1991
COLOGNE
1995
1999
2003
BARCELONA
2007
ROTTERDAM
2011
RENNES
2015
NAPOLI
2019
STELLENBOSCH
2023
BONN

CARME

CORRESPONDENCE ANALYSIS AND RELATED METHODS

www.carme-n.org

Volume of Abstracts

DFG Deutsche Forschungsgemeinschaft

UNIVERSITÄT BONN

Scientific organizing committee

Jörg Blasius (Bonn)

Michael Greenacre (Barcelona)

Alfonso Iodice d'Enza (Naples)

Angelos Markos (Alexandroupuli)

Local organizing committee

Jörg Blasius

Alice Barth

Rebekka Atakan

Manuela Schmidt

We gratefully acknowledge the financial support by
Deutsche Forschungsgesellschaft, DFG
(German Research Foundation), Bonn

Welcome to Bonn, Germany's picturesque city located on the banks of the Rhine River

As you embark on your journey to this historic city, you'll discover a place steeped in rich history, known for its status as the former capital of West Germany, and its association with the legendary composer Ludwig van Beethoven, one of the world's greatest composers. The visionary statesman Robert Schuman, a key figure in the European integration process, spent much of his life in Bonn. He played a pivotal role in the creation of the European Coal and Steel Community, a precursor to the European Union, and is celebrated for his contributions to European unity. Bonn's history as the capital of West Germany from 1949 to 1990 is a testament to its significance in the country's political landscape. During this time, it played a pivotal role in the reunification of Germany, which remains a defining moment in European history. Many of the city's landmarks, such as the Federal Republic of Germany's former seat of government, still bear witness to this era, offering a glimpse into the city's political past.

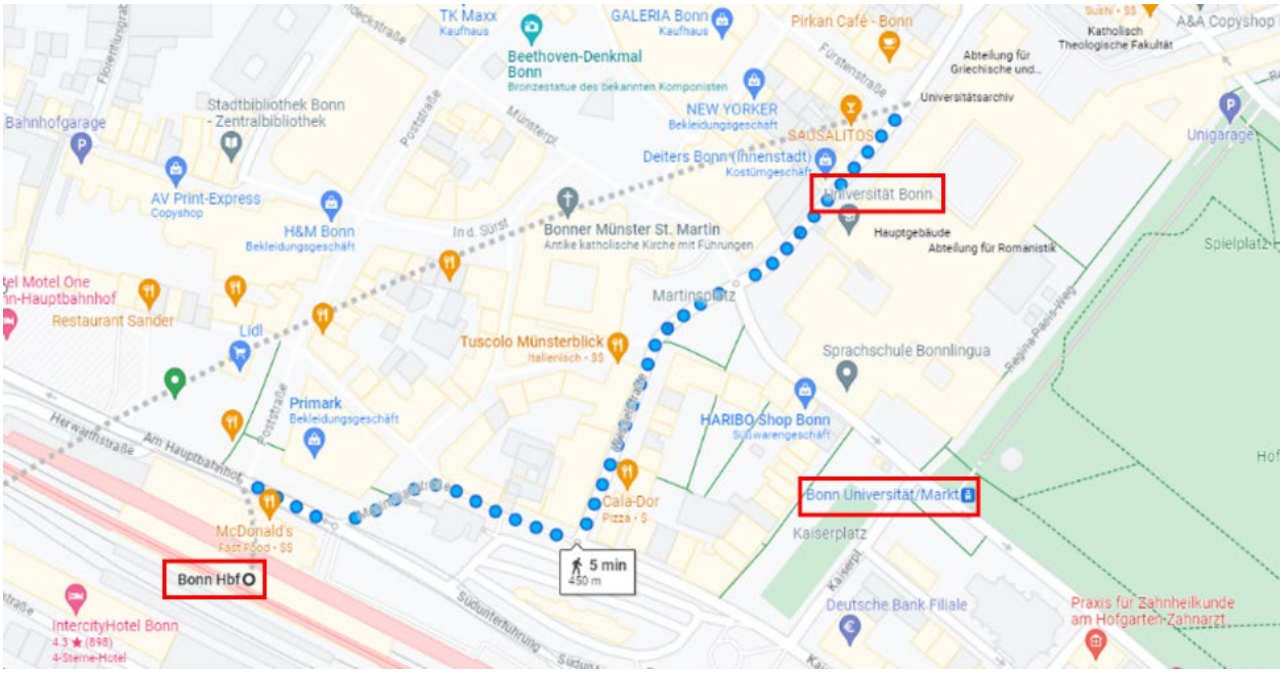
Bonn and especially the university boast a heritage of notable personalities who have shaped the worlds of politics, science, and economics. At the University of Bonn, Nobel Prize laureates Wolfgang Paul and Reinhard Selten made groundbreaking contributions to the field of physics and economics respectively. Selten's research on game theory earned him this prestigious award in 1994 and continues to influence statisticians and economists worldwide. Joseph Schumpeter, a pioneering economist renowned for his theories on entrepreneurship and innovation, served as a professor at the University of Bonn. His concepts, including "creative destruction," continue to shape economic thought. The university has also been the institution of the only Germans who have won the Fields medal for mathematics: Gerd Faltings (in 1986) and Peter Scholze (in 2018, at the age of 30), who is regarded as one of the leading mathematicians in the world.

The city has an even richer scientific legacy. August Kekulé, the renowned chemist, conducted groundbreaking work on the structure of benzene while at the University of Bonn. His discoveries have been foundational in the field of organic chemistry. In the realm of physics, Heinrich Hertz, another Bonn luminary, made groundbreaking contributions to our understanding of electromagnetic waves and wireless communication. The unit of frequency, "hertz" (Hz), is named in his honor.

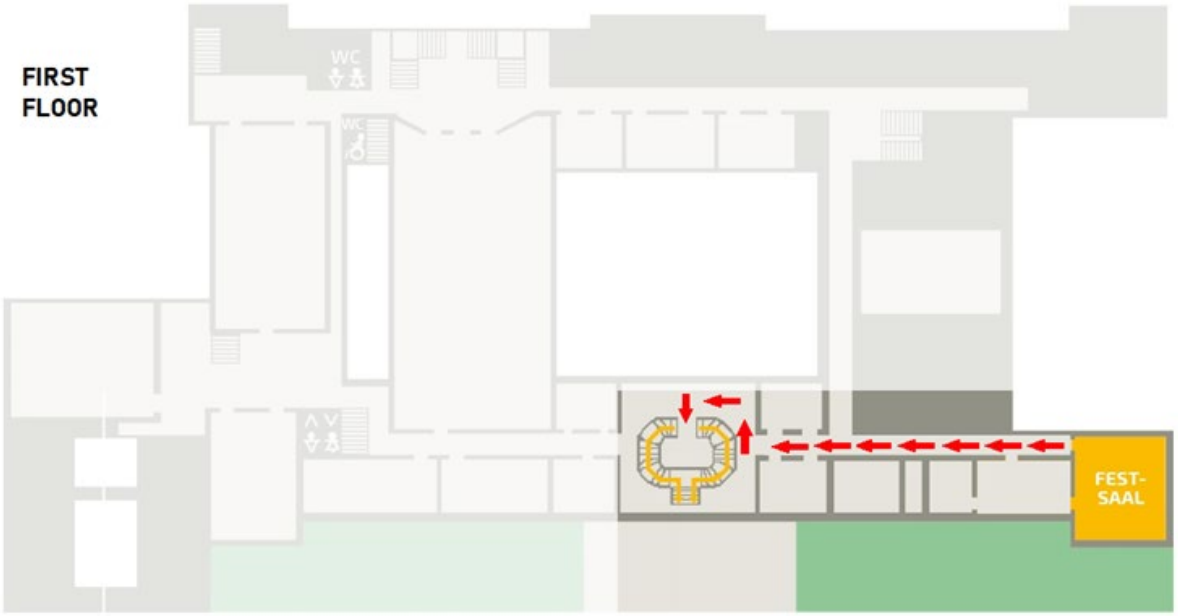
But it is probably the association with Beethoven that most marks the city. As you explore Bonn's charming streets, you'll encounter Beethoven's legacy at every corner, from the Beethoven House museum, his birthplace, to numerous statues and memorials dedicated to his life and work. Enjoy your stay!

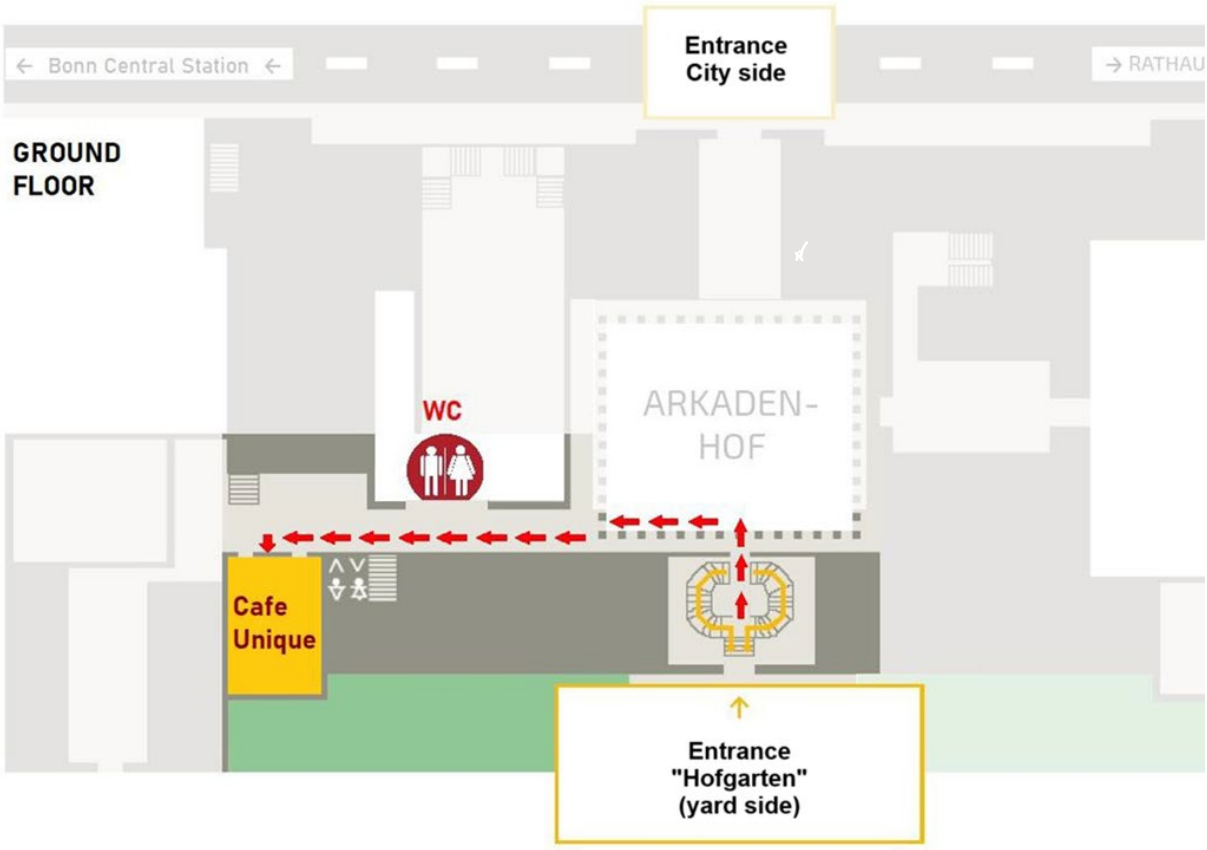
From Bonn main station to the University of Bonn

The University of Bonn is within walking distance (5 mins) from Bonn main station (450 m)



Directions from Festsaal (1st floor) to Café Unique (ground floor)





CARME 2023 IN BONN, PROGRAM

TUESDAY, SEPTEMBER 26

10.00 – 18.00	Excursion: Tour of the former "Government Bunker", lunch and wine tasting in the Ahr valley

WEDNESDAY, SEPTEMBER 27

8.30 – 17.00	Registration
9.15 – 9.45 Room 1 (Festsaal)	Opening session *Chair: Jörg Blasius and Alice Barth* Stephan Conermann (Dean of Faculty of Arts, University of Bonn) **Welcome Address** Michael Friendly (York University, Canada) **In memory of Antoine de Falguerolles** Patrick Groenen (Erasmus University Rotterdam, The Netherlands) **In memory of John Gower**
9.45 – 10.45 Room 1 (Festsaal)	Invited speakers *Chair: Jörg Blasius* Christian Hennig (University of Bologna, Italy) **Data analytic understanding of statistics** Hervé Abdi (The University of Texas at Dallas, USA), Vincent Guillemot (Institut Pasteur, Université Paris Cité, France), Luke Moraglia (The University of Texas at Dallas, USA), Ju-Chi Yu (Campbell Family Mental Health Research Institute, Toronto, Canada) **Interpreting correspondences: how to keep it simple**
10.45 – 11.15	COFFEE BREAK (Senatssaal)
11.15 – 12.30 Room 1 (Festsaal)	Session 1.1: CARME and digital economy. Social space and digital data value chain analysis *Session organizers: Ulf Wuggenig and Christian Tarnai* *Chair: Ulf Wuggenig & Christian Tarnai* 1: Ulf Brefeld, Dennis Fassmeyer, Cheryce von Xylander (Leuphana University Lüneburg, Germany) **AI-based prediction of automatable occupations, or how machines pick teams?** 2: Philippe Saner (University Lucerne, Switzerland) **The data sciences as a space of opportunities. Investigating the double construction of a key field in the digital economy** 3: Ulf Wuggenig, Cheryce von Xylander (University Lüneburg, Germany), Christian Tarnai (Bundeswehr University Munich, Germany) **Digital economy, social space and symbolic power. Correspondence and cluster analysis results of secondary analyses of Eurobarometer surveys**

11.15 – 12.30	Session 1.2: Mixed and ordinal categorical variables
	Chair: Paul H.C. Eilers
Room 2 (HS IV)	1: Angelos Markos (Democritus University Thrace, Greece), Alfonso Iodice D'Enza (Università degli Studi di Napoli Federico II), Michel van de Velden, Carlo Cavicchia (Erasmus University Rotterdam)
	Bridging differences: A holistic framework for distance quantification in mixed-type variables
	2: Alfonso Iodice D'Enza (University of Naples Federico II, Italy), Angelos Markos (Democritus University of Thrace, Greece), Michel van de Velden, Carlo Cavicchia (Erasmus University of Rotterdam, The Netherlands)
	Association-based learning for mixed data
	3: Leyao Zhang, Peter Xuekun Song (University of Michigan, United States of America)
	Supervised learning of outcome-relevant items from a questionnaire via mixed integer optimization
11.15 – 12.30	Session 1.3: CARME in ecology
Room 3 (HS V)	*Session Organizer: Michael Greenacre* *Chair: Michael Greenacre*
	1: Stephane Dray, Lisa Nicvert (CNRS - Université Lyon 1, France)
	Correspondence analysis of ecological networks
	2: Raul Primicerio (UiT, The Arctic University of Norway, Norway)
	Spatial patterns and trait dependency in fish diet
	3: Cajo J F ter Braak (Wageningen University & Research, The Netherlands)
	Progress and regression in constrained ordination
12.30 – 14.00	LUNCH: CAFÉ UNIQUE
14.00 – 15.30	Session 2.1: Correspondence analysis
	Chair: Gilbert Saporta
Room 1 (Festsaal)	1: Carlos M. Cuadras (University of Barcelona, Spain)
	Continuous correspondence analysis with applications
	2: Jörg Breitung (University of Cologne, Germany)
	Alternative distance measures for analysing contingency tables
	3: Brigitte Le Roux (MAP5, Université Paris Cité, France), Frédérik Cassor, Flora Chanvril (CEVIPOF, Sciences-Po Paris, France)
	A generalized multiple correspondence analysis: bi weighted multiple correspondence analysis
	4: Koen Plevoets (Ghent University, Belgium)
	Correspondence analysis of power-law distributed data via ranks

14.00 – 15.30	Session 2.2: Sociological Application of CARME 1
	Session Organizer: Alice Barth *Chair: Alice Barth*
Room 2	1. Anton Grau Larsen (Roskilde University, Denmark), Jacob Lunding, Christoph Ellersgaard (Copenhagen Business School, Denmark)
(HS IV)	**Invariant structures? Enduring relations of oppositions in the field of the Danish business elite from 1910-2020**
	2: Joschka Baum (University of Mainz, Germany)
	Unraveling the social patterns of video game preferences: integrating big data and survey responses for multiple correspondence analysis
	3: Valentina Petrović, Jörg Rössel (University of Zürich, Switzerland)
	The advent of neoliberal cultural policy? The curious case of the Zurich opera
	4: Martin Fritz, Dennis Eversberg (Friedrich Schiller University Jena, Germany)
	A cartography of the social-ecological transformation in Germany: Exploring the interconnections between classes, mentalities, and modes of living
14.00 – 15.30	Session 2.3: Analysing social spaces and fields 1: Analyses of science and related fields
	Session Organizers: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok *Chair: Andreas Schmitz*
Room 3	1: Séverine Marguin (Technische Universität Berlin, Germany)
(HS V)	**Research and design. On the refiguration of an ambivalent relationship in the academic field of architecture**
	2: Jonas Volle (Otto-von-Guericke-Universität Magdeburg, Germany)
	Integrating topic modeling and block models into correspondence analysis: the example of the sociological field in Germany
	3: Sebastian Diemer Mørk, Anton Grau Larsen (Roskilde University, Denmark)
	Designed for success or failure: differences in funding and rejection in the space of applications to the Danish Art Foundation among craftsmen and designers
	4: Tomasz Warczok (University of Warsaw, Poland), Hanna Dębska (Pedagogical University of Cracow, Poland)
	Academia, market, and symbolic exchanges: the specificity of the legal field Poland
15.30 – 16.00	COFFEE BREAK (Senatssaal)
16.00 – 17.30	Session 3.1: Analysing social spaces and fields 2: Construction and analyses of social space
	Session Organizers: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok *Chair: Tomasz Warczok*
Room 1	1: Lucas Page Pereira (ENS Paris-Saclay, France)
(Festsaal)	**The evolution of social inequalities in Brazil**
	2: Naoki Iso (Tokyo University of the Arts, Japan)
	Social space and principle of differentiations in Japan
	3: Jan Rasmus Riebling (Bergische Universität Wuppertal, Germany)
	From geometric analysis to analytical geometry: making sense of [(habitus)(capital)] + field = practice

16.00 – 17.30	Session 3.2: Analysis and models of categorical data
	Chair: Jörg Blasius
Room 2	1: Fernanda Mazzotta, Lavinia Parisi, Monica Consuelo Rodriguez Torres (University of Salerno, Italy)
(HS IV)	**Poverty, happiness, and economic freedom: a correspondence analysis with a doubling technique on Latin American countries**
	2: Jan Fredrik Hovden (University of Bergen, Norway)
	Well-bred and well-spoken: on the role of class origins for children's experiences with rhetorical speech in schools public speech competitions
	3: Jörg-Henrik Heine (Technical University of Munich, Germany), Christian Tarnai (Universität der Bundeswehr München, Germany)
	Inglehart index revisited: a comparative scaling of materialism – postmaterialism using multiple correspondence analysis and the generalized graded unfolding model
	4: Vincent Guillemot (Université Paris Cité, France), Hervé Abdi (The University of Texas at Dallas, USA,) Victoire Baillet (Université Paris Cité, France), Ju-Chi Yu (Centre for Addiction and Mental Health, Toronto, Canada)
	Applying sparse DiSTATIS to categorical data
16.00 – 17.30	Session 3.3: Compositional data analysis
	Session organizer: Michael Greenacre *Chair: Michael Greenacre*
Room 3	1: Paul H.C. Eilers (Erasmus University Medical Center, The Netherlands)
(HS V	**Improving the ternary diagram for compositions**
	2: Germà Coenders (University of Girona, Spain), Ana de Vega-Ventura, Berta Ferrer-Rosell, Eva Martín-Fuentes (University of Lleida, Spain),
	Decomposing the market share contributions of price and units sold. A compositional visualization
	3: Anna Maria Fiori (University of Milano-Bicocca, Italy), Germà Coenders (University of Girona, Spain)
	Systemic risk in Europe from a compositional perspective
	4: Karel Hron (Palacky University Olomouc, Czech Republic)
	Analyzing multivariate densities in Bayes spaces with applications
17.40 – 18.40	Invited speakers
	Chair: Patrick Groenen
Room 1	3: Sugnet Lubbe (Stellenbosch University, South Africa)
(Festsaal)	**The biplot zone for exploring truths in mixed scaled measurements**
	4: Michael Friendly (York University, Canada)
	My life in pictures: tales of a graphic developer and amateur historian of #dataviz
18.45 – 21.00	Reception: Café Unique

THURSDAY, SEPTEMBER 28

9.00 – 18.00	Registration
9.30 – 10.30 Room 1 (Festsaal)	Invited speakers *Chair: Angelos Markos* 5: Vartan Choulakian, Jacques Allard (Université de Moncton; CA, Canada) **Taxicab correspondence analysis of sparse contingency tables** 6: Paula Brito (University of Porto & LIAAD INESC TEC, Portugal) **Linear models for distributional data analysis**
10.30 – 11.00	COFFEE BREAK (Senatssaal)
11.00 – 12.30 Room 1 (Festsaal)	Session 4.1: CARME and clustering *Chair: Brigitte Le Roux* 1: Jean-Luc Durand (Université Sorbonne Paris Nord, France) **Cluster analysis of categorical variables based on pairwise regressions** 2: Michael Greenacre (Universitat Pompeu Fabra, Spain), Maurizio Vichi (Università di Roma La Sapienza, Italy) **Combining categories for each variable in a multivariate categorical data set, with minimum loss of variable association** 3: Maurizio Vichi (Università di Roma La Sapienza, Italy), Michael Greenacre (Universitat Pompeu Fabra, Spain) **Latent categorical variables of a set of observed categorical variables** 4: Stéphanie Bougeard (Anses, France), Xavier Bry (University of Montpellier, France), Thomas Verron (Danais, Orléans, France), Ndeye Niang (CEDRIC CNAM, Paris, France) **Combined-information criterion for clusterwise elastic-net regression. Application to omic data**
11.00 – 12.30 Room 2 (HS IV)	Session 4.2: Sociological Application of CARME 2 *Session Organizer: Alice Barth* *Chair: Alice Barth* 1: Jörg Blasius (University of Bonn, Germany), Susanne Vogl (University of Stuttgart, Germany) **Identifying outliers to improve survey statistics** 2: Rebekka Damla Atakan, Alice Barth (University of Bonn, Germany) **The association of spatial perceptions and expected neighborhood change – a multiple factor analysis** 3: Miriam Trübner, Judith Lehmann, Rasmus Hoffmann, Alexander Patzina (University of Bamberg, Germany) **Understanding patients' preference: a study on the diversification of medical practices in the contemporary healthcare system** 4: Kazuo Fujimoto (Tsuda University, Japan) **Extend the use of supplemental variables in GDA by applying machine learning to the free text descriptive response portion and combining it with MCA analysis**

11.00 – 12.30 Room 3 (HS V)	Session 4.3: Analysing social spaces and fields 3: Analysis of higher education *Session Organizer: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok* *Chair: Tomasz Warczok* 1: Pablo Antonio Lillo Cea (Uppsala University, Sweden) **The world-class ordination: studying a global sub-field of universities** 2: Jan Thorhauge Frederiksen (University of Copenhagen, Denmark) **Pedagogical and disciplinary positions in the Danish field of higher education** 3: Ylva Bergström (Uppsala University, Sweden) **Education capital – a landmark in the Swedish political landscape** 4: Ida Lidegran (Uppsala University, Sweden) **The educational space of preschools in a cultural stronghold in Sweden**
12.30 – 14.00	LUNCH: CAFÉ UNIQUE
14.00 – 15.30 Room 1 (Festsaal)	Session 5.1: Related methods *Chair: Mark de Rooij* 1: Mark Stemmler (University of Erlangen-Nuremberg, Germany), Jörg-Henrik Heine (Technical University Munich, Germany) **Person-centered data-analysis with covariates and the R-package confreq** 2: Lucio Palazzo, Maria Iannario, Francesco Palumbo (University of Naples Federico II, Italy) **Combining archetypal analysis with latent profile analysis to assess financial knowledge** 3: Rosanna Verde, Antonio Balzanella (University of Campania "Luigi Vanvitelli", Italy) **A Factorial method for multiple distributional data based on LDQ transformation** 4: Se-Kang Kim (Baylor College of Medicine, United States of America) **Seeking pure pattern effects: latent level and pattern variance in profile analysis is evidenced by reparametrized observed level and pattern variance**
14.00 – 15.30 Room 2 (HS IV)	Session 5.2: CARME and education *Session Organizer: Johannes Hjellbrekke* 1. Stylianos Papalexandris, Sofia Anastasiadou (University of Western Macedonia, Greece) **Correspondence analysis of midwifery students' beliefs and attitudes towards statistics** 2: Simone Mejding Poulsen (Copenhagen University, Denmark) **Meritocracy and selectivity in the Danish field of higher education; how the admission system effectively perpetuates horizontal stratification in a seemingly egalitarian system** 3: Jose G. Clavel (Universidad de Murcia, Spain), Roberto De la Banda (Universidad de Educación a Distancia, UNED, Spain) **What CARME has to offer to the ESCS index in the PISA assessment** 4: Andreas Roaldsnes (University of Bergen, Norway) **Social capital and the intergenerational transmission of cultural capital: how parents' social networks influence children's accumulation of cultural capital**

14.00 – 15.30 Room 3 (HS V)	Session 5.3: Analysing social spaces and fields 4: Analysis and methodological advances in the realm of economy and politics *Session Organizers: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok* *Chair: Andreas Schmitz* 1. Bruno Cautres, Brigitte Le Roux (CEVIPOF/Sciences Po – France) **Open-ended questions analysis in public opinions research** 2: Christian Schmidt-Wellenburg (Universität Potsdam, Germany) **Experts in the field of European politics 1966-2018: changes in European expert groups analyzed using multiple correspondence analysis and hierarchical agglomerative clustering** 3: Rob Timans (Nivel, The Netherlands) **The field of private equity in the Netherlands** 4: Frédéric Lebaron, Brigitte Le Roux, Aykiz Dogan (ENS Paris-Saclay, France) **Applying combinatorial inference in GDA. The case of European Central Bank governing council members (1999-2022)**
15.30 – 16.00	COFFEE BREAK (Senatssaal)
16.00 – 17.50 Room 1 (Festsaal)	Session 6.1: Biplots and related methods *Chair: Niel Le Roux* 1: Laura Vicente González, José Luis Vicente Villardón (Universidad de Salamanca, Spain) **Partial least squares for mixed continuous and binary variables and its associated biplot** 2: Jose Luis Vicente Villardon, Laura Vicente González (Universidad de Salamanca, Spain) **Biplots of mixed continuous and categorical data** 3: Jules de Tibeiro, Vartan Choulakian (University of Moncton, Canada), Pasquale Sarnacchiro (University of Naples, Italy) **On the choice of weights in aggregate compositional data analysis** 4: Ruiping Liu (Beijing Information Science and Technology University, China), Ndeye Niang (CEDRIC CNAM, Paris, France), Gilbert Saporta (Conservatoire national des arts et métiers, France) **Sparse non-symmetrical correspondence analysis** 5: Nguyen-Khang Pham (Can Tho University, Vietnam), Éric Grosso (University of Edinburgh, Scotland), Hélène Noizet (Université Paris 1 Panthéon-Sorbonne, France), Jean-Luc Pinol (École Normale Supérieure de Lyon, France) Jean-Hugues Chauchat (Université Lumière Lyon 2, France) **Visualization of geographical data: AMADO-carto to integrate Bertin diagram, clustering trees and colored map**

16.00 – 17.50

Room 2

(HS IV)

Session 6.2: Data science in social and political research

Session organizer: Theodore Chadjipadelis
Chair: Angelos Markos

1: Qianqian Qi, David J. Hessen, Tejaswini Deoskar, Peter van der Heijden (Utrecht University, The Netherlands)

A comparison of latent semantic analysis and correspondence analysis of document-term matrices

2: Theodore Chadjipadelis (Aristotle University Thessaloniki, Greece)

What was really the case? The analysis of 2023 elections in Greece

3: Georgia Panagiotidou, Theodore Chadjipadelis (Aristotle University of Thessaloniki, Greece)

A comparative tool for analyzing electoral behavior and political culture: a multivariate methodology leveraging hierarchical cluster analysis and factorial correspondence analysis

4: Vasiliki Bouranta, Georgia Panagiotidou, Theodore Chadjipadelis (Aristotle University Thessaloniki, Greece)

Exploring and comparing political marketing strategies in Greek elections: a multivariate analysis on candidates, parties, issues and the impact of electoral systems

5: Cassandra Handan-Nader (Stanford University, United States of America)

Graph embeddings with influential outliers using correspondence analysis

16.00 – 17.50

Room 3

(HS V)

Session 6.3: Quantifying digital social spaces and fields. Towards a Bourdieusian methodology of computational social sciences and digital humanities

Session organizer: Andreas Schmitz & Mattia Samory
Chair: Andreas Schmitz & Mattia Samory

1: Mikael Kasimir Wallin (Tampere University, Finland)

Meanings on working class on Finnish social media

2: Langyi Tian (Columbia University, United States of America), Aurélien Boucher (The Chinese University of Hong Kong (Shenzhen), China)

Stratifying lifestyle and social class in urban China

3: Yuanmo He, Milena Tsvetkova (London School of Economics and Political Science, United Kingdom)

A method for estimating individual socioeconomic status of Twitter users

4: Andreas Schmitz (Bonn University, Germany), Mattia Samory (Sapienza University of Rome)

Modelling reddit as social space – a mixed methods iterative procedure

5: Oliver Wieczorek (University of Kassel, Germany)

Asian whizkids, black basketballers, and white male professors? The visual embedding of different (minority-) groups in Youtube videos of US-universities

18.00 – 18.30

Room 1

(Festsaal)

Invited speakers

Chair: Alice Barth

7: Johannes Hjellbrekke (University of Bergen, Norway)

Analysing temporal stability and temporal change by way of class specific MCA (CSA). An exploratory presentation

FRIDAY, SEPTEMBER 29

9.00 – 17.00	Registration
9.30 – 10.30 Room 1 (Festsaal)	Invited speakers *Chair: Michael Greenacre* 8: Peter van der Heijden (Utrecht University / University of Southampton) **Some relationships of correspondence analysis with other methods and models** 9: Mark de Rooij (Leiden University, The Netherlands) **Logistic multidimensional data analysis with an emphasis on ordinal variables**
10.30 – 11.00	COFFEE BREAK (Senatssaal)
11.00 – 12.30 Room 1 (Festsaal)	Session 7.1: Analysing social spaces and fields 5: Relational approaches to culture *Session Organizers: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok* *Chair: Frédéric Lebaron* 1: Mikael Börjesson (Uppsala University, Sweden) **The Swedish space of culture and media practices: structure and trends** 2: Željka Tonković (University of Zadar, Croatia) **Unravelling structural properties of cultural production and consumption: a multiple correspondence approach** 3: Dave Benjamin Balzer, Gunnar Otte, Nico Sonntag (University of Mainz, Germany) **Measuring symbolic boundaries – cultural, economic, and moral distinctions in Germany** 4: Jörg Rössel (University of Zurich, Switzerland), Patrick Schenk (University of Lucerne, Switzerland) **Dimensions of cultural openness in the wine field**
11.00 – 12.30 Room 2 (HS IV)	Session 7.2: CARME in medicine and biology *Chair: Cajo J.F. ter Braak* 1: Martin Paries (Oniris, INRAE, StatSC, Nantes, France; Anses, Ploufragan, France), Evelyne Vigneau (Oniris, INRAE, StatSC, Nantes, France), Adeline Huneau (Anses, Ploufragan, France), Olivier Lantz (Clinical Immunology Laboratory, Institut Curie, Paris, France), Stéphanie Bougeard (Anses, Ploufragan, France) **MBPCA-OS: an exploratory multiblock method for mixed variables. Application to study the immune response to SARS-CoV-2 infection and vaccination** 2: Yan Zhao, Clemens Kohl, Martin Vingron (Max Planck Institute for Molecular Genetics, Germany) **CAbiNet: Joint visualization of cells and genes based on a gene-cell graph** 3: Ionas Erb (Centre for Genomic Regulation (CRG), Spain) **Optimal power transformations for the correspondence analysis of relative counts** 4: Lauren Hsu (Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, USA; Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA), Aedin Culhane (University of Limerick, Limerick, Ireland) **Correspondence analysis in gene expression of single cells**

11.00 – 12.30 Room 3 (HS V)	Session 7.3: CARME and regression modelling *Chair: Maurizio Vichi* 1: Patrick Groenen (Erasmus University Rotterdam, The Netherlands), Michael Greenacre (Universitat Pompeu Fabra, Barcelona, Spain) **Nonlinear prediction by kernels made explainable** 2: Modesto Escobar, Cristina Calvo-López (University of Salamanca, Spain) **Interactive network graphs online to analyze surveys** 3: Pieter Cornelis Schoonees (Erasmus University Rotterdam, Econometric Institute, The Netherlands, Niel J. le Roux (Stellenbosch University, Department of Statistics and Actuarial Science, South Africa) **Linear regression estimation and model selection under data aggregation**
12.30 – 14.00	LUNCH: CAFÉ UNIQUE
14.00 - 15.30 Room 1 (Festsaal)	Session 8.1: The spectrum of CARME *Chair: Michael Friendly* 1: Aldo Corbellini, Marco Riani (Dipartimento di Scienze Economiche e Aziendale; Interdepartmental Centre for Robust Statistics, Università di Parma, Italy), Anthony Atkinson (London School of Economics, London, UK), Francesca Torti (European Commission, Joint Research Centre, Ispra, Italy) **Applications of robust correspondence analysis** 2: Ivailo Partchev (Cito, The Netherlands) **Of music, politics, visual perception and, of course, SVD** 3: Rainer Diaz-Bone (University of Lucerne, Switzerland) **Is multiple correspondence analysis failing in the social sciences? – Lessons from „science wars" and statistical education** 4: Karl Michael van Meter (École Normale Supérieure, France) **CARME – The rise and fall of a sociological space in the East**
14.00 – 15.30 Room 2 (HS IV)	Session 8.2: Sociological application of CARME 3 *Session Organizer: Alice Barth* *Chair: Alice Barth* 1. Philippe Bonnet (Université Paris Cité, France) **From application to publication** 2: Yasemin El Menouar (Bertelsmann Stiftung, Germany) **Muslims in Europe: the diversity of religious belief and practice** 3: Rahim Hajji (Hochschule Magdeburg-Stendal, Germany) **Social identities and sociocultural adaptation of Moroccan and Turkish immigrants and their descendants - A study using cluster and correspondence analysis** 4: Alice Barth, Jörg Blasius (University of Bonn, Germany) **Using MCA to show changes of cultural and economic capital over time**

14.00 – 15.30 Room 3 (HS V)	Session 8.3: Analysing social spaces and fields 6: Expanding and refining the construction of spaces/fields *Session Organizers: Andreas Schmitz, Frédéric Lebaron & Tomasz Warczok* *Chair: Andreas Schmitz* 1: Will Atkinson (University of Bristol, UK) **Social space, ethno-racial space and symbolic space in the US: investigating the interplay** 2: Stephanie Beyer (Leibniz Universität Hannover, Germany) **Social stratification, lifestyles and health – Exploring the German social space with specific multiple correspondence and hierarchical cluster analysis** 3: Henk Roose (Ghent University, Belgium), Roger Friedland (University of California, Santa Barbara, United States) **The institutional logics of love: using geometric data analysis to explore the social structuring of intimate practices among American college students** 4: Magne Flemmen, Nora Essahli, Jan Fredrik Hovden (University of Oslo, Norway) **Political taste and its homologies**
15.30 – 16.00	COFFEE BREAK (Senatssaal)
16.00 – 17.00 Room 1 (Festsaal)	Invited speakers *Chair: Alfonso Iodice D'Enza* 10: Rosaria Lombardo (University of Campania "Luigi Vanvitelli", Italy), Eric Beh (University of Wollongong, Australia; Stellenbosch University, South Africa) **The Cressie-Read family of divergence statistics and correspondence analysis for two-way and multi-way contingency tables** 11: Ron Kenett (KPA Ltd., SNI Technion, Israel), Chris Gotwalt (JMP Statistical Discovery, LLC, USA) **Predictive analytics, causality and sensitivity analysis of association rules**
17.00-17:30 Room 1 (Festsaal)	Closing session *Chair: Jörg Blasius, Alfonso Iodice D'Enza, Michael Greenacre, Angelos Markos*
18.30	Conference dinner Cruise on the river Rhine: 19.00 – 23.00 (Moby Dick starts boarding at 18.30)

Table of Contents

Interpreting correspondences: how to keep it simple

Hervé Abdi[1], Vincent Guillemot[2], Luke Moraglia[1], Ju-Chi Yu[3]

[1]The University of Texas at Dallas, United States of America; [2]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France; [3]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada; herve@utdallas.edu

Plain and multiple correspondence analyses (CA and MCA) were mostly defined as descriptive multivariate methods, and early practitioners (e.g., Benzécri) were explicitly avoiding inferences, in part because standard assumptions such as random sampling were rarely (if ever) met. But, even within this descriptive framework, several rules of thumbs were used to select important factors, observations, or variables. These procedures were often imported from the psychometric tradition. For example, Kaiser's "little jiffy" stipulation of keeping eigenvalues larger than 1 translated into "keep dimensions with an eigenvalue larger than average." Along the same lines, important items contribute more than the average or than their mass.

However, in the eighties, several authors (e.g., Greenacre, 1984, Lebart et al. 1984) suggested to add inferences (e.g., bootstrap based) to evaluate result stability. A case in point, in this framework, Malinvaud (1987, see also Saporta, 2011) encouraged practitioners (here, mostly econometricians) to embrace the multivariate complexity of real-world data but also to integrate inferences to avoid factors that could simply reflect "noise." There, Malinvaud was working with real contingency tables and derived asymptotic analytical solutions based on the chi-square distribution to select the relevant factors.

Interestingly, these approaches—often inspired by the psychometric tradition—did not import standard procedures such as rotation (e.g., varimax routinely used with Kaiser's "little jiffy") to simplify the interpretation, possibly because these approaches necessitated two steps: first, find the dimensionality, and then rotate.

Recently the problem of obtaining simple factorial structures and identifying important dimensions and items (e.g., row and columns) has been re-ignited with methods imported from statistical learning and artificial intelligence, such as LASSO based sparsification. These methods often reframe interpretation as an optimization problem under constraints (e.g., orthogonality of dimensions). In this talk we review and compare these early and recent methods and offer some guidelines.

Keywords: rotation, simple structure, sparsification, sparse generalized SVD, inferences

The association of spatial perceptions and expected neighborhood change – a multiple factor analysis

Rebekka Damla Atakan[1], Alice Barth[1]

[1]Rheinische Friedrich-Wilhelms-Universität Bonn, Germany; ratakan@uni-bonn.de

Neighborhood gentrification implies that residents with low socio-economic status are replaced by young urban professionals and wealthy upper-class dwellers. During this process, at least for some time, long-term residents and newly moved-in residents live in close proximity while the neighborhood adopts more and more to the needs of the new affluent dwellers. As spatial perceptions and environmental needs are highly influenced by education and socialization, it can be expected that these social, physical, functional, and symbolic changes of a neighborhood are perceived differently by various social groups. For example, while newcomers to the neighborhood welcome developments, long-term residents may feel increasingly uncomfortable around the new neighbors, their upscale boutiques and fancy restaurants, and develop fears of displacement.

In this paper, three questions will be explored. First, we want to find out how spatial perceptions and evaluations are associated with expectations of the future development of a neighborhood. Second, we want to investigate how patterns of spatial perception and expectation are related to socio-structural characteristics. Third, we will assess whether patterns and relations change during the gentrification process.

Using data from the first (2010, N=1009) and the fifth (2022, N=915) waves of the Cologne Dwelling Panel, spatial perceptions and evaluations of residents of two residential areas in Cologne as well as expected socio-spatial changes are used as active variables in a multiple factor analysis, separate for the two waves. We compare the structure of the resulting “spaces of neighborhood perception” between the two time points and assess how socio-structural characteristics are related to perception patterns by adding the former as supplementary variables. As such, we will analyze interrelations and changes between spatial perceptions, expected neighborhood change and socio-structural position during more than 10 years of gentrification.

Keywords: gentrification, neighborhood change, spatial perceptions

Social space, ethno-racial space and symbolic space in the US: investigating the interplay

Will Atkinson[1]

[1]University of Bristol, United Kingdom; w.atkinson@bristol.ac.uk

This paper presents work aimed at exploring the homology between the social space and lifestyle space in the US and their mediation by ethnicity/race. The starting point is the conjecture that ethnicity/race, themselves conceivable in relational or geometric terms, exert an autonomous effect on lifestyles. Using a nationwide survey fielded in the US in 2017-18, I construct a model of the social space and a model of the symbolic space using multiple correspondence analysis. I then explore the homology of the two spaces and its mediation or complication by ethnicity/race by employing elements of structured data analyses. That includes using multiple linear regression, on which I will make some remarks given its uneasy relationship with geometric data analysis in research on social spaces and lifestyles.

Keywords: social space, lifestyle space, ethnicity, MCA, regression

Measuring symbolic boundaries – cultural, economic, and moral distinctions in Germany

Dave Benjamin Balzer[1], Gunnar Otte[1], Nico Sonntag[1]

[1]Johannes Gutenberg University Mainz, Germany; dave.balzer@uni-mainz.de

Our contribution examines symbolic boundaries in Germany using survey data. Following Lamont, symbolic boundaries are understood as the perception and evaluation of persons, objects, or practices which leads to various forms of distinction via stereotyping and prejudice formation. If the boundaries become solidified and actors differentiate between “us” and “them”, this may lead to social closure and discrimination. We examine three expressions of symbolic boundaries: Socio-economic (e.g., wealth/occupational success), cultural (e.g., sophistication/education) and moral (e.g., discipline/moderation).

We opt for a relational approach that emphasizes the step of the data collection. Rather than simply asking respondents to self-assess their own lifestyles, we seek to encourage comparison with others to explore the extent to which respondents identify with or distinguish themselves from the individuals presented, which is the first, necessary condition for symbolic boundaries.

In the second wave of the panel study “Cultural Education and Cultural Participation in Germany,” a random sample of the German population aged 15 years and older were presented with short descriptions of the lifestyles of fictitious persons (e.g., “For someone, a measured and righteous life is more important than success and prestige” or “Someone likes to surround himself/herself with exclusive material things”). Similar to Schwartz's PVQ, the respondents were asked to indicate how similar or dissimilar the person described is to them.

We take a deductive approach and formulate hypotheses about which forms of identification or distinction are to be expected depending one’s volume and structure of capital (e.g., the greater the endowment with cultural capital, the greater the identification with sophistication/education, and the stronger the distinction from morality-focused and wealth-focused lifestyles).

We then use multiple correspondence analysis to situate individuals and their indications of (dis)similarity to different lifestyles in social space derived from multiple indicators of cultural and economic capital.

Keywords: relational sociology, measurement, symbolic boundaries, Bourdieu, social space

Using MCA to show changes of cultural and economic capital over time

Alice Barth[1], Jörg Blasius[1]

[1]University of Bonn, Department of Sociology; alice.barth@uni-bonn.de; jblasius@uni-bonn.de

Based on Bourdieu’s “social space” approach, we will demonstrate how lifestyles in two gentrified neighborhoods in Cologne (Germany) change over time. We use indicators such as preferred sources of furniture, food preferences, and interior style of the dwellings to create a two-dimensional space by applying multiple correspondence analysis (MCA). The axes of the resulting space can be interpreted as cultural and economic capital, or respectively, capital volume and composition. Further, socio-demographic characteristics are included as supplementary variables to confirm the interpretation of the axes. By doing this, we can assess which lifestyle characteristics are particularly indicative of high or low cultural capital and whether their relational position, i.e. their power of distinction, changes over time.

Our data source is the Cologne Dwelling Panel (2010-2022), a five-wave face-to-face survey in two residential neighborhoods that are differentially affected by gentrification processes. The data are especially suitable for the study of neighborhood change, as the sample units are dwellings (apartments or houses) instead of persons or households, as in “classical” panel studies. For each dwelling, a resident is interviewed as its “spokesperson”. As such, new residents (in-movers) are included in the sample by design, which allows for a representative depiction of population changes over time.

We will assess changes in cultural and economic capital with data from the first wave (2010) and the fifth wave (2022), using only those dwellings that participated in both time points (N=482). The joint evaluation of these two waves allows us to study changes in lifestyles in the neighborhoods. Further, we can assess changes in the positions of social groups within the space, e.g. whether in-movers' capital volume differs from that of staying residents or whether the relationship between educational achievements and certain indicators of cultural capital changes over time.

Keywords: changes in life-styles over time, dwelling panel, social space, MCA, gentrification

Unraveling the social patterns of video game preferences: integrating big data and survey responses for multiple correspondence analysis

Joschka Baum[1]

[1]University of Mainz, Germany; jobaum@uni-mainz.de

The social structuration of taste in music, art, and film is a well-established area of research in sociology. However, there is only scant knowledge on the question of how taste in the recently emerged social field of video gaming is socially determined. This presentation fills the gap by uncovering social patterns of video gaming in Germany. For this purpose, data from the second wave of the panel survey on cultural education and cultural participation in Germany (KuBiPaD) is used. The survey is based on nationwide two-stage random sampling of the adult population (N=2455) and includes a variety of items on cultural participation, cultural tastes and video game genre preferences. One shortcoming of these video game genre items is that they only convey information on the game mechanics but not aesthetic features of games. To include aesthetic features, survey participants were asked about the title of their favorite video game. This information is used to link information on aesthetic properties of the favorite game to the survey data via the gaming platform Steam. The linked data will be used to look for patterns of video game preferences within Bourdieu's and Schulze's theoretical frameworks. For this purpose, video game preferences will be projected into a social space and a space of cultural tastes.

Keywords: video game preferences, big data, multiple correspondence analysis, cultural tastes

Education capital - a landmark in the Swedish political landscape

Ylva Bergström[1]

[1]Uppsala University, Sweden; ylva.bergstrom@edu.uu.se

For a long time, the scholarly debate on voting, party preferences, and political attitudes, paid attention to the decline in the relevance of class to political position takings and voting behavior and a shift in political values. Social class has not only been claimed 'dying', and 'dead' but captured in new conceptualizations, new methods, and kept alive through attempts to explain its decline. The structure of political attitudes has become more complex emphasizing alternatives to traditional left-right oppositions, recognizing alternative value and identity issues. The basic argument in this paper is that adherence to a different set of political attitudes and values is still related to social origin, but in a more complex structure that can be explored with a multidimensional concept of social class and political attitudes. Here data from the Swedish panel of the World Value Survey of 1991, 1996, 2006, 2011, and 2017 are analyzed, using 14 variables, and 38 response modalities (posed in the same way throughout 5 survey waves) to conduct multiple correspondence analysis. Examination discloses a pattern that reflects oppositions between clusters of the traditional left-right wing as well as new green alternatives and authoritarian attitudes. Despite changes over time, the space of political attitudes corresponds with a Bourdieusian conception of a two-dimensional social class. Volume of educational capital corresponds strongly with the structure that distinguishes attitudes on trust in political institutions and internationalization, whereas the volume of economic capital corresponds strongly with the structure that distinguishes attitudes on redistribution and privatization. Capital composition reflects divisions a space of position takings between academic left-liberals and merchant radical right.

Keywords: MCA, cluster analysis, structuring factors, educational capital, political opinion, World Value Survey / European Value Study

Social stratification, lifestyles and health – exploring the German social space with specific multiple correspondence and hierarchical cluster analysis

Stephanie Beyer[1]

[1]Leibniz Universität Hannover, Germany; stephanie.beyer@lcss.uni-hannover.de

Numerous studies show the nexus of social status and health but only relatively few works rely on a relational approach to study social stratification, lifestyles and health inequalities. Therefore, this contribution examines the connection of social structure, lifestyles and health in Germany from a relational perspective. To this end, I draw on the German General Social Survey (ALLBUS) that includes rich information on 3471 respondents' social position, their self-reported health, health-relevant practices such as dietary, smoking and exercise habits, as well as living conditions, attitudes, and leisure time activities. Based on these variables, I first map a multidimensional social space using specific multiple correspondence analysis. In a second step, this reconstructed space allows me to identify distinctive lifestyle clusters with hierarchical cluster analysis.

Keywords: social space, lifestyles, health, specific multiple correspondence analysis, cluster analysis

Identifying outliers to improve survey statistics

Jörg Blasius[1], Susanne Vogl[2]

[1]Universität Bonn, Germany; [2]Universität Stuttgart, Germany; jblasius@uni-bonn.de, susanne.vogl@sowi.uni-stuttgart.de

Already a small number of outliers can affect the entire solution in a large data set. This is especially true when analyzing sets of variables within a multivariate approach, for example, item batteries measuring latent attitudes. In both principal component and factor analyses, rotation of results is a standard feature of multivariate data analysis. In many cases, the unrotated solutions remain unpublished. Most often applied is the varimax rotation, which usually allows a better interpretation of the content, since it better adapts to the variable clusters. In this presentation we show that such a rotation often only optimally adapts to the outliers of a survey, i.e., respondents who are often characterized by a (strong) satisficing behavior. As an empirical example we use data from a Viennese study with 14 to 16-year-old pupils from non-advanced schools; the data have been collected in 2018 within a web survey. With low educational attainment and more than half of the respondents having German as their second language only, we can show that some pupils gave arbitrary answers that affect the entire solution.

Keywords: data quality, satisficing, principal component analysis

From application to publication

Philippe Bonnet[1]

[1]Université Paris Cité, France; philippe.bonnet@u-paris.fr

How does one go from presenting some sixty papers at a conference to publishing a book containing only a third of them? In addition to the scientific quality that made them eligible for the congress, what other criteria, more or less explicit, led to the publication of only a limited number of these papers.

The "Empirical Investigation of Social Space II" conference held in Bonn in October 2015 has been the focus of our attention. Some sixty papers were presented, covering a wide range of sociological fields. What they had in common was the use of empirical methods. Based on an analysis of the abstracts of these papers, we tried to identify a geography of thematic networks that does not overlap with the simple division into fields of sociology. We also used the available characteristics of the authors of these papers and of the institutions to which the authors belong. In addition to textual analysis, we made extensive use of geometric data analysis, such as multiple correspondence analysis and hierarchical clustering. Even if the answers to initial questions are not self-evident, we will ask ourselves how to use textual analysis techniques, remembering that there is no neutrality of techniques. The aim of this study is not simply to use these software programs as revealing tools, in the photographic sense, but to illustrate that each question corresponds to a certain type of analysis of the corpus of texts. The result is an interpretative space largely defined by the choice of techniques used.

Keywords: textual analysis, geometric data analysis, reflexivity

Combined-information criterion for clusterwise elastic-net regression. Application to omic data.

Stephanie Bougeard[1], Xavier Bry[2], Thomas Verron[3], Ndeye Niang[4]

[1]Anses, France; [2]University of Montpellier, France; [3]Danais, Orléans, France; [4]CEDRIC CNAM, Paris, France; stephanie.bougeard@anses.fr

Many research questions pertain to a regression problem assuming that the population under study is not homogeneous with respect to the underlying model. A common strategy is to use clusterwise regression (a.k.a., typological regression) to identify homogeneous clusters of observations having their own sets of regression coefficients; each observation is assigned to one cluster. To this end, several solutions have been proposed.

These are based either on a K-means-like criterion applied to the least square error, yielding models referred to in the sequel as deterministic models; or on a conditional mixture likelihood criterion, yielding (stochastic) mixture models. These latter models are popular and research on mixture models is very active. In this setting, we propose an original method denoted combined information criterion CLUSterwise elastic-net regression (CICLUS). This method handles several methodological and application-related challenges. It is derived from both the information theory and the microeconomic utility theory and maximizes a well-defined criterion combining three weighted sub-criteria, each being related to a specific aim: getting a parsimonious partition, compact clusters for a better prediction of cluster-membership and a good within-cluster regression fit. The solving algorithm is monotonously convergent under mild assumptions. The CICLUS method provides an innovative solution to two key issues: the automatic optimization of the number of clusters and the issue of a prediction model. We applied it to elastic-net regression in order to be able to manage high-dimensional data involving redundant explanatory variables. CICLUS is illustrated through a real example in the field of omic data, showing how it improves the quality of the prediction and facilitates the interpretation. It should therefore prove useful whenever the data involve a population mixture as for example in biology, social sciences, economics, or marketing.

Exploring and comparing political marketing strategies in Greek elections: a multivariate analysis on candidates, parties, issues and the impact of electoral systems

Vasiliki Bouranta[1], Georgia Panagiotidou[1], Theodore Chadjipadelis[1]

[1]Aristotle University Thessaloniki, Greece; vickybouranta@gmail.com

The proposed presentation explores the evolving domain of political marketing, a field that extends beyond communication methods and public relations, encapsulating activities that influence the political behavior of parties and individuals. Drawing on theoretical frameworks and methodologies, we explore the application of marketing mix theory (product, price, place, promotion) within this political context. The focal point of our research is an in-depth examination of the political marketing strategy employed by Greek political parties during the Greek parliamentary elections of 2023.

The analysis scrutinizes the strategic patterns used in terms of selecting promotion tools, prioritizing political agenda issues, and focusing on the candidate versus the party. This involves advanced multivariate analysis methods such as cluster analysis, multiple correspondence analysis, and principal component analysis, which are utilized to detect and analyze in a comparative perspective the different strategies of the candidates and the parties in the Greek parliamentary elections of 2023. Moreover, the analysis focuses on how parties incorporated the newly implemented simple proportional representation system into their marketing strategies and their pre-electoral campaigns.

Our data is derived from various sources including newspapers, mass media (TV, radio), and social media, allowing us to scrutinize the political product (party program and candidates), the 'price' (the voter's vote), the distribution strategies and promotion activities at both local and national level. Furthermore, we explore the relation between candidate profiles, their political marketing strategies, their political characteristics, and their probability of being elected or not. The study ultimately suggests that political and electoral competition pivots on three pillars—candidates, parties, issues—which interact within the institutional framework as configured by the electoral law. This research bridges the gap between political marketing strategies, electoral systems, and their impact on campaign success, contributing significantly to the independent scientific scope of political marketing.

Keywords: political marketing, electoral campaign, Greek elections, electoral systems, multivariate analysis

The Swedish space of culture and media practices: structure and trends

Mikael Börjesson[1]

[1]Uppsala University, Sweden; mikael.borjesson@edu.uu.se

This paper draws on a Swedish survey, the annual SOM (society, opinion and media) habits survey, which has been distributed since 1986 to a representative sample of the Swedish population. A space of cultural and media practices is constructed using specific multiple correspondence analysis (Le Roux & Rouanet 2004, 2010) with inspiration from Bourdieu's classical study presented in *La Distinction* (1979). In the plane of principal axes 1 and 2, a set of configurations of properties can be identified, including, along the first axis, those who invest heavily in highbrow art practices such as attending art galleries, opera, and classical music concerts in combination with extensive reading habits and radio listening versus those who do not invest in such practices at all, and an opposition along the second axis more relating to digital, social, and commercial media, on the one hand, and traditional print and radio media on the other. In addition, a third polarity is established between a mundane lifestyle comprising frequent restaurant and sports event going combined with some cultural activities, as opposed to an ascetic pole. This space has a clear social logic, where there is a strong link between extensive cultural practices and substantial investments in education, while economic resources tend to be more oriented along the dimension of mundane/ascetic. Furthermore, the space is not only socially structured but also geographically order in a way that the largest concentration of cultural and economic resources is found in the metropolitan areas, and the rural areas are predominantly defined by a lack of such resources. These differences have become more pronounced over time. A special attention is given to the impact of the Covid-19 pandemic.

Keywords: social space, geometric data analysis, Sweden, culture and media practices, cultural capital

Progress and regression in constrained ordination

Cajo J F ter Braak[1]

[1]Wageningen University & Research; cajo.terbraak@wur.nl

The world is not 2-d, but users of multidimensional scaling and other unconstrained ordination methods nevertheless usually try to display it in 2-d ordination diagrams and biplots. This is where constrained ordination is of help: it allows to focus the analysis on effects of particular interest, based on meta-information coded as external predictors. Constrained ordination combines the power of regression analysis and ANOVA with that of (unconstrained) ordination. Principal component analysis with regression gives redundancy analysis or reduced rank regression, and correspondence analysis similarly gives canonical correspondence analysis. Their power can be increased by specifying covariates which enter the model in full rank so as to adjust for effects of confounders or factors that are the focus in another analysis.

This talk discusses progress (permutation testing (ter Braak & te Beest 2022) and extension to generalized linear (mixed) models (Yee 2015; van der Veen *et al.* 2023)), and a return to regression in a less developed state in chemometrics (ter Braak 2023).

References

ter Braak, C.J.F. (2023) *Redundancy analysis includes ASCA and outperforms ASCA+ and WE-ASCA* (preprint). *SSRN*, 1-91. http://dx.doi.org/10.2139/ssrn.4375930

ter Braak, C.J.F. & te Beest, D.E. (2022) *Testing environmental effects on taxonomic composition with canonical correspondence analysis: alternative permutation tests are not equal*. Environmental and Ecological Statistics*,* 29, 849-868. https://doi.org/10.1007/s10651-022-00545-4

van der Veen, B., Hui, F.K.C., Hovstad, K.A. & O'Hara, R.B. (2023) *Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling*. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.14035

Yee, T.W. (2015) *Vector Generalized Linear and Additive Models With an Implementation in R*. Springer, New York.

Keywords: redundancy analysis, canonical correspondence analysis, supervised learning, ASCA

AI-based prediction of automatable occupations, or how machines pick teams?

Ulf Brefeld[1], Dennis Fassmeyer[1], Cheryce von Xylander[1]

[1]Leuphana Universität Lüneburg, Germany; dennis.fassmeyer@leuphana.de

The automation of labor continues to be a pivotal research area, especially in the context of evolving markets and artificial intelligence. Traditional methodologies (e.g., Frey & Osborne, 2017) use machine learning approaches to classify occupations as either automatable or non-automatable or estimate automation potentials. However, these approaches rely on a great deal of manual annotations, which poses a limiting factor on scalability and resource allocation. Secondly, by focusing on occupations as a whole, they ignore that automation predominantly affects specific tasks within jobs, not entire occupations. To address these limitations, we introduce a novel transformer-based approach, leveraging multiple instance learning (MIL) on raw text data instead of hand-engineered features. Specifically, we treat jobs as collections of tasks and aim to reduce the need for manual labels by utilizing occupation-level ratings of knowledge, skills, and abilities as surrogates for representation learning. The resulting architecture significantly outperforms traditional approaches in predicting job automation potential, demonstrating its effectiveness on the comprehensive O*Net database. In addition, our emphasis on task-level analysis not only informs about potential job displacements but also illuminates opportunities for job enhancements, thereby offering a fresh perspective on the implications of automation in the labor market. This recombination of tasks is set to have profound mereological implications for conceptions of ‘agency and structure’ and its commodified labor form. For a comparative methodological perspective, we also subject the O*Net database to a multiple correspondence analysis on textual data to examine notions of task granularity and collaboration (e.g., teamwork) in the task-based digital economy (e.g., Husson 2017).

Keywords: multiple instance learning (MIL), MCA, digital economy, job & task automation, text mining

Alternative distance measures for analysing contingency tables

Jörg Breitung[1]

[1]University of Cologne, Germany; breitung@statistik.uni-koeln.de

This paper argues that the usual chi-square metric inherent in correspondence analysis possesses some undesirable features and it is therefore interesting to consider alternative metrics. Specifically, it is shown that the "standardized residuals" used as a starting point for correspondence analysis are not properly standardized as their variances depend on cell probabilities. As an alternative, the deviation from independence is measured by the underlying correlation coefficient. Interestingly, correlations are constructed similarly as the chi-square distance but imply a different denominator. Another measure is the likelihood ratio (LR), which seems most natural from a statistical point of view. Moreover, LR statistics allow us to find out whether some specific cell frequency (or row/column) is significantly different from the expected cell frequency under the assumption of independent outcomes. To illustrate the findings, I consider the empirical example of Blasius and Greenacre (2006) and compare different distance measures and visualization techniques. In particular I present the correlation matrix in the form of a ``balloon plot'' that highlights the distances from the expected cell frequencies. Furthermore, I visualize the correlation matrix in a biplot that reveals interesting differences with the plot in standard correspondence analysis.

Keywords: chi-square distance, correspondence analysis, correlation matrix

Linear models for distributional data analysis

Paula Brito[1]

[1]Fac. Economia, University of Porto & LIAAD INESC TEC, Portugal; mpbrito@fep.up.pt

The classical data representation model is too restrictive when the data to be analysed are not real numbers or single categories, but comprise variability. In this talk, we are interested in numerical distributional data, where units are described by histogram or interval-valued variables. Linear models for such distributional variables are proposed, which rely on the representation of histograms or intervals by the associated quantile functions, under specific assumptions. These then allow for multivariate analysis of distributional-valued data, e.g. multiple linear regression or linear discriminant analysis. Principal component analysis may also be addressed under this framework. Applications of the proposed approach will be presented.

Keywords: symbolic data analysis, histogram data, interval data

Open-ended questions analysis in public opinions research

Bruno Cautres[1], Brigitte Le Roux[1]

[1]CEVIPOF/Sciences Po - France; bruno.cautres@gmail.com, brigitte.leroux@sciencespo.fr

Academic opinion surveys (in political science) are often characterized by a very large number of indicators for measuring values, attitudes or behavior. The aim of this multiplicity of indicators is to make it possible to analyze complex systems of social representations. One of the main methodological concerns of these batteries of items is to capture the multidimensionality of political and social attitudes. This generates survey questionnaires that are often very long, a real methodological challenge insofar as these surveys are increasingly administered online, a mode of survey administration that makes long interviews potentially problematic. Curiously, these surveys make little use of open-ended questions, even though the use of open-ended questions makes it possible to capture the complexity and multidimensionality of opinions. The aim of this paper is to review the status of open-ended questions in social science methods today, and to use a French example to show how geometric data analysis can be used to capture citizens' mental representation patterns and their multidimensionality.

Keywords: public opinion, open-ended questions, geometrical data analysis, textual analysis, politics

What was really the case? The analysis of 2023 elections in Greece

Theodore Chadjipadelis[1]

[1]Aristotle University Thessaloniki, Greece; chadji@polsci.auth.gr

In this paper we present a detailed study of the 2023 Parliamentary Elections in Greece (2023A & 2023B). This study includes correspondence analysis on the election results and ecological inference estimates of voter transition rates. Correspondence analysis reveals four dimensions of political competition in Greece. Ecological inference provides estimates of loyalty and defection rates between parties for each of the 59 electoral districts and for the entire country. We are also able to estimate abstention rates for each group of voters.

This is a significant outcome that is not easily extracted by exit-polls. Geographical presentation of local estimates reveals patterns. These patterns can serve as the basis for further research on their association with other demographic and socio-economic characteristics.

In addition, through hierarchical cluster analysis we form clusters of geographical areas in order to present and check the aforementioned associations.

The Greek Parliament consists of 300 members who are elected for a period of four consecutive years through direct, universal and secret ballot by the citizens. The last parliamentary elections in Greece were held on 21/5/2023 and 25/6/2023. This date was one month before the end of the four-year government term in a period that electoral polls awarded the government the top position in voting preferences. After a short pre-election period, which began with the so called “return to normality” after the COVID pandemic, and with the opposition parties accusing government for incapability, it was common belief that the government would win, but it was doubtful if it would get a parliamentary majority. Therefore (cause also of the electoral system), the government received 40% of the votes in the first round, didn’t get the majority and called for second elections.

Keywords: correspondence analysis, hierarchical cluster, ecological inference, elections

Visualization of geographical data: AMADO-carto to integrate Bertin diagram, clustering trees and colored map

Nguyen-Khang Pham[1], Éric Grosso[2], Hélène Noizet[3], Jean-Luc Pinol[4], Jean-Hugues Chauchat[5]

[1]Can Tho University; [2]Mapping Edinburgh; [3]Université Paris 1 Panthéon-Sorbonne; [4]École Normale Supérieure de Lyon; [5]Université Lumière Lyon2; jhc@aasma.fr

AMADO-carto simultaneously produces the Bertin chart, the row and column classification trees and the cartographic representation of a table when one of the dimensions of the table is a list of geographic areas. AMADO-carto is an extension, of the AMADO-online application that outputs the Bertin chart, reordering the rows and columns manually or according to hierarchical clustering or according to the first factor of correspondence analysis.

We thus obtain, side by side, the graphical representation of the diagonalized table and the map; the zones of the same class are visualized with the same color on the map and on the Bertin graph. AMADO-carto is available as a freely downloadable extension under QGIS (free geographic information system that can be installed on a local computer) and can be integrated into the "PTM for QGIS" toolbox.

Example: Paris, 2015, employed labour force, age 25 to 54, by occupation and arrondissement

AMADO-online is a free multilingual application that displays and analyses data matrix (binary, counts, responses to Likert-type items, or measures of heterogeneous variables...) by combining Bertin's visualization method with factorial analysis (to find an approximately diagonal structure, if it exists in the data) and hierarchical classification (to find block-models).

The home page offers the choice of nine languages, a demonstration video in English or French, and a user guide in English or French.

References:

Bertin Jacques (1967), *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*, Paris/La Haye, Mouton ; Paris, Gauthier-Villars. (Translation 1983. *Semiology of Graphics* by William J. Berg.)

Chauchat J-H & A. Risson 1998, *Bertin's Graphics and Multidimensional Data Analysis*, in Visualization of Categorical Data, J. Blasius, M. Greenacre (eds). Chapman & Hall / CRC Press.

Greenacre, M. 2007. *Correspondence Analysis In Practice, chap 15*. Chapman & Hall / CRC Press.

Projet Time Machine (ex. Paris Time Machine) paris-timemachine.huma-num.fr

QGIS Documentation www.qgis.org/en/site/

AMADO-online is available at paris-timemachine.huma-num.fr/amado/

Keywords: statistical graphics, Jacques Bertin, maps, clustering, QGIS

Taxicab correspondence analysis of sparse contingency tables

Vartan Choulakian[1], Jacques Allard[1]

[1]Université de Moncton, Canada; vartan.choulakian@umoncton.ca

The presentation covers two topics:

Part 1: We present an overview of taxicab correspondence analysis, a robust variant of correspondence analysis, for visualization of sparse contingency tables. In particular, I visualize an extremely sparse textual data set of size 590×8260, concerning fragments of eight sacred books recently introduced by Sah and Fokoué (2019) and studied quite in detail by (12 + 1) dimension reduction methods (t-SNE, UMAP, PHATE,...) in machine learning by Ma, Sun and Zou (2023).

Part 2: We examine some theoretical results concerning Greenacre's theorem, Goodman's marginal-free correspondence analysis and the Sinkhorn-Knopp (RAS-IPF) algorithm.

References:

Sah, Preeti and Fokoué, Ernest (2019*). What do Asian and non-Asian scriptures have in common? An applied statistical machine learning inquiry*. Mathematics for Applications 8(2), 151-171.

Ma, Rong, Sun, Eric D., and Zou, James (2022). Ma, Sun and Zou (2022). *A spectral method for assessing and combining multiple data visualizations*. Nature Communications 14, 780.

What CARME has to offer to the ESCS index in the PISA assessment.

Jose G. Clavel[1], Roberto De la Banda[2]

[1]Universidad de Murcia, Spain, Spain; [2]Universidad de Educación a Distancia, UNED, Spain; jjgarvel@um.es

The index of economic, cultural, and social status (ESCS) is perhaps one of the most used variables in secondary analyses of data from Large Scale Assignments on Education. Specifically, its presence is widespread on the Programme for International Student Assessment (PISA) research. This ESCS index is based on student responses to the context questionnaires that they complete in the PISA test. Once obtained, it is used as an international standard for the measurement of socioeconomic status of the student.

In the last decade, some scholars (see Avvisati, 2020) have challenged ESCS validity as an indicator of the socioeconomic status, questioning its components (for PISA 2018 they were: home possessions including books at home, parental education, and highest parental occupation).

The presence of massive qualitative information obtained from the different PISA questionnaires (not only student context questionnaires but also those of the principal of the school, teachers and parents are available) offer a rich field where CARME techniques have much to offer. In this paper, alternatives to the traditional ESCS index with be presented, and their capacity to measure the socioeconomic status contrasted. Adding new inputs to the index will allow a more accurate measure of the socioeconomic status of the students' families and consequently a more accurate secondary analyses of the PISA dataset.

References:

Avvisati, F. (2020). *The measure of socio-economic status in PISA: a review and some suggested improvements.* Large-scale Assessments in Education, 8(1), 8. https://doi.org/10.1186/s40536-020-00086-x

OECD. (2019). *PISA 2018 technical report*. www.oecd.org/pisa/data/pisa2018technicalreport/

Keywords: PISA, ESCS, dual scaling

Applications of robust correspondence analysis

Aldo Corbellini[1], Marco Riani[1], Anthony Atkinson[2], Francesca Torti[3]

[1]Dipartimento di Scienze Economiche e Aziendale and Interdepartmental Centre for Robust Statistics, Università di Parma, Italy; [2]London School of Economics, London, UK; [3]European Commission, Joint Research Centre (JRC), Ispra, Italy aldo.corbellini@unipr.it

In this paper we introduce a robust form of correspondence analysis based on the tools of robust statistics. This leads to the systematic deletion of outlying rows of the table and to plots of greatly increased informativeness (Riani et al., 2022). The robust method requires that a specified proportion of the data be used in fitting. To accommodate this requirement, we provide an algorithm that uses a subset of complete rows and one row partially, both sets of rows being chosen robustly. We prove the convergence of this algorithm. We also discuss some methods to adaptively select the optimal proportion of trimming which has to be used. Using a variety of examples, coming from different sources, we show how the application of the suggested method can easily identify subsets of the sample which are behaving in a different way from the bulk of the data.

Reference:

Riani, M., Atkinson A.C., Torti, F., Corbellini, A. (2022). *Robust correspondence analysis*. Journal of the Royal Statistical Society: Series C (Applied Statistics), 71, 1063–2041.

Continuous correspondence analysis with applications

Carles M. Cuadras[1]

[1]University of Barcelona, Spain; ccuadras@ub.edu

The continuous version of correspondence analysis starts with Lancaster's expansion of a bivariate distribution in terms of canonical correlations and functions. This expansion is useful for constructing families of bivariate distributions with given marginals, which are generalizations of the Farlie-Gumbel-Morgenstern (FGM) family. However, the representation of the cumulative distribution function does not give enough graphical information to recognize and distinguish different distributions. For instance, Gaussian and Clayton look very similar. Also, the scatterplot of a bivariate sample can be very similar for distributions such as Clayton and Raftery. Lancaster's expansion can also be interpreted via the chi-squared distance between observations of a random variable. Inspired by this geometrical property, we propose a graphical procedure for representing families of distributions.

We fit a generalized FGM to a given family (e.g. Gaussian). Then we obtain the two first canonical correlations of the generalized FGM but computed with respect to the given family, which depends

on a parameter. This provides a parametric curve, a kind of signature, which describes the whole family and help us to distinguish different families.

Keywords: bivariate density expansion, continuous MDS, generalized FGM, visualization of bivariate distributions

Is multiple correspondence analysis failing in the social sciences? Lessons from "science wars" and statistical education

Rainer Diaz-Bone[1]

[1]University of Lucerne, Switzerland; rainer.diazbone@unilu.ch

At a first glance, there is no other statistical approach which fits so obviously to sociological thinking as multiple correspondence analysis (MCA). MCA is well suited to detect social structures and has been introduced into the social sciences by the influential work of Pierre Bourdieu and his collaborators. Although it can be regarded in French sociology as an established method, MCA cannot be regarded as a standard approach in the social sciences elsewhere. Instead, MCA is marginalized by other dominant statistical approaches and epistemic cultures in the social sciences.

There are different reasons for this. On the one hand, there are dominant "epistemic couples" of (methods and paradigms) which gain more influence and contradict the explorative and factorial culture of MCA. On the other hand, there are specific MCA-properties, which still impede MCA to be regarded as part of the core of social science analysis tools and which make its training challenging.

The standing of MCA and its teaching cannot be understood and promoted when MCA is recognized as a statistical tool only. The presentation addresses two interrelated tensions: the "science wars" in the field of the social sciences (i.e., the conflicts between scientific paradigms), and the problem of statistical education in social sciences. MCA has to be conceived as being positioned in the epistemological "quarrels" between sociological paradigms, and its teaching has to be conceived as demanding for a (neo)structuralist way to train students in the social sciences.

All in all, this presentation will address topics linked to philosophy of science, sociology of social research and statistical education. The aim is to develop strategies to improve MCA's positioning and training in the social sciences.

Keywords: science wars, statistical education

Correspondence analysis of ecological networks

Stephane Dray[1], Lisa Nicvert[1]

[1]CNRS - Université Lyon 1, France; stephane.dray@univ-lyon1.fr

Ecological networks are increasingly used to depict biotic interactions (edges) between species (nodes) in ecological communities. Among them, bipartite network consists of two sets of nodes and a set of edges in which each edge only joins nodes in different sets. These types of networks are used, for instance, to depict plant-pollinator or host-pathogen interactions. Graphical representation is often a first step in the analytical pipeline but is challenging for this type of data if no external information is used to position or order the nodes. Choosing node positions (called a graph layout) is apparently a simple task but is in fact a very hard combinatorial problem and consists of searching for the optimal layout for a given set of objectives (e.g., maximizing attractions between connected nodes or minimizing edge crossings).

In this presentation, we will show how correspondence analysis (a widely used method in community ecology) and related methods can be useful to represent ecological networks. Using the weighted average and variance concepts, it can be demonstrated that correspondence analysis offers a nice way to reconcile mathematical properties of the graphical representation and the ecological theory of networks.

Keywords: correspondence analysis, (double) canonical correspondence analysis, bipartite network, interaction niche

Cluster analysis of categorical variables based on pairwise regressions

Jean-Luc Durand[1]

[1]Université Sorbonne Paris Nord, France; jean-luc.durand@univ-paris13.fr

In this talk, we introduce a new method of hierarchical cluster analysis of categorical variables. This method is intended to be in harmony with multiple regression and correspondence analysis.

Assuming that a categorical variable can be geometrically represented by the Euclidean cloud of its categories, as is the case in multiple correspondence analysis ("MCA-cloud"), we define the regression of a categorical variable B onto a categorical variable A as the multiple regression of each of the coordinate variables of the MCA-cloud of B onto the indicator variables of categories of A. The fitted cloud obtained by the regression of B onto A is nothing else than the cloud of categories of A given by the correspondence analysis of the contingency table A×B.

We define the mutual goodness of fit of two categorical variables A and B as the global rate of variance of the two MCA-clouds of A and B explained by the regressions of B onto A and A onto B.

Given a partition of a set of categorical variables, the within-cluster homogeneity criterion is a weighted average of the mutual goodness of fit of variables belonging to the same cluster.

At each step of the ascending classification, the within-cluster homogeneity criterion is maximized. At the last step, the homogeneity criterion of the single cluster can be expressed thanks to the matrix of the mean square contingency coefficients (phi-2) between variables: it is equal to the average of the diagonal cells divided by the average of the off-diagonal cells.

Two interesting features of this method are:

- the homogeneity criterion can be interpreted in terms of variance rate;

- likewise, one can perform a hierarchical cluster analysis of subsets of categorical variables corresponding to different topics.

Keywords: categorical variables, correspondence analysis, hierarchical cluster analysis, multiple regression

Improving the ternary diagram for compositions

Paul H.C. Eilers[1]

[1]Erasmus University Medical Center, Netherlands, The; p.eilers@erasmusmc.nl

The ternary diagram is a familiar display for compositions with three components. It uses the fact that the sum of the distances to the sides of an equilateral triangle is constant for all points inside the triangle. It is a convenient device, as long as none of the three fractions are close to zero or one. If that is not the case, data points are pushed into a corner or to one of the sides of the triangle.

De Rooij and Eilers (2013 proposed an alternative display, which they called TrioScale. In the ternary diagram, the projection on each axis is determined by the arithmetic difference of one pair of the three components. The TrioScale display is based on the differences of their logarithm, i.e. the log-ratio.

Log-ratios do not accept zeros. If a data set contains low counts, with one or more zeros, the TrioScale display cannot be used directly. As a remedy, I propose to compute a penalized log-linear fit to the data and to replace counts by their expected values. This idea is inspired by the PRIDE model of Perperoglou and Eilers (2010}. Let $\mu_{ij} = \mathrm{E}(\mathrm{y}_{ij})$. We fit the saturated log-linear model $\eta_{ij} = \log(\mu_{ij}) = \alpha_i + \beta_j + \gamma_{ij}$, with a ridge $\lambda \sum_i \sum_j \gamma_{ij}^2$ on γ. The penalty parameter λ is tuned using AIC.

The display of data is much improved. It shows no corners and sides of the triangle. Also, in the Trioscale diagram, log-linear relations show up as straight lines. This may help suggesting models for data.

References

Perperoglou, A. and Eilers, P. H. C. (2010) *Penalized regression with individual deviance effects*, Computational Statistics, 25, 341-361.
Rooij, M. and Eilers, P. (2013) TrioScale: *A new diagram for compositional data. Proceedings of CODAWork*, 2013, Vorau.

Keywords: TrioScale, log-ratios, penalty

Muslims in Europe: the diversity of religious belief and practice

Yasemin El Menouar[1]

[1]Bertelsmann Stiftung, Germany; yasemin.el-menouar@bertelsmann-stiftung.de

Muslims living in European countries still endure a negative perception. This can be traced back to the negative perception of their religion. Public debates and controversies about the hijab, Mohammed cartoons and violent attacks committed in the name of Islam have shaped a negative image of Islam. But this does not reflect the diverse reality of the Muslim population. The Bertelsmann Stiftung's Religion Monitor shows that many Muslims in Europe are more religious than other groups. However, Muslims in Europe are far from being a homogeneous group. Instead, the Religion Monitor data show a wide range of different religious practices and orientations. Conceptual and methodological aspects measuring Muslim religiosity in quantitative research will be discussed.

This paper is based on the data of the Religion Monitor 2017 which contain a variety of different indicators to measure Muslim religiosity. It includes more then 3 000 Muslim respondents surveyed in Germany (n=1 113), Austria (n=503), Switzerland (n=501), United Kingdom (n=500) and France (n=502). The application of PCA, cluster analysis and MCA shows different pattern and types of Muslim religiosity across Europe.

Keywords: religion, diversity, Islam, religion monitor

Optimal power transformations for the correspondence analysis of relative counts

Ionas Erb[1]

[1]Centre for Genomic Regulation (CRG), Spain; ionas.erb@crg.eu

Data matrices containing relative counts or proportions (a.k.a. compositional data) can be analysed with both logratio analysis (LRA) and correspondence analysis (CA). In fact, there is a known connection between both methods that is mediated by the Box-Cox transformation, i.e., CA of power-transformed data converges to LRA for powers tending to zero. Power-transformations are sometimes preferred over simple logarithmic or logratio transformations because they do not need zero replacements, and they usually represent a gentler, tunable transformation of the original data sets.

It is therefore of interest to obtain power parameters that are optimal in a well-defined sense. Previous work by Greenacre has focused on obtaining powers that maintain maximum coherence as measured by similarity with the LRA solution. We have recently shown that power transformations

of relative counts can be considered shrinkage estimators of compositions (when these are viewed as multinomial parameters), which leads to an alternative justification of these transformations as well as a different optimisation scheme for the power parameter. We have also shown how shrinkage can be used to obtain better logratio covariance estimates. Here we propose to use the optimised shrinkage intensity for logratio covariance matrices as the power for the transformation of column profiles of a CA. While estimating the compositions concerns the rows of the data matrix, estimating the covariance makes use of its columns. Both approaches should ideally be considered together as a dual problem. We explore how the two points-of-view can be reconciled when applying power transformations to relative count data. CA of power-transformed counts can be used to normalise and visualise data obtained from RNA sequencing experiments. We provide arguments in favour of this approach and explore benchmark options for single-cell transcriptomic data.

Keywords: Box-Cox transformation, relative counts, logratio covariance, shrinkage estimates, data normalization

Interactive network graphs online to analyze surveys

Modesto Escobar[1], Cristina Calvo-López[1]

[1]University of Salamanca, Spain; modesto@usal.es

Graphs have been employed not only to solve topographic problems and to represent social structures, but also to show the correlation between variables according to casual models. Path analysis and structural equations models are indeed well known by social scientists, but both were restricted to quantitative variables at their early stages. In this paper, we will propose a new way to display the connections between qualitative variables in a similar way to the correspondence analysis, but using another set of multivariate techniques, such as lineal and logistic regression, mixed with network analysis.

Network coincidence analysis is to be used for the exploratory analysis of survey data. For this purpose, nodes represent the different categories of the selected variables, while links symbolize the relationships between the different variables. One of the specific uses of this analysis technique involves the characterization by diverse sociodemographic variables of different response profiles. Besides correlation measures, the proposed analysis can estimate log-linear models to study multivariate relationships including interactions.

Furthermore, to increase the analytic power of these tools, they have interactive characteristics online, which include either the selection of the elements according to their size or attributes, and the filter of the most central and strongest links.

The first part of the paper deals with the statistical basis of these representations; the second proposes a web page to apply this analysis to your own data, and the third gives examples of their use in international comparative surveys, such as the European Social Survey.

Keywords: network coincidence analysis, multivariate graphs, interactive graphs on line, coincidence and regression graphs

Systemic risk in Europe from a compositional perspective

Anna Maria Fiori[1], Germà Coenders[2]

[1]University of Milano-Bicocca, Italy; [2]University of Girona, Spain; anna.fiori@unimib.it

In response to the global financial crisis, the traditional firm-level approach to banking regulation has been progressively supplemented with macro-prudential policies aimed at ensuring the stability of the whole financial system. The focus of these policies is on measurement and monitoring of systemic risk, which relates to the risk of capital shortage for financial institutions under stress, and can thus be quantified by the amount of money that a firm (bank or insurance company) would need to raise in order to function properly in a crisis event. Nationwide measures of systemic risk can be obtained by aggregation of firm-level measures of capital shortfall (cf. Engle, “Systemic risk ten years later”, Annual Review of Financial Economics, 10, 125–152, 2018). This aggregation mechanism points to a compositional data (CoDa) perspective, in which different countries may be viewed as parts of a wider financial framework and their relative systemic risk contributions become the focus of the analysis.

In this study, we investigate the evolution of systemic risk in the Euro area during the period 2009-2022, representing the single Eurozone members as parts of a compositional dataset with monthly observations. Based on a constrained hierarchical clustering procedure (cf. Grimm, “CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis…”, Computers and Geosciences, 13, 13-35, 1987), we classify the dataset into a sequence of time-adjacent CoDa samples and we identify some trajectories of systemic risk concentration with the help of CoDa biplots. We also perform a cluster analysis of countries according to Ward’s hierarchical method, showing that a few groups retain a relevant proportion of total variability in the CoDa set. These findings suggest new interpretations for the interplay of core and peripheral Eurozone countries in alternating phases of the global financial crisis and the subsequent sovereign debt crisis.

Keywords: CoDa analysis, SRISK, constrained clustering, Ward’s method

Political taste and its homologies

Magne Flemmen[1], Nora Essahli[1], Jan Fredrik Hovden[1]

[1]University of Oslo, Norway; m.p.flemmen@sosgeo.uio.no

Recent years have seen an emerging literature is drawing on the ideas of Pierre Bourdieu to cast new light on questions of class and politics. One point of this literature is that political proclivities may be rather similar to cultural lifestyle and taste, so that they might be conceptualized as "political taste". However, it is unclear if political preferences actually abide by the same logic as cultural taste, which questions the usefulness of the concept. One way of investigating this is to look for a homology between position-takings in these different spheres and pay particular attention its concrete empirical patterning. This involves questioning which actors actually exhibit homologous position-takings, and which don't.

However, there are also methodological issues. Much of the work that looks for "political taste" relies on questions on political attitudes. But the uneven cognitive and statutory competence of actors and the different modes of production of political opinion suggests that attitudinal questions might be ill suited to construct political taste empirically. In this paper, we extend this line of reasoning. First, we try to provide a clearer and theoretically sounder conception of "political taste", linking it to Daniel Gaxie's concept of political market. In this understanding, political taste should refer to a taste for the products on offer on that market – parties, politicians, policies, etc. Drawing on unusually detailed survey data, we then use MCA to construct a space of political tastes, before examining its homology with the space of lifestyles. This shows that while there are certainly overlaps between political and cultural tastes, this is less systematic and complete than both of their connections to the social space. This raises questions about whether – and which! – actors orient themselves in politics with the same set of dispositions they draw on when navigating the space of lifestyles.

Pedagogical and disciplinary positions in the Danish field of higher education

Jan Thorhauge Frederiksen[1]

[1]University of Copenhagen, Denmark; janjaf@hum.ku.dk

Most correspondence analyses of national fields of education use enrolled students (through properties such as capital volume, composition, etc.) as cases of the analysis, in accordance with Bourdieu's analyses of social space. This design allows the researcher to examine selectivity, and hierarchies of status within the field, and determine how these relate to student positions. However, students are not the only agents within higher education (HE).

Examining the actions of other agents, specifically institutions, disciplines or individual programs of HE is complicated with such a construction of the field. Yet the relation between pedagogical practices and selectivity is of great sociological interest. Numerous pedagogical practices, e.g. whether to teach through lectures, work in project-based groups, require work-practice and so on relates to local disciplinary, institutional and cultural assumptions about students and learning. The relation between such pedagogical practices and selectivity within HE has not been studied systematically.

In this paper the Danish HE programs serve as cases of the analysis (n=927). I collected a prosopography of data describing both quantitative aspects of these programs (e.g. admission criteria, gender distribution, age distribution, number of exams and institutional affiliations, dropout rates, national quality measurements etc.) and qualitative characteristics of the programs categorized (e.g. teaching formats, exam forms, degree of professionalization, work practice, etc.). Through a specific MCA of these data, I construct a space of pedagogical positions within the Danish field of HE. Using structured data analysis and Euclidean classification, I show how the usage of strongly classified and framed pedagogy (cf. Bernstein) is structured by intersections between program status, student cultural capital, gender. Using additional data from qualitative student group interviews (n=27), I show students boundary work, quality norms and performativity in the classroom (cf. Lamont, Boltanski and Macfarlane) to also be structured by the above factors.

Keywords: Higher education, pedagogy, selectivity, disciplinary cultures, multiple correspondence analysis

My life in pictures: tales of a graphic developer and amateur historian of #dataviz

Michael Friendly[1]

[1]York University, Canada; friendly@yorku.ca

Come join me for a smorgasbord menu of topics in data visualization and its history. In my day job, I work on developing and implementing graphical methods for areas which were initially lacking: particularly categorical data and multivariate linear models. A chef's selection of appetizers will be served: all low-fat but sometimes high in dimensionality. All modern graphical methods have deep, but often forgotten, roots. The main course stems from a chance encounter with a magical graphical display from 1884 that caused me to catalog and study the vast history of data visualization. I'll serve a tasting menu of small bites of this history. For dessert, I'll provide a tasting menu of vignettes from my favorite period, the Golden Age of Data Graphics (1850-1900). Please do not overeat.

Keywords: categorical data, multivariate models, Golden Age

A cartography of the social-ecological transformation in Germany: exploring the interconnections between classes, mentalities and modes of living

Martin Fritz[1], Dennis Eversberg[1]

[1]Friedrich Schiller University Jena, Germany; martin.fritz@uni-jena.de

In our paper we apply a Bourdieusian relational methodology to study current societal struggles around a social-ecological transformation in Germany and try to answer the question why there is, despite scientific evidence that points to the urgent need, still not enough climate action. While knowledge and awareness of the causes and solutions are both very high, societies are not succeeding in the necessary change of infrastructures, consumption and production patterns. There is a lack of mobilization, in Latour's words, which is more than just an attitude-behavior gap on the individual level, it also reflects power structures in society and the hegemony of mentalities oriented at growth and nature domination.

In this context, we ask: What socio-ecological mentalities and modes of living exist among the general population? Do the societal elites, i.e. the economic upper or ruling class, hold mentalities that are opposed to a social-ecological transformation? What other potentially more pro-transformative mentalities exist? And do everyday practices correspond to the socio-ecological mentalities or are there mentality-practice gaps similarly to the attitude-behavior gap?

In our empirical study we try to answer these questions with data from a representative survey we conducted in 2021/22 among the German population. Applying methods such as principal component analysis, multiple correspondence analysis and cluster analysis on a wide range of questions about respondents' socio-demographic characteristics, socio-ecological attitudes and practices we distinguish class fractions, mentalities and modes of living. These three elements represent the analytical layers in Bourdieu's theory of practice: positions, dispositions and position-takings (or positionings). By locating them in social space we discover their homologies and heterologies and draw a detailed and empirically rich cartography of the state of the social-ecological transformation in Germany. Political and societal conflicts are revealed as well as potential alliances for more ambitious climate action.

Keywords: climate change, Bourdieu, mentalities, modes of living, social space, social-ecological transformation

Extend the use of supplemental variables in GDA by applying machine learning to the free text descriptive response portion and combining it with MCA analysis

Kazuo Fujimoto[1]

[1]Tsuda University, Japan; kazuo.fujimoto2007@gmail.com

The practice of linking the distribution of individuals within the space revealed by MCA with qualitative surveys has been mentioned in the book [1] and practiced in research activity [2]. In Japan, KH Coder [3] as a text analysis tool has been remarkably popularized and used in many social surveys. It is possible to link this text analysis with the selected answers using functions within KH Coder. Our first attempt as a mixed research method is to use this functionality. The next step is to add the frequently occurring words (important words) obtained at this stage to the individual coordinates as supplementary variables in the MCA and to analyze them by a GDA method [4].

In this report, as the next step, we report an example [5] in which frequently occurring words (important words) were tagged as positive/negative by the machine learning process and analyzed as supplementary variables.

This approach extends the use of supplementary variables in GDA.

References:

[1] Le Roux, Brigitte, & Henry Rouanet. 2010. "*Multiple correspondence analysis*.", Quantitative applications in the social sciences 163. Thousand Oaks, Calif: Sage Publications. "Between quantity and quality, there is geometry."p1
[2] Tony Bennett, Mike Savage, Elizabeth Silva, Alan Warde, Modesto Gayo-Cal and David Wright al, "Culture, Class, Distinction",2009,2010, Routledge
[3] https://khcoder.net/en/
[4] with [1] and using the GDAtools package of R. Robette N. (2023), *GDAtools : Geometric Data Analysis in R*, version 2.0, https://nicolas-robette.github.io/GDAtools/
[5] Kazuo Fujimoto and Kazuya Ohata, "Development of a method for analyzing participant satisfaction survey data that combines MCA and Aspect Based Sentiment Analysis."(in Japanese), NLP2023 (https://www.anlp.jp/proceedings/annual_meeting/2023/pdf_dir/Q1-11.pdf)

Keywords: MCA, GDA, mixed methods, machine learning

Combining categories for each variable in a multivariate categorical data set, with minimum loss of variable association

Michael Greenacre[1], Maurizio Vichi[2]

[1]Universitat Pompeu Fabra, Spain; [2]Università di Roma La Sapienza, Italy; michael.greenacre@upf.edu

The combination of categories in a multivariate categorical data set is investigated, with the technical objective of reducing the number of categories for each variable while having a minimum effect on the variable associations. The variable associations are measured by the "adjusted inertia" in multiple correspondence analysis, i.e. the average of the inertias of all pairwise contingency tables (Greenacre 2017, chapter 19). The overall aim is to arrive at fewer categories (and, thus, lower data dimensionality) without affecting the essential features of the data set, thereby simplifying the interpretation of any analysis using the categorical variables. The categories can be of an ordinal or nominal nature. Categories of an ordinal variable can only be combined with adjacent categories, while any categories of a nominal variable can be combined. The case of ordinal categories along with a "free" category, for example indicating a missing value, is also considered, in which case the ordinal categories have to combine with adjacent categories while the free category can combine with any of the ordinal ones. Results are given in the form of a dendrogram, indicating which categories are successively merged. This approach, which generalizes that of Greenacre (1988), is illustrated using sociological survey data from the International Social Survey Program.

In a companion paper, Vichi & Greenacre consider the aggregation of categories of different variables, in order to identify a reduced set of latent categorical variables with minimum loss of inertia.

References:

Greenacre, M. (1988). Clustering the rows and columns of a contingency table. *Journal of Classification* 5: 39-51.

Greenacre, M. (2017). *Correspondence Analysis in Practice. Third Edition*. Chapman & Hall / CRC Press. Boca Raton, Florida.

Keywords: chi-square, contingency table, inertia, multiple correspondence analysis, Ward clustering

Nonlinear prediction by kernels made explainable

Patrick Groenen[1], Michael Greenacre[2]

[1]Erasmus University Rotterdam, The Netherlands; [2]Universitat Pompeu Fabra, Barcelona, Spain; groenen@ese.eur.nl

The use of kernels for nonlinear prediction is widespread in machine learning. They have been popularized in support vector machines (SVMs) and used in kernel ridge regression (KRR), amongst others. Kernel methods share three aspects. First, instead of the original matrix of predictor variables, each row is mapped into a high dimensional feature space. Second, a ridge penalty term is used to shrink the weights (coefficients) on the predictors in the high-dimensional feature space. Third, the solution is not obtained in this feature space, but through solving a dual problem. So far, a major drawback in the present use of kernels is that the interpretation in terms of the original predictor variables is lost.

In this presentation, we argue that in the case of a wide $n \times p$ matrix of predictor variables (with p greater or equal to n), the kernel solution can be re-expressed in terms of a linear combination of the original matrix of predictor variables and a ridge penalty that involves a special metric. Consequently, the exact same predicted values can be obtained as a weighted linear combination of the predictor variables in the usual manner and thus can be interpreted. In the case p smaller than n, we discuss a least-squares approximation of the kernel matrix that still allows the interpretation in terms of a linear combination. It is shown that these results hold for any function of a linear combination that minimizes the weights and has a ridge penalty of these weights, such as in kernel logistic regression and kernel Poisson regression. As a consequence, the kernel methods can be seen as linear methods that are penalized by a ridge penalty in a peculiar metric that is directly derived from the kernel matrix.

Keywords: nonlinear prediction, ridge regression, interpretable AI

Applying sparse DiSTATIS to categorical data

Vincent Guillemot[1], Hervé Abdi[2], Victoire Baillet[1], Ju-Chi Yu[3]

[1]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France; [2]School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, TX, USA; [3]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada; vincent.guillemot@pasteur.fr

DiSTATIS (Distance-based STATIS) is a well-established multivariate analysis method that builds on the use of a compromise matrix derived from positive semidefinite transformations of distance matrices. DiSTATIS has been widely used for continuous data. However, its application to categorical data remains challenging. Here, we present ScaDiSTATIS (Sparsity-based Categorical DiSTATIS), a novel method expanding the DiSTATIS framework to categorical data. Using adapted distances and incorporating sparsity, our method offers a powerful approach to unravel complex patterns and extract meaningful information from high-dimensional categorical datasets.

ScaDiSTATIS handles categorical variables through the use of tailored metrics, such as Hellinger or chi-squared distances, to accurately capture dissimilarities between the variables under scrutiny. The optimization algorithm is then performed in a vector space equipped with these new metrics. Compared to a more classical strategy, namely, replacing the conventional RV coefficient matrix with a distance-based similarity matrix, our approach provides a more meaningful analysis of categorical data.

In addition to addressing the categorical data challenge, ScaDiSTATIS incorporates sparsity to effectively handle high-dimensional datasets by incorporating sparsity constraints in the optimization problem. This extension allows for the extraction of simple yet informative features from complex categorical data, facilitating interpretation and enhancing computational efficiency.

As a proof-of-concept, we applied ScaDiSTATIS to neuroimaging and genetics datasets. In the former, ScaDiSTATIS enabled the identification of key features in brain imaging studies, aiding in the understanding of neurological disorders and patterns of brain activity. In the latter, ScaDiSTATIS provided insights into the relationships between genetic markers, facilitating the identification of important genetic variations associated with specific phenotypes.

Albeit preliminary, these analyses demonstrate the robustness and relevance of ScaDiSTATIS for high-dimensional, categorical datasets of various kinds, paving the way for its applicability in a wide range of research fields.

Keywords: sparsity, DiSTATIS, categorical distances

Social identities and sociocultural adaptation of Moroccan and Turkish immigrants and their descendants: a study using cluster and correspondence analysis

Rahim Hajji[1]

[1]Hochschule Magdeburg-Stendal, Germany; rahim.hajji@h2.de

The issue of social identity and sociocultural adaptation of immigrants and their descendants has been the subject of repeated controversial debates in research. This presentation discusses the relationship between social identity and sociocultural adaptation of Moroccan and Turkish immigrants and their descendants in Germany, France, Belgium and in the Netherlands. The lecture will present the various theories that attempt to explain social identity and sociocultural adaptation, including assimilation theory, segmented assimilation theory, and acculturation theory. Assimilation theory states that adaptation processes lead to sociocultural adaptation of immigrants and their descendants over time. Segmented assimilation theory emphasizes that sociocultural adaptation depends on the social environment of immigrants and their descendants. Acculturation theory posits that people who develop a bicultural identity are more likely to adapt socioculturally to the host society.

However, selected research findings suggest that immigrants and their offspring who identify more strongly with the host country and behave more assimilatively tend to have better sociocultural adjustment than those who develop a bicultural identity. Against this background, the question arises as to which theoretical approaches are suitable for describing the social identity and sociocultural adaptation of Moroccan and Turkish immigrants and their descendants. The basis for the analysis of the explanatory power of the theoretical approaches is the SCIICS data set, which is based on surveys among Turkish and Moroccan immigrants and their descendants in Germany, France, Belgium and in the Netherlands. The results of the cluster analysis show that the social identities of Moroccan and Turkish immigrants and their descendants can only be partially described by acculturation theory. This is because bicultural identities are hardly observed. The results of the correspondence analysis suggest that the segmented assimilation theory better describes social identities and sociocultural adaptations.

Keywords: social Identity, sociocultural adaptation, Moroccan and Turkish immigrants and their descendants

Graph embeddings with influential outliers using correspondence analysis

Cassandra Handan-Nader[1]

[1]Stanford University, United States of America; slnader@stanford.edu

Correspondence analysis (CA) and canonical correspondence analysis (CCA) are popular methods in the natural, social, and biomedical sciences to summarize the interactions between two categorical variables in a contingency table. In this paper, I use a graph embedding derivation of correspondence analysis to show that both methods require a connectedness assumption on the one-mode similarity graphs that emerge from the contingency table in order to work as intended. I show that, when the connectedness assumption is violated, CA and CCA estimate completely different quantities that may not be obvious to the researcher, and even when the assumption is technically met, influential outliers in nearly disconnected segments of the graph can pull the solution towards them in undesirable ways. I offer two approaches to address disconnection and near-disconnection driven by small coalitions of influential outliers: a novel conductance index based on a Markov Chain interpretation of CA to identify and remove influential outliers in graphs following preferential attachment processes, which are common in the ecological and social sciences, and using covariates to integrate outliers into a well-defined reference space using either CA or CCA. I demonstrate the efficacy of these approaches on a large and complex political fundraising network that is frequently used to study ideological polarization in the pool of candidates for U.S. federal office, which spans three decades and includes nearly 20,000 candidates for federal office and 3 million of their donors.

Keywords: graph embeddings, outlier detection, preferential attachment processes, political fundraising

A method for estimating individual socioeconomic status of Twitter users

Yuanmo He[1], Milena Tsvetkova[1]

[1]London School of Economics and Political Science, United Kingdom; y.he54@lse.ac.uk

The rise of social media has opened countless opportunities to explore social science questions with new data and methods. However, research on socioeconomic inequality remains constrained by limited individual-level socioeconomic status (SES) measures in digital trace data. Following Bourdieu, we argue that the commercial and entertainment accounts Twitter users follow reflect their economic and cultural capital.

Adapting a political science method for inferring political ideology, we use correspondence analysis to estimate the SES of 3 482 652 Twitter users who follow the accounts of 339 brands in the United States. We validate our estimates with data from the Facebook marketing API, self-reported job titles on users' Twitter profiles, and a small survey sample. The results show reasonable correlations with the standard proxies for SES, alongside much weaker or non-significant correlations with other demographic variables. The proposed method opens new opportunities for innovative social research on inequality on Twitter and similar online platforms.

Keywords: socioeconomic status, Twitter, correspondence analysis, measurement, social media, cultural capital

Some relationships of correspondence analysis with other methods and models

Peter van der Heijden[1]

[1]Utrecht University / University Of Southampton; p.g.m.vanderheijden@uu.nl

We compare correspondence analysis (CA) with some statistical models that have a different form but related aims.

First, we compare CA with loglinear models with bilinear terms, and show that results of CA and these models are remarkably similar. Second, if we view CA as a method for factorization of a matrix, we will show that results are remarkably similar to non-negative factorization methods of a matrix such as latent class analysis of a two-way table and latent budget analysis. Also, in the same way, for higher-way arrays joint CA provides results that are remarkably similar to the non-negative higher-way factorization provided by latent class analysis.

CA and these statistical models are well known in statistics. In comparing these models with (J)CA, we also mention relations to work in the computing sciences, where loglinear models with bilinear terms, latent class and latent budget analysis are well-known, be it under different names, but CA has received much less attention.

Inglehart index revisited: a comparative scaling of materialism – postmaterialism using multiple correspondence analysis and the generalized graded unfolding model

Jörg-Henrik Heine[1], Christian Tarnai[2]

[1]Technical University of Munich, Germany; [2]Universität der Bundeswehr München, Germany; joerg.heine@tum.de

Postmaterialistic value orientation according to Inglehart is typically operationalized in the form of a ranking. In the *German General Social Survey* (ALLBUS), respondents are given four statements, with the instruction to rank them in order of importance. Two indicators refer to materialism and two to post-materialism. An index with four levels is formed from the rankings, after which respondents are classified into four groups: Postmaterialists, PM-mixed type, M-mixed type, and Materialists.

In the present analyses, two forms of scaling are used alternatively for index construction. One is multiple correspondence analysis (MCA) using the method of doubling the data matrix appropriate for rank data. Second, scaling is performed using the generalized graded unfolding model (GGUM). All analyses are performed using the 'FactoMineR' and 'GGUM' packages for the free R statistical environment.

The analyses are based on the ALLBUS data from 1991 to 2018. The results of the MCA show that the items can be mapped according to their ranking respectively between the poles materialism – postmaterialism in a two-dimensional space. In this space, the manifest index can be projected as a supplementary variable.

The results from unfolding scaling can be graphically represented in a comparable way as a common, but in contrast to MCA unidimensional, trait continuum for items and groups of persons. In this trait continuum, the items arrange between materialism and postmaterialism according to their rank order. The single integer categories of the manifest index can be represented by linking across persons with the GGUM person estimators in this unidimensional trait continuum. The diagnostic quality of the three measures – the manifest index, the object scores (MCA), and the person estimators (GGUM) – is discussed comparatively.

Keywords: multiple correspondence analysis, generalized graded unfolding model, Inglehart index, postmaterialism, ranking, preference data

Data analytic understanding of statistics

Christian Hennig[1]

[1]University of Bologna, Italy; christian.hennig@unibo.it

The title refers to the pursuit of understanding how the methods of statistics and data analysis process and transform the data, and what this implies. Statistics is predominantly taught and interpreted based on probability models, but, as pioneers of data analysis such as Tukey, Gower, and Benzécri knew, there is far more to the understanding of statistical methods than their performance assuming certain models.

This includes some rather elementary considerations with wide-ranging consequences that are rarely taught or discussed, such as the different effects of different standardization techniques when aggregating variables, or what a hypothesis test actually does in case the assumed model is not true (pretty much always, that is).

My presentation will make an appeal to look at our methods from the angle of a direct interpretation of what they do to the data. I will also discuss how this is philosophically different from setting up probability models and trying to infer "truths" assuming them, and what the role of models can be for the data analytic understanding of statistics if not assumed to be true.

Keywords: foundations of statistics, data science, probability models, robustness, data visualization

Analysing temporal stability and temporal change by way of Class Specific MCA (CSA). An exploratory presentation.

Johannes Hjellbrekke[1]

[1]University of Bergen, Norway; johs.hjellbrekke@uib.no

The problem of how to analyse temporal stability and temporal change remains a key question in sociological analyses, e.g in comparative studies of social mobility (see for instance Breen [ed.] 2004). When data are analysed by way of multiple correspondence analysis (MCA), the perhaps most commonly used alternative has been to project individuals, variables and categories measured at timepoint 2 into a space constructed from variables measured at timepoint 1, to calculate the mean modality points within the space of reference and to compare the distances between two or more points at timepoints 1 and 2 (see for instance Le Roux 2014, Rosenlund 2019).

This presentation seeks to explore some of the potentials that are inherent in the combination of MCA and a technique developed by prof. Brigitte Le Roux, class-specific MCA (CSA) (Le Roux, op.cit.). By analysing two different data sets from Norway, one on cultural participation and one on media preferences, three questions will be addressed:

(i) Have the dimensionality and the structure of primary axes within these spaces been stable or subject to change?

(ii) Are the oppositions within these spaces weakening or strengthening over time?

(iii) How has the connection between cultural and media divisions and class divisions developed over time.

Our exemplary case is Norway and the data stem from three different rounds of The Culture and The Media Surveys, conducted by Statistics Norway between 2012 and 2021.

References:

Breen, Richard (ed.) (2004). *Social Mobility in Europe.* Oxford: Oxford U.P.

Le Roux, Brigitte (2014). *Analyse géometrique des données.* Paris: Dunod.

Rosenlund, Lennart (2019). "The persistence of inequalities in an era of rapid social change. Comparisons in time of social spaces in Norway." In *Poetics*, Volume 74, June 2019, https://doi.org/10.1016/j.poetic.2018.09.004

Keywords: social space, stability, change, class specific analysis (CSA)

Well-bred and well-spoken: on the role of class origins for children's experiences with rhetorical speech in schools public speech competitions

Jan Fredrik Hovden[1]

[1]University of Bergen, Norway; jan.hovden@uib.no

How do young adults experience public speaking, and how does their experience depend on their social background? Using a national speech competition in Norwegian secondary schools as a natural experiment for studying the relationship between class background and the use of language, a survey (43 school classes, N=940) identifies their experiences of participating in the course as varying much. Their responses, organised after the classical steps in rhetorics from inventio to actio, are used to construct a space of experienced rhetorical mastery using principal component analysis, which structuration is then explored in light of class backgrounds, indicators for the use of language in the family, and the content of their speeches.

Working-class pupils in general found the writing and delivery of a public speech often much more difficult, boring, meaningless and stressful, and less often reported a family culture favourable to the cultivation of public speaking, e.g. less often having parents who liked public speaking, discussed societal issues with them or helped them with their school writing and oratory assignments. Working-class pupils also tended to choose private, non-controversial themes rather than political themes. Girls enjoyed the writing process more than the oral delivery, while it was the reverse for boys. The study in this way identifies basic class-based differences in the relation to language and public speaking, which in important ways appears as an internalisation of more or less probable futures, of being speakers or listeners.

Keywords: class, rhetoric, language, social reproduction, Bourdieu

Analyzing multivariate densities in Bayes spaces with applications

Karel Hron[1]

[1]Palacky University Olomouc, Czech Republic; karel.hron@upol.cz

Probability density functions can be embedded in the geometric framework of Bayes spaces, which respect their relative nature and enable further modeling and analysis. Specifically, the Hilbert space structure of Bayes spaces (of which compositional data are one specific instance) has several important implications for classical and Bayesian inference, as well as functional data analysis. In this contribution, an orthogonal decomposition of multivariate densities in Bayes spaces using a distributional analog of the Hoeffding–Sobol identity is constructed. The decomposition is based on a reformulation of the standard (arithmetic) margins into so-called geometric margins. These are orthogonal projections into one-dimensional space of information contained in multivariate densities and coincide with the arithmetic margins in case of independence. More generally, the decomposition contains an independent part and all possible interaction terms. The orthogonality of the decomposition results in an analog of Pythagoras' Theorem for squared norms of the decomposed densities and margin-free property of the interaction terms. Because the squared norms from the Pythagoras' decomposition are in essence compositional in nature, all tools of the log-ratio methodology, including visualization using compositional biplots, can be used. Theoretical results will be illustrated with empirical geochemical data. This talk is based on joint work with Christian Genest and Johanna Nešlehová from McGill University, Montréal, Canada.

Correspondence analysis in gene expression of single cells

Lauren Hsu[1, 2], Aedin Culhane[3]

[1]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, USA; [2]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA; [3]Limerick Digital Cancer Research Centre, Health Research Institute, School of Medicine, University of Limerick, Limerick, Ireland; laurenhsu@hsph.harvard.edu

Gene expression is a key biological measure for understanding and modeling the function and state of biological systems. Single cell mRNA sequencing (scRNAseq) simultaneously measures the counts of more than 20,000 genes expressed in an individual cell. This technology is applied to measure gene expression in thousands of functional diverse cells in a tissue or biological experiment. A single study typically contains at least 10,000 and can include 1 million or more cells. Due to biological heterogeneity and specialization of cell functions, many genes can be cell-type specific, leading to sparse, overdispersed high dimensional data. Effective dimension reduction is essential for scRNAseq analysis. Principal component analysis (PCA) has been widely used, but since it requires continuous, normally-distributed data, it is often coupled with log-transformation in scRNAseq applications, which can distort the data and obscure meaningful variation. We study the application of correspondence analysis (CA) to scRNAseq as an approach to avoid distortive log-transformation. However, standard CA, singular value decomposition (SVD) of the Pearson residuals, has known limitations when applied to overdispersed data. We describe and benchmark five fast, scalable, adaptations of CA applied to overdispersed, highly sparse scRNAseq data. Each computes cell embeddings with more performant or comparable clustering accuracy in 8 out of 9 test datasets to standard CA with Pearson residuals. We report that CA with Freeman–Tukey residuals performs well across a diverse range of datasets. We describe a CA extension to multi-table analysis; corralm for integrative multi-table dimension reduction of scRNAseq data. We implement CA for scRNAseq data in, corral, an R/Bioconductor package which interfaces directly with single cell classes in Bioconductor and computationally efficient partial SVD methods such as irlba. Switching from PCA to CA is achieved through a simple pipeline substitution and improves dimension reduction of scRNAseq datasets.

Keywords: bioinformatics, overdispersed, high-dimensional, biology, sparse

Association-based learning for mixed data

Alfonso Iodice D'Enza[1], Angelos Markos[2], Michel Van de Velden[3], Carlo Cavicchia[3]

[1]University of Naples Federico II, Italy; [2]Democritus University of Thrace; [3]Erasmus University of Rotterdam; iodicede@unina.it

Several statistical learning methods, both supervised (e.g. k-nearest neighours) and unsupervised (e.g. spectral clustering, partitioning around medoids), take a matrix of pairwise distances as input. Obviously, the distance measure adopted affects the obtained results, and it should be carefully chosen. There is, however, a wide range of distance measures to choose from, more so if the observations at hand are described by continuous and categorical attributes. On one side of the spectrum there is the subjective selection of well-established distance measures: Euclidean for continuous data, matching-based dissimilarity for categorical data, and the Gower index for mixed data. On the other side of the spectrum there is distance metric learning (DML), a field of machine learning based on a fully data driven definition of custom distance measures. DML seems a reasonable approach to define the ‘best' distance measure, given a classification or a regression task. What is meant by ‘best' is not easily defined when it comes to unsupervised tasks, such as clustering. Furthermore, most DML methods are designed for continuous data only, as the categorical attributes, if any, are re-coded to numerical attributes and processed as such. The present proposal falls somewhere in between the two sides of the spectrum mentioned above: that is, not subjective nor fully data-driven distance selection. Exploiting a general framework that embeds several distances proposed for categorical (and, mixed) data, we propose a task-coherent approach to refine the distance measure selection both in the case of supervised learning (KNN) and unsupervised learning (spectral clustering).

Keywords: mixed data, spectral clustering, nearest neighbours

Social space and principle of differentiations in Japan

Naoki Iso[1]

[1]Tokyo University of the Arts, Japan; iso.naoki@pc.geidai.ac.jp

In this research presentation, we will construct the social space of Japan by using MCA of data from a postal survey we conducted in January 2023. The problem in this report is the choice of variables when using MCA: while considering the methods of constructing 'social space' by Flemmen et al. (2018, 2019) and Atkinson (2020, 2022), I chose my own variables, which differ from theirs.

As a result, we found the total volume of capital on the first axis, attitudes toward schooling on the second axis, and cultural excellence on the third axis.

We analyse the relationship between culture and inequality in contemporary Japan from a theoretical and empirical sociological perspective, in terms of 'culture-caused inequality' and 'inequality in access to culture'. In other words, it addresses two types of issues: i) what is culture-caused inequality, and ii) what is the unequal distribution of cultural resources. In order to study these issues, the cultural capital concept is theoretically developed as a theory of social differentiation. Postal and interview surveys are conducted in the Kanto region in Japan. The data collected in the survey are analysed using a mixed research method that uses the characteristics of multiple correspondence analysis. This study captures the relationship between culture and inequality by applying the cultural capital concept but does not aim to find a proxy variable for cultural capital. Instead, the primary aim of the research is to elucidate cultural capital as a component of one of the principles of social differentiation, and the causal relationship between culture and inequality is also in its scope.

Keywords: capital, social space, MCA, Japan

Predictive analytics, causality and sensitivity analysis of association rules

Ron Kenett[1], Chris Gotwalt[2]

[1]KPA Ltd., SNI Technion, Israel; [2]JMP Statistical Discovery, LLC, USA; ron@kpa-group.com

Many modern organizations generate a large amount of unstructured textual transaction data, on a daily basis. These include logbooks of systems and processes, technician reports of product failures and social media blogs. Transactions typically include semantic descriptors that require specialized methods for analysis. Exploring association rule (AR) is a powerful semantic data analytic technique used for extracting information from transaction databases. AR analysis was originally developed in shopping basket analysis where the combination of items in a shopping cart are evaluated. To generate an AR, the collection of more frequent itemsets, a set of two or more items—is first detected. Then, as a second step, all possible ARs are generated from each itemset in this group. The ARs are then ranked using measures of association such as support, confidence, and lift. These measures are labelled “measures of interest”. The R package “arules” provides more than a dozen such measures including the relative linkage disequilibrium (RLD) which normalizes classical Euclidean distances of the itemset from a surface of independence. JMP (www.jmp.com) computes lift, support and confidence. The talk will show how to conduct a sensitivity analysis of such measures by repeatedly splitting the data into a training and validation sets. This helps determine the level of overfitting which affects the predictive analysis performance.

Minimal overfitting indicates that the AR performance in the training set is generalizable in future and similar data sets. Another approach, more similar to case control assessments, will be presented. It relies on the concept of fair data sets and enables a discussion based on counterfactuals.

Keywords: keywords association rule; relative linkage disequilibrium (RLD), befitting cross validation (BCV), sensitivity analysis (SA), fair data sets

Seeking pure pattern effects: latent level and pattern variance in profile analysis is evidenced by reparametrized observed level and pattern variance

Se-Kang Kim[1]

[1]Baylor College of Medicine, United States of America; Se-Kang.Kim@bcm.edu

The goal of the present study is to understand the central response patterns of cognitive ability over latent dimensions by identifying core profiles from an $I \times J$ rectangular-shaped data set where I rows represent person profiles of J column measurements. These central response patterns, however, have been confounded with row, column, or row-by-column interaction effects. The row effect refers to the row totals, and the column effect refers to the column totals. Consequently, the desired central pattern effects have been attenuated by the row, column, and row by column interaction effects. In this study, two parameterization techniques are presented to determine the optimal method for identifying "pure" pattern effect profiles that characterize central response patterns. The first method is to ipsatize core profiles around their means derived from the data, while the second is to identify core profiles from row-centered data around person (or row) means, proposing that either method or both can identify the pattern effect profiles that are not confounded by row, column, or interaction. To demonstrate, PCA was performed on the seven domain scores of the Woodcock-Johnson III tests of cognitive abilities, and four pattern effect core profiles were identified from the responses of 3,825 participants on the seven ability domains. Then, we estimated latent level and pattern variance from the core profiles and compared them with observed level and pattern variance estimated from the data; they were virtually identical. These results support comparability of the two parameterization methods, indicating that both methods are equally valid for identifying pattern-effect profiles that are uncontaminated by row, column, and interaction effects. Four validation checks were conducted to confirm the significance of the pattern-effect profiles. The insufficient psychometric properties of previous profile analyses are contrasted with the advantages of the current paradigm for profile analysis.

Invariant structures? Enduring relations of oppositions in the field of the Danish business elite from 1910-2020

Anton Grau Larsen[1], Jacob Lunding[2], Christoph Ellersgaard[2]

[1]Roskilde University, Denmark; [2]Copenhagen Business School, Denmark; agraul@ruc.dk, jlunding@gmail.com

This paper proposes a new analytical methodological framework for longitudinal studies of social fields. Within Bourdieusian field analysis, it is assumed that fields change over time as a result of struggles between the agents and institutions. Some elements of the structure of the fields may be stable and homologous over time, while others are less stable. But there is not yet a clear tradition for how to empirically show stability and change over longer time periods within field analysis. In this paper we analyse prosopographical data from the Danish equivalent to Who's Who with geometric data analysis to construct the field of the inner circle of the Danish business elite from 1910 until 2020. The study population is the core of a career network which varies in size from 80 to 300 business leaders, which is similar to what Michael Useem defines as the inner circle of the business elite. With multiple correspondence analysis we analyse 17 variables measuring various forms of capital tied to careers, social trajectory, symbolic resources and education. To strike a balance between comparability and a sensitivity to historical variations, we propose a dynamic relational coding scheme where the active categories adapt to the population each year resulting in between 48 to 60 active categories. We identify enduring oppositions between key categories that are opposed for more than 80 years. With a clique census between categories within enduring oppositional pairs we construct enduring homological structures of opposed profiles. Empirically we show that very few oppositions are stable over 110 years, except for those tied to recognition from the state, the royal court and international experience – indicating the long relationship between the structure of the field and the prevailing Danish growth model. We can construct three sets of homological structures that are stable over time in specific time periods.

Keywords: elites, history, fields, MCA

A generalized multiple correspondence analysis: bi-weighted multiple correspondence analysis

Brigitte Le Roux[1, 2, 3], Frédérik Cassor[2], Flora Chanvril[2]

[1]MAP5, Université Paris Cité, France; [2]CEVIPOF, Sciences-po Paris, France; [3]ENS Paris-Saclay, Université Paris-Saclay, France; brigitte.leroux@sciencespo.fr

This paper presents a general method for analyzing a table of individuals × categorical variables, with a weighting on individuals and a weighting on categorical variables. We give the properties and the matrix formulas for MCA as well as for specific MCA and class-specific analysis.

Keywords: multiple correspondence analysis, class-specific analysis, bi-weighting

Applying combinatorial inference in GDA. The case of European central bank governing council members (1999-2022)

Frédéric Lebaron[1], Brigitte Le Roux[1], Aykiz Dogan[1]

[1]ENS Paris-Saclay, France; flebaron@yahoo.fr

This contribution proposes an integrated methodology applied to an economic sociological issue, in the spirit of Bourdieu (2000). It is based on combinatorial inference in GDA, allowing to deal with the particular situation of a very small but exhaustive data set, where we detect a statistical effect of sufficient descriptive magnitude in a multidimensional framework.

In this situation, we develop an approach following a set of successive steps:

- Construction of a social space using MCA;

- Studying a structuring factor in the cloud of individuals and assessing their descriptive effects;

- Using the framework of combinatorial inference, in order to study: the typicality of a particular subgroup; the compatibility zone around the mean point; the homogeneity between two groups;

The subject of study is the factors determining the position-takings (stylised in three categories) of members of the governing council of the ECB since 1999. 85 individuals have composed this committee which decides over the interest rates and various components of the public policy of the institution, the central pilot of monetary policies in the Eurozone.

Members of the governing council are coded according to three categories of position-takings: hawk / dove / intermediary or uncertain position-takings. This approach allows to deal with very low frequencies and relatively scarce data, without leaving aside the multidimensionality and complexity of the object.

Keywords: MCA, social space, central bankers, combinatorial inference, low frequencies

The educational space of preschools in a cultural stronghold in Sweden

Ida Lidegran[1]

[1]Uppsala universitet, Sweden; ida.lidegran@edu.uu.se

In Sweden, the vast majority of families with children choose a preschool already during their child's first or second year of life. This is a very decisive choice; if you do not actively choose a preschool, you will not be offered a place. Thus, all families who need childcare must execute an educational related choice very early in the child's life, long before the child has an opinion on the matter. The early choice demanded of families is conditioned by the availability of specific preschools and of different types of preschools. With inspiration from Bourdieu's sociology of education the purpose of this paper is to examine the educational space of preschools in Uppsala, Sweden's fourth largest city and the one most dominated by social groups with large amounts of educational capital. Correspondence analysis has been used to construct the space of preschools in Uppsala. Data is provided by Statistics Sweden (SCB) and is anonymized, on individual level and contains information about children who attend preschool in 2017 and information about their parents. The results of the correspondence analysis show a three-polar structure: Social groups with the largest amount of educational assets are drawn to a special form of preschools, so called parent cooperatives, often smaller in size and with very homogenous social recruitment. In the other end of the space, the families that have the least of educational assets place their children in municipal preschools. In between these extreme groups, the economic fractions of the lower middle class and of the middle class are overrepresented in the part of the space where the private and for-profit preschools are most widespread. This structure corresponds well with the educational strategies that these social classes and class fraction develop at later stages of the educational system.

Keywords: preschool, correspondence analysis, Sweden, Bourdieu

The world-class ordination: studying a global sub-field of universities

Pablo Antonio Lillo Cea[1]

[1]Uppsala University, Sweden; pablo.lillo_cea@edu.uu.se

The notion of "world-class university" suggests the existence of a world-encompassing space where a select group of universities take positions in the struggles over the acquisition of the assets indicative of this class. Global university rankings are pivotal in validating these positions through a set of indicators presumed to reflect institutional excellence. Drawing on Bourdieu's field theory, this study investigates the symbolic order of universities and nations reproduced through ranking evaluations and identifies the key features that differentiate and allow the understanding of the relationships between competing institutions. The analysis incorporates the top 100 universities from the 2022 ARWU, QS, and THE rankings, amounting to 140 unique institutions. Data from these rankings are assessed using multiple correspondence analysis (MCA). Preliminary findings confirm the continued dominance of American and European institutions and underscore the significant roles of internationalisation and current reputation in distinguishing institutions, revealing diverging institutional strategies. An East-West divide is discernible, with Eastern universities generally exhibiting less international academic recruitment and collaboration compared to their Western counterparts. Nordic institutions stand out for having a lower academic reputation and a higher scientific reputation reflected by alumni and staff holding Nobel prizes and other awards. Through this exploration, the study advances a fresh perspective on the dialogue surrounding world-class universities and advocates for a more nuanced comprehension of the evolving landscape of higher education. Ultimately, this research calls for a rethinking of the implications and connotations of the world-class label within our increasingly globalised academic environment.

Keywords: world-class university, globalisation, higher education, rankings

Sparse non-symmetrical correspondence analysis

Ruiping Liu[2], Ndeye Niang[1], Gilbert Saporta[1]

[1]Conservatoire national des arts et métiers, France; [2]Beijing Information Science and Technology University, China; gilbert.saporta@cnam.fr

Sparse methods inspired by the Lasso have been successfully applied to principal component analysis and only recently to correspondence analysis in its multiple and contingency table versions. The principle consists in forcing a significant number of row and column weights to be zero to facilitate the interpretation of the components, in the spirit of the search for simple structures, especially for large arrays. This simplicity comes at the expense of the loss of characteristic properties such as orthogonality and double barycentric relations, hence the need for a trade-off.

Liu et al. (2023) developed a specific deflation algorithm to obtain orthogonal successive weight vectors, in a constrained generalized SVD. Although it is possible to sparsify only the rows (or columns) of a contingency table, their method remains fundamentally symmetric.

We propose here a sparse version of the non-symmetrical correspondence analysis (NSCA) introduced by Lauro and D'Ambra (1984), cf Beh and Lombardo (2021). Since NSCA is a weighted PCA of centered column profiles, or a generalized SVD, both row (the responses) and column (the predictors) weights may be sparsified, or only one set according to user's choice. Sparse NSCA will be illustrated on several data sets and compared to standard CA, sparse CA and standard NSCA.

References:

Beh, E.J. and Lombardo, R. (2021). *Non-symmetrical correspondence analysis.* In *An Introduction to Correspondence Analysis*, Chapter 3, 71-97, John Wiley.

Lauro, N. and D'Ambra, L. (1984). *L'analyse non symétrique des correspondances.* In E. Diday, ed.*, Data Analysis and Informatics III*, 433–446, Elsevier North-Holland.

Liu, R., Niang, N., Saporta, G., and Wang, H. (2023). *Sparse correspondence analysis for large contingency tables.* Advances in Data Analysis and Classification*,* 17: 1–20.

Keywords: non-symmetrical correspondence analysis, sparse PCA, generalized SVD

The Cressie-Read family of divergence statistics and correspondence analysis for two-way and multi-way contingency tables

Rosaria Lombardo[1], Eric Beh[2,3]

[1]University of Campania "Luigi Vanvitelli", Italy; [2]University of Wollongong, Australia; [3]Stellenbosch University, South Africa; rosaria.lombardo@unicampania.it

Correspondence analysis is a popular technique for visualising the association between categorical variables and has been extensively studied (Greenacre, 1984; Beh and Lombardo, 2014). It generally involves using Pearson's chi-squared statistic as its numerical foundation, however, the statistic is only one member of the Cressie-Read family of divergence statistics (Cressie and Read, 1984) which also includes the Freeman-Tukey statistic, the likelihood ratio statistic, and modified versions of these measures. Therefore, we discuss a general framework that explores how correspondence analysis can be performed using this family to suit different data structures. Special cases of the framework include the correspondence analysis using the Freeman-Tukey statistic (Beh et al., 2018), which is linked to the Hellinger distance decomposition (Cuadras and Cuadras, 2006) method, and the log-likelihood ratio statistic, which leads to log-ratio analysis (Greenacre, 2010).

Additionally, we will explore some modifications to this framework, including its application to multi-way correspondence analysis (Lombardo et al., 2021).

References:

Beh, E.J. and Lombardo, R. (2014). *Correspondence analysis: Theory, Practice and New Strategies*. Wiley, Chichester.

Beh, E. J., Lombardo, R. and Alberti, G. (2018). Correspondence analysis and the Freeman-Tukey statistic: A study of archaeological data. *Computational Statistics and Data Analysis*, 128, 73 – 86.

Cressie, N. A. C. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46, 440 – 464.

Cuadras, C. M. and Cuadras D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, 417, 64 – 74.

Greenacre, M.J. (1984). *Theory and Application of Correspondence Analysis*. Academic Press, London.

Greenacre, M. (2010). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, 42, 129 – 134.

Lombardo, R., Beh, E.J. and Kroonenberg P.M. (2021). Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables. *Statistical Science*, 36, 542 – 561.

Keywords: Cressie-Read family of divergence statistics, correspondence analysis, data visualization

The biplot zone for exploring truths in mixed scaled measurements

Sugnet Lubbe[1]

[1]Stellenbosch University, South Africa; slubbe@sun.ac.za

Traditionally, principal component analysis (PCA) biplots are associated with continuous scale measurements and multiple correspondence analysis (MCA) biplots with categorical scale measurements. In this paper we will look at different methods for visually representing data, some continuous measurements, some with categorical measurements, which can be either nominal or ordinal scale and sometimes a mixture of different data types.

Biplots such as the generalised biplot, introduced by Gower in 1992, make provision for a mixture of measurements scales in a single data set. Other methods, convert one measurement scale into the other. Categorical principal component analysis (CatPCA) assigns optimal scores to the categorical variables to allow for constructing a visualisation with PCA.

On the other hand, in 2011 Aşan and Greenacre introduced biplots of fuzzy coded data. In these biplots, where continuous data are visually represented using MCA.

We will delve deeper into the similarities and differences of various biplot methods and compare visualisations and the quality of fit. Specific attention will be given to the visual representation of data with a mixture of different measurement scales.

Keywords: biplots, principal component analysis, multiple correspondence analysis

Research and design. On the refiguration of an ambivalent relationship in the academic field of architecture

Séverine Marguin[1]

[1]Technische Universität Berlin, Germany; severine.marguin@tu-berlin.de

As an applied discipline, architecture traditionally has its curricula and therefore the composition of the professorships orchestrated by the core discipline - design - surrounded by auxiliary disciplines that bring the required knowledge from outside the discipline, i.e. from the humanities as well as from the engineering sciences. This composition corresponds to the allocation of roles between designing on the one hand and researching on the other. The current context of a scientification of German-speaking universities affects heavily the faculties of architecture. To what extent does this conjuncture affect the processes of recruitment at the faculties of architecture and the balance between research and design?

This article takes a closer look at the recognition processes of architects at German-speaking universities. For this purpose, a mixed-methods research design was used, consisting of ethnographic field research and multiple correspondence analysis. The aim was to use this method to reconstruct the German-speaking academic field of architecture in order to work out autonomisation factors, weightings and polarities or congruencies. A set of variables on all professors (n=478) at all German-speaking university faculties of architecture in 2020 was built: besides demographic variables (gender, year of birth, field of education), the data were collected along three areas, relevant in the legitimation process at work in this context : a) academic profile of the professorship (subject area, degree of education, generation of appointments, publications, third-part funding) ; b) practice profile (architectural chamber, own architectural practice, year of foundation, number of employees, orientation of the practice, awards and competitions); c) centrality of city in the field of architecture (size, budget and age of university, number and internationalization of architectural practices). The analysis shows a spatialised refiguration of formative dualities (research vs. design; engineering vs. art; artistic vs. political) for the discipline of architecture.

Keywords: MCA, sociology of science, architecture, design, centrality

A game theory application in the Greek elections of 2023. Political competition, mass media and public opinion data management.

Evangelia Nikolaos Markaki[1], Theodore Chadjipantelis[1]

[1]Aristotle University of Thessaloniki, Greece; chadji@polsci.auth.gr

The political elections in Greece 2023 were of prime importance and included very interesting aspects as far as the political competition is concerned. The dynamic interrelation of the political parties, the candidates and the issues were presented, considered and commented extensively during the pre-election period. In the present study we use a game theory approach of competition analysis to represent in matrices the competition between different parties and strategies toward different issues such as: economy, health, education, foreign affairs and defense, environment, new generation and young people, cooperation, behavior towards the opposite party/candidate.

We collected data from different sources of information: online newspapers, traditional newspapers, informational sites, debates, tv interviews, radio presentation, social media information and campaign publications. With group data analysis we find how many positive references the different strategies in the competition have in mass media and public opinion. The game theory matrices give us information about the best clear strategies when a Nash equilibrium exists and if not, the mixed strategy approach is applied so as to calculate the possibilities and the payoffs of each strategy used in the pre-election period.

This method creates a tool of expertise in the development of high-performance political campaigns because it gives the political parties the information needed about the strategies used in their presentation, their intensity and their efficacy.

Keywords: game theory, decision making, political competition, data management, political marketing

Bridging differences: a holistic framework for distance quantification in mixed-type variables

Angelos Markos[1], Alfonso Iodice D'Enza[2], Michel van de Velden[3], Carlo Cavicchia[3]

[1]Democritus University Thrace, Greece; [2]Università degli Studi di Napoli Federico II; [3]Erasmus University Rotterdam; amarkos@eled.duth.gr

Variability among subjects regarding specific properties, as quantified by a set of variables, is a critical consideration in many statistical methodologies. This is evident in techniques such as classification, clustering, and data visualization, which necessitate a measure of disparity in observed values. Such disparity measures, or 'distances', are contextually defined based on data characteristics and the particular problem under investigation. For numerical variables, many distance definitions exist, typically contingent on the magnitude of observed variations. However, determining a distance measure for categorical data is inherently more complicated due to the lack of a direct quantification method for observed differences. More so if both numeric and categorical attributes are considered. As a result, numerous proposed measures are available to gauge differences based on categorical variables. In this talk, we present a comprehensive framework designed to efficiently and transparently implement distances between observations on mixed-type variables. Our proposed framework accommodates various pre-existing distance measures and intuitively fosters the development of novel distance formulations. It also offers the potential for implementing flexible distance definitions tailored to specific cases and data.

Keywords: categorical data, distance, cluster analysis, classification

Poverty, happiness and economic freedom: a correspondence analysis with a doubling technique on Latin American Countries

Fernanda Mazzotta[1], Lavinia Parisi[2], Monica Consuelo Rodriguez Torres[3]

[1]University of Salerno, Italy; [2]University of Salerno, Italy; [3]University of Salerno, Italy; mazzotta@unisa.it

The paper aims to analyse the relationship among poverty, happiness, economic freedom and other institutional aspect (democracy, trust, corruption, inflation and growth) in 15 Latin American countries. The periods under consideration are two, the first is the time span between 2007-2009 and the second is between 2017-2019.

The countries selected are considered as the less poor among the Latin American ones. The method used to analyse these relationships is correspondence analysis. The findings show that in the countries where the poverty index (absolute definition) is low, there are higher levels of freedom (economically and politically) compared with countries where the poverty is high. However, having freedom does not indicate having more happiness, in fact the index considered as a proxy of happiness (i.e. the suicide rate) is high. The corruption index is positively linked with poverty and inflation as expected. On the other hand, poorer countries and with lower economic freedom show a higher level of happiness.

Keywords: poverty, happiness, economic freedom, developing countries, correspondence and cluster analysis, Jelcode: I32, D73, O15, C38

CARME – the rise and fall of a sociological space in the East

Karl Michael van Meter[1]

[1]Ecole normale supérieure, France; karl.vanmeter@ens.fr

When Mikhail Gorbachev in 1985 came to power in the USSR, his wife, Raisa, often described as an activist and a philanthropist interested in fostering new talent, decided she wanted to defend a doctoral thesis in sociology, which was then a sub-branch of the Marxist-Leninist Philosophy of the Soviet Academy of Science. Mikhail Gorbachev had an independent Institute of Sociology created in May 1988 (formerly the Institute for Concrete Social Research) under the direction Vladimir Jadov, who became the supervisor of Raisa's thesis. To establish the bona-fide scientific nature of sociology in the USSR, it was necessary to organize an internationally recognized sociology conference in the USSR, but the German Democratic Republic (GDR) had already a lead in this endeavor through German relations with the International Sociological Association (ISA) Research Committee 33 (RC33), "Logic and Methodology". The USSR seems to have "pulled rank" on the GDR and with UNESCO so that RC33 chaperoned the first – and the only – international sociology conference in the Soviet Union, in Moscow (24-27 October 1988), before the first – and the only – international sociology conference in the GDR (Holzhau, 2-6 October 1989). CARME work was used as the "litmus test" in both cases to show the world that empirical sociology existed in the East, but unfortunately not for long [the Berlin Wall opened on 9 November 1989, German reunification on 3 October 1990, and dissolution of the Soviet Union 26 December 1991].

Keywords: CARME, USSR, DDR, ISA RC33, BMS

Designed for success or failure: differences in funding and rejection in the space of applications to the Danish Art Foundation among craftsmen and designers

Sebastian Diemer Mørk[1], Anton Grau Larsen[1]

[1]Roskilde University, Denmark; sdmoerk@ruc.dk, agraul@ruc.dk

We map the exercise of aesthetic authority by looking at the funding practices of the Danish Art Foundation with a unique dataset of both accepted and rejected applications for funding for artistic crafts and design. Combining stochastic block modelling and GDA (Hyland et al., 2021; Le Roux & Rouanet, 2004), we construct a space of topics with the most prevalent topics in the applications for funding received by the Danish Art Foundation between 2014-2022 (n=4091). The stochastic block models approach identifies communities and when used on a textual corpus, these communities function as topics. This approach to topic modelling allows us to make fewer assumptions regarding the corpus beforehand. Hence, this approach develops the capabilities of both topic modelling and geometric data analysis. We relate the space of topics to social characteristics of the applicants and the sequence of applications highlighting the relational differentiation and hierarchy in the space. We are thus able to expound on the process of the creation and maintenance of aesthetic authority in the field of craft and design production (Bourdieu, 1996; Rössel & Weingartner, 2015). The comparison between accepted and rejected applications, not only positions them relationally in the same space, but also allows us to highlight how themes and topics differ between them. Often the successful applications define the field, hence using the rejected ones will allow a negative constitution of the field. With this paper we will highlight how two dynamics interact: artists who use the right topics have an easier time receiving funding, which triggers a Matthews-effect for their subsequent applications (Alexander, 2018; Peters & Roose, 2022a, 2022b).

Keywords: art, topic models, stochastic block models, aesthetic authority

The evolution of social inequalities in Brazil

Lucas Page Pereira[1]

[1]ENS Paris-Saclay, France; lucaspagepereira@gmail.com

How can geometric data analysis contribute to the better understanding of the evolution of social inequalities? Our study aims to investigate the evolution of the position of different social classes within the Brazilian social space during the 2000s, a period characterized by an overall improvement in living conditions, a significant increase in income (especially for the vulnerable strata) and in the level of education of Brazilians. After a brief presentation of the Brazilian context and the academic

debates about the changes experienced by the country at this time, we will present components of the space of living and working conditions in Brazil, specifically the six indicators related to the insertion of individuals in the labor market, six indicators related to the characteristics of the household's housing, and five indicators reactive to family life. Then, we will discuss the interpretative consequences, from a sociological point of view, of taking as reference the space situated at the end of the analyzed period (2015) or the space situated at the beginning of this period (2002). Finally, we will show how the approach adopted allows us to apprehend and objectify aspects of the evolution of inequalities in Brazil that are barely visible through traditional approaches to social stratification. For this, we will use data from the National Household Sample Survey (PNAD-IBGE) from 2002-2015.

Keywords: social inequalities, social stratification, social classes, Brazil

Combining archetypal analysis with latent profile analysis to assess financial knowledge

Lucio Palazzo[1], Maria Iannario[1], Francesco Palumbo[1]

[1]University of Naples Federico II, Italy; lucio.palazzo@unina.it

Knowledge is defined as a multi-facetted latent variable not directly measurable but through manifest variables, namely items. Then latent variable models are widely used for this aim. Finding homogeneous groups of units according to their knowledge levels turns out helpful to policymakers and to any other who has to take decisions into the domain. Therefore, latent variables models are combined within integrated approaches to find homogeneous groups. The aim of this work is to propose a coordinated strategy combining item response theory (IRT) models with archetypal analysis (AA). The proposed method is applied to a dataset of 625 Italian respondents to a European project entitled “Fintech and Artificial Intelligence in Finance”. Empirical evidence demonstrates that the proposed method is an effective and helpful tool to get respondents’ homogeneous profiles according to their responses to the questionnaire.

Keywords: archetypal analysis, multidimensional item response theory, graded response model, partial credit model

A comparative tool for analyzing electoral behavior and political culture: a multivariate methodology leveraging hierarchical cluster analysis and factorial correspondence analysis

Georgia Panagiotidou[1], Theodore Chadjipadelis[1]

[1]Aristotle University of Thessaloniki, Greece; gvpanag@polsci.auth.gr

This presentation explores the correlations between political culture and political and electoral behavior, using a multivariate methodological framework comprising hierarchical cluster analysis (HCA) and factorial correspondence analysis (AFC). Rooted in Benzécri's two-step approach, the study aims to detect and investigate behavioral discourses and typologies based on variables such as voting intention, participation in the elections, ideology, political interest, mobilization, and information, but also moral values, perception of democracy and trust in institutions.

Our unique approach leverages symbolic representation methodology, employing visual imageries representing concepts such as "democracy" and "moral values". The demonstration of our approach is substantiated by data analysis and results derived from a survey conducted amongst students from Aristotle University in Thessaloniki in spring 2023.

The HCA facilitates the first step of our analysis by deciphering variables for the symbolic representation of the democratic self and the moral self. This process results in clusters of images and respondents, providing crucial insights into the profile of each cluster. The second step involves AFC, which synchronously examines these cluster membership variables with other political behavioral variables.

The process culminates in a correspondence analysis, restructuring all categories along multiple dimensions, encapsulating the overall data inertia. This, in turn, aids in clustering together the coordinates of categories to form specific sets reflecting the existing discourses and their associated behavioral traits.

Finally, the presentation will depict these behavioral discourses within a two-dimensional system. Here, categories are positioned on axes according to their coefficients, enabling clear visualization and comprehension of discourses. Furthermore, the platform allows the assessment of inner discourse antagonisms, informed by their positioning within the Cartesian field, producing a visualized behavioral map. Thus, the proposed presentation offers a comprehensive comparative tool which enables the exploration of the various typologies and the dynamics of electoral, political and cultural behavior through a robust, multivariate methodological lens.

Keywords: electoral behavior, multivariate methodology, comparative analysis, hierarchical cluster analysis (HCA), factorial correspondence analysis (AFC)

Correspondence analysis of midwifery students' beliefs and attitudes towards statistics

Stylianos Papalexandris[1,2], Sofia Anastasiadou[1]

[1]University of Western Macedonia, Greece; [2]Democritus University of Thrace, Greece; sanastasiadou@uowm.gr

Statistics, biostatistics, medical statistics play a significant part in the health sciences because the interpretation, comprehension and decision-making based of real evidence related to diseases, severe or not, are extremely important as scientists struggle to improve health outcomes based on regarding clinical research, experimental trials surveys and observational studies. Medical research data are analyzed and interpreted in many fields of health science although they are disgracefully disliked among health science students. Statistical reasoning, based both on mathematical and probabilistic models, provokes fear and anxiety health sciences students and are strongly related to law academic achievement. Lately, the interest in health sciences student's attitudes toward statistics has increased considerably. Therefore, the current study evaluates midwifery students' beliefs and attitudes towards statistics by the help of correspondence analysis. A sample of 160 Greek midwifery students from the University of Western Macedonia in Greece participated in the study. The study used a survey of attitudes toward statistics (SATS) questionnaire of 36 items requiring answers based on a 7-point Likert scale. The SATS scale consists of six conceptual constructs called affect, cognitive competence, value, difficulty, interest and effort.

Reliability in terms of internal consistency was greater than the cutoff point of 0.70, both for the whole scale and its six conceptual constructs. Correspondence analysis verified the construct validity of the SATS scale. The results demonstrated negative attitudes for almost all SATS scale' conceptual constructs. Since the data were stored from only one University department there is a need for further research from the three midwifery departments in Greece, evaluating more widely beliefs and perceptions toward statistics.

Keywords: students, beliefs, attitudes, towards, statistics

MBPCA-OS: an exploratory multiblock method for mixed variables. Application to study the immune response to SARS-CoV-2 infection and vaccination.

Martin Paries[1,2], Evelyne Vigneau[1], Adeline Huneau[2], Olivier Lantz[3], Stéphanie Bougeard[2]

[1]Oniris, INRAE, StatSC, 44300 Nantes, France; [2]Anses, Epidemiology, Health and Welfare, Laboratory of Ploufragan-Plouzané-Niort, France; [3]Clinical Immunology Laboratory, Institut Curie, Paris, France; martin.paries@oniris-nantes.fr

The analysis of multiblock data, which involves studying numerous variables measured on the same observations and organized in blocks, is now standard in several domains, particularly in biology. To explore the relationships between these variables at both the block and variable levels, several unsupervised component-based multiblock methods were proposed in the literature. However, most of them were designed only for numeric variables, despite the fact that most real-life datasets include mixed variables (nominal, ordinal and numeric). In this context, we propose an unsupervised multiblock method, denoted multi-block principal component analysis with optimal scaling (MBPCA-OS), which accounts for the appropriate measurement level of each variable. The method uses an optimal scaling strategy to integrate mixed variables, and searches for the best quantifications possible of the variables that optimize the MBPCA criterion. The objectives are threefold: (i) deal with the multivariate structure of the data by seeking latent components, thus allowing dimension reduction (ii) seek for common information between blocks of variables, and (iii) handle variables at their appropriate measurement level (numeric, nominal and ordinal). MBPCA-OS shares these objectives with OVERALS or MFAmix, with differences in the way to deal with within-block multicolinearity or ordinal variables. MBPCA-OS is applied to multiblock data from the CURIE-O-SA French cohort collected on 1,917 participants between April 2020 and November 2021. In this study, variables are of different measurement levels and organized in four blocks. The objective is to study the immune responses (block 1) according to the SARS-CoV-2 infection and vaccination statuses (block 2), the symptoms (block 3) and the participant's characteristics (block 4). Thresholds effects of serological assays were highlighted while applying appropriate scaling. The differences in immunization between vaccinated and unvaccinated participants, and the higher protection provided by a joint infection and vaccination are shown.

Keywords: multiblock analysis; exploratory analysis; mixed data; optimal scaling; SARS-CoV-2

Of music, politics, visual perception and, of course, SVD

Ivailo Partchev[1]

[1]Cito, Netherlands, The; partchev@gmail.com

Matrix decompositions, especially SVD, are at the heart of many powerful techniques for data analysis. Some of the most popular among these, such as principal component analysis (PCA) or correspondence analysis (CA), focus on geometric representations of their findings. Skillicorn (2007) considers four interpretations of matrix decompositions: factor, geometric, component, and graph, arguing that they are adapted to search for specific hidden aspects of the data. To this we may add that persons vary considerable in their capacity for visual perception, mental rotation and other faculties helpful in geometry. We present a PCA of a small data set, comparing the biplot interpretation with a series of clustered image plots representing the original data, individual components, and sums thereof. While conclusions are essentially the same, we feel that the component representation will be appealing to many users.

Reference:

Skillicorn, D.B. (2007) Understanding Complex Datasets: Data Mining with Matrix Decompositions. Chapman & Hall / CRC Press.

Keyword: SVD

The advent of neoliberal cultural policy? The curious case of the Zurich Opera

Valentina Petrović[1], Jörg Rössel[1]

[1]UZH, Switzerland; petrovic@soziologie.uzh.ch

In September 1994, the electorate of the Canton of Zurich had the opportunity to vote on the future financing of the Zurich Opera House. A vast majority of 73% accepted the way to finance the Opera House primarily by cantonal public funds. How does this high level of support come about for an art form that is often seen as particularly steeped in tradition and, above all, as requiring particularly substantial public subsidies (Reuband 2009)? To answer this question, we conducted a content analysis based on more than 400 newspaper articles that appeared in the period before the vote in September 1994. We used a combination of a theory-guided approach and exploratory content analysis for the coding (Gerhards/Lindgens 1995).

Not only did our analysis confirm the theory guided arguments referring to economic advantages or cultural aspects of classical high culture, but arguments were discovered in an explorative way based on meritocratic criteria, in line with the new public management school. Beyond the analysis of the arguments, socioeconomic characteristics of the population in the circulation areas of the newspapers were also collected. Finally, we analysed the data using multiple correspondence analysis, which allows us to look at the arguments' spatial configuration against the background of socioeconomic characteristics.

Keywords: sociology of culture, new public management school, opera house, correspondence analysis, content analysis

Correspondence analysis of power-law distributed data via ranks

Koen Plevoets[1]

[1]Ghent University, Belgium; koen.plevoets@ugent.be

Although one of the first applications of correspondence analysis was on language data (see the historical overviews in e.g. Bécue-Bertaut 2014; Murtagh 2005: 8-9), linguistic frequencies are a challenging data type for the technique. Frequency tables of language cooccurrences usually exhibit both a high degree of sparsity (i.e. zero frequency cells) and heavy skewness, with linguistic frequencies typically having a Zipf's law distribution (Zipf 1936; 1949). Such data involve various low-frequency items and, as Greenacre (2013; 2017: 297-298) has pointed out, these often end up as outliers in a biplot, while the bulk of the other items appears cluttered around the origin. This paper proposes to tackle this issue by transforming the cooccurrence frequencies to their ranks and applying ordinary correspondence analysis to the table of these ranks. The use of ranking is partly motivated by Zipf's law, which relates linguistic frequencies to their ranks, but also by its wide-spread application in nonparametric statistics to focus on order differences in data regardless of the underlying distribution (Corder & Foreman 2014: 6-8). Because the rank transformation is a monotonic step function, it naturally extends to analyzing the associations between rows and columns in a frequency table. The correspondence analysis of ranks is illustrated on the basis of various linguistic datasets, available in the R packages *svs* (Plevoets 2015) and *languageR* (Baayen & Shafaei-Bajestan 2019). The results show elegant and interpretable configurations in the biplot. These are compared with the outputs of untransformed correspondence analysis, several power-transformed analyses (Greenacre 2009) as well as multidimensional unfolding, which is closely related to the technique proposed here. These comparisons, which are done by means of Procrustes analysis in the *smacof* package (Mair, Groenen & de Leeuw 2022), reveal high degrees of congruence among the various alternatives.

Keywords: language data, power laws, outliers, ranking, unfolding

Meritocracy and selectivity in the Danish field of higher education; how the admission system effectively perpetuates horizontal stratification in a seemingly egalitarian system

Simone Mejding Poulsen[1]

[1]Copenhagen University, Denmark; smp@hum.ku.dk

The Danish system of higher education (HE) is often perceived as meritocratic, egalitarian, and with no clear-cut elite. Yet, researchers argue that unified admission systems, like the Danish, inadvertently construct a hierarchy among the HE programs by determining selectivity based on the applicants; the more the applicants, the higher the selectivity. The exclusivity of high selectivity leads to perceived prestige and desirability thereby creating an elite group of study programs within the same level of HE.

Utilizing correspondence analysis, prior Nordic research has examined stratification within seemingly egalitarian HE systems. By focusing on the socioeconomic background and gender of the students across different study programs, they demonstrate how these systems still contain elite programs and institutions. Thus far, however, little attention has been devoted to how this horizontal stratification is a result of the internal mechanisms of unified admission systems.

This paper sets out to examine the relationship between horizontal stratification and selectivity perpetuated by the Danish admission system, considering gender, academic attainment, and socioeconomic background. This is done via simple correspondence analysis using microdata from Statistics Denmark on all applications (N = 207 398) for Danish university BA programs from 2021. By constructing the space of university bachelor programs on the admitted applications and plotting the entire scope of applications as a supplementary variable, the analysis shows how the admission system perpetuates horizontal stratification through application-based selectivity. Applicants of all backgrounds seek admission into the same handful of highly selective study programs, yet only the meritocratic elite gets admitted.

Keywords: higher education, selectivity, stratification, admission systems, meritocracy

Spatial patterns and trait dependency in fish diet

Raul Primicerio[1]

[1]UiT, The Arctic University of Norway, Norway; raul.primicerio@uit.no

Dietary variation within and among species is influenced by spatial distribution, seasonality and biological characteristics of predators and prey. An extensive dataset from the Barents Sea, consisting of over 25 000 fish diets from 70 species, allows to address the three components of dietary variation jointly. Double constrained correspondence analysis is used to address the effects of season, space and predators' traits on dietary contributions by different prey taxa, while accounting for prey traits. Seasonal variation in diet is most pronounced among fish species eating plankton, and less prominent among fish eating other fish or benthos. Ample spatial variation in diet is found among most investigated fish species, likely influenced by prey availability and cooccurring fish species. Predator and prey traits are correlated indicating niche specialization.

Keywords: double constrained correspondence analysis, fish, diet, traits

A comparison of latent semantic analysis and correspondence analysis of document-term matrices

Qianqian Qi[1], David J. Hessen[1], Tejaswini Deoskar[1], Peter van der Heijden[1,2]

[1]Utrecht University, Netherlands; [2]University of Southampton, UK; q.qi@uu.nl

Both latent semantic analysis (LSA) and correspondence analysis (CA) can use a singular value decomposition (SVD) for dimensionality reduction. Here, we present a theoretical analysis and comparison of the two techniques in the context of document-term matrices. We show that CA has some attractive properties as compared to LSA, for instance that effects of margins, i.e. sums of row elements and column elements, arising from differing document-lengths and term-frequencies are effectively eliminated, so that the CA solution is optimally suited to focus on relationships among documents and terms. We empirically compare CA to various LSA based methods on text categorization in English and authorship attribution on historical Dutch texts, and find that CA performs significantly better. We also apply CA to a long-standing question regarding the authorship of the Dutch national anthem Wilhelmus and provide further support that it can be attributed to the author Datheen, amongst several contenders.

Keywords: text data mining, text classification, authorship attribution, information retrieval, statistical methods, singular value decomposition

From geometric analysis to analytical geometry: making sense of [(habitus)(capital)] + field = practice

Jan Rasmus Riebling[1]

[1]Bergische Universität Wuppertal, Germany; riebling@uni-wuppertal.de

In his extensive and groundbreaking work on 'La Distinction', Bourdieu gives an equation to summarize the process underlying observable social phenomena: "[(habitus)(capital)] + field = practice". It is not immediately clear what this tries to convey, since very little context is provided directly and it seems to be somewhat at odds with standard notation. It has even been suggested that this specific formula can be considered an inside joke regarding rational choice theory and its ubiquitous utility function. I suggest a more charitable view: when we attempt to make sense of this formula mathematically, we will find, that it could serve as an essential bridge between sociological theory, computational methods and mathematical analysis.

I propose to view the aforementioned formula as a Hamiltonian function of an Ising model, which has been extensively used in the study of social dynamics. More simply put, "[(habitus)(capital)]" can be seen as modeling a local interaction depending on the properties of a single element (i.e., habitus) as well as the properties of interacting elements (i.e., capital). This happens within a field propagated through the interacting elements. Or, as Bourdieu would put it: "the set of agents who are placed in homogeneous conditions of existence imposing homogeneous conditionings and producing homogeneous systems of dispositions capable of generating similar practices"

This interpretation connects Bourdieu's theory to statistical mechanics and mean field theory which have both been strongly influenced by the Ising model. Thereby allowing for a new analytical way to reason about Bourdieu's model and maybe even to derive testable predictions from it. It also has important implications for longstanding arguments in social theory. Most prominently, it can be shown that it would exhibit the property of "macroscopic uniformity". Thereby exhibiting stable behavior on the macro-level while being irreducible to micro-level interactions.

Keywords: field theory, Ising model, social dynamics, statistical mechanics, mathematical modeling

Social capital and the intergenerational transmission of cultural capital: how parents' social networks influence children's accumulation of cultural capital.

Andreas Roaldsnes[1]

[1]University of Bergen, Norway; andreas.roaldsnes@uib.no

The aim of this study is to further our understanding of how parents' social resources influence the process of cultural transmission to their children. Research on this process focuses mainly on within-family mechanisms, and the role of social networks has received comparatively little scrutiny. Secondly, the study aims to contribute to the theoretical understanding of Bourdieu's notions of social capital and how they may be applied to empirical research. The link between cultural capital acquired in childhood, educational attainment and higher social positions is well established, but we do not fully understand how this cultural capital is transmitted from parents to children. Research on the intergenerational transmission of cultural capital is often conducted on within-family dynamics, paying less attention to external influences. This article studies the role that parents' social networks play in the process of transmitting cultural resources to children, analyzing data on parents' social ties, and parents' and children's culture, leisure and athletic practices, using geometric data analysis and traditional statistical techniques on families in Bergen, Norway (N=4754). I find that cultural consumption is strongly linked to Bourdieu's notion of social homophily. Parents with ties to higher status occupations are associated with their children's higher participation and exposure to legitimate culture, also when other influences are controlled for. Parents who have social ties to working class occupations correlate to having children who are far less exposed to cultural practices. Crucially, I find evidence of Bourdieu's multiplier effect, that social resources multiply returns from other resources. Parents' investments in culture for their children is dependent on parental cultural capital and multiplied by parents' social capital. At the same time, there is evidence of a divider-effect, where having cultural orientations is hampered by having working class ties, suggesting social closure and class membership through cultural participation.

Keywords: social capital, cultural transmission, cultural capital, concerted cultivation, social networks

Logistic multidimensional data analysis with an emphasis on ordinal variables

Mark de Rooij[1]

[1]Leiden University, Netherlands, The; rooijm@fsw.leidenuniv.nl

In this presentation, a general framework for logistic multidimensional data analysis is laid-out for supervised and unsupervised analysis of categorical response variables. Dominance and proximity variables are distinguished, where dominance variables are analyzed using inner product models and proximity variables are analyzed using distance models. Biplot representations of both type of models will be discussed. We emphasize ordinal response variables, for which a continuous underlying latent variable is assumed, leading to cumulative logit models. An expectation-majorization-minimization algorithm is derived for maximum likelihood estimation of the parameters of the models. The methodology will be illustrated with data from the International Social Survey Programme.

Keywords: categorical data, PCA, multidimensional unfolding, visualization

The institutional logics of love: using geometric data analysis to explore the social structuring of intimate practices among American college students

Henk Roose[1], Roger Friedland[2]

[1]Ghent University, Belgium; [2]University of California, Santa Barbara, United States; henk.roose@ugent.be

Contemporary fields of sexually intimate life are replete with heterogeneous understandings of what kinds of intimacy one can or ought to pursue and through what kinds of practices one “loves”. In this paper we explore how geometric data analysis (GDA) can be deployed to identify institutional logics, here of sexual intimacy, and measure these patterns of meaningful practice. We understand institutional logics to be constellations of meaningful practices which presume and produce particular substantive goods: institutional logics of love refer to divergent modes of doing and being, thinking and wanting, regularities that may produce romantic love, sexual pleasure, friendship, marriage, or piety and/or sacrality. So, how one loves, whether one loves, and whom one loves is conditioned by one’s engagement in these logics. We analyze whether and how students “make love” and whether these intimate practices are associated with gender, social class and various indicators of sexual capital.

Using an internet survey representative of the student population at a Pacific University (n = 1,315) we find three dimensions to be structuring the field of intimacy. “Making love” is done in a variety of ways: loving sex is literally a different kind of sex from hook-up sex and sex as understood as part of the sacred and moral obligations of marriage, not only in terms of what it means , but what it does, from talking intimately to having an orgasm.

Keywords: field, social space, intimate life, gender, sexual capital

Dimensions of cultural openness in the wine field

Jörg Rössel[1], Patrick Schenk[2]

[1]University of Zurich, Switzerland; [2]University of Lucerne, Switzerland; roessel@soziologie.uzh.ch, patrick.schenk@unilu.ch

Our paper puts wine consumption in its social context, and conceptualizes it as a symbolic practice. We follow Bourdieu’s insight that such symbolic behaviors, such as the music someone listens to or the choice of dress, are intimately connected with a person’s socio-economic position. Based on Peterson and Kern, we assume that such practices are structured along a dimension differentiating between a low and a high level of openness towards a broad range of practices. Going one step further, it needs to be emphasized that the notion of openness can be realized in several ways: Openness can refer to (1) the range of products of a similar kind a person finds enjoyable (omnivorousness), (2) the range of products of a similar kind a person actually consumes (eclecticism), (3) the frequency of consumption of different products of a similar kind (voraciousness), (4) the consumption of foreign vs. domestic products of a similar kind (cosmopolitanism) and (5) the ways in which a product is consumed. Hence, our paper seeks to examine which of these five dimensions of cultural openness can be identified in the field of wine consumption, how they are related among each other, and how they are related to economic, social, and cultural resources.

To analyze these dimensions of cultural openness, we use data from a standardized mail survey conducted in four German cities (Mainz, Wiesbaden, Hamburg, Cologne). This allows us to investigate the relationship between different aspects of wine consumption, especially its more symbolic and more material aspects, other forms of cultural consumption (such as leisure activities and musical tastes) as well as socio-demographic characteristics. By means of multiple correspondence analysis, we can visualize the interrelations among these variables in a two-dimensional space. This helps us to identify the primary dimensions of cultural openness and their relations to socio-economic position.

Keywords: wine, openness, omnivorism, class

The data sciences as a space of opportunities. Investigating the double construction of a key field in the digital economy

Philippe Saner[1]

[1]University of Lucerne, Switzerland; philippe.saner@unilu.ch

This paper investigates the recent emergence of the data sciences as a new interdisciplinary and transversal field of knowledge that connects and transcends multiple social sectors of the digital society. Previous research has analyzed and interpreted the data sciences as a new profession (Avnoon 2021; Brandt 2016; Dorschel 2021; Dorschel and Brandt 2021), a new fraction of the middle class (Dorschel 2022), or a new (inter-)discipline (Murtagh and Devlin 2018). Here, we take a different approach and analyze it from a field theory perspective (Eyal 2013; Saner 2022), drawing on Pierre Bourdieu's theoretical framework. Our aim is to adapt and apply Bourdieu's theoretical concepts, in particular his work on economic sociology and on the French housing market (Bourdieu 2005), to the analysis of the data sciences as a key driver of the digital economy. We argue that the recent emergence of the data sciences builds on a 'double construction' of the new field, on both the demand and the supply side, supported and reinforced by a variety of actors in the tech industry, academia, research and higher education policy, and the media.

On the 'demand side', we investigate job advertisements as constructions of employers' demands for data science expertise and skills, combining a cultural sociology approach to topic modelling (DiMaggio, Nag, and Blei 2013; Mützel 2015) with multiple correspondence analysis (Le Roux and Rouanet 2010). On the 'supply side', we study degree programs in data sciences, and more broadly, in all study fields related to digital economy topics, as constructions of data science knowledge in higher education, using multiple correspondence analysis. We base our analyses on unique datasets on the labor market and the higher education sector in Switzerland, which we collected in two empirical projects.

Keywords: digital economy, data sciences, labor market, higher education, topic modelling

Experts in the field of European politics 1966-2018: changes in European expert groups analyzed using multiple correspondence analysis and hierarchical agglomerative clustering

Christian Schmidt-Wellenburg[1]

[1]Universität Potsdam, Germany; cschmidtw@uni-potsdam.de

European expertise is an effect of the field of European politics, opens up access to sources of knowledge and authority beyond the political context and varies depending on the state of European integration. This can be exemplified by a prosopography of members of political economic European expert groups (1966-2018). A multiple correspondence analysis of 261 professional careers allows to reconstruct the structure of European expertise in this policy area, to determine six types of experts by hierarchical agglomerative clustering, and in the light of the observed diversity to trace the decline of scientific and the increase of financial expertise. This development can be interpreted as a strengthening of the economic-heteronomous pole in the field, which goes hand in hand with the rise of neoliberal forms of government, the autonomisation of the field, and the professionalisation of expert consultations.

Keywords: MCA, EU, HAC, political field, experts

Modelling Reddit as social space - a mixed methods iterative procedure

Andreas Schmitz[1], Mattia Samory[2]

[1]Bonn University, Germany; [2]Sapienza University of Rome; andreas.schmitz@uni-bonn.de, mattia.samory@uniroma1.it

We apply Bourdieu's logic of iterative object construction to digital process data from the platform Reddit. In order to construct a social space and a dimensional representation, we make use of a procedure that combines deductive, inductive, and abductive steps. We present the theoretical conception of the object, discuss qualitative steps, the formalization by natural language processing, and social computing (with recourse to textual data analysis, network analysis, and variants of singular value decomposition), and the mutual interrelation of these steps. Thanks to this relational methodology, a social space of Reddit is constructed. Its main structural features are presented including general structural traits as well as features specific to this particular field under study.

We conclude with an outlook on possible applications in future research and on the implications of a consistent cooperation between sociology and computer science.

Keywords: Reddit, social space, textual data, process data, qualitative methods, singular value decomposition

Linear regression estimation and model selection under data aggregation

Pieter Cornelis Schoonees[1], Niel J. le Roux[2]

[1]Erasmus University Rotterdam, Econometric Institute, The Netherlands; [2]Stellenbosch University, MuViSU (Centre for Multi-Dimensional Data Visualisation), Department of Statistics and Actuarial Science, South Africa; schoonees@ese.eur.nl

Aggregating over individuals belonging to different groups before performing linear regression is known to induce a so-called aggregation bias in the ordinary least-squares (OLS) coefficient estimates compared to those obtained without aggregation. There are however situations where this problem is unavoidable, such as when data from different views are merged. For example, neuroscientific studies may combine a sample of individuals who supply EEG data on a set of stimuli (i.e., the groups) with data which reports the aggregated behaviour of a second sample of individuals.

In this work, we study the effect linear aggregations have on model estimates and model selection procedures. Relying on matrixvariate normal distributional assumptions, we report an expression for the bias of the maximum likelihood estimator of the covariance matrix when the aggregated data is used instead of the original data. Moreover, it is noted that theoretically the non-intercept OLS estimates derived from the aggregated data are statistically independent of those derived from the original data. This is illustrated with a simple simulation experiment. The adverse effects of aggregation on common model selection procedures are investigated in a second simulation study. Situations are investigated where not all explanatory variables may contain information about the outcome variable. The findings and implications are briefly discussed.

Keywords: aggregation bias, matrixvariate normal, ordinary least squares, model selection

Person-centered data-analysis with covariates and the R-package confreq

Mark Stemmler[1], Jörg-Henrik Heine[2]

[1]University of Erlangen-Nuremberg, Germany; 2Technical University Munich;
mark.stemmler@fau.de, jhheine@googlemail.com

Configural Frequency Analysis (CFA) is a useful statistical method for the analysis and modelling of multivariate categorical data perceived as multiway contingency tables. CFA is an appropriate tool for person-centered research (Stemmler, 2020). In complex contingency tables, patterns or configurations are analyzed while comparing observed cell frequencies with expected frequencies. Significant differences between observed and expected frequencies lead to the emergence of types and antitypes. Types are patterns or configurations that are observed significantly more often than expected; *antitypes* are configurations that are observed less often than expected. Here, CFA is presented based on the log-linear modeling approach (Stemmler & Heine, 2017) which allows flexible formulations of different CFA base models. CFA may also be used together with interval-level variables, which can be added as covariates into the design matrix to bring the estimated cell frequencies closer to the observed cell frequencies. In this presentation, real data examples are analyzed with the R package *confreq* (Heine, et al. 2022).

References:

Heine, J.-H. Alexandrowicz, R. W., & Stemmler, M. (2022). *confreq: Configural Frequencies Analysis using log-linear modeling (R-package).* Retrieved from https://CRAN.R-project.org/package=confreq (R-package version 1.6.1-1).

Stemmler, M. (2020). *Person-centered methods: Configural Frequency Analysis (CFA) and Other Methods for the Analysis of Contingency Tables (2nd edition)*. Series: Springer Briefs in Statistics. New York: Springer Publishing Company.

Stemmler, M. & Heine, J.-H. (2017). Using Configural Frequency Analysis as a Person-Centered Analytic Approach with Categorical Data. *International Journal of Behavioural Development*. DOI: https://doi.org/10.1177%2F0165025416647524.

Keywords: analysis and modelling of categorical data, R package, configural frequency analysis, log-linear modeling, multiway contingency tables

Stratifying lifestyle and social class in urban China

Langyi Tian[1], Aurélien Boucher[2]

[1]Columbia University, United States of America; [2]The Chinese University of Hong Kong (Shenzhen), China; lt2710@columbia.edu

This study adopts Bourdieu's relational analytical strategy to investigate the relation between lifestyle and social stratification in the Chinese urban population. Multiple correspondence analysis and clustering illuminate a typology of four lifestyles: (1) disengagement from consumer culture, (2) self-restrained consumption of highbrow products, (3) preference for intellectualized and ephemeral leisure activities, and (4) economically and culturally well-established consumer-buyers. A multinomial logistic regression validates the association of lifestyle with cultural and economic capital. Capital structure seems less influential in China than in Western European societies, while generation emerges as a unique differentiator under rapid social transformation.

On the choice of weights in aggregate compositional data analysis

Jules de Tibeiro[1], Vartan Choulakian[1], Pasquale Sarnacchiro[2]

[1]University of Moncton, Canada; [2] University of Naples, Italy; jules.de.tibeiro@umoncton.ca

In this talk, we distinguish between two kinds of compositional data sets: elementary and aggregate. This fact will help us to decide the choice of the weights to use in log-interaction analysis of aggregate compositional vectors. We show that in the aggregate case, the underlying given data form a paired data sets composed of responses and qualitative covariates; this fact helps us to propose two approaches for analysis-visualization of data named log-interaction of aggregates and aggregate of log-interactions and their first-order approximations. Furthermore, we also use the QSR index within taxicab log-ratio analysis for the choice of the best map for visualisation of a data set.

The field of private equity in The Netherlands

Rob Timans[1]

[1]Nivel, Netherlands, The; robtimans@gmail.com

Within the field of finance, private equity (PE) firms have become increasingly important players. Exemplars of what has been called the 'second wave of financialization', PE firms use debt to finance takeovers of private companies. The aim typically is to increase returns of these companies and to sell them after a holding period of some five to seven years. The take-over of a firm by a PE investor can have consequences for employees and other stakeholders, as well as for other firms in the industry. Moreover, PE firms have become influential economic and political players.

In the (economics) literature on the subjects, PE firms are mostly treated as uniform actors. Despite their growing economic, political and symbolic importance, little is known about their salient characteristics or the differences between PE firms. What are their characteristics and how do their strategies differ? On what type of knowledge networks do they rely? How do these features affect their position vis-à-vis each other? In this contribution, we present an analysis of the field of PE in the Netherlands to answer these questions. Using a database containing information on the largest PE firms in the Netherlands, we employ geometric data analysis (GDA) and social network analysis (SNA) to examine the structure of the field and the position-taking of PE firms.

Keywords: fields, finance, geometric data analysis, economic sociology

Unravelling structural properties of cultural production and consumption: a multiple correspondence approach

Željka Tonković[1]

[1]University of Zadar, Croatia; zeljka.tonkovic@unizd.hr

Multiple correspondence analysis (MCA) has been widely used in the research of cultural production and consumption. As a relational technique, whose primary aim is not to estimate the effect of some explanatory variables on the dependent variable, it has been considered close to sociological thinking of Pierre Bourdieu. In order to explore patterns of cultural practices and taste, MCA has been recognised as a tool in various contexts and fields, including publishing and literary field, music and TV production. This presentation will engage with the recent empirical contributions, focusing on data collected in several research projects. The application of MCA in the field of cultural production

is based on ongoing research of the field of literary translation in Croatia, while the study of cultural consumption is based on survey data from the project "Social stratification in Croatia: structural and subjective aspects", co-funded by the Croatian Science Foundation. The data was collected from 1000 respondents living in Croatia and above the age of 18 in 2017. The data was analyzed using MCA and hierarchical cluster analysis. Application of cluster analysis was particularly useful in order to establish boundaries in the field of cultural tastes. The findings are reflected in the light of ongoing theoretical conversations about the interrelationship between social structure and cultural hierarchies.

Keywords: cultural production, cultural consumption, tastes, social class, field analysis

Understanding patients' preference: a study on the diversification of medical practices in the contemporary healthcare system

Miriam Trübner[1] Judith Lehmann[2,] Rasmus Hoffmann[3], Alexander Patzina[4]

[1]University of Bamberg, Germany; [2]University of Bamberg, Germany; [3]University of Bamberg, Germany; [4]University of Bamberg, Germany; miriam.truebner@uni-bamberg.de

Contemporary societies are undergoing a marked process of individualization leading to differentiation of numerous areas of life, such as consumer behavior and lifestyle choices. This is particularly evident in the medical field, where a broad array of evolving therapeutic practices, ranging from conventional medical approaches to more unconventional medical approaches (e.g., acupuncture or TCM), illustrates the diversity within the healthcare system. Against this backdrop, we ask the following research questions: How are preferences towards diverse medical practices distributed in the German population? How are these different preferences related to each other? And what background characteristics are decisive for having specific medical preferences?

To answer these questions, we apply latent class analysis (LCA), a technique that enables us to exploratively reveal the most common preference patterns regarding diverse combinations and levels of preference towards naturopathy, conventional medicine, complementary medicine, integrative medicine, and alternative medicine. For this, we perform analyses based on a sample retrieved by a nationwide study on the acceptance and use of naturopathy in Germany (respondents aged 18-75 years; N~4000). As a result, we can identify five distinct preference patterns which we can further relate to a wide range of socio-structural traits. Preliminary results establish that low socioeconomic background is associated with a tendency to be generally indifferent towards medical practices, while higher socioeconomic status is associated with a stronger preference for specific practices. Moreover, older people and women show a greater receptivity towards less conventional medical practices such as a naturopathy and non-evidence-based methods.

Considering that some more customized and specialized medical practices and treatments have a positive effect on health and the prevention of diseases, this social gradient in the use of unconventional medical practices may contribute to further increases in health inequality.

Keywords: latent class analyses, health, social inequality, social stratification

Decomposing the market share contributions of price and units sold. A compositional visualization

Ana de Vega-Ventura[1], Berta Ferrer-Rosell[1], Eva Martín-Fuentes[1], Germà Coenders[2]

[1]University of Lleida, Spain; [2]University of Girona, Spain; germa.coenders@udg.edu

Market share data are compositional by nature, expressed as percentages adding up to 100. They can be effectively visualized by means of a compositional biplot, corresponding to a principal component analysis after a transformation by means of centred logratios. Market share data can be computed using different criteria, leading to compositions by percentage of units sold or by revenue. The ratio between two compositions is an operation called the perturbation difference and leads to another composition. In the market-share case, the ratio between the compositions by revenue and by units sold is the composition indicating the contribution of price to market share, which can be visualized together with the original market-share data. Revenue market share is thus decomposed into two compositions indicating the contributions of volume and price.

We present an illustration from the hospitality industry with market share by distribution channels (hotel direct sales, call centre, hotel chain's web, traditional travel agencies, internet intermediaries). Data come from a worldwide hotel chain company with more than 650 hotels, excluding the US, with yearly data from 2013 until 2020. When using data by room-nights booked, the market share of internet intermediaries has increased over the period considered, at the expense of all other distribution channels. However, when using revenue data, hotel direct sales have had the most favourable evolution at the expense of the remaining channels, which is explained by the contribution of the increasing relative price of direct sales being more relevant than that of the diminishing volume sold.

Keywords: market share, perturbation difference, biplot, compositional data, hospitality distribution channels

A factorial method for multiple distributional data based on LDQ transformation

Rosanna Verde[1], Antonio Balzanella[1]

[1]University of Campania "Luigi Vanvitelli", Italy; rosanna.verde@unicampania.it

This work deals with a factorial analysis method for multiple distributional-valued data ([1], [2]). Let E be a set of objects observed on p distributional variables. Each object is represented by p probability density functions (PDFs), or empirical ones. In consideration of the most recent developments in distributional data analysis (DDA), we introduce a transformation of the quantile functions associated with the PDFs, named logarithm derivative quantile (LDQ) functions. A similar transformation was introduced by (3) to map density probability functions in a Hilbert space. Our proposal is based on functional principal component analysis of density functions, introduced by the same authors, and on the definition of variability measures and averages function according to Frèchet. We consider an extension of the factorial approach to multivariate distributional data, providing a multiple factorial analysis on LDQ-transformed data (MFA-LDQ). Some interpretative aids (e.g., contributions to the new axes) are then introduced. Some first results on real data have shown the effectiveness of the proposed method.

References:

[1] Bock H., Diday E.: *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer (1999).
[2] Brito P., Dias S.: *Analysis of Distributional Data*. Chapman Hall (2022).
[3] Petersen A., Müller H.: *Functional data analysis for density functions by transformation to a Hilbert space*. The Annals of Statistics, Ann. Statist. 44(1), 183-218, (2016).

Keywords: symbolic data analysis, distributional data, quantile density functions

Partial least squares for mixed continuous and binary variables and its associated biplot

Laura Vicente González[1], José Luis Vicente Villardón[1]

[1]Universidad de Salamanca, Spain; laura20vg@usal.es

In this study, we present a novel approach that extends partial least squares regression to handle continuous and binary variables simultaneously. The method can use combination of binary and continuous variables in both sets, predictors and responses. We visualize the results using biplots. To develop the final algorithm, we adapt the NIPALS method to accommodate binary and continuous variables using a gradient descent procedure.

To demonstrate the effectiveness of our method, we apply it to classify different strains of Colletotrichum graminicola using RNA-seq data. We investigate the variations among nine strains originating from different countries and identify the genes that contribute to their characterization. The calculations are performed using the statistical software *R*, and we have incorporated new functions into the *MultBiplotR* package.

Keywords: PLS, binary data, NIPALS, biplot

Biplots of mixed continuous and categorical data

Jose Luis Vicente-Villardon[1], Laura Vicente-González[1]

[1]Universidad de Salamanca, Spain; villardon@usal.es

The term used to describe the combined display of individuals and variables from a data matrix is known as a biplot.
When the variables are binary, nominal, or ordinal, a traditional linear biplot representation is not suitable. More recently, biplots for categorical data have been introduced, which are based on logistic response models. These models are applicable to binary, nominal, or ordinal data. The coordinates of individuals and variables are calculated to exhibit logistic responses along the biplot dimensions. These methods are akin to logistic regression, just as classical biplots are associated with linear regression. Consequently, they are referred to as logistic biplots. Similar to the connection between linear biplots and principal component analysis, logistic biplots are associated with latent trait analysis, factor analysis for categorical variables or item response theory. The geometry of such biplots, as well as its relationship with the factorization of various correlation matrices, is examined.

For continuous, binary, or ordinal data, the representation of variables results in straight lines. In the case of nominal data, the representation of variables on the biplot is not a straight line, but rather a "prediction region." Algorithms employing gradient descent methods are also provided for constructing the graphical representations, prepared to mix different types of data on the same representation. The practicality and interpretation of logistic biplots are demonstrated through several real-world data applications.

Keywords: biplot, mixed types of data, gradient descent

Latent categorical variables of a set of observed categorical variables.

Maurizio Vichi[1], Michael Greenacre[2]

[1]Sapienza Univrsity of Rome, Italy; [2]Universitat Pompeu Fabra, Bercelona, Spain; maurizio.vichi@uniroma1.it

The combination of categories in a multivariate observed categorical data set is investigated, with the fundamental objective of identifying a reduced set of latent categorical variables having a minimal loss of dependence on the observed data structure. The basic premise is that the observed associations between manifest variables are explained or reconstructed by a reduced set of latent variables. The overall association (inertia) of the categorical variables is measured by the average of the inertias of all unique pairwise cross-tabulations of the variables (Greenacre 2017, chapter 19). The association between categories is computed starting from the square of the standardized Burt matrix, which includes the standardized contingency tables between pairs of observed variables. A hierarchical clustering algorithm that minimizes loss of inertia by aggregating categories of different variables is used to identify the order of the combined categories that define latent categorical variables. A test is used to find a partition of categories into subgroups with the slightest decrease in inertia. Each subgroup identifies a latent categorical variable. The disjoint confirmatory factor analysis model specifies latent categorical variables along with their quantifications. In a companion paper, Greenacre & Vichi consider the aggregation of categories of the same variables, in order to reduce the number of response categories with minimum loss of inertia.

References:

Greenacre, M. (2017). *Correspondence Analysis in Practice*. Third Edition. Chapman & Hall / CRC Press. Boca Raton, Florida.

Keywords: latent categorical variables, chi-square, contingency table, inertia, multiple correspondence analysis

Integrating topic modeling and block models into correspondence analysis: the example of the sociological field in Germany

Jonas Volle[1]

[1]Otto-von-Guericke-Universität Magdeburg, Germany; jonas.volle@ovgu.de

Bourdieusian theory and methodology have a long history of application within the field of science (Baier and Schmitz 2019; Bourdieu 1988; Rossier and Benz 2021; Schmidt-Wellenburg 2018; Schmidt-Wellenburg and Schmitz 2022; Schmitz et al. 2019; Schwemmer and Wieczorek 2020; Warczok and Beyer 2021). Often, prosopographic data is collected and analyzed using correspondence analysis to identify social power structures.

In the sociology of science, new data sources, such as bibliometric data from databases Scopus or WoS, for example, have been used for some time. These data contain network information such as collaborative relationships, citation relationships, or shared citations of references. In addition, these databases also provide textual information in the form of abstracts. These data and methods are already used with methods such as block modeling and topic modeling (Choe and Lee 2017; Cugmas, Ferligoj, and Kronegger 2019; Daenekindt and Huisman 2020; Peixoto 2019; Schwemmer and Wieczorek 2020). For example, topic modeling can identify thematic groups, while block modeling allows the construction of author clusters that share co-authors or references.

This contribution aims to integrate these innovative data sources and procedures into correspondence analysis employing Bourdieusian relational methodology. This is illustrated by the example of the field of sociology in Germany. I use data from an ongoing research project funded by the German research foundation: individual data from 1200 sociologists in Germany (Schmitz et al. 2019) are used and combined with bibliometric data from Scopus (Volle et al. 2023). The information obtained by topic and block modeling is then projected into space relationally to the individual data.

Keywords: field of sociology, correspondence analysis, topic model, block model

Meanings of working class on Finnish social media

Mikael Kasimir Wallin[1]

[1]Tampere University, Finland; mikael.wallin@tuni.fi

Recent cultural sociology research has argued that there are various forms of everyday culture that contribute to the wellbeing of people's lives but are not traditionally recognized as culture neither in research nor society at large, due to what is coined as middle-class bias. This also echoes the notion in social stratification research that working-class culture is often depicted as vulgar which is used to justify cultural hierarchies. This study will use a mixed methods approach, combining latent Dirichlet allocation (LDA) topic modeling and membership category analysis, to examine online discussions of the terms "working class" and "middle class" in Finnish online communities. The dataset will include tens of thousands of text comments from the online forum Suomi24 and the microblogging site Twitter from 2001 to 2023. The first research question will explore the dominant notions with which the working class is represented in Finnish social media. The second research question will focus on whether the working class is discussed online as culturally non-participant and lacking taste. This study contributes to the previous discourse on the use of computational methods in sociological research by outlining a novel mixed methods approach that combines LDA topic modeling with membership category analysis.

Keywords: cultural capital, membership category analysis, social class, social media, topic modeling

Academia, market and symbolic exchanges: the specificity of the legal field in Poland

Tomasz Warczok[1], Hanna Debska[2]

[1]University of Warsaw; [2]Pedagogical University of Cracow; tomaszwarczok@gmail.com, hannadebska.up@gmail.com

In Bourdieu's classic approach, the legal field is clearly divided into a pole occupied by "theorists" (legal scholars) and the pole of "practitioners" (where legal advisers and judges are localized). This division of labor is homologous to the key divide of the field of power (dominant class) into spiritual power and temporal power. In Poland, as our research shows, such opposition is less distinct, which is due to the fact that legal scholars perform various forms of professional activity outside the academic field: most of them practice full-time as legal advisers, advocates, or, less frequently, judges.

This leads to a very different construction of the economics of symbolic exchanges in the Polish legal academic field than in Western countries, as well as the specific forms and content of juridical capital and the related hierarchy between academia and the market.

We reveal the hidden dimensions of these structures and forms by using a geometric analysis of prosopographical data of all individuals located in the Polish space of legal philosophy (the most autonomous part of the academic legal subfield). More specifically, we carry out a multiple correspondence analysis along with Euclidian clustering and, particularly to identify the links between academia and the legal market, structured data analysis.

Keywords: field, MCA, GDA, space, capital

Asian whizkids, black basketballers, and white male professors? The visual embedding of different (minority-) groups in YouTube videos of US-universities

Oliver Wieczorek[1]

[1]University of Kassel, Germany; oliver.wieczorek@incher.uni-kassel.de

Around the world, the competition for the best and brightest students is fierce and is getting progressively more intense over time. To remain competitive in this environment, US universities increasingly seek to recruit foreign scholars and students, and the brightest members of ethnic minorities. Given this struggle for the best and brightest students, and to comply with affirmative action, universities within the US academic field increasingly use videos to target different audiences to increase their pool of (affluent) students. Against this backdrop, the contribution at hand addresses this research gap by exploring how videos can be incorporated into mixed-methods higher education research. To this end, the paper examines 450 Youtube promotional videos from US universities. More specifically, the contribution addresses the following research questions:

1) To what extent are different minorities (e.g., African Americans) portrayed separately by gender at universities of different rank?

2) In what situations (e.g., research, teaching, sports, party life) are different groups embedded in the videos?

In order to answer these questions, the project firstly relies on convolutional neural networks RetinaFace for face-recognition and DeepFace for the extraction of gender and ethnicity in the university YouTube-videos. Second, deductive qualitative content analysis is employed to encode the situations and objects depicted. Finally, multiple factor analysis is employed to combine video data with university data, such as financial endowments, student body composition, faculty, and ranking positions.

Preliminary results indicate that in the videos of high-ranking elite universities, predominantly white males dominate in research contexts, while minorities remain primarily in the role of face for students on campus and in recreational activities. For mid- to lower-ranked universities, the composition tends to become more diverse, but focuses on seminar contexts for female faculty and on sports as well as campus recreational activities for minority members.

Keywords: convolutional neural networks, US-universities, visual analysis, minorities, mixed methods

Digital economy, social space and symbolic power. Correspondence and cluster analysis results of secondary analyses of Eurobarometer surveys

Ulf Wuggenig[1], Cheryce von Xylander[1], Christian Tarnai[2]

[1]Leuphana University Lüneburg, Germany; [2]Universität der Bundeswehr München, Germany; ulf.wuggenig@leuphana.de, cheryce.von_xylander@leuphana.de, christian.tarnai@unibw.de

The discourse on the regulation of the digital economy is gaining momentum both in Europe and in other countries and regions. At the European level, which is the focus of this paper, discussion initially centered on the Digital Market Act and the Digital Services Act. Attention has increasingly shifted to the AI act being prepared in the European Commission and the European Parliament. Controversy abounds on effective policy measures to regulate the digital transformation. The interests of different agents in bureaucratic/political/economic fields are continually mobilized in lobbying efforts. Within the framework of these disputes – latently as well as openly contested conflicts – data is weaponized in the shaping of a public opinion that, so Bourdieu, 'does not exist' (1972). Habermas (1962), by contrast, regarded public opinion surveys as integral to a public sphere, which he deems constitutive of democracy as such, but also dangerously subject to biased suasion. Though Habermas continues to revise his theory of communicative action, most recently in light of perceived threats from digitization and social media (Habermas 2022), Bourdieu arguably provides a better field guide to navigating the virtual arena. However, Bourdieu's radical scepticism and critique of bracketed publics and causal reductionist quantification seems apt as ever. On this theoretical background, the paper examines how public opinion is invoked in discourses on digitization and digital economy in the form of Eurobarometer surveys. In particular, indicators of social space and symbolic power (Wacquant 2019; Blasius, Lebaron, LeRoux, Schmitz eds. 2019; Atkinson, Schmitz 2022) are extrapolated using simple and multiple correspondence analyses as well as cluster analysis to reveal the epistemic salience as well as limitations of such surveys (Bourdieu, 1972; Blasius and Thiessen 2006).

Keywords: digital economy, public opinion surveys, Eurobarometer, MCA, cluster analysis

Supervised learning of outcome-relevant items from a questionnaire via mixed integer optimization

Leyao Zhang[1], Peter Xuekun Song[2]

[1]University of Michigan, United States of America; [2]University of Michigan, United States of America; pxsong@umich.edu

Questionnaires are among the oldest and most widely used instruments in practice to measure variables relevant to traits of interest that cannot be easily measured by physical devices, for example, depression. In many practical settings, the scope of an existing questionnaire is often unfit to apply to a new study population, whose underlying characteristics are different from those of the original population used for the questionnaire's development and/or validation. Motivated by a cohort study of elderly asthma patients, we aim to examine associations between health outcomes and quality of life (QoL) measured by a QoL questionnaire. To increase comparability, we consider a supervised learning method to identify a subset of questions whose summary score is maximally associated with a specific health outcome under investigation. The resultant set of selected items gives an optimal summary metric of the questionnaire, which improves both statistical power and interpretation. Our item extraction procedure is built upon the best subset algorithm implemented by a mixed integer programming, which enjoys both theoretical guarantee of selection consistency and flexibility of handling non-response missing data. Our methodology is first evaluated by extensive simulation studies with comparisons to existing methods, and then applied to derive tailored QoL scores adaptive to two outcomes of lung function measure (FEV1) and asthma control test (ACT), respectively, among elderly people with persistent asthma.

Keywords: dimension reduction, ordinal categorical data, regression analysis, supervised learning

CAbiNet: joint visualization of cells and genes based on a gene-cell graph

Yan Zhao, Clemens Kohl[1], Martin Vingron[1]

[1]Max Planck Institute for Molecular Genetics, Germany; zhao@molgen.mpg.de, kohl@molgen.mpg.de

Contrary to conventional cell clustering approaches, biclustering algorithms aim to simultaneously find functionally related sets of genes and the corresponding cells in a single step. However, existing biclustering algorithms struggle with the higher noise and large size of single-cell RNA-sequencing (scRNA-seq) data and lack tools for a practical and concise visualization of results. Here we present correspondence analysis based biclustering on networks, abbreviated as CAbiNet, a graph-based biclustering approach for scRNA-seq data.

CAbiNet consolidates several analysis steps through efficient co-clustering of cells and their respective marker genes and joint visualization of the biclustering results in a non-linear embedding. It fills the need for a well-performing biclustering algorithm for scRNA-seq and spatial transcriptomics data, streamlines existing workflows and allows for easy and interactive visual exploration of cells and their corresponding marker genes in a single plot for cell type annotation.

Keywords: correspondence analysis, biological data, nonlinear embedding

CARME

CARME in mythology

Carme, the Latinized form of Greek Karmê ("shearer"), was a female Cretan spirit. According to Olympian mythology, she was the mother, by Zeus, of the virginal huntress Britomartis. Carme was the daughter of either Phoenix and Cassiopeia, or of the divine ploughman Euboulos, son of Karmanor. The name Karmanor is simply "the man of Karme", an epithet with the masculine '-or' suffix describing his role.

CARME in astronomy

Carme, is a retrograde irregular satellite of Jupiter. It was discovered by Seth Barnes Nicholson at Mount Wilson Observatory in California in July 1938. It is named after the mythological Carme, mother by Zeus of Britomartis, a Cretan goddess. Before 1975 it was simply known as Jupiter XI. The **Carme group** is a group of retrograde irregular satellites of Jupiter that follow similar orbits to Carme and are thought to have a common origin.

CARME in Catalonia

Carme is a municipality in the county of Anoia in Catalonia, Spain.

Carme is a feminine given name of two separate origins. The first is a Galician and Catalan form of Hebrew karmel, "garden". The second is from Greek Karmē, whose name means "she who cuts the grain", from keirein, "to cut".

CARME in statistics

CARME is the abbreviation of "Correspondence Analysis and Related Methods". It describes a series of quadriennial conferences, started in 1991 at the Central Archive of Empirical Social Research, Cologne, Germany, and repeated in Cologne 1995 and 1999, then in Barcelona (2003) and Rotterdam (2007). The 6th CARME conference took place in Rennes, France, February, 2011, celebrating 50 years of the original work in Rennes itself on correspondence analysis by Jean-Paul Benzécri. The 7th CARME conference was in Naples, Italy, September, 2015 and the 8th CARME conference in Cape Town, South Africa, February 2019. The 9th CARME conference takes place in Bonn, Germany, September, 2023. The website of the **CARME network** is

www.carme-n.org

Michael Greenacre, Barcelona
Source: Wikipedia
Jean-Marie Monget, Sophia-Antipolis

CARME

CARME dans le chrétienté

Le terme "**Carme"** désigne un membre masculin de l'Ordre religieux du Carmel. Cet ordre est né sur le mont Carmel en Galilée à la fin du XIIéme siècle. Les premiers frères étant des anciens croisés demeurés en "Terre Sainte" et vivant en ermites dans des grottes. Rapatrié en Europe, cet ordre est surtout connu pour la branche féminine des carmélites inspirée par St Thérèse d'Avila. A la Révolution française, le Couvent des Carmes à Paris, est le théâtre d'un massacre de religieux. Il n'en reste aujourd'hui que l'église St Joseph des Carmes, 72 Rue de Vaugirard.

CARME en Anglais

En Anglais, le terme orthographié en "**Carme ou Karme**" désigne la nymphe grecque "**Κάρμη**" parfois liée dans la mythologie crétoise à Demeter, le dieu des moissons. Il s'orthographie "**Carmé**" en Français.

CARME en Catalan

Carme est une municipalité du conté d'Anoia en Catalogne. En Catalan et en Galicien le prénom féminin ortographié en "Carme" est très commun. Mais il proviendrait dans ce cas du terme hébreux "Karmel" désignant un jardin. Il pourrait aussi être la version catalane du prénom "Carmen" en espagnol

CARME en probabilités et théorie des jeux

Le **Carme** est un coup au jeu de tric-trac où deux quatre sortent simultanément. Le tric-trac est un jeu de dés semblable à celui du backgammon ou du jeu de "tavli" très populaire en Grèce et en Turquie.

CARME en statistique

CARME est l'abréviation du terme anglais "Correspondence Analysis and Related Methods". Il désigne une série de congrès quadrienneaux, initiée en 1991 au *Zentralarchiv für Empirische Sozialforschung* de Cologne, repété à Cologne en 1995 et 1999, et puis à Barcelone (2003), où le nom CARME a été utilisé pour la première fois, Suivit ensuite celui de Rotterdam, Pays Bas en 2007. Le 6ème congrès CARME s'est déroulé à Rennes, France en février 2011, célébrant les 50 ans de la création à l'Université de Rennes de l'Analyse des Correspondances par Jean-Paul Benzécri. Le 7ème a eu lieu à Naples, Italie en septembre 2015 et le 8ème à Stellenbosch, Afrique du Sud en février 2019. Le 9ème congrès CARME se tient à Bonn, Allemagne en septembre 2023.. Le site Web du réseau CARME est **www.carme-n.org**

Source: Wikipedia

Jean-Marie Monget, Sophia-Antipolis
Michael Greenacre, Barcelona