data. When included in Ordered Probit models to measure the strength of their effect on the self-reported health, the indicators built for the two sets of risks turn out to be significant, both together and alone.

The benefit of building these synthetic indicators relies on that they allow for simplifying a model-based analysis and may help in disentangling specific drivers of work-related well-being, as long as they are actually carriers of information, with the additional advantage of removing redundant information, obtaining more robust models.

Keywords: Psychosocial risks, Working conditions, Occupational health, Composite indicators

References:


Statistical analysis of copy number alteration from next-generation sequence data

Arief Gusnanto
Department of Statistics, University of Leeds, Leeds LS2 9JT, United Kingdom

Next-generation sequencing (NGS) has transformed and improved our investigation of genomic profiles such as copy number alteration (CNA). However, there are many steps need to be taken to make sense the information contained in the sequence data. It starts from the moment we obtain mapped sequence data to the use of the data for clinical subtyping or identification of important genomic regions. In each and every steps involved, statistical modelling and analysis are needed to address those challenges (Gusnanto et al., 2012). In this talk, I will discuss and present the challenges and the statistical methods to address those challenges. This includes segmentation and normalisation to estimate CNA (Gusnanto et al., 2012), modelling strategies based on CNA profile (Gusnanto et al., 2015), and developing new statistical test per genomic region (Telford et al., 2019).

Keywords: Next-generation sequencing, copy number alteration, segmentation, normalisation, prediction

References:
From a black box of regression trees ensemble to a reliable Representative Tree

Marika Vezzoli, Marco Sandri, Stefano Renzetti, Stefano Calza

Department of Molecular and Translational Medicine, Unit of Biostatistics, University of Brescia, Brescia, Italy

DMS StatLab, University of Brescia, Brescia, Italy

Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

How to choose the best regression tree from a “forest” of models obtained by an ensemble method? A possible answer to this question comes from a recent algorithm proposed in Vezzoli et al. (2017) called RERT (REpresentative Regression Tree). In fact, it looks inside the blackbox generated by an ensemble method and identifies the best performer in order to restore the tree structure usually lost with the tree forests. In this perspective, RERT helps to solve the eternal conflict between accuracy and interpretability, celebrated in the well-known Occam’s razor problem, since it stabilizes the results across multiple replications and provides accurate predictions ($\hat{y}$) also offering a definitive model for decision making. 

Starting from RERT, we look inside the blackbox of regression trees ensemble clustering them through different measures proposed by Banerjee et al. (2012) to identify homogenous tree groups. The optimal number $K$ of clusters is defined using the silhouettes method. From each cluster, we extract a tree medoid (namely, the best group representative) and its related AUC. In order to find a single tree starting from $K$ predictions, we combine $\hat{y}_k$ weighting them through their AUCs:

$$\bar{y} = \sum_{k=1}^{K} \hat{y}_k \cdot \frac{AUC_i}{\sum_{l=1}^{K} AUC_i}.$$

Hence, $\bar{y}$ is used as outcome in a representative tree grown on the same variables used for the ensemble. The result obtained is a stable and accurate model which maintains the simplicity and interpretability of well-known CART.

We applied this novel algorithm on subsample of the dataset (164 patients) used by Ray et al. (2007). They collected archived plasma samples from individuals with presymptomatic to late-stage Alzheimer’s disease (85 subjects) and from various controls (79 subjects) and measured the abundance of 120 known signaling proteins in these samples with filter-based, arrayed sandwich ELISAs. From preliminary analysis on real data, we obtain a prediction, $\hat{y}$, which is a good competitor respect to other ensemble methods. Moreover, we restore the simple tree structure useful for decision making approach in many fields.

Keywords: RERT, Ensemble Method, Tree Medoid, Model Combination.

References:


Modeling the joint effect of intensity and duration of alcohol drinking with bivariate spline models

Gioia Di Credico\textsuperscript{a}, Jerry Polesel\textsuperscript{b}, Luigino Dal Maso\textsuperscript{b}, Carlo La Vecchia\textsuperscript{c}, Francesco Pauli\textsuperscript{a}, Nicola Torelli\textsuperscript{a}, Valeria Edefonti\textsuperscript{c}

\textsuperscript{a} Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy;
\textsuperscript{b} Cancer Epidemiology Unit, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, Italy;
\textsuperscript{c} Branch of Medical Statistics, Biometry and Epidemiology "G. A. Maccacaro”, Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milano, Italy;

When modeling the relationship between a response and some continuous covariates, an assumption of linearity may be too restrictive in many contexts. Spline functions, defined as piecewise polynomials with a fixed degree, whose joint points are called knots, offer a flexible solution to this problem (Ruppert et al., 2003). In the current paper, we will introduce and review possible approaches to spline models, to assess how to modeling in the best way a potential non-linear association between a continuous exposure and disease risk in epidemiology. In detail, we will explain why a “fewer-but-better-knots” approach may be desirable in epidemiology, as compared to traditional penalisation-based approaches. In addition, we will explore advantages of fixed and free-knots options and we will consider a free-knots approach where the maximum number of knots is fixed and all the models up to the maximum number of knots are fitted and compared. We will also introduce a novel two-step Bayesian procedure called Stochastic search variable selection, which generalizes the previous ones (O’Hara et al., 2009). In the two-step procedure we will first select the optimal number of knots considering a large, possibly overparameterised model, and then we will fit the final model on a restricted set of knots by simultaneously estimating knot locations, regression and spline coefficients. Finally, we will show how this framework is applied in a bivariate context, where the aim is to modeling the joint effect of intensity and duration of alcohol drinking on the risk of cancer of the oral cavity within the International Head and Neck Cancer Epidemiology Consortium (4839 cancer cases and 25,871 controls) (Winn et al., 2015). Calculations were carried out using the open-source Stan program (Stan, 2019) within the open-source R language.

Keywords: Bayesian analysis, Bivariate splines, Cancer of the oral cavity, Free-knots regression splines, Stochastic search variable selection approach.

References
Potential impact fraction for a continuous risk factor: assessing the burden of oral and pharyngeal cancer according to the adherence to the healthy eating index

Matteo Di Maso\textsuperscript{a}, Tomaino Laura\textsuperscript{a}, Monica Ferraroni\textsuperscript{a}, Carlo La Vecchia\textsuperscript{a}, Valeria Edefonti\textsuperscript{a}, Francesca Bravi\textsuperscript{a}

\textsuperscript{a} Department of Clinical Sciences and Community Health, Branch of Medical Statistics, Biometry and Epidemiology “G.A. Maccacaro”, University of Milan, Milan, Italy

A generalization of the attributable fraction is the potential impact fraction (PIF) defined as the proportion of disease burden that would occur if the distribution of the risk factor in the population was modified according to a counterfactual scenario (Ezzati M et al. 2003). We will estimate the PIF for oral and pharyngeal cancer attributable to healthy eating index (HEI) adherence according to 6 counterfactual scenarios. Data came from an Italian multicentre hospital-based case control study on the risk of oral cavity and pharynx cancers. Using dietary information collected by a food frequency questionnaire, we calculated the HEI for each study subject. We designed the following 6 scenarios:

1) \textit{maximum prevalence reduction} i.e., assigning the highest HEI value for all study subjects;
2) \textit{mild global prevalence intervention} i.e., 10-point increment in the HEI for all subjects;
3) \textit{strong prevalence intervention on low and lower-middle HEI subjects} i.e., 1/3 increment for subject in the 1\textsuperscript{st} and 2\textsuperscript{nd} quartile of the HEI distribution;
4) \textit{mild prevalence intervention on low and lower-middle HEI subjects} i.e., 1/3 prevalence increment for subjects in the 1\textsuperscript{st} quartile and 1/4 prevalence increment for subjects in the 2\textsuperscript{nd} quartile of the HEI distribution;
5) \textit{strong prevalence intervention on lower- and upper-middle HEI subjects} i.e., 1/3 prevalence increment for subjects in the 2\textsuperscript{nd} quartile and 1/4 increment for subjects in the 3\textsuperscript{rd} quartile in the HEI distribution;
6) \textit{mild prevalence intervention on lower- and upper-middle HEI subjects} i.e.; 1/4 prevalence increment for subjects in the 2\textsuperscript{nd} and 3\textsuperscript{rd} quartile in the HEI distribution.

The fraction of oral and pharyngeal cancer cases attributable to HEI adherence under the \textit{maximum prevalence reduction} scenario was 42.0% (25.1%-59.9%). This figure was 18.8% (8.3%-26.5%) according to the \textit{mild global prevalence intervention} scenario. Whereas, the PIF estimates were 18.4% (10.0%-26.1%) and 16.4% (7.4%-23.2%) according to the \textit{strong prevalence intervention on the low and lower-middle HEI subjects} scenario and the \textit{mild prevalence intervention on the low and lower-middle HEI subjects} scenario, respectively.

The fraction of attributable cases were 15.6% (8.0%-21.6%) under the \textit{strong prevalence intervention on the lower- and upper-middle HEI subjects} scenario and 13.6% (6.6%-18.9%) under \textit{mild prevalence intervention on the lower- and upper-middle HEI subjects} scenario.

The need of the epidemiologist and public health officers to translates risk factor prevalence and disease occurrence in useful numbers under realistic and feasible counterfactual scenarios makes the PIF to be in a mounting epidemiological importance.

\textbf{Keywords:} Attributable shares, Continuous exposure, Counterfactual scenarios.

\textbf{References:}
The information of the Web for local administrations

Paola Chiesa, Marco Palamenghi
Municipality of Brescia - USCI

The sudden changes that characterize the current scenario of information and the interaction via the Web are changing the approach of the public sector to the Network, opening up new scenarios, new perspectives and new challenges. Local administrations must also be made aware of the use of "Big Data" in order to extrapolate the data informative potential. According to this perspective, searching for online data through common search engines by common users can be used as a tool for analysis and forecasting. The fields of application may be the most varied: from social to marketing, from economy to tourism.

The present work, which is born inside the working group dedicated to communication of the cultural area, aims to make a more effective marketing and communication campaign of the Municipality of Brescia related to the cultural and tourist events planned in the municipal area. This aims to provide new elements in order to gain better communication and to verify the achievement of the objectives through the use of free tools such as Google Trends, UberSuggests and AnswerThePublic. These tools base their statistics on users web search activity and can extract data and "food for thought" for decision-makers and domain experts to be interpreted.

The case study examined is CidneOn (International Festival of Lights) which takes place at Brescia Castle in February and which attracted about 340,000 visitors in the 2019 edition. The results obtained, shared with the experts of communication, underline the importance of the value of data analyzing. In particular, the importance of using a brand easily recognizable and unique to characterize an event has emerged. Furthermore, information search let to plan the timing of the communication, the organization of some services useful for the citizen to be able to access the events, and the possibilities to share the communication efforts with other stakeholder.

**Keywords:** Big Data, Google Trends, Web Marketing, Search Engine

**References:**

https://trends.google.it/trends, Web
The mobile phone big data tell the story of the impact of Christo's *The Floating Piers* on the Lake Iseo

Maurizio Carpita
Department of Economics and Management, University of Brescia, Italy;
DMS StatLab - Data Methods and Systems Statistical Laboratory, University of Brescia, Italy.

"Those who experienced *The Floating Piers* felt like they were walking on water – or perhaps the back of a whale. The light and water transformed the bright yellow fabric to shades of red and gold throughout the sixteen days."

Christo

Figure 1: Christo's *The Floating Piers* (christojeanneclaude.net/projects/the-floating-piers).

From June 18 through July 3 2016, the Italy's Lake Iseo was reimagined, with an international event free and open to the people, *The Floating Piers* (Figure 1). It was a temporary 3-kilometer-long walkway that crossed the shores of Lake Iseo (about 100 kilometers east of Milan), from Sulzano to Monte Isola and to the island of San Paolo created - using canvases, cables and metal structures - by the contemporary artist Christo (Figure 1). Has been estimated that more than 1,2 million people visited the site in the sixteen days of *The Floating Piers* event, an average of 72,000 visitors daily in an area where usually there are about 12,000 residents.

In this talk we present analytics about this big social event obtained using the statistical approach with mobile phone big data (Carpita and Simonetto, 2014; Metulini and Carpita, 2019).

**Keywords:** big data, raster analytics, histogram of oriented gradients, cluster analysis

**References:**

Evaluating museum visitor experience in Brescia: the Capitolium Archaeological Area and the Santa Giulia museum complex

Paola Biggi\textsuperscript{a}, Maria Elena Comune\textsuperscript{b}, Marica Manisera\textsuperscript{a}, Loredana Scassa\textsuperscript{a}, Paola Zuccolotto\textsuperscript{a},

\textsuperscript{a} Department of Economics and Management, University of Brescia, Brescia, Italy
\textsuperscript{b} Unità di Staff Statistica, Comune di Brescia, Brescia, Italy

In the last few decades, museums have expanded in variety and they are becoming more and more popular. Together with this evolution, the role of museums in society has notably changed, moving from cultural centers devoted to collections and research to institutions for public learning. In museums, education plays a crucial role, also because nowadays the financial support for museums essentially comes from the public sector, which requires attainment of public education goals. In this context, it is very important to understand and evaluate the museum experience from the visitor’s perspective (Falk and Dierking, 2016).

In this contribution, together with Brescia Municipality and Fondazione Brescia Musei, we developed a survey aimed at investigating the museum visitor’s experience, with a focus on the persons visiting the Capitolium Archaeological Area and the Santa Giulia Museum complex. In the period June-September 2019, visitors were asked to fill in a questionnaire addressing the reasons why people go to museums, what they expect to do there, what is their immediate experience, what are their previous knowledge, interests and expectations, satisfaction, together with some socio-demographic characteristics.

Several facets of the visitors’ behaviour are investigated by means of latent variables, measured by means of several items with ordered response categories. From a statistical point of view, the resulting data must be analysed by means of appropriate models for latent variables (among others, Finch and French, 2015) and multivariate data analysis techniques (see, for example, Husson et al., 2017).

Results from this survey will give very important recommendations for museum professionals, who can adapt museum interpretation and exhibition design in order to improve the visitors’ museum experience and help to reach their educational goals, but also to attract more and more people with different behaviour and to convince visitors to repeat their experience.

Keywords: Museum experience, Survey, Visitor behaviour, Latent variables, Multivariate data analysis

References:
Tangible and multisensory museum visiting experience: the Tosio Martinengo Gallery in Brescia

Elisa Cirillo\textsuperscript{a}, Marica Manisera\textsuperscript{a}, Marco Palamenghi\textsuperscript{b}, Giulia Zanoletti\textsuperscript{a}, Paola Zuccolotto\textsuperscript{a}

\textsuperscript{a} Department of Economics and Management, University of Brescia, Brescia, Italy
\textsuperscript{b} Unità di Staff Statistica, Comune di Brescia, Brescia, Italy

Evaluating the museum visiting experience in order to give useful insights to museum professionals is gaining more and more importance (Falk and Dierking, 2016). The understating of visitors’ beliefs, feelings, sensations, perceptions and emotions can indeed give crucial suggestions to professionals who need knowledge and skills to re-organize the museum experience and design the exhibition path in order to enhance the visiting experience.

Another important information source is given by sensory evaluations experienced by persons during their visit. The rising of sensory experimentation in contemporary museums and exhibitions gave birth to the so-called “Sensory Museology” (Howes, 2014). Besides sight, the sense of touch is the most frequently investigated in museum studies, but also hearing, smell and taste are now being studied, by means of multisensory museum objects. The aim of this contribution is to describe the results of a survey developed in collaboration with Brescia Municipality and Fondazione Brescia Musei, addressed to evaluating the museum visiting experience. The focus is on visitors of the Tosio Martinengo Gallery in Brescia, which has been completely renewed and re-opened last year. Twenty-five exhibition rooms contain a collection including masterpieces which establish the international importance of the Brescian collection. Our challenge is to evaluate not only tangible aspects of the museum experience, typically expectations from the visit, reasons leading to visit the Gallery, time dedicated, satisfaction, but also the emotional impact and the multisensory experience, not from the fruition of multisensory museum objects but focusing on the visitors’ emotions and sensory sensations when visiting the Gallery.

Data were collected by administering two separate questionnaires to a stratified random sample of visitors in June-September 2019. The use of questionnaires with Likert-type scales and questions with responses measured on a rating-point semantic differential scales makes it necessary to use appropriate statistical methods, including multivariate data analysis techniques (see, for example, Husson et al., 2017).

Results from this survey can be fruitfully used by museum professionals in order to understand the visitors’ evaluations about their experience and think about new communication languages that can be used to enhance all the facets of the visitor experience.

Keywords: Museum experience, Sensory analysis, Emotions, Multivariate data analysis

References:
We consider the issue of estimating the effect of schools on student achievement when a pre-test is available. Based on Invalsi data (Martini, 2018), our focus is on the value added of the lower secondary schools on test scores at the 8th grade (post-test), accounting for the student test scores at the 5th grade (pre-test), and controlling for the available student and school characteristics by multilevel models (Goldstein, 2010). In this framework, the school value added can be estimated by either inserting the pre-test score as a covariate (conditioning approach), or by considering the difference between post-test and pre-test scores as the response (gain score approach). We discuss the assumptions underlying each approach (Kim and Steiner, 2019), and the consequences of their violations, and we present a comparison of the results obtained with the two approaches on Invalsi data.

**Keywords:** Invalsi achievement tests, Multilevel model, Random effects model, Value added.

**References**


Strengths, constraints and issue of progress test

Laura Antonucci\textsuperscript{a}, Rosanna Cataldo\textsuperscript{b}, Corrado Crocetta\textsuperscript{c}, Gabriella Grassia\textsuperscript{d}

\textsuperscript{a} Department of Clinical Science, University of Foggia, Foggia, Italy;
\textsuperscript{b} Department of Statistics and Economics, University of Naples, Naples, Italy;
\textsuperscript{c} Department of Economics, University of Foggia, Foggia, Italy;
\textsuperscript{d} Department of Social Science, University of Naples, Naples, Italy;

There has been an increasing use of assessment tests in medical education. The progress test of applied knowledge has been applied for the first time, in late 1970s, at Maastricht University and at University of Missouri. It is currently used in United Kingdom, Italy, Netherlands, Germany, Austria etc. The Progress Test is used in many ways and with several formats to reflect the variety of curricula and assessment purposes. In Italy since 2006 most Italian medical schools use this longitudinal, progressive method of assessment to evaluate the curricula. This method helps to explore improvements in the quality of learning, teaching and in general of the educational standards. The Progress Test made up of 300 multiple-choice questions, has been administered every year to all the Italian medical students. In order to fit the Italian Core Curriculum, the questions have been written by many Italian university teachers. For each multiple-choice question, the Facility Index was obtained (number of correct answers/number of students). We analysed the results for academic years 2017-18 and 2016-117 for the 31 dentistry courses involved in this test.

**Keywords:** Progress test, PLS PM, Performance indicators

**References:**


Measuring Schein’s career anchors on University students to promote self-realization and organizational well-being

Cristiana Martini, Giovanna Galli, Marcello Tedeschi
Department of Communication and Economics, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Schein’s Career Orientation Inventory (1990) identifies eight anchors that drive employees’ career paths and orientations: general managerial competence, technical/functional competence, autonomy/independence, security/stability, entrepreneurial creativity, dedication to a cause, pure challenge, life-style. Schein affirms that a career anchor is “that one element in a person self-concept, which he or she will not give up even in the face of difficult choices” (1990) and that, along his/her work experience, each individual develops only one career anchor. Consistency between crucial career anchor and work experience helps people to attain positive careers outcomes and self-realization. Measuring career anchors among university students is particularly meaningful since it allows to raise awareness on how expectations may shape work decisions and priorities, and helps the institutions to develop proactive and comprehensive strategies to prepare the students for their future careers.

Schein model has been extensively adopted but also widely criticized. One of the main criticisms refers to the coexistence of more than one dominant anchor in employees’ self-concept, even when they reach a high work experience. Moreover, the construct validity of Schein scale has sometimes appeared controversial.

This paper aims at assessing construct validity of Schein scale and at evaluating the existence of one or more crucial career anchors at university students’ level. Since students usually do not have long work experience, the measurement scale has been reworded in terms of desires, needs and expectations; this scale has then been submitted to 686 students at the University of Modena and Reggio Emilia attending scientific, humanistic and social undergraduate and graduate study programmes. The construct validity of the new scale has been tested, comparing results with eight or nine emerging dimensions. Results show that not all the original Career Orientation Inventory belong to the same construct and that students use more than one dimension to depict their desires and needs. Moreover, different students’ vocations emerge from a difference on dominant anchors, depending on the study programmes they are attending, with some unexpected convergence of engineering and business students’ anchors.

**Keywords:** Schein’s career anchors, Organizational well-being, Scale validation, University students’ vocation.

**References:**
Internationality as a psychological and social trigger to participants to Erasmus+ VET international mobility

Elena Bortolato\textsuperscript{a}, Luigi Fabbris\textsuperscript{a}, Manuela Scioni\textsuperscript{a}

\textsuperscript{a}Department of Statistical Sciences, University of Padua, Italy

VET international mobility under the shield of Erasmus+ is a mobility that students of any type of school and apprentices from firms realise in an abroad organisation. A survey was carried out in 2018 to measure the effects of mobility on a large sample of participants. The final evaluation of the experience was evaluated on a 1 to 10 scale and a set of descriptors of the experience was also collected to understand which groups of participants particularly appreciated or did not appreciate their abroad experience. The survey was conducted in four countries: Germany, Italy, Portugal and Spain.

A nested regression model was applied on evaluations. Regressors were selected through a stepwise procedure. The analysis showed that the activity participants were doing before mobility, the duration of the internship, the duties at destination as compared with the home duties, working in an international environment, the languages used at work and in the leisure time, the skills they improved, and socio-psychological descriptors of mobility on participants enable to explain a large portion of variability in participants’ evaluations. Gender, which was significant taken alone, is no longer significant if socio-psychological variables are inserted into the model. Also, costs to families and time and personal relationships “sacrificed” by participants to realise the experience almost vanish once positive aspects enter the model.

The understanding of the evaluation given by participants may address policy making to improve VET international mobility.

Keywords: VET international mobility, Abroad internship, Nested model, Socio-psychological descriptors.
Three is a magic number
Evidence on the three-point rule application to Italian soccer league

Vincenzo Alfano\textsuperscript{a}, Giuseppe Lucio Gaeta\textsuperscript{b}, Michele Gallo\textsuperscript{b}, Francesca Rotondo\textsuperscript{b}

\textsuperscript{a} Consiglio Nazionale delle Ricerche, Istituto per il Mediterraneo, Napoli, Italy;
\textsuperscript{b} Dipartimento di Scienze Umane e Sociali, Università degli Studi di Napoli L'Orientale, Napoli, Italy;

In 1995 the \textit{Fédération Internationale de Football Association} (FIFA) raised from two points to three points the reward for the winning team in soccer matches. At the same time, the reward for teams involved in draw matches (one single pint) was left unaltered.

By increasing the difference in reward between winning and drawing, such a choice aimed to promote more offensive play by the teams and ultimately to make matches more exciting.

This paper empirically studies whether such objectives were actually achieved when the rule was introduced into the Italian \textit{Serie A}. Recent papers inspected the effects of the introduction of the three points rule in some national soccer leagues (Guedes and Machado 2002 inspected the case of Portugal; Palacious-Huerta 2004; Dilger and Geyer 2009 as well as Hon and Parinduri 2016 examined the case of Germany;)) but, to the best of our knowledge, no previous study has specifically inspected the case of the Italian \textit{Serie A}. Nevertheless, investigating the impact of the three points rule on Italian matches is definitely worthwhile especially because Italian teams have a long tradition of ultra-defensive strategy.

Our analysis uses official data on the \textit{Serie A} matches played from the beginning of the 80s until recent days; following previous contributions (Hon and Parinduri 2016) our empirical elaborations rely on a regression discontinuity design.

We find evidence that the introduction of the three-points rule increased the number of wins in the Italian \textit{Serie A} as well as the number of goals scored. Furthermore, we find evidence that it increased the number of goals scored during the second half of those matches that reported a draw at the end of the first half.

Overall these results suggest that in the case of the \textit{Serie A} the three-points rule worked as FIFA intended.

\textbf{Keywords:} Three-point rule, Soccer, Regression discontinuity design.
Evaluating karate performances using a wearable wireless Inertial Measurement Unit

Simone Di Zio\textsuperscript{a}, Lara Fontanella\textsuperscript{a}, Luigi Ippoliti\textsuperscript{b}
\textsuperscript{a} Department of Legal and Social Sciences, University “G. d’Annunzio”, Chieti-Pescara; \textsuperscript{b} Department of Economics, University “G. d’Annunzio”, Chieti-Pescara.

Modern Karate consists of two different disciplines: kumite (fight) and kata (form). While the first falls into the category of “open skill” sports, where the environment is constantly changing, the second one belongs to the class of “closed skills”, where the environment is stable and predictable, and the athlete knows exactly what to do and when. In a kata, the motor actions follow predefined patterns, having a clear beginning and end, made of an encoded sequence of attacks and defences with a clear rhythm. During the last fifteen years, karate has undergone a radical transformation, thanks to growing research and increasingly sophisticated training methods. The athletes have reached very high technical execution speeds, therefore the quality of katas are very difficult to assess with the naked eye, and the work of the referees (in the competitions) and trainers (in the gym) have become increasingly hard. The possibility of using automatic procedures and minimally invasive tools to assist both coaches and referees would be highly desirable to date. Technologies and statistical procedures for quantitative assessment of karate-kata performances - and more generally for closed skills sports - which can be used during trainings and competitions could be highly beneficial. Among the many tools that the current technology makes available to us, the wearable wireless Inertial Measurement Units (IMU) appears to be a suitable tool for the above purposes (Preece \textit{et al.}, 2009; Worsey \textit{et al.}, 2019). Indeed, they provide an easy, low cost and minimally invasive way to measure kata performance and deliver feedbacks in real-time, therefore ideal for athletes assessment, both in training and competition. In this work, we discuss an application with kata athletes of various levels and schools, and propose a prototype experimental protocol for data collection and analysis. Specifically, the data are obtained through a wearable nine axis Bluetooth IMU, which integrates a three axis Accelerometer (±16g), a three axis Gyroscope (±2000°/s) and a three axis Magnetometer (±180°). The device (Model BWT901CL, Wit Motion), has a baud rate of 115200 and an output frequency up to 200Hz which favours the use of functional data analysis (Ramsay and Silverman, 1997) both to assess the individual athlete performances and to classify kata executions. It will be shown that results from our analysis contribute to improve karate techniques, kata performances and motor efficiency, as well as to enhance training methods, to assist both referees and coaches, and to reach higher quality performances for better sports results.

\textbf{Keywords:} Karate Kata, Closed Skill, Inertial Measurement Units, Accelerometer, Functional Data Analysis.

\textbf{References:}


Application of multivariate statistics in sports: Exploration of recall and recognition of UEFA Champions league sponsors

Milica Maricic\textsuperscript{a}

\textsuperscript{a} Department of Operational Research and Statistics, University of Belgrade, Belgrade, Serbia

So far, multivariate statistical analysis has been employed with a lot of success in various fields of study such as medicine, business, and economic (Mertler and Reinhart, 2016). The field of study which is slowly but surely developing is the application of statistics in sports (Albert and Koning, 2007). The question that has not received significant attention in sponsorship literature is the recall of embedded advertisements (Maricic et al., 2019). Attention has been placed on the recognition of brands whose commercials are shown during the context exposure, but not on the advertisements shown in the context itself (Walraven et al., 2014). The question attempted to be solved is to explore which factors facilitate the recall of advertisements that are integrated in the sports broadcast and is there at all a significant level of recall of these messages? The research aimed to explore how involvement in a particular sport and involvement with a sponsored event impact the recall and recognition of embedded advertisements. Nevertheless, the further analysis aimed to explore whether the recall and recognition of embedded advertisements has a positive effect on attitude towards sponsorship and actual sponsors and towards purchase intentions and experience (Ko et al., 2008). To investigate the above-mentioned relations, a novel conceptual model was proposed. To verify the proposed conceptual model on a particular sports competition, a questionnaire related to the UEFA Champions league sponsors in the season 2016/17 was designed. The data were analysed using structural equation modelling (SEM). The results provided support to the claim that higher involvement in a sport and higher exposure to a sponsored event lead to higher recall and recognition of embedded advertisements. Conversely, the same two factors alongside sponsor awareness, have no effect on the attitude towards sponsors. While on the other hand, sponsor awareness and attitude toward sponsors have a direct positive role on the future purchase intention of sponsor products/services. The findings illustrate that to maximise the effectiveness of their embedded advertisements, sponsors should not simply concentrate on brand awareness, but should go beyond and attempt to improve the consumers’ positive perception of their products and services. Moreover, by sponsoring a sport event, the sponsors can be sure their message will be effectively transmitted to those who are fans of the sport and sponsored competition.

Keywords: UEFA Champions League, Sports marketing, Sports analytics, SEM analysis

References:
RERT: a novel approach for extracting a synthetic representation from an ensemble. The case study of the Italian wine quality

Marika Vezzoli\textsuperscript{a} and Eugenio Brentari\textsuperscript{b}

\textsuperscript{a}Department of Molecular and Translational Medicine, Unit of Biostatistics, University of Brescia, Brescia, Italy
\textsuperscript{b}Department of Economics and Management, University of Brescia, Brescia, Italy

In this study we apply and evaluate the performance of a new algorithm, recently proposed in a paper (Vezzoli et al. 2007) and called RERT (namely REpresentative Regression Tree). The main idea of the algorithm is that from an ensemble of Regression Trees, which is a “black box” that provides good predictions losing the interpretability of a single tree, we can extract a simple, interpretable and accurate model (Banerjee et al. 2012). In other terms, RERT could be a possible solution to the eternal conflict between accuracy and simplicity (interpretability). The RERT stabilizes the results across multiple replications providing accurate predictions also offering a definitive model for decision making. The algorithm runs in three steps as follows: (i) we draw a high number of bootstrap samples from the original data set; (ii) in each sample we grow a tree, and (iii) from the realized ensemble, we select the representative tree (RERT), namely the Regression Tree which provides the best performance in terms of Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC) curve computed on the cross-validated predictions.

In this paper, RERT is used to evaluate the quality of Italian red and white wines trying to identify which of the sensorial and chemical variables have a major impact. The topic is extremely relevant due to the multifaceted nature of the wine quality, where subjective valuations (obtained from sensorial test conducted by panel of experts) and objective features are mixed together in order to obtain fair judgments on the overall quality. In more depth, using data provided by Altroconsumo, an Italian independent consumer’s association, and specifically obtained joining the data used for the Guida Vini from 2006 to 2012 containing 35 sensorial and chemical variables for 2031 wines, we try to predict the Global score of quality (an ordinal variable ranging from 0 - lowest quality - to 100 - highest quality) attributed by Altroconsumo to each wine, also identifying the major determinants of the wine quality.

Keywords: RERT, ensemble of trees, wine quality, chemical and sensory variables.

References:
Prediction of wine taste preferences based on physicochemical properties and data mining models

Paulo Cortez
Department of Information Systems, University of Minho, Guimarães, Portugal

Wine is increasingly enjoyed by a wider range of consumers. In this work, a data mining approach was employed to analyse easily available analytical tests at the certification step of Portuguese white and red “vinho verde”. Using large datasets (compared to previous studies), several regression models were adjusted to map physicochemical properties with wine taste preferences, as evaluated by three sensory assessors via blind tastes. The best results were achieved by a support vector regression model. Under the considered 0 (worst wine) to 10 (best wine) scale, the support vector regression achieved a high quality classification accuracy, around 87.

Keywords: Sensory preferences, Machine learning, Support vector machines, Neural networks
Automatic identification of wines via supervised text classification

Massimo Aria\textsuperscript{a}, Michelangelo Misuraca\textsuperscript{b}, Alfonso Sorrentino\textsuperscript{a}, Maria Spano\textsuperscript{a}

\textsuperscript{a} DiSES, University of Naples Federico II, Naples, Italy; \textsuperscript{b} DiScAG, University of Calabria, Arcavacata di Rende, Italy;

Aim of this paper is to explore the use of text classification in a sensory analysis context. The research question on which the proposal is based considered if it is possible to automatically detect a specific product from the analysis of the reviews written by a group of users, assuming that each product has peculiar distinctive characteristics.

In particular, we focused on wines’ reviews written by experts which are trained to give a professional evaluation of the different products. The task described above can be accomplished by considering the text categorisation techniques developed in the framework of Text Mining. Several algorithms have been developed in literature, but we particularly focused on supervised algorithms. Following a so called \textit{k-fold cross validation} procedure, a selected algorithm is trained on a labeled subset of reviews, then is tested on different subsets. In this way it is possible to estimate how accurately the predictive model performs.

The results obtained by analysing 130,000 reviews scraped from \textit{WineEnthusiast} magazine are presented and discussed, showing the effectiveness of our proposal.

\textbf{Keywords:} Online reviews, Text Categorisation, Sensory Analysis.
Dati e indicatori a supporto delle politiche sociali

Felice Scalvini
Ex-assessore al Welfare del Comune di Brescia

Nella mia esperienza di assessore al welfare della città di Brescia ho cercato di introdurre la cultura dei dati come base per assumere orientamenti e decisioni. Lo sviluppo della conoscenza attraverso l’acquisizione di informazioni adeguate ha costituito un punto fermo del programma strategico dell’Assessorato. Allo stesso modo, con la redazione del bilancio sociale si è avviata una prassi di rendicontazioni confrontabili.

Di particolare interesse è risultata la rilevazione della consistenza e delle caratteristiche del Terzo Settore cittadino, che ha permesso di fondare alcune scelte politiche di fondo.
Social enterprises are covering an increasing role in the production of social interest services. In Italy, the recent reform of third sector organizations (Law 206/2016) has defined them as enterprises devoted to “pursuing the common good, raising the levels of citizenship, cohesion and social protection, encouraging participation, inclusion and the full development of the person, enhancing the potential for growth and employment”. Social wellbeing: this should be the keyword describing the mission of third sector organizations and this aim should especially define social enterprises – within this sector – since the social aim must be conceived by social enterprises as a strict boundary to their entrepreneurial nature, to their management and in their strategic planning.

A need for social enterprises therefore becomes the identification of concrete indexes of their social aims, of wellbeing generated to their users and to the community as a whole, of social impact produced. Having specific indicators of social aims and social impact allows to define own benchmarks and goals of wellbeing and to periodically verify achievements and gaps. In line with this constructive approach, also Law 206/2016 has provided for the implementation in social enterprises of social impact assessment. Indicators should be established and monitored by social enterprises in order to demonstrate and evaluate “qualitative and quantitative, short-run and long-run effects of their activity on the community”.

The paper illustrates some results of the ImpACT method applied to a sample of social cooperatives in north-Est of Italy in order to measure social impact produced. The method consists of questionnaires collecting data on the activity performed and on questionnaires and interviews to the organizational stakeholders, specifically to users of social services and workers. The attention will be devoted to indexes of wellbeing produced, with definition of social cohesion, social inclusion, changes in wealth and psychological wellbeing. The paper will conclude on the ability of social enterprises to develop innovative strategies and processes in order to increase wellbeing and to better answer the need of fragile people.
Bessimo is a Social Cooperative funded in 1976 in order to offer therapeutic rehabilitation programmes to drug-addicted people. Bessimo is a non-profit organization, a social enterprise accredited by Lombardia Region for addictions treatment since 2008. Bessimo cares about addictions from Prevention to Harm Reduction: about 200 people are employed in the Organization. As a social cooperative, Bessimo is based on member social participation and cooperation: the social assembly is composed by 105 members. Bessimo Social Cooperative helps people suffering from other forms of addiction including alcohol, gambling and Internet.

The activities carried out by the Social Enterprise are:

- 14 residential therapeutic communities: Therapeutic Rehabilitation, Orientation, and Specific Treatments for families, alcohol-addiction and multiple addiction. It can also offer an alternative to prison, such as external criminal enforcement corrective measure.
  - Bessimo offers an amount of 265 credited seats
    - 7 Centres for Males
    - 3 Centres for females
    - 1 Centre for Single women with children
    - 3 Centres for Addicted Couples (with or without children)
- 1 Educative Residential Centre for children
- 1 Specialistic Centre for Gamblers
- Social Housing Service
- 4 Harm-Reduction Services in 4 different cities (Brescia, Cremona, Bergamo, Mantova)
- 1 Risk-Reduction Service operating in night-life events
- Counselling Service for families, teenagers,
- 2 Support Services for inmates in Brescia Prisons.

Bessimo Social Cooperative is based in Lombardy. Nowadays, Bessimo is one of the most important Italian organization in the field of addiction care, the services are in four Lombardy provinces: Brescia, Bergamo, Cremona and Mantova. The experience and the number of services and people treated by Bessimo Social Enterprise leaded to an organization really interested in data. Each person is registered, each program is defined and reported. Since 2017, the whole information system about users and beneficiary data is digitalized. This allow the Social Enterprise’s management to use and read the data in order to improve the activity. Each month the management can check the budget consumption, each semester it is possible to verify the outcome of users' therapeutic pathways, to understand improvements, to define needs. At the end of the year, the management can analyse the performance of each service (each residential therapeutic community, each harm-reduction service). This is important not only for internal purposes - knowledge, continuous improvement, innovation – but also for an external use. The size of Bessimo cooperative makes it clear that it has a sample of importance in relation to the issue of drug addiction at the regional, and therefore national, level. Bessimo's data could allows policy makers to read the phenomenon of addictions, to know more about the need and, therefore, to study more effective responses.

During the conference, the manager of the Bessimo Cooperative will present the type of data available, the current internal use for business purposes and the possible use at the policy level.

**Keywords:** Addiction, Healthcare, Social Enterprise, Communities, Harm-reduction services
Impact of changes in the price of meals at university canteens on students’ eating behaviours: evidence from an Italian university using the difference-in-differences approach

Valentina Lorenzoni\textsuperscript{a}, Lucio Masserini\textsuperscript{b}, Mailde Bini\textsuperscript{c}

\textsuperscript{a} Institute of Management, Sant’Anna School of Advanced Studies, Pisa, Italy;
\textsuperscript{b} Department of Economics and Management, University of Pisa, Pisa, Italy;
\textsuperscript{c} Department of Human Sciences, European University of Rome, Rome, Italy.

University canteens play an important role in academic life since they allow students to benefit from subsidized food services and to select meals at lower prices than those commonly available at other local eateries. Therefore, any intervention that changes the conditions of food service delivery is expected to impact not only the students’ eating habits, but also other aspects of the students’ life. This study aims at evaluating the impact of income-based tariff system on students’ behaviors related to the use of the university canteen, in terms of frequency of access and meal choices, by means of a quasi-experimental design and a difference-in-differences approach. Using panel data from an Italian university, this study proves that users who experienced a tariff increase significantly reduced their patronage of university canteens.

\textbf{Keywords:} Difference-in-differences, Tariff change, University canteen meals, University finance.
The blockchain for the certification of the dairy supply chain, the “Lucanum” basket and the bakery products for well-being

Antonio Notarnicola\textsuperscript{a}, Vito Santarcangelo\textsuperscript{a}, Nicola Martulli\textsuperscript{b}, Francesco Abbondanza\textsuperscript{c}

\textsuperscript{a} iInformatica Srls (PMI innovativa), Corso Italia, Trapani, Italy;
\textsuperscript{b} Capurso Azienda Casearia S.r.l., II Trav. Vicinale Le Strettole, Gioia del Colle (BA), Italy;
\textsuperscript{c} L’Abbondanza Lucana, Via B. Buozzi,11, Matera, Italy;

Paper shows innovative blockchain-oriented approaches aimed at complying with the product specification and improving product quality within the dairy supply chain applied to DOP processes of Gioia del Colle’s mozzarella. Alongside this application there is a revolutionary blockchain-oriented initiative applied to the “Lucanum” basket of typical Basilicata products (in order to respect the territorial enogastronomic peculiarities) and to the traceability of bakery products, focused on a well-being target.

**Keywords:** Blockchain, Well-being, Mozzarella Gioia del colle, Lucanum, Traceability

**References:**
**Family lifestyle habits: what is passed down from adults to children?**

Riccardo Borgia\(^a\), Elena Castellari\(^a\), Paolo Skokai\(^a\)

\(^a\) Dipartimento di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza, Italy

Noncommunicable diseases are responsible for 70% of deaths worldwide, primarily induced by the use of tobacco, unhealthy dietary habits, lack of physical activity (PA) and abuse of alcohol (WHO, 2017). Among these behaviours, the adoption of healthy eating patterns and a regular physical activity initiated in childhood may reduce the risk of occurrence of overweight and obesity, immediately, and the insurge of chronic diseases, into adulthood (Nicklas et al., 2001). To promote the early adoption of such healthy practices family has always played a key role (Patrick & Nicklas, 2005). Nevertheless, the children’s daily life environment is deeply and rapidly changing and the central role of family in shaping children lifestyle habits is being called into question.

The study focuses on four behaviors indicated as highly – negatively and positively – related to the occurrence of overweight and obesity. The advisable behaviors are i) eating food for breakfast, ii) doing regular PA and iii) consuming five portions of fruit and vegetable (FV) per day. The unadvisable behavior is consuming one or more savory snacks a day. Each habit is studied individually but by means of the same model (i.e. including the same variables). The model employs a multinomial logistic (MNL) regression to relate the occurrence of the four investigated behaviors to the characteristics of the children (age and gender) and the respective household. Specifically the latter outline health and socio-demographic features of the household: share of obese adults, number of children and adults, adults’ education level, share of females, and geographical area. Moreover, to assess the household behavioral environment – main focus of the research – the share of adults manifesting the investigated children behaviors is included in the model. Lastly, the year of data collection is also observed – even if the study is cross-sectional – to monitor any significant variation from 2013 to 2016. The study is performed on the microdata of the *Italian Multipurpose Survey on Households Daily Life Aspects* provided by ISTAT (years 2013, 2014, 2015, 2016). The dataset consists of 25,265 children belonging to 16,893 households. According to UN, children are identified in the dataset as individuals younger than 18 years old. The main insight provided by the study is that the likelihood of observing the children behaviors increases with the share of adults manifesting the same behavior (except for PA) suggesting that family still plays a role in influencing some children lifestyle habit, although unevenly among behaviors. Indeed, results show that the habit of practicing regular PA is much more likely to be passed down than breakfast-eating, and they are both less likely than daily-snacking. This might suggest that unadvisable behaviors are more liable to be emulated than the advisable ones. Finally, the research suggests priority targets for children health promotion policies: Southern Italy and – limited to some habit – 14-17 years old.

**Keywords:** Children’s habits, Family environment, Lifestyle behaviour adoption, Eating behaviour.

**References:**


Intelligent systems to support patients

Vito Santarcangelo\textsuperscript{a}, Emilio Massa\textsuperscript{a}, Diego Sinit\textsuperscript{a},c, Giuseppe Scavone\textsuperscript{b}

\textsuperscript{a} iInformatica Srls (PMI innovativa), Corso Italia, Trapani, Italy;
\textsuperscript{b} Centro Rham S.r.l., Via I Maggio 60, Matera, Italy;
\textsuperscript{c} University of Catania, Italy;

Paper shows innovative approaches and tools patented by the Centro Rham applied to orthoptics, to DSA support and alzheimer patients. High innovation system implemented are focused on improving awareness of visual coupling and monocular response. Traditional tools based on innovative communication methods improve the patient analysis. The revolutionary methods and products object of industrial research and experimental development carried out are the result of the integration of multi-faceted engineering and health skills, and aspire to become a reference point for the sector.

\textbf{Keywords:} DSA, Orthoptics, alzheimer, Patents, Innovative methods

\textbf{References:}

Intelligent device and method for a better electroneurographic examination

Lara Gallicchio

a Via Don Luigi Sturzo, Matera, Italy;

Paper shows a patented ulnar goniometer device, characterized by the use of a base from a variable angle managed by a suitable guide to position the ulna firmly, to record the positioning angle of the ulna tracking all the operations in a ledger of a private or public blockchain. Using a process of mining phase applied on data recorded in blockchain, it allows to better determine the parameter C that defines the value of the performance of the test found as the parameters of the patient vary and the angle used for the protractor.

Keywords: Blockchain, Process Mining, Ulna, Ledger, Examination

References:
Gallicchio, L., (2019). Dispositivo goniometro ulnare e relativo metodo per l’ottimizzazione dell’esame elettroneurografico, UIBM, 102019000009912, Patent
The determinants of vaccination behaviour of General Practitioners in South Tyrol: differences and similarities between Italian and German respondents

Giulia Cavrini\textsuperscript{a}, Andrea Lazzerini\textsuperscript{b}

\textsuperscript{a} Faculty of Education, Free University of Bozen/Bolzano, Bolzano, Italy
\textsuperscript{b} Faculty of Statistics, University of Bologna, Bologna, Italy

For decades, vaccinations, particularly for infants, have been decisive tools in the battle against infant morbidity and mortality. Despite this, in recent years there has been huge disinformation campaign throughout western nations in which anti-vaccine campaigns have been purported to be extremely harmful and dangerous for the population, and unfortunately with too often deadly results. Moreover, for many decades in Italy and in other European countries, associations against vaccination practices have been spreading, in relation to cases of supposed permanent and disabling effects on children that have been attributed to vaccinations. In Italy, in recent years, several surveys have been conducted to better understand the determinants of vaccination behaviours. Of the Italian regions, the province of Bolzano has the lowest vaccination coverage and for this, an analysis of this behaviour in the territory is required. The general objective of the project, conducted in collaboration with the Health Authority of Bolzano, is to identify what ideas this resistance and the arguments against vaccination are based on, and what influence doctors and health professionals may have on this lack of confidence in Science.

In order to achieve the objectives, three different surveys have been planned: a) a collection of the opinions of 398 General Practitioners and Free Choice Paediatricians that play a key role in the vaccination program; b) a collection of the opinions of parents and prospective parents; 3) an opinion survey of the younger population - age group 18-24 years. The first survey was conducted online in 2018; the second and the third surveys will be conducted next autumn. Of the 248 respondents who successfully filled out the questionnaire, without significant differences between Italian and German speakers, about 70% are general practitioners while the remaining 30% are paediatricians, with the same proportion of the Italian/German speakers. German speaking physicians tend to see less value in vaccination as a defence against infectious diseases and even consider them less safe. Most of both the Italian and German speakers would like to have more information on vaccines to talk to parents who are against vaccination. There was a significant disagreement between the two groups of respondents when asked if mandatory rather than recommended vaccination was more important. Most of the Italian speakers answered in the affirmative in contrast to the German speakers. Descriptive analyses and the preliminary logistic models show significant differences in vaccination behaviour among Italian and German speaking physicians.

\textbf{Keywords:} Vaccination, South Tyrol, General Practitioners, Italian and German speakers

\textbf{References:}

Balancing multi-class imbalanced data into a training dataset using SCUT method

Rafaela Bueno, Luiz Sá Lucas
Theopinionedge, Brazil.

The present paper is a result from a paper from Carpita et. al. (2015). The idea was to improve the accuracy of results for a predictive model for the Italian Soccer Championship (2008-2012) where the objective is to estimate a probability of win / draw / loss of the home teams. To split the dataset into training and test ones we have a drawback: the response variable of the training sample is unbalanced since the sizes of each category (win, loss, and draw) are different: 40% / 30% / 30%. For unbalanced binary response variables, there are four available options in CARET (R software package). They are: Up, Down, SMOTE, and ROSE. However, when an unbalanced response variable is multi-categorical, it is necessary to search for other tools besides Up and Down. One of the authors of the cited paper (Carpita) called our attention for a technique called SCUT that could help with the matter. Other authors developed the theory, but we developed a R software function, also called SCUT. Then the purpose of this paper is to present this SCUT R function for unbalanced multi-categorical response in training datasets. Two indicators were used to measure and compare the used methods - Accuracy and Kappa – and three different methods - Down, Up and SCUT. In addition, three distinct training datasets sizes for SCUT output were tested: 380, 600, and 1,200 observations. As a result, it is expected to: i) present the SCUT function and show that it is a very good method to compensate for the unbalanced multicategorical variables; ii) indicate the best method in this case; iii) the number of observations for the SCUT training dataset that suggests the best sample size for the method.

Keywords: SCUT, Multicategorical variable, CARET, Training dataset balancing techniques

References:
Exploring the statistical structure of soccer team performance variables using the Principal Covariates Regression

Maurizio Carpita\textsuperscript{a}, Enrico Ciavolino\textsuperscript{b}, Paola Pasca\textsuperscript{b}

\textsuperscript{a} Department of Economics and Management, University of Brescia, Italy; \\
\textsuperscript{b} Department of History, Society and Human Studies, University of Salento, Lecce, Italy.

In the Data Science panorama, great room for indicators building, as well as predictive modeling is represented by sports data. Match outcome lends itself to the application of statistical learning models, while players’ performance represents a topic of particular interest for decision making and best choices in the competitive framework. The European Soccer database, available on Kaggle (KES database) incorporates players’ and teams’ data of about 20,000 soccer matches for seasons 2009-2015 in 10 European countries (Carpita et al., 2019b-c). Experts of the EA Sports FIFA videogame (see the website sofifa.com) state that the performance of a soccer player is made up of 7 broad dimensions (\textit{power, mentality, skill, movement, attacking, defending} and \textit{goalkeeping}), each of which incorporates, in turn, more specific skills to be developed and mastered by players on the pitch. Relying on experts’ suggestion, Carpita et al. (2019c) modified the original sofifa indicators by incorporating the four player roles (forward, midfielder, defender, goalkeeper): results showed that performance skills might play a different role according to where players are located in the pitch. However, no statistical inquiry has been carried out on sofifa experts’ performance indicators. Correlations among them revealed an unclear dimensional structure, making their statistical structure worth to be examined in detail. As a first development, Carpita et al. (2019a) used a non-supervised clustering technique for multivariate data which, however, did not consistently improve prediction of match results. For this reason, it is worth to examine the KES database with clustering techniques that also encompass prediction objectives. \textit{Principal Covariates Regression} (PCovR) fits this purpose: it simultaneously reduces the predictors to a few components and regresses the criterion on these components (De Jong and Kiers, 1992). The predictive performance of PCovR components is compared with experts’ sofifa indicators via \textit{Skellam Model}, a regression variation that best fits the distribution of home and team goal differences (Karlis and Ntzoufras, 2008).

Keywords: Principal Covariates Regression, Skellam Model, Composite Indicators, KES.

References:


Comparing statistical models and Machine Learning algorithms in predicting football outcomes

Leonardo Egidi a, Nicola Torelli a

a Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy

Nowadays, modelling football outcomes is widespread and popular and the challenge to include relevant predictors along with new possible correlations is strong. From a statistical point of view, two approaches are designed to achieve this task: the goals-based (direct) models (Dixon and Coles, 1997; Karlis and Ntzoufras, 2003) for the number of goals scored by two competing teams; and the results-based (indirect) models, for the probability of the categorical outcome of a win, a draw, or a loss, the so-called three-way process. Both the frameworks have pro and cons; a long debate has been produced to state which approach is better, and many agreed that any direct comparison between the forecasting abilities of the two types of models must be based on forecasts of match results (Goddard, 2005).

Machine Learning tools such as Classification and Regression Trees (CART, Breiman et al. (1984)) and Random Forests represent alternatives to predict new match results (Groll et al., 2019) and in some cases have proved to be successful.

In this paper we develop a broad comparison between some statistical results-based models and some results-based Machine Learning algorithms, to explore predictive performance for future matches. Although not conclusive, we believe our comparison review may be beneficial for future scholars to discern between goals-based and results-based models.

Keywords: Football model, Forecasting, CART, Predictive performance.

References


Adopting an identity perspective in studying food choices could be of interest in analysing how vegetarians, and meat-eaters as well, feel toward dietary styles involving the reduction of meat and dairy products consumption. Therefore, the main objective of the present paper is the validation of the Italian version of the Dietary Identity Questionnaire (Rosenfeld and Burrow, 2018). Moreover, we want to test the effectiveness of the DIQ to separate and characterize the different eating behaviour, which as a whole define a spectrum of omnivorous food choices. For the purpose of our research, a forward translation followed by a back-translation of the DIQ was performed via two bilingual researchers, independently of each other. An online questionnaire was administered via Qualtrics to a sample of the Italian population. The questionnaire is composed by: a) the Italian version of the DIQ, consisting of 33 items; b) a subset of items aimed at the characterization and classification of food choices; c) socio-demographics.

Data collection is still ongoing and, until now, 180 individuals completed the questionnaire. As a first analysis, an Exploratory Factor Analysis (EFA) was performed, using maximum likelihood factoring and promax rotation, in order to extract eight factors from the DIQ items. Then, according to Rosenfeld, 2019, a Confirmatory Factor Analysis (CFA) will be conducted. Moreover, we will assess if different dietary patterns (i.e. vegetarians, omnivores, flexitarians and so on) differ significantly from one another on several DIQ variables in order to envisage the use of DIQ as a valid instrument for the consumer segmentation along the vegetarian-omnivorous continuum.

In this work, we will validate the Italian version of the DIQ and we’ll verify whether different dietary styles are accompanied by significant differences in the mean scores of the DIQ’s subscales.

**Keywords:** dietary identity, vegetarianism, food sustainability, confirmatory factor analysis

**References:**

Validation of a food insecurity scale through structural equation models
Elena Grimaccia\textsuperscript{a}, Alessia Naccarato\textsuperscript{b}
\textsuperscript{a} Istat and Roma Tre University, Rome, Italy;\textsuperscript{b} Department of Economics, Roma Tre University, Rome, Italy;

The aim of the study is to test the internal consistency, convergent and construct validity of a food insecurity scale, by means of a Confirmatory Factor Analysis. Food insecurity is at the basis of sustainable development, and its measurement is still a policy and academic issue. The Food Insecurity Experience Scale (FIES) is a recently developed tool, created by FAO, in order to measure food insecurity at the individual level all over the world. The instrument has been included among the indicators used to monitoring Goal 2 of the Sustainable Development Goals (“End hunger, achieve food security and improved nutrition and promote sustainable agriculture”). The validation of such a scale, surveyed worldwide, is really important because of very large social and cultural differences in the access to food.

In general, measurement involves assigning scores so that they represent some characteristic of the individuals. Therefore, to assess the validity of FIES, it is necessary to demonstrate - not only to assume - that the FIES score represents the underlying latent construct “food insecurity”. In order to validate the tool, after analysing the frequency distributions and measuring the internal consistency, an Exploratory Factor Analysis (EFA) has been applied. Based on the latent factors identified by EFA, a Confirmatory Factor Analysis (CFA) has been estimated through the application of Simultaneous Equation Models. CFA is a type of structural equation modelling that deals with the relationships between observed measures or indicators (in our case test items) and latent factors. Finally, external validity has been evaluated by a micro econometric analysis of FIES in relation with extreme poverty and other relevant factors of food insecurity.

Results show that FIES presents a good level of reliability and internal consistency, so it can be employed successfully, and the measure of food insecurity associated with a respondent can be calculated by the number of positive responses to items. However, the cumulability of the scale is not perfect, and two items can be deleted without losing fundamental information, in order to reduce the statistical burden on respondents and to improve the cumulability of the scale. Moreover, the results of the CFA models show two different latent constructs: one measuring ‘Perceived’ aspects of food insecurity and a sub-scale related to ‘Actual experienced’ activities. Therefore, a proposal for two sub-scales has been made, offering a better tool for policies.

Keywords: Food insecurity, Confirmatory Factor Analysis, Structural Equation Models, Validity.

References:
Perception of the food quality in children: 
a compared approach between Bayesian Network and 
Structural Equation Model

Anna Simonetto a, Silvia Golia b, Buirma Malo a and Gianni Gilioli a

a Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy;
b Department of Economics and Management, University of Brescia, Brescia, Italy

The evaluation of the concept of multidimensional food quality in children (up to 14 years) is of particular interest for political institutions, educational agencies and health professionals as children are in the most sensitive periods for the formation of orientation in food choices and the ability to respond to educational proposals aimed at promoting the health and quality of life of today's and tomorrow's citizens (Holsten et al., 2012).

The objective of the study is to assess the children knowledge of the six dimensions of food quality identified in our multidimensional approach: Nutritional value, Organoleptic quality, Food Security, Food safety, Environmental sustainability, Social responsibility. The research team developed a tool to detect the concept of food quality in children (11-14 years old, first grade secondary school). This tool was administered to 696 students from 6 schools in the Province of Brescia.

The results were analysed with a Bayesian Network (BN) model (Kjærulff and Madsen, 2013) to identify the independence and association relationships existing between the identified dimensions. Results of BN were compared with the results of Structural Equation Models (SEM; Bollen, 1989). The 6 dimensions are correctly identified by both models, although the items relevant to each dimension are modified. The bootstrap approach applied to the BN selects fewer associations between dimensions than SEMs. Finally, the scores estimated with the SEM were used to assess the level of knowledge of food quality in children, considering gender, class attended and previous experiences in the field of food education.

Keywords: Structural Equation Models, Bayesian Networks, Food quality, Multidimensional approach.

References:


City Prosperity Index: A comparative analysis of Latin American and Mediterranean cities based on well-being and social inclusion features

Alessio Surian\textsuperscript{a}, Andrea Sciandra\textsuperscript{b}

\textsuperscript{a} Department of Philosophy, Sociology, Education and Applied Psychology (FISPPA), University of Padova, Padova, Italy;

\textsuperscript{b} STAR.Lab - Socio Territorial Analysis and Research, University of Padova, Padova, Italy;

The paper analyses recent data provided by the City Prosperity Index (CPI). CPI (UN Habitat, 2016) made available baseline data to monitor Sustainable Development Goal 11 aiming at making cities and human settlements inclusive, safe, resilient and sustainable. It is an instrument based on UN-Habitat concept of urban prosperity as being composed by six dimensions: productivity; infrastructure; quality of life; equity and social inclusion; environmental sustainability, and governance and legislation.

The City Prosperity Index is computed using city level data deriving from a set of commonly available indicators that exist among all cities. The global CPI is a weighted mean of standardized indices from each of the six dimensions.

The study focuses on Quality of Life (QOL) and Equity & Social Inclusion (ESII) indicators as key data in order to monitor well-being and urban inclusion. Based on CPI’s units of analysis, this study performs a comparative analysis of cities from different countries belonging to two areas: Latin America and Mediterranean (including both European and Northern African cities). This allowed us to explore the different stages of modernization (Inglehart & Welzel, 2005) in relation to the dimensions of social inclusion and well-being.

Cluster analysis grouping data by QOL and ESII indicators shows that Mexican and Brazilian cities have very similar performances, while Moroccan cities show a gap in QOL indicators compared to Southern European cities, even though Moroccan cities accomplish better results in relation to poverty rate and youth unemployment.

An interesting result occurred using t-tests with 1000 bootstrap samples to compare the Latin American cities with the Mediterranean ones. The analysis does not show statistically significant differences concerning indicators relating to Education (Literacy Rate and Mean Years of Schooling), Social Inclusion (Slum Households) and Gender Inclusion (Equitable Secondary School Enrolment). Instead, youth employment rate turned out to be significantly higher in Brazilian and Mexican cities when compared to the cities belonging to the Mediterranean area.

These analyses may suggest an incomplete but visible improvement of well-being and social inclusion features in the two most populous Latin American countries or, in other words, a shift towards a post-materialist perspective.

**Keywords:** Well-being, Social inclusion, Health, Education, Social ties.

**References:**


Assessing Mental Health Therapeutic Communities Functioning

Anna Maria Parroco\textsuperscript{a}, Laura Mancuso\textsuperscript{a}, Vincenzo G. Genova\textsuperscript{b}, Francesca Giannone \textsuperscript{a}

\textsuperscript{a}Department of Psychology, Educational Science and Human Movement, University of Palermo, Italy
\textsuperscript{b}Department of Economics, Business and Statistics, University of Palermo, Italy

Empirical and clinical evidence shows that therapeutic communities produce changes in people’s mental health and well-being. Treatment in communities has effects on improving interpersonal relationships, acceptance among members, ability to recognize other people's feelings, symptomatology, life satisfaction and self-esteem. But in which way the treatment works? Which conditions, which therapeutic factors contribute to the treatments effectiveness? Therapeutic communities are complex settings where numerous organizational and relational variables act (structures, activities, care characteristics, relationship between members, group dynamics). The empirical assessment of their functioning is an important clinical and social issue for promoting effective interventions and improving the quality of care.

This paper aims to investigate the formation pathways of the evaluation of Mental Health Therapeutic Communities functioning. In order to analyse functioning survey data have been collected. Units under study are represented by staff members belonging to 19 therapeutic communities. All participants have been involved in the Visiting DTC project on voluntary basis. In order to analyse the evaluation of the functioning, multilevel models have been used to take into account for information related to individual and contextual levels. Our research questions can be summarized as follow: How much do therapeutic communities vary regarding the degree of the functioning evaluation of their staff members? Do staff personal characteristics predict a higher functioning degree? Is the connection between functioning evaluation and individual predictors stable across therapeutic communities? Or does the relationship show substantial variation? According to our empirical evidence neither the individual nor the contextual level, separately considered, can appropriately explain the observed variability. Some preliminary results have shown weak relationships among level 2 variables, while differences among mean-groups, tested by the simplest multilevel model, are significant. This paper contributes to the possibility of interpreting complex data and it provides meaningful insights into the functioning of mental health therapeutic communities.

**Keywords:** Mental Health Therapeutic Communities, Empirical Evaluation, Effectiveness, Multilevel Models.

**References:**


Cyberbullying: A Threat for Relationships and Social Health

Angela Maria D'Uggento\textsuperscript{a}, Ignazio Grattagliano\textsuperscript{b}, Nunziata Ribecco\textsuperscript{a}, Ernesto Toma\textsuperscript{a}

\textsuperscript{a}Department of Economics and Finance, University of Bari Aldo Moro, Bari, Italy
\textsuperscript{b}Department of Education Science, Psychology, Communication Science, University of Bari Aldo Moro, Bari, Italy

The rapid spread of powerful digital and technological devices such as smartphones and tablets among young people is one of the main factors responsible for the attitude of new adolescents generation to replace personal face-to-face relationships with virtual ones, preferring messages, tweet in social groups and images posted on the web (Instagram, Facebook, etc). Even though this lifestyle change facilitates and amplifies the contact opportunities among people, it might represent a new possibility for the phenomenon of bullying to spread, assuming an even more subtle and cowardly modality, the cyberbullying.

Cyberbullying consists in using technology to hurt, threaten, upset or harass someone else by means of aggressive messages, posts of personal information, pictures or videos. Intimidation and unpleasant comments may deal with a person's behaviour, sexual orientation, physical differences or any other kind of discrimination. Cyber-bullying is even more dangerous and underhanded than other face-to-face acts of bullying as it's difficult to control and to stop and the victim can be tormented all day long, violating his/her privacy trough any device or computer. The cyber-bully is able to reach a potentially infinite public through the web, preserving his anonymity and without being physically reachable, then can strike even more aggressively and subtly with offenses and insults the victim who cannot defend himself. Cyber-bullying is, undoubtedly, a serious problem mainly affecting young people, particularly adolescents and pre-adolescents, and adults, parents and educators must counter it with every means. In order to promote a greater awareness of the relational dynamics and the risks connected to the incorrect use of digital devices, to assess the phenomenon of cyber-bullying and to understand the psychological, social and cultural aspects connected to it, some researchers of the University of Bari carried out the survey "Le determinanti sociali del cyberbullismo". A questionnaire composed of 32 questions has been administered to a sample of about 3,700 students attending several Apulian high schools during 2018. Bullying is a complex phenomenon, involving psychological and sociological aspects, often originating from a profound discomfort afflicting both the bully and the victim, therefore, it requires strategies capable of capturing and managing this discomfort. The research allowed us to draw the profiles of the victim and of the persecutor and to understand how to promote the adoption of proactive behaviours inspired by the respect of rules in the school network as well as in social networks. The main results highlight the students' awareness and empathy towards problems concerning their age and allow to identify good practices to support teachers for the prevention of bullying and cyber-bullying in the educational context.

Keywords: Cyberbullying, Statistical analysis, Logistic regression, Social health.

References:
A functional data analysis of Google Trends on Health and Wellness

Francesca Fortuna, Giulia Caruso, Tonio Di Battista
Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University G. D’Annunzio of Chieti-Pescara, Italy

Nowadays, the use of Google Search as source of information has become incredibly pervasive. Through the use of Google Trends, a tool that analyzes the popularity of search queries entered in Google, it is possible to obtain deep insights into population behaviour. Interestingly, Google Search is increasingly being used also to obtain Health and Wellness-related information. Thus, we analysed the search traffic related to this dynamic and attractive sector, using Google Trends data as proxies to measure its attractiveness and analyzing them through the functional data analysis (FDA) approach.

Keywords: Google Trends, Functional Clustering, FDA, Health, Wellness.

References:
Emotional Text mining and health psychology: the culture of organ donation in Spain

Francesca Greco a, Silvia Monaco b, Michela Di Trani b, Barbara Cordella b

a Dipartimento Comunicazione e Ricerca Sociale, University of Sapienza, Rome, Italy;
b Dipartimento di Psicologia Dinamica e Clinica, University of Sapienza, Rome, Italy

Organ donation is an extremely important issue in our society as patients on the waiting lists for transplantation die every year due to the lack of organs. The Spanish healthcare system is a best practice in Europe, reaching 48 donors per million people. A best practice compared to the Italian one in which only 28 donors for million people make this choice. In order to understand the cultural elements influencing the choice to donate, this paper presents the application of Emotional Text Mining (ETM) in the field of health psychology and, in particular, on organ donation. ETM is an unsupervised procedure aiming to profile media and social media discourses that can be considered a proxy of the Spanish culture. It is a fast and simple procedure to extract meaningful information from a large collection of texts. It is based on a bottom-up approach to classify unstructured data by means of a multivariate analysis for the identification of the representation, the cultural symbolization, and the sentiment of organ donation, as those elements influence people’s decision to donate.

We collected all the articles (n= 342) of three most popular Spanish newspapers, the El Mundo, El Pais and Marca, published in the last ten years (2010-2019) containing the keyword “donacion de organos”. In order to check whether it was possible to statistically process data, we calculated two lexical indicators of the large size corpus, the type-token ratio and the percentage of hapax (TTR = 0.092; Hapax% = 0.515). Results show six representations characterizing the Spanish culture of organ donation. While one representation connects the donation to the death, all the other representations highlight positive elements associated to life, which could be a relevant factor explaining the high rate of the Spanish donors.

Keywords: Emotional Text Mining, Organ Donation, Health Psychology, Public Health

References:
Overhearing Italian subjective well-being on Twitter

S.M. Iacus\textsuperscript{a}, G. Porro\textsuperscript{b}, S. Salini\textsuperscript{a}, E. Siletti\textsuperscript{a}

\textsuperscript{a} Department of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy;
\textsuperscript{b} Department of Law, Economics and Culture, Università degli Studi dell’Insubria, Como, Italy;

In the last decades, the definition of social well-being has been the topic of a controversial debate, due partly to a deeper awareness of societies, partly to a larger availability of information and data. A turning point in the debate has occurred with the so called Stiglitz Commission Report (2009), that invites institutions and analysts to “go beyond GDP” in the measurement of social well-being: this implies, mainly, the development of a range of indicators as clear and appealing as GDP, representing both objective and subjective assessment of well-being. At the same time, on the side of information availability, scholars have become familiar with a new and continuous flow of data coming from social network sites (SNS): these data are a huge source of information, almost free, and with an high time frequency. Despite several positive traits of SNS data, some criticisms - related to their non-random sampling origin - require statisticians to face new methodological challenges, in order to benefit from this rich source.

With the aim to propose a subjective well-being index exploiting new social media information, our team extracted from Twitter data a composite index that captures various aspects and dimensions of individual and collective life perception. The Subjective Well-being Index (SWBI) is a multidimensional indicator evaluated by applying iSA, a human supervised algorithm for sentiment analysis. It consists of eight dimensions that describe three different areas: personal well-being, social well-being and well-being at work.

The Italian Subjective Well-being Index (SWBI\textsubscript{ITA}), that we present here, overheard the Italian subjective well-being revealed by tweets acquired via the public Twitter API, written in Italian language, and posted from Italy from January 2012 to December 2017. Moreover, since around 1 to 5\% of the data includes geo-referenced information, it is also possible to provide an index at local level. In this work, we describe the subjective well-being in Italy at a regional level, weighting the SWBI\textsubscript{ITA} dimensions by two variables (broadband coverage and the Twitter rate), in order to partially overcome the selection bias caused by the use of SNS data. Moreover, we compare our evaluations with those provided by ISTAT in its “Aspects of daily life” report.

Because of the features of iSA algorithm, it should be noted that SWBI is not the result of the aggregation of individual well-being measurement, but it directly estimates the aggregate composition of sentiment within a society. Since, as previously pointed out, the traditional indicators of well-being often lack the subjective or self-perceived dimension, we propose this index as complementary statistics, to integrate the official measurement of social well-being.

**Keywords:** Well-being, Social media data, Sentiment analysis
Quantity and mood of final open-ended comments on an Erasmus+ VET mobility questionnaire

Marco Vivian\textsuperscript{a}, Elena Bortolato\textsuperscript{a} and Luigi Fabbris\textsuperscript{a}
\textsuperscript{a}Department of Statistical Sciences, University of Padua, Padua, Italy

VET international internships are work experiences carried out by students and apprentices in an abroad organisation, either a school, a training centre, or a company. In a CAWI survey carried out in 2018 in four European countries, the data from a large sample of participants who moved abroad to realise an Erasmus+ internship were collected. After a series of questions aimed to describe and evaluate their experience, interns were asked to suggest, with their words, possible ways to improve the aims or ease the mobility experience of future participants. Almost one third of respondents added their suggestions. This work is concerned with a categorisation of the obtained suggestions and the elicitation of the information contained in these “open-ended” responses that was not contained in previous “closed” responses. Moreover, the information offered and the mood shown by these collaborative respondents was correlated with some indicators of questionnaire quality. The analysis shows that suggestions are indeed relevant to understand the final evaluation given by participants about their internship. Moreover, the suggestions alone cannot subrogate in full the description of the experience as deduced from the other responses, even in case the verbatim description is long. As expected, the interest in the topical issue correlates with the number of words composing the suggestions.

**Keywords**: VET international mobility, Final suggestions, Open ended responses, Text mining, Interest in survey issue
Museum preferences analysis: an item response model applied to network data

Silvia Bacci\textsuperscript{a}, Bruno Bertaccini\textsuperscript{a}, Alessandra Petrucci\textsuperscript{a}
\textsuperscript{a} Dipartimento di Statistica, Informatica, Applicazioni “Giuseppe Parenti”, Università degli Studi di Firenze, Firenze (IT)

Firenzecard is the official museum pass of the municipality of Florence that allow the visit of more than eighty collections and exhibitions located in Florence and in the surrounding area. Firenzecard provides a huge amount of information concerning the paths of visits followed by tourists as well as some individual characteristics (e.g., country of origin). In this contribution we focus on the data relating to the 127,092 cards sold in the year 2018 which correspond to a total of 884,389 visits to museums. First, we use the instruments proper of the (social) network analysis (Kolaczyk, 2009) to provide a description of the relations among the museums in terms of tourists’ preferences. Second, we estimate a binary latent class item response model (Bartolucci et al., 2016) to detect unobservable (latent) classes of tourists that are homogenous with respect to their propensity to visit museums. In particular, this analysis is aimed at identifying museums whose attractiveness differs among latent classes of tourists.

Keywords: Item response theory, Latent class model, Network analysis

References


A multi-inflated hurdle regression model for the total number of overnight stays of Italian tourists in the years of the economic recession

Chiara Bocci, Laura Grassini, Emilia Rocco
Department of Statistics, Computer Science, Applications “G. Parenti”,
University of Florence, Florence, Italy

This contribution is concerned with the tourism behavior of Italian residents in the period covering the last economic recession: it investigates whether and how the economic recession has affected the total number of overnight stays in a quarter by modelling it through a hurdle multi-inflated regression model.

The assumptions of the hurdle model are consistent with the phenomenon under study, in which firstly a person decides whether to have a vacation trip and then, conditionally to a positive decision, he decides the number of overnight stays. Therefore the binary process concerning the decision to have at least a vacation in a given quarter is modelled through a logit regression model. Then the total number of overnight stays, for those who had at least a vacation, is modelled. Since this variable is naturally concentrated on some specific values (like 2, 6, 7, 14, 20 nights), we use a Multi-Inflated Truncated Negative Binomial regression model in order to control for this peculiar peaks.

We analyse data from the quarterly survey on Trips and Holidays in Italy and Abroad carried out by the Italian National Institute of Statistics, in the period 2004 to 2013.

The empirical results show that socio-economic characteristics of the individuals and of their families have an important effect on their tourism participation; that these factors, together with some trips-related characteristics, affect the total number of overnight stays; and that the economic recession impacted negatively on both aspects of tourism behavior.

Keywords: Count data, Multimodal distribution, Tuncated-at-zero models

References

Ecotourism and food geographic areas

Fabrizio Antolini\textsuperscript{a}, Francesco Giovanni Truglia\textsuperscript{b},
\textsuperscript{a} Department of Communication Sciences, University of Teramo, Teramo, Italy;
\textsuperscript{b} National Institute of Statistics, Istat, Agricultural Units, Roma, Italy;

The scope of this work is to analyse a particular segment of the tourist flows that are aimed at the farms-holiday. It is a specific area of hospitality that, thanks to the increasing sensitivity to environmental issues and aspects related to psychophysical well-being, is proving to be very successful and important both economically and culturally. The growing demand is fuelled by a radical transformation of the supply services by operators as shown by the local production of DOP-IGP products.

From the data of ISTAT surveys on farms and quality products it seemed interesting to analyse if at territorial level exists a possible spatial convergence, in specific geographic areas.

The second step of the work will compare these areas to tourism flows in the near municipalities. If the correspondence will be confirmed in terms of high flows of tourists, the local areas could be defined as touristic with a “food vocation”.

The analysis concerns the regions of Tuscany and Apulia.

\textbf{Keywords:} Tourism; Slow Tourism; Ecotourism; Food; DOP-IGP products.

\textbf{References:}
Computing ordinal consistency thresholds for pairwise comparison matrices

Pietro Amenta a, Antonio Lucadamo a, Gabriella Marcarelli a
a Department of Law, Economics, Management and Quantitative Methods, University of Sannio, Benevento, Italy

Pairwise comparisons are the most common tool used for representing preferences in multi-criteria decision problems. They are used, for instance, in Analytic Hierarchy Process (AHP) and its generalisation. An important issue related to pairwise comparisons matrix is the consistency of judgments: they may be irrational and/or intransitive. The consistency of judgments is strictly connected to the reliability of preferences. Less consistent judgments lead to poor priority vector estimates. If judgments are perfectly consistent, then all prioritisation methods give the same result. However, if they are not consistent, then each method derives a different priority vector. Given that perfect consistency is unattainable in practice, a degree of inconsistency is acceptable. It is, therefore, obviously necessary to measure the consistency of judgments before deriving the priority vector. Several consistency indices have been proposed in the literature to measure the level of inconsistency in a set of pairwise judgements. For instance, the Consistency Index (CI) and the Consistency Ratio (CR) that were introduced by Saaty, the Consistency Measure (CMK) that was introduced by Koczkodaj, the Consistency Measure (CMSH) that was proposed by Salo-Hamalainen, and the Geometric Consistency Index (GCI) that was proposed by Crawford and Williams. For these indices thresholds defining the consistency or inconsistency of the matrix have been proposed. The consistency indices and their thresholds may be useful to face cardinal consistency but they do not take into account the ordinal consistency (transitivity). In this work we focus on this aspect and propose transitivity thresholds that could be useful because they may provide meaningful information about the reliability of the preferences.

Keywords: pairwise comparisons, inconsistency indices, consistency thresholds, transitivity thresholds, Analytic Hierarchy Process

References

The odds ratio represents a simple and very popular measure of association between two or more categorical variables. In literature, a large amount of models and methods have been developed around this measure (Andersen 1980, Goodman 1985, Agresti 1996, Kateri 2014). Here our attention is to study the association among three categorical variables by a generalization of the log-ratio analysis (Aitchison and Greenacre 2002) which concerns the study between two categorical variables and is based upon odds ratios.

The main advantage of this new method with respect to the classical log-linear models (Andersen 1980, Agresti 1996) is to provide a quick, simple graphical summary of the global association. As a result, the visual interpretation of the association will be done using a distance and an inner product rule rested upon odds ratios (de Rooij and Heiser 2005). The point distances will be computed using the principal and standard coordinates resulting from the Tucker3 decomposition method (Tucker 1966; Kroonenberg 2008) of a particular three-way table.

As an example, we propose the three-way log-ratio analysis for assessing the players' performance of different Italian football clubs.

**Keywords:** Odds Ratios, Three-Way Contingency Table, Log-Ratio Analysis, Sport Performance.

**References:**


A composite indicator via hierarchical disjoint factor analysis for measuring the Italian football teams’ performances

Carlo Cavicchia a,b, Pasquale Sarnacchiaro a, Maurizio Vichi b

a Department of of Law and Economics, University of Rome Unitelma Sapienza, Rome, Italy; b Department of Statistical Sciences, University of Rome La Sapienza, Rome, Italy.

In the last years, with the data revolution and the use of new technologies, phenomena are frequently described by a huge quantity of information useful for making strategical decisions. In the current "big data" era, the interest of statistics into sports is increasing over the years, sportive and economic data are collected for all teams which use statistical analysis in order to improve their performances.

For dealing with all this amount of information, an appropriate statistical analysis is needed. A priority is having statistical tools useful to synthesise the information arised from the data. Such tools are represented by composite indicators, that is, non-observable latent variables and linear combination of observed variables. The strategy of construction of a composite indicator used in this paper is based on a non-negative disjoint and hierarchical model for a set of quantitative variables. This is a factor model with a hierarchical structure formed by factors associated to subsets of manifest variables with positive loadings.

In this paper, a composite indicator for measuring the Italian football teams’ performances, in terms of sportive and economic variables, is proposed.

Keywords: composite indicators, factor analysis, hierarchy, sports, football.
Analysis of the financial performance in Italian football championship clubs via longitudinal count data and diagnostic test

Anna Crisci\textsuperscript{a}, Luigi D’Ambra\textsuperscript{a}

\textsuperscript{a} Department Economic, Management, Institutions, University of Naples Federico II, Naples, Italy.

Football is undoubtedly the most powerful and most popular sport in Italy, linking communities and stirring emotions. The main goal of any Football Championship club is to achieve sport results. The study of the relationship between sport and economic results attracts the interest of many scholars belonging to different disciplines. Very informative is considered the connection, over short or long periods of time, between the points in the championship and the resource allocation strategies. The aim of this paper is to give an interpretation of this last link using the longitudinal count data. Some diagnostic test will be presented and used.

Keywords: Italian Football championship clubs, Longitudinal count data, random effects, Diagnostic test.

References:
Another look at the relationship between perceived well-being and income satisfaction

Omar Paccagnella\textsuperscript{a}, Ilaria Zanin\textsuperscript{a}

\textsuperscript{a} Department of Statistical Sciences, University of Padova, Padua, Italy.

Well-being is not a physical phenomenon that can be readily and objectively measured, especially if we place the individual at the center of our investigation. Faced with the impossibility of applying statistical indicators to abstract concepts, the subjective assessment of life satisfaction seems to be a valid criterion for attempting to quantify an individual’s well-being. Moreover, well-being can be globally considered as an aggregate concept that is explained by means of several domains.

Efforts to compare self-assessments relating to well-being, health, or other subjective aspects of quality of life or society often suffer from the fact that different groups of respondents use rating scales differently (DIF – Differential Item Functioning). Anchoring vignettes have recently been developed as a tool that enables to identify and adjust for the heterogeneity deriving from this issue. The underlying idea is to compare respondents’ self-assessments with a gold standard (or anchor) created within the context of a given investigation.

To this aim, in this work we aim at showing that analysing subjective well-being measures without correcting for the heterogeneity in the response scales may lead to some misleading conclusions. In particular, the relationship between life and income satisfaction is investigated by means of different bivariate analyses, exploiting the richness of SHARE (Survey of Health, Ageing and Retirement in Europe) data.

The main finding is that, correcting for DIF, the correlation between life and income satisfaction is lower than the one observing without correcting for DIF. Differences are not due to sample sizes of the analysed samples. At the same time, investigating uncorrected variables, DIF may introduce some forms of sample selection that have not reasons to exist.

\textbf{Keywords:} Anchoring vignette, Differential Item Functioning, Income satisfaction, Life satisfaction, SHARE data
Profile Patterns of Italians NEET by Nonlinear PCA

Anna Parola\textsuperscript{a}, Francesco Palumbo\textsuperscript{b}

\textsuperscript{a}Department of Humanities, University of Naples Federico II, Naples, Italy;
\textsuperscript{b}Department of Political Sciences, University of Naples Federico II, Naples, Italy

The current labor market complexity has had consequences for school-to-work transition. In Italian context, the group who during the critical period of the late teens spend a substantial amount of time outside any form of education, employment, or training (NEET) takes on ever-increasing aspects of risk of predicting and planning the working future (Savickas, 2012), and of psychological well-being (Parola and Donsi, 2018; Paul and Moser, 2009). Cluster analysis (CA) allows identifying homogeneous groups with respect the symptom-profiles related to the NEET condition. Generally, to get more stable results CA is performed on a reduced subspace obtained by a principal component analysis (PCA) on the original variable set (Fabrigar, Wegener, MacCallum, and Strahan, 1999; Linting, Meulman, Groenen, and van der Kooij, 2007). However, in the present work, we empirically show that in the analysis of our NEET data a nonlinear PCA according to Gifi (1990) offers better results, as it accounts non-linear relationships between variables. Such a non-linear association depends largely on the highly skewed variable distributions. Therefore, nonlinear PCA better highlights the different symptom profiles of the young NEET. In particular, we compare the results of the two-step analyses, PCA, non-linear PCA and CA, performed on the data collected through the administration of the Adult Self Report 18-59 (ASR; Achenbach and Rescorla, 2001) to a sample of 150 Italian NEET’s. Performing non-linear PCA, the basis expansion method is proposed to recode the original variables via a three-knots order two b-splines. Where the median corresponds to the central node. This approach is referenced in the literature as homals (homogeneity analysis, Gifi 1990). Then a hierarchical clustering algorithm is applied using as input variables the principal coordinates of the statistical units on the factorial subspace.

Keywords: Nonlinear PCA, Homogeneity Analysis, Clustering, Psychological well-being, NEET.

References:
Infertility is a major psychosocial crisis as well as being a medical problem. The factors that predict psychosocial consequences of infertility may vary in different gender, education level, socio-economic status. The primary purpose of this study was to investigate the relationship between sociodemographic characteristics and levels of depression and anxiety in infertile couples by exploring the role of each partner and of the related perceived levels of depression and of quality of dyadic adjustment.

This paper analyses these components by means of a mixture model for ordinal rating responses, allowing for uncertainty in answering (Piccolo, 2003). In responding to rating questions as the latent components which lead the perception of depression and/or anxiety, an individual may give answers either according to his/her feeling or to his/her level of indecision, typically motivated by a response style. Since ignoring this uncertainty may lead to misleading results, we define the distribution of the ordinal responses via a mixture model which weights both components in answering. The study allows also to model the actor/partner interdependence in case of categorical dyadic data by presenting an alternative approach with respect to the current used methods (see Kenny et al. (2006), among others).

The effectiveness of the model is attested through the analysis of a cross-sectional study of 206 infertile couples interviewed from 2014 to 2016. A gynecologist evaluated participants for demographic and medical data and then they were visited by a psychologist to perform questionnaire scales which were the Dyadic Adjustment Scale, the Edinburgh Depression Scale and the State-Trait Anxiety Inventory for the evaluation of the perceived levels of psychological disease (Zurlo et al, 2017, 2018)

**Keywords:** Ordinal scales, Mixture models, Latent variables, Infertility, Depression

**References**


Quantile Composite-based path modelling to handle differences in territorial well-being

Cristina Davino\textsuperscript{a}, Pasquale Dolce\textsuperscript{b}, Stefania Taralli\textsuperscript{c}, Domenico Vistocco\textsuperscript{d}
\textsuperscript{a} Department of Economics and Statistics, University of Naples Federico II, Naples, Italy; 
\textsuperscript{b} Department of Public Health, University of Naples Federico II, Naples, Italy; 
\textsuperscript{c} ISTAT, Ancona, Italy; 
\textsuperscript{d} Department of Political Science, University of Naples Federico II, Naples, Italy;

The Italian system of indicators on Equitable and Sustainable Well-being (Benessere Equo e Sostenibile - BES) proposed by the National Institute of Statistics represents a well-established reference database in the national and international debate on the research on alternative well-being measures. The main strengths of this set are represented by the broad coverage of all the components of this complex concept and the availability of information not only at the aggregate level but also at the provincial level (NUTS3 level) (Istat, 2019; Taralli et al. 2015). In this framework, it is possible to consider not only the levels of well-being but also the differences in their distribution thus highlighting differences in the territories.

The paper proposes an advancement of work elaborated in Davino et al. (2018), where a hierarchical model was used to study the relationships among components of the BES. The proposed hierarchical model allows us to synthesize individual indicators into single indexes, in order to construct composite indicators at a global and a partial level. Partial Least Squares path modeling (Lohmöller J.B., 1989) and a recent method, called Quantile Composite-based path modeling (Davino and Vinzi, 2016), were used respectively to estimate average effects in the network of relationships among variables and to explore whether the magnitude of these effects changes across different parts of the variables distributions. The present contribution aims to deepen the study taking into account that living conditions are quite different according to an unobserved or observed heterogeneity (for example according to the geographic area of the province).

Keywords: Well-being, Composite Indicators, Quantile Regression, Statistical models, Territorial differences.

References:
A comparison of different gender equality composite indicators: an application to Italian regions

Enrico di Bella\textsuperscript{a}

\textsuperscript{a} Department of Political Sciences, University of Genoa, Genova, Italy

Gender equality is a topic of growing relevance on the international panorama both on the political and on the academic perspective. While it is important to define indicator systems that take into account the complexity of the phenomenon, it is also necessary to create synthetic measures that facilitate communication and comparison between different statistical units. However, composite indicators built on the same data can lead to different results depending on the construction technique used for the composite indicator. If, for example, the objective is to create a ranking of statistical units (countries, regions, municipalities, ...) on the basis of a set of variables or indicators, the choice of the technique to be applied becomes an element that can influence the final order of statistical units and consequently also the possible policy decisions.

Using a set of data on gender equality for the Italian regions derived from the original set at European level of the European Institute on Gender Equality (EIGE), in this work different approaches to the construction of a composite indicator of gender equality in Italy are compared. The first approach that is recalled, and that covers the role of benchmark, is the replicate at regional level of the one used by EIGE at national level. After, some techniques for the construction of a synthetic indicator of gender equality, alternative to the one proposed by EIGE, are then briefly recalled. The first set of techniques used is linked to the theory of partially ordered sets (POSAC and POSET) which are not aggregative by nature and which try to keep separate all the information inherent to each indicator. The second set of techniques refers instead to the approach of the Mazziotta-Pareto Index and the approach of the Benefit of Doubt. Although the gender equality data can be described by a low number of dimensions, the results obtained by applying different techniques lead to different orderings of Italian Regions. This highlights the non-neutral role with respect to the results of the researcher's analysis and choice of the indicator's construction technique. Finally, the work proposes a brief reflection on some operational limits of the theory of partially ordered sets for the construction of composite indicators with some proposals on how they can be overcome by placing them in relation to other more traditional techniques.

Keywords: Gender Equality, Indicators, Partially Ordered Sets, Regional Studies

References:
A composite indicator to stratify old people resident in Piedmont according to their frailty level

Margherita Silan\textsuperscript{a}, Giovanna Boccuzzo\textsuperscript{a}, Giada Signorin\textsuperscript{a}
\textsuperscript{a} Department of Statistical Science, University of Padua, Padua, Italy;

The average age of the Italian population is rising steeply. This implicates an increase also of chronic diseases attributable to ageing. In this framework, the planning and the management of health care resources and policies is a huge challenge for the country's health system. The National Program for Chronic Diseases is focused on stratifying the population according to health care needs. Frail subjects are often older, they have special and wider care needs, but not always they are known and assisted by the health and social services. This is why their identification and the quantification of their frailty level become really important in order to improve both the distribution of public resources and the quality of life of older individuals.

The aim of this work is to construct an indicator that measures the frailty level of each individual in the population using administrative health data-flows of the Piedmont region. Frailty is a multidimensional concept and there is no agreement on a unique definition for it. However, some definitions proposed in literature mention the fact that a frail individual is associated to an increased risk of adverse outcomes. Thus, the indicator we propose classifies individuals with respect to their probability of experimenting adverse outcomes, such as death, access to the emergency room with red code, emergency hospitalization, avoidable hospitalization, hip fracture, disability. Through the inclusion of multiple health outcomes, the indicator achieves the additional goal of representing the complexity of frailty condition.

The first challenge we faced consists in variables selection, the main goal is to identify a parsimonious set of variables that is also able to predict simultaneously the six outcomes we considered. We used 100 logistic regression models, for each outcome, with stepwise selection on balanced sub-samples of the total population. The final set of variables were then selected observing for every variable both the median order of entrance in the model and the percentage of presence of the variable in the models. It includes 7 variables that assemble the frailty indicator through partially ordered set (poset) theory (Silan, 2019).

Variables included in the indicator are age, poliprescriptions, Charlson Index, disability, number of accesses to the emergency room with yellow code, accesses to the emergency room with green code and the presence of the Parkinson disease.

The indicator is associated to high AUC with respect to all the six outcomes (0.82 for death, 0.71 for access to the emergency room with red code, 0.70 for emergency hospitalization, 0.77 for avoidable hospitalization, 0.75 for hip fracture and 0.78 for disability).

This indicator can be particularly useful for Italian Local Health Units because it is built with a small set of variables obtained from current administrative health care data flows, available in all Italian Local Health Units. Moreover, even if used methods are quite sophisticated, the indicator results to be easy to calculate thanks to the construction of a user-friendly application that returns the individual frailty indicator after a guided loading of a suitable input data file.

Keywords: Frailty indicator, Administrative healthcare data, Poset theory, Measurement, Aging.

References:
Missing values in Social Media: an application on Twitter data

Paolo Mariani a, Andrea Marletta a, Nicholas Missineo a
 a Department of Economics, Management and Statistics, University of Milano-Bicocca, Milano, Italy

The main aim of this work is to study the behaviour of users in social media analysis. In particular, the work involves Twitter users about the perception of the Italian guaranteed minimum income (“Reddito di cittadinanza”) on the basis of different categories of users. The main distinction about users is made between verified Twitter users and not verified users. The first category is related to politicians, institutional authorities and other official stakeholders. The second one is represented by citizens and other subjects not directly involved in the process of realization of this measure. A classification method based on tweets, retweets and quotes posted by users with hashtag #redditodicittadinanza will be able to discern between verified and not verified users. Moreover, an analysis of the KPI (Key Performance Indicators) will be conducted using their presence and absence through the use of the complementary values. This tool is very useful to give a meaning to an absence of behaviour distinguishing between no interest and a negative opinion. Data have been collected in Italy in April 2019 using all tweets containing #redditodicittadinanza using the official Twitter API. Data will be analysed using R and Python. The use of this classification method allowed also to propose a model to recognize the presence of a Bot, a simulated Twitter user that login using same channels of the humans. Bot are often used to spread the effect of messages and topic in order to influence positive or negative effects.

Keywords: Missing values, Twitter, Social Media analysis, guaranteed minimum income

References

Finger, L (2013), Ask, Measure, Learn: Using Social Media Analysis to Understand and Influence Costumer Behaviour.
Harmonised Administrative Databases: a new approach in the era of Big Data

Vittorio Nicolardi^a, Caterina Marini^a

^a Department of Economics and Finance, University of Bari Aldo Moro, Bari, Italy;

The real challenge that in the nowadays society needs to be scientifically faced is to accurately handle the enormous flow of information that in an IT world can be tremendously powerful to analyse the social and economic changes. The official statistics that the National Institutes of Statistics yield are facing their limits in all the research areas and the problem is already recognised and discussed in the international scientific context. Undoubtedly, the Public Administration datasets can be a very useful source of additional and detailed data to complete the statistical information about phenomena that are still partially depicted by means of the official data. In some cases, limitations in dealing with the real size of many socio-economic territorial developments are significant. In this context, however, it is important to underline that the administrative databases present two main problems that need to be considered: their purpose, which is not statistical, and the huge amount of data they store. An additional problem can be the entirety of the administrative data on the phenomenon analysed when its complete information requires the merging of two or more databases that often belong to independent PA offices. In this paper, we focus the analysis on the Italian real estate phenomenon and how the administrative data are powerful in adding new information on the phenomenon in terms of both volume and value. In particular, the analysis is circumscribed to the city of Bari, in the South Italy because it is part of a national research project. We built a unique administrative database starting from 4 independent administrative databases, normally managed by independent PA offices (the Italian Real Estate Registry and the National Revenue Agency), which provide autonomous information. The record linkage has required the basic practise of the big data methods to deal with both missing data, duplication and erroneous information, and the identification of the best variables to merge the 4 data sources. Although we have restricted the analysis to one city, the amount of data has also required the application of GIS processes to guarantee the exact matching of data and depict the real estate framework in detail. In fact, the results of our work allow researchers and policy makers to deeply analyse the territory, even the single building. And the differentials between the real estate market monetary values and the real estate values have shown significant results in terms of potential revaluations of city districts. The importance of the analysis we yielded in this paper is unique and original in its attempt to describe an economic phenomenon that, in Italy, still suffers the consequences of the dearth of a complete and harmonised data-warehouse. Our work is the first attempt in this sense and the outcomes can be perfectly replicated in any dimensional territorial area.

**Keywords:** Administrative database, Official Statistics, Big Data, Real Estate.

**References:**


Inferring Twitter users home location based on trend topics

Paola Zola a, Costantino Ragno b, Paulo Cortez c

a Department of Economics and Management, University of Brescia, Brescia, Italy;
b School of Science and Technology, University of Camerino, Camerino, Italy;
c Department of Information Systems, University of Minho, Guimarães, Portugal.

Inferring Twitter users home location is a growing interest research topic given its importance for many decision support system. For example knowing the people location is fundamental in event detection studies, recommendation systems, friendship network analysis and so on. Some studies have been done in order to estimate the users’ home country given the world distribution (Zubiaga et al., 2017) while other focus on more fine grain location prediction relatively to a specific country or region (Eisenstein et al., 2010).

In this study we propose a novel approach that aim to infer Twitter users home location at the finest grain (coordinates) considering all the world surface. The proposed approach follows the one in Zola et al. (2019) where for each Twitter account the nouns distribution is evaluated and, passing thought Google Trends data, the estimation is performed. However, while in (Zola et al., 2019) the prediction is determined at country level, in this research we apply clustering algorithms to assign to each user a probability distribution over the whole world area in order to identify an unique couple of coordinate associated to the user home location.

The dataset is composed by 3,298 Twitter account and for each we consider his/her historical 3,200 tweets. For each account we extract the cities distribution from Google Trends using as keyword the nouns (generic and proper) hold in the tweets collection. Having the cities distribution we define the cities polygons sampling points in order to respect the Google Trends data distribution. Then two clustering algorithm are performed: Gaussian Mixture Models and k-means model. The results are evaluated using mean and median absolute error computed on Euclidian and Haversine distance from the ground truth users home location derived in the work of Zola et al. (2019).

Keywords: Geoparsing, Natural Language Processing, Google Trends.

References


Invariance in the structural topic models

Emma Zavarrone a, Maria Gabriella Grassia b, Rocco Mazza b

a Department of Business, University IULM, Milan, Italy;
b Department of Social Sciences, University of Naples Federico II, Naples, Italy;

Twitter is now a platform that hosts a vast community of users and, as such, can be used as a valuable research tool on the many themes of discussion that animate the network, in fact in recent years a large number of analysis tools have been developed that allow explore the meaning of these contents. The researcher who applies these methods cannot ignore the role of keywords used for the extraction and collection of texts. The choice of selected search keywords is crucial in order to identify the corresponding semantic content, however the presence of single word in a group of tweets could not assure the capture of relevant latent construct. We propose an innovative approach through structural topic models (Roberts et al., 2014) for the study of semantic invariance between the selected keywords and the analysed conceptual dimension. In detail, the research questions concern: 1) the identification of topics sequences conceptually connected to the latent semantic dimensions, 2) development of construct invariance forms in order to measure the invariance (Cheung et al., 2002) among the tweets groups. The model proposed in this paper is applied to a dataset composed of 325111 tweets focused on the violence against women and women hate speech (Malmasi and Zampieri, 2018). Italian tweets were extracted using specific search keywords, based on previous content analysis, in the time period from July 2018 to May 2019. Latent dimension that one wants to trace is the women hate speech. From these texts, subsequently the cleaning and pre-treatment phases, through the application of a structural topic modelling, the latent topics are extracted and in conclusion the invariance between the different constructs of the model is studied.

Keywords: Construct Invariance, Structural Topic Model, Hate speech, Textual analysis

References

Elderly with and without Children: Do They Report Different Health Conditions?

Chiara Bocci, Silvana Salvini
Department of Statistics, Computer Science, Applications “G. Parenti”,
University of Florence, Florence, Italy

The world is aging rather rapidly. As both the proportion of older people and the length of life increase throughout the world, key questions arise. Will population aging be accompanied by a longer period of good health, a sustained sense of well-being, and extended periods of social productivity, or are associated with more illness, disability, and dependency? And which is the relationship between childlessness and health, in a world where fertility does decline?

We intend to study the relationship between childlessness and perceived health for elderly in four European countries: France, The Netherland, Poland and Italy. While Italy is a familistic society, the other three countries present few intergenerational exchanges and an early exit of children from the parental home. We have chosen to analyse these countries, which are characterised by different culture and welfare systems, to understand if the context is important to determine the relationship. The influence of family behavior on health in old age has been increasingly recognised in literature. Apart from the physiological and psychological effects of pregnancy and childbirth, the health of both women and men may be influenced by stress, role changes, and changes in allocation of personal and family resources associated with child-rearing and by the emotional and social support benefits of parenthood.

To examine France, The Netherland and Poland we use data from the Generations and Gender Surveys (for the years of 2005, 2002-2004 and 2010-2011, respectively), while for Italy we use the Italian Multipurpose Survey carried out in 2009. Even if the Surveys collect information on the whole adult population, we focus our analysis on people aged 50 and over, that are 18,033 in Italy, 3,331 in Netherland, 10,402 in Poland and 4,446 in France. To investigate the possible relation between the perceived health status of old people and the number of children (controlling for age, sex, marital status, education, work status, region and ownership of household) we apply a multinomial logistic regression.

Generally, in every country we examined, we did not observe a significant association between the number of children and the perceived health of elderly, whereas covariates are often linked with health by a strong relationship. In summary children, and this is the result we outline, do not have any impact on health, but for a weak positive effect of having 1 child (a lower probability of “fair” vs “good” health). Therefore, people with or without children do not seem to perceive a different health status.

Keywords: Elderly health, Children, European countries, Multinomial regression

References

Short-run and Long-run Persistence of Bad Health Among Elderly

Daria Mendola\(^a\), Paolo Li Donni\(^a\)

\(^a\) Department of Economics, Business and Statistics (SEAS), University of Palermo, Palermo, Italy.

The assessment of elderly health and quality of life is receiving increasing importance in both social and economic health policy planning for which longevity is not the only primary goal, while well-being is nowadays assuming a central role. In 2016 21.8 per cent of non-institutionalized persons aged 65 and over are in fair or poor health in US, and 6.4% need help with personal care from other persons (National Center for Health Statistics, 2017). Persistence in bad health (not limited to chronic disease) is one of the main assailant to the elderly chance to a successful aging, threatening elderly ability to fulfil occupational, social and family roles. We study the health dynamics among older Americans using ten waves of the Health and Retirement Study comparing two approaches dealing with the measurement of persistence: a spell-approach and a regression-based approach. The former is fully non-parametric synthesizing the sequences of health status into a Health Persistence Index (HPI) drawn by Mendola et al. (2011). The latter approach relies on a latent Markov model (Bartolucci and Farcomeni, 2009) capturing persistence in poor health by modelling time-varying unobserved heterogeneity. Our results show that only few elders experiences persistently a poor health status. Three latent states are identified. The first state is characterized by a lower probability of reporting poor health over time as compared to the other states; the second one has a lower probability to report poorer health over time as compared to the third; and the third one refers to individuals frequently sick, hence with the highest propensity of being sick. In particular we observed a decreasing relationship with education level indicating that better educated elders are able to live a healthier life. Persistence in bad health decreases with income and increases with lowering of activity intensity in the labour market. The U-shaped relationship observed with the individual weight indicates that persistence in bad health increases with any deviations from the regular weight interval. Finally the lower the labour intensity the higher the persistence in bad health. Consistently, higher values of the HPI are observed with the main socio-demo-economic risk factors, indicating that individuals with lower education, unhealthier life styles and limited participation in the labour market experience higher persistence in bad health. Hence, interestingly, the factors affecting poor health persistence do not differ from a short-run to a long-run perspective. This suggests that whether a factor is associated with poor health in the short-run, it is also likely to be persistently associated with subsequent health statuses in the long-run period.

**Keywords:** Health and retirement study, Older Americans, Latent Markov model, Health persistence index, State dependence

**References:**


Selecting Features for Machine Learning in Alzheimer’s Diagnostics

Luiz Sá Lucas\textsuperscript{a}, Rafaela Bueno\textsuperscript{a}, Ana Carolina Lucas\textsuperscript{a}

\textsuperscript{a} Theopinionedge, Brazil;

Alzheimer’s disease (AD) is the most common cause of neurodegenerative dementia. Despite its higher prevalence in the older population, it is still the most frequent form of dementia under the age of 65 (Tábuas-Pereira et al., 2016). Clinicopathological studies suggest that Alzheimer’s disease (AD) pathology begins 10–15 years before the resulting cognitive impairment draws medical attention. Biomarkers that can detect AD pathology in its early stages and predict dementia onset would, therefore, be invaluable for patient care and efficient clinical trial design. It is possible to use a targeted proteomics approach to discover novel cerebrospinal fluid (CSF) biomarkers that can augment the diagnostic and prognostic accuracy of current leading CSF biomarkers (Ab42, tau, p-tau181) (Craig-Schapiro et al., 2011). On the other hand, usually machine learning datasets, such as the one studied in our paper, have a number of features that are too numerous for a good model performance. The present paper aims to compare several feature selection techniques in this area: simulated annealing optimization, a variable clustering tool that we developed for the paper, importance of variables as given by the Random Forest algorithm, independent component analysis, selection using the literature etc. We used the dataset given in the AppliedPredictiveModelling R package (Kuhn and Johnson,), also used in the “Multiplexed” study. Two indicators were used to measure and compare the methods - Accuracy and Kappa. As a result, it is expected to: (i) present a useful method combining the described feature selections techniques; (ii) give a warning about performing diagnosis based only on machine learning tools (the so called ‘weapons of math destruction’).

\textbf{Keywords:} Machine learning, Feature selection, Alzheimer diagnosis

\textbf{References:}


The use of play-by-play data analysis for team managing in basketball

Luca Grassetti\textsuperscript{a}, Ruggero Bellio\textsuperscript{a}, Giovanni Fonseca\textsuperscript{a}, Paolo Vidoni\textsuperscript{a}

\textsuperscript{a} Department of Economics and Statistics, University of Udine, Udine, Italy.

The sports analytics literature regarding basketball is vast but the analyses based on disaggregated data, such as the play-by-play match data, are not very common. The analysis of the whole sequence of play-by-play match events has an undeveloped potential, yet most of the available methods focus on the final match results. The present work aims at finding a model-based strategy for the analysis of the match progress, which may support the decision-making process of the team staff and, to a lesser extent, be used for the prediction of match results. The proposed approach is built upon the literature of Adjusted Plus Minus (Rosenbaum, 2004) and of its regularized version (Engelmann, 2017). The main aim of these methods is to describe the outcome of the observed plays (the scored points) as a function of the efficiency of the players on the field and, eventually, of some contextual variables. The RAPM approach is extended here in two main directions. The first extension consists in the adoption of a response variable which considers the most relevant events in the game, and not only the number of scored points. This offers some useful advantages, including the possibility of obtaining separate estimates about different complementary aspects. Further, next to player efficiency effects, the efficiency of five-man lineups is estimated. An ulterior peculiarity of the present work regards the adoption of the empirical Bayes approach to model estimation, which provides a suitable regularization.

For the empirical analyses, we consider a dataset regarding the Italian Basketball League (Serie A1), focusing on the matches of the first round of the 2018/2019 championship. The dataset collects the play-by-play information and the matches box scores, which are made available by the league website (www.legabasket.it). To gather the required data we use the R statistical software (R Core Team, 2018). In particular, the packages rvest (Wickham, 2016) and stringi (Gagolewski et al., 2019) are used for scraping the necessary information from the web-pages and for data-cleansing process, respectively. To consider a more manageable outcome measure, the plays are then aggregated by shifts. The shifts can be defined as periods between subsequent substitutions. The results of the analysis could be used to support the decision-making process of team management and some illustrations on this point are provided.

Keywords: Basketball Analytics, Statistical Model, Play-by-play data, Web-crawling, Data-driven decision process.

References

Wickham H. (2016). rvest: easily harvest (scrape) web pages. R package version 0.3.2. \url{https://CRAN.R-project.org/package=rvest}
Assessment of game actions performance in water polo: a data analytic approach

Alessandro Lubisco\textsuperscript{a}, Stefania Mignani\textsuperscript{a}, Carlo Trivisano\textsuperscript{a}

\textsuperscript{a} Department of Statistical Sciences, University of Bologna, Bologna, Italy;

The aim of this study is to analyse the performance of an Italian second division water polo team at the end of the 2018 national championship. In the literature the previous studies are dedicated to matches from major tournaments (Olympics, World and European Championship). In this work we compare the team’s offense and defence performance in 17 matches (12 at home and 5 away). Data have been collected during the match through an original program for water polo match analysis specifically designed and developed in MS Excel by the first author. A subsequent video analysis was required to supplement any data not detected during the match. The dataset consists in more than 1800 plays distinguished by playing situation (even, counterattack and powerplay), area of the field where the event occurs, outcome (goal, penalty, exclusion, no goal shot, lost possession,…), quarter of the match, difference between the teams in terms of goals and exclusions and so on. A first descriptive analysis shows that 49.3\% of the goals resulted from powerplay (41.9\%) or penalty (7.4\%) situations, while, respectively 33.8\% and 16.9\% resulted from even and counterattack situations, percentages are not so different as compared to those observed in top team international tournaments. We estimated a logistic regression model to evaluate which variables affect the outcome of the play, defining it as “GoodOutcome” (goal, penalty or exclusion) and “BadOutcome” otherwise (lost ball, shoot out or saved,…). At the end of the championship, it turns out that the analysed team exhibit the best defense among the 12 teams, and the 5\textsuperscript{th} best attack. The model results underline that, when the team is in an offensive phase, the probability to complete a play with a GoodOutcome is higher when the event occurs near the two-meter line in front of the goalkeeper, between the two posts, where usually plays the center forward. This happens also during a powerplay or counterattack situation. This could appear obvious. But, when the team is defending, something different happens. There are only two significant coefficients. The first coefficient refers to the playing situation and the positive sign meaning that, all else being equal, the GoodOutcome is more likely for counterattack if compared with other situations. The other significant coefficient is related to the covariate called Zone 2, usually occupied by a left handed player ant it is negative sign meaning that for the opposite team the GoodOutcome is less likely rom that position if compared with other positions. In fact, the strategy of the analysed team during all the year was to force the opponents to try a shot from that position. This seemed to be a good strategy, resulting in the best defense of the championship. This model, even if suffering from the heterogeneity due to the presence of many factors is a promising tool as confirmed by a percentage of cases correctly classified around 70\%.

Keywords: Waterpolo analytics, Match analysis, Tactical indicators, Logistic regression

References:


Athletes’ mental skills, personality and other drivers to assess the performance in a study on volleyball

Daniela Caso\(^a\), Maria Iannario\(^b\), Francesco Palumbo\(^b\)

\(^a\) Department of Humanities, University of Naples Federico II, Naples, Italy; 
\(^b\) Department of Political Science, University of Naples Federico II, Naples, Italy.

Several studies investigating the relationship among sports performance, personality and athletes’ mental skills have suggested the positive relationship between self-esteem and conscientiousness to identify higher performance and a low level of neuroticism to predict their athletic success (see Allen et al. (2013), among others). This work is aimed at studying personality measured with the five-factor inventory (Costa and McCrae, 1992) and mental skills assessed through the sport performance psychological inventory (IPPS-48) (Robazza et al., 2009), and how they can successfully affect the performance of team athletes. Moreover, measured with Rosenberg’s Self-Esteem Scale (Rosenberg, 1965), the self-esteem was considered to evaluate the athletes’ unidimensional feeling about the self in the sport context. The analysis was based on a mixture model that allows identifying the level of feeling of respondents and possible heterogeneity in the responses (Piccolo, 2003). It is about a sample of young female athletes enrolled in women’s volleyball teams of several Series. By using the mixture model we explored not only some psychological traits that reveal new elements to assess the performance but also other drivers as the practice of other sports or time dedicated to the study, among others.

**Keywords:** CUB models, Ordinal data, Sport psychology.

**References**


Dynamic of efficiency: an analysis of Italian geriatric wards

Marcella Mazzoleni a, Mariangela Zenga a

a Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milano, Italy.

The attention placed on healthcare systems is increasing constantly, in recent years. In fact the definition of "quality" plays an important role in the work of evaluating the healthcare service. This paper aims to investigate the relationship between this combination of effectiveness and appropriateness in the geriatric wards in Italy. Firstly an analysis regarding the healthcare outcomes related to the principal business healthcare indexes will be conducted in order to produce a rank of those geriatric wards into a league table. Secondly the paper will investigate the problem for appropriateness in hospitals wards. The analysis shows the dynamics of the efficiency and appropriateness for the geriatric wards respect to 4 years.

Keywords: Health care, Geriatric Wards, Appropriateness.

References

Spatial inequality and urban poverty: Evidence from U.S. metropolitan areas

Francesco Andreoli, Mauro Mussini, Vincenzo Prete

Department of Economics, University of Verona, Verona, Italy;
Department of Economics, Management and Statistics, University of Milan-Bicocca, Milan, Italy;
Department of Economics, University of Verona, Verona, Italy;

Income inequality has increased in most of American cities in the last decades (Watson, 2009), even though with different patterns and trends between and within cities. Inequalities between neighborhoods of a city are often substantial (Wheeler and La Jeunesse, 2008), with some neighborhoods having a high proportion of poor individuals and others with a low proportion of poor people (Andreoli, 2018). This poses the question of which indexes are suitable to measure poverty in the whole city and at the same time to take heterogeneity in the distribution of poverty across city neighborhoods into account. A common approach to urban poverty measurement is the use of the concentrated poverty index, which expresses the proportion of a city’s poor population living in high-poverty neighborhoods. We argue that such an approach has some limitations since (i) heterogeneity in the distribution of poor people across neighborhoods is not fully evaluated, (ii) spatial clustering of neighborhoods with similar poverty incidence is not captured. To overcome such limitations we develop a decomposition framework for examining urban poverty which can capture several aspects of the distribution of poverty across neighborhoods and the components of the change in urban poverty over time. We show that changes in urban poverty in a city are additively decomposable into the contributions of demographic, convergence, re-ranking and spatial effects. Patterns and trends of urban poverty across American cities over the last 35 years are analyzed.

Keywords: Concentrated Poverty, Decomposition, Inequality, Urban Poverty

References

Socioeconomic inequalities and cancer risk: the challenges and opportunities of worldwide epidemiological data consortia

Carlotta Galeone\textsuperscript{a}, Rossella Bonzi\textsuperscript{a}, Claudio Pelucchi\textsuperscript{a}, Federica Turati\textsuperscript{a}, Carlo La Vecchia\textsuperscript{a}

\textsuperscript{a} Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

There is increasing evidence that low socioeconomic position (SEP) is a strong determinant of morbidity and premature mortality from selected non-communicable diseases, including several cancers. SEP reflects the availability of cultural, material and social resources that translate into advantages in terms of decision-making, social network, lifestyle habits and health services access. Thus, an accurate quantification of the impact of SEP on cancer risk is of major importance to plan public health interventions for cancer incidence and socioeconomic disparities reduction. During the last decade, the advent of collaborative and interdisciplinary research with the proliferation of multi-institutional consortia allowed to define and quantify the associations of interest with a higher degree of accuracy, to explore subgroups of the population and to investigate the interactions between environmental, genetic and socio-economic factors. The Stomach Cancer Pooling (StoP) Project and the International Head and Neck Cancer Epidemiology (INHANCE) are two large data consortia, in which the University of Milan is proactively involved. Their large sample size allowed investigators to address the effects of education and household income on the onset and evolution of the disease. INHANCE findings suggested that low education and low income are risk factors for head and neck cancer, independently of tobacco smoking and alcohol consumption. The collaborative pooled-analysis within the StoP consortium showed a strong inverse relation between SEP indicators and gastric cancer risk, with a ~40% decreased risk among individuals with intermediate/high education status as compared to less educated study subjects. In conclusion, SEP is a strong determinant of cancer. Effective interventions to reduce socioeconomic inequalities at local, national and international level are needed to reduce cancer risk among the more vulnerable groups of the population. Being cancer strongly related to low SEP, these interventions will reduce the burden of the disease in the whole population.

**Keywords:** Socioeconomic inequalities, Cancer, Risk factors, Consortia, Epidemiology

**References:**


A data analytics framework:
medical prescription pattern dynamics

Ilaria Giordani¹, Gaia Arosio², Ilaria Battiston³, Francesco Archetti⁴,⁵

¹ Department of Computer Science, Systems and Communication, University of Milano
Bicocca, Milano, Italy;
² Consorzio Milano Ricerche, Milano, Italy;

Prescription pattern monitoring studies (PPMS) exploits medical information to improve the prescribing practices and thus the standards of medical treatments at all levels of healthcare. Data analytics can give a major contribution to PPMS, tackling health and socioeconomic challenges: by means of insights originated from analytical results health authorities can plan informed actions, promote appropriate use and reduce the abuse/misuse of monitored drugs. This study summarizes relevant insights obtained by the analysis of data related to about 1,500 general practitioners (GPs) and around 1,015,000 patients, during the period 2000 to 2018. The available database was collected, managed and complied with current privacy regulations. It contains pseudonymised recorded medical data about outpatient visits at GPs’ ambulatories located in several Health Districts within the Campania region in the southern Italy. After a quality assessment of the whole database, the sample was reduced to about 720,000 patients, balanced in terms of gender, age, diagnosis and prescriptions. This study presents a data analytics framework composed by different computational modules. Starting from the explorative and descriptive analytics modules, the data analysis process allows the reconstruction of diagnosis and prescriptions global trending, extracting patient journeys and underline changes. Through the “patient journey” component the framework can reconstruct the patient’s history and relationships with primary healthcare, identifying prescription patterns and changes. Finally, a deeper insight on antibiotic prescriptions dynamics, one of the major challenges to public health despite guidelines and calls by government agencies, is presented. The prescription patterns of the 5 most frequently prescribed antibiotics is analysed highlighting the heterogeneous trends.

Keywords: Prescription patterns, Antibiotic prescriptions dynamics, Patient journey.

References:
Applying network modelling to uncover the relationships among well-being dimensions

Laura Giuntoli\textsuperscript{a}, Giulio Vidotto\textsuperscript{a}

\textsuperscript{a}Department of General Psychology, University of Padova, Padova, Italy

Confirmatory factor analysis (CFA) is the most widely used statistical method to test the structure of psychological constructs, but it requires strong a-priori hypotheses on the model structure and at least three indicators for each latent variable. Furthermore, CFA falls short in that any two observed variables are conditionally independent given the latent variable (i.e., local independence). Indeed, such assumption is frequently unlikely to hold for items depicting a psychological construct and this is particularly true for the well-being construct. Individual well-being can be interpreted as the amalgam of the different components proposed in the literature by the hedonic stream, measuring positive emotional states, and the eudaimonic stream, that proposed several dimensions of optimal psychological functioning such as self-acceptance, meaning in life and positive relationships. According to the analogy with mental illness proposed by Keyes (2002), mental health may be operationalized as an emerged condition based on the concept of a syndrome of symptoms of individuals' subjective well-being, that is individuals' perception and evaluation of their life in terms of affective states and psychological and social functioning. Thus, well-being can be conceived as a network of interacting and self-reinforcing symptoms, not an underlying entity that produces symptoms (i.e., positive individual traits and experiences). For instance, sense of competence may activate other positive psychological features (e.g., self-acceptance and engagement), likely in circular, self-reinforcing ways. In this view, hedonic and eudaimonic dimensions of well-being are not mere passive psychometric indicators, but are active causal ingredients of mental health. Probably, the latent variable approach is not the best instrument to test such a theoretical model of well-being. Instead of pursuing a theory-driven approach, the structure of well-being dimensions can be retrieved by means of network psychometrics (Epskamp et al., 2017) that is a highly informative, data-driven approach that allows the model structure to spontaneously emerge from the relationships among indicators. Network modeling is based on the estimation of the weight matrix, that depicts the relationships among individual indicators as partial correlation coefficients. Strength centrality, also named Expected Influence (EI), represents the influence of each individual observed variable on the whole network. To describe the dimensionality of a psychological construct, we propose a strength centrality decomposition that is able to summarize the influence of a specific indicator within its dimension (within-EI) and between the other dimensions (between-EI). We provide a discussion of the proposed formulae and a practical example using the well-being operationalization by Diener et al. (2010).

**Keywords:** Network analysis, Expected Influence, within-EI, between-EI, Diener’s model.

**References:**


Evaluating Health Performance and Inequalities in Marche Region of Italy

Alberto Franci\textsuperscript{a}, Pietro Renzi\textsuperscript{b}

\textsuperscript{a} Department of Economy, Science and Law, University of the Republic of San Marino, Rep. San Marino - Department of Sociology, University of Urbino “Carlo Bo”;
\textsuperscript{b} Department of Economy, Science and Law, University of the Republic of San Marino, Rep. San Marino;

A worldwide selection of models for measuring performance in health services was appraised, with the internationally recognised Health System Performance Assessment tool chosen for testing in a local health authority in the Marche Region of Italy, utilising local, regional, national and international comparisons.

A complementary means of measuring health inequality involving the Concentration Index enabled a holistic evaluation of the local health environment.

Whilst the approach addressed a comprehensive range of issues, limitations with data availability were found to present genuine constraints that require future action. Nevertheless, valuable lessons were learned for policy makers, with the relationship between socio-economic inequalities and systematic variations in health indicators highlighted.

The Health System Performance Assessment tool presents the opportunity for strategic alignment in performance measurement. Our article presents an overview of extensive piece of research.

**Keywords:** Benchmarking, Concentration Index, Health Inequalities, Health System Performance, Assessment Tool.

**References:**


The determinants of physical inactivity throughout the analysis of well-being and sustainability indicators: the Italian BES framework

Laura Murianni\(^a\), Rita De Carli\(^a\), Emanuela Bologna\(^a\), Laura Iannucci\(^a\), Alessandra Tinto\(^a\)

\(^a\) The Italian National Institute of Statistics- Istat

Regular physical activity is associated with a healthier, longer life. Physical inactivity is widely recognized as a major risk factor for chronic diseases, and ranks among the most important risk factors in contributing to the population burden of disease in western societies. Physical inactivity is currently identified as a major contributor to the increasing levels of obesity, and other serious medical conditions. The increased political, media and scientific interest in obesity since the late 1990’s has placed physical activity high among current public health issues. To act for the reduction of physical inactivity is one of the priorities fixed by the WHO in many countries in the International First Global Plan. In Italy in 2017 37.9% of the population is sedentary, the value has started slightly declining since 2013, but it is still high and many other actions are needed to reduce it further. In Italy in 2010 the National Statistical Office Istat started a joint initiative with the National Council for Economics and Labour with the aim of setting up a framework for the measurement of Equitable and Sustainable well-being in Italy (Bes framework). The framework is structured in 12 domains, measured through a total of 130 indicators. One of the domains is the Health domain, in which physical inactivity, excess body weight and other life style behaviours are considered as fundamental elements to describe all the components of the Health Dimension (Stiglitz et al., 2009).

In this study we analyze the relation between life style indicators (among which physical inactivity, smoking and BMI) and well-being indicators, with the intent of showing the impact of improvement of life style behaviours on well-being. Estimates of the impacts of life styles indicators on well-being confirm the validity of the BES measurement framework and the importance of considering it as a mirror of a multiplicity of objective situations on which to intervene with political multisectorial actions.

**Keywords:** BES, Well-being, Physical inactivity, Life style indicators.

**References:**


Household wealth and income in Italy: Analysis by locally weighted quantile regression

Ilaria Lucrezia Amerise¹, Agostino Tarsitano¹

¹ Dipartimento di Economia Statistica e Finanza, University of Calabria, Rende (Cs), Italy

Household wealth and income in Italy: Analysis by locally weighted quantile regression.
With the re-awakening of interest in the social-welfare aspect of economic change, the study of the relationships between Wealth and income has begun to draw renewed attention. In this paper we examine various features of the wealth distribution in Italy at household level, using data drawn from the Bank of Italy’s survey of household income and wealth (SHIW).
The purpose of our analysis is twofold. One is the notion that there is some relation between the distributions of household wealth, income, family size gender and age of the head of family. The other is that such distributions can be better portrayed and analyzed with the aid of weighted quantile regression rather than with the tools currently being used such as least squares.

As a robust data analysis technique, quantile regression has attracted extensive interest. Substantial efficiency might be gained by incorporating an appropriate weight function to account for inhomogeneity of the variances in the model specification and to attenuate the impact of heavy tails in model fitting. In this study, two systems of weights are considered: one based on sparsity function and the other based on the first-moment sparsity function.

Weighted quantile regression analysis data confirms that wealth in Italy remains concentrated in the hands of a wealthy few. The minority sitting at the higher end of the income scale have reaped the benefits of the economic growth in recent decades by disproportionately increasing their personal wealth. This has been at the expense of the majority at the other end of the income scale, whose share of personal wealth has declined relative to those high-income earners. Large wealth disparities also exist between different occupational and industry groupings. The latter are particularly indicated when positive variables are involved.
The SHIW data highlights the persistence of disparities in the wealth of Italian men and women. Average levels of wealth for men and women remain highly unequal across the categories of age, income, occupation and industry. Even when women and men appear to be relatively equal according to some measurements, there are other gendered dimensions of economic inequality. For example, while the wealth levels of those working in high-status occupations are not dissimilar on average, there are many fewer women within those occupations. Such discrepancies often translate into relative poverty for women during retirement, as well as increased reliance on government support and pensions.

Keywords: Income and wealth distributions, Quantile regression, Quantile density function, Heteroskedasticity, Outliers
Understanding local administrations policies effects on well-being in Italian inner areas

Luca Romagnoli\textsuperscript{a}, Luigi Mastronardi\textsuperscript{a}
\textsuperscript{a} Department of Economics, University of Molise, Campobasso, Italy;

The paper proposes an analysis of local administrations policies aimed at contrasting poverty and inequalities in Italy, basing on the classification of the municipalities defined under the National Inner Areas Strategy (\textit{Strategia Nazionale per le Aree Interne}, SNAI). Poverty and inequality are multidimensional concepts, and are strong determinants of well-being. Inner areas are territorial contexts characterized by a significant distance from the main supply poles of essential services (health, education, mobility). They have recently been the subject of specific policies, which find their main reference in the SNAI; in 2014, when SNAI was defined, municipalities classified as “inner areas” were 4,185, and represented almost 52\% of total Italian municipalities. Resident population amounted to more than 13 million inhabitants (approx 22\% of national population), mainly residing in small municipalities.

Territorial dimension plays a central role in explaining poverty and inequalities and their induced threats. Italy, in fact, presents significant differences at a territorial level: North/South, City/Countryside, Plain/Mountain, and so on. These differences reflect on well-being levels, depending on the presence/absence of environmental, social and economic resources and public and private services.

In Italy, economic inequalities are relatively greater in cities easily accessible, where services are concentrated and directional centres are located, rather than in inner areas, more distant from poles. In inner areas, income distribution is more balanced, maybe owing to the weakness of socioeconomic structure which determines lower income levels and job opportunities mainly in agricultural sector.

In inner area development municipal administrations contribute, together with Regions and Ministries, to implement optimal levels of “essential services of citizenship”. These latter are considered pre-requisites for inner area development.

Municipalities expenditures, again in 2014, summed to € 64 billion circa, 7,7\% of total Public Administrations (PA) expenditures. This is a significant amount, though representing only a 39,5\% of Regions expenditures (€ 162 billion). Municipalities expenditures have a multisectoral valence (i.e. social assistance, infrastructural works, education, local police services, and so on), while Regions spend their budget mostly for health care.

Our study takes into account: a) the per-capita incomes at municipality level, and b) the budgets of Municipalities, in particular for what concerns the expenditures part, as subdivided in “Missions”. Income data come from Revenue Agency, while budget data have been gathered on AIDA database of Bureau van Dijk.

We will analyse and discuss the effects of local policies on poverty and inequality in Italian inner areas by means of some statistical tools, such as Gini index at municipality level, and a regression model, capable to highlight the social and spatial dynamics of well-being.

\textbf{Keywords:} Well-being, Poverty, Inequality, Welfare policies, Inner areas.

\textbf{References:}


Family and personal determinants of gambling risk among Italian adolescent students

Luigi Fabbris\textsuperscript{a}, Alessandra Andreotti\textsuperscript{b}, Bruno Genetti\textsuperscript{b}, Paolo Vian\textsuperscript{b}, Claudia Mortali\textsuperscript{c}, Luisa Mastrobattista\textsuperscript{c}, Adele Minutillo\textsuperscript{c}, Roberta Pacifici\textsuperscript{c}

\textsuperscript{a} Department of Statistical Sciences, University of Padua, Italy; \\
\textsuperscript{b} Explora Center for Research and Statistical Analysis. Vigodarzere (PD), Italy; \\
\textsuperscript{c} Centro Nazionale Dipendenze e Doping, Istituto Superiore di Sanità, Roma, Italy

This paper discusses the social and personal dimensions that may influence youth risk of gambling. We highlight attitudinal, behavioural, family-related, peers-related, and local community characteristics that may influence gambling beyond the situational and demographic descriptors. According to this, we analysed the data collected in a national survey carried out in 2017 and representative of the Italian students aged 14 to 17. The sampling procedure followed a three-stage PPS (Probability Proportional to Size) model with stratification in the first-, second- and third-stage units. First sampling stage was that of municipalities, second stage was that of schools within sampled municipalities and third stage was that of classes within sampled schools. The survey was conducted through a Computer-Assisted Self Interview (CASI) technique using a non-replicable, unique, and anonymous access identification. The students’ propensity to gamble was assessed through an Italian validated version of the South Oaks Gambling Screen-Revised Adolescent (SOGS-RA) scale (Winters et al., 1993; Colasante et al., 2014), which identifies three types of gamblers: social, at risk, and problem (Winters et al., 1995). In order to elicit the possible relationships between the students’ propensity to gamble, on the one hand, and students’ attitudes and behaviours and various external stimuli (such as those from peers and other social environments), on the other hand, we applied a multilevel ordinal logistic model (Agresti, 2002). The analysed data comprising 15,602 students attending 201 secondary schools in 97 municipalities distributed throughout the Italian territory. The multivariate analysis showed that attitudes of students and certain family-related and community-related stimuli altogether influence the propensity to gamble beyond biological and situational characteristics of youngsters.

Keywords: Problem gambling, Students’ propensity to gamble, Socio-ecological model, Multilevel ordinal logistic regression analysis.

References:
Does industry change affect strategic, governance, and financial configurations of private hospital providers? A survey of Italian private healthcare organizations.

Massimo Aria\textsuperscript{a}, Corrado Cuccurullo\textsuperscript{b}

\textsuperscript{a} Department of Economics and Statistics, University of Naples Federico II, Italy
\textsuperscript{b} Department of Management and Economics, University of Campania Luigi Vanvitelli, Italy

The Italian National Health Service (I-NHS) has been affected by a deep evolution of its institutional systems, due to the pursuit of economic sustainability. In the last decade, 10 Regions adopted Health Cutback Plans. The other 10 Regions started processes of grip-back of health policies with the aim of rebalancing the system's hospital-centered vision. Under these circumstances, private hospitals have been pushed by financial and regulatory pressures towards processes of industry concentration (Cuccurullo and Pennarola, 2017).

Italian private hospitals have 30.2\% of the beds of the overall offer (Lega et al., 2018). Their ownership is private, but they operate in a regulated sector and are subject to the dynamics and rules of public sector (Carbone 2013). Moreover, they generally are SMEs (only 12\% exceed 200 beds) with concentrated ownership (typically family business) and face the industry transition with difficulties.

Recently, Italian scholars pay attention to private hospitals, especially in terms of firm size and scope through qualitative methods, such as case studies. The financial and governance configurations of private hospitals are little explored. Our purpose is to bridge the gap, mapping statically and dynamically the configurational archetypes (Sarto et al., 2014).

Methodologically, we have collected financial data of all Italian private hospitals (Ateco 86.10.10) and built a dataset (drop-off criteria: 1-the legal form other than SpA and Srl; 2-the activity other than hospitalization for acute cases; 3-lack of data) with data from 2008, 2012 and 2016. We have analyzed the dataset through factorial analysis and clustering techniques to trace the profile of private hospital in their evolution from 2008 to 2016. The level of analysis is twofold: Region (the I-NHS is decentralized) and hospital.

Findings allow us to understand the intra-industry dissimilarities, with useful implications for policymakers, management, consultants and analysts, who play a professional role in the industry concentration process (Cuccurullo et al., 2016). Our paper ends with a methodological proposal for using statistical analysis of financial data for SMEs with concentrated ownership.

**Keywords:** Italian National Health Service, Performance valuation, Survey; Factorial analysis, Clustering.

**References:**


Demand vs Estimated Burden of Health Care: a comparative evaluation based on spatial analysis

Giuseppe Alfonzetti\textsuperscript{a}, Laura Rizzi\textsuperscript{a}, Luca Grassetti\textsuperscript{a}, Michele Gobbato\textsuperscript{b}  
\textsuperscript{a} Department of Economics and Statistics, University of Udine, Udine, Italy;  
\textsuperscript{b} S.O.C. Epidemiologia Oncologica – CRO, IRCCS, Aviano (PN), Italy.

In the context of increasing life expectancy and growth of the elderly population, the assessment of time and spatial patterns of over 65 population health care need is a key step in order to better manage public resources (Gray, 2005). The aim of this study is to highlight the existence of spatial heterogeneity in the elderly healthcare burden, comparing alternative modelling approaches, in the context of Regione Friuli Venezia Giulia (FVG). Data on estimated health burden in 2017 and 2018 were aggregated on age classes within each municipality. The population size, the ratio between males and females, and the death rate, the counts of 21 chronic conditions, the Resource Utilization Band (RUB) indicator, and the expenditures for healthcare services (Pharmaceutical, Hospital, and Outpatient types) in years from 2002 to 2017 were also collected.

A descriptive analysis both of ageing phenomenon and of health care expenditures trends has been performed. The availability of the RUB indicator, provided in the John Hopkins ACG System (version 11.1.2), allows comparing observed healthcare expenditures (HCE) with the estimated healthcare burdens. In particular, different spatial econometrics models (Elhorst, 2014; Moscone and Tosetti, 2014; LeSage and Pace, 2009, such as those discussed in) have been compared to explore spatial heterogeneity of the differences between demand and health need. The analyses are developed on the full population and also focusing on the elderly population only. The empirical evidence shows that while HCE does not present any spatial pattern, the RUB indicator is characterized by some strong geographical clusterization even after controlling for the demographical structure of municipalities. In order to model the spatial heterogeneity, an SDM specification is chosen after an appropriate set of tests. The spatial patterns of morbidities play an important role in the explanation of the healthcare burden, together with the economic characteristic of the municipality. The model estimation, based on the elderly subpopulation, provides further insights on the diseases mostly influencing the healthcare burden, namely age macular degeneration, human immunodeficiency virus and low back pain. Surprisingly, the focus on the subpopulation points out that elderlies living in areas with higher shares of elderly population are healthier and needs fewer resources than their peers in other areas.

Keywords: Health Expenditures, Panel Data Models, Population Ageing, Resource Utilization Band, Spatial Econometrics.

References

Partial Least Squares - Path Modeling approach for Sustainability using qualitative information

Rosanna Cataldo, Maria Gabriella Grassia, Marina Marino
Department of Social Science, University of Naples Federico II - Napoli, Italy

Sustainability is defined as a complex multidimensional phenomenon and it can be viewed from the perspective of the three pillars model: economic, environmental and socio-cultural pillars respectively. Starting from the successful implementation of Sustainabile Development Agenda 2030 and from the existing 169 indicators, higher-order Partial Least Squares - Path Modeling has been applied in order to compute a global sustainability index. The analysis has been conducted on data derived from the World Bank database with reference to the 28 European community countries. Qualitative information concerning European community countries have been considered in order to perceive differences relating to this issue. In particular, geographic area, type of policy adopted by the countries and currency information have been used in the model.

Keywords: PLS-PM, Sustainability, Qualitative Information

References