



DAGStat 2022

Statistics under one umbrella

6TH JOINT STATISTICAL MEETING

March 28 – April 1, 2022, UKE Hamburg

MIND THE GAP –

INTERPLAY BETWEEN THEORY AND PRACTICE



Book of Abstracts

DAGStat 2022

Book of Abstracts

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliographische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

1. Aufl. – Göttingen: Cuvillier, 2022

Electronic Version 25.03.2022

Slight differences to the printed version due to some recent changes in the program.

© CUVILLIER VERLAG, Göttingen 2022

Nonnenstieg 8, 37075 Göttingen

Telefon: 0551-54724-0

Telefax: 0551-54724-21

www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.

1. Auflage, 2022

Gedruckt auf umweltfreundlichem, säurefreiem Papier aus nachhaltiger Forstwirtschaft.

ISBN 978-3-7369-7586-6

eISBN 978-3-7369-6586-7

PROGRAM

Monday, March 28, 7:30 pm – 9:00 pm

Statistik in der Öffentlichkeit

Chair(s): Tim Friede, Ralf Münnich, Room: HSN55

Walter J. Radermacher

Statistik für Klimaschutz und Gesundheit – (mehr) Fortschritt wagen!

7:30pm – 9:00pm (CTID 159 on page 130)

Tuesday, March 29, 9:00 am – 10:40 am

Design of Experiments and Clinical Trials I

Chair(s): Kirsten Schorning, Geraldine Rauch, Room: HSN30

Cornelia Ursula Kunz

Adaptive phase 2/3 drug development programs – The pros and cons

9:00am – 9:40am (CTID 118 on page 107)

Marius Placzek, Tim Friede

Blinded sample size recalculation in adaptive enrichment designs

9:40am – 10:00am (CTID 157 on page 129)

Stephen Schüürhuis, Kevin Kunzmann, Cornelia Ursula Kunz

Alternative approaches to Delayed Response Group Sequential Design

10:00am – 10:20am (CTID 190 on page 153)

Maximilian Pilz, Meinhard Kieser

Optimality of the inverse normal combination test for two-stage group-sequential designs with early stopping

10:20am – 10:40am (CTID 156 on page 129)

Network Analysis

Chair(s): Göran Kauermann, Steffen Nestler, Room: HSW30

Carter Tribley Butts

Parametric Network Models for Sets of Relations

9:00am – 9:40am (CTID 37 on page 57)

Cornelius Fritz, Marius Mehrl, Paul W. Thurner, Göran Kauermann

All that Glitters is not Gold: Relational Events Models with Spurious Events

9:40am – 10:00am (CTID 66 on page 76)

Robert Wilhelm Krause, Mark Huisman

Estimation of ERGMs under Missing Attribute Data

10:00am – 10:20am (CTID 115 on page 105)

Markus Schepers, Jochem König, Konstantin Strauch

Modelling of infectious disease spread with spatio-temporal networks

10:20am – 10:40am (CTID 182 on page 148)

Statistics of High Dimensional Data I

Chair(s): Nestor Parolya, Room: HSN61

Holger Dette

Testing relevant hypotheses for functional data

9:00am – 9:40am (CTID 47 on page 63)

Fabian Mies, Ansgar Steland

Sequential Gaussian approximation for nonstationary time series in high dimensions

9:40am – 10:00am (CTID 143 on page 121)

Stanislav Anatolyev, Mikkel Mikkel Soelvesten

Testing many restrictions under heteroskedasticity

10:00am – 10:20am (CTID 2 on page 35)

Martin Spindler, Philipp Bach, Sven Klaassen, **Jannis Kück**

Uniform Inference in High-Dimensional Generalized Additive Models

10:20am – 10:40am (CTID 201 on page 160)

Statistical Literacy and Statistical Education I

Chair(s): Rolf Biehler, Room: HSN43

Mine Çetinkaya-Rundel, Johanna Hardin, Benjamin Baumer, Amelia McNamara, Nicholas Horton, Colin Rundel

An educator's perspective of the tidyverse

9:00am – 9:40am (CTID 42 on page 60)

Maximilian Michael Mandl, Roman Hornung, Christina Nießl, Theresa Ullmann, Anne-Laure Boulesteix

A didactic experiment on fishing expeditions

9:40am – 10:00am (CTID 133 on page 116)

Henrike Weinert, Katja Ickstadt, Henrik Müller

Data Competence Network – Data Literacy Education an der Technischen Universität Dortmund

10:00am – 10:20am (CTID 227 on page 176)

Joachim Engel, Laura Martignon

Data Science for informed citizens: On synergies between Digital Literacy and Civic Statistics

10:20am – 10:40am (CTID 56 on page 69)

Survey Methodology I

Chair(s): Heinz Leitgöb, Carina Cornesse, Room: HSN55

Annette Jäckle

Asking respondents to do more than answer survey questions

9:00am – 9:40am (CTID 93 on page 92)

Florian Berens, Sebastian Hobert

Influence of the Design of Open-Ended Survey Questions Using Conversational Agents or Voice-Based Input on the Responses

9:40am – 10:00am (CTID 18 on page 45)

Henning Silber, Barbara Felderer, Carina Cornesse

When and Why Do Respondents Consent to Passive Data Collection – Evidence from Mouse Clicks and GPS Data

10:00am – 10:20am (CTID 195 on page 156)

Zaza Zindel, Simon Kühne

Opportunities, Challenges, and Limitations of Social Media Recruitment for Rare and Hard-to-Reach Populations

10:20am – 10:40am (CTID 239 on page 183)

Tuesday, March 29, 11:10 am – 12:35 pm

Plenary I & Opening

Chair(s): Antonia Zapf, Tim Friede, Room: HSN55

Miguel Hernan

Because there is no other way: Estimating vaccine effectiveness using observational data

(CTID 81 on page 84)

Tuesday, March 29, 1:30 pm – 2:50 pm

Bioinformatics and Systems Biology I

Chair(s): Tim Beißbarth, Anne-Laure Boulesteix, Room: HSW30

Mark Robinson

Adventures in benchmarkizing computational biology

1:30pm – 2:10pm (CTID 170 on page 140)

Clemens Kreutz

A powerful curve fitting model for approximation of the transient differential equations dynamics

2:10pm – 2:30pm (CTID 117 on page 106)

Daniel Fridljang

Better multiple Testing: Using multivariate co-data for hypotheses

2:30pm – 2:50pm (CTID 65 on page 75)

Design of Experiments and Clinical Trials II

Chair(s): Kirsten Schorning, Geraldine Rauch, Room: HSN30

Frank Miller, Renata Eirini Tsirpitz

Optimal dose-finding for efficacy-safety-models of Emax-type

1:30pm – 1:50pm (CTID 145 on page 122)

Fritjof Freise, Heinz Holling, Rainer Schwabe

On D-Optimal designs for repeated item response testing

1:50pm – 2:10pm (CTID 64 on page 74)

Leonie Schürmeyer, Kirsten Schorning, Jörg Rahnenführer

Comparison of Different Designs for Dose-Response Gene Expression Data

2:10pm – 2:30pm (CTID 188 on page 152)

Parisa Parsamaram, Rainer Schwabe, Heinz Holling

Design and Analysis in a non-linear longitudinal Poisson regression model

2:30pm – 2:50pm (CTID 153 on page 127)

Statistics of High Dimensional Data II

Chair(s): Jörg Rahnenführer, Room: HSN61

Fernando Castro-Prado, Dominic Edelmann, Jelle J Goeman

A kernel and distance-covariance perspective on the genomics of quantitative traits

1:30pm – 1:50pm (CTID 41 on page 59)

Franziska Kappenberg, Wiebke Albrecht, Jan G. Hengstler, Jörg Rahnenführer

Prediction of hepatotoxicity based on gene expression data

1:50pm – 2:10pm (CTID 99 on page 96)

Hannah Klinkhammer, Christian Staerk, Carlo Maj, Peter Krawitz, Andreas Mayr

Component-wise L2-boosting for polygenic risk scores based on large cohort data

2:10pm – 2:30pm (CTID 107 on page 100)

Shuo Wang, Edwin Kipruto, Willi Sauerbrei

Using methods to identify influential points in high-dimensional data

2:30pm – 2:50pm (CTID 222 on page 173)

Statistical Literacy and Statistical Education II

Chair(s): Rolf Biehler, Room: HSN43

Yannik Fleischer, Rolf Biehler, Susanne Podworny

Teaching and learning machine learning using educationally designed Jupyter Notebooks

1:30pm – 1:50pm (CTID 62 on page 73)

Nicole Steib, **Karin Binder**, Büchter Theresa

TrainBayes - A training study for medical and law students to improve Bayesian reasoning

1:50pm – 2:10pm (CTID 206 on page 163)

Florian Berens, Sebastian Hobert

Why Are the Motivated Successful in Our Classes: A Learning Analytics Study on the Effects of Attitudes Toward Statistics on Learning Behavior

2:10pm – 2:30pm (CTID 19 on page 46)

Angelina Hofacker, **Markus Zwick**

The Covid-19 pandemic as a driver of change in official statistics: New roles of the Federal Statistical Office illustrated by the Covid-19 data hub

2:30pm – 2:50pm (CTID 85 on page 87)

Time Series Analysis I

Chair(s): Dominik Wied, Room: HSN55

Piotr Fryzlewicz

Narrowest Significance Pursuit: inference for multiple change-points in linear models

1:30pm – 2:10pm (CTID 67 on page 76)

Marvin Borsch, Dominik Wied, Alexander Mayer

Consistent Estimation of Multiple Breakpoints in Dependence Measures

2:10pm – 2:30pm (CTID 28 on page 52)

Claudia Kirch, **Philipp Klein**, Marco Meyer

Anomaly detection based on MOSUM statistics in large image data

2:30pm – 2:50pm (CTID 103 on page 98)

Tuesday, March 29, 3:20 pm – 4:40 pm

Bioinformatics and Systems Biology II

Chair(s): Tim Beißbarth, Anne-Laure Boulesteix, Room: HSW30

Martin Treppner, Harald Binder

scVIDE: Single Cell Variational Inference for Designing Experiments

3:20pm – 3:40pm (CTID 212 on page 166)

Johannes Ostner, Salomé Carcy, Christian L. Müller

Bayesian tree-aggregated analysis of compositional amplicon and single-cell data

3:40pm – 4:00pm (CTID 150 on page 125)

Maren Hackenberg, Harald Binder

Combining deep learning and modeling for time-series single-cell RNA-sequencing data

4:00pm – 4:20pm (CTID 73 on page 79)

Chiara Balestra, Carlo Maj, Emmanuel Müller, Andreas Mayr

A game-theoretic approach for unsupervised ranking and selection of pathways in gene sets

4:20pm – 4:40pm (CTID 9 on page 39)

Official and Survey Statistics I

Chair(s): Hanna Brenzel, Ralf Münnich, Room: HSN43

Danny Pfeffermann

Accounting for mode effects and proxy surveys in survey sampling inference with nonignorable nonresponse

3:20pm – 4:00pm (CTID 154 on page 127)

Jan Pablo Burgard

Estimation of domain specific consumer baskets in Germany

4:00pm – 4:20pm (CTID 35 on page 56)

Patrick Krennmair, Nora Würz, Timo Schmid

Analysing Opportunity Cost of Care Work using Mixed Effects Random Forests under Aggregated Census Data

4:20pm – 4:40pm (CTID 116 on page 106)

Statistics in Agriculture and Ecology, Environmental Statistics I

Chair(s): Alessandro Fassò, Karin Hartung, Room: HSN30

Kelly McConville

Estimating land use and land cover change: An exploration of change measures, estimation tools, and data sources for change detection

3:20pm – 4:00pm (CTID 138 on page 119)

Patricia Szokol, Sándor Baran, Marianna Szabó

Truncated generalized extreme value distribution based EMOS model for calibration of wind speed ensemble forecasts

4:00pm – 4:20pm (CTID 210 on page 165)

Mária Lakatos, Sándor Baran

Restoration of Temporal Dependence in Statistical Post-processing of Ensemble Weather Forecasts

4:20pm – 4:40pm (CTID 121 on page 109)

Statistics in Finance I

Chair(s): Roxana Halbleib, Lars Winkelmann, Room: HSN61

Roberto Renò

Zeros

3:20pm - 4:00pm (CTID 166 on page 138)

Anine E. Bolko, Kim Christensen, Mikko S. Pakkanen, Bezirgen Veliyev

Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures

4:00pm - 4:20pm (CTID 27 on page 51)

Markus Bibinger

Inference on jumps in high-frequency order-price models with one-sided noise

4:20pm - 4:40pm (CTID 20 on page 47)

Time Series Analysis II

Chair(s): Christian Weiß, Room: HSN55

Vivien Less, Philipp Sibbertsen

A Multivariate Perturbation Robust Test Against Spurious Long Memory

3:20pm – 3:40pm (CTID 125 on page 112)

Simon Wingert, **Mwasi Paza Mboya**, Philipp Sibbertsen

Distinguishing between breaks in the mean and breaks in persistence under long memory

3:40pm – 4:00pm (CTID 232 on page 179)

Malte Knüppel, Fabian Krüger, **Marc-Oliver Pohle**

Score-based calibration testing for multivariate forecast distributions

4:00pm – 4:20pm (CTID 110 on page 102)

Rouven Michels, Roland Langrock

Using tensor product B-splines for nonparametric inference in multivariate hidden Markov models

4:20pm – 4:40pm (CTID 142 on page 121)

Tuesday, March 29, 5:00 pm – 6:20 pm

Official and Survey Statistics II

Chair(s): Hanna Brenzel, Ralf Münnich, oom: HSN55

Ulrich Rendtel, Marcus Gross, Timo Schmid, Kerstin Erfurth

Kernel density smoothing of composite spatial data on administrative area level: An application to incidence maps of COVID19 infections in Germany

5:00pm – 5:20pm (CTID 165 on page 138)

Sandra Hadam

Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsforschreibung und Mobilfunkdaten

5:20pm – 5:40pm (CTID 74 on page 80)

Shari Stehrenberg, Alexandra Ils

Trusted Smart Surveys: Solutions for the European Statistical System - Objectives and Main Challenges

5:40pm – 6:00pm (CTID 205 on page 162)

Jannik Schaller

Data Fusion in the Context of Microsimulation: Combining Semi-parametric Methods with Statistical Learning Approaches

6:00pm – 6:20pm (CTID 179 on page 146)

Statistics in Agriculture and Ecology, Environmental Statistics II

Chair(s): Alessandro Fassò, Karin Hartung, Room: HSN30

Ágnes Baran, Sándor Baran

Calibration of wind speed ensemble forecasts for power generation

5:00pm – 5:20pm (CTID 10 on page 40)

Dany Djeudeu, Michael Schweitzer, Susanne Moebus, Katja Ickstadt

Dynamic of the urban acoustic environment during COVID-19 Lockdown in Urban Ruhr Area, Germany

5:20pm – 5:40pm (CTID 51 on page 66)

Marianna Szabó, Sándor Baran

Parametric post-processing of dual-resolution precipitation ensemble forecasts

5:40pm – 6:00pm (CTID 209 on page 164)

Paolo Maranzano, Philipp Otto, Alessandro Fassò

Penalized likelihood adaptive-LASSO algorithm for feature selection in functional HDGM

6:00pm – 6:20pm (CTID 134 on page 117)

Statistics in Finance II

Chair(s): Roxana Halbleib, Lars Winkelmann, Room: HSN61

Tobias Fissler, Marc-Oliver Pohle

Generalized Correlation Measures

5:00pm – 5:20pm (CTID 61 on page 72)

Alexander Mayer, Dominik Wied

Estimation and Inference in Factor Copula Models with Exogenous Covariates

5:20pm – 5:40pm (CTID 137 on page 118)

Timo Dimitriadis, Yannick Hoga

Co-Quantile Regression

5:40pm – 6:00pm (CTID 49 on page 64)

Ekaterina Kazak, Roxana Halbleib, Winfried Pohlmeier

Bagged Value-at-Risk Forecast Combination

6:00pm – 6:20pm (CTID 101 on page 97)

Statistics in Science, Technology and Industry + Digital and

Sensor Data

Chair(s): Rüdiger Pryss, Sonja Kuhnt, Room: HSW30

Susan A. Murphy

Assessing Personalization in Digital Health

5:00pm – 5:40pm (CTID 147 on page 123)

Bradley Lubich, **Daniel R. Jeske**, Weixin Yao

Statistical Inference for Method of Moments Estimators of a Semi-Supervised Two-Component Mixture Model

5:40pm – 6:20pm (CTID 129 on page 114)

Time Series Analysis III

Chair(s): Christian Weiß, Room: HSN55

Guy-Niklas Brunotte

Dependent Wild Bootstrap based Test for Independence under Local Stationarity

5:00pm – 5:20pm (CTID 32 on page 54)

Carina Beering

A Test of Independence for Locally Stationary Processes Using a Weighted Characteristic Function-based Distance

5:20pm – 5:40pm (CTID 15 on page 43)

Maxime Faymonville, Carsten Jentsch, Christian Weiß, Boris Aleksandrov

Penalized estimation of INAR models

5:40pm – 6:00pm (CTID 58 on page 70)

Anne Leucht, Efstathios Paparoditis, Daniel Rademacher, Theofanis Sapatinas

Testing equality of spectral density operators for functional linear processes

6:00pm – 6:20pm (CTID 126 on page 112)

Wednesday, March 30, 9:00 am – 10:40 am

Bayesian Statistics I

Chair(s): Thomas Kneib, Room: HSN43

Andrea Riebler

Prior elicitation for variance parameters in Bayesian hierarchical models

9:00am – 9:40am (CTID 168 on page 139)

Boyao Zhang, Colin Griesbach, Elisabeth Bergherr

Bayesian Boosting for Simultaneous Estimation and Selection of Fixed and Random Effects in High-Dimensional Mixed Models

9:40am – 10:00am (CTID 237 on page 182)

Anja Rappl, Sebastiaan Klokman, Manuel Carlan, Thomas Kneib, Elisabeth Bergherr

Bayesian Variable and Effect Selection for Quantile Regression

10:00am – 10:20am (CTID 161 on page 134)

Sophie Tchanyou Ganme, Katja Ickstadt

A Bayesian Inference for Intensity Based Load-Sharing Models with Damage Accumulation

10:20am – 10:40am (CTID 211 on page 166)

Clustering and Classification

Chair(s): Christian Hennig, Bettina Grün, Room: HSN43

Iven Van Mechelen

Benchmarking in cluster analysis: Preview of a white paper

9:00am – 9:40am (CTID 218 on page 170)

Christian Hennig

An empirical comparison and characterisation of nine popular clustering methods

9:40am – 10:00am (CTID 79 on page 83)

Theresa Ullmann, Anna Beer, Maximilian Hünemörder, Thomas Seidl, Anne-Laure Boulesteix

Over-optimistic evaluation and reporting of novel cluster algorithms: An illustrative study

10:00am – 10:20am (CTID 215 on page 168)

Amanda Fernández-Fontelo, Felix Henninger, Pascal J. Kieslich, Frauke Kreuter, Sonja Greven

A new ensemble model for multivariate functional data classification with an application to survey research

10:20am – 10:40am (CTID 60 on page 72)

Statistics in Behavioral and Educational Sciences

Chair(s): Philipp Doebler, Stefan Krauss, Room: HSW30

Sven Hilbert

Specifics of analyzing educational data

9:00am – 9:40am (CTID 84 on page 86)

Annette Möller, Ann Cathrice George, Jürgen Groß

Predicting school transition rates in Austria with classification trees

9:40am – 10:00am (CTID 146 on page 123)

Felix Zimmer, Rudolf Debelak

Simulation-based Design Optimization for Statistical Power: Utilizing Machine Learning

10:00am – 10:20am (CTID 238 on page 183)

Marie Beisemann, Philipp Doebler

Understanding Ability Differences Measured with Count Items and their Reliability: The Distributional Regression Test Model and the Count Latent Regression Model

10:20am – 10:40am (CTID 17 on page 45)

Survey Methodology II

Chair(s): Heinz Leitgöb, Carina Cornesse, Room: HSN61

Peter Selb, Sina Chen, John Körtner, Philipp Bosch

Bias and variance in multiparty election polls

9:00am – 9:20am (CTID 193 on page 155)

Matthias Sand, Christian Bruch, Barbara Felderer, Ines Schaurer, Jan-Philipp Kolb, Kai Weyandt

Design weights in panel surveys with multiple refreshment samples. A general discussion with an application to the GESIS panel.

9:20am – 9:40am (CTID 176 on page 144)

Barbara Felderer, Jannis Kück, Martin Spindler
Using Double Machine Learning to Understand Nonresponse in the Recruitment of a Mixed-mode Online Panel

9:40am – 10:00am (CTID 59 on page 71)

Julian B. Axenfeld, Christian Bruch, Christof Wolf
General-purpose imputation of planned missing data in social surveys: different strategies and their effect on correlations

10:00am – 10:20am (CTID 6 on page 37)

Julian Schüssler, **Peter Selb**
Graphical causal models for survey inference
10:20am – 10:40am (CTID 189 on page 153)

Wednesday, March 30, 11:10 am – 12:35 pm

Plenary II & DAGStat Medals

Chair(s): Tim Friede, Room: HSN55

Tamara Broderick
An Automatic Finite-Sample Robustness Metric: Can Dropping a Little Data Change Conclusions?
(CTID 31 on page 54)

Wednesday, March 30, 1:30 am – 2:50 pm

Young Statistician Session – Talks

Chair(s): Stefanie Peschel, Annika Möhl Room: HSW40

Kristin Blesch, David S. Watson, Marvin N. Wright
Conditional Feature Importance for Mixed Data
1:20pm – 1:40pm (CTID 24 on page 49)

Michael Lau, Claudia Wigmann, Sara Kress, Tamara Schikowski, Holger Schwender
Evaluation of tree-based statistical learning methods for constructing genetic risk scores
1:40pm – 2:00pm (CTID 124 on page 111)

Maximilian Kertel, Markus Pauly
Learning the Joint Distribution with Missing Data under the Gaussian Copula Model
2:00pm – 2:20pm (CTID 102 on page 97)

Julian Martin Rodemann
Prior-mean-RObust Bayesian Optimization (PROBO)
2:20pm – 2:40pm (CTID 171 on page 141)

Jana Kleinemeier, Nadja Klein

Scalable Estimation for Structured Additive Distributional Regression Through Variational Inference

2:40pm – 3:00pm (CTID 106 on page 100)

Bayesian Statistics II

Chair(s): Katja Ickstadt, Room: HSN43

Paul Bürkner, Sergey Oladyshkin, Ilja Kröker, Wolfgang Nowak

The Sparse Polynomial Chaos Expansion: A Fully Bayesian Approach with Joint Priors on the Coefficients and Global Selection of Terms

1:30pm – 1:50pm (CTID 36 on page 56)

Oliver Grothe, **Jonas Rieger**

Combining point forecasts to calibrated probabilistic forecasts using copulas

1:50pm – 2:10pm (CTID 71 on page 78)

Sonja Jäckle, Elias Röger, Volker Dicken, Benjamin Geisler, Jakob Schumacher, Max Westphal

Supporting COVID-19 quarantine decisions with a statistical risk assessment model

2:10pm – 2:30pm (CTID 94 on page 92)

Ilya Zarubin

Bayesian Learning in a General Class of Multivariate Nonlinear Dynamic Panel Data Models

2:30pm – 2:50pm (CTID 236 on page 181)

Education – Statistics in Practice I

Chair(s): Theresa Keller, Thomas Schmelter, Willi Sauerbrei, Room: HSN55

Richard D. Keele, Gary S. Collins

Prognosis Research in Healthcare: initiatives to improve methodology standards

1:30pm – 2:50pm (CTID 164 on page 136)

Marketing and E-Commerce

Chair(s): Friederike Paetz, Daniel Guhl, Room: HSW30

Stephan Seiler

The use of machine learning methods for targeted marketing

1:30pm – 2:10pm (CTID 191 on page 154)

Carsten Schultz, Friederike Paetz

Good to hear your voice! Interactions of personality traits and smart speaker brand preferences in voice commerce

2:10pm – 2:30pm (CTID 187 on page 151)

Tobias R. Rebholz, Marco Biella, Mandy Hütter

Mixed-Effects Regression Weights of Advice: Individual Differences in Judgment Formation and Sampling

2:30pm – 2:50pm (CTID 162 on page 134)

Robust and Nonparametric Statistics I

Chair(s): Christine Müller, Melanie Birke, Room: HSN30

Catherine Aaron

Geometric inference and robustness

1:30pm – 2:10pm (CTID 1 on page 35)

Maximilian Wechsung, Konietzschke Frank

Efficient multivariate inference in general factorial diagnostic studies

2:10pm – 2:30pm (CTID 224 on page 174)

Jan Speller, Christian Staerk, Andreas Mayr

Robust statistical boosting with quantile-based adaptive loss functions

2:30pm – 2:50pm (CTID 199 on page 159)

Statistics in the Pharmaceutical and Medical Device Industry

Chair(s): Werner Brannath, Room: HSN61

James Wason

Innovative design and analysis approaches for master protocols

1:30pm – 2:10pm (CTID 223 on page 174)

Maria Stark, Antonia Zapf

Blinded sample size re-estimation in a paired diagnostic study
2:10pm – 2:30pm (CTID 203 on page 161)

Reinhard Vonthein, Christina Hölsken
Power of Logrank and RMST tests for non-constant hazard ratio
2:30pm – 2:50pm (CTID 221 on page 172)

Wednesday, March 30, 3:20 am – 4:40 pm

Advanced Regression Modelling I

Chair(s): Andreas Groll, Andreas Mayr, Room: HSN61

Brian D. Marx, Paul H. C. Eilers
Variations on Varying-Coefficient Signal Regression
3:20pm – 4:00pm (CTID 135 on page 117)

Thomas Kneib, Alexander Silbersdorff, Benjamin Säfken
Rage Against the Mean - A Review of Distributional Regression Approaches
4:00pm – 4:20pm (CTID 108 on page 101)

Almond Stöcker, **Sonja Greven**
Functional additive models on manifolds of planar shapes and forms
4:20pm – 4:40pm (CTID 207 on page 163)

Computational Statistics and Statistical Software I

Chair(s): Gero Szepannek, Room: HSW30

Julie Josse
Supervised learning with missing values
3:20pm – 4:00pm (CTID 98 on page 95)

Alexander Bauer, Maximilian Weigert, Hawre Jalal
APCtools: An R Package for Descriptive and Model-based Age-Period-Cohort Analysis
4:00pm – 4:20pm (CTID 11 on page 40)

Sheila Görz, Alexander Dürre, Herold Dehling, Roland Fried
R package 'robcp' for robust detection of change points
4:20pm – 4:40pm (CTID 69 on page 77)

Education – Statistics in Practice II

Chair(s): Theresa Keller, Thomas Schmelter, Willi Sauerbrei, Room: HSN55

Richard D. Keele, Gary S. Collins
Prognosis Research in Healthcare: initiatives to improve methodology standards
3:20pm – 4:40pm (CTID 164 on page 136)

Robust and Nonparametric Statistics II

Chair(s): Christine Müller, Melanie Birke, Room: HSN30

Leonie Selk, Jan Gertheiss

Nonparametric Regression and Classification with Functional, Categorical, and Mixed Covariates

3:20pm – 3:40pm (CTID 194 on page 156)

Lea Wegner, Martin Wendler

Robust Detection for Change-Points in Functional Time Series based on Spatial Signs and Bootstrap

3:40pm – 4:00pm (CTID 225 on page 175)

Martin Wendler, Herold Dehling, Kata Vuk

Weighted Change-Point Tests based on 2-Sample-U-Statistics

4:00pm – 4:20pm (CTID 229 on page 177)

Eftychia Solea, Kuang-Yao Lee, Bing Li

High-dimensional nonparametric functional graphical models via the additive partial correlation operator

4:20pm – 4:40pm (CTID 198 on page 158)

Survival and Event History Analysis I

Chair(s): Jan Feifel, Matthias Schmid, Room: HSN61

Maja Pohar Perme

Pseudo-observations in survival analysis

3:20pm – 4:00pm (CTID 158 on page 130)

Dina Voeltz, Amelie Forkel, Anke Schwandt, Oliver Kuss, Annika Hoyer

A parametric additive hazard model for time-to-event analysis

4:00pm – 4:20pm (CTID 219 on page 171)

Alina Schenk, Matthias Schmid

A pseudo-value approach for building regression models with time-dependent covariate effects

4:20pm – 4:40pm (CTID 181 on page 147)

Wednesday, March 30, 5:00 pm – 6:20 pm

IBS-DR Price Session

Chair(s): Annette Kopp-Schneider, Room: HSN55

Moritz Fabian Danzer, Tobias Terzer, Frank Berthold, Andreas Faldum, Rene Schmidt

Adaptive group sequential designs for single-arm phase II studies with multiple time-to-event outcomes

5:00pm – 5:15pm (CTID 44 on page 61)

Nadja Klein, Michael Stanley Smith

Bayesian variable selection for non-Gaussian responses: a marginally calibrated copula approach

5:15pm – 5:30pm (CTID 105 on page 99)

Saide Atmaca

Evaluation of Misspecified Linear Regression Models for Subgroup Analysis

5:30pm – 5:45pm (CTID 5 on page 36)

Eric Samuel Knop

Robust confidence intervals for mixed-effects meta-regression with interaction

5:45pm – 6:00pm (CTID 109 on page 102)

Sabrina Schmitt, Ann-Kathrin Ozga

*Time-to-event analysis with competing risks considering cluster structures - Comparison of methods based on a simulation study**

6:00pm – 6:15pm (CTID 185 on page 150)

Advanced Regression Modelling II

Chair(s): Andreas Mayr, Room: HSN43

Hendrik van der Wurp, Andreas Groll

Introducing Regularisation to Generalised Joint Regression Modelling with an Application to Football

5:00pm – 5:20pm (CTID 217 on page 170)

Colin Griesbach, Andreas Mayr, Elisabeth Bergherr

Variable Selection and Allocation in Joint Models via Gradient Boosting Techniques

5:20pm – 5:40pm (CTID 70 on page 78)

Alexander Volkmann, Nikolaus Umlauf, Sonja Greven

Flexible joint models for multivariate longitudinal and time-to-event data using a functional principal components representation of shared random effects

5:40pm – 6:00pm (CTID 220 on page 172)

Alena Skolkova

Ridge Model Averaging

6:00pm – 6:20pm (CTID 196 on page 157)

Computational Statistics and Statistical Software II

Chair(s): Roland Fried, Room: HSW30

Niklas Koenen, Marvin N. Wright

*Interpreting Deep Neural Networks with the R Package *innsight**

5:00pm – 5:20pm (CTID 112 on page 103)

Michel Lang, Raphael Sonabend, Franz Kiraly, Andreas Bender, Bernd Bischl

mlr3proba: A unified interface for machine learning with survival tasks

5:20pm – 5:40pm (CTID 122 on page 110)

Marc Becker, Michel Lang, Bernd Bischl

mlr3tuning: A general framework for ML hyperparameter tuning

5:40pm – 6:00pm (CTID 13 on page 42)

Philipp Bach, Victor Chernozhukov, Malte Kurz, Martin Spindler

DoubleML - An Object-Oriented Implementation of Double Machine Learning in R

6:00pm – 6:20pm (CTID 8 on page 38)

Robust and Nonparametric Statistics III

Chair(s): Christine Müller, Melanie Birke, Room: HSN30

Felix Gnottner, Claudia Kirch, Alicia Nieto-Reyes

Depth-based two-sample testing

5:00pm – 5:20pm (CTID 68 on page 77)

Hannah Blocher, Georg Schollmeyer, Christoph Jansen

Statistical Models for Partial Orders based on Data Depth and Formal Concept Analysis

5:20pm – 5:40pm (CTID 25 on page 50)

Georg Schollmeyer

On the uniform control of the Vapnik-Chervonenkis dimension in subgroup discovery using formal concept analysis

5:40pm – 6:00pm (CTID 186 on page 151)

Kerstin Rubarth, Frank Konietzschke

Estimation and Testing of Wilcoxon-Mann-Whitney Effects in factorial clustered data designs

6:00pm – 6:20pm (CTID 174 on page 143)

Survival and Event History Analysis II

Chair(s): Moritz Berger, Room: HSN61

Sandra Frank, Sabrina Dietz, Arthur Allignol, Jan Beyersmann

A stabilised Aalen-Johansen estimator with internal left-truncation and overly small risks sets

5:00pm – 5:20pm (CTID 63 on page 74)

Ina Dormuth, Markus Pauly, Geraldine Rauch, Carolin Herrmann

Addressing hazards in application – a sample size calculation approach for the average hazard ratio

5:20pm – 5:40pm (CTID 53 on page 67)

Lars Beckmann, Anke Schulz, Guido Skipka

Firth correction in Cox proportional hazard analysis in the presence of zero events

5:40pm – 6:00pm (CTID 14 on page 42)

Nikolai Spuck, Matthias Schmid, Moritz Berger

Flexible tree-structured regression models for discrete event times

6:00pm – 6:20pm (CTID 202 on page 160)

Thursday, March 31, 9:00 am – 10:40 am

Artificial Intelligence and Machine Learning I

Chair(s): Friedhelm Schwenker, Room: HSN30

Frank Köster

AI enables Innovation – Safe and Secure AI is a must for many AI-based Applications

9:00am – 9:40am (CTID 113 on page 104)

Maximilian Weigert, Magdalena Mittermeier, David Rügamer, Ralf Ludwig, Helmut Küchenhoff

Classification of atmospheric circulation patterns using a smoothed deep learning approach

9:40am – 10:00am (CTID 226 on page 175)

Anna Malinovskaya

Statistical monitoring of deep learning models

10:00am – 10:20am (CTID 132 on page 115)

Vi Thanh Pham

Statistical Learning of ECG Based on Functional Neural Networks

10:20am – 10:40am (CTID 155 on page 128)

Meta-Analysis I

Chair(s): Annika Hoyer, Room: HSW30

Dan Jackson

, Robbie C. M. van Aert

Multi-step estimators of between-study variances and covariances and their relationship with the Paule-Mandel estimator

9:00am – 9:40am (CTID 95 on page 93)

Cynthia Huber

, Norbert Benda, Tim Friede

Subgroup identification in individual participant data meta-analysis using model-based recursive partitioning

9:40am – 10:00am (CTID 91 on page 91)

Heinz Holling

, Katrin Jansen, Dankmar Böhning

Estimation of effect heterogeneity in rare events meta-analysis

10:00am – 10:20am (CTID 87 on page 88)

Justyna Wünstel

, **Peter Schlattmann**

Meta-analysis of mean values with unreported variances: estimation bias due to variance heteroscedasticity when common joint variance is assumed

10:20am – 10:40am (CTID 235 on page 181)

Statistical Methods in Epidemiology I

Chair(s): Irene Schmidtman, Ralph Brinks, Room: HSW30

Helene Jacqumin-Gadda, Reza Hashemi

Development of prediction tools for health events from multiple longitudinal predictors

9:00am – 9:40am (CTID 96 on page 94)

Nadine Binder, **Anika Schlosser**, Martin Schumacher

Reevaluating dementia incidence trends: The critical role of adequate design and methodology

9:40am – 10:00am (CTID 22 on page 48)

Sabrina Tulka, Annika Hoyer, Ralph Brinks

Statistical properties of a prevalence estimator for chronic diseases based on a differential equation: Simulation study in the illness-death-model

10:00am – 10:20am (CTID 214 on page 167)

Anastasiia Holovchak, **Ruben Wisskott**, Ralph Brinks, Annika Hoyer

Weighted Generalized Estimating Equations for Longitudinal Binary Response: Prevalence Estimation of Health Limitations in the SHARE Study

10:20am – 10:40am (CTID 88 on page 89)

STRATOS Session

Chair(s): Heiko Becher, Willi Sauerbrei, Room: HSN55

James Carpenter

Handling missing data in the analysis: practical guidance for structuring the analysis, choosing the tools, and reporting the results

9:00am – 9:25am (CTID 39 on page 58)

Veronika Deffner

Measurement error and misclassification of covariates: Should we worry?

9:25am – 9:50am (CTID 45 on page 62)

Georg Heinze

Selection of variables and functional forms for multivariable models

9:50am – 10:15am (CTID 78 on page 83)

Willi Sauerbrei, Heiko Becher

STRATOS – aims, tasks, support of the initiative

10:15am – 10:40am (CTID 177 on page 145)

Visualisation and Exploratory Data Analysis

Chair(s): Adalbert F.X. Wilhelm, Helmut Küchenhoff, Room: HSN43

Catherine Hurley

Investigating model adequacy, predictor effects and higher-order interactions for machine-learning models

9:00am – 9:40am (CTID 92 on page 91)

Moritz Hess, Dominik Scheuer, Harald Binder

A pattern extraction app for biomedical data

9:40am – 10:00am (CTID 82 on page 85)

Linda Krause, Caroline Seifert, Antonia Zapf
What implications do analysis choices have on study results?
10:00am – 10:20am (CTID 114 on page 105)

Moritz N. Lang, Reto Stauffer, Achim Zeileis
Visualizing Goodness of Fit of Probabilistic Regression Models
10:20am – 10:40am (CTID 123 on page 110)

Thursday, March 31, 11:10 am – 12:30 am

Artificial Intelligence and Machine Learning II

Chair(s): Sarah Friedrich, Room: HSN30

Philip Buczak, Markus Pauly
Analyzing the Influence of Missing Data Imputation on the Predictive Performance of Classifiers
11:10am – 11:30am (CTID 113 on page 55)

Björn-Hergen Laabs, Ana Westenberger, Inke R. König
Construction of artificial most representative trees by minimizing tree-based distance measures
11:30am – 11:50am (CTID 226 on page 108)

Lena Schmid, Alexander Gerharz, Andreas Groll, Markus Pauly
Machine Learning for Multi-Output Regression: Comparing multivariate approaches with separate univariate ones
11:50am – 12:10pm (CTID 183 on page 149)

Max Westphal, Rieke Alpers
The role of estimands in machine learning algorithm evaluation
12:10pm – 12:30pm (CTID 230 on page 178)

Advanced Regression Modelling III

Chair(s): Andreas Groll, Room: HSN55

Nicolai Hans, Nadja Klein, Andreas Mayr
Boosting copulas with continuous margins
11:10am – 11:30am (CTID 75 on page 81)

Annika Strömer, Nadja Klein, Christian Staerk, Andreas Mayr
Boosting Bivariate Structured Additive Distributional Regression Models
11:30am – 11:50am (CTID 208 on page 164)

Jonas Meier
Multivariate Distribution Regression
11:50am – 12:10pm (CTID 139 on page 119)

Charlotte C. Behning, Monika Fleckenstein, Maximilian Pfau, Christine Adrion, Lukas Goerd, Moritz Lindner, Steffen Schmitz-Valckenberg, Frank G. Holz, Matthias C. Schmid

Modeling trajectories of slowly progressing diseases: A mixed-model-based algorithm for variable transformation, prediction and age-of-onset estimation
12:10pm – 12:30pm (CTID 16 on page 44)

Meta-Analysis II

Chair(s): Heinz Holling, Room: HSW30

Theodoros Papakonstantinou, Gerta Rücker, Guido Schwarzer, Daniela Zöller, Adriani Nikolakopoulou

A mechanical analogue of network meta-analysis

11:10am – 11:30am (CTID 152 on page 126)

Georgios Seitidis, Stavros Nikolakopoulos, Ioannis Ntzoufras, Dimitris Mavridi
A stochastic search variable selection approach for identifying inconsistency in network meta-analysis

11:30am – 11:50am (CTID 192 on page 154)

Theodoros Evrenoglou, Ian R. White, Sivem Afach, Dimitris Mavridis, Anna Chaimani

Network meta-analysis of rare events using penalized likelihood regression

11:50am – 12:10pm (CTID 57 on page 70)

Manuel Wiesenfarth, Jasmin Wandel

Bayesian meta-analysis for exact and interval censored binomial outcomes

12:10pm – 12:30pm (CTID 231 on page 178)

Survival and Event History Analysis III

Chair(s): Jan Feifel, Matthias Schmid, Room: HSN43

Jasmin Rühl, Jan Beyersmann, Sarah Friedrich

General independent censoring in event-driven trials with staggered entry

11:10am – 11:30am (CTID 175 on page 143)

Marc Ditzhaus, Dennis Dobler, Jon Genuneit, Arnold Janssen, Markus Pauly

GFDsurv: A flexible toolbox for factorial survival designs as an alternative to Cox models

11:30am – 11:50am (CTID 50 on page 65)

Ghislain Sofack, Soumya Banerjee, Thodoris Papakonstantinou, Demetris Avraam, Paul Burton, Tom R. P. Bishop, Daniela Zöller

Implementing disclosure controls in DataSHIELD demonstrated by the dsSurvival package

11:50am – 12:10pm (CTID 197 on page 157)

Steffen Unkel, Steven Abrams, Andreas Wienke, Niel Hens

Random-effects models for quantifying heterogeneity in infectious disease transmission involving two types of contacts

12:10pm – 12:30pm (CTID 216 on page 169)

Statistical Methods in Epidemiology II

Chair(s): Irene Schmidtmann, Ralph Brinks, Room: HSN61

Johannes Bracher, Daniel Wolfram, Sam Abbott, Matthias an der Heiden, Sebastian Funk, Felix Günther, Davide Hailer, Stefan Heyder, Thomas Hotz, Helmut Küchenhoff, Diella Syliqi, Jan van de Kassteede, Maximilian Weigert, Alexander Ullrich, Tilmann Gneiting, Melanie Schienle

Collaborative real-time nowcasting of COVID-19 hospitalization incidence rates
11:10am – 11:30am (CTID 30 on page 53)

Lukas Refisch, Fabian Lorenz, Clemens Kreutz

Data-Driven Prediction of COVID-19 Cases in Germany for Decision Making
11:30am – 11:50am (CTID 163 on page 135)

Thomas Hotz, Stefan Heyder

Predicting COVID-19 Hospitalisation from Incidences
11:50am – 12:10pm (CTID 90 on page 90)

Stefan Heyder, Thomas Hotz, Tyll Krüger, Jan Pablo Burgard

Regional estimates of reproduction numbers with application to COVID-19
12:10pm – 12:30pm (CTID 83 on page 85)

Thursday, March 31, 1:30 pm – 2:50 pm

Advanced Regression Modelling IV

Chair(s): Andreas Groll, Andreas Mayr, Room: HSN55

Andreas Haupt, Øyvind Nicolay Wiborg, Nicolai Topstad Borgen

A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model
1:30pm – 1:50pm (CTID 77 on page 82)

Tim Kutzker, Nadja Klein, Dominik Wied

Flexible Specification Testing in Quantile Regression Models
1:50pm – 2:10pm (CTID 119 on page 108)

Tobias Hepp, Jakob Zierk, Sarem Seitz

Distributional latent class modelling for the indirect estimation of reference distributions using mixture density networks
2:10pm – 2:30pm (CTID 80 on page 84)

Fabian Otto-Sobotka, Jessica Koschate, Tania Zieschang

Modeling post-lockdown exercise training development during the COVID-19 pandemic in Germany using mixed distributional regression
2:30pm – 2:50pm (CTID 151 on page 125)

Empirical Economics and Applied Econometrics I

Chair(s): Daniel Gutknecht, Room: HSN30

Hugo Bodory, **Martin Huber**, Lukas Laffers
Evaluating (weighted) dynamic treatment effects by double machine learning
1:30pm – 2:10pm (CTID 26 on page 51)

Giovanni Mellace, Roberta Di Stefano
The inclusive Synthetic Control Method
2:10pm – 2:50pm (CTID 140 on page 120)

Mathematical Statistics I

Chair(s): Markus Pauly, Hajo Holzmann, Room: HSW30

Sebastian Doehler, Etienne Roquain
Controlling the False Discovery Exceedance for heterogeneous statistical tests
1:30pm – 1:50pm (CTID 52 on page 66)

Dominik Liebl, Matthew Reimherr
Fast and Fair Simultaneous Confidence Bands for Functional Parameters
1:50pm – 2:10pm (CTID 127 on page 113)

Marléne Baumeister, Marc Ditzhaus, Markus Pauly
Quantile-based MANOVA: A new tool for inferring multivariate data in factorial designs
2:10pm – 2:30pm (CTID 12 on page 41)

Frank Röttger, Sebastian Engelke, Piotr Zwiernik
Total positivity in multivariate extremes
2:30pm – 2:50pm (CTID 172 on page 141)

Survival and Event History Analysis IV

Chair(s): Steffen Unkel, Room: HSN43

Nadine Binder, **Kathrin Möllenhoff**, August Sigle, Holger Dette
Similarity of competing risks models with constant intensities in an application to clinical healthcare pathways involving prostate cancer surgery
1:30pm – 1:50pm (CTID 21 on page 47)

Angela Carollo, Hein Putter, Paul H. C. Eilers, Jutta Gampe
Smooth hazards with multiple time scales
1:50pm – 2:10pm (CTID 38 on page 57)

Dominic Edelmann, Tobias Terzer, Axel Benner
The time to progression ratio in molecular tumor trials - a critical examination of current practice and suggestions for alternative methods
2:10pm – 2:30pm (CTID 55 on page 68)

Rafael Weißbach, Dominik Wied
Truncating the exponential with a uniform distribution
2:30pm – 2:50pm (CTID 228 on page 177)

Statistical Methods in Epidemiology III

Chair(s): Irene Schmidtmann, Ralph Brinks, Room: HSN61

Patrick Schmidt

Inference under Superspreading: Determinants of SARS-CoV-2 Transmission in Germany

1:30pm – 1:50pm (CTID 184 on page 149)

Anabelle Wong, Sarah Kramer, Marco Piccininni, Jessica L. Rohmann, Tobias Kurth, Sylvie Escolano, Ulrike Grittner, Matthieu Domenech de Cellès

Using LASSO regression to estimate the population-level impact of pneumococcal conjugate vaccines

1:50pm – 2:10pm (CTID 233 on page 179)

Roman Schefzik, Leonie Boland, Bianka Hahn, Thomas Kirschning, Holger A. Lindner, Manfred Thiel, Verena Schneider-Lindner

A control selection strategy for differential network testing in intensive care: Revealing diverging dynamics of organ system interactions for survivors and non-survivors

2:10pm – 2:30pm (CTID 180 on page 147)

Nicolas Apfel, Xiaoran Liang

Agglomerative Hierarchical Clustering for Selecting Valid Instrumental Variables

2:30pm – 2:50pm (CTID 4 on page 36)

Thursday, March 31, 3:20 pm – 4:40 pm

Data Science

Chair(s): Berthold Lausen, Bernd Bischl, Room: HSN43

Rebecca Nugent

Demystifying and Optimizing Data Science

3:20pm – 4:00pm (CTID 149 on page 124)

Nadja Klein, Clara Hoffmann

Marginally calibrated response distributions for end-to-end learning in autonomous driving

4:00pm – 4:20pm (CTID 104 on page 98)

Lucas Kock, Nadja Klein, David J. Nott

Variational Inference and Sparsity in High-Dimensional Deep Gaussian Mixture Models

4:20pm – 4:40pm (CTID 111 on page 103)

Empirical Economics and Applied Econometrics II

Chair(s): Robert Jung, Room: HSN30

Claudia Noack, Tomasz Olma, Christoph Rothe

Flexible Covariate Adjustments in Regression Discontinuity Designs

3:20pm – 3:40pm (CTID 148 on page 124)

Mirjam Reutter, Gregor Pfeifer, Kristina Strohmaier
Does Smoking Affect Wages?

3:40pm – 4:00pm (CTID 167 on page 139)

Simone Maxand

A panel SVAR model for European climate policy

4:00pm – 4:20pm (CTID 136 on page 118)

Matei Demetrescu, Mehdi Hosseinkouchak, Paulo Rodrigues
Cross-Sectional Error Dependence in Panel Quantile Regressions

4:20pm – 4:40pm (CTID 46 on page 63)

Mathematical Statistics II

Chair(s): Markus Pauly, Hajo Holzmann, Room: HSW30

Ismaël Castillo

Some theoretical properties of Bayesian Tree methods

3:20pm – 4:00pm (CTID 40 on page 59)

Paul Bach, Nadja Klein

Marginally calibrated response distributions for end-to-end learning in autonomous driving

4:00pm – 4:20pm (CTID 7 on page 38)

Maximilian Steffen, Mathias Trabs

PAC-Bayes training for sparse neural networks

4:20pm – 4:40pm (CTID 204 on page 162)

Text Mining and Content Analysis

Chair(s): Andreas Blätte, Tilman Becker, Room: HSN61

Tomáš Mikolov

Statistical language modeling explained

3:20pm – 4:00pm (CTID 144 on page 122)

Jonas Rieger, Carsten Jentsch, Jörg Rahnenführer

Improving the reliability of LDA results using LDAPrototype as selection criterion

4:00pm – 4:20pm (CTID 169 on page 140)

Thursday, March 31, 5:00 pm – 6:30 pm

Plenary III & Poster/Talklet Prices

Chair(s): Hajo Holzmann, Room: HSN55

Trevor John Hastie

Some comments on CV

(CTID 76 on page 81)

Friday, April 1, 9:00 am – 10:40 am

Causal Inference

Chair(s): Uwe Siebert, Helmut Farbmacher, Room: HSN55

Martin Spindler, Jannis Kück, Zigan Wang, Ye Luo

Estimation and Inference of Treatment Effects with L2-Boosting in High-Dimensional Settings

9:00am – 9:40am (CTID 200 on page 159)

Nicolas Apfel, Helmut Farbmacher, Martin Huber, Henrika Langen

Retrieving grouped LATEs via Classifier-Lasso

9:40am – 10:00am (CTID 3 on page 35)

Franz Leon Classe, Christoph Kern

Survey Scale Forests: Estimating Unconfounded Latent Variable Effects

10:00am – 10:20am (CTID 43 on page 61)

Vanessa Didelez, Ronja Foraita, Janine Witte

Causal discovery with incomplete cohort data

10:20am – 10:40am (CTID 48 on page 64)

IQWiG & IQTIG Session

Chair(s): Tim Friede, Room: HSN43

Michael Höhle, Jona Cederbaum

Statistical Challenges in the Quality Assurance of Healthcare

9:00am – 9:20am (CTID 86 on page 87)

Maurilio Gutzeit, **Johannes Rauh**, Jona Cederbaum

Modelling Volume-Outcome Relationships in Health Care

9:20am – 9:40am (CTID 72 on page 79)

Christian Röver, Jona Lilienthal, Sibylle Sturtz, Ralf Bender, Tim Friede

Elicitation of empirical information on between-study heterogeneity in Bayesian meta-analysis

9:40am – 10:00am (CTID 173 on page 142)

Jona Lilienthal, Sibylle Sturtz, Christian Röver, Tim Friede, Ralf Bender

Application of a hierarchical Bayesian model to determine the empirical distribution of the heterogeneity parameter in IQWiG reports

10:00am – 10:20am (CTID 128 on page 113)

Matthias Maiworm, **Christoph Schürmann**, Guido Skipka, Ralf Bender

Properties of Bayesian meta-analyses in evidence syntheses of very few studies

10:20am – 10:40am (CTID 130 on page 114)

Latent Variable Modelling

Chair(s): Steffi Pohl, Martin Elff, Room: HSW30

Francis Tuerlinckx

Data, parameters and models: Some insights in statistical modeling

9:00am – 9:40am (CTID 213 on page 167)

Sina Mews, Roland Langrock, Marius Ötting, Houda Yaqine, Jost Reinecke

Continuous-time latent-state modelling of delinquent behaviour in adolescence and young adulthood

9:40am – 10:00am (CTID 141 on page 120)

Aisouda Hoshiyar, Henk A.L. Kiers, Jan Gertheiss

Penalized Non-Linear Principal Components Analysis for Ordinal Data

10:00am – 10:20am (CTID 89 on page 89)

Jonas Bjermo, Ellinor Fackle Fornius, Frank Miller

Optimal Item Calibration in the Context of the Swedish Scholastic Aptitude Test

10:20am – 10:40am (CTID 23 on page 49)

Open Topics

Chair(s): Sven Knoth, Room: HSN61

Willi Sauerbrei, Tim Haeussler, Marianne Huebner

Structured reporting – low hanging fruit to improve completeness and transparency of analyses in medical and methodological research

9:00am – 9:20am (CTID 178 on page 145)

Jörg Rahnenführer

The role of modern statistical methodology in toxicological research

9:20am – 9:40am (CTID 160 on page 133)

Anne-Laure Boulesteix, Sabine Hoffmann

To adjust or not to adjust: a simple unifying criterion

9:40am – 10:00am (CTID 29 on page 52)

Julia Christin Duda, Jörg Rahnenführer

An intuitive time-dose-response model for cytotoxicity data with varying exposure times

10:00am – 10:20am (CTID 54 on page 68)

Spatial and Spatiotemporal Statistics

Chair(s): Philipp Otto, Room: HSN30

Matthias Katzfuss

Scalable Gaussian-Process Inference Using Vecchia Approximations

9:00am – 9:40am (CTID 100 on page 96)

Malte Jahn, **Christian Weiß**, Hee-Young Kim

Approximately Linear INGARCH Models for Spatio-Temporal Counts

9:40am – 10:00am (CTID 97 on page 94)

Steffen Maletz, Konstantinos Fokianos, Roland Fried

On the Detection of Changes in Spatio-Temporal GLMs for Count Data

10:00am – 10:20am (CTID 131 on page 115)

Timo Budszuhn, **Philipp Doeblner**, Tobias Waldemar Meißner, Sarah Weigelt, Jona Lilienthal, Roland Fried

Independence test for functional variables based on functional canonical correlation

10:20am – 10:40am (CTID 34 on page 55)

Friday, April 1, 11:10 am – 12:30 pm

Plenary IV & Closing

Chair(s): Sven Knoth, Room: HSN55

William H. Woodall

The Evolution of Statistical Process Monitoring

(CTID 234 on page 180)

CONTENTS

1 Geometric inference and robustness

Catherine Aaron

Université Clermont Auvergne, France

When observing an iid sample in \mathbb{R}^D , drawn according to a distribution \mathbb{P}_X supported by a compact d dimensional set S we aim at estimating S (when $d = D$ we will speak of set estimation, when $d < D$ we will speak of manifold estimation) its levels sets $L_t = \overline{\{x, f(x) \geq t\}}$, its possible boundary ∂S and related quantities such as its volume $|S|_d$ or its perimeter $|\partial S|_{d-1}$. A paradoxal point being that geometric inference can be used to detect outliers. It was the point of one of the first papers in the topic : ("Detection of abnormal behaviour via nonparametric estimation of the support" by Luc Devroye and Gary L Wise in 1980). But geometric inference method are, mostly, not robust. The aim of the talk consists in presenting some tools related to this topics, to discuss there (mostly lack of) robustness and, as perspective propose some clues to deal with robustness.

RNS1

2 Testing many restrictions under heteroskedasticity

Stanislav Anatolyev¹, Mikkel Mikkelsen²

1 – CERGE-EI, Czech Republic; 2 – University of Wisconsin-Madison, USA

We propose a hypothesis test that allows for many tested restrictions in a heteroskedastic linear regression model. The test compares the conventional F statistic to a critical value that corrects for many restrictions and conditional heteroskedasticity. The correction utilizes leave-one-out estimation to correctly center the critical value and leave-three-out estimation to appropriately scale it. Large sample properties of the test are established in an asymptotic framework where the number of tested restrictions may be fixed or may grow with the sample size and can even be proportional to the number of observations. We show that the test is asymptotically valid and has non-trivial asymptotic power against the same local alternatives as the exact F test when the latter is valid. Simulations corroborate the relevance of these theoretical findings and suggest excellent size control in moderately small samples also under strong heteroskedasticity.

SHDA1

3 Retrieving grouped LATEs via Classifier-Lasso

Nicolas Apfel¹, Helmut Farbmacher², Martin Huber³, Henrika Langen³

1 – University of Surrey, UK; 2 – TU Munich, Germany; 3 – University of Fribourg, Switzerland

In the context of an endogenous binary treatment with heterogeneous effects and multiple instruments, we propose to apply a classifier-Lasso (C-Lasso, Su, Shi, Phil-

ipps, 2016) and an Agglomerative Hierarchical Clustering (AHC, Ward, 1963) procedure to identify complier groups with identical local average treatment effects (LATE), in spite of relying on distinct instruments. Our procedure is based on the fact that the LATE needs to be homogeneous for any two or multiple instruments that (i) satisfy the LATE assumptions and (ii) generate identical complier groups in terms of treatment probabilities given the respective instruments. Under the assumption that a majority or plurality of instruments with identical complier groups satisfies the LATE assumptions, our procedure permits identifying the valid instruments (in the sense that they satisfy the exclusion restriction) in a data driven way. The choice of tuning parameters is discussed and for this we propose the use of an information criterion. We also provide a simulation study investigating the finite sample properties of our LATE C-Lasso and AHC approach and an empirical application investigating the effect of incarceration on recidivism in the US with judge assignments serving as instruments.

CI

4 Agglomerative Hierarchical Clustering for Selecting Valid Instrumental Variables

Nicolas Apfel¹, Xiaoran Liang²

1 – University of Surrey, UK; 2 – University of Bristol, UK

We propose a novel procedure for Mendelian Randomization (MR) with pleiotropic SNPs which combines agglomerative hierarchical clustering and a test of overidentifying restrictions for selecting valid instrumental variables (IV) from a large set of candidate IVs. Some of these IVs may be invalid in the sense that they fail the exclusion restriction. We show that if the largest group of IVs is valid, our method achieves oracle properties. Unlike existing techniques, our method can deal with weak instruments, multiple endogenous regressors as in Multivariable MR and heterogeneous causal pathways. In simulations, we show that our procedure outperforms the two closest methods, the Hard Thresholding and the Confidence Interval method.

SME3

5 Evaluation of Misspecified Linear Regression Models for Subgroup Analysis

Saide Atmaca

Ulm University of Applied Sciences in Cooperation with Boehringer Ingelheim Pharma GmbH & Co. KG, Germany

In clinical trials, subgroup analyses are routinely performed to assess the treatment effect in certain subpopulations. For linear models with continuous endpoints, the use of linear regression represents a classical approach to determine the influence

of subgroup status on treatment effect. For a valid interpretation of the resulting regression estimates, correct model specification is a fundamental assumption. One source of model misspecifications is that the choice of included variables does not represent the true underlying model, i.e. the omission of relevant (underfitting) or the inclusion of irrelevant (overfitting) variables.

We investigate model misspecification in the context of subgroup analyses by comparing a full population model, including interaction terms for the subgroup status, with a subgroup model that represents a separate analysis within each subgroup category. A full population model that is not fully interacted with the subgroup status may be underfitted compared to the subgroup model, which can cause bias to the parameter estimates (the so-called omitted variable bias). In this work, we assess the estimators in underfitted full population analysis models. Using a simulation study the relationship between the two types of models is examined under different parameter settings. In a second step, an analytical approach is taken to better understand which question the estimators may answer. The simulation results indicate that the absence of an interaction term in the full population model only leads to underfitting when the effect of the variables is different across the subgroup categories. Moreover, the analytical results underpin that by adding interaction terms between the variables and the subgroup status, the estimators consider the observations of the included variables for the respective groups separately. In the absence of the grouping variable or the corresponding interactions, the consideration of observations is expanded across groups.

IBSDRP

6 General-purpose imputation of planned missing data in social surveys: different strategies and their effect on correlations

Julian B. Axenfeld¹, Christian Bruch², Christof Wolf²

1 – University of Mannheim, Germany; 2 – GESIS Leibniz Institute for the Social Sciences, Germany

Planned missing survey data, for example stemming from split questionnaire designs, are becoming more and more common in large-scale social surveys, making imputation indispensable to obtain reasonably analyzable data. This is especially because surveys are facing pressures to shorten questionnaires: Long questionnaires are associated with low response rates, poor response quality, and are particularly considered inappropriate for the increasingly popular online mode. However, these data can be difficult to impute due to common features of social survey data, such as low correlations, predominantly categorical data, and relatively small sample sizes available to support imputation models with many potential predictor variables.

In this presentation, we discuss findings from a Monte Carlo simulation in which we simulate split questionnaire designs, evaluating different imputation methods based on data from the German Internet Panel (GIP). In this simulation, we also experiment with predictor set specifications in which imputation models are restricted exclusively to variables that have correlations to the imputed variable clearly larger than zero. Our results show that strategies that simplify the imputation exercise

(for instance, predictive mean matching procedures with restricted predictor sets) perform well, while some established strategies lead to strong biases.

SM2

7 Posterior Concentration Rates for Bayesian Penalized Splines

Paul Bach, Nadja Klein

Humboldt-Universität zu Berlin, Germany

Despite their widespread use in practice, the asymptotic properties of Bayesian penalized splines have not been investigated so far. We close this gap and study posterior concentration rates for Bayesian penalized splines in a Gaussian nonparametric regression model. A key feature of the approach is the hyperprior on the smoothing variance, which allows for adaptive smoothing in practice but complicates the theoretical analysis considerably. Our main tool for the derivation of posterior concentration rates with a general hyperprior on the smoothing variance is a novel spline estimator that projects the observations onto the first basis functions of a Demmler-Reinsch basis. Our results show that posterior concentration at near optimal rate can be achieved if the hyperprior on the smoothing variance strikes a fine balance between oversmoothing and undersmoothing. Another interesting finding is that the order of the roughness penalty must exactly match the regularity of the unknown regression function in order to achieve posterior concentration at near optimal rate. Overall, our results are the first posterior concentration results for Bayesian penalized splines and can be generalized in many directions.

MS2

8 DoubleML - An Object-Oriented Implementation of Double Machine Learning in R

Philipp Bach¹, Victor Chernozhukov², Malte Kurz¹, Martin Spindler¹

1 - University of Hamburg, Germany; 2 - MIT, USA

The R package DoubleML implements the double/debiased machine learning framework of Chernozhukov et al. (2018). It provides functionalities to estimate parameters in causal models based on machine learning methods. The double machine learning framework consist of three key ingredients: Neyman orthogonality, high-quality machine learning estimation and sample splitting. Estimation of nuisance components can be performed by various state-of-the-art machine learning methods that are available in the mlr3 ecosystem (Lang et al, 2019). DoubleML makes it possible to perform inference in a variety of causal models, including partially linear and interactive regression models and their extensions to instrumental variable estimation. The object-oriented implementation of DoubleML enables a high flexibility for the model specification and makes it easily extendable. This presentation serves as an introduction to the double machine learning framework and the R package DoubleML. In reproducible code examples with simulated and real data sets, we

demonstrate how DoubleML users can perform valid inference based on machine learning methods.

References

[1] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21: C1-C68, <https://doi.org/10.1111/ectj.12097>.

[2] Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L. and Bischl, B. (2019), mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*, <https://doi.org/10.21105/joss.01903>.

CSASOFT2

9 A game-theoretic approach for unsupervised ranking and selection of pathways in gene sets

Chiara Balestra^{1,3}, Carlo Maj², Emmanuel Müller¹, Andreas Mayr³

1 – Technical University of Dortmund, Germany; 2 – Institute for Genomic Statistics and Bioinformatics IGSB, University Hospital Bonn, Germany; 3 – Institut für Medizinische Biometrie, Informatik und Epidemiologie, University Hospital Bonn, Germany

Shapley values provide a way to assign fair values to players within a cooperative game. We interpret sets within families of sets as players of a game and extend the application of game theory to unsupervised rankings of pathways in gene sets; each pathway, i.e., each set, has a value dependant on the distribution of its elements within the family of sets. However, the Shapley values do not consider possible overlaps among sets and possibly rank overlapping sets in similar positions.

We propose Shapley values-based feature importance scores aware of redundancy for sets. Inspired by the necessity of reducing the overlap among pathways while still retaining a high coverage of the gene sets, our method assigns an importance score to each set. We apply the obtained ranking to reduce the dimension of families of sets. Our case study for pathways and gene sets shows that the obtained selection of pathways contains lower redundancy while still keeping a high coverage of the genes when compared to naive Shapley values. Moreover, a naive application of Shapley values leads to a positive correlation of the position in the ranking and the pathways' size; with our method, instead, the aforementioned correlation is reduced.

Furthermore, we paired the Shapley values-based feature importance scores with gene set enrichment analysis where multiple hypothesis testing for association with phenotypic traits is an issue. We use our ranking for unsupervised pre-selection of pathways and study the effect on the number of significant pathways found for the single association traits.

BSB2

10 Calibration of wind speed ensemble forecasts for power generation

Ágnes Baran, Sándor Baran

University of Debrecen, Hungary

In 2020, 36.6% of the total electricity demand of the world was covered by renewable sources, whereas in the EU (UK included) this share reached 49.3%. A substantial part of green energy is produced by wind farms, where accurate short range power predictions are required for successful integration of wind energy into the electrical grid. Accurate predictions of the produced electricity require accurate forecasts of the corresponding weather quantity, where the state-of-the-art method is the probabilistic approach based on ensemble forecasts. However, ensemble forecasts are often uncalibrated and might also be biased, thus require some form of post-processing to improve their predictive performance.

To calibrate (hub height) wind speed ensemble forecasts we propose a novel flexible machine learning approach, which results either in a truncated normal or a log-normal predictive distribution [1]. In a case study based on 100m wind speed forecasts of the operational AROME-EPS of the Hungarian Meteorological Service, the forecast skill of this method is compared with the predictive performance of three different ensemble model output statistics approaches and the raw ensemble predictions. We show that compared with the raw ensemble, post-processing always improves the calibration of probabilistic and accuracy of point forecasts, and from the five competing methods the novel machine learning based approaches result in the best overall performance.

Reference

[1] Baran, S., Baran, Á., Calibration of wind speed ensemble forecasts for power generation. *Időjárás* 125 (2021), 609–624.

SAES2

11 APCtools: An R Package for Descriptive and Model-based Age-Period-Cohort Analysis

Alexander Bauer¹, Maximilian Weigert¹, Hawre Jalal²

1 – Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany; 2 – Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, USA

Age-Period-Cohort (APC) analysis aims to differentiate relevant drivers for long-term developments and is used in many fields of science. The critical challenge in APC analysis is that the main components age, period and cohort are linearly dependent (e.g., cohort = period – age). Accordingly, flexible methods and visualization techniques are needed to properly disentangle observed temporal association structures. We introduce APCtools, the first sophisticated R package that covers all aspects of APC analysis. The package comprises functions both for descriptive

and regression model-based analyses. For the former, we use density (or ridgeline) matrices and adapted heatmaps as innovative visualization techniques that build on the concept of Lexis diagrams. Since Lexis diagrams and derived visualizations share the problem of visually underrepresented diagonals we offer hexagonally binned heatmaps (hexamaps) as an alternative in which all three axes have similar visual weight. Model-based analyses build on the separation of the temporal dimensions based on generalized additive models and can be used to simultaneously account for further control variables. The main temporal effect is estimated based on a tensor product interaction surface between two dimensions, representing the third dimension on its diagonal. Additional to heatmap and hexamap visualizations of the resulting tensor product estimate, partial APC plots allow for detailed insights in the dependency structure among the temporal dimensions. We showcase the functionality of the package by analyzing how the travel behavior of German travelers changed over the last decades.

CSASOFT1

12 Quantile-based MANOVA: A new tool for inferring multivariate data in factorial designs

Marléne Baumeister, Marc Ditzhaus, Markus Pauly

TU Dortmund University, Germany

Multivariate Analysis of Variance (MANOVA) allows for the joint comparison of multiple responses and is commonly used in fields such as medicine, ecology or psychology. However, the classical MANOVA relies in assumptions, e.g. normality or homogeneous covariance matrices, which are often difficult to justify. To overcome these difficulties there are less restrictive mean-based MANOVA concepts proposed for testing global hypothesis about multivariate expectations, e.g. (Friedrich and Pauly, 2018). In case of outliers or distributions with larger tails, however, non-robust estimators like the mean can have some drawbacks. Despite the usage of quantiles is intuitive in that case and often applied in descriptive statistics, e.g. boxplots, quantiles "[appear] to be quite underused in medical research" (Beyerlein, 2014). Therefore we developed a flexible quantile-based MANOVA method. The approach is adaptable to general factorial designs and has the advantage that it fits to median and other quantile-based statistical methods. To achieve this, we considered two quadratic-form type test statistics and three different strategies for estimating the covariance. The test statistics' distribution is approximated via resampling. We prove by empirical processes that our method is valid in theory and even works in case of general heterogeneous or heteroscedastic data beyond normality. In a simulation study, we compare the novel procedures with state-of-the-art mean-based approaches.

References

- [1] Beyerlein, A. (2014). Quantile Regression—Opportunities and Challenges From a User's Perspective. *American Journal of Epidemiology*, 180(3):330–331.
- [2] Friedrich, S. and Pauly, M. (2018). MATS: Inference for potentially singular and

13 mlr3tuning: A general framework for ML hyperparameter tuning

Marc Becker¹, Michel Lang², Bernd Bischl¹

1 – Ludwig-Maximilians-Universität München, Germany; 2 – Technische Universität Dortmund, Germany

The performance of modern ML algorithms depends on their hyperparameters. While R offers packages to solve general optimization problems, it lacks a common interface for hyperparameter tuning.

mlr3tuning offers exactly that and already includes these tuners: random search, grid search, iterated racing, GenSA and nloptr. Packages mlr3hyperband and mlr3mbo extend this list with multifidelity approaches and Bayesian optimization. Furthermore, mlr3tuning offers a flexible system for termination, a generic archive for logging of evaluations and the capacity for nested resampling. Search space parameters can be of different types, including mixed categorical-numerical spaces, including hierarchical dependencies. These can be optimized on arbitrary scales via transformations. This is further simplified by mlr3tuningspaces which offers a extensible collection of common search spaces for learners.

An important aspect of mlr3tuning is its modularity integration into the mlr3 ecosystem (Lang et al., 2019). Learners can be tuned for a large number of metrics and resampling methods and the tuning process can be parallelized via different backends. Furthermore, arbitrary ML pipelines from mlr3pipelines (Binder et al., 2021) can be jointly optimized to handle the combined algorithm selection and hyperparameter optimization (CASH) problem.

References

- [1] Binder et al. (2021). mlr3pipelines-Flexible Machine Learning Pipelines in R. Journal of Machine Learning Research, 22(184), 1–7.
- [2] Lang et al. (2019). mlr3: A modern object-oriented machine learning framework in R. Journal of Open Source Software, 4(44), 1903.

CSASOFT2

14 Firth correction in Cox proportional hazard analysis in the presence of zero events

Lars Beckmann, Anke Schulz, Guido Skipk

IQWiG, Germany

For the early benefit assessment of drugs in Germany, the pharmaceutical company

must describe the extent of an added benefit of the drug to be assessed compared with an appropriate comparator therapy [1]. The confidence interval of a significant effect must lie completely outside a certain corridor around the null effect for the extent of the effect to be regarded as minor, considerable or major. The corridors are defined by different thresholds depending on outcome category. For time to event analyses, the 95% confidence interval for the hazard ratio is required to determine the extent.

If no events are observed in one of the study arms, the standard Cox proportional hazard regression does not provide effect estimates with corresponding confidence intervals, while the log rank test provides appropriate p-values. Thus, in the case of a statistically significant effect, the extent cannot be determined and the overall assessment of the added benefit might be hampered.

Heinze and Schemper proposed an adaption of the Firth correction to reduce bias from maximum likelihood estimation for the Cox proportional hazard [2]. To assess the applicability of this approach, we performed a simulation study of time to event analyses with zero events. We will present results from this study and discuss the situations, in which the application of the Firth correction provides reliable estimates that can be used for the assessment of the extent of an added benefit.

References

[1] IQWiG. General Methods 6.0 [online]. 2020. URL: <https://www.iqwig.de/en/about-us/methods/methods-paper/>

[2] Heinze and Schemper (2001). *Biometrics* 57(1):114–119.

SEHA2

15 A Test of Independence for Locally Stationary Processes Using a Weighted Characteristic Function-based Distance

Carina Beering

Helmut Schmidt University Hamburg, Germany

We propose a testing procedure for independence of locally stationary processes based on a weighted distance composed of characteristic functions (CF) and its empirical version. The essential idea of this concept is inspired by the distance covariance defined by Székely et al. (2007) and was taken up by Jentsch et al. (2020). For the purpose of compiling a testing procedure at the end, we provide the needed results with the notion of the beneficial effects of a bootstrap analogue. Therefore, we establish the bootstrap versions of the previously presented findings. Beforehand, we transfer the concept of empirical weighted CF-distance to the bootstrap world. Finally, a simulation study is performed using our testing procedure to detect dependence of different forms.

References

[1] Jentsch, C., Leucht, A., Meyer, M. and Beering, C. (2020). Empirical characteristic functions-based estimation and distance correlation for locally stationary

processes. *Journal of Time Series Analysis* 41, 110-133.

[2] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35, 2769-2794.

TSA3

16 Modeling trajectories of slowly progressing diseases: A mixed-model-based algorithm for variable transformation, prediction and age-of-onset estimation

Charlotte C. Behning¹, Monika Fleckenstein², Maximilian Pfau³, Christine Adrion⁴, Lukas Goerdts⁵, Moritz Lindner⁵, Steffen Schmitz-Valckenberg², Frank G. Holz⁵, Matthias C. Schmid¹

1 – Department of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Germany; 2 – John A. Moran Eye Center, University of Utah, Salt Lake City, USA; 3 – Ophthalmic Genetics and Visual Function Branch, National Eye Institute, Bethesda, MD, USA; 4 – Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany; 5 – Department of Ophthalmology, University Hospital Bonn, Bonn, Germany

Modeling trajectories of slowly progressing diseases is essential to understand the course of the disease, to identify associated risk factors and to design clinical (interventional) trials. Determining the age of disease also remains a largely unsolved problem. Here, we aim to model the progression of geographic atrophy (GA) in patients with age-related macular degeneration (AMD) by building a suitable statistical regression model. Based on theoretical considerations, we develop a linear mixed-effects model for GA size progression that incorporates covariable-dependent enlargement rates as well as correlations between longitudinally collected GA size measurements. Instead of simply including a patient specific intercept and slope, we directly incorporate the (risk-factor dependent) age of disease at study entry into our model. This way, we account for the fact, that patients enter a clinical study at different stages in their disease history. To capture nonlinear progression in a flexible way, we systematically assess Box-Cox transformations with different transformation parameters λ . The selection of an optimal transformation parameter is based on Akaike information criterion (AIC). We evaluate our model using data collected for two longitudinal, prospective, multicenter cohort studies of GA size progression that included 150 eyes from 101 patients. The model allows for age-of-onset estimation, identification of risk factors and prediction of future GA size. Based on the assessment of the transformation approach, we recommend a square-root transformation of atrophy size before model fitting.

Reference

[1] <https://doi.org/10.1186/s12874-021-01356-0>

ARM3

17 Understanding Ability Differences Measured with Count Items and their Reliability: The Distributional Regression Test Model and the Count Latent Regression Model

Marie Beisemann, Philipp Doebler

TU Dortmund University, Germany

In psychology and education, various tasks and tests lead to observations of counts (e.g., error counts in reading assessment). The first Item Response Theory (IRT) model for count data is Rasch's Poisson Counts Model (RPCM), a log-linear model which assumes Poisson distributed item scores depending on the item's difficulty and conditional on a test taker's latent ability. Recent work has proposed generalizations of the RPCM which allow for loosening the Poisson distribution's equidispersion assumption, with necessary advantages for model implied reliability. Proposed models use the Conway-Maxwell-Poisson (CMP) distribution in its mean parameterization by Huang. The two-parameter CMP model (2PCMPM), a generalization of the first proposed CMP Counts model (CMPCM) and the RPCM, allows researchers to estimate item-specific difficulty (resp. easiness), discrimination (resp. factor loading), and dispersion parameters. When studying tests' properties, it is often not only of interest to estimate item and person parameters, but also to be able to explain differences between them. With regard to item parameters, explaining them also allows for explaining differences in item information, i.e., item reliability, which is important for item selection. This work introduces two explanatory count IRT models as extensions of the 2PCMPM: The Distributional Regression Test Model (DRTM) which allows for inclusion of item covariates, and the Count Latent Regression Model (CLRM) which models person covariates. These models can be understood as a combination of univariate factor analysis and distributional regression. Estimation methods are provided and statistical properties assessed in simulation studies.

SBES

18 Influence of the Design of Open-Ended Survey Questions Using Conversational Agents or Voice-Based Input on the Responses

Florian Berens, Sebastian Hobert

University of Göttingen, Germany

Online surveys allow large groups of people to be surveyed easily and cost-effectively. In addition to closed questions, such standardized surveys often use open-ended questions to give respondents room to answer more freely. The open-ended questions are usually presented as free text fields into which respondents type their answers using their keyboard. However, experience in survey research shows that many respondents shy away from this effort and that the amount of item nonresponse is high and the data quality of the answers given is rather poor. Therefore, innovative

approaches are needed that reduce the effort for respondents or create incentives to motivate respondents to give more and better answers.

In a survey experiment with a 2×2 design, first-year students were asked about their expectations for university and learning. Questions were randomly assigned either in a traditional, static format or in the form of a dialogue with a conversational agent. Depending on a second randomization, respondents could either type the answer into a text box using their keyboard or verbally report the answer using a microphone.

The analysis of the answers shows that verbally given answers are longer and also have a higher answer quality in other parameters. The same applies to answers that were requested in chat format. At the same time, however, it is also evident that verbal responses in particular significantly increase the item nonresponse rate. By contrast, interaction effects of technology affinity with the aforementioned effects are hardly confirmed.

SM1

19 Why Are the Motivated Successful in Our Classes: A Learning Analytics Study on the Effects of Attitudes Toward Statistics on Learning Behavior

Florian Berens, Sebastian Hobert

University of Göttingen, Germany

Negative attitudes toward introductory courses in statistics have multiple negative consequences. Students with negative attitudes at the beginning of a course score lower on exams in the course initially and are less willing to take more in-depth courses later on or to use statistics in their careers or everyday lives. While a correlation between attitudes and course success seems plausible, the explanation of this relationship has been poorly explored.

For this study, students of an introductory statistics course for social scientists were surveyed in detail about their attitudes toward statistics before the course began. During the course, students learned using a very extensive digital learning platform. From the usage data of this learning platform, learning behavior can be represented in many dimensions. Analyses of the correlations between attitudes and learning behavior show that students with positive attitudes not only study more quantitatively, but also learn differently than students with more negative attitudes. Here, for example, the learning rhythm can be referred to, which differs depending on the attitude. The study thus shows that positive attitudes not only have an influence on learning success through the pure volume of studying, but also have a more profound effect on the learning behavior of students.

SLSE2

20 Inference on jumps in high-frequency order-price models with one-sided noise

Markus Bibinger

Julius-Maximilians-Universität Würzburg, Germany

For high-frequency intra-day price processes it is well-known that market microstructure dilutes the underlying dynamics of a semi-martingale model. Different to the classical model for traded prices with additive, centred market microstructure noise, we consider a stochastic boundary model with one-sided noise for best-ask prices from the limit order book. We construct methods to estimate, locate and test for jumps in this model. The different structure of irregular noise leads us to statistics based on local minima instead of local averages as used for regular noise in the literature. We first discuss inference on jumps at some given (stopping) time. We provide a local test and show that we can consistently estimate price jumps. We develop a global test for jumps based on extreme value theory. We establish the asymptotic distribution of a maximum statistic on which the test is based on under the null hypothesis and consistency under the alternative hypothesis. The rate of convergence for local alternatives is determined and faster than the optimal rate in the regular noise model what allows the identification of smaller jumps based on discrete observations. In the process, we establish pointwise central limit theorems for spot volatility estimation as well as uniform consistency for volatility estimation.

SiF1

21 Similarity of competing risks models with constant intensities in an application to clinical healthcare pathways involving prostate cancer surgery

Nadine Binder¹, Kathrin Möllenhoff², August Sigle³, Holger Dette⁴

1 – Institute for General Practice/Primary Care, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany; 2 – Mathematical Institute, Heinrich-Heine University, Düsseldorf, Germany; 3 – Department of Urology, Faculty of Medicine, Medical Center, University of Freiburg, Freiburg, Germany; 4 – Department of Mathematics, Ruhr University Bochum, Bochum, Germany

The recent availability of routine medical data, especially in a university-clinical context, may enable the discovery of typical healthcare pathways by assessing their similarity for ultimately improving the standard of care. We understand the pathway as a temporal process with possible transitions from a single initial treatment state to hospital readmission of different types, which constitutes a competing risk setting. In this talk, we propose a multi-state model-based approach to uncover pathway similarity between two groups of individuals. Precisely, we derive a hypothesis test based on a constrained parametric bootstrap for assessing the similarity of two competing risk models assuming constant transition intensities. Our methodological approach is motivated by an application to hospital readmission for competing

reasons after prostate cancer surgery. As the readmission intensities are low, the question is whether they are sufficiently similar for patients with prior in-house diagnostics versus without prior in-house diagnostics such that the two populations can be combined for further outcome analyses. Our suggested approach allows to identify thresholds for which the transition intensities are to be considered similar, enabling patient groups to be pooled for further investigations.

SEHA4

22 Reevaluating dementia incidence trends: The critical role of adequate design and methodology

Nadine Binder¹, Anika Schlosser², Martin Schumacher²

1 – Institute for General Practice/Primary Care, Medical Center – University of Freiburg, Germany; 2 – Institute of Medical Biometry and Statistics, Medical Center – University of Freiburg, Germany

A seeming decline in dementia incidence in Western nations has been a topic of continuous debate resulting in a recently published analysis of data from seven population-based cohort studies (Wolters et al., 2020). Constructing several non-overlapping 5-year epochs, the corresponding design and analysis closely follows a framework previously used within the Framingham Heart Study (FHS) cohort. However, we challenged the finding of the FHS cohort on the basis that bias may have resulted from the failure to adequately account for potential disease onset in the period between last observation and death. Re-analyzing the FHS data using spline-based analytic methods, we did not find convincing evidence for a decline in dementia incidence over the epochs (Binder et al., 2019). Yet, there is further room for improvement.

As a more suitable approach for analyzing the question how dementia incidence has evolved over time, we divide the FHS cohort into birth cohorts (decades of birth years), consider age as natural time scale, and estimate transition probabilities and age-conditional probabilities. Still, we have to deal with cases of missing disease information due to death and, moreover, lost-to-follow-up. This requires the use of statistical methods based on the illness-death multi-state model, such as spline-based penalized likelihood as employed in our earlier study. We will present the findings of the proposed design and analysis strategy, aiming for a realistic quantification of the dementia incidence trend in the Framingham Heart Study cohort.

References

- [1] Wolters FJ et al. *Neurology*. 2020;95(5):e519-e531.
- [2] Binder N et al. *Eur J Epidemiol*. 2019;34(11):1075-1083.

SME1

23 Optimal Item Calibration in the Context of the Swedish Scholastic Aptitude Test

Jonas Bjeremo, Ellinor Fackle Fornius, Frank Miller

Stockholm University, Sweden

Large-scale achievement tests require the existence of item banks with items for use in future tests. Before an item is included in the bank, its parameters need to be estimated as efficiently as possible. The precision of the item parameters depends on the abilities of the examinees, which is not known and must be estimated beforehand.

Methods of optimal design have been developed to allocate calibration items to examinees with the most suited ability. In this presentation, we investigate the performance of an optimal ability-dependent allocation in the context of the Swedish Scholastic Aptitude Test (SweSAT) and quantify the gain from using the optimal allocation. We apply a method based on an optimal design algorithm, that shows advantages in theory. The algorithm utilizes a so-called optimal restricted design, using the D-optimality criterion. The computations involved in the optimal design algorithm use some approximations, and the optimal design needs to be evaluated in a real testing situation.

We performed a simulation study to explore the optimal design strategy in a realistic setting and compare the results to a random design strategy, where items are instead randomly allocated to examinees. The evaluation measures used in the simulation study are developed carefully to mimic the measures used in optimal design theory. We have quantified the gain from using the optimal allocation and, on average over all items, we see an improved precision of calibration. We are also able to identify for what kind of items the method works well.

LVM

24 Conditional Feature Importance for Mixed Data

Kristin Blesch^{1,2}, David S. Watson³, Marvin N. Wright^{1,2,4}

1 – Leibniz BIPS, Germany; 2 – Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany; 3 – Department of Statistical Science, University College London, London, UK; 4 – Department of Public Health, University of Copenhagen. Copenhagen, Denmark

Despite the popularity of feature importance measures in interpretable machine learning, the statistical adequacy of these methods is rarely discussed. From a statistical perspective, a major distinction is between analyzing a variable's importance before and after adjusting for covariates – i.e., between marginal and conditional measures. For conditional feature importance (CFI), only few methods are available and practitioners have hitherto been severely restricted in method application due to mismatching data requirements. Most real-world data exhibits complex feature dependencies and incorporates both continuous and categorical data (mixed data). However, most CFI measures are designed for just a single type of data, or make

simplifying assumptions about feature dependencies.

To fill this gap, we propose to combine the conditional predictive impact (CPI) framework [1] with sequential knockoff sampling [2]. The CPI enables CFI measurement that controls for any feature dependencies by sampling valid knockoffs – hence, generating synthetic data with similar correlation structure – for the data to be analyzed. Sequential knockoffs were deliberately designed to handle mixed data and thus allow us to extend the CPI approach to such datasets.

We demonstrate through numerous simulations that our proposed workflow controls type I error and achieves high power, while other marginal and CFI metrics result in misleading interpretations. Our findings highlight the necessity of developing specialized methods for mixed data such that interpretable machine learning methods behave as intended on real-world data.

References

[1] Watson & Wright (2021), *Mach Learn* 110:2107-2129.

[2] Kormaksson et al. (2021), *Stat Med* 40:3313-3328.

YSS1

25 Statistical Models for Partial Orders based on Data Depth and Formal Concept Analysis

Hannah Blocher, Georg Schollmeyer, Christoph Jansen

Ludwig-Maximilians-Universität München, Germany

We introduce new and generalize already existing statistical models for partial orders whose incomparable part cannot be understood as induced by a missing process. Most existing statistical models for partial orders are based on the assumption that incomparability follows from unobserved data and, thus, rely on the idea of relaxing classical models for total orders. For this reason, we first discuss the conceptual differences between statistical models for total rankings versus partial rankings. In particular, we address distance-based models. Afterwards, we demonstrate that by using the notion of data depth, we can introduce new or generalize already existing statistical models for partial orders that avoid relying on the assumption of stemming from some missing process. For this, we use the rich vocabulary of formal concept analysis by representing the set of partial orders by a formal context. After a concise definition of unimodality and unimodal statistical models for partial orders, the proposed statistical models are analysed and discussed in terms of unimodality. Therefore, we present and generalize several depth functions and introduce different algorithms for efficiently sampling from unimodal models. Finally, we demonstrate the proposed statistical models in a concrete data situation.

RNS3

26 Evaluating (weighted) dynamic treatment effects by double machine learning

Hugo Bodory¹, Martin Huber², Lukas Laffers³

1 – University of St Gallen, Switzerland; 2 – University of Fribourg, Switzerland;
3 – Matej Bel University, Slovakia

We consider evaluating the causal effects of dynamic treatments, i.e. of multiple treatment sequences in various periods, based on double machine learning to control for observed, time-varying covariates in a data-driven way under a selection-on-observables assumption. To this end, we make use of so-called Neyman-orthogonal score functions, which imply the robustness of treatment effect estimation to moderate (local) misspecifications of the dynamic outcome and treatment models. This robustness property permits approximating outcome and treatment models by double machine learning even under high dimensional covariates. In addition to effect estimation for the total population, we consider weighted estimation that permits assessing dynamic treatment effects in specific subgroups, e.g. among those treated in the first treatment period. We demonstrate that the estimators are asymptotically normal and root-n-consistent under specific regularity conditions and investigate their finite sample properties in a simulation study. Finally, we apply the methods to the Job Corps study.

EEA1

27 Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures

Anine E. Bolko¹, Kim Christensen¹, Mikko S. Pakkanen²,
Bezirgen Veliyev¹

1 – Aarhus University, Denmark; 2 – Imperial College London, United Kingdom

In this paper, we develop a generalized method of moments approach for joint estimation of the parameters of a log-normal stochastic volatility model driven by a fractional Brownian motion with unrestricted Hurst exponent. We show that an estimator based on integrated variance is consistent. Moreover, under stronger conditions we derive a central limit theorem. The results continue to hold even when integrated variance is replaced with a noisy measure of volatility calculated from discrete high-frequency data. However, in practice a realized estimator contains sampling error, the effect of which is to skew the fractal coefficient toward “roughness”. We construct an analytical approach to control this error. In a simulation study, we demonstrate convincing small sample properties of our approach based both on integrated and realized variance over the entire memory spectrum. We show the bias correction attenuates any systematic deviance in the estimated parameters. Our procedure is applied to empirical high-frequency data from numerous leading equity indexes. With our robust approach the Hurst index is estimated around 0.05,

confirming roughness in the volatility process.

SiF1

28 Consistent Estimation of Multiple Breakpoints in Dependence Measures

Marvin Borsch, Dominik Wied, Alexander Mayer

Institute of Econometrics and Statistics, University of Cologne, Germany

We extend the fluctuation test by Manner et al. [2019] to detect multiple breaks in general dependence measures. We will use their moment based statistic to detect breaks in the dependence measures, mainly focusing on Spearman's rho. We will observe consistency of the estimates for the location of the breakpoints as well as the number of breaks. We will use a factor copula model as our leading model as it is useful for analyzing data in high dimensions. In order to find multiple breaks Binary Segmentation (BS) as well as a more recent approach the Wild Binary Segmentation (WBS) introduced by Fryzlewicz [2014] is used. We will see in general better results in our Monte Carlo simulation for multiple breaks using WBS. We will also analyze different locations in the breaks for both algorithms. We find that our method works as well when we first filter the conditional variances in the marginals. Additionally we propose an Information Criterion incorporating WBS, which is shown to be consistent for the number of breaks and its locations. Monte Carlo simulations considering size and power of this procedure and a real data application will be presented.

TSA1

29 To adjust or not to adjust: a simple unifying criterion

Anne-Laure Boulesteix, Sabine Hoffmann

LMU Munich, Germany

The question as to whether to adjust or not for multiple testing has been extensively discussed in the last decades in various scientific fields including health sciences. Here we use the term multiple testing for the sake of simplicity, but our arguments equally apply to adjustments performed in the context of other types of analyses, for example the adjustment of confidence intervals in case of multiple comparisons. While much of the recent advances and discussions on multiple testing has focused on complex issues such as high-dimensional data, adaptive designs and the so-called "p-hacking" problem, we note that even in more classical settings (typically: up to a few dozens of tests corresponding to different treatments, outcomes, biomarkers, time points, etc.) the question of whether and how to adjust or not is not yet definitively settled. In some situations the community agrees that adjustment is required, while in other situations the necessity of adjustment is debated. We suggest a unique criterion which has the advantages that 1) many previously proposed "rules" are special cases of it, 2) it is easily understandable for applied researchers

thus enabling efficient interactions between statisticians and applicants regarding the decision to adjust or not, 3) it is not a complex set of rules, but a single criterion which is valid in various situations. Its particularity is that it relates the decision to adjust or not to the discrepancy between conducted analyses and the way they are reported.

OT

30 Collaborative real-time nowcasting of COVID-19 hospitalization incidence rates

Johannes Bracher¹, **Daniel Wolfram**¹, **Sam Abbott**², **Matthias an der Heiden**³, **Sebastian Funk**², **Felix Günther**⁴, **Davide Hailer**¹, **Stefan Heyder**⁵, **Thomas Hotz**⁵, **Helmut Küchenhoff**⁶, **Diella Syliqi**⁶, **Jan van de Kassteel**⁷, **Maximilian Weigert**⁶, **Alexander Ullrich**³, **Tilmann Gneiting**⁸, **Melanie Schienle**¹

1 – Karlsruhe Institute of Technology, Germany; 2 – London School of Hygiene and Tropical Medicine, United Kingdom; 3 – Robert Koch Institute, Berlin, Germany; 4 – Stockholm University, Sweden; 5 – Technische Universität Ilmenau, Germany; 6 – Ludwig Maximilian University of Munich, Germany; 7 – Dutch National Institute for Public Health and the Environment, Bilthoven, the Netherlands; 8 – Heidelberg Institute for Theoretical Studies, Germany

The seven-day hospitalization incidence is one of the main indicators used to assess the pandemic situation in Germany. It is defined as the number of COVID-19 cases reported over a seven-day period and subsequently hospitalized. Due to delays between the reporting of a case and a possible hospitalization, and between the hospitalization date and appearance in the surveillance data, the most recent values of this indicator are typically incomplete and need to be corrected upwards over the following days and weeks. As a consequence, recent trends cannot be read directly from the raw values of the seven-day hospitalization incidence. Statistical nowcasting methods can be used to adjust these, and obtain a more realistic picture of recent tendencies. We report on a collaborative nowcasting platform called the German COVID-19 Nowcast Hub (<https://covid19nowcasthub.de/>) which unites probabilistic nowcasts of the seven-day hospitalization incidence from eight distinct models run independently by different teams of researchers. Nowcasts are collected on a daily basis and combined into an ensemble, which represents the main output of the project. The different methods and ensembles will be compared systematically in a pre-registered evaluation study (running from November 2021 through April 2021). We will discuss preliminary evaluation results and the potential and pitfalls of collaborative real-time analyses during the pandemic.

SME2

31 An Automatic Finite-Sample Robustness Metric: Can Dropping a Little Data Change Conclusions?

Tamara Broderick

MIT, United States of America

One hopes that data analyses will be used to make beneficial decisions regarding people's health, finances, and well-being. But the data fed to an analysis may systematically differ from the data where these decisions are ultimately applied. For instance, suppose we analyze data in one country and conclude that microcredit is effective at alleviating poverty; based on this analysis, we decide to distribute microcredit in other locations and in future years. We might then ask: can we trust our conclusion to apply under new conditions? If we found that a very small percentage of the original data was instrumental in determining the original conclusion, we might expect the conclusion to be unstable under new conditions. So we propose a method to assess the sensitivity of data analyses to the removal of a very small fraction of the data set. Analyzing all possible data subsets of a certain size is computationally prohibitive, so we provide an approximation. We call our resulting method the Approximate Maximum Influence Perturbation. Our approximation is automatically computable, theoretically supported, and works for common estimators — including (but not limited to) OLS, IV, GMM, MLE, MAP, and variational Bayes. We show that any non-robustness our metric finds is conclusive. Empirics demonstrate that while some applications are robust, in others the sign of a treatment effect can be changed by dropping less than 0.1% of the data — even in simple models and even when standard errors are small.

P2

32 Dependent Wild Bootstrap based Test for Independence under Local Stationarity

Guy-Niklas Brunotte

University of Bamberg, Germany

In the last decades, several tests for independence have been established under various assumptions. However, most of the known independence tests assume that the underlying random variables are i. i. d. or consider stationary processes. These assumptions allow the evolution of an extensive theory but are often not fulfilled in practice. Generalizing stationary processes to locally stationary processes allows wider applications. These processes are non-stationary processes which are characterized by the fact that for each point in time exists a stationary process that approximates the locally stationary process at this point in time.

The presented talk introduces a consistent level α test which examines whether two locally stationary processes are asymptotically independent at the same point in time. This test belongs to a wide class of adapted locally stationary processes which includes, inter alia, many time-varying random coefficient models and time-varying ARCH-processes. The test procedure will be obtained by extending the dependent

wild bootstrap approach proposed by Shao (2010) to a locally stationary situation and the given simulation study will show a good performance of this test under the null hypothesis and the alternative.

TSA3

33 Analyzing the Influence of Missing Data Imputation on the Predictive Performance of Classifiers

Philip Buczak, Markus Pauly

Technische Universität Dortmund, Germany

In biostatistics, missing data is a reoccurring challenge. Patients' measurements cannot always be observed due to random or deterministic factors. As missing data can contain valuable information, omitting it may lead to information loss. Thus, missing data is often imputed through different techniques such as simple (mean, mode or LOCF) or advanced imputation methods such as MICE (multiple imputation by chained equation) or random forest based imputation.

However, data imputation may also affect the quality of prediction or inference from resulting models. It is therefore crucial to analyze the extent to which imputation methods influence subsequent regression or classification models obtained from imputed data. In the present work we focus on classification. Classification problems frequently occur in biostatistics, e.g. in the context of diagnosing the presence of certain conditions or diseases. Using common imputation algorithms, we study their impact under increasing missing rates on the predictive performance of widely used classifications algorithms such as random forests, gradient boosting, and (penalized) logistic regression. We evaluate the performance of the respective classification models in terms of the mean misclassification error and the area under the curve (AUC).

AIML2

34 Independence test for functional variables based on functional canonical correlation

Timo Budszuhn, Philipp Doeblér, Tobias Waldemar Meißner, Sarah Weigelt, Jona Lilienthal, Roland Fried

Technische Universität Dortmund, Germany

Data are functional if observations can be naturally understood as mappings. Functional domains are often time periods or spatial coordinates. Functional data arise when aggregation does not do justice to the features, especially when integration over a spatial or temporal direction would lead to a significant loss of information. The association of two functional variables can be examined with the help of functional canonical correlation analysis (FCCA; Leurgans et al., 1993). Despite the age of the method, hypothesis tests for functional correlations have hardly been studied

so far. We compare two approaches to permutation testing of functional variable independence, and both are intended for situations with missing data: The first approach uses local polynomial kernel smoothing to preprocess (noisy) functional data and proceeds by classic canonical correlation analysis (CCA) on equidistant points. This requires the tuning of smoothing parameters and the number of equidistant points. In the second approach, the functional observations are projected on the eigenfunctions of their autocovariance operator, followed by a CCA of the eigenbasis coefficients. Simulation experiments compare the two approaches, including comparisons with Dubin and Müller's (2005) dynamic correlation. The independence tests are applied to study the quality of corrections for head motion artifacts in fMRI experiments using the association of spatial displacement and strength of the BOLD signal. A second application studies the association of the number of COVID-19 infections and excess mortality in the last two years in European countries.

STS

35 Estimation of domain specific consumer baskets in Germany

Jan Pablo Burgard

Trier University, Germany

For inflation measurement, two core components are needed. On the one hand, the change in prices, and on the other hand, the consumer basket. Taking the mean basket overall households in Germany, will lead to an inflation measure, that does not necessarily reflect individual inflation experience. One example is the impact of a rise in food prices on the total spending of high- and low-income households. In this talk, new advances in compositional small area estimation will be used to estimate domain-specific consumer baskets, which in their nature are a simplex. The domains comprise different household incomes and education levels. From these estimated domain-specific consumer baskets, domain-specific inflation rates will be deduced. The mean squared error is approximated using a parametric bootstrap.

OAS1

36 The Sparse Polynomial Chaos Expansion: A Fully Bayesian Approach with Joint Priors on the Coefficients and Global Selection of Terms

Paul Bürkner, Sergey Oladyshkin, Ilja Kröker, Wolfgang Nowak

University of Stuttgart, Germany

Polynomial chaos expansion (PCE) is a versatile tool widely used in uncertainty quantification and machine learning, but its successful application depends strongly on the accuracy and reliability of the resulting PCE-based response surface. High accuracy typically requires high polynomial orders, demanding many training points especially in high-dimensional problems through the curse of dimensionality.

So-called sparse PCE concepts work with a much smaller selection of basis polynomials compared to conventional PCE approaches and can overcome the curse of dimensionality very efficiently, but have to pay specific attention to their strategies of choosing training points. Furthermore, the approximation error resembles an uncertainty that most existing PCE-based methods do not estimate. In this study, we develop and evaluate a fully Bayesian approach to establish the PCE representation via joint shrinkage priors and MCMC. The suggested Bayesian PCE model directly aims to solve the two challenges named above: achieving a sparse PCE representation and estimating uncertainty of the PCE itself. The embedded Bayesian regularizing via the joint shrinkage prior allows using higher polynomial degrees for given training points due to its ability to handle underdetermined situations, where the number considered PCE coefficients could be much larger than the number of available training points. We also explore multiple variable selection methods to construct sparse PCE expansions based on the established Bayesian representations, while globally selecting the most meaningful orthonormal polynomials given the available training data. We demonstrate the advantages of the our Bayesian PCE and the corresponding sparsity-inducing methods on several benchmarks.

BS2

37 Parametric Network Models for Sets of Relations

Carter Tribley Butts

UC Irvine, United States of America

Parametric models for networks with heterogeneity and/or complex dependence have seen considerable progress over the past two decades, opening the door to further modeling opportunities. Among these is the modeling of sets of networks drawn either from a single common generating process (i.e., replication), or from mixtures of such processes (as is frequently the case in observational studies of populations of networks). The multiple-network setting poses a number of computational and statistical challenges, particularly in dependence models (for which intractable normalizing factors are the norm). Here, I discuss several examples of progress on this front, all within the context of exponential family random graph models (ERGMs). These include strategies for efficient analysis of large numbers of pooled networks, approximate continuous mixtures to account for heterogeneity in reciprocity and mean degree, and discrete mixtures of ERGMs for populations with higher levels of heterogeneity. These approaches are illustrated with applications to protein structure and dynamics, intra-organizational networks, and political interactions.

NA

38 Smooth hazards with multiple time scales

Angela Carollo^{1,2}, Hein Putter², Paul H. C. Eilers³, Jutta Gampe¹

1 – Max Planck Institute for Demographic Research, Rostock, Germany; 2 – Leiden University Medical Center, Leiden, The Netherlands; 3 – Erasmus Medical Center,

Rotterdam, The Netherlands

Survival analysis models the time until an event of interest. In many applications more than one time scale is important. Clinical examples include the time since disease onset and duration of the treatment or the age of the patient.

Usually, time-to-event data are described by means of hazard models, which require the choice of one main time scale, according to which the (baseline) hazard changes. Potential other time scale(s) are then included through time-varying covariate(s). Hazard models that allow to include two time scales simultaneously have been proposed, however, they mostly lack full flexibility in the way the hazard can vary along the two time scales.

Here we propose a model in which the hazard is specified as a smooth function of two time scales treated on an equal footing. Estimation is achieved by binning the two-dimensional time plane in many small tiles of equal size, expressing the hazard via bivariate P-splines and maximizing the resulting penalized Poisson likelihood of the event counts and exposure times in the time tiles. To include individual level covariates in a proportional hazard setting we use a modified version of the GLAM algorithm. The optimal choice of smoothing parameters is discussed. The approach is illustrated by an application to the study of mortality after recurrence of colon cancer.

SEHA4

39 Handling missing data in the analysis: practical guidance for structuring the analysis, choosing the tools, and reporting the results

James Carpenter

The London School of Hygiene & Tropical Medicine, United Kingdom

Missing data are a pervasive problem in data analysis. Although there is increasing guidance on how to handle missing data, practice is changing slowly and misapprehensions abound, particularly in observational research. Further, the lack of transparency around methodological decisions is threatening the validity and reproducibility of modern research. Drawing on the recent TARMOS ‘Treatment And Reporting of Missing data in Observational Studies’ [1] framework from the STRATOS initiative, and using data from the Youth Cohort (Time) Series (YCS) for England, Wales and Scotland, 1984-2002, we first discuss how to structure the analysis.

Then, we consider three common ‘tools’ for handling missing data[2]: (a) complete-case analysis, where only units that are complete on the variables in an analysis are included; (b) weighting, where the complete cases are weighted by the inverse of an estimate of the probability of being complete; and (c) multiple imputation (MI), where missing values of the variables in the analysis are imputed as draws from their predictive distribution under an implicit or explicit statistical model, the imputation process is repeated to create multiple filled-in data sets, and analysis is carried out using simple MI combining rules. We give a non-technical discussion of

the strengths and weakness of these approaches.

Finally, we discuss the arguments for systematic reporting of analysis which make use of these techniques, and sketch how this might work.

References

[1] <https://doi.org/10.1016/j.jclinepi.2021.01.008>

[2] Little et al. (2022). A comparison of three popular methods for handling missing data: complete case ... Revision submitted to Sociological Methods and Research.

STRATOS

40 Some theoretical properties of Bayesian Tree methods

Ismaël Castillo

Sorbonne Université, France

Methods based on trees and forests are a popular choice in the statistician's toolbox. Since the introduction of CART, many such methods have been developed and are currently in use, such as Bayesian CART, random forests and BART, to name a few.

In this talk we focus on the properties of posterior distributions based on trees for curve estimation, in regression and density estimation settings.

We first briefly discuss contraction rates of posterior distributions using tree or forest priors when a quadratic-type loss is considered. We then move to the supremum-norm loss, which is particularly appealing for practical applications, as it corresponds to visual closeness of curves. We present a unified framework that enables to derive results for a number of tree prior classes: among others Bayesian CART and variants in regression settings, and Optional Polya trees in density estimation. We show that the corresponding posterior distributions contract near-optimally in terms of rates as well as quantify uncertainty in terms of confidence bands.

We will also discuss one of the possible interests of forest posteriors over single-tree counterparts: using a simple toy example of a forest prior, we show that posteriors based on forests can exhibit a smoothing effect: namely, although individual forest posterior samples are (correlated) random histograms, they contract at the minimax (near-)optimal rate for true functions of arbitrary high Hölder smoothness levels in terms of quadratic or Hellinger losses.

MS2

41 A kernel and distance-covariance perspective on the genomics of quantitative traits

Fernando Castro-Prado¹, Dominic Edelmann², Jelle J. Goeman³

1 – University and Health Research Institute, Santiago de Compostela, Spain; 2

– German Cancer Research Center, Heidelberg, Germany; 3 – Leiden University Medical Center, the Netherlands

Distance covariance is a general dependence measure that was firstly introduced in Euclidean spaces (Székely et al., 2007) and was later extended to general semimetric spaces, having also been shown to be equivalent to the Hilbert–Schmidt independence criterion (Sejdinovic et al., 2014). Most recently, this concept has been shown to present some interesting dualities with the so-called "global tests" in Gaussian process regression (Edelmann and Goeman, 2021).

In this talk, we present some novel distance-covariance methods for testing the association of single nucleotide polymorphisms (SNPs) with quantitative responses. We show that certain versions of distance covariance correspond to locally most powerful tests for specific statistical models leading to insights in which situations these tests perform well, which are of paramount interest in the application fields. What is more, the kind of alternative the test is most powerful against can be specified a priori.

Closed form expressions for the distributions of the test statistics and corresponding estimators are derived from the spectral decomposition of the corresponding operators. The performance of the approach is investigated in various simulation studies and a real world example. Extensions to survival data and testing groups of SNPs are discussed.

SHDA2

42 An educator's perspective of the tidyverse

Mine Çetinkaya-Rundel¹, Johanna Hardin², Benjamin Baumer³, Amelia McNamara⁴, Nicholas Amelia⁵, Colin Rundel¹

1 – Duke University, United States of America; 2 – Pomona College, United States of America; 3 – Smith College, United States of America; 4 – University of St Thomas, United States of America; 5 – Amherst College, United States of America

Computing makes up a large and growing component of data science and statistics courses. Many of those courses, especially when taught by faculty who are statisticians by training, teach R as the programming language. A number of instructors have opted to build much of their teaching around the use of the tidyverse. The tidyverse, in the words of its developers, is a collection of R packages that share a high-level design philosophy and low-level grammar and data structures, so that learning one package makes it easier to learn the next". The shared principles have led to the widespread adoption of the tidyverse ecosystem. No small part of this usage is because the tidyverse tools have been intentionally designed to ease the learning process and cognitive load for users as they engage with each new piece of the larger ecosystem. Moreover, the functionality offered by the packages within the tidyverse spans the entire data science cycle, which includes data import, visualisation, wrangling, modeling, and communication. In this talk, we discuss how the tidyverse provides an effective and efficient pathway to data science mastery for students at a variety of different levels of experience. Specifically, we touch on the what (a brief introduction to the tidyverse), the why (pedagogical benefits, opportunities, and challenges), the how (scoping and implementation options), and the

where (details on courses, curricula, and student populations).

SLSE1

43 Survey Scale Forests: Estimating Unconfounded Latent Variable Effects

Franz Leon Classe¹, Christoph Kern²

1 – Bayerisches Institut für Hochschulforschung und Hochschulplanung, Germany;
2 – School of Social Sciences, University of Mannheim, Germany

We develop a *survey scale forest* (SC Forest) algorithm for the estimation of heterogeneous and causally unbiased effects of a latent variable on an endogenous variable. In the SC Forest framework, the latent variable can be measured by or predict dichotomous, ordinal, or numeric variables. We utilize the *model based recursive partitioning* (MOB) algorithm to detect subgroups with homogenous parameters. This way, SC Forest takes effect heterogeneity in latent variable modeling into account and captures the decision rules that lead to subgroups with good model fit. Furthermore, we were influenced by the *random forest* algorithm and repeat MOB with *random split selection* and *bagging* to form a tree ensemble with high tree diversity. Additionally, we take the requirements for causality of the estimated latent variable effect into account. We do this by implementing *tree honesty* and testing for conditional independence of the latent variable from potential confounders using a non-parametric kernel-based independence criterion.

SC Forest provides various insights about the detected subgroups with homogeneous latent variable effects: A list of terminal nodes with fitted models with acceptable model fit is compiled. Information is provided about whether causal unconfoundedness is given for each of these subgroups. Furthermore, individual predictions of latent variable scores are computed at each iteration, but only for those subgroups for which acceptable model fit indices have been estimated. The individual predictions are then averaged across all trees.

We apply SC Forest to simulated and real data and show that effect heterogeneity and causal unconfoundedness can be detected by the algorithm.

CI

44 Adaptive group sequential designs for single-arm phase II studies with multiple time-to-event outcomes

Moritz Fabian Danzer¹, Tobias Terzer², Frank Berthold³, Andreas Faldum¹, Rene Schmidt¹

1 – Institute of Biostatistics and Clinical Research, University of Münster, Germany; 2 – Division of Biostatistics, German Cancer Research Center, Germany; 3 – Department of Pediatric Oncology and Hematology, Center for Integrated Oncology,

University of Cologne, Germany

Adaptive group sequential designs for single-arm phase II studies with multiple time-to-event outcomes

Existing methods concerning the assessment of long-term survival outcomes in one-armed trials are commonly restricted to one primary endpoint. Corresponding adaptive designs suffer from limitations regarding the use of information from other endpoints in interim design changes.

Here we provide adaptive group sequential one-sample tests for testing hypotheses on the multivariate survival distribution, while making provision for data-dependent design modifications based on all involved time-to-event endpoints. In general, the reference distribution defining the null hypothesis can take any form. However, a derivation under a multi-stage model arises naturally in this context. For this approach, we explicitly elaborate application of the methodology to one-sample tests for the joint distribution of (i) progression-free survival (PFS) and overall survival (OS) in the context of an illness-death model, and (ii) time to toxicity and time to progression while accounting for death as a competing event.

Large sample distributions are derived using a counting process approach. Small sample properties are studied by simulation. An already established multi-state model for non-small cell lung cancer is used to illustrate the adaptive procedure.

In support of this, we present an R package that accompanies the entire workflow of the developed testing procedure. Going beyond the methodology for single-arm studies, we highlight the problems in the development of analogous methods for two-arm comparisons and outline a corresponding solution.

IBSDRP

45 Measurement error and misclassification of covariates: Should we worry?

Veronika Deffner

Federal Office for Radiation Protection, Germany

In order to quantitatively investigate various phenomena in the real world using statistical models, researchers rely on measurements of the variables that are of interest. But what if these measurements have errors? Should we worry about their impact on the validity of the results? Topic group 4 of the STRATOS initiative developed theoretical and practical guidance to help evaluate the impact of the problem and deal with measurement error and misclassification in statistical modeling. The guidance document and the Shiny application “MEM-Explorer” will be presented in the talk with a focus on multiple linear regression with error-prone covariates.

Covariate measurement error may affect the analysis of the association of the covariate with the outcome variable, with regard to three aspects: the estimate of the regression coefficient, the test of the null hypothesis of no association and the power to detect the association. The magnitude of the impact depends on the magnitude and type of the error, as well as on the characteristics of the regression model and the data. In particular, the role of the type of error is often misjudged. Various

methods are available to account for covariate error in a regression analysis, but they all require information about the type and magnitude of the covariate error. “MEM-Explorer” allows interactive exploration of the impact of covariate measurement error: data, based on user-specified settings, are generated and the resulting estimates from the regression models are visualized, with and without the application of methods to adjust these estimates for covariate error.

STRATOS

46 Cross-Sectional Error Dependence in Panel Quantile Regressions

Matei Demetrescu, Mehdi Hosseinkouchak, Paulo Rodrigues

University of Kiel, Germany

We argue that a factor structure of the disturbances in panel quantile regression [QR] models may have a biasing effect on the QR estimator even if factors and loadings are strictly exogenous. Therefore, cross-sectional dependence is an indicator of misspecification in panel quantile regressions. We furthermore propose a test for misspecification based on this argument. The proposed test is a version of the familiar Breusch-Pagan test based on residuals from individual-unit quantile regression estimation at the quantile of interest. To this end, we distinguish between pooled slope coefficient estimation and individual-unit estimation. The test possesses a standard normal limiting distribution under joint N, T asymptotics with restrictions on the relative rate at which N goes off to infinity. A finite-sample correction improves the applicability of the test for larger cross-sectional dimension N .

EEA2

47 Testing relevant hypotheses for functional data

Holger Dette

Ruhr-Universität Bochum, Germany

We develop methodology for detecting relevant hypotheses in functional time series. Instead of testing for exact equality, for example for the equality of the mean functions or principal component functions from two samples, we propose to test the null hypothesis of no relevant "deviation". This means that the null hypothesis is formulated in the form that a metric between the mean functions (or other parameters) is smaller than a given threshold. This formulation is motivated by the fact that small discrepancies might not be of importance in many applied situations.

We consider the L^2 -distance (in the Hilbert space framework) and use the concept of self-normalization to develop (asymptotically) pivotal tests, which does not require long-run variance estimation. We investigate the sample properties by means of a

simulation study and a data example.

(joint work with A. Aue, G. Dierickx, K. Kokot, T. Kutta, S. Volgushev)

SHDA1

48 Causal discovery with incomplete cohort data

Vanessa Didelez, Ronja Foraita, Janine Witte

BIPS, Germany

Causal discovery algorithms estimate causal graphs from observational data. This can provide a valuable complement to analyses focussing on the causal relation between individual treatment-outcome pairs. Constraint-based causal discovery algorithms rely on conditional independence testing when building the graph. Until recently, these algorithms have been unable to handle missing values. In this presentation, we present two alternative solutions: Test-wise deletion (or available-case analysis) and multiple imputation. More specifically, the methods we investigate include random forest imputation and a hybrid procedure combining test-wise deletion and multiple imputation.

We establish necessary and sufficient conditions for the recoverability of causal structures under test-wise deletion, and argue that multiple imputation is more challenging in the context of causal discovery than for estimation. The methods are illustrated and compared in an extensive simulation from benchmark causal graphs: as one might expect, we find that test-wise deletion and multiple imputation both clearly outperform complete-case-analysis and single imputation. Crucially, our results further suggest that multiple imputation is especially useful in settings with a small number of either Gaussian or discrete variables, but when the dataset contains a mix of both neither method is uniformly best.

CI

49 Co-Quantile Regression

Timo Dimitriadis¹, Yannick Hoga²

1 – Heidelberg University, Germany; 2 – University of Duisburg-Essen, Germany

The popular systemic risk measure CoVaR (Co-Value-at-Risk) is widely used in financial risk management, banking regulation, and macroeconomics. Formally, it is defined as an (extreme) conditional quantile of the asset return Y_t , given that the return of another asset X_t is beyond its (extreme) conditional quantile. It measures how extreme observations of X_t influence extreme observations of Y_t and hence, it quantifies the spillover risks from one asset to another. Despite its popularity, the estimation of (models for) the CoVaR is often based on naive, sequential quantile regressions. This practice results in incorrect statistical inference and is insufficient to estimate joint dynamic models, which are, among others, essential to generate reliable CoVaR forecasts.

In this article, we propose a “Co-Quantile Regression” which jointly models the VaR

and CoVaR semiparametrically and propose joint GMM estimation by drawing on recently proposed joint identification functions for the VaR and CoVaR. This allows for the estimation of joint dynamic models, efficient semiparametric estimation, and valid statistical inference. Our method can further be extended to the Marginal Expected Shortfall (ES) and the Co-Expected Shortfall (CoES) in a straight-forward fashion. We apply our co-quantile regression to correct the statistical inference in the existing literature on CoVaR, and to generate CoVaR forecasts for real financial data, which show to be superior to existing methods.

SiF2

50 GFDsurv: A flexible toolbox for factorial survival designs as an alternative to Cox models

Marc Ditzhaus¹, Dennis Dobler², Jon Genuneit³, Arnold Janssen⁴, Markus Pauly¹

1 – TU Dortmund University, Germany; 2 – Vrije Universiteit Amsterdam, The Netherlands; 3 – Leipzig University, Germany; 4 – Heinrich Heine University of Düsseldorf, Germany

While the log-rank test and hazard ratios were the gold standard in time-to-event analysis for a long time, there is a recent trend towards alternative methods not relying on the proportional hazard assumption. The reason for this change are violations of the proportional hazard assumption frequently observed in real data, among others, in oncology. To tackle this, we developed, in a series of publications [1,2,3], different strategies to handle non-proportional hazards in factorial designs and implemented them in the R-package GFDsurv including a user-friendly shiny app. These strategies cover a nonparametric approach and procedures based on novel estimands, such as concordance probabilities, survival medians and, in the near future, restricted mean survival times. In this talk, we present the methodology behind this flexible toolbox including appropriate resampling strategies for a better performance under small sample sizes. Moreover, its application is illustrated by analysing a recent study on asthma [4], for which the assumption of proportional hazards is not justifiable.

References

- [1] Ditzhaus, Genuneit, Janssen and Pauly (2021). CASANOVA: Permutation inference in factorial survival designs. *Biometrics*, 1-13: <https://doi.org/10.1111/biom.13575>.
- [2] Ditzhaus, Dobler and Pauly (2021). Inferring median survival differences in general factorial designs via permutation tests. *Statistical Methods in Medical Research*, Vol. 30, 875-891.
- [3] Dobler and Pauly (2019). Factorial analyses of treatment effects under independent right-censoring. *Statistical Methods in Medical Research*, Vol. 29, 325-343.
- [4] Genuneit (2014) Sex-specific development of asthma differs between farm and nonfarm children: a cohort study. *American Journal of Respiratory and Critical*

51 Dynamic of the urban acoustic environment during COVID-19 Lockdown in Urban Ruhr Area, Germany

Dany Djeudeu¹, Michael Schweitzer¹, Susanne Moebus², Katja Ickstadt¹

1 – TU Dortmund university, Germany; 2 – Institute for Urban Public Health, University Hospital Essen, Germany

The severe restrictions during the COVID-19 pandemic changed the traffic in several places, leading to changes in the acoustic environment.

Data from 22 automated sound recording devices located in the city of Bochum, Germany was analysed. Using the soundscape approach and respective ecoacoustic indices, the development of specific sound pressure level indices and ecoacoustic indices was explored. The indices for the noise level were investigated to make statements about changing noise volumes. The ecoacoustic indices Normalized Difference Soundscape Index (NDSI) and bioacoustic index were used to study the changing proportion of sound sources.

Preliminary analyses of sound indices suggest change point occurring directly after the first lockdown in Germany. This analysis aims to model the growth trajectories of acoustic indices and noise levels across two distinct periods (before and during the COVID-19 pandemic).

We use a Piecewise growth curve model to account for possible change points during the restriction period. We consider the unconditional model to examine the pure time effect on the response (NDSI a example) and the conditional model to examine the effects of covariates on the change. The slope of the linear growth trajectories decreases slightly after the first change points for the acoustic indices.

Reduced traffic led to a distinct decrease in noise levels after the restrictions dues to the COVID-19 pandemic had been imposed.

SAES2

52 Controlling the False Discovery Exceedance for heterogeneous statistical tests

Sebastian Doehler¹, Etienne Roquain²

1 – Hochschule Darmstadt, Germany; 2 – Sorbonne Universite, France

Several classical methods exist for controlling the false discovery exceedance (FDX) for large-scale multiple testing problems, among them the Lehmann-Romano procedure (Lehmann and Romano 2005) and the Guo-Romano procedure (Guo and

Romano 2007). While these two procedures are the most prominent, they were originally designed for homogeneous test statistics, that is, when the null distribution functions of the p -values are all equal. In many applications, however, the data are heterogeneous which leads to heterogeneous null distribution functions. Two prominent examples are when the test statistics are continuous but the p -values are weighted by some known independent weight vector or when the test statistics are discretely distributed, as is the case for data representing frequencies or counts.

In this talk, we present new procedures that deal efficiently with such heterogeneous tests while still guaranteeing rigorous FDX control.

References

- [1] Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.*, 33:1138–1154.
- [2] Guo, W. and Romano, J. (2007). A generalized Sidak-Holm procedure and control of generalized error rates under independence. *Stat. Appl. Genet. Mol. Biol.*, 6:Art. 3, 35 pp.

MS1

53 Addressing hazards in application – a sample size calculation approach for the average hazard ratio

Ina Dormuth¹, Markus Pauly¹, Geraldine Rauch², Carolin Herrmann²

1 – TU Dortmund, Germany; 2 – Institute of Biometry and Clinical Epidemiology, Charité, Germany

Time-to-event endpoints are of primary interest in many clinical studies. Often, they are reported together with hazard ratios and their differences are quantified using the log-rank test. However, in many trials, the proportional hazard assumption does not hold. In this case, applying the log-rank test may lead to a loss in power and hazard ratios vary over time. Therefore, average hazard ratios (Kalbfleisch and Prentice, 1981) were introduced as a time-independent effect measure. It employs a flexible weight function to take the time-dependency into account. Even though the average hazard ratio is interpretable in the proportional and non-proportional hazards setting, it is rarely used in practice (Rauch, Brannath, Brücker and Kieser, 2018). One reason might be the lack of well-established sample size calculation approaches. We address this gap by suggesting an easily applicable approach for calculating the sample size with a corresponding test statistic. The approach is evaluated in extensive simulation studies, regarding multiple survival and censoring distributions as well as their misspecification. Moreover, we give recommendations for application to support using average hazard ratios in practice.

SEHA2

54 An intuitive time-dose-response model for cytotoxicity data with varying exposure times

Julia Christin Duda, Jörg Rahnenführer

TU Dortmund University, Germany

Modeling approaches for dose-response analyses are becoming more popular in toxicological applications. For cytotoxicity assays, not only the concentration but also exposure period of the compound administered to cells can be varied. A popular concentration-response model is the two-parameter log-logistic (2pLL) model.

We propose a two-step procedure and a new time-concentration-response model for cytotoxicity data in which both concentration and exposure period are varied. The parameter of interest for the estimation is the EC50 value. The procedure consists of a testing step and a modeling step. In the testing step, a nested ANOVA test is performed to decide if the exposure time has an effect on the shape of the concentration-response curve. If no effect is identified then a classical 2pLL model is fitted. Otherwise, a new time-concentration-response model called td2pLL is fitted. In this model, we incorporate exposure duration information into the 2pLL model by making the EC50 parameter dependent on the exposure duration.

In simulation studies based on a real data set, we compare the proposed procedure against various alternatives with respect to the precision of the estimation of the EC50. In all simulations, the new procedure provides estimates with higher or comparable precision, which demonstrates its universal applicability in corresponding toxicological experiments. In addition, we show that the use of optimal designs for cytotoxicity experiments further improves the EC50 estimates in most scenarios considered. In order to facilitate the application in toxicological practice, the developed methods are available to practitioners via the R package td2pLL.

OT

55 The time to progression ratio in molecular tumor trials – a critical examination of current practice and suggestions for alternative methods

Dominic Edelmann, Tobias Terzer, Axel Benner

German Cancer Research Center, Germany

The time to progression ratio (TTP_r) is a novel endpoint in Phase I/II oncology trials, which is frequently applied to evaluate the efficacy of molecular targeted treatments in late stage patients. The general idea of the design is that a patient serves as their own control. To calculate the TTP_r for an individual, the time to progression (TTP) under the experimental targeted treatment is divided by the last TTP under standard treatment. If the TTP_r exceeds a certain value (typically 1.3), the person is considered a responder. Subsequently, a binomial test is performed, investigating if the proportion of responders is significantly higher than a certain

threshold (typically 15%). In this work, the current practice for the TTPr is critically examined. Using elementary calculations and simulations, we point out numerous shortcomings of the current methodology. Notably, the applied threshold values will often lead to even if the experimental treatment is harmful. On the other hand, the approach features little power when the thresholds are specified correctly. As a remedy to these shortcomings, we propose an alternative methodology using a Cox model or accelerated failure time (AFT) model for clustered time-to-event data. Using a comprehensive simulation study, we show that this methodology clearly outperforms the current approach in terms of power.

SEHA4

56 Data Science for informed citizens: On synergies between Digital Literacy and Civic Statistics

Joachim Engel, Laura Martignon

Ludwigsburg University of Education, Germany

Digital media and the availability of data of sheer unlimited scope and magnitude change our access to information in radical ways. Emerging data sources provide new sorts of evidence, provoke new sorts of questions, make new sorts of answers possible and shape the ways in which evidence is used to influence decision making in private, professional and public life. Data science as a practical science has been conceived to address tangible problems in science, technology and society. Problems related to society require skills beyond the technical mastery of algorithms such as data ethics, measurability issues, classification of results under risk and uncertainty and in specific contexts and awareness of implications for policy and society. Data literacy can be understood as the ability that a responsible citizen (as “mündiger Bürger”) needs in digitization in order to find his/her way around an oversupply of data and information and to make well-founded decisions - in everyday life and at various political levels. This decision-making competence requires the ability to differentiate in order to separate data and information from interpretations and opinions. These features relate data science closely to the recently developed field of civic statistics. Many civic statistics competencies needed to make sense of data about society apply also to digital literacy. This conceptual paper looks at the common grounds between the two disciplines. It investigates how digital literacy and data science can enhance civic statistics and, vice versa, how civic statistics concepts have the potential to enrich digital literacy.

SLSE1

57 Network meta-analysis of rare events using penalized likelihood regression

Theodoros Evrenoglou¹, Ian R. White², Sivem Afach³, Dimitris Mavridis⁴, Anna Chaimani¹

1 – *Université de Paris, Research Center of Epidemiology and Statistics (CRESS-U1153), INSERM, Paris, France*; 2 – *MRC Clinical Trials Unit, University College London, London, UK*; 3 – *Université Paris-Est Créteil, UPEC, EpiDermE EA 7379, Créteil, F-94010, France*; 4 – *Department of Primary Education, University of Ioannina, Greece*

Network meta-analysis (NMA) of rare events has attracted little attention in the literature. When there are studies with rare events, the normal approximation assumptions made by the conventional inverse variance NMA model are compromised and effect estimates are potentially biased. An approach of penalizing the likelihood function has been previously proposed for bias reduction of the maximum likelihood estimator (MLE) for the analysis of individual studies with rare events. To improve the accuracy and the precision of the NMA relative effect estimates in the presence of rare events, we extended the penalized likelihood approach to the context of NMA. We focused on the common effect logistic regression expression of the NMA model and we removed the first term from the asymptotic bias expansion of the MLE. This approach has the ability to address separation issues and to always provide finite estimates. Thus, it can also be applied in NMAs including studies with extremely rare and even zero events. We further used a two-stage approach to incorporate a heterogeneity parameter in our model through a multiplicative overdispersion term. We evaluated the performance of our approach using 32 simulation scenarios and we compared it with other 10 NMA approaches in terms of bias, coverage probability and different uncertainty measures. The simulations suggested that our approach performs consistently well across all tested scenarios and most often results in smaller bias than the other NMA methods especially in networks with very few studies per comparison and very low control group risks.

MA2

58 Penalized estimation of INAR models

Maxime Faymonville¹, Carsten Jentsch¹, Christian Weiß², Boris Alexandrov²

1 – *TU Dortmund University, Germany*; 2 – *Helmut Schmidt University Hamburg, Germany*

Integer time series have received increasing attention in research over the decades. Well-known models for integer-valued autoregressive data are INAR models, for which the literature mainly deals with parametric estimation. In a remarkable paper, Drost et al. (2009) on the other hand introduced a semiparametric estimation approach, which allows for estimation of the INAR models without any parametric assumption on the innovation distribution. However, with a low sample size of observations, a semiparametric estimation approach may lead to inferior estimation

performance. Therefore, we propose a penalized version of the semiparametric estimation approach of Drost et al. (2009), which exploits the fact that the innovation distribution is often considered to be smooth. This is the case, for example, in the frequently used INAR models with Poisson, negative binomially or geometrically distributed innovations. In Monte Carlo simulations, we illustrate the superiority of the proposed penalized estimation approach and argue that a combination of penalized and unpenalized estimation approaches results in overall best INAR model fits.

Reference

[1] Drost, F., Van den Akker, R., and Werker, B. “Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models”. In: *Journal of the Royal Statistical Society. Series B* 71, Part 2 (2009), pp. 467-485.

TSA3

59 Using Double Machine Learning to Understand Nonresponse in the Recruitment of a Mixed-mode Online Panel

Barbara Felderer¹, Jannis Kück², Martin Spindler²

1 – GESIS, Germany; 2 – Universität Hamburg, Germany

The recruitment to a probability-based panel is arguably the most important and most expensive part of the panel life cycle. The recruited sample needs to well represent the target population and the sample size needs to be large enough to obtain precise estimates. Losing potential panel participants during the recruitment phase challenges the quality of the panel. To prevent nonresponse, it is crucial to understand nonresponse behavior.

For mixed-mode online panels, it is common to conduct a first recruitment interview in an interviewer-administered mode and invite recruited respondents to participate in a subsequent welcome survey in a self-administered mode of their choice. Participating in the welcome survey thus includes a switch in response mode that might provoke systematic nonresponse.

In this paper, we study nonresponse in the welcome survey of the recruitment for the GESIS panel. We make use of a rich set of variables collected in the recruitment survey which poses challenges to nonresponse modeling. Classical machine learning, that is frequently used to predict nonresponse in high-dimensional settings, is not applicable when the goal is to explain rather than to predict nonresponse. Causal inference can, however, be drawn using the double machine learning approach which we apply in this paper.

Our analysis focuses on the role of the chosen survey mode in interaction with socio-demographic characteristics. Comparing nonresponse models from the initial recruitment in 2013 and the refreshment in 2016, we evaluate whether relationships change over time, e.g., due to societal change like increasing digitalization

and Internet coverage.

SM2

60 A new ensemble model for multivariate functional data classification with an application to survey research

Amanda Fernández-Fontelo¹, Felix Henninger², Pascal J. Kieslich³, Frauke Kreuter², Sonja Greven¹

1 – Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Germany; 2 – Institute for Statistics, LMU München, Germany; 3 – Mannheim Centre for European Social Research, University of Mannheim, Germany

We introduce a new ensemble method for d -variate functional data classification based on semi-metric-type methods by Ferraty and Vieu (2003) and Fuchs et al. (2015). We particularly extend their methods in two ways: First, we generalize their methods from the univariate to the multivariate functional case and propose new semi-metrics to compute distances between functions. Second, we look at more flexible options for combining weak classifiers in ensemble models, going beyond Fuchs' ideas of linear combinations and considering alternatives such as stacked generalization methods using tree-based random forest and gradient boosting models. Our ensemble models can thus be viewed as flexible methods for d -variate functional data classification, allowing, e.g., for more complex dependencies on functional features.

We apply our ensemble model to improve survey data quality by identifying respondents facing difficulties with several questions. The goal of survey researchers is to collect reliable data from respondents' questionnaires while reducing potential sources of errors. One of these sources is respondents' difficulty in answering survey questions appropriately. For this purpose, we apply our ensemble models to user interaction data from a survey in which we experimentally manipulated multiple questions to create two controlled levels of difficulty. After gaining consent from all users, we collected the mouse movement trajectories as bivariate functional data and used them to classify the two difficulty levels. We then validate the predictive performance of our models, showing promising results for difficulty detection and thus helping improve the quality of online survey data.

CaC

61 Generalized Correlation Measures

Tobias Fissler¹, Marc-Oliver Pohle²

1 – Vienna University of Economics and Business (WU), Austria; 2 – Goethe University Frankfurt, Germany

We introduce and examine generalised measures of covariance and correlation. The classical Pearson covariance between two random variables X and Y is the inner product of the two errors $X - E(X)$ and $Y - E(Y)$. These errors measure the joint

deviation of X and Y from their means. If one is interested in joint deviations around functionals T other than the mean, e.g. the median, one should not resort to a linear deviation measure, i.e., $X - T(X)$, but should rather utilise a generalised error $e_T(X) = v_T(X, T(X))$. Here, v_T is an identification function for T which implies that $E[e_T(X)] = 0$, ensuring that the generalised covariance between X and Y relative to T is 0 if X and Y are independent.

This new measure allows to assess the joint behaviour of X and Y not only around measures of central tendency, but also around measures of the tail behaviour, such as quantiles or expectiles. This can be particularly interesting in the context of risk management, and it resembles the rationale of the recent systemic risk measure CoVaR (conditional value at risk). This generalisation of the covariance is akin to the historically older extension of mean or least squares regression to quantile and expectile regression (also called asymmetric least squares regression).

We further discuss appropriate normalisations, attainability issues, and the dependence or independence of the marginal distributions of X and Y . Finally, we extend this concept to distributional correlations and discuss the link to famous rank-based correlation measures such as Spearman's ρ or Kendall's τ .

SiF2

62 Teaching and learning machine learning using educationally designed Jupyter Notebooks

Yannik Fleischer, Rolf Biehler, Susanne Podworny

Universität Paderborn, Germany

As part of the Project Data Science and Big Data at School (www.prodabi.de/en), we develop and implement data science courses in different scenarios: a yearlong course for upper secondary level and teaching modules for grades 9/10 and 5/6. In these different grades, we use Jupyter Notebooks (JNB) with Python for various applications. A main topic is data-based predictive modeling. We designed a Python library (PyTree) for machine learning with decision trees using JNB in our educational settings. In class, we use JNBs with PyTree in different ways: as tool JNB for teaching and learning specific aspects and concepts, as interactive worked example JNB that offer a framework for carrying out more complex tasks, and as computational essay template JNB that students are to use for documentation and explanation of their work in a reproducible and transparent way. The JNB for younger students hide Python commands: students work in menu-based environments for exploring decision trees and for using professional libraries for creating decision trees. Students at upper secondary level also work in a code-based environment with Python commands.

In our presentation, we will explain the design and use of JNB with PyTree and present different activities that students carried out with different multivariate data. For example, we use data on media behavior of adolescents and data on occupancy of parking lots to create decision trees as predictive models. The educational approach can also be adapted for tertiary level courses, which use JNB with Python or the

markdown language in R environments.

SLSE2

63 A stabilised Aalen-Johansen estimator with internal left-truncation and overly small risks sets

Sandra Frank¹, Sabrina Dietz², Arthur Allignol³, Jan Beyersmann¹

1 – Institute of Statistics, University of Ulm, Germany; 2 – DATAMAP GmbH, Freiburg, Germany; 3 – Daiichi Sankyo Europe GmbH, Munich, Germany

Standard survival estimators like the Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen estimator account for left-truncated data. However, with left-truncated data there could occur overly small risk sets which may lead to big jumps in the estimates and overly wide confidence intervals. Therefore, Lay&Ying (1) suggest a modified Kaplan-Meier estimator which Friedrich et al. (2) extend to competing risks data with a modified Aalen-Johansen estimator.

Even if there is no left-truncated data, every multistate model is subject to internal left-truncation because patients arrive one by one in intermediate states. E.g. stem-cell transplanted patients can drip into the status of GvHD and suffer a relapse or die immediately afterwards. Here, too small risk sets can occur which are not known at time 0 and may lead to big jumps in the Aalen-Johansen estimator. Therefore, we generalize the estimator from Friedrich et al.(2) to multistate model data. We detect in 6 out of 12 freely available R datasets that overly small risk sets occur and our stabilised Aalen-Johansen estimator should be considered to estimate transition probabilities. The key quantity of the stabilised estimator contains a threshold depending on the number at risk which choice we discuss in a simulation study and apply our results to data from stem cell transplantation registry. A special challenge to specify the threshold in the multistate framework is that a transition probability depends also on the hazards of the other transitions.

References

- [1] LAI, Tze Leung; YING, Zhiliang; 1991, AoS
- [2] FRIEDRICH, Sarah, et al.; 2017, AoAS

SEHA2

64 On D-Optimal designs for repeated item response testing

Fritjof Freise¹, Heinz Holling², Rainer Schwabe³

1 – Stiftung Tierärztliche Hochschule Hannover, Germany; 2 – Westfälische Wilhelms-Universität Münster, Germany; 3 – Otto-von-Guericke-Universität Magdeburg, Germany

We present findings on D-optimal designs for estimating the mean ability over time in repeated measures growth curve models. We develop a new equivalence theorem

for this situation and present analytical solutions for some simplified growth curve models. Motivated from studying the improvement in psychological item response testing results, when participants are allowed to take retests, we further consider models incorporating an increasing mean ability and a saturation effect, i.e. the mean ability approaches a maximal achievable value as an upper limit. For these more realistic nonlinear models, computational results are given for the D-optimal designs which are validated by the equivalence theorem.

DECT2

65 Better multiple Testing: Using multivariate co-data for hypotheses

Daniel Fridljang

European Molecular Biology Laboratory, Germany

Consider a multiple testing task, where for each test we have access to its p-value and additional information represented by a covariate that may be informative on prior probabilities of null and alternative hypotheses and/or on the test's power. The independent hypothesis weighting (IHW) framework uses these covariates to stratify the tests into a finite number of bins which are then assigned different weights for use in a weighted multiple testing method such as that of Benjamini and Hochberg. IHW guarantees false discovery rate control and increases overall power compared to the unweighted method.

Ignatiadis and Huber propose performing the stratification by quantile slicing the covariates. However, this simple approach does not take full advantage of the data. First, quantiles are not able to capture the heterogeneities among tests. The hypotheses within each bin should be as homogeneous as possible. Second, the procedure becomes computationally infeasible for high-dimensional covariates, due to an exponential increase of the number of bins with the covariates' dimension.

We address this challenge by introducing a random forest based approach, where the leaves of the trees replace the bins. The objective function is chosen such that the splits are sensitive to the prior probability and shape of the postulated conditional density. This yields homogeneous bins and hence increases power. The trees can handle high-dimensional covariates and an orthonormal series expansion makes growing the trees computationally efficient. We apply the updated implementation to RNA-seq data and show the benefits of our method when performing differential expression analysis.

BSB1

66 All that Glitters is not Gold: Relational Events Models with Spurious Events

Cornelius Fritz, Marius Mehrl, Paul W. Thurner, Göran Kauermann

LMU Munich, Germany

As relational event models are an increasingly popular model for studying relational structures, the reliability of large-scale event data collection becomes increasingly important. Automated or human-coded events often suffer from relatively low sensitivity in event identification. At the same time, most sensor data is primarily based on actors' spatial proximity for predefined time windows; hence, the observed events could relate either to a social relationship or random co-location. Both examples lead to false positives in the observed events that may bias the estimates and inference. We propose the Relational Event Model for Spurious Events (REM-SE) as an extension to existing approaches for interaction data. The model provides a flexible solution for modeling data while controlling for false positives. Estimation of our model is carried out in an empirical Bayesian approach via data augmentation. In a simulation study, we investigate the properties of the estimation procedure. Consecutively, we apply this model to combat events from the Syrian civil war and to student co-location data. Results from both the simulation and the application identify the REM-SE as a suitable approach to modeling relational event data in the presence of measurement error.

NA

67 Narrowest Significance Pursuit: inference for multiple change-points in linear models

Piotr Fryzlewicz

London School of Economics, United Kingdom

We propose Narrowest Significance Pursuit (NSP), a general and flexible methodology for automatically detecting localised regions in data sequences which each must contain a change-point, at a prescribed global significance level. Here, change-points are understood as abrupt changes in the parameters of an underlying linear model. NSP works by fitting the postulated linear model over many regions of the data, using a certain multiresolution sup-norm loss, and identifying the shortest interval on which the linearity is significantly violated. The procedure then continues recursively to the left and to the right until no further intervals of significance can be found. The use of the multiresolution sup-norm loss is a key feature of NSP, as it enables the transfer of significance considerations to the domain of the unobserved true residuals, a substantial simplification. It also guarantees important stochastic bounds which directly yield exact desired coverage probabilities, regardless of the form or number of the regressors. NSP works with a wide range of distributional assumptions on the errors, including Gaussian with known or unknown variance, some light-tailed distributions, and some heavy-tailed, possibly heterogeneous distributions via self-normalisation. It also works in the presence of autoregression. The

mathematics of NSP is, by construction, uncomplicated, and its key computational component uses simple linear programming. In contrast to the widely studied post-selection inference approach, NSP enables the opposite viewpoint and paves the way for the concept of post-inference selection". The R package `nsp` is available from CRAN.

TSA1

68 Depth-based two-sample testing

Felix Gnettner¹, Claudia Kirch¹, Alicia Nieto-Reyes²

1 – Otto-von-Guericke-Universität Magdeburg, Germany; 2 – Universidad de Cantabria Santander, Spain

Depth functions provide measures of the deepness of a point with respect to a given set of observations. This non-parametric concept can be applied in spaces of any dimension and entails a center-outward ordering for the given data. A two-sample test has been previously proposed that is based on depth-ranks and offers opportunities for further investigations: Observing that the corresponding test statistic $\mathcal{L}\mathcal{S}(X, Y)$ is not symmetric with respect to the two samples X and Y , the power can be greatly increased if $\mathcal{L}\mathcal{S}(X, Y)$ and $\mathcal{L}\mathcal{S}(Y, X)$ are jointly considered. Within the last years, depths with respect to functional data have been established that we combine with this procedure to obtain new non-parametric two-sample tests for functional data. We investigate the asymptotic behaviour of this modified test procedure for several classes of depths including depths for functional data.

RNS3

69 R package 'robcp' for robust detection of change points

Sheila Görz¹, Alexander Dürre², Herold Dehling³, Roland Fried¹

1 – TU Dortmund, Germany; 2 – Université Libre de Bruxelles, Belgium; 3 – Ruhr-Universität Bochum, Germany

This talk reports our progress on the R package 'robcp', which provides robust tests for the detection of change points in time series in the possible presence of outliers and heavy tails. Our R package includes tests based on Huberized versions of CUSUM tests, or on statistics with suitable robustness properties like medians, quantiles, U -statistics, and U -quantiles. Focus is on the detection of abrupt changes of location, but changes of scale or correlation can be detected as well. The Huberized CUSUM tests can cope with both uni- and multivariate time series.

CSASOFT1

70 Variable Selection and Allocation in Joint Models via Gradient Boosting Techniques

Colin Griesbach¹, Andreas Mayr², Elisabeth Bergherr¹

1 – *University of Göttingen, Germany*; 2 – *University of Bonn, Germany*

Modelling longitudinal data and risk for events separately, even though the underlying processes are related to each other, leads to loss of information and bias. Hence, the popularity of joint models for longitudinal and time-to-event data has grown rapidly in the last few decades. Gradient boosting is a statistical learning method that has the inherent ability to select variables and estimate them at the same time. We construct a data-driven allocation algorithm for basic joint models by applying recent developments from gradient boosting for distributional regression. Instead of specifying beforehand which covariate has an influence on which part of the joint model, the algorithm allocates the covariates to the appropriate sub-model. A simulation study shows that this is possible when using a non-cyclic updating scheme for the boosting algorithm. In addition, recent findings regarding adaptive step lengths and early stopping based on probing are incorporated in order to improve allocation accuracy and reduce the computational effort.

ARM2

71 Combining point forecasts to calibrated probabilistic forecasts using copulas

Oliver Grothe, Jonas Rieger

Karlsruhe Institute of Technology (KIT), Germany

We formulate a copula-based algorithm for combining multiple point forecasts with complicated forecast error dependencies to a calibrated density forecast through a Bayesian representation. The combination proceeds by a copula times series model of the joint forecast error density and thus accounts for the forecasts' serial dependence and the forecast errors' dependence structure. We show that the algorithm can produce calibrated forecasts for any forecast errors, particularly if the actual margins and copula are known. In the case of multivariate Gaussian forecast errors, the presented algorithm embeds the results of the point forecast combination by Bates & Granger (1969). We compare the algorithm to other forecast combination methods in a simulation study and provide a worked-through example on temperature forecast data for the Zugspitze, the highest mountain in Germany.

Reference

[1] Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, 20, 451. <https://doi.org/doi:10.2307/3008764>.

BS2

72 Modelling Volume-Outcome Relationships in Health Care

Maurilio Gutzeit, Johannes Rauh, Jona Cederbaum

Federal Institute for Quality Assurance and Transparency in Healthcare (IQTIG) / Biometrics Unit, Berlin, Germany

Despite the ongoing strong interest in associations between quality of care and the number of cases (volume) of healthcare providers, a unified statistical framework for analyzing them is missing and many studies suffer from poor statistical modelling choices such as the discretization of volume into groups.

In this talk, we present a flexible, additive mixed model on the level of the individual patients for studying volume-outcome associations in health care. We treat volume as a continuous variable and model its effect on the considered outcome through penalised splines. We take account of different case-mixes by including patient specific risk factors and of clustering on the provider level through random intercepts. Using that approach, we obtain a smooth volume effect as well as volume-independent provider effects. A comparison of these two quantities gives insight into the sources of variability of quality of care. All effects are estimated in a unified framework allowing for adequate uncertainty quantification.

Depending on the estimated association from data, our approach also enables the estimation of potential threshold values for the volume based on a break point model. This is, for example, of interest when investigating administrative requirements on the minimum provider volume. Furthermore, given a potential minimum provider volume, it is also possible to evaluate the statistical effect on the number of adverse outcomes.

We illustrate our approach through an example based on German health care data.

IQW/TIG

73 Combining deep learning and modeling for time-series single-cell RNA-sequencing data

Maren Hackenberg, Harald Binder

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

Single-cell RNA-sequencing data from multiple time points promises insights into mechanisms controlling differentiation and cell fate decisions at the level of individual cells. Yet, due to the destructive nature of the sequencing protocol, at each time point a different, heterogeneous sample of cells from diverse types and developmental stages is obtained. This complicates the identification of specific developmental trajectories across multiple time points.

To address this challenge, we propose a modeling approach that integrates neural network-based dimension reduction with inference of the temporal dynamics. We use a deep learning approach to infer a low-dimensional, latent representation of gene expression. In this latent space, we optimize an explicit dynamic model to

describe cell type-specific trajectories. Specifically, we alternate between assigning cells into groups based on the current dynamic model predictions, and optimizing the model parameters by matching the predicted and true distributions in each group with a quantile-based loss function.

Based on a simulation study and an application on neural differentiation in mice, we show that this approach allows for inferring distinct developmental trajectories despite the lack of one-to-one correspondence between cells at different time points. Jointly optimizing the neural network for dimension reduction and the dynamic model allows for learning an improved low-dimensional representation specifically adapted to the underlying dynamics.

The approach thus more broadly exemplifies the potential of combining neural networks and dynamic modeling to address complex modeling challenges, such as the joint identification of developmental trajectories and a suitable dimension reduction from time-series single-cell RNA-seq data.

BSB2

74 Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten

Sandra Hadam

Statistisches Bundesamt, Germany

Kleinräumige und aktuelle Bevölkerungszahlen sind für politische Entscheidungsfindungen unerlässlich. Die Bevölkerungsfortschreibung ermöglicht die Angabe aktueller Einwohnerzahlen auf geografischer Ebene der Gemeinden und wird auf Basis des Zensus 2011 laufend fortgeschrieben. Um den wachsenden Bedarf an kleinräumigeren Bevölkerungszahlen kurzfristig zu decken, wird die Bevölkerungsfortschreibung mit einem neuen, experimentellen Lösungsansatz ergänzt. Bisherige Analysen zur Bevölkerungsdarstellung mit Mobilfunkdaten zeigen grundsätzlich, dass die Bevölkerungsverteilung mit Mobilfunkdaten gut und zeitnah abgebildet werden kann (Hadam et. al 2020). Im Projekt "experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten" wird darauf aufbauend erforscht, inwieweit anhand der Mobilfunkdaten die Einwohnerzahl der Bevölkerungsfortschreibung von der Gemeindeebene bundesweit auf 1×1 km Gitterzellen umverteilt werden kann bis die erste georeferenzierte Bevölkerungszahl auf Basis des Zensus 2022 vorliegt. Zu diesem Zweck stehen Mobilfunkdaten für ganz Deutschland auf einem INSPIRE-konformen 1×1 km Raster zur Verfügung, die aufgrund ihrer speziellen Aufbereitung die potentielle Wohnbevölkerung bundesweit abbildet. Im Rahmen eines Verfahrens werden die Ergebnisse der Bevölkerungsfortschreibung daraufhin anhand der Verteilung der ermittelten Wohnbevölkerung aus den Mobilfunkdaten von der Gemeindeebene bundesweit auf 1×1 km Gitterzellen umverteilt. Anhand der experimentellen georeferenzierten Bevölkerungszahl kann der Bedarf an aktuellen und georeferenzierten Bevölkerungszahlen gedeckt werden. Der experimentelle Charakter erlaubt die Vorteile der Mobilfunkdaten und ihrer flächendeckenden kleinräumigen Verfügbarkeit zur Verteilung der Bevölkerungszahlen

unterhalb der Gemeindeebene zu nutzen.

Referenz

Hadam, S., Schmid, T., Simm, J. (2020): Kleinräumige Prädiktion von Bevölkerungszahlen basierend auf Mobilfunkdaten aus Deutschland. In: Klumpe, B., Schröder, J., Zwick, M. (Hrsg.) Qualität bei zusammengeführten Daten – Befragungsdaten, Administrative Daten, Neue digitale Daten: Miteinander besser? Springer, Wiesbaden, S 27-44. https://doi.org/10.1007/978-3-658-31009-7_3

OAS2

75 Boosting copulas with continuous margins

Nicolai Hans¹, Nadja Klein¹, Andreas Mayr²

1 – Humboldt-Universität zu Berlin, Germany; 2 – Universität Bonn, Germany

Capturing complex dependence structures between outcome variables is of high relevance in contemporary biomedical data problems. Conditional copula regression provides a flexible tool to model the joint distribution of multiple outcome variables by disentangling the marginal response distributions and their dependence structure. In a regression setup each parameter of the copula model, i.e. the marginal distribution parameters and the copula dependence parameters, can be related to covariates via structured additive predictors. While there exist Bayesian and penalized likelihood-based frameworks to estimate the additive predictors of this model class, we propose a model-based boosting algorithm to fit copula regression models with continuous margins. Model-based boosting is a modern estimation technique that incorporates useful features like an intrinsic variable selection mechanism, parameter shrinkage and the capability to fit regression models in high dimensional data setting, i.e. situations with more covariates than observations. An interesting consequence of the variable selection mechanism is that also independence between the outcome variables can be detected by a lack of selected variables in the predictors associated with the dependence parameters. The performance of the boosting algorithm in the context of copula regression models with continuous margins is evaluated in simulation studies that cover low and high dimensional data settings and situations with and without dependence between the responses. Moreover, copula boosting is used to jointly analyze the length and the weight of newborns conditional on sonographic measurements during pregnancy and multiple clinical variables.

ARM3

76 Some comments on CV

Trevor John Hastie

Stanford University, United States of America

Cross-validation is ubiquitous in data science, and is used for both model selection and assessment. Yet in some regards it is poorly understood. In this talk we discuss

three aspects of CV.

- What CV estimates?
- Confidence intervals for prediction error using nested CV.
- OOB error for random-forests and standard error estimates.

This talk is dedicated to Leo Breiman and Colin Mallows.

(Joint work with Stephen Bates, Samyak Rajanala and Rob Tibshirani)

P3

77 A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model

Andreas Haupt¹, Øyvind Nicolay Wiborg², Nicolai Topstad Borgen²

1 – Karlsruhe Institute of Technology, Germany; 2 – University of Oslo, Norway

The identification of unconditional quantile treatment effects (QTE) has become increasingly popular within social sciences. However, current methods to identify unconditional QTEs of continuous treatment variables are incomplete. Contrary to popular belief, the unconditional quantile regression model introduced by Firpo, Fortin, and Lemieux (2009) does not identify QTE, while the propensity score framework of Firpo (2007) allows for only a binary treatment variable, and the generalized quantile regression model of Powell (2020) is unfeasible with high-dimensional fixed effects. This paper introduces a two-step approach to estimate unconditional QTEs. In the first step, the treatment variable is regressed on the control variables using an OLS. Then, in the second step, the residuals from the OLS are used as the treatment variable in a quantile regression model. The intuition is that the first step decomposes the variance of the treatment variable into a piece explained by the observed control variables and a residual piece independent of the controls. Since the control variables purge the treatment of confounding, they are redundant in the second step. Therefore, our RQR approach circumvents the problem that the inclusion of controls together with the treatment variable in CQR changes the interpretation of the treatment coefficients. Unlike much of the literature on quantile regression, this two-step residualized quantile regression framework is easy to understand, computationally fast, and can include high-dimensional fixed effects.

ARM4

78 Selection of variables and functional forms for multivariable models

Georg Heinze

Medizinische Universität Wien, Austria

In hardly any statistical topic the gap between cutting-edge methodology and daily practice of data analysis is larger than in multivariable model building. For example, according to recent systematic reviews, univariate screening of variables considered for multivariable models is still common practice and sometimes even believed to be a precondition for inclusion. Furthermore, while there are sophisticated and well-investigated methods for considering nonlinear functional forms of continuous variables in such models, dichotomization of such variables is still popular. If continuous variables are not dichotomized, the linearity assumption is hardly ever questioned. We give an overview of methods for selection of variables and functional forms, discuss recommendations for practitioners and point at open gaps in research. Furthermore, we present some activities of our initiative to guide and educate researchers with different levels of statistical education, including our workshops, short videos and interactive shiny apps.

STRATOS

79 An empirical comparison and characterisation of nine popular clustering methods

Christian Hennig

University of Bologna, Italy

Nine popular clustering methods are applied to 42 real data sets. The aim is to give a detailed characterisation of the methods by means of several cluster validation indexes that measure various individual aspects of the resulting clusters such as small within-cluster distances, separation of clusters, closeness to a Gaussian distribution etc. 30 of the data sets come with a trueclustering. On these data sets the similarity of the clusterings from the nine methods to the trueclusterings is explored. Furthermore, a mixed effects regression relates the observable individual aspects of the clusters to the similarity with the trueclusterings, which in real clustering problems is unobservable. The study gives new insight not only into the ability of the methods to discover trueclusterings, but also into properties of clusterings that can be expected from the methods, which is crucial for the choice of a method in a real situation without a given trueclustering.

CaC

80 Distributional latent class modelling for the indirect estimation of reference distributions using mixture density networks

Tobias Hepp^{1,2}, **Jakob Zierk**³, **Sarem Seitz**⁴

1 – Institut für Medizininformatik, Biometrie und Epidemiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; 2 – Professur für Raumbezogene Datenanalyse und Statistische Lernverfahren, Georg-August-Universität Göttingen, Germany; 3 – Kinder- und Jugendklinik, Universitätsklinikum Erlangen, Germany; 4 – Professur für Angewandte Informatik, insbes. Kognitive Systeme, Otto-Friedrich-Universität Bamberg, Germany

Reference intervals represent the expected range of physiological test results in a healthy population and are essential to support medical decision making. Particularly in the context of pediatric reference intervals, what is considered healthy or not depends on covariates such as age. Estimation approaches that aim to accurately describe the distribution of healthy values are therefore very demanding in terms of sample size, a challenge that is difficult to meet because of strict regulations applying to the recruitment of children.

Currently established indirect methods enable robust identification of the distribution of healthy samples from laboratory databases that also include unlabeled pathologic cases, but are severely limited when adjusting for essential patient characteristics such as age. To overcome this limitation, we propose the use of conditional finite mixture models in the form of latent class distributional regression. Earlier implementations based on the EM-algorithm provided promising results but were limited to estimating the mixture component weights as constants. Mixture density networks provide a promising alternative based on the framework of artificial neural networks that is able to estimate all parameters including component weights as non-linear functions of the covariate(s). Different implementations are evaluated in varying simulation settings as well as on a large dataset of hemoglobin concentration and compared with the results of the EM-framework.

ARM4

81 Because there is no other way: Estimating vaccine effectiveness using observational data

Miguel Hernan

Harvard University, United States of America

Randomized trials cannot provide precise estimates of vaccine effectiveness and safety for rare outcomes, multiple subgroups, new virus variants, and different vaccination schedules. To answer many of these questions, we need to rely on analyses of observational databases that emulate (hypothetical) target trials. This talk describes key challenges to emulate target trials of vaccine effects using observational data and propose some solutions. The methods are illustrated with world examples

based on the analyses of nationwide databases from several countries.

P1

82 A pattern extraction app for biomedical data

Moritz Hess, Dominik Scheuer, Harald Binder

Faculty of Medicine and Medical Center - University of Freiburg, Germany

Explorative analyses play a huge role in biomedical research. A common scenario is, that scientists generate a small data set (pilot data) to identify multivariable signals (patterns) in their experimental setting, e.g., to study if cells at different time-points differ in the expression of specific genes.

We have developed methods for extracting such multivariable patterns with deep generative models and log-linear models (Hess et al. 2020). However, as with similar published approaches, applying those requires the user to be familiar with data science techniques and text based interfaces.

Here we propose an iOS / web app which is capable to extract multivariable patterns from potentially high dimensional omics data.

The app provides an interface for visually studying the data, e.g. with cluster heatmaps, as well as techniques for variable selection such as boosting, including resampling techniques for assessing the randomness of a pattern. The user can interactively gain an understanding of how the implemented variable selection techniques work by studying, how likely synthetic patterns are picked up dependent on signal strength. By importing her real experimental pilot data, the user can then extract patterns, i.e. combinations of related variables. By providing variables of interest, the user can study how the identified patterns relate to his expert knowledge.

Reference

[1] Hess, M., Hackenberg, M., & Binder, H. (2020). Exploring generative deep learning for omics data using log-linear models. *Bioinformatics*, 36(20), 5045-5053.

VEDA

83 Regional estimates of reproduction numbers with application to COVID-19

Stefan Heyder¹, Thomas Hotz¹, Tyll Krüger², Jan Pablo Burgard³

1 - TU Ilmenau, Germany; 2 - Wrocław University of Science and Technology, Poland; 3 - Freie Universität Berlin, Germany

In the last year many public health decisions were based on real-time monitoring the spread of the ongoing COVID-19 pandemic. For this one often considers the reproduction number which measures the amount of secondary cases produced by a single infectious individual. While estimates of this quantity are readily available on the national level, subnational estimates, e.g. on the county level, pose more

difficulties as there are only few incidences. As countermeasures to the pandemic are usually enforced on the subnational level, such estimates are of great interest to assess the efficacy of the measures taken, and to guide future policy.

We present a novel extension of the well established estimator [1] of the country level reproduction number to the county level by applying techniques from small-area estimation. This new estimator yields sensible estimates of reproduction numbers both on the country and county level. It can handle low and highly variable case counts on the county level, and may be used to distinguish local outbreaks from more widespread ones. We demonstrate the capabilities of our novel estimator by a simulation study and by applying the estimator to German case data.

Reference

[1] C. Fraser. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic, *PloS one*, 2(8), e758, 2007.

SME2

84 Specifics of analyzing educational data

Sven Hilbert

Universität Regensburg, Germany

Modeling in the educational sciences is constantly evolving along with statistics, data science, and methodology development in general. Even though, the field is prone to profit from all sorts of advancements, some are of particular relevance for educational research, such as in multilevel modeling, decision theory, latent variable models, and replicability. Combining these elements with a focus on questions and directions from the educational sciences will further advance the field and pave the way for reliable empiric research. Moreover, analytical approaches and models have to be adapted to meet the specific needs of educational research. While log-data analysis constitutes a standard procedure in data science, learning platforms and data derived from devices in the classroom require additional steps due to privacy requirements and the specific needs of their users. Part of these needs is the interaction of different age-groups with the platform, which has to be adequately modeled, and the user feedback, which requires a minimum of data literacy to be effective. Finally, in the educational context, decisions for single-cases have to be focused to use data analytic tools for a range of practical purposes.

SBES

85 The Covid-19 pandemic as a driver of change in official statistics: New roles of the Federal Statistical Office illustrated by the Covid-19 data hub

Angelina Hofacker, Markus Zwick

Statistisches Bundesamt, Germany

As a customer-oriented information provider and digital data manager, the Federal Statistical Office (Destatis) wants to exploit the potential of official and non-official data that are available and extend the range of data it is offering in order to close data gaps in the statistical landscape. Destatis also wishes to improve the data literacy, or statistical literacy, of its users because data as an information basis relying on facts can make debates within society more objective.

A practical example: In the course of the Covid-19 pandemic, we have increased the range of data we offer in specific areas by providing, for instance, current and regional data on infections, the situation in hospitals, mobility and the measures taken by governments to contain the pandemic. Here Destatis draws upon existing data stocks compiled by the market and social research institute Infas on behalf of the Federal Ministry for Economic Affairs and Energy (BMWi), which have so far been published only for research purposes and are now made available to the public.

This is an example of the questions Destatis has to answer again and again to be able to meet emerging data needs and user expectations more rapidly and flexibly. (What data are there? What is the quality of the data available, especially with regard to comparability, harmonisation? What sources are reliable? How do the data change in the course of the pandemic?)

SLSE2

86 Statistical Challenges in the Quality Assurance of Healthcare

Michael Höhle, Jona Cederbaum

Federal Institute for Quality Assurance and Transparency in Healthcare (IQTIG) / Biometrics Unit, Berlin, Germany

The Federal Institute for Quality Assurance and Transparency in Healthcare (IQTIG) is the central institution in Germany for the statutory quality assurance in health care. In accordance with its statutes, it is scientifically independent and works for, in particular, The Federal Joint Committee providing its expertise in various tasks of quality assurance of medical care. Its main responsibilities include the evaluation of the quality of care provided by healthcare providers on the basis of quality indicators. The IQTIG develops and manages quality criteria in different clinical areas, such as: gynaecological operations, perinatology, cardiac pacemaker implantations, hip joint endoprosthesis operations and transplants.

In this talk, we give a short introduction to the IQTIG, what data sources we work with and the type of statistical methodology we apply in our work. We illustrate this through two examples: The first example is on the use of statistical

process control techniques to identify providers, which do not meet requirements. A particular challenge is the quantitative-qualitative nature of the decision-making process. The second example is the development of statistical methodology in order to analyse patient surveys. We end the talk by a discussion of how we try to achieve transparency about the statistical methods and analyses used in our context.

References

[1] Hengelbrock J., Rauh J., Cederbaum J., Kähler M., Höhle M. (2021), Hospital profiling using Bayesian decision theory. <https://doi.org/10.1101/2021.06.23.21259367>

[2] IQTIG (2018), Ch. 8 & 12, Entwicklung von Patientenbefragungen im Rahmen des Qualitätssicherungsverfahrens Perkutane Koronarintervention Koronarangiographie, Abschlussbericht

[3] IQTIG (2021), Methodische Grundlagen Entwurf für Version 2.0

IQW/TIG

87 Estimation of effect heterogeneity in rare events meta-analysis

Heinz Holling¹, Katrin Jansen¹, Dankmar Böhning²

1 – Westfälische Wilhelms-Universität Münster, Germany; 2 – Math. Sciences and Southampton Stat. Sciences Research Institute, University of Southampton, UK

In this talk, we present several approaches for dealing with meta-analyses of count outcome data. These counts are the accumulation of occurred events and these events might be rare, so a special feature of the meta-analysis is dealing with low counts including zero-count studies. Emphasis is put on approaches which are state-of-the-art for count data modelling including mixed log-linear (Poisson) and mixed logistic (binomial) regression as well as nonparametric mixture models for count data of Poisson and binomial type. A simulation study investigates the performance and capability of discrete mixture models in estimating effect heterogeneity. The approaches are exemplified on a meta-analytic case study investigating the acceptance of bibliotherapy.

MA1

88 Weighted Generalized Estimating Equations for Longitudinal Binary Response: Prevalence Estimation of Health Limitations in the SHARE Study

Anastasiia Holovchak¹, Ruben Wisskott¹, Ralph Brinks², Annika Hoyer³

1 – Department of Statistics, Ludwig-Maximilians-University Munich, Germany; 2 – Chair for Medical Biometry and Epidemiology, University Witten-Herdecke, Germany; 3 – Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Germany

Based on data from the Study of Health, Ageing and Retirement in Europe (SHARE) the aim of this work is to estimate the prevalence of health limitations due to chronic diseases together with the corresponding standard errors while properly accounting for the longitudinal structure of the data. To account for the presence of non-ignorable missingness and to obtain unbiased estimators of the population-specific prevalence of health limitations, we propose to use a marginal longitudinal model. For this, we modeled binary correlated responses for estimating the prevalence of health limitations by solving weighted generalized estimating equations with flexible non-linear effects. We used the inverse probability of an individual to be still part of SHARE, i.e. to be not lost to follow-up, as weight. They are estimated using a discrete-time survival model. Two different methods are applied to determine the standard errors of prevalence of health limitations. First, we use the delta method, which is based on an asymptotic normal distribution of the regression coefficient estimator. Second, we use the blockwise bootstrap algorithm on weighted subject-clustered observations, which allows applying flexible non-linear models. We found that the prevalence of health limitations is markedly increasing over calendar time indicating that as people live longer, they also spend more time with limitations.

SME1

89 Penalized Non-Linear Principal Components Analysis for Ordinal Data

Aisouda Hoshiyar¹, Henk A. L. Kiers², Jan Gertheiss¹

1 – School of Economics and Social Sciences, Helmut Schmidt University, Hamburg, Germany; 2 – Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands

Ordinal data occur frequently in the social sciences. When applying principal components analysis (PCA), however, those data are often treated as numeric implying linear relationships between the variables at hand, or non-linear PCA is applied where the obtained quantifications are sometimes hard to interpret. Non-linear PCA for categorical data, also called optimal scoring/scaling, constructs new variables by assigning numerical values to categories such that the proportion of variance in those new variables that is explained by a predefined number of principal components is

maximized. We propose a penalized version of non-linear PCA for ordinal variables that is an intermediate between standard PCA on category labels and non-linear PCA as used so far. The new approach is by no means limited to monotonic effects, and offers both better interpretability of the non-linear transformation of the category labels as well as better performance on validation data than unpenalized non-linear PCA and/or standard linear PCA. The methods are used on simulated and publicly available data sets, with a special focus on the International Classification of Functioning, Disability and Health (ICF). Further information on the methodology and a detailed analysis can be found at <https://arxiv.org/pdf/2110.02805.pdf>. Finally, we introduce the R package `ordPens` (Gertheiss and Hoshiyar, 2021), which implements the algorithm for penalized non-linear PCA and offers – among various other methods for ordinal data analysis – cross-validation to choose a proper amount of regularization.

Reference

[1] Gertheiss, J. and A. Hoshiyar (2021). `ordPens`: Selection, Fusion, Smoothing and Principal Components Analysis for Ordinal Variables. R package version 1.0.0, <https://CRAN.R-project.org/package=ordPens>

LVM

90 Predicting COVID-19 Hospitalisation from Incidences

Thomas Hotz, Stefan Heyder

TU Ilmenau, Germany

The COVID-19 pandemic endangers not only those getting infected but society as a whole by threatening a collapse of its health care system. Monitoring hospitalisation of COVID-19 cases is therefore of the utmost importance, leading to the introduction of the so-called "hospitalisation incidence" as a key indicator in Germany determining which counter measures are to be taken. However, due to its specification as the count of new cases on a certain day which will be hospitalised any time in the future, this indicator is heavily affected by delays not only due to reporting but also due to the time it takes to develop severe symptoms leading to hospitalisation. This calls for a prediction of the number of hospitalised cases out of the reported ones. We propose to estimate the proportion of such cases stratified by age using the most recent historic data, and to use those proportions to predict the hospitalised cases to come. Uncertainty of the predictions is quantified based on their performance in the past. Since November 2021, these predictions are being contributed to and, according to a preregistered study protocol, evaluated by the KIT's NowcastHub [1]. We will report on the performance of the proposed predictor. Moreover, we will discuss the usefulness of the predictions, and of the indicator itself, for monitoring the disease, in particular in comparison to other indicators such as the number of cases hospitalised on a given day.

Reference

[1] <https://covid19nowcasthub.de/>

SME2

91 Subgroup identification in individual participant data meta-analysis using model-based recursive partitioning

Cynthia Huber¹, Norbert Benda^{2,1}, Tim Friede¹

1 – Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany; 2 – Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany

Model-based recursive partitioning (MOB) can be used to identify subgroups with differing treatment effects. The detection rate of treatment-by-covariate interactions and the accuracy of identified subgroups using MOB depend strongly on the sample size. Using data from multiple randomized controlled clinical trials can overcome the problem of too small samples. However, naively pooling data from multiple trials may result in the identification of spurious subgroups as differences in study design, subject selection and other sources of between-trial heterogeneity are ignored. In order to account for between-trial heterogeneity in individual participant data (IPD) meta-analysis random-effect models are frequently used. Commonly, heterogeneity in the treatment effect is modelled using random effects whereas heterogeneity in the baseline risks is modelled by either fixed effects or random effects. In this article, we propose metaMOB, a procedure using the generalized mixed-effects model tree (GLMM tree) algorithm for subgroup identification in IPD meta-analysis. Although the application of metaMOB is potentially wider, e.g. randomized experiments with participants in social sciences or preclinical experiments in life sciences, we focus on randomized controlled clinical trials. In a simulation study, metaMOB outperformed GLMM trees assuming a random intercept only and model-based recursive partitioning (MOB), whose algorithm is the basis for GLMM trees, with respect to the false discovery rates, accuracy of identified subgroups and accuracy of estimated treatment effect. The most robust and therefore most promising method is metaMOB with fixed effects for modelling the between-trial heterogeneity in the baseline risks.

MA1

92 Investigating model adequacy, predictor effects and higher-order interactions for machine-learning models

Catherine Hurley

Maynooth University, National University of Ireland, Ireland

Machine learning models are well-known to be high on performance and low on interpretability. This talk describes some techniques we have developed for visualization of these models, which help address the interpretability deficit. We describe improved methods for exploring predictor importance, predictor interaction and partial dependence model summaries. We use interactive visualisation to dig deeper into model fits by focusing on slices of predictor space, thus investigating local lack of fit, local predictor effects and higher-order interactions. Our techniques are model agnostic and are appropriate for any regression or classification problem. The

methods presented are implemented in the R packages `vivid` and `condvis2`.

References

- [1] Catherine B. Hurley, Mark O’Connell & Katarina Domijan (2021) Interactive Slice Visualization for Exploring Machine Learning Models, *Journal of Computational and Graphical Statistics*, <https://doi.org/10.1080/10618600.2021.1983439>
- [2] Alan Inglis, Andrew Parnell & Catherine B. Hurley (2022) Visualizing Variable Importance and Variable Interaction Effects in Machine Learning Models, *Journal of Computational and Graphical Statistics*, <https://doi.org/10.1080/10618600.2021.2007935>

VEDA

93 Asking respondents to do more than answer survey questions

Annette Jäckle

University of Essex, United Kingdom

Social surveys primarily collect data by asking respondents questions. The data that can be collected in this way are limited to information respondents know, can recall, and are willing to report; the level of detail is limited by the number of questions that can feasibly be asked in an interview. To expand the quality and range of data collected, social surveys are increasingly asking respondents to do more than answer questions. In this presentation I will examine what determines whether respondents participate in additional tasks for a survey, using data from the UK Household Longitudinal Study: Understanding Society. Over the years we have asked respondents to download mobile apps and use them to collect additional data, to consent to link their survey data with government administrative data, to provide blood samples, hair samples, and COVID-19 tests, and to answer additional monthly questions between the annual interviews. The results from these studies suggest that the respondent’s decision whether to participate not only depends on the nature of the task (e.g. the topic or content of what is being linked/asked/collected) and the protocols for implementing that task (e.g. incentives for the task, placement and wording of the request, offering feedback), but also on the features of the survey in which the invitation to the additional task is made. This suggests a need to design the initial interview in a way that respondents view the additional task as being within the remit of what they have agreed to do for the survey.

SM1

94 Supporting COVID-19 quarantine decisions with a statistical risk assessment model

Sonja Jäckle¹, Elias Röger², Volker Dicken³, Benjamin Geisler³, Jakob Schumacher⁴, Max Westphal³

1 – Fraunhofer Institute for Digital Medicine MEVIS, Lübeck, Germany; 2 – Fraun-

hofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany; 3 – Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany; 4 – Health Department Berlin-Reinickendorf, Berlin, Germany

In Germany, the local health departments are responsible for managing of COVID-19 cases. Direct contacts of infectious persons at group meetings have to be traced and potentially quarantined. Such quarantine requirements may be revoked after certain days, when all contact persons obtain a negative test result. However, contact tracing and testing is time-consuming, costly and not always feasible. We present a statistical model for the probability that the primary infected person has not infected any contact person given a number of negative PCR-test results among contact persons [1]. A parametric Bayesian model is used which combines an adaptable Beta-Binomial prior and two likelihood components in a novel fashion. For the likelihood, the time-dependent sensitivity of the test (e.g. polymerase chain reaction, rapid antigen) are utilized. The prior can be adapted to specific situations such as school classes or children daycare. Furthermore, an extension for more than one primary infected persons is presented. This model is integrated into a web application developed for German health departments and will be part of an R package. Our model was tested with a real-world dataset and the result showed that our approach can support decisions whether the quarantine can be reduced. For future work, a broad study with health departments is planned to evaluate quarantine decisions based on our tool.

Reference

[1] Jäckle, S.; Röger, E.; Dicken, V.; Geisler, B.; Schumacher, J.; Westphal, M. A Statistical Model to Assess Risk for Supporting COVID-19 Quarantine Decisions. *Int. J. Environ. Res. Public Health* 2021, 18, 9166

BS2

95 Multi-step estimators of between-study variances and covariances and their relationship with the Paule-Mandel estimator

Dan Jackson¹, Robbie C. M. van Aert²

1 – AstraZeneca, United Kingdom; 2 – Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

Moment based estimation methods for the between-study variance are well established in univariate meta-analysis. Two-step moment based estimators have been proposed as a way to improve their performance, whilst retaining computational and conceptual simplicity. Multi-step estimators are a further natural extension proposed by Aert and Jackson (2018), who also establish a new relationship with the Paule-Mandel estimator. This talk will review their main findings but our focus will be the multivariate case. We will begin by extending the generalised method of moments estimator, so that it is applicable to multivariate meta-regression, and so derive a new class of estimators of the between-study covariance matrix. We will then show how these new estimators can be used to motivate the use of multivariate

multi-step estimators. We will tentatively propose that their limiting form (where the number of steps becomes large) may be presented as Paule-Mandel type estimators. We will illustrate our methodology using two examples and simulation study findings will also be summarised.

Reference

[1] Aert and Jackson (2018) Multistep estimators of the between-study variance: The relationship with the Paule-Mandel estimator", *Statistics in Medicine* 37(17), 2616-2629.

MA1

96 Development of prediction tools for health events from multiple longitudinal predictors

Helene Jacqmin-Gadda^{1,2}, Reza Hashemi³

1 – Université de Bordeaux, Bordeaux, France; 2 – Inserm, U1219- Bordeaux Population Health, Bordeaux, France; 3 – Department of Statistics, Razi University, Kermanshah, Iran

The development of precision medicine relies heavily on the availability of individual prediction tools for clinical events. Dynamic prediction models based on repeated measures of predictors (markers or risk factors) are especially useful as they make it possible to update the individual prediction at each new measure of one of the markers. Dynamic prediction tools are built using either joint models for time-to-events and repeated measures of longitudinal variables or the landmark approach. Joint models have some advantages as they are estimated using all the available information from the learning data set and can then be used for prediction at any time. In contrast, in the landmark approach, a model should be estimated at each landmark time (time of prediction) using only past measures of the markers. However, estimation of joint models with multiple longitudinal variables is computationally challenging limiting the use of this approach to predict the risk of health events using a large set of longitudinal markers. In this talk, we made an overview of previously proposed solutions for developing prediction tools based on multiple longitudinal predictors. Then, we describe a new model averaging approach for computing individual predictions based on repeated measures of several markers from estimated joint models including only one marker.

SME1

97 Approximately Linear INGARCH Models for Spatio-Temporal Counts

Malte Jahn¹, Christian Weiß¹, Hee-Young Kim²

1 – Helmut Schmidt University Hamburg, Germany; 2 – Korea University, Sejong,

South Korea

Models for count time series are often defined as counterparts to the well-known ARMA models, such as INARMA and INGARCH models (referred to as BINGARCH models if bounded counts), both having an ARMA-like ACF. Here, we are concerned with multivariate count data resulting from a spatio-temporal process. Although there also exist a few spatio-temporal modifications of INARMA models, most contributions are related to the more flexible INGARCH models. Among others, some spatio-temporal INGARCH models for unbounded counts of infectious diseases have been proposed in the literature, where a conditionally linear structure is particularly relevant. But exactly linear model specifications go along with limiting parameter constraints that prevent negative parameter and ACF values. Thus, we consider two recent classes of approximately linear (B)INGARCH models, which allow for negative ACF values while (nearly) preserving the linear model structure. We combine the spatio-temporal and the approximately-linear (B)INGARCH approaches in a unique way to obtain a class of rather flexible spatio-temporal (B)INGARCH-type models. We discuss their stochastic properties and analyze the finite-sample performance of maximum likelihood estimation. The general need for modeling bounded spatio-temporal counts is illustrated with a couple of real applications from climatology. An example on the cloud coverage of the sky demonstrates that solutions for dealing with missing data are crucial. When modeling hourly precipitation, we have to account for zero-inflation, and data on daily measurements of precipitation are used to illustrate how additional cross-correlation might be included into our novel model families.

STS

98 Supervised learning with missing values

Julie Josse

INRIA, France

An abundant literature addresses missing data in an inferential framework: estimating parameters and their variance from incomplete tables. Here, I will review recent works on supervised-learning settings: predicting a target when missing values appear in both training and testing data.

First, we study the seemingly-simple case where the target to predict is a linear function of the fully-observed data and show that multilayer perceptrons with ReLU activation functions can be consistent but highly complex. Based on an Neumann series approximation of the optimal predictor, we propose a new principle architecture, called NeuMiss networks. Their originality and strength come from the use of a new type of non-linearity: the multiplication by the mask. We provide an upper bound of the Bayes risk of NeuMiss networks, and we show that they have a good predictive accuracy. Then, we go beyond the linear regression setting and show how imputation can be coupled with powerful learners such as random forest to achieve consistency.

CSASOFT1

99 Prediction of hepatotoxicity based on gene expression data

Franziska Kappenberg¹, Wiebke Albrecht², Jan G. Hengstler², Jörg Rahnenführer¹

1 – TU Dortmund, Germany; 2 – IfADo, Germany

Drug induced liver injury (DILI) is an important concern in drug development. The risk of DILI depends on the dose in which a drug is administered and thus on the maximal blood concentration after intake. We present an in vitro/in silico system which has the following two objectives:

1. Prediction of the DILI risk based on an in vitro concentration and the maximal blood concentration
2. Estimation of the acceptable maximal blood concentration based on an in vitro concentration and the hepatotoxicity status of the compound.

For this, the statistical measures “Toxicity Separation Index” and “Toxicity Estimation Index” were introduced in Albrecht et al. (2019), which quantify these two goals for a set of compounds.

The performance of these indices depends on the chosen in vitro concentration. To determine this concentration, a combination of cytotoxicity and high-dimensional gene expression data is considered: So-called “alert concentrations”, often defined as the smallest concentration where some pre-defined effect in the response is observed, are determined for the cytotoxicity and the gene expression data. Different combinations of these alerts serve as in vitro concentration for the calculation of the indices. We propose different methods to combine results from high-dimensional gene expression data in a meaningful way in order to optimize the prediction of the hepatotoxicity status with the indices.

Reference

[1] Albrecht, W., Kappenberg, F., Brecklinghaus, T., et al. (2019). Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood concentrations. *Archives of toxicology*, 93(6), 1609–1637. <https://doi.org/10.1007/s00204-019-02492-9>

SHDA2

100 Scalable Gaussian-Process Inference Using Vecchia Approximations

Matthias Katzfuss

Texas A&M University, USA

Gaussian processes (GPs) are popular, flexible, and interpretable probabilistic models for functions in geospatial analysis, computer-model emulation, and machine learning. However, direct application of GPs involves dense covariance matrices and is computationally infeasible for large datasets. We consider a framework for fast GP inference based on the so-called Vecchia approximation, which implies a sparse

Cholesky factor of the inverse covariance matrix. The approximation can be written in closed form and computed in parallel, and it includes many popular existing approximations as special cases. We discuss applications of the framework to noisy and non-Gaussian data, to emulation of computer experiments, and to nonlinear regression and variable selection. Further extensions allow nonparametric inference on the covariance matrix and even on nonlinear dependence structures, with applications in climate-model emulation and data assimilation.

STS

101 Bagged Value-at-Risk Forecast Combination

Ekaterina Kazak¹, Roxana Halbleib², Winfried Pohlmeier³

1 – University of Manchester, UK; 2 – University of Freiburg, Germany; 3 – University of Konstanz, Germany

Recent developments in the financial econometrics literature on joint scoring functions for Value-at-Risk and Expected Shortfall allow for consistent implementation of statistical tests based on the Model Confidence Set (MCS). MCS is shown to be an excellent tool for model comparison, both in-sample and out-of-sample. Another branch of literature focuses on the superior performance of convex forecast combinations, which often outperform stand-alone forecasting models. This paper combines both results and proposes a novel approach to a forecast combination of Value-at-Risk and Expected Shortfall based on the MCS. We exploit the statistical properties of bootstrap aggregation (bagging) and combine competing models based on the bootstrapped probability of the model being in the Confidence Set. The resulting forecast combination allows for a exible and smooth switch between the underlying models and outperforms the corresponding stand-alone forecasts.

SiF2

102 Learning the Joint Distribution with Missing Data under the Gaussian Copula Model

Maximilian Kertel^{1,2}, Markus Pauly²

1 – BMW Group, Germany; 2 – Technische Universität Dortmund, Germany

Missing data is an obstacle for many statistical analyses in practice. If the researcher is interested in the underlying joint distribution of an experiment and data is Missing at Random, it is popular to assume a multivariate normal distribution and to apply the corresponding Expectation Maximization (EM) algorithm.

A generalization can be found in the Gaussian copula model for which various approaches in the literature apply the following steps: First, they estimate the marginal distributions using the observed data points only and derive transformations. Afterwards, the copula is found by deploying the EM algorithm for multivariate normality on the transformed data.

Our example reveals that this approach neither leads to consistent estimates for the marginals nor for the joint distribution. To overcome those difficulties, we present a rigorous application of the EM algorithm that alternates repeatedly between the estimation of the copula and the marginals and offers insights how both are interconnected. It comes at the price that a parametrization of the marginals must be chosen a priori. We propose to circumvent this restriction by using a semiparametric specification of the marginals.

A simulation study indicates that our method finds better estimates of the underlying joint distribution than the existing approach. The estimated distribution can also be used for inference about the marginals and to sample fill-ins for Multiple Imputation.

YSS1

103 Anomaly detection based on MOSUM statistics in large image data

Claudia Kirch¹, Philipp Klein¹, Marco Meyer^{1,2}

1 – Otto-von-Guericke-University Magdeburg, Germany; 2 – TU Braunschweig, Germany

Being able to identify anomalies in image data is important in many practical applications. This includes the detection of cancerous cells in medicine, obstacles in autonomous driving and cracks in the construction material of buildings. Based on the idea of moving sum (MOSUM) statistics, we develop a pre-processing procedure for the detection of cracks in 2D image data. In particular, based on the observation of geometric properties of cracks, we use a combination of moving semicircle- and rectangle-shaped windows in order to identify regions that potentially have cracks. Under the null hypothesis of no anomalies for image data in dimension P , we are able to show convergence for MOSUM scan statistics using a general class of sets towards a functional of P -parameter Gaussian processes. Particularly, this class includes 2D convex sets which we use for crack detection.

TSA1

104 Marginally calibrated response distributions for end-to-end learning in autonomous driving

Nadja Klein, Clara Hoffmann

Humboldt-Universität zu Berlin, Germany

End-to-end learners for autonomous driving are deep neural networks that predict the instantaneous steering angle directly from images of the ahead-lying street. These learners must provide reliable uncertainty estimates for their predictions in order to meet safety requirements and initiate a switch to manual control in areas of high uncertainty. Yet end-to-end learners typically only deliver point predictions, sin-

ce distributional predictions are associated with large increases in training time or additional computational resources during prediction. To address this shortcoming we investigate efficient and scalable approximate inference for the implicit copula neural linear model of Klein, Nott and Smith (2021) in order to quantify uncertainty for the predictions of end-to-end learners. The result are densities for the steering angle that are marginally calibrated, i.e. the average of the estimated densities equals the empirical distribution of steering angles. To ensure the scalability to large n regimes, we develop efficient estimation based on variational inference as a fast alternative to computationally intensive, exact inference via Hamiltonian Monte Carlo. We demonstrate the accuracy and speed of the variational approach in comparison to Hamiltonian Monte Carlo on two end-to-end learners trained for highway driving using the comma2k19 data set. The implicit copula neural linear model delivers accurate calibration, high-quality prediction intervals and allows to identify overconfident learners. Our approach also contributes to the explainability of black-box end-to-end learners, since predictive densities can be used to understand which steering actions the end-to-end learner sees as valid.

DS

105 Bayesian variable selection for non-Gaussian responses: a marginally calibrated copula approach

Nadja Klein¹, Michael Stanley Smith²

1 – Humboldt-Universität zu Berlin, Germany; 2 – Chair of Management (Econometrics), Melbourne Business School, The University of Melbourne, Australia

We propose a new highly flexible and tractable Bayesian approach to undertake variable selection in non-Gaussian regression models. It uses a copula decomposition for the joint distribution of observations on the dependent variable. This allows the marginal distribution of the dependent variable to be calibrated accurately using a nonparametric or other estimator. The family of copulas employed are “implicit copulas” that are constructed from existing hierarchical Bayesian models widely used for variable selection, and we establish some of their properties.

Even though the copulas are high dimensional, they can be estimated efficiently and quickly using Markov chain Monte Carlo. A simulation study shows that when the responses are non-Gaussian, the approach selects variables more accurately than contemporary benchmarks. To illustrate the full potential of our approach, we extend it to spatial variable selection for fMRI. Using real data, we show our method allows for voxel-specific marginal calibration of the magnetic resonance signal at over 6000 voxels, leading to an increase in the quality of the activation maps.

IBSDSP

106 Scalable Estimation for Structured Additive Distributional Regression Through Variational Inference

Jana Kleinemeier, Nadja Klein

Humboldt-Universität zu Berlin, Germany

Structured additive distributional regression models provide a very general framework to estimate complete conditional distributions by relating all parameters of a parametric distribution to covariates through additive predictors. The flexibility of these models allows to fully utilize information in big and complex data but typically also induces highly-parameterized models with many unknowns. Thus, the models are often difficult and costly to estimate using standard approaches such as fully Bayesian approaches based on Markov chain Monte Carlo methods. To overcome these issues, we suggest a fast and scalable alternative based on variational inference. Our approach combines a parsimonious parametric approximation for the posteriors of regression coefficients, with the exact conditional posterior for hyperparameters. In order to minimize the Kullback-Leibler divergence between the true posterior and the approximation we use a stochastic gradient ascent method combined with an efficient way to reduce the variance of the estimators. We investigate global and local annealing to make estimation even more robust, especially against data outliers. Our implementation is very general so that any parametric one-dimensional response distribution can be estimated, while allowing for different functional effects such as penalized splines or more complex tensor product interactions. In a simulation study we prove the efficacy of the new method on four different data sets with respect to accuracy and computation time. Finally, we benchmark our approach in two illustrating examples that deal with outlier detection in brain activity and the challenging task of modeling daily precipitation in Germany.

YSS1

107 Component-wise L2-boosting for polygenic risk scores based on large cohort data

Hannah Klinkhammer^{1,2}, Christian Staerk¹, Carlo Maj², Peter Krawitz², Andreas Mayr¹

1 – Institute for Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Germany; 2 – Institute for Genomic Statistics and Bioinformatics University Hospital Bonn, Germany

Polygenic risk scores (PRS) evaluate the individual genetic liability to a certain trait and play an increasingly important role in the field of risk stratification. Most often, PRS models are based on summary statistics of univariate effects derived from genome-wide association studies. To improve the prediction performance of PRS, it is desirable to fit multivariable models directly on the genotypic data. Due to the large and high-dimensional data, efficient algorithms have to be developed to overcome the computational burden.

We implemented a component-wise L2-boosting algorithm to fit genotyped data

from large cohort studies to continuous outcomes using the genotypic variants as linear base-learners. Similar to the `snpnet` approach for the lasso [1] we iteratively work on smaller batches of variants. By restricting the set of possible base-learners in a data-driven way in each boosting step, we are able to increase the computational efficiency without losing prediction accuracy. Furthermore, we show both in a simulation study and via different traits from the UK biobank data, that our method yields competitive results in comparison to other methods such as the lasso. Due to the modular structure of boosting, the method can be extended to construct PRS for different outcomes such as binary or time-to-event data.

Reference

[1] Qian J. et al. (2020) A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the UK Biobank. *PLoS 16*(10): e1009141. <https://doi.org/10.1371/journal.pgen.1009141>.

SHDA2

108 Rage Against the Mean - A Review of Distributional Regression Approaches

Thomas Kneib, Alexander Silbersdorff, Benjamin Säfken

Georg-August-Universität Göttingen, Germany

Distributional regression models that overcome the traditional focus on relating the conditional mean of the response to explanatory variables and instead target either the complete conditional response distribution or more general features thereof have seen increasing interest in the past decade. The current state of distributional regression will be discussed, with a particular focus on the four most prominent model classes: (i) generalized additive models for location, scale and shape, (ii) conditional transformation models and distribution regression, (iii) density regression, and (iv) quantile and expectile regression. Characteristics of the different distributional regression approaches will be provided to establish a structured overview on the similarities and differences with respect to the required assumptions on the conditional response distribution, theoretical properties, and the availability of software implementations. In addition, challenges arising in the interpretability of distributional regression models will be discussed and all four approaches will be illustrated with an application analyzing determinants of income distributions from the German Socio-Economic Panel (GSOEP).

ARM1

109 Robust confidence intervals for mixed-effects meta-regression with interaction

Eric Samuel Knop

Department of Statistics, TU Dortmund University, Germany

The power of statistical tests and confidence intervals in mixed-effects meta-regression models depends strongly on the method used for the variance estimation of the parameters of interest. Since standard approaches perform poorly when basic assumptions like homogeneity or normality are violated, robust methods are required. In previous research such methods were compared to the commonly used Knapp-Hartung method for models with one moderator. The Knapp-Hartung method usually turned out to be the best, even when those basic assumptions were violated.

In this work, the performance of confidence intervals based on the Knapp-Hartung method is compared to confidence intervals based on robust heteroscedasticity consistent methods in a model where an interaction between two moderators is present. By a simulation study it could be shown that in the absence of normality the Knapp-Hartung method reaches its limits and a heteroscedasticity consistent method might be more adequate.

IBSDRP

110 Score-based calibration testing for multivariate forecast distributions

Malte Knüppel¹, Fabian Krüger², Marc-Oliver Pohle³

1 – Deutsche Bundesbank, Germany; 2 – Karlsruhe Institute of Technology, Germany; 3 – Goethe University Frankfurt, Germany

Multivariate distributional forecasts are becoming increasingly popular due to demands to go beyond point forecasts and to characterize dependencies between multiple variables. To assess the quality of those forecasts, suitable evaluation methods are needed. When it comes to absolute forecast evaluation, calibration tests based on the probability integral transform (PIT) are routinely used in the univariate case. However, multivariate extensions of PIT-based calibration tests face various challenges. We therefore propose two new approaches that arise from a generalized framework for calibration testing in the multivariate case. Both approaches use proper scoring rules and are simple to implement even in large dimensions. The first approach employs the PIT of the score. The second approach is based on comparing the expected performance of the forecast distribution (i.e. the expected score) to its actual performance based on realized observations (i.e. the realized score). We discuss the precise notion of calibration that is considered by the tests (auto-calibration) and propose a principled implementation. The tests have good size and power properties in simulations and solve various problems of extant tests. We illustrate the new tests using Bayesian (vector) autoregressive forecast distributions

for economic time series data.

TSA2

111 Variational Inference and Sparsity in High-Dimensional Deep Gaussian Mixture Models

Lucas Kock¹, Nadja Klein¹, David J. Nott²

1 – Humboldt-Universität zu Berlin, Germany; 2 – National University of Singapore, Singapore

Gaussian mixture models are a popular tool for model-based clustering, and mixtures of factor analyzers are Gaussian mixture models having parsimonious factor covariance structure for mixture components. There are several recent extensions of mixture of factor analyzers to deep mixtures, where the Gaussian model for the latent factors is replaced by a mixture of factor analyzers. This construction can be iterated to obtain a model with many layers. These deep models are challenging to fit, and we consider Bayesian inference using sparsity priors to further regularize the estimation. A scalable natural gradient variational inference algorithm is developed for fitting the model, and we suggest computationally efficient approaches to the architecture choice using overfitted mixtures where unnecessary components drop out in the estimation. In a number of simulated and two real examples, we demonstrate the versatility of our approach for high-dimensional problems, and demonstrate that the use of sparsity inducing priors can be helpful for obtaining improved clustering results.

DS

112 Interpreting Deep Neural Networks with the R Package innsight

Niklas Koenen^{1,2}, Marvin N. Wright^{1,2,3}

1 – Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany; 2 – Faculty of Mathematics and Computer Science, University of Bremen, Germany; 3 – Department of Public Health, University of Copenhagen, Denmark

In the last decade, deep neural networks have explosively spread and established themselves in almost all areas of research, industry and public life. However, for sensitive applications — besides predictive performance — an understanding of the black box decision process is essential to assess its reliability or to gain insights into the data. Several interpretability methods for neural networks have been proposed, e.g. feature attribution [1]. A major limitation so far is that most of these are Python-exclusive and not directly available in R. Further, it is unclear which method performs best in a given setting and no neutral comparison study is available.

We present the R package `innsight` (<https://github.com/bips-hb/innsight/>) providing the most common feature attribution methods in a unified framework. Con-

trary to the implementations in Python (e.g. iNNvestigate), our package works independently of the deep learning library the model was trained with. Additionally, it offers a rich set of visualization methods.

Assessing explanatory quality usually relies on visual impressions inspired by image data, or on perturbation-based [2] or axiomatic approaches [3]. All of these evaluation methods have in common that the ground truth is unknown and thus sensitivity and specificity cannot be assessed. We contribute to this active field of research with a statistical perspective: we conduct a proper simulation study with a known data generating process, with the long-term goal of providing a comprehensive neutral comparison study of neural network interpretability methods.

References

[1] Ancona et al. (2019) https://doi.org/10.1007/978-3-030-28954-6_9

[2] Samek et al. (2017) <https://doi.org/10.1109/TNNLS.2016.2599820>

[3] Montavon (2019) https://doi.org/10.1007/978-3-030-28954-6_13

CSASOFT2

113 AI enables Innovation – Safe and Secure AI is a must for many AI-based Applications

Frank Köster

DLR-Institute for AI Safety and Security, Sankt Augustin and Ulm, Germany

Artificial Intelligence is highly relevant for many application areas – it is one of the key technologies that paves the ground for innovative applications in the course of advancing digitalisation. This is also true for applications that have to fulfill high requirements in the areas of safety and security. Accordingly, the field of AI safety and security is more and more of relevance. Only a deeper investigation of safety and security in the context of AI-based technologies and applications can enable their comprehensive use for the benefit of the economy and society. In this talk, an approach to ensure high levels of safety/security will be sketched. In this context, we rely on extensive experiences in domains like aeronautics, space, transportation and energy etc. – thus the core-application domains des DLR. The approach integrates current safety/security-oriented methodologies and processes with approaches from the area of explainable and trustworthy AI. In addition to that, data infrastructures and future data- and service-ecosystems for AI-based applications will also be taken into account.

AIML1

114 What implications do analysis choices have on study results?

Linda Krause, Caroline Seifert, Antonia Zapf

University Medical Center Hamburg-Eppendorf, Germany

Which variables should we include in the regression analysis as adjusting variables? Using which criteria should we define the binary endpoint? These and more are questions we as statisticians often have to discuss when consulting clinicians who want to plan or analyse clinical trials or observational studies. In clinical trials, those decisions have to be made during writing of the study protocol before collection of any data or at latest during formulating the statistical analysis plan before database lock. In observational studies those choices are often made after data collection.

We aim to systematically investigate the implications of analysis choices on study results in the sense of a robustness analysis using one clinical trial and one cross-sectional study as examples. In the cross-sectional study we examine the influence of covariate sets to include in the regression analyses for adjustment. In the clinical trial we explore the impact of derivation of binary endpoints from numerical values. We inspect the implications on study results by comparing estimators and confidence intervals for all possible combinations of potential analysis choices. Depending on the respective study and hypothesis, we aim to combine the different results to obtain one answer to the initial medical questions or to show that such a combination is not sensible and why. In addition, we visualize all results in a joint manner to assist clinicians and collaborators in understanding the potential implications of analysis choices on study results and their influence on study interpretation.

VEDA

115 Estimation of ERGMs under Missing Attribute Data

Robert Wilhelm Krause¹, Mark Huisman²

1 – Freie Universität Berlin, Germany; 2 – University of Groningen, The Netherlands

In this talk, I will present latest work on the estimation of exponential random graph models (ERGM) under missing nodal attribute data. As of yet, missing data in a nodes attribute (e.g., age, race, gender...) makes estimation of the model difficult, for instance, the most commonly used function to estimate ERGMs (`ergm()` in R from the `ergm` package) does not run at all under missing attribute data. We propose to integrate a network and covariate dependent imputation of missing attributes within the ERGM estimation, which can be easily implemented within a Bayesian framework. The proposed function can also be used to obtain joined imputations of missing network data (edges) and missing nodal attribute values. These imputations are compared with stationary stochastic actor-oriented model based imputations of missing network and attribute data. The results suggest that the proposed algorithm provides reliable estimates of network parameters under missing attribute data.

NA

116 Analysing Opportunity Cost of Care Work using Mixed Effects Random Forests under Aggregated Census Data

Patrick Krennmair¹, Nora Würz¹, Timo Schmid²

1 – Freie Universität Berlin, Germany; 2 – Otto-Friedrich-Universität Bamberg, Germany

Reliable estimators of the spatial distribution of socio-economic indicators (like opportunity cost of care work) are essential for evidence-based policy-making. As sample sizes are small for highly disaggregated domains, the accuracy of the direct estimates is reduced. To overcome this problem small area estimation approaches are promising. In this work we propose a small area methodology using machine learning methods. The semi-parametric framework of mixed effects random forest combines the advantages of random forests (robustness against outliers and implicit model-selection) with the ability to model hierarchical dependencies. Existing model-based methods using random forest for small area estimation require access to auxiliary information on population-level. We present a methodology that deals with the lack of population micro-data. Our strategy adaptively incorporates auxiliary information through calibration-weights - based on empirical likelihood - for the estimation of area-level means. In addition to our point estimator, we provide a non-parametric bootstrap estimator measuring its uncertainty. Extensive model-based simulations show the advantages of the proposed point estimator and its uncertainty measure. Finally, the proposed methodology is applied to the 2011 Socio-Economic Panel and aggregate census information from the same year to estimate the average opportunity cost of care work for 96 regional planning regions in Germany.

OAS1

117 A powerful curve fitting model for approximation of the transient differential equations dynamics

Clemens Kreutz

Medical Center – University of Freiburg, Germany

In systems biology, ordinary differential equations (ODEs) are frequently applied and typically defined by translating relevant biochemical interactions into rate equations. A disadvantage for such *mechanistic* dynamic models is, that they become very large in terms of the number of dynamic variables and parameters if entire cellular pathways are described. Moreover, analytical solutions of the ODEs are not available and the dynamics is nonlinear which is a challenge for numerical approaches as well as for statistically valid reasoning.

I recently introduced a complementary modeling approach based on curve fitting of a tailored *retarded transient response function (RTF)* [1] which exhibits amazing capabilities in approximating ODE solutions in case of transient dynamics as it is typically observed for cellular signaling pathways. Besides the broad and easy applicability, a benefit of the RTF is the self-explanatory interpretations of the parameters as response time, as amplitudes, and time constants of a transient and

a sustained part of the response.

Nine benchmark problems for cellular signaling were analyzed to demonstrate the approach in realistic systems biology settings. The presented approach offers a data-driven alternative modelling strategy for situations where classical ODE modeling is cumbersome or even infeasible. Moreover, it enables valuable interpretations of traditional ODE models and is also applicable for the analysis of time course data of omics experiments which is also illustrated in this presentation.

Reference

[1] <https://doi.org/10.3389/fphy.2020.00070>

BSB1

118 Adaptive phase 2/3 drug development programs – The pros and cons

Cornelia Ursula Kunz

Boehringer Ingelheim Pharma GmbH & Co. KG, Germany

To accelerate the drug development process, recent statistical research has focused on adaptive designs including seamless Phase 2/3 and enrichment designs. The general idea is that these designs lead to a reduction in sample size or a gain in power as compared to classical designs as they use the data more efficiently. However, cases exist where the complexity of an adaptive design might not be worthwhile a small gain in power as compared to less complex fixed designs. Depending on the design specifics there might not even be a gain in power.

For example, seamless Phase 2/3 designs often incorporate many treatment arms in Phase 2 to not select a promising treatment arm but also to establish a dose-response relationship. In this setting, it is not uncommon to have four or even five treatment arms in Phase 2 while one or two arms will be carried forward to Phase 3. Using p-value combination tests will ensure control of the type I error rate. However, due to the many arms, the adjustments to be made are often leading to conservative test statistics. As a result, the overall power might be lower than the overall power for conducting separate Phase 2 and Phase 3 trials or a single-stage design comparing all treatment arms against control and without selecting an arm – assuming a fixed total sample size across both phases. Hence, when planning a trial, it might be worthwhile to compare different strategies in terms of overall power across the two phases.

DECT1

119 Flexible Specification Testing in Quantile Regression Models

Tim Kutzker¹, Nadja Klein¹, Dominik Wied²

1 – Humboldt-Universität zu Berlin, Germany; 2 – Universität zu Köln, Germany

We propose three novel consistent specification tests for quantile regression models which generalize former tests in three ways. First, we allow the covariate effects to be quantile-dependent and nonlinear. Second, we allow parameterizing the conditional quantile functions by appropriate basis functions, rather than parametrically. We are hence able to test for functional forms beyond linearity, while retaining the linear effects as special cases. In both cases, the induced class of conditional distribution functions is tested with a Cramér-von Mises type test statistic for which we derive the theoretical limit distribution and propose a bootstrap method. Third, to increase the power of the tests, we further suggest a modified test statistic.

We highlight the merits of our tests in a detailed MC study and two real data examples. Our first application to conditional income distributions in Germany indicates that there are not only still significant differences between East and West but also across the quantiles of the conditional income distributions, when conditioning on age and year. The second application to data from the Australian national electricity market reveals the importance of using interaction effects for modelling the highly skewed and heavy-tailed distributions of energy prices conditional on day, time of day and demand.

ARM4

120 Construction of artificial most representative trees by minimizing tree-based distance measures

Björn-Hergen Laabs¹, Ana Westenberger², Inke R. König¹

1 – Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Germany; 2 – Institute of Neurogenetics, University of Lübeck, Lübeck, Germany

Selecting most representative trees from a random forest is one way to improve the interpretability of the otherwise very complex ensemble. For this, pair-wise tree distances are estimated, and the tree with the minimal mean distance is selected to represent the forest. To decorrelate the trees within the random forest, at each node in a tree only a small subset of independent variables is used to select the optimal split point. This leads to the problem that each split is only a local and not necessarily a global optimum. Thus, the interpretation of the selected most representative trees is still challenging.

To overcome this issue, we developed an algorithm which generates an artificial most representative tree using a greedy approach selecting the split point at each node that minimizes the distance to all trees of the original random forest. It iteratively adds new splits to the artificial representative tree until no new split is found, which decreases the mean distance to the forest.

In an extensive simulation study, we compared artificial most representative trees

with those directly selected from the forest regarding their prediction performance, ability to condense the information of the ensemble and coverage of the meaningful predictors. Additionally, we applied both methods to a genetic data set of X-linked dystonian-parkinsonism (XDP) and evaluated the resulting most representative trees with regard to recent results on genetic modifiers of age at onset in XDP.

Finally, we added the new method to our existing R package `timbR` (<https://github.com/imbs-hl/timbR>).

AIML2

121 Restoration of Temporal Dependence in Statistical Post-processing of Ensemble Weather Forecasts

Mária Lakatos, Sándor Baran

University of Debrecen, Hungary

An influential step in weather forecasting was the introduction of ensemble forecasts in operational use due to their capability to account for the uncertainties in the future state of the atmosphere. However, ensemble weather forecasts are often under-dispersive and might also contain bias, which calls for some form of post-processing. A popular approach to calibration is the ensemble model output statistics (EMOS) resulting in a full predictive distribution for a given weather variable. However, this form of univariate post-processing may ignore the prevailing spatial and/or temporal correlation structures among different dimensions. Since many applications call for spatially and/or temporally coherent forecasts, multivariate post-processing aims to capture these possibly lost dependencies.

Our main objective is the comparison of different nonparametric multivariate approaches to modeling temporal dependence of ensemble weather forecasts with different forecast horizons. We investigate two-step methods, where after univariate post-processing, the EMOS predictive distributions corresponding to different forecast horizons are combined to a multivariate calibrated prediction using an (empirical) copula. Based on global ensemble predictions of the European Centre for Medium-Range Weather Forecasts from January 2002 to March 2014 we investigate the forecast skill of different versions of Ensemble Copula Coupling and Schaake Shuffle. In general, compared with the raw and independently calibrated forecasts, multivariate post-processing substantially improves the forecast skill; however, there is no unique winner, the best-performing approach strongly depends on the weather variable at hand.

Joint work with Sándor Baran, Sebastian Lerch, Stephan Hemri, Anette Möller and Roman Schefzik

SAES1

122 mlr3proba: A unified interface for machine learning with survival tasks

Michel Lang^{1,2}, Raphael Sonabend³, Franz Kiraly⁴, Andreas Bender², Bernd Bischl²

1 – TU Dortmund, Germany; 2 – LMU Munich, Germany; 3 – Imperial College London, UK; 4 – Shell Research UK, University College London

While there are many extensive comparative studies for regression problems and especially classification problems, systematic benchmarks of different modeling approaches in survival time analysis often still prove difficult. On the one hand, this is due to the large number of different measures and different interpretations, which require different information from the learned models. On the other hand, there is a substantial technical hurdle, since techniques for survival time analysis are either completely neglected or not incorporated tightly into existing machine learning frameworks.

Both issues are addressed by the mlr3proba package. The R package supports more than 20 survival time measures for discrimination, calibration and time prediction and assists in the extraction and conversion of different model predict types required for this purpose. This applies to classical survival regression models as well as to neural networks and tree-based models like random survival forests.

Due to the tight integration in the machine learning framework mlr3, numerous ML techniques become directly applicable to survival analysis tasks, e.g., parallelized resampling and benchmarking, variable selection using filters and wrappers, tuning, preprocessing, pipelining or ensembles.

The talk will cover a brief introduction to mlr3 and showcase (a) how convenient practical benchmark studies can be conducted with our package, and (b) how to tackle some more advanced but common problems in survival analysis.

CSASOFT2

123 Visualizing Goodness of Fit of Probabilistic Regression Models

Moritz N. Lang¹, Reto Stauffer^{1,2}, Achim Zeileis¹

1 – Department of Statistics, Universität Innsbruck, Austria; 2 – Digital Science Center, Universität Innsbruck, Austria

Modeling and forecasting in terms of entire probability distributions - as opposed to capturing only the mean of a certain variable - is of prime importance in many different disciplines from natural sciences to social sciences and beyond. Hence, likelihood-based models and distributional regression have been receiving increasing interest over the last decades. In order to assess the goodness of fit along with potential deficits of such models, graphical evaluations in terms of visualizations are an important complement to numerical evaluations (e.g., based on proper scoring rules).

Various types of diagnostic graphics have been suggested in the literature for this purpose, e. g., PIT (probability integral transform) histograms, Q-Q (quantile-quantile) plots of (randomized) quantile residuals, rootograms, reliability diagrams, and worm plots, among others. However, discussion in the literature and usage in practice is somewhat scattered across different scientific communities, different types of variables (continuous vs. discrete vs. categorical), and different software packages. To overcome these problems, we present the visualizations above in a common conceptual framework accompanied by a flexible and object-oriented software implementation in the R package `topmodels` (<https://topmodels.R-Forge.R-project.org/>). It is discussed how all these types of graphics can be understood as different strategies for visualizing either the marginal fit (on the original scale of the observations) or the conditional fit (on a probability or quantile scale). Using several empirical examples, it is highlighted which types of displays are particularly useful for bringing out which kind of model deficits.

VEDA

124 Evaluation of tree-based statistical learning methods for constructing genetic risk scores

Michael Lau^{1,2}, **Claudia Wigmann**², **Sara Kress**², **Tamara Schikowski**², **Holger Schwender**¹

1 – Heinrich Heine University Düsseldorf, Germany; 2 – IUF-Leibniz Research Institute for Environmental Medicine, Germany

Genetic risk scores (GRS) summarize genetic features, in particular single nucleotide polymorphisms (SNPs), in a single statistic with respect to a given trait such as a disease status. Typically, GRS are constructed by generalized linear models, e.g., fitting an elastic net. However, these linear models are not able to incorporate gene-gene interactions unless prior biological knowledge was supplied about which loci might interact. To address this shortcoming, tree-based statistical learning methods such as random forests and logic regression may be an alternative for constructing GRS due to their well-known ability of autonomously detecting interactions. In this work, we evaluate random forests, logic regression, and appropriate extensions of these statistical learning procedures in the construction of GRS and compare these methods to elastic net as reference procedure.

In an extensive simulation study and an application to a real data set from a German cohort study, we show that both tree-based approaches can outperform elastic net when constructing GRS for binary traits. In particular, the ensemble methods random forests and logic bagging could induce high predictive powers. Even when considering no epistatic interaction effects but only marginal genetic effects, elastic net leads in most cases to inferior results.

If little biological information about the relationship between the regarded genetic features and the disease trait is available or if it can be assumed that epistasis between possibly unknown SNPs is present, random forests and logic bagging, therefore, provide beneficial alternatives to standard procedures for constructing GRS.

YSS1

125 A Multivariate Perturbation Robust Test Against Spurious Long Memory

Vivien Less, Philipp Sibbertsen

Leibniz Universität Hannover, Germany

This paper introduces a multivariate extension to the local Whittle with noise estimator, as well as a modified score-type test against spurious long memory. The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle with noise estimator. By explicitly addressing the noise term when approximating the spectral density near the origin improves the efficiency of the estimator and the size and power properties of the test. We prove consistency and asymptotic normality of the local Whittle estimator and we derive the limiting distribution and show consistency of the procedure.

TSA2

126 Testing equality of spectral density operators for functional linear processes

Anne Leucht¹, Efstathios Paparoditis², Daniel Rademacher³, Theofanis Sapatinas²

1 – Universität Bamberg, Germany; 2 – University of Cyprus; 3 – Universität Heidelberg, Germany

In this talk, the problem of testing equality of the entire second order structure of two independent functional processes is considered. A fully functional L2-type test is developed which evaluates, over all frequencies, the Hilbert-Schmidt distance between the estimated spectral density operators of the two processes. Under the assumption of a linear functional process, the asymptotic behavior of the test statistic is investigated. Furthermore, a novel frequency domain bootstrap method is developed which leads to a more accurate approximation of the distribution of the test statistic under the null than the large sample Gaussian approximation derived. Asymptotic validity of the bootstrap procedure is established under very general conditions and consistency of the bootstrap-based test under the alternative is proved. Numerical simulations show that, even for small samples, the bootstrap-based test has a very good size and power behavior.

TSA3

127 Fast and Fair Simultaneous Confidence Bands for Functional Parameters

Dominik Liebl^{1,2}, Matthew Reimherr³

1 – Universität Bonn, Germany; 2 – Hausdorff Center for Mathematics, Bonn, Germany; 3 – Penn State University, USA

Quantifying uncertainty using confidence regions is a central goal of statistical inference. Despite this, methodologies for confidence bands in Functional Data Analysis are underdeveloped compared to estimation and hypothesis testing. This work represents a major leap forward in this area by presenting a new methodology for constructing simultaneous confidence bands for functional parameter estimates. These bands possess a number of striking qualities: (1) they have a nearly closed-form expression, (2) they give nearly exact coverage, (3) they have a finite sample correction, (4) they do not require an estimate of the full covariance of the parameter estimate, and (5) they can be constructed adaptively according to a desired criteria. One option for choosing bands we find especially interesting is the concept of fair bands which allows us to do fair (or equitable) inference over subintervals and could be especially useful in longitudinal studies over long time scales. Our bands are constructed by integrating and extending tools from Random Field Theory, an area that has yet to overlap with Functional Data Analysis.

MS1

128 Application of a hierarchical Bayesian model to determine the empirical distribution of the heterogeneity parameter in IQWiG reports

Jona Lilienthal¹, Sibylle Sturtz¹, Christian Röver², Tim Friede², Ralf Bender¹

1 – Institute for Quality and Efficiency in Health Care (IQWiG), Köln, Germany; 2 – University Medical Center Göttingen, Germany

The frequentist approach of meta-analysis often reaches its limit when few studies are available. Due to imprecise estimation of the heterogeneity parameter, the effect estimation according to Knapp-Hartung can lead to conservative confidence intervals. An alternative is the application of a Bayesian meta-analysis, but the result depends on the specification of the prior distribution for the heterogeneity parameter.

To determine an adequate prior distribution for the heterogeneity parameter, the empirical information of existing data can be used. This is done by extending the Bayes model to describe all available meta-analyses jointly (Turner et al., 2015). A parametric distribution for the heterogeneity parameter can then be derived from the results of the hierarchical Bayes model.

This methodology is applied to a dataset of meta-analyses performed in IQWiG

reports until 2020 to find an appropriate distribution of the heterogeneity parameter. The results are presented and the influence of different distribution families is analysed. Also, the resulting distributions of the heterogeneity parameter are compared for different effect measures, as well as to the distribution of heterogeneity estimates.

Reference

[1] Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., & Higgins, J. P. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6), 984-998.

IQW/TIG

129 Statistical Inference for Method of Moments Estimators of a Semi-Supervised Two-Component Mixture Model

Bradley Lubich, Daniel R. Jeske, Weixin Yao

UC Riverside, United States of America

A mixture of a distribution of responses from untreated patients and a shift of that distribution is a useful model for the responses from a group of treated patients. The mixture model accounts for the fact that not all the patients in the treated group will respond to the treatment and consequently their responses follow the same distribution as the responses from untreated patients. The treatment effect in this context consists of both the fraction of the treated patients that are responders and the magnitude of the shift in the distribution for the responders. In this paper, we investigate asymptotic properties of method of moment estimators for the treatment effect based on a semi supervised two-component mixture model and demonstrate their superior statistical inference performance compared to the computationally intensive bootstrap intervals and their Bias-Corrected versions.

STIuDSD

130 Properties of Bayesian meta-analyses in evidence syntheses of very few studies

Matthias Maiworm, Christoph Schürmann, Guido Skipka, Ralf Bender

IQWiG - Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Germany

According to its current methods paper, IQWiG applies frequentist procedures for conducting meta-analyses to summarize evidence of multiple studies. Effect estimation is preferably done using the approach of Knapp and Hartung (KH). The KH-method, however, has unfavourable properties in the case of very few studies, because an imprecise estimate of the heterogeneity parameter τ may result in overly wide confidence intervals for the treatment effect. Therefore, the final conclusion on

an effect is often drawn case-specific and by use of various assessment approaches. Bayesian methods of meta-analysis have been focused on with increased attention in recent years for the situation of very few studies. Our objective was to investigate if IQWiG benefit assessments could profit from employing these methods. Data of all meta-analyses presented in IQWiG reports until 2020 have been re-analysed according to the current standard on the one hand and a Bayesian analysis on the other hand. When performing Bayesian analyses, we tried various prior distributions for the heterogeneity parameter based on results in the literature and estimates obtained from our own data. Both narrow and wide distributions were considered. We compare the conclusions that would be drawn from the current approach with those from the Bayesian one using the various prior distributions under consideration. Finally, we make first suggestions on if and how Bayesian meta-analyses could play a part in future IQWiG assessments.

IQW/TIG

131 On the Detection of Changes in Spatio-Temporal GLMs for Count Data

Steffen Maletz¹, Konstantinos Fokianos², Roland Fried¹

1 – TU Dortmund, Germany; 2 – University of Cyprus

We construct models for spatio-temporal count time series based on the multivariate extensions of INGARCH and log-linear time series models of Fokianos et al. (2020). For this we replace the matrices describing the effects of the past on the current observations in the multivariate by sparsely parameterized linear combinations of a priori known matrices, which represent spatial neighbourhoods and dependencies, similar as in the STARMA models of Pfeifer and Deutsch (1980). In this way we achieve a strong parameter reduction, so that the models are even applicable in high-dimensional scenarios. We provide a Wald test and a score test, which can be used, for example, to retrospectively test for covariate or intervention effects. We investigate the usefulness of the tests in simulations.

STS

132 On the Detection of Changes in Spatio-Temporal GLMs for Count Data

Anna Malinovskaya

Leibniz University Hannover, Germany

The rapid advancement of models based on artificial intelligence demands innovative monitoring techniques which can operate in real time with low computational costs. In machine learning, especially if we consider neural networks, which are also known as "deep learning" algorithms, the models are often trained in a supervised manner

and thus highly rely on the training dataset. Consequently, the learned relationship between the input data and the output must remain valid during the model's deployment. If this stationarity assumption holds, we can conclude that a neural network can generate accurate predictions. Otherwise, the retraining or rebuilding of the model is inevitable. In the structure of a neural network, the vital operation before generating the output is to represent complex features of a data sample as a vector in a latent space. This new representation of the processed data (called "embedding") has a beneficial practical usage to determine the time point when the data stream starts being non-stationary, leading to the deviating performance of the model. In this case, retraining or careful evaluation of the current model's configuration becomes necessary. In this work, we propose to monitor embeddings applying multivariate statistical process control based on the calculation of the data depth and normalised ranks. The quality of the introduced method is evaluated by applying various neural networks with different underlying data formats.

AIML1

133 A didactic experiment on fishing expeditions

Maximilian Michael Mandl, Roman Hornung, Christina Nießl, Theresa Ullmann, Anne-Laure Boulesteix

Ludwig-Maximilians-Universität München, Germany

In the empirical sciences, different researchers working on the same research question may make different analytical choices, which can lead to a variety of results. These free choices in the statistical analysis process are generally known as researcher degrees of freedom. An interdisciplinary overview in relation to the replication crisis is given by Hoffmann et al. (2021). Experiments that empirically measure the variation between multiple individual teams analysing the same data have already been described in the literature. This variability of results arising from the researcher degrees of freedom provides the opportunity to selectively report the most impressive result, a practice known as "p-hacking" in the context of hypothesis testing, or more generally "fishing expeditions". We conduct an experiment (still ongoing) on a sample of $n = 27$ bachelor students studying statistics. The students each received a simulated dataset and were then asked to estimate the effect of a covariate of interest in a multivariable regression setting that required several analytical choices such as model selection, imputation of missing values, and handling of outliers. The procedure will be repeated following a didactic intervention that aims to increase the awareness of the issue. We are then able to not only quantitatively assess the magnitude of p-hacking in this setting, but also evaluate whether the intervention leads to a reduction in these questionable practices.

Reference

[1] Sabine Hoffmann, Felix Schönbrodt, Ralf Elsas, Rory Wilson, Ulrich Strasser, and Anne-Laure Boulesteix (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *Royal Society Open Science*, 8(4).

SLSE1

134 Penalized likelihood adaptive-LASSO algorithm for feature selection in functional HDGM

Paolo Maranzano¹, Philipp Otto², Alessandro Fassò¹

1 – University of Bergamo, Italy; 2 – Leibniz University of Hannover, Germany

We present a novel model selection algorithm based on a penalized maximum likelihood estimator (PMLE) for functional hidden dynamic geostatistical models (f-HDGM). This class of models employs a mixed-effect regression structure with embedded spatio-temporal dynamics to model geo-referenced data observed in a functional domain. Thus, all parameters are functions across this domain. The algorithm simultaneously selects the relevant set of spline functions and regressors used to model the fixed-effects relationship between the response variable and the covariates. In this way, it automatically shrinks irrelevant parts of functional coefficients or the entire effect of irrelevant regressors to zero. The algorithm is based on an iterative Fisher-scoring optimization and uses an adaptive LASSO penalty function, where the weights are obtained by the unpenalized f-HDGM maximum likelihood estimators. We test the algorithm's performances using three simulation setups with increasing degree of complexity and realism: spatio-temporal uncorrelated data with uncorrelated covariates, spatio-temporal correlated data with uncorrelated covariates and spatio-temporal correlated observations with correlated covariates. Eventually, the algorithm is applied to a real case study in which the hourly airborne NO₂ concentrations in Lombardy (Italy) observed during the COVID-19 lockdown period imposed in 2020 are modeled as a functional process with several weather covariates.

SAES2

135 Variations on Varying-Coefficient Signal Regression

Brian D. Marx¹, Paul H. C. Eilers²

1 – Louisiana State University, United States of America; 2 – Erasmus University Medical Centre, Dordrecht, The Netherlands

Penalized signal regression (PSR) is an important part of functional data analysis. The goal is to predict a scalar outcome from a vector input, with values along a domain with a natural order, like time, wavelength, or position. Ordinary least squares regression cannot be used for a calibration model, because of severe multicollinearity in the matrix of signals. The solution applied here is forcing the coefficient vector to be smooth.

Many chemical and physical instruments produce signals. In this presentation they are near infrared optical reflectance spectra. They are used to predict soil properties like the concentrations of organic matter, nitrogen, and carbon.

An extension of signal regression allows the coefficient vector to vary along an additional dimension. This leads to varying-coefficient PSR. An example is to include the measured temperature to improve predictions. Another is to use the locations of samples. The data set at hand was collected in different locations and the goal was

to use spatial information to improve predictions. There may be more information in the collection of spectra that is not covered by explicit spatial information. To test this idea, principal components can be used as spatial proxies. Theory, application to a real data set and performance evaluations will be presented.

This is joint work with Bin Li (& Paul Eilers).

ARM1

136 A panel SVAR model for European climate policy

Simone Maxand

Humboldt-Universität Berlin, Germany

Carbon emissions are priced in Europe in two ways: first, as a cap-and-trade-system the EU emission trading system sets a price on carbon. Following prior literature, this price is determined by energy prices, the macroeconomic development, weather and renewables. We analyse these interactions by a monthly smooth transition SVAR model identified by non-Gaussianity and narrative sign restrictions and find heterogeneous effects over two regimes of economic activity. Second, individually set carbon taxes cover a certain share of emissions in the separate member countries. We study these tax effects in a panel SVAR model which accounts for the joint market development. Joining the two models brings the development and effects of the identified structural carbon price/tax shocks on EU and country level together. This allows to study the interaction and separate development of the two carbon pricing tools over time by means of mixed frequency approaches. This model framework additionally enables to set further country specific environmental policies in relation to the European macroeconomy.

EEA2

137 Estimation and Inference in Factor Copula Models with Exogenous Covariates

Alexander Mayer, Dominik Wied

University of Cologne, Germany

A factor copula model is proposed in which factors are either simulable or estimable from exogenous information. Point estimation and inference are based on a simulated methods of moments (SMM) approach with non-overlapping simulation draws. Consistency and limiting normality of the estimator is established and the validity of bootstrap standard errors is shown. Doing so, previous results from the literature are verified under low-level conditions imposed on the individual components of the factor structure. Monte Carlo evidence confirms the accuracy of the asymptotic theory in finite samples and an empirical application illustrates the usefulness of the model to explain the cross-sectional dependence between stock returns.

SiF2

138 Estimating land use and land cover change: An exploration of change measures, estimation tools, and data sources for change detection

Kelly McConville

Harvard University, United States of America

Precise and accurate information about land use and land cover (LULC) change is vital to the sustainable management of the Earth's resources. In this talk, I will present the findings of a study that sought to explore definitions of change, a variety of estimation strategies, and potential remotely sensed data sources that could augment traditional forest inventory data. For our study area, we focused on the north central portion of the state of Georgia in the US as it has experienced significant loss in forest land and tree cover over the last three decades. I will provide both our study conclusions and overarching advice on improving the imprecision of change estimates.

SAES1

139 Multivariate Distribution Regression

Jonas Meier

University of Amsterdam, Netherlands, The

This paper introduces multivariate distribution regression (MDR), a semi-parametric approach to estimate the joint distribution of outcomes. The method allows studying complex dependence structures and distributional treatment effects without making strong parametric assumptions. I show that the MDR coefficient process converges to a Gaussian process and that the bootstrap is consistent for the asymptotic distribution of the estimator. Methodologically, MDR contributes by offering the analysis of many functionals of the CDF. For instance, this includes counterfactual distributions. Compared to copula models, MDR achieves the same accuracy but is (i) more robust to misspecification, (ii) allows to condition on many covariates, thus ensuring a high degree of flexibility and (iii) does not require comparably strong parametric assumptions. Finally, an application analyzes shifts in spousal labor supply in response to a health shock. I find that if low-income individuals receive disability insurance benefits, their spouses respond by increasing their labor supply. Whereas the opposite holds for high-income households, likely because they are well insured and can afford to work fewer hours.

ARM3

140 The inclusive Synthetic Control Method

Giovanni Mellace¹, Roberta Di Stefano²

1 – *University of Southern Denmark, Denmark*; 2 – *University of Rome - Sapienza, Italy*

The Synthetic Control Method (SCM) estimates the causal effect of a policy intervention in a panel data setting with only a few treated units and control units. The treated outcome in the absence of the intervention is recovered by a weighted average of the control units. The latter cannot be affected by the intervention, neither directly nor indirectly. We introduce the inclusive synthetic control method (iSCM), a novel and intuitive synthetic control modification that allows including units potentially affected directly or indirectly by an intervention in the donor pool. Our method is well suited for applications with multiple treated units where including treated units in the donor pool substantially improves the pre-intervention fit and/or for applications where some of the units in the donor pool might be affected by spillover effects. Our iSCM is very easy to implement, and any synthetic control type estimation and inference procedure can be used. Finally, as an illustrative empirical example, we re-estimate the causal effect of German reunification on GDP per capita allowing for spillover effects from West Germany to Austria.

EEA1

141 Continuous-time latent-state modelling of delinquent behaviour in adolescence and young adulthood

Sina Mews, Roland Langrock, Marius Ötting, Houda Yaqine, Jost Reinecke

Bielefeld University, Germany

We develop a flexible modelling framework and associated inferential tools for general continuous-time state-space models (SSMs), which is motivated from a longitudinal data set on delinquent behaviour of adolescents in Germany. The study aim is to investigate the persistence of an individual's delinquency level over time. We assume the latter to be a latent trait underlying the trajectories of adolescents' delinquency, which is observed at irregular time intervals. For analysing such irregularly sampled sequential observations that are driven by an underlying state process, continuous-time SSMs constitute a flexible tool. In particular, we specify the state process as an Ornstein-Uhlenbeck process, such that an individual's delinquency level is persistent over time and changes gradually. We further include the adolescents' gender and age as covariates in the observation process, which is assumed to follow a negative binomial distribution (conditional on the states). Statistical inference is carried out by maximum approximate likelihood estimation, where multiple numerical integration within the likelihood evaluation is performed via a fine discretisation of the state process. The corresponding reframing of the SSM as a continuous-time hidden Markov model enables us to apply the associated efficient algorithms for parameter estimation and state decoding. The results reveal temporal persistence in the deviation of an individual's delinquency level from the

population mean. While the presented continuous-time SSM is tailored to the analysis of adolescents' delinquent behaviour, the modelling framework generally allows both the observation and the state process to be non-linear and non-Gaussian, thus enabling various possible model specifications.

LVM

142 Using tensor product B-splines for nonparametric inference in multivariate hidden Markov models

Rouven Michels, Roland Langrock

Bielefeld University, Germany

Hidden Markov models (HMM) comprise an observed time series that is driven by an unobserved Markov chain. The class of state-dependent distributions — e.g. normal, Poisson, or gamma — is typically chosen before modelling. An unfortunate choice of this class can lead to a poor fit, to wrong inference regarding the number of states, and to unsatisfactory state decoding. When fitting HMMs to multivariate time series, challenges already present in the univariate setting, such as skewness, heavy tails or outliers, are amplified by possibly complex dependence structures between the variables considered. To avoid the potential pitfalls associated with the use of a parametric class of distributions we estimate the state-dependent distributions nonparametrically. To this end, we discuss the use of multivariate tensor product B-splines within HMMs, thus estimating the multivariate state-dependent density in a data-driven way, i.e. without having to make a distributional assumption.

In a simulation study, we demonstrate the general feasibility of the suggested non-parametric approach, and showcase a scenario in which it would be superior to a parametric approach. The practical use of the nonparametric approach is further illustrated in a case study using football data. Specifically, we model the bivariate data on length and angle of the passes of goalkeepers during the 2021 UEFA European Championship, thereby detecting match phases in which teams apply different strategies to start an attack. By incorporating covariates into the state-switching probabilities, we gain information about potential tactical adjustments by the team managers.

TSA2

143 Sequential Gaussian approximation for nonstationary time series in high dimensions

Fabian Mies, Ansgar Steland

RWTH Aachen University, Germany

Gaussian couplings of partial sum processes are derived for the high-dimensional regime $d = o(n^{1/3})$. The coupling is derived for sums of independent random vectors and subsequently extended to nonstationary time series. Our inequalities depend

explicitly on the dimension and on a measure of nonstationarity, and are thus also applicable to arrays of random vectors. To enable high-dimensional statistical inference, a feasible Gaussian approximation scheme is proposed. Applications to sequential testing and change-point detection are described.

SHDA1

144 Statistical language modeling explained

Tomáš Mikolov

Charles University, Czech Republic

Modeling natural languages using statistical methods has a long and rich history in the research community. During the last decade, we have seen an explosion of novel models and techniques based on deep learning, which can sometimes be confusing. In this talk, we will review some of the main concepts of language modeling, and show how various concepts such as neural networks, n-grams and text generation are related.

TMCA

145 Optimal dose-finding for efficacy-safety-models of Emax-type

Frank Miller, Renata Eirini Tsirpitz

Stockholm University, Sweden

The knowledge of how optimal designs look like is beneficial for the planning of dose-finding trials. In many cases, dose-finding trials consider both efficacy and safety. Therefore, we analyse a bivariate model for these two outcomes and consider different models of Emax-type for each of the two outcomes. We review existing and derive new results for the number of design points for the c- and D-optimal designs. Then, we derive these optimal designs, and our algebraic results offer insight in the structure of the optimal designs. They help also to understand the results in cases where only numeric results could have been derived. We see that the drug's therapeutic index has a significant impact on the shape of the optimal design both for the c- and D-optimal designs. While our focus in this talk is on locally optimal designs, we discuss handling of unknown parameter values by sequential/adaptive designs or other approaches.

DECT2

146 Optimal dose-finding for efficacy-safety-models of Emax-type

Annette Möller¹, Ann Cathrice George², Jürgen Groß³

1 – Bielefeld University, Germany; 2 – Federal Institute for Quality Assurance of the Austrian School System (IQS), Austria; 3 – University of Hildesheim, Germany

Machine learning based methods become increasingly popular in many research areas as they allow to fit models in a highly-data driven fashion, and often show comparable or even increased performance in comparison to traditional statistical models.

In the area of educational sciences the application of machine learning methods is still less common, however, recently first approaches have been made in this direction with promising results.

This work investigates the benefit of using classification trees for analysing data from educational sciences. An application to data on school transition rates in Austria indicates different aspects of interest in the context of educational sciences. The selection of variables within the tree growing procedure is in strong accordance with existing confirmatory theories from educational sciences. This allows to utilize the trees not only for prediction of school transition rates but also to perform variable selection for classical regression models typically employed in the educational sciences. Additionally, a case study shows that the classification performance of trees is comparable to that of binary regression models. At the same time the hierarchical nature of a tree provides an appealing interpretation in the context of educational sciences.

The results obtained so far indicate that trees and possibly other machine learning methods may be helpful to explore high-dimensional educational data sets, where no confirmatory theories have been developed yet and thus no insights exist about what variables should be chosen for building a statistical model to analyse the data.

SBES

147 Assessing Personalization in Digital Health

Susan A. Murphy

Harvard University, United States of America

Reinforcement Learning provides an attractive suite of online learning methods for personalizing interventions in a Digital Health. However after a reinforcement learning algorithm has been run in a clinical study, how do we assess whether personalization occurred? We might find users for whom it appears that the algorithm has indeed learned in which contexts the user is more responsive to a particular intervention. But could this have happened completely by chance? We discuss some first approaches to addressing these questions.

STIuDSD

148 Flexible Covariate Adjustments in Regression Discontinuity Designs

Claudia Noack¹, Tomasz Olma², Christoph Rothe³

1 – Department of Economics and Nuffield College, University of Oxford, United Kingdom;; 2 – University College London, United Kingdom; 3 – University of Mannheim, Germany

Empirical regression discontinuity (RD) studies often use covariates to increase the precision of their estimates. In this paper, we propose a novel class of estimators that use such covariate information more efficiently than the linear adjustment estimators that are currently used widely in practice. Our approach can accommodate a possibly large number of either discrete or continuous covariates. It involves running a standard RD analysis with an appropriately modified outcome variable, which takes the form of the difference between the original outcome and a function of the covariates. We characterize the function that leads to the estimator with the smallest asymptotic variance, and show how it can be estimated via modern machine learning, nonparametric regression, or classical parametric methods. The resulting estimator is easy to implement because tuning parameters can be chosen as in a conventional RD analysis. An extensive simulation study illustrates the performance of our approach.

EEA2

149 Demystifying and Optimizing Data Science

Rebecca Nugent

Carnegie Mellon University, United States of America

The Data Science Pipeline - far more than just a set of AI/ML algorithms. The first questions we ask, the early decisions we make, the final use and interpretation of our results - all of these play a crucial role when leveraging data-informed decision making for any problem. In this talk, we'll unpack what data science is with an emphasis on thinking about the entire data life cycle. We'll explore how data science is being used to tackle problems in transportation logistics, retail, travel, professional sports, as well as take an insider's look at modeling influenza and the COVID-19 pandemic. We'll also describe our current research on "the science of data science" supported by the Integrated Statistics Learning Environment (<https://www.stat.cmu.edu/isle>). An interactive, web-based e-learning platform used by thousands of students and industry practitioners, ISLE tracks the entire data analysis process including written work and group collaboration, helping us capture behavioral information that allows us to study data science: How should we teach it? Where are the most common mistakes made and why? How do people best collaborate using data? What impact do early decisions (e.g. EDA/graphs) have on the final results? Optimizing the use of data science requires understanding the people who do it and the decisions they make. Because, at its heart, data science is all about people.

DS

150 Bayesian tree-aggregated analysis of compositional amplicon and single-cell data

Johannes Ostner^{1,2}, **Salomé Carcy**³, **Christian L. Müller**^{1,2,4}

1 – Helmholtz-Zentrum München, Germany; 2 – Department of Statistics, Ludwig-Maximilians-Universität München, Germany; 3 – Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA; 4 – Center for Computational Mathematics, Flatiron Institute, New York, New York, USA

Accurate generative statistical modeling of count data is of critical relevance for the analysis of biological datasets from high-throughput sequencing technologies. Important instances include the modeling of microbiome compositions from amplicon sequencing surveys and the analysis of cell type compositions derived from single-cell RNA sequencing. Microbial and cell type abundance data share remarkably similar statistical features, including their inherent compositionality and a natural hierarchical ordering of the individual components from taxonomic or cell lineage tree information, respectively.

The recent Bayesian model for tree-aggregated amplicon and single-cell Compositional Data Analysis (tascCODA; Ostner et al., 2021) seamlessly integrates hierarchical information and experimental covariate data into the generative modeling of compositional count data. By combining latent parameters based on the tree structure with spike-and-slab Lasso penalization, tascCODA can determine covariate effects across different levels of the population hierarchy in a data-driven, parsimonious way.

In the context of differential abundance testing, tascCODA shows excellent performance on a comprehensive set of synthetic benchmark scenarios. Analyses on human single-cell RNA-seq data from ulcerative colitis patients and amplicon data from patients with irritable bowel syndrome, respectively, identified aggregated cell type and taxon compositional changes that were more predictive and parsimonious than those proposed by other schemes.

In summary, we posit that tascCODA and its implementation in Python constitute a valuable addition to the growing statistical toolbox for generative modeling and analysis of compositional changes in microbial or cell population data.

BSB2

151 Modeling post-lockdown exercise training development during the COVID-19 pandemic in Germany using mixed distributional regression

Fabian Otto-Sobotka, **Jessica Koschate**, **Tania Zieschang**

Carl von Ossietzky University Oldenburg, Germany

During the first lockdown of the COVID-19 pandemic in Germany, the fitness centers were closed from mid-March to June. People had fewer possibilities to exercise

during that period and physical fitness levels declined. This is a problem especially for people of advanced age. We modeled the development of the exercise performance of older people after the re-opening of fitness centers to find variables associated with a faster return to their previous performance.

The data were exported from a fitness circuit system, that automatically stores the parameters of the exercise training sessions in a cloud system. Participants were included if they had at least 3 measured exercise sessions before the lockdown and more than 10 exercises post lockdown. The response variable was a training score on a leg extension device (product of weight and repetitions). In the primary analysis we constructed a mixed regression model predicting the growth of the exercise score with weight, repetitions, number of sets, prior performance, age and gender. Further, we explored a mixed distributional regression model using GAMLSS and expectile regression.

The mean loss in the training score was 600kg during the lockdown. However, an additional training set per session increased the score by 600kg on average. Training with less than 50kg during the first 50 post-lockdown trainings led to an additional loss of around 1000kg points. 20 repetitions were necessary to maintain the current performance. Guidance for training settings helping to approach a prior performance could be derived from the results.

ARM4

152 A mechanical analogue of network meta-analysis

Theodoros Papakonstantinou, Gerta Rücker, Guido Schwarzer, Daniela Zöller, Adriani Nikolakopoulou

Institute of Medical Biometry and Statistics, University of Freiburg, Germany

Pairwise meta-analysis can be expressed as a system of linear springs (a linear spring exerts force proportional to displacement from its natural length) [1]. Under this representation, the length of a spring corresponds to a study treatment effect and the stiffness of the spring corresponds to its inverse variance. As more precise studies are stiffer than less precise ones, it is more “difficult” to substantially change their effect size. To synthesise study springs we use two main instruments: connection in parallel and connection in series. Connection of springs in parallel corresponds to the inverse variance weighted average and connection of springs in series corresponds to the synthesis of independent successive statistical quantities.

Drawing on previous work for pairwise meta-analysis [1], we extend the method to represent network meta-analysis as a system of springs. We also show how models that make less use of the normality assumption can be represented by the use of non-linear springs. Furthermore, by investigating the thermodynamics of the proposed system of springs, we show how statistical properties of a network, expected values and variances correspond to thermodynamic quantities such as energy and entropy. The analogue functions as an intuitive visualisation of the synthesis process and can be used to investigate differences in models as well as sources of heterogeneity and

inconsistency.

Reference

[1] Papakonstantinou, T, Nikolakopoulou, A, Egger, M, Salanti, G. Meta-analysis as a system of springs. Res Syn Meth. 2021; 12: 20–28. <https://doi.org/10.1002/jrsm.1470>

MA2

153 Design and Analysis in a non-linear longitudinal Poisson regression model

Parisa Parsamaram¹, Rainer Schwabe¹, Heinz Holling²

1 – Otto-von-Guericke University, Germany; 2 – University of Münster, Germany

Optimum design for the estimation of regression models is of importance in various studies, such as in medicine or in psychology. It provides an optimum estimation of the model. In this talk, we consider a non-linear longitudinal Poisson regression model. This model can be used to analyze the development of the ability effect of subjects over time in a study where the number of correct answers is observed for several tasks. For this purpose, the tasks are given to the subjects at several time points. The interest is in the estimation of the mean ability parameter, while the subject specific ability may vary randomly across subjects and over time. Therefore, we have to take random effects part into account and we assume a multivariate Gamma distribution for the random effects.

To find a D-optimum design, we would have to maximize the determinant of the Fisher information matrix with respect to the design. However, In the present model, the Fisher information matrix does not have an explicit analytical form. Therefore, we take the quasi-Fisher information matrix as an approximation. Then, the D-optimum design is obtained based on this approximation. Since the quasi-Fisher information matrix depends on the model parameters, we need to fix them to suitable nominal values and obtain locally D-optimum design. Finally, we examine the sensitivity of the locally D-optimum designs with respect to the perturbations of the model parameters by their D-efficiency. The locally D-optimum designs turn out to be quite robust against these perturbations.

DECT2

154 Accounting for mode effects and proxy surveys in survey sampling inference with nonignorable nonresponse

Danny Pfeffermann^{1,2}

1 – University of Southampton, United Kingdom; 2 – National Statistician and CBS Director, Hebrew University of Jerusalem, Israel

We propose a new, model-based methodology to address two major problems in

survey sampling: The first problem is known as mode effects, under which responses of sampled units possibly depend on the mode of response, whether by internet, telephone, personal interview, etc. The second problem is of proxy surveys, whereby sampled units respond not only about themselves but also about other units. For example, in many familiar household surveys, one member of the household provides information for all other members, possibly with measurement errors. Ignoring the existence of mode effects and/or possible measurement errors in proxy surveys could result in possible bias of point estimators and subsequent inference. Our approach accounts also for nonignorable nonresponse. We illustrate the proposed methodology by use of simulation experiments and real sample data, with known true population values.

OAS1

155 Statistical Learning of ECG Based on Functional Neural Networks

Vi Thanh Pham

University of Copenhagen, Denmark

An electrocardiogram (ECG) is a screening tool for heart disease. It is interesting to study to what extent a machine can extract information from an ECG.

We aim to develop functional neural networks that can provide clinically relevant predictions of heart disease related outcomes based on data from ECG. Through this data-driven approach, we aim at elucidating different patterns of the ECG signals that are most predictive for different outcomes and thereby improving the traditional ways of extracting information from the signals and also to obtain further knowledge about the underlying etiologies.

We use the multivariate functional structure of the standard 12-lead ECGs as an input to neural networks. In this talk, we present and compare different approaches for handling multivariate functional data as predictors. One such approach is obtained by extending the methods of Yao et al. (2021), who explored the single function input layer, to multivariate input layers. An important aim is to improve the explainability of neural networks so that clinicians would use them.

We illustrate the methods using large cohort data from the Danish population nested into the Danish registries.

Reference

[1] Junwen Yao, Jonas Mueller, Jane-Ling Wang Proceedings of the 38th International Conference on Machine Learning, PMLR 139:11898-11908, 2021.

AIML1

156 Optimality of the inverse normal combination test for two-stage group-sequential designs with early stopping

Maximilian Pilz, Meinhard Kieser

Institute of Medical Biometry, University of Heidelberg, Germany

Adaptive trial designs, i.e. dividing a clinical trial into two or more stages by an unblinded interim analysis with the option of data-driven design modifications, become more and more popular. In this case, the question arises of how to combine the test statistics of the different design stages without type I error rate inflation. The most frequently applied method in practice is the inverse normal combination test (Lehmacher & Wassmer, 1999). While the optimality of this procedure for independent stages is well-known, it is unclear whether it holds true for dependent stages as well. However, in practice, the stages are usually dependent since the option to terminate the trial early at an interim analysis for overwhelmingly good or bad results implies a dependency between the data of two different stages.

In this talk, we present a proof of the assertion that the inverse normal method is the power-maximizing strategy for two-stage group-sequential designs including the option of early stopping at the interim analysis. Since this is the most frequently applied strategy in practice, this result gives a strong rationale for the utilization of the inverse normal combination test in group-sequential designs. Moreover, we show examples of how the presented proof technique can be applied to investigate further properties of optimal adaptive two-stage designs. This gives additional insights into the behavior of optimal adaptive two-stage designs and offers the opportunity for further research.

DECT1

157 Blinded sample size recalculation in adaptive enrichment designs

Marius Placzek, Tim Friede

Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

Adaptive designs promise flexibility and efficiency in trials with subgroup analyses. Adaptations include (unblinded) interim analyses or blinded sample size reviews. An interim analysis offers the possibility to select promising subgroups and reallocate sample size in further stages. Trials with these features are known as adaptive enrichment designs. Such complex designs comprise many nuisance parameters, such as prevalences of the subgroups and variances of the outcomes in the subgroups. Additionally, a number of design options including the timepoint of the sample size review and timepoint of the interim analysis have to be selected. Here, for normally distributed endpoints, a strategy combining blinded sample size recalculation and adaptive enrichment at an interim analysis is proposed, i.e. at an early timepoint nuisance parameters are reestimated and the sample size is adjusted while subgroup selection and enrichment is performed later. Implications of different scenarios con-

cerning the variances as well as the timepoints of blinded review and interim analysis are discussed and design characteristics are presented via simulations.

DECT1

158 Pseudo-observations in survival analysis

Maja Pohar Perme

IBMI, Medical faculty, University of Ljubljana, Slovenia

Pseudo-observations present a general tool that can simplify a wide range of analyses in the survival field. The basic idea is to first handle the issue of censoring by redefining an outcome that is available for each individual at each follow-up time. Having defined pseudo-observations, the basic need for special survival analysis methods is removed and one can use standard approaches available outside the survival field. We shall consider three different problems of survival analysis where pseudo-observations provide simple and flexible solutions: graphical methods for assessing goodness-of-fit for regression models, estimation in relative survival and regression with years lost or saved as the outcome of interest. We will use these examples to discuss the properties of pseudo-observations, their advantages and their issues. While it shall often turn out that compared to model specific solutions, pseudo-observation approach may be less efficient, their importance becomes clear when considering models for which no method has yet been developed. We shall use both simulated and real data to illustrate each of the examples.

SEHA1

159 Statistik für Klimaschutz und Gesundheit – (mehr) Fortschritt wagen!

Walter J. Radermacher

FENStatS, Germany

1. Wie Statistiken helfen können

Angesichts der Dringlichkeit der Corona Pandemie scheint es naheliegend, die Frage nach der Rolle der Statistik hierauf zu konzentrieren. In diesem Beitrag wird allerdings ein breiterer Blickwinkel eingenommen und ein längerer Horizont befüwortet. Es ist vor allem die Notwendigkeit der Transformation zu einer nachhaltigen Entwicklung in vielen Lebensbereichen, in Produktion und Konsum, in Mobilität und Ernährung usw., die in den kommenden Jahren nach neuen Statistiken als Grundlage für politische Entscheidungen und den politischen Diskurs verlangen werden. Hierauf müssen wir uns so rasch wie möglich einstellen, damit die Qualität der Fakten (rechtzeitig) den Ansprüchen genügt.

2. Was ist Statistik

„Meinungsfreiheit ist eine Farce, wenn die Information über die Tatsachen nicht

garantiert ist.“ stellte Hannah Arendt bereits Mitte der 1960er Jahre fest. „Die Tatsacheninformation ... inspiriert das Denken und hält die Spekulation in Schranken.“ Hiermit ist die Rolle von Statistik in der Politik definiert: Sie soll als Sprache für den öffentlichen Diskurs dienen, helfen Entscheidungen zu verbessern und sie soll die Teilhabe aller an diesen Prozessen ermöglichen. Was hätte uns mehr vor Augen führen können, wie wichtig gute Statistiken sind, wenn nicht die pandemischen Erfahrungen der jüngeren Vergangenheit.

Statistiken sind von Natur her auf spezielle Art und Weise politisch; sie sind der Gegenstand von Meinung, ohne jedoch selbst durch Meinungen und Interessen beeinflusst zu werden. Wie kann Statistik dieser schwierigen Aufgabe gerecht werden, Tatsacheninformation im Sinne von Hannah Arendt zu sein? Wie können Fakten politisch relevant, nicht aber politisch getrieben sein?

Die Antwort auf diese Fragen ist mehrschichtig:

- Sie umfasst zunächst und vor allen Dingen gute mathematisch-statistische **Methodik**

sowie

- eine moderne Technologie basierend auf verlässlichen Datenquellen und effizienten Verarbeitungsprozessen.

Damit aber nicht genug:

- Statistische Prozesse bedürfen einer nachhaltig gepflegten **Infrastruktur**, sei es für die Forschung an Methoden und Entwicklung von Technologien, sei für die ‚Fabrikation‘ von statistischen Informationen.
- Qualität von statistischen Informationen beweist sich beim ‚Fitness for Purpose‘ Test: Sind die bereitgestellten Fakten relevant, werden sie richtig verstanden, erfüllen sie (realistische) Erwartungen? Diese Dimension von Qualität wird dann als Ziel erreicht, wenn alle Beteiligten, also die Produzenten und die Nutzer von Fakten, Daten, Trends, eine gemeinsame **Sprache** sprechen.

3. Statistik: Old School oder voll im Trend?

Statistik ist aufgefordert, sich als eigenständige Disziplin im Verbund mit anderen Datenwissenschaften zu positionieren und zu behaupten. Hier geht es um Kooperationen, neue Möglichkeiten und Synergien. Aus einer rein methodisch-technischen Perspektive betrachtet, erscheinen die Differenzen zwischen Datenwissenschaften, ‚Artificial Intelligence‘ und Statistik weniger gravierend. Vielmehr sind es andere Faktoren, wie Werte, Qualitätsdefinition, oder professionelles Selbstverständnis, welche den Unterschied ausmachen. Sowohl auf der Seite der Produzenten von quantitativen Informationen wie auch auf der der Nutzer gibt es in dieser Hinsicht sehr verschiedene ‚Datenwelten‘, was besonders im Vergleich der ‚Big Data Welt‘ mit der amtlichen Statistik deutlich wird.

4. Was steht auf der politischen Agenda?

- Koalitionsvertrag der Ampelregierung
- Europas digitale Dekade, Europäische Datenstrategie

5. (Mehr) Fortschritt wagen

5.1 Methodik

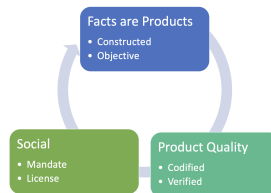
- Stärkung der Zusammenarbeit von Methodenentwicklung und Anwendung
- Verbindungen zwischen Datenwissenschaften und Statistik fördern
- Komplementarität zwischen Survey Statistik und Analyse vorhandener Quellen nutzen

5.2 Technologie

- Effiziente Technologien in allen möglichen Formen (insbesondere KI, maschinelles Lernen) vorantreiben und miteinander verbinden
- Qualitätsmanagement für neuere Technologien entwickeln
- Open Data Methoden (Schutz der Vertraulichkeit + Zugang für Forschung)
- Ethik und Good Governance Prinzipien für neue Technologie entwickeln

5.3 Infrastruktur

Bei der amtlichen Statistik handelt es sich um eine öffentliche Infrastruktur, die (wie der Schienenverkehr) gepflegt, ausgebaut und modernisiert werden muss.



- Wieviel dezentral und wieviel zentral darf/sollte es sein, bei der Umsetzung der Digitalisierungsoffensive der Bundesregierung?
- Data Stewardship: Wer übernimmt die Koordination?
- Von der amtlichen zur öffentlichen Statistik: Koordinierung eines größeren Systems durch ein(e) Chefstatistiker(in) (siehe Fachgespräch, S. 38)
- ‚Dateninstitut‘: Was sind die Erwartungen, wer und wo könnte es am besten angesiedelt sein?
- Kontinuierliche Investitionen in solche Infrastruktur sowie vorausschauende Ausrichtung auf neue Informationsbedürfnisse sind für solche Infrastruktur essentiell. Wird an dieser (falschen) Stelle gespart, erzeugt man ein gravierendes Problem, wenn auch vielleicht nicht unmittelbar, dafür aber umso mehr auf längere Sicht.
- Vorausschauende Programmplanung der öffentlichen Statistik
- Vernetzung zwischen wissenschaftlichen Beiräten und empirischer Statistik

- Governance: Grundlegende Reform des BstatG
- Verbesserung der wissenschaftlichen Interessenvertretung in DE (Benchmark: RSS)

5.4 Sprache

- Literacy, Standardisierung, Charta
- Statistische Expertise in Expertengremien einbeziehen

Si0

160 The role of modern statistical methodology in toxicological research

Jörg Rahnenführer

TU Dortmund, Germany

The statistical analysis of experimental data is crucial part of research in toxicology. This applies to classical low-dimensional data generated by many widely used toxicological assays, but even more so to high-dimensional data such as gene expression or proteomic measurements. The rapid developments in biotechnology and computer technology have enabled the generation of such complex data and therefore driven the paradigm shift in molecular biology, from hypothesis-driven research to hypothesis-generating research.

In the first part of the talk, we discuss statistical methods for low-dimensional data analysis. Many suitable approaches are available, but often, not sufficient statistical expertise is available in the toxicology teams. This even concerns the presentation of results with respect to visualization as well as summary measures. For example, instead of confidence intervals, only providing p-values is still common. In studies with different increasing concentrations and a negative control, we discuss testing each concentration against the control versus modelling the dose-response curve with continuous functions, and we will contrast the insights with common practice observed in toxicological publications.

In the second part of the talk, we will give an overview of typical necessary steps in the analysis of high-dimensional data in toxicology. First, preprocessing and normalization steps are of great importance and require statistical expertise. Then, the more complex data structures require tailor-made statistical approaches, such as penalization techniques in high-dimensional regression and classification, and enrichment tests to identify important biological processes and molecular functions that are common to gene or protein lists identified in toxicological experiments.

OT

161 Bayesian Variable and Effect Selection for Quantile Regression

Anja Rappl¹, Sebastiaan Klokman³, Manuel Carlan², Thomas Kneib²,
Elisabeth Bergherr²

1 – *Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*; 2 – *Georg-August-Universität Göttingen, Germany*; 3 – *NN*

The question, which variables to include in a model, has always been at the core of regression and in order to facilitate this decision increasing interest lies with automated variable and effect selection. In case the model in question is to regress an outcome not part of the exponential family or the analytical focus is on the more extreme outlier values a well established option is to use quantile regression. However, for this particular model class automated variable and/or effect selection is limited. We propose effect selection with an NBPSS prior in Bayesian quantile regression as an alternative to statistical boosting in quantile regression. We've conducted extensive simulation studies on data including informative and (quantile specific) non-informative linear, smooth and spatial effects and compared the results of Bayesian quantile regression with NBPSS priors to the results of boosted quantile regression. Further both methods were used to analyse childhood malnutrition in Nigeria. We found the Bayesian approach to outperform the statistical boosting approach in terms of effect selection accuracy and prediction performance, but to be sensitive towards selection of hyperparameters of the prior. Still this new approach makes a valuable contribution to the tools available for analysts.

BS1

162 Mixed-Effects Regression Weights of Advice: Individual Differences in Judgment Formation and Sampling

Tobias R. Rebholz, Marco Biella, Mandy Hütter

Psychology Department, Eberhard Karls University of Tübingen, Germany

New information technologies and social networks make it increasingly convenient to access a wide variety of opinions and advisors across different contexts. We propose mixed-effects regression weights to increase flexibility in the analysis of cognitive processing in various judgment tasks such as advice taking and anchoring. Mixed-effects regression modeling is shown to be a statistically adequate conceptualization of weighting various exogenous sources of information in endogenous (i.e., externally influenced) judgment formation. In addition, it essentially enables individual differences in weighting sequentially sampled advice, investigations of nonlinear weighting policies, as well as model comparisons to elaborate on the best fitting cognitive processing schemes. The new approach thereby overcomes the necessity to rely on problematic data preprocessing techniques such as weighting outlier identification, and its inherent shrinkage entails predictive improvement. The results of an extensive simulation study hence reveal that the multitude of technical merits of mixed-effects regression weights eventually implies better estimation performance

than with ratio-of-differences-based weighting indices (i.e., current state-of-the-art) in many typical advice taking scenarios. Moreover, the practical relevance of the proposed method is substantiated by demonstrating revised and extended evidence to selected empirical findings. By explicit modeling of the endogeneity in judgment formation, mixed-effects regression weights inaugurate new substantive areas of research.

MEC

163 Data-Driven Prediction of COVID-19 Cases in Germany for Decision Making

Lukas Refisch^{1,2}, Fabian Lorenz^{1,3}, Clemens Kreutz^{1,3}

1 – Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Germany; 2 – Institute of Physics, University of Freiburg, Germany; 3 – CIBSS - Centre for Integrative Biological Signalling Studies, University of Freiburg, Germany

The COVID-19 pandemic has led to a high interest in mathematical models describing and predicting the diverse aspects and implications of the virus outbreak. Model results represent an important part of the information base for the decision process on different administrative levels. The Robert-Koch-Institute initiated a project whose main goal is to predict COVID-19-specific occupation of beds in intensive care units. The incidence of COVID-19 cases is a crucial predictor for this occupation.

We developed a model based on ordinary differential equations for the COVID-19 spread with a time-dependent infection rate described by a spline. The model explicitly accounts for weekday-specific reporting and adjusts for reporting delay. The model is calibrated in a purely data-driven manner by a maximum likelihood approach. Uncertainties are evaluated using the profile likelihood method. The uncertainty about the appropriate modelling assumptions can be accounted for by merging results of different modelling approaches.

The model is calibrated based on incident cases on a daily basis and provides daily predictions of incident COVID-19 cases for the upcoming three weeks including uncertainty estimates for Germany and its subregions. Derived quantities such as cumulative counts and 7-day incidences with corresponding uncertainties are computed. The estimation of the time-dependent infection rate leads to an estimated reproduction factor that is oscillating around one. Data-driven estimation of the dark figure purely from incident cases is not feasible. Results of the incidence modeling are made available to various decision makers and used as a predictor for forecasting the need of intensive care units.

SME2

164 Prognosis Research in Healthcare: initiatives to improve methodology standards

Richard D. Riley¹, Gary S. Collins²

1 – Centre for Prognosis Research, School of Medicine, Keele University, Keele, UK; 2 – Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

In healthcare, prognosis research is the study of future outcomes in individuals with a particular disease or health condition.^[1] Statistical methods are fundamental to prognosis research, for example to appropriately summarise, explain and predict outcomes, and to provide reliable results that inform clinical practice and personalized healthcare decisions. However, methodology standards within prognosis research are often sub-standard, with issues including small sample sizes, overfitting, dichotomisation of continuous variables, lack of validation, and selective or incomplete reporting.^[2] These problems have exacerbated with the growth of AI and machine learning methods.

Nevertheless, positive initiatives are being made to help improve prognosis research. In this talk, we will describe a number of these initiatives and encourage participants to adopt and disseminate better practice. Over two 90-minute sessions, we will cover four broad topics and illustrate the issues using real examples.

The PROGRESS Framework

The PROGNosis RESEARCH Strategy (PROGRESS) provides a framework of four key themes within prognosis research: overall prognosis, prognostic factors, prognostic (prediction) models, and predictors of treatment effect.^[3–6] We will describe the rationale for this framework, outline the scope of each theme and why they are important, explain the limitations of current statistical practice in each, and signpost guidance for methodological improvements.

Reporting Guidelines for Prognosis Research

It is crucial for prognosis research studies to be fully and transparently reported (e.g. in terms of their rationale, design, methodology and findings), so that their findings can be critically appraised and utilised as appropriate. We will provide evidence along with examples of poor reporting in current prognosis and prediction studies (including machine learning studies), and showcase new and upcoming reporting guidelines that aim to address current shortcomings.^[7–8]

Sample Size Calculations For Prognostic Model Research

Sample size calculations are rarely undertaken in prognosis model research; if they are, overly-simplistic rules of thumb are often used. In terms of sample size for model development, a well-known rule of thumb is to have at least 10 events per predictor variable, but we will describe a more principled approach based on minimising expected overfitting and ensuring precise parameter estimation.^[9] In terms of sample size for model validation, a rule of thumb is to ensure at least 100 events and 100 non-events. Again, a more principled approach is possible, and we will describe how it uses the distribution of the model's linear predictor, and targets precise estimation of key model performance measures (calibration, discrimination and clinical

utility).^[10–11]

Individual Participant Data Meta-Analysis for Prognosis Research

One way to increase sample size is to undertake an IPD meta-analysis project, where the participant-level data from multiple existing studies are obtained, checked, harmonised, and synthesized. We will describe what an IPD meta-analysis project entails, and give examples for how it has improved prognosis research. In particular, we demonstrate how it enables non-linear prognostic relationships to be modelled; allows the development and validation of prognostic models across multiple settings and populations; and allow a more powerful and appropriate assessment of predictors of treatment effect.^[12–13]

References

- [1] Riley RD, van der Windt D, Croft P, et al., editors. Prognosis Research in Healthcare: Concepts, Methods and Impact. Oxford, UK: Oxford University Press, 2019.
- [2] Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009;339:b4184.
- [3] Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
- [4] Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10(2):e1001380.
- [5] Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10(2):e1001381.
- [6] Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
- [7] McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93(4):387-91.
- [8] Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multi-variable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;162:55-63.
- [9] Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
- [10] Riley RD, Collins GS, Ensor J, et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. *Stat Med* 2021
- [11] Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med* 2021;40(19):4230-51.
- [12] Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Stat Med* 2020;39(15):2115-37.
- [13] Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction

models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140.

EDU1, EDU2

165 Kernel density smoothing of composite spatial data on administrative area level: An application to incidence maps of COVID19 infections in Germany

Ulrich Rendtel¹, Marcus Gross², Timo Schmid³, Kerstin Erfurth⁴

1 – FU Berlin, Germany; 2 – INWT-Statistics, Berlin, Germany; 3 – Universität Bamberg, Germany; 4 – Amt für Statistik Berlin-Brandenburg, Germany

Composite spatial data on administrative area level are often presented by maps. The aim is to detect regional differences in the concentration of subpopulations. The standard presentation is by Choropleth maps where each administrative unit is represented by a single value. These maps can be criticized under three aspects: the implicit assumption of a uniform distribution within the area, the instability of the resulting map with respect to a change of the reference area and the discontinuities of the maps at the borderlines of the reference areas which inhibit the detection of regional clusters.

We use a density approach for the construction of maps. This approach does not enforce a local uniform distribution. It does not depend on a specific choice of area reference system and there are no discontinuities in the displayed maps. A standard estimation procedure of densities are Kernel density estimates. However, these estimates need the geo-coordinates of the single units which are not at disposal as we have only access to the aggregates of some area system. To overcome this hurdle, we use a statistical simulation concept.

We apply this statistical concept to incidence maps of COVID19 infections in Germany to detect infection clusters and their development over time. The concept was realized in a free web animation, see Rendtel et al (2021).

Reference

[1] Rendtel, U.; Neudecker, A.; Fuchs, L. (2021): Die Darstellung von Inzidenzgebieten mit simulierten Geokoordinaten. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 15, 93-106, <https://doi.org/10.1007/s11943-021-00288-x>

OAS2

166 Zeros

Roberto Renò

University of Verona, Italy

Asset prices can be stale. We define price staleness as a lack of price adjustments

yielding zero returns (i.e., zeros). We show that zeros are a genuine economic phenomenon linked to the dynamics of trading volume and, therefore, liquidity. Zeros are, in general, not the result of institutional features, like price discreteness. In essence, spells of idleness or near idleness are stylized facts suggestive of a key, omitted market friction in the modeling of asset prices. Even moderate amounts of zero returns in financial data, associated with stale prices, are heavily detrimental for reliable jump inference. Using the pricing of a large cross-section of individual stocks as an economic metric, we finally demonstrate that accounting for the theoretical interplay between trade inaction, leading to staleness, and the asymptotic properties of realized moments is revealing about the pricing of liquidity and high-order moments, the subjects of large applied literatures.

SiF1

167 Does Smoking Affect Wages?

Mirjam Reutter¹, Gregor Pfeifer², Kristina Strohmaier³

1 – European University Institute, Italy; 2 – Universität Hohenheim, Germany; 3 – Universität Tübingen, Germany

Previous studies have not reached consensus on whether there exists a causal relationship of smoking on wages. This study aims at filling that gap by providing new empirical evidence from a rich survey panel of German individuals. On average, smokers earn 15% less. This raw gap shrinks to 1.5 – 5% after controlling for a large set of observable characteristics, comparing siblings, or using past smoking behavior. To deal with unobserved time-varying factors, we propose a novel instrument, which provides exogenous variation in smoking behavior: Smoking bans at schools, introduced by the federal states at different years. While OLS estimates appear to be negatively biased due to unobserved heterogeneity, instrumental variable estimates indicate no causal overall effect of smoking on wages. However, significantly different effects in different directions appear for men (positive) and women (negative).

EEA2

168 Prior elicitation for variance parameters in Bayesian hierarchical models

Andrea Riebler

Norwegian University of Science and Technology, Trondheim, Norway

The selection of prior distributions is a central component of Bayesian inference, but often challenging even for statistical experts. In practice often rather ad hoc procedures are used to specify prior distributions. In this talk, I will first review common methods for setting priors for hierarchical models before presenting a novel tool for prior elicitation that can be used as part of the Bayesian workflow. Our new framework allows the user to formulate joint priors for variance parameters by taking

the entire model structure into account. It is implemented in the user-friendly R-package "makemyprior", in which the user utilizes prior beliefs to proportion the total variance present in the model to different random effect components in an intuitive and transparent way. A relation to specifying a prior distribution on the model fit as measured by the coefficient of determination, R^2 , will be outlined.

BS1

169 Improving the reliability of LDA results using LDAPrototype as selection criterion

Jonas Rieger, Carsten Jentsch, Jörg Rahnenführer

TU Dortmund University, Germany

For organizing large text corpora topic modeling provides useful tools. A widely used method is Latent Dirichlet Allocation (LDA). The assignments of LDA usually relies on a (random) initialization such that the outcome is also to some extent random. In particular, replicated runs on the same text data lead to different results such that the LDA is not fully reproducible. This instability of the LDA approach is often neglected in everyday practice, where text data analysis is commonly based only on a single LDA run.

We propose a method called LDAPrototype to select the most representative run from a set of replicated LDA runs applied on the same dataset. By this, we improve the reliability of conclusions drawn from LDA results, since replications of LDA-Prototype are more similar to each other than replications of single LDA runs, or even as LDA runs selected by perplexity or NPMI.

Our study on real world datasets show that the reproducibility of LDA results using LDAPrototype increases, and so does the reliability of empirical findings based on topic modeling. Overall, the new approach outperforms perplexity and NPMI regarding reliability and is justified in terms of its application by an easy usage and computational feasibility due to its implementation as an R package. The algorithm is also generalizable to be applied to other topic modeling procedures with topics characterized by word distributions.

TMCA

170 Adventures in benchmarkizing computational biology

Mark Robinson

University of Zurich, Switzerland

Computational methods represent the lifeblood of modern molecular biology. The extensive generation of complex omics datasets has resulted in an explosion of bespoke approaches to model and interpret large-scale data. Method benchmarking is critical to dissect important steps of an analysis pipeline, formally assess performance across common situations and edge cases, and ultimately guide users. However,

current benchmarking strategies need a rethink, because they are inefficient (conclusions are quickly outdated, method developers repeat similar benchmarks), not easily extensible (missing intermediate steps, assumptions unclear in simulations), authoritative (not easily challenged), and disconnected from practical workflows; the field is churning out “novel” computational methods at a rapid pace with overly-optimistic claims of improved performance, but we have limited know-how of what works in which situation. This talk will highlight, using several examples, some of our early steps to implement a next-generation platform for continuous neutral benchmarking and engage the community to actively participate.

BSB1

171 Prior-mean-RObust Bayesian Optimization (PROBO)

Julian Martin Rodemann

LMU Munich, Germany

Bayesian optimization (BO) with Gaussian processes (GP) as surrogates is used to optimize functions that are expensive to evaluate and lack analytical description, e.g. in hyperparameter-tuning of predictive models. In my presentation at the YSS, I would like to propose Prior-mean-RObust Bayesian Optimization (PROBO).

In the first part of my talk, the effect of the GP prior specifications on classical BO’s convergence is studied. I find the prior’s mean parameters to have the highest influence on convergence among all prior components. In response to this result, I introduce PROBO that aims at rendering BO more robust towards GP prior mean parameter misspecification. This is achieved by explicitly accounting for GP imprecision via a prior near-ignorance model. At the heart of this is a novel acquisition function, the generalized lower confidence bound (GLCB). In the second part of my talk, I will test my approach against classical BO on a real-world problem from material science and observe PROBO to converge faster.

YSS1

172 Total positivity in multivariate extremes

Frank Röttger¹, Sebastian Engelke¹, Piotr Zwiernik²

1 – Université de Genève, Switzerland; 2 – University of Toronto, ON, Canada

Engelke and Hitz (2020) recently introduced a general theory for conditional independence and graphical models for extremes. For Hüsler–Reiss distributions, the extremal analogue of Gaussians, they showed that extremal conditional independences are encoded in a precision matrix that results from a transformation of their parameter matrix Γ .

Multivariate total positivity of order 2 (MTP₂) is a strong form of positive dependence that induces many interesting properties in graphical modeling, as well as weaker forms of positive dependence like association. A multivariate Gaussian distribution is MTP₂ when its precision matrix is an M-matrix, i.e. when all the

non-diagonal entries in the precision matrix are non-positive. It is associated when all entries of its covariance matrix are non-negative.

We introduce the notions of extremal MTP_2 ($EMTP_2$) and extremal association and show that many classical models are always $EMTP_2$. Hüsler–Reiss distributions are $EMTP_2$ if and only if their precision matrix is the Laplacian matrix of a connected graph with positive edge weights. It is extremal associated when the parameter matrix Γ satisfies the triangle inequality $\Gamma_{ij} \leq \Gamma_{ik} + \Gamma_{jk}$ for all i, j, k , which reveals an interesting connection with research on effective resistance in electrical networks.

We propose an estimator for the parameters of the Hüsler–Reiss distribution under $EMTP_2$ as the solution of a convex optimization problem with Laplacian constraint. We prove that this estimator is consistent and typically yields a sparse model. We further construct a block descent algorithm and apply our method to real data.

This is joint work with Sebastian Engelke and Piotr Zwiernik.

MS1

173 Elicitation of empirical information on between-study heterogeneity in Bayesian meta-analysis

Christian Röver¹, Jona Lilienthal², Sibylle Sturtz², Ralf Bender², Tim Friede¹

1 – University Medical Center Göttingen, Germany; 2 – Institute for Quality and Efficiency in Health Care (IQWiG), Köln, Germany

In Bayesian meta-analysis, the specification of prior probabilities for the between-study heterogeneity is commonly required, and is of particular benefit in certain cases such as only few studies (Röver, 2020). Among the considerations in the setup of such prior distributions, the consultation of available empirical data on a set of relevant past analyses sometimes plays a role (Röver et al., 2021). How exactly to summarize historical data sensibly is not immediately obvious; in particular, the investigation of an empirical collection of heterogeneity estimates will not target the actual problem and will usually only be of limited use (Turner et al., 2015).

Using an example data set, we demonstrate how to fit a distribution to empirically observed heterogeneity data from a larger set of meta-analyses, including the choice of a parametric distribution family. We focus on simple and readily applicable approaches to then translate these into (prior) probability distributions for future meta-analyses.

References

- [1] C. Röver. Bayesian random-effects meta-analysis [1] using the bayesmeta R package. *Journal of Statistical Software*, 93(6):1–51, 2020. <https://doi.org/10.18637/jss.v093.i06>
- [2] C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods*, 12(4):448–474,

2021. <https://doi.org/10.1002/jrsm.1475>

[3] R. M. Turner, D. Jackson, Y. Wei, S. G. Thompson, P. T. Higgins. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6):984–998, 2015. <https://doi.org/10.1002/sim.6381>

IQW/TIG

174 Estimation and Testing of Wilcoxon-Mann-Whitney Effects in factorial clustered data designs

Kerstin Rubarth^{1,2}, Frank Konietzschke^{1,2}

1 – Charité - Universitätsmedizin Berlin, Germany, Institute of Biometry and Clinical Epidemiology; 2 – Berlin Institute of Health, Germany

Clustered data occur frequently, especially in biomedicine. Whenever subjects are repeatedly observed under a certain condition and time point, this inherent dependency should not be neglected when analyzing such data sets. Ignoring a possible existent correlation would lead to estimation bias and increased type-I error rates. Naïve approaches such as aggregating the information within a cluster by using summary measures, e.g. mean or medians, would lead to a possibly substantial loss of power. Well-established methods for the analysis of clustered data in the parametric framework are Mixed Models or Generalized Estimating Equations. However, they have certain assumptions such as linearity of predictors or homogeneity of variances and verifying these assumptions is often impossible in practical applications. Therefore, nonparametric methods are suitable options, especially in case of ordinal, ordered categorical, highly skewed metric or binary data, as well as in scenarios with small to moderate sample sizes. We therefore propose to use so called Wilcoxon-Mann-Whitney effects and set up a model for the analysis of factorial repeated measures designs with possibly dependent replicates and present quadratic- as well as multiple contrast test procedures for this setting. Additionally, the framework allows for the occurrence of missing data and makes use of all-available information instead of incorporating the information from only completely observed subjects. To investigate the newly proposed methods, we conduct an extensive simulation study, which indicates, that the method can be applied in various scenarios. A real world data example exemplifies the application of the procedure.

RNS3

175 General independent censoring in event-driven trials with staggered entry

Jasmin Rühl¹, Jan Beyersmann², Sarah Friedrich¹

1 – Mathematical Statistics & AI in Medicine, University of Augsburg, Germany;

Randomised clinical trials with time-to-event endpoints are frequently stopped after a pre-specified number of events has been observed. This practice leads to dependent data and non-random censoring, though, which can generally not be solved by conditioning on the underlying baseline information. If the observation period starts at the same time for all of the subjects, the assumption of independent censoring in the counting process sense is valid, and the common time-to-event methods can be applied. In case of staggered study entry, matters are complicated substantially, however.

We demonstrate that the study design at hand still entails general independent censoring, provided that the analysis is based on study time information only. To illustrate that the filtrations must not use abundant information, we present the results of a simulation study where we evaluated event-driven trials with staggered entry by means of Cox regression models with covariates for the calendar times. The associated Breslow curves of the cumulative baseline hazard show considerable deviations. This implies that the analysis is disturbed by the condition on calendar time variables. Besides, we consider a second simulation study which shows that Efron's classical bootstrap, unlike the (martingale-based) wild bootstrap, may lead to biased results in the given setting, since the assumption of random censoring is violated.

SEHA3

176 Design weights in panel surveys with multiple refreshment samples. A general discussion with an application to the GESIS panel

Matthias Sand, Christian Bruch, Barbara Felderer, Ines Schaurer, Jan-Philipp Kolb, Kai Weyandt

GESIS - Leibniz-Institute for the Social Sciences, Germany

All panel surveys suffer from dropout, e.g., panel members do not continue to participate in the panel. To still be able to infer panel findings to the general population and to increase the sample size, the panel population is often refreshed by recruiting new members. However, newly integrated members often stem from a more current sample that relies on an altered target population (persons may grow into and out of the target population). Since a panel with refreshments may also contain members that stem from multiple sampling frames, there is a need to adjust each member's inclusion probability accordingly. The inclusion probabilities need to account for members being viable for more than one sample, e.g., at different points in time, while others may only be able to enter via one frame. This can be conducted by adjusting the inclusion probabilities according to common multiple fame approaches.

In this presentation, we discuss how to appropriately address such problems using the design weighting approach developed by Kalton and Anderson (1986). We apply the weighting approach to the complex recruiting scheme of the GESIS panel

consisting of an initial sample and two refreshment samples. We illustrate the importance of multiple-frame weighting by comparing design weighted estimates from the GESIS panel to unweighted ones. We further compare multiple-frame weighting to a weighting procedure that does not correctly integrate the refreshment samples but uses design weights for each sample and does not account for the possibility of panel members' multiple inclusion.

SM2

177 STRATOS – aims, tasks, support of the initiative

Willi Sauerbrei¹, Heiko Becher²

1 – Medical Center - University of Freiburg, Germany; 2 – Universitätsklinikum Hamburg-Eppendorf, Germany

In 2013 the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative was created with the aim to systematically evaluate existing methodologies, identify unresolved issues, stimulate research in these areas, develop guidance and education material. The overarching aim is to enhance methodological accuracy of real-life data analyses.

STRATOS

178 Structured reporting – low hanging fruit to improve completeness and transparency of analyses in medical and methodological research

Willi Sauerbrei¹, Tim Häussler¹, Marianne Huebner²

1 – Medical Center - University of Freiburg, Germany; 2 – Michigan State University, USA

For prognostic marker studies reporting guidelines and a corresponding 'explanation and elaboration' paper are available for many years (Altman et al 2012, <https://doi.org/10.1371/journal.pmed.1001216>). However, a recent review has shown that even basic items of the study population and relevant details of statistical analyses are often not provided (Sekula et al 2017, <https://doi.org/10.1371/journal.pone.0178531>). The two-part REMARK profile, a structured display which summarizes key aspects of a study, especially the derivation of the sample, and information about all analyses performed, has been proposed to improve completeness and transparency of reporting, specifically of statistical analyses (Altman et al 2012). Created prospectively it helps to develop the statistical analysis plan and it increases transparency of all analyses conducted (Winzer et al 2016, <https://doi.org/10.1371/journal.pone.0149977>). Aiming to examine what analysis steps were performed and whether sufficient details of each analysis are given in published biomarker studies, we reviewed a selection of observational studies published in clinical journals on cancer research. From each of five journal we se-

lected three papers. Our brief review shows that reporting of analyses is insufficient in nearly all 15 studies and consequently interpretation of results is subject to severe bias, an important problem hindering suitable meta-analyses (Sauerbrei and Haeussler 2018, <https://doi.org/10.1038/s41416-018-0023-z>).

We will argue that the REMARK profile and suitable adaptations are suitable instrument to improve transparency in reporting many types of medical (eg. prediction) and methodological (eg simulation) studies.

OT

179 Data Fusion in the Context of Microsimulation: Combining Semi-parametric Methods with Statistical Learning Approaches

Jannik Schaller

Federal Statistical Office of Germany, Germany

Microsimulations are typically based on synthetic data sets. Since relevant variables are often not collected in one but in different data sources, it is necessary to jointly consider different data sets of official statistics as basis for synthetic data generation. Due to the lack of unique identifiers and legal restrictions, direct record linkage is not possible, which is why we have to resort to data fusion methods. Data fusion scenarios are defined as a specific missing data pattern that occurs through ~~ktacking~~ of two (or more) data sources. Typically, data fusions are performed by means of the common variables collected in both data sets (e.g. gender, age, etc.). Non-parametric nearest neighbour techniques seem to be the most widespread data fusion implementation in practice. Semi-parametric approaches, however, use a model to identify most similar observations and, thus, implicitly take into account the varying explanatory power of the common variables with regard to the specific variables to be fused. This advantage could be further exploited by combining common semi-parametric approaches, which are typically based on linear regression models, with performant Statistical or Machine learning methods, such as Random Forest. Therefore, the aim of this presentation is to jointly use the strength of the semi-parametric methods on the one hand and the Statistical Learning methods on the other. For this purpose, different possibilities for combination are illustrated and their effects on the fusion performance are discussed.

OAS2

180 A control selection strategy for differential network testing in intensive care: Revealing diverging dynamics of organ system interactions for survivors and non-survivors

Roman Schefzik, Leonie Boland, Bianka Hahn, Thomas Kirschning, Holger A. Lindner, Manfred Thiel, Verena Schneider-Lindner

Department of Anesthesiology and Surgical Intensive Care Medicine, Medical Faculty Mannheim, Heidelberg University, Germany

Testing for differences between statistical networks is well-established in many scientific disciplines, but so far has only rarely been applied in intensive care medicine. Here, we compare networks based on parameters representing main organ systems to evaluate differential dynamics associated with survival in a cohort of critically ill patients from the surgical intensive care unit (ICU) of the University Medical Centre Mannheim.

We apply the following control selection strategy aiming to ensure valid cross-sectional and longitudinal network comparisons on admission and at an event stage prior to death: First, for each non-survivor, we identify risk sets of potential control patients alive after the same ICU treatment duration before death. We then sample from each risk set as survivor control the patient with the closest propensity score for network parameters and mortality on admission. To ensure sufficient variability of survivor admission network parameters, a patient serves as control only once. The networks are compared using our R package DNT (<https://github.com/RomanSchefzik/DNT>), providing a comprehensive framework for differential network testing procedures. Expectedly, the non-survivors' and survivors' admission networks do not significantly differ. However, the organ system interactions of the survivors then stabilize at the event stage, revealing significantly more network edges, whereas those of the non-survivors do not. Results from 100 control sets randomly drawn from the risk sets without propensity score-based matching confirm the validity of our control selection approach, allowing for meaningful and fair comparisons. We therefore suggest application of comparable control selection strategies for other network comparisons in related contexts.

SME3

181 A pseudo-value approach for building regression models with time-dependent covariate effects

Alina Schenk, Matthias Schmid

Department of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Germany, Germany

Pseudo-value regression has become a convenient tool for the direct modeling of survival probabilities [1]. In contrast to many other techniques for survival analysis, pseudo-value models do not rely on restrictive assumptions but allow for the specification of flexible predictor-response relationships. Furthermore, they avoid

reducing the available data to a working data set including uncensored observations only (as done e.g. by methods based on inverse-probability-of-censoring weighting). Currently, the standard approach is to estimate the coefficients of a main-effects model within the generalized estimating equations (GEE) framework.

Here, we propose a semiparametric approach to extend standard pseudo-value models by higher-order interaction terms and time-dependent covariate effects. Our method is based on the application of a recursive partitioning algorithm to identify relevant interactions between covariates, followed by a series of gradient boosting steps that optimize a GEE-type weighted least squares loss. To avoid overfitting, boosting models are successively updated via node-specific offset values, inspired by logistic model trees [2]. We illustrate our method by modeling survival probabilities in observational data provided by the Department of Anaesthesiology and Intensive Care Medicine at the University Hospital Bonn [3].

References

- [1] Andersen PK, Pohar Perme M (2010). Pseudo-observations in survival analysis. *Stat Methods Med Res* 19, 71-99.
- [2] Landwehr N et al (2005). Logistic model trees. *Mach Learn* 59, 161-205.
- [3] POSE-Study group (2021). Peri-interventional outcome study in the elderly in Europe. *Eur J Anaesthesiol*, <https://doi.org/10.1097/EJA.0000000000001639>.

SEHA1

182 Modelling of infectious disease spread with spatio-temporal networks

Markus Schepers, Jochem König, Konstantin Strauch

IMBEI, Unimedizin, Johannes-Gutenberg-University Mainz, Germany

The global spread of covid-19 has triggered tremendous research efforts to gain a better understanding of infectious diseases, in particular their mathematical modelling. Indeed historically, mathematical reasoning helped, for instance, in the eradication of smallpox by showing the overall benefits of vaccination. A current research question is: how do time-varying contact patterns influence the spread of an infectious disease?

Here, we investigate spatio-temporal networks as a model for time-varying contact patterns. Nodes representing individuals are embedded into a latent space with contacts existing between nearby nodes. Hyperbolic geometry has emerged as a natural choice of latent space to generate networks with common real-world properties such as clustering and degree heterogeneity. Clustering reflects densely connected communities, like private households, and degree heterogeneity captures the phenomenon of super spreading. For the disease process on the network, we include a compartment for temporal immunity.

We will derive several observations regarding the epidemic threshold, i.e. conditions on the model parameters under which a major outbreak occurs. Theoretical results will be assessed using human mobility data and face-to-face interaction data. Inclu-

ding time-varying contact patterns into the modelling might help to make better predictions or might contribute to a better understanding of the effectiveness of non-pharmaceutical interventions.

NA

183 Machine Learning for Multi-Output Regression: Comparing multivariate approaches with separate univariate ones

Lena Schmid, Alexander Gerharz, Andreas Groll, Markus Pauly

TU Dortmund University, Germany

Tree-based ensembles such as Random Forests or Extra Trees are modern classics among statistical learning methods. Compared to other statistical learning methods these tree-based ensembles usually rank very good regarding their predictive performance. Random Forests and Extra Trees are usually used for predicting univariate responses. In case of multiple outputs, the question arises whether multiple univariate models should be fitted separately or whether a multivariate approach should be considered directly. In this talk, these methods are compared in extensive simulations to help in answering the primary question when to use multivariate ensemble techniques. The presented study compares the performance of separate univariate Random Forests and Extra Trees for each component, multivariate Random Forests for all components simultaneously as well as two different multivariate extensions of Extra Trees, again for all components simultaneously, on simulated data sets with multiple outputs. The multivariate Random Forests are computed with a multivariate L2-Loss function instead of the univariate L2-Loss function typically used in the univariate Random Forest. For multivariate Extra Trees, two different implementations are compared. In the first approach, the different outputs are considered as different tasks and the data set can be split not just with regard to a specific feature, but also with regard to the different tasks. The second approach is similar to the approach for multivariate Random Forests and based on using a multivariate L2-loss function. The simulation designs cover different dependency structures and different dimensionalities to cover a wide range of models. Illustrative data analyses complement our investigation.

AIML2

184 Inference under Superspreading: Determinants of SARS-CoV-2 Transmission in Germany

Patrick Schmidt

University of Zurich, Switzerland

Superspreading, under-reporting, reporting delay, and confounding complicate statistical inference on disease surveillance data. I propose a structural model that improves inference under those factors. In a Bayesian framework, the model is esti-

mated on German Covid-19 data featuring date of symptom onset and age group. The data allows to identify age-specific determinants of transmission, including the effect of weather, reported prevalence, and testing and tracing. Several factors were associated with a strong reduction in transmission: public awareness rising, information on local prevalence, testing and tracing, and high temperature. A smaller reduction was found for school and restaurant closures, stay-at-home orders, and mandatory face covering. Reported local prevalence reduced transmission up to 44% (95% – CI : [40%, 48%]), which suggests a prominent role of behavioral adaptations to risk of infection and that disease spread is costly irrespective of policy interventions. Extrapolating weather effects, I estimate that transmission increases by 53% (95% – CI : [43%, 64%]) in colder seasons. On a broader perspective, the study illustrates the potential of surveillance data with demographic information and date of symptom onset to improve inference in the presence of under-reporting and reporting delay.

SME3

185 Time-to-event analysis with competing risks considering cluster structures - Comparison of methods based on a simulation study

Sabrina Schmitt¹, Ann-Kathrin Ozga²

1 – University of Applied Science – RheinAhrCampus, Germany; 2 – University Medical Center Hamburg-Eppendorf, Germany

Models of event time analysis are often used to demonstrate therapy efficacy. Here, one evaluates the time until a certain event of interest occurs. However, in clinical practice, another competing absorbing event may occur first. In addition, clinical trials are usually conducted simultaneously at multiple hospitals (clusters). It is assumed that the cluster structure arising here leads to dependency between failure times.

Since these mentioned cluster structures are mostly still neglected in the analysis of competing risks in clinical trials, the question arises how such cluster structures can be thereby considered.

Therefore, the aim of this work was to compare statistical methods by performing a Monte Carlo simulation study and to derive recommendations for future analysis of event times considering competing events and existing cluster structures. The methods considered have already been described in the literature and they are based on the Cox proportional hazards model or methods modeling subdistribution hazards.

In summary, there were only marginal differences between the models considered in terms of bias, mean square error, and empirical power. However, the approach by Katsahian et al. [1] showed the best performance in most scenarios based on these

values.

Reference

[1] Katsahian, S., Boudreau, C. (2011). Estimating and testing for center effects in competing risks, *Statistics in Medicine*, 30(13):1608-1617.

YSS1

186 On the uniform control of the Vapnik-Chervonenkis dimension in subgroup discovery using formal concept analysis

Georg Schollmeyer

Department of Statistics, LMU Munich, Germany

In this paper we analyze two methods of regularization in subgroup discovery. The first approach constrains the description length of the envisaged subgroup descriptions. The second approach is in the spirit of formal concept analysis: One firstly identifies large contranominal scales in the data structure, which correspond to large shatterable sets and thus a high VC dimension of the subgroup space. Then, the impact of the largest contranominal scales is reduced by temporarily deleting the corresponding objects to obtain a less complex subcontext of the given formal context. Based on the subcontext, we analyze two variants of reincorporating the deleted objects: The first variant looks at the closed item sets of the subcontext and their induced subgroups in the original formal context. The second variant instead restructures the original context by projecting the removed objects onto the concept lattice of the subcontext. We analyze the behavior of the methods with respect to a local version of the VC dimension and the actual size of the regularized subgroup spaces. It turns out that for the first method the VC dimension cannot be guaranteed to be controlled, but one can still obtain generalization bounds through the analysis of the size of the regularized subgroup space. While for the first variant of the second method the situation is similar, the second variant is able to (uniformly) control the VC dimension to a given nominal dimension. Finally, we analyze the methods by means of a data example from the German General Social Survey (GGSS).

RNS3

187 Good to hear your voice! Interactions of personality traits and smart speaker brand preferences in voice commerce

Carsten Schultz¹, Friederike Paetz²

1 – FernUniversität in Hagen, Germany; 2 – TU Clausthal, Germany

The competitive market for digital voice assistants, i.e., smart speakers, necessitates

tes an all-encompassing characterization of potential consumers by brands to derive efficiently targeted marketing strategies for voice commerce. Exploring users' personality is one particular option because research has generally established interdependencies between brand personality and consumer personality: Users tend to prefer brands with congruent personalities. Whereas brand personalities of digital voice assistants have seen initial research, personality traits of brand-specific users of digital voice assistants are understudied so far.

To close this research gap, we conducted an empirical discrete choice experiment in the product category of smart speakers also capturing the respondents' personalities by the five-factor model. We, then, estimated mixed logit models with user's personality traits as observed heterogeneity variables. In general, the brand attribute turned out as the main driver for smart speaker's selection—closely followed by the price attribute. Further attributes, i.e. language performance and data storage location, showed only minor effects on users' preference building. While the personality traits conscientiousness and agreeableness generally hinder the purchase of smart speakers, open personalities are comparatively likely to buy smart speakers. However, if conscientious and agreeable users enter the smart speaker market, they prefer the brand Amazon. The same holds for extravert users. Smart speakers of Google are preferred by neurotic, less open, but agreeable personalities. Such knowledge contributes to brand-specific targeting strategies of smart speaker users who commence into the market volume of voice commerce.

MEC

188 Comparison of Different Designs for Dose-Response Gene Expression Data

Leonie Schürmeyer, Kirsten Schorning, Jörg Rahnenführer

TU Dortmund, Germany

Understanding the dose-response relationship of a candidate drug is one of the main goals in toxicology. Therefore several authors emphasized the importance of optimal design regarding dose-response experiments. Using classical optimal design approaches significantly enhances the precision of the estimated curves or rather specific parameters. Thus optimal designs can evidentially improve comprehension in toxicological context.

We consider classic approaches like D-optimal and Bayes-optimal designs for dose-response relationships applied to gene expression data. The experimental conditions of microarray data are new challenges in planning those since all genes were elevated at the same doses. Thousands of dose-response relationships need to be determined, so the key question is which design works best for the analysis. Thereby gene expression data of valproic acid applied to human embryonic stem cells is analyzed. A simulation study is conducted to demonstrate the differences of the various designs and accessory the number of support points vary.

The results actively demonstrate to support the consideration of using optimal design approaches for gene expression data. Especially the Bayes-optimal design performs considerably better than the original used design. Measured by the root

mean squared error the original and equidistant design have an inferior performance compared to the optimal designs. Besides the precision of the model fits highly increases by enlarging the number of support points. Furthermore the Bayes-optimal design led to the most exact model fits, accordingly it should be preferred over the other designs.

DECT2

189 Graphical causal models for survey inference

Julian Schüssler¹, Peter Selb²

1 – University of Aarhus, Denmark; 2 – University of Konstanz, Germany

Directed acyclic graphs (DAGs) are now a popular tool to inform causal inferences. We discuss how DAGs can also be used to encode theoretical assumptions about nonprobability samples and survey nonresponse and to determine whether population quantities including conditional distributions and regressions can be identified. We describe sources of bias and assumptions for eliminating it in selection scenarios familiar from the missing data literature. We then introduce and analyze graphical representations of multiple selection stages in the data collection process, and highlight the strong assumptions implicit in using only design weights. Furthermore, we show that the common practice of selecting adjustment variables based on correlations with sample selection and outcome variables of interest is ill-justified and that nonresponse weighting when the interest is in causal inference may come at severe costs. Finally, we identify further areas for survey methodology that can benefit from advances in causal graph theory.

SM2

190 Alternative approaches to Delayed Response Group Sequential Design

Stephen Schüürhuis, Kevin Kunzmann, Cornelia Ursula Kunz

Boehringer Ingelheim Pharma GmbH & Co. KG, Germany

Common theory of group-sequential designs (GSD) assumes that the responses of subjects to treatments or control can be obtained immediately or at least shortly after administration justifying the assumption that at interim as well as final analysis the number of recruited subjects equals the number of observed subjects (assuming no missing data). However, in reality there is a latency between administration of treatment and availability of response so that the number of recruited subjects is larger than the number of observed subjects.

To address this issue, the following Delayed Response GSD (DR-GSD) has been proposed by Hampson and Jennison (2013): At interim analysis, one performs statistical testing on whether or not to continue recruitment. Having recruitment stop, the design intends to wait until additional responses from pipeline subjects have been obtained to subsequently test the actual hypothesis of interest. If recruitment

is to be continued, the study ordinarily proceeds to the next (if applicable) interim and procedure repeats up to final analysis.

We investigated alternative methods to handle the delayed response in group-sequential designs including a design using conditional probabilities and a design with additional rejection regions. The design characteristics are compared with respect to power, expected sample size and trial duration.

Reference

[1] Hampson, L.V. and Jennison, C. (2013). Group sequential tests for delayed responses (with discussion). *Journal of the Royal Statistical Society, Series B* (75), 3-54. <https://doi.org/10.1111/j.1467-9868.2012.01030.x>

DECT1

191 The use of machine learning methods for targeted marketing

Stephan Seiler

Imperial College London, United Kingdom

The use of targeted marketing activity has increased dramatically over recent years due to more available data on consumer behavior and advances in statistical techniques that allow the researcher to target based on a high-dimensional vector of customer characteristics. I summarize recent work at the intersection of statistics, computer science, and quantitative marketing that has applied machine learning techniques to issues of targeting. I focus on aspects of machine learning applications that are specific to targeting problems, namely (1) targeting on incremental impact rather than base rates, (2) the use of economic objective functions such as profits rather than statistical measures of fit, (3) cross-model comparison of targeting policies based on a model-free evaluation framework.

MEC

192 A stochastic search variable selection approach for identifying inconsistency in network meta-analysis

Georgios Seitidis¹, Stavros Nikolakopoulos^{1,2}, Ioannis Ntzoufras³, Dimitris Mavridis¹

1 – Department of Primary Education, School of Education, University of Ioannina, Ioannina, Greece; 2 – Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 3 – Department of Statistics, Athens University of Economics and Business, Greece

The reliability of the network meta-analysis (NMA) results lies in the plausibility

of its key assumption, transitivity, implying that the distribution of effect modifiers is similar across treatment comparisons. Transitivity is statistically manifested through the consistency assumption, suggesting that direct and indirect evidence are in agreement. Several methods have been suggested to evaluate consistency. A popular approach suggests adding inconsistency factors to certain places (e.g. each consistency equation or design estimating the same treatment effects). Both NMA and NMA with inconsistency factors are linear regression models and, in the Bayesian framework, choice of model is typically evaluated based on the DIC criterion. We use a different approach by assuming each inconsistency factor as a candidate covariate whose choice relies on variable selection techniques.

Our method, termed Stochastic Search Inconsistency Factor Selection (SSIFS), evaluates the consistency assumption both locally and globally, by applying the stochastic search variable selection method to determine whether the inconsistency factors should be kept in the model or not. An inclusion probability for each inconsistency factor reflects how likely inconsistency is for the specific comparison. Posterior odds or the median a posteriori model can be used for deciding both if there is inconsistency and where it can be found. We also constructed an informative prior on how likely a network is to be consistent based on an empirical study on 201 published network meta-analyses. The performance of our proposed method is evaluated in two published network meta-analyses.

MA2

193 Bias and variance in multiparty election polls

Peter Selb¹, Sina Chen¹, John Körtner², Philipp Bosch¹

1 – University of Konstanz, Germany; 2 – University of Lausanne, Switzerland

Recent polling failures highlight that election polls are prone to biases that the “margin of error” customarily reported with polls does not capture. However, such systematic errors are difficult to assess against the background noise of sampling variance. Shirani-Mehr, Rothchild, Goel and Gelman (JASA 2018) developed a hierarchical Bayesian model to disentangle random and systematic errors in poll estimates of two-party vote shares at the election level. We use a multivariate logistic normal distribution to adapt the model to multiparty elections, and we include temporal splines to improve model fit and variance estimation. We then estimate bias and variance in 5,000+ German national election polls 1994-2021. Such analyses can inform realistic assessments of poll accuracy.

SM2

194 Nonparametric Regression and Classification with Functional, Categorical, and Mixed Covariates

Leonie Selk, Jan Gertheiss

Helmut-Schmidt-University Hamburg, Germany

We consider nonparametric prediction with multiple covariates, in particular categorical or functional predictors, or a mixture of both. The method proposed bases on an extension of the Nadaraya-Watson estimator where a kernel function is applied on a linear combination of distance measures each calculated on single covariates, with weights being estimated from the training data. The dependent variable can be categorical (binary or multi-class) or continuous, thus we consider both classification and regression problems. The methodology presented is illustrated and evaluated on artificial and real world data. Particularly it is observed that the data-driven weights downgrade those covariates that are irrelevant, noise variables whereas relevant covariates are weighted distinctly higher. Thus variable selection is automatically performed and prediction accuracy is increased.

Reference

[1] Selk & Gertheiss (2021). Nonparametric Regression and Classification with Functional, Categorical, and Mixed Covariates. <https://arxiv.org/abs/2111.03115>

RNS2

195 When and Why Do Respondents Consent to Passive Data Collection – Evidence from Mouse Clicks and GPS Data

Henning Silber¹, Barbara Felderer¹, Carina Cornesse²

1 – GESIS - Leibniz Institute for the Social Sciences, Germany; 2 – German Institute for Economic Research (DIW Berlin), Germany

Web surveys offer the possibility of collecting passive data alongside survey data. However, respondents must consent to the additional data collection, which, as previous research showed, is challenged by respondents' data sharing hesitancy. Willingness to share additional data is influenced by factors such as data sensitivity, the data sharing method, and personal privacy concerns. In this experimental study, we explore under which conditions and for which reasons respondents do or do not consent to the collection of two common types of passive data: Keystroke paradata, which are relatively unobtrusive, and GPS location data, which is more intrusive from a data privacy standpoint. The experiment was implemented in 2019 in the German Internet Panel (GIP), a probability-based online access panel, and included about 3,500 respondents. First results show that respondents were significantly more likely to consent to the collection of mouse clicks (62 percent) than GPS data (51 percent), likely because the latter is considered more sensitive. We also reproduced the common finding that question order influences consent rates (effect sizes between 8 and 13 percentage points), depending on which of the two consent requests was asked first in the question sequence. In addition, we will ex-

plore respondents' understanding of the consent process as well as reasons given by consenters and non-consenters using closed and open questions, which followed the respective consent requests.

SM1

196 Ridge Model Averaging

Alena Skolkova

CERGE-EI, Czech Republic

Model averaging is an increasingly popular alternative to model selection. Ridge regression and model averaging serve a similar purpose, minimization of a mean squared error through shrinkage, though in different ways. In this paper, we propose the ridge-regularized modifications of Mallows model averaging (Hansen, 2007) and heteroskedasticity-robust Mallows model averaging (Liu & Okui, 2013) to utilize the capabilities of averaging and ridge regularization simultaneously. Via a simulation study, we examine the finite-sample improvements obtained by replacing least-squares with a ridge regression. Ridge-based model averaging is especially useful when one deals with sets of moderately to highly correlated predictors. We also demonstrate the superiority of the ridge-regularized modifications via empirical examples, focused on wages and economic growth.

References

- [1] Hansen, B. (2007). Least squares model averaging. *Econometrica*, 75(4), 1175-1189.
- [2] Liu, Q. & Okui, R. (2013). Heteroscedasticity-robust C_p model averaging. *The Econometrics Journal*, 16(3), 463-472.

ARM2

197 Implementing disclosure controls in DataSHIELD demonstrated by the dsSurvival package

Ghislain Sofack¹, Soumya Banerjee², Thodoris Papakonstantinou¹, Demetris Avraam³, Paul Burton³, Tom R. P. Bishop², Daniela Zöllner¹

1 – Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Germany; 2 – University of Cambridge, Cambridge, United Kingdom; 3 – Newcastle University, Newcastle, United Kingdom

The Cox proportional hazards model is one of the most popular survival analysis models used to determine the importance of predictors in survival. Achieving sufficient power in survival analysis usually requires large amounts of data from several sites or institutions. However, due to the ethical and practical considerations related to data transmission, and institutional policies, individual-level data cannot be

shared in many situations. As an alternative, the DataSHIELD framework (based on the statistical programming language R) can be used. Here, the individual-level data remain at each site, and only anonymous, aggregated data are shared. By carefully implementing disclosure controls in each DataSHIELD function, the privacy of the individual-level data is ensured. It is complex to build and verify the disclosure controls and therefore a limited number of DataSHIELD functions are available.

To address this problem, we have developed the package ‘dsSurvival’ for building Cox proportional hazards models using individual-patient data that are distributed across several sites, without moving those data to a central site. We will present the package and discuss potential disclosure risks and how we addressed those. Specifically, we propose a framework, for ensuring privacy control, where only the relevant non-disclosive statistical summaries will be shared. Additionally, we add a privacy control that prevents the fitting of Cox models where the number of parameters are less than a pre-set fraction of the number of entries in the survival data. Our tool can be of great use in domains where there is a need for building privacy-preserving survival models.

SEHA3

198 High-dimensional nonparametric functional graphical models via the additive partial correlation operator

Eftychia Solea¹, Kuang-Yao Lee², Bing Li³

1 – ENSAI; France; 2 – Temple University, USA; 3 – Penn State University, USA

This article develops a novel approach for estimating a high-dimensional and non-parametric graphical model for functional data. Our approach is built on a new linear operator, the functional additive partial correlation operator, which extends the partial correlation matrix to both the nonparametric and functional setting. We show that its nonzero elements can be used to characterize the graph, and we employ sparse re- regression techniques for graph estimation. Moreover, the method does not rely on any distributional assumptions and does not require the computation of multi-dimensional kernels, thus avoiding the curse of dimensionality. We establish both estimation consistency and graph selection consistency of the proposed estimator, while allowing the number of nodes to grow with the increasing sample size. Through simulation studies, we demonstrate that our method performs better than existing methods in cases where the Gaussian or Gaussian copula assumption does not hold. We also demonstrate the performance of the proposed method by study of an electroencephalography dataset to construct a brain network.

RNS2

199 Robust statistical boosting with quantile-based adaptive loss functions

Jan Speller, Christian Staerk, Andreas Mayr

University of Bonn, Germany

Robust regression methods are getting more and more important due to the rising availability of complex and large data sets, which are difficult to examine for noise or corruption. High-dimensional data, where the number of potential (co-)variables exceeds the number of observations, further complicates the analysis and entails additional challenges regarding regularization and variable selection. To address the challenges of high-dimensional and potentially corrupted data, we combine statistical gradient boosting algorithms with robust loss functions. Statistical boosting is particularly suitable for modelling high-dimensional data as it incorporates automated variable selection and regularization. To achieve robustness against outliers in the outcome variable, we consider different robust composite loss functions together with base-learners for linear regression. For these composite losses, such as the Huber loss and the Bisquare loss, a threshold parameter has to be specified that controls the robustness. In the context of boosting algorithms, we propose an approach that adapts the threshold parameter of these robust losses in each iteration to the current sizes of the residuals, based on a fixed quantile level. We compare the performance of our approach to classical M-regression and boosting approaches with standard losses regarding prediction accuracy and variable selection in different simulation settings. We also analyse a high-dimensional application of skewed protein expression data, where we are able to demonstrate that the new adaptive loss functions perform favourably compared to standard losses regarding prediction accuracy, while resulting in very sparse models.

RNS1

200 Estimation and Inference of Treatment Effects with L2-Boosting in High-Dimensional Settings

Martin Spindler¹, Jannis Kück¹, Zigan Wang², Ye Luo²

1 – Universität Hamburg, Germany; 2 – Hong Kong University, Hong Kong

Empirical researchers are increasingly faced with rich data sets containing many controls or instrumental variables, making it essential to choose an appropriate approach to variable selection. In this paper, we provide results for valid inference after post- or orthogonal L2-Boosting is used for variable selection. We consider treatment effects after selecting among many control variables and instrumental variable models with potentially many instruments. To achieve this, we establish new results for the rate of convergence of iterated post-L2-Boosting and orthogonal L2-Boosting in a high-dimensional setting similar to Lasso, i.e., under approximate sparsity without assuming the beta-min condition. These results are extended to the 2SLS framework and valid inference is provided for treatment effect analysis. We give extensive simulation results for the proposed methods and compare them

with Lasso. In an empirical application, we construct efficient IVs with our proposed methods to estimate the effect of pre-merger overlap of bank branch networks in the US on the post-merger stock returns of the acquirer bank.

CI

201 Uniform Inference in High-Dimensional Generalized Additive Models

Martin Spindler, Philipp Bach, Sven Klaassen, Jannis Kück

Universität Hamburg, Germany

We develop a method for uniform valid confidence bands of a nonparametric component f_1 in the general additive model $Y = f_1(X_1) + \dots + f_p(X_p) + \varepsilon$ in a high-dimensional setting. We employ sieve estimation and embed it in a high-dimensional Z-estimation framework allowing us to construct uniformly valid confidence bands for the first component f_1 . As usual in high-dimensional settings where the number of regressors p may increase with sample, a sparsity assumption is critical for the analysis. We also run simulation studies which show that our proposed method gives reliable results concerning the estimation properties and coverage properties even in small samples. Finally, we illustrate our procedure with an empirical application demonstrating the implementation and the use of the proposed method in practice.

SHDA1

202 Flexible tree-structured regression models for discrete event times

Nikolai Spuck, Matthias Schmid, Moritz Berger

Institute of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn

Time-to-event analysis is an essential tool to model data where the outcome variable is the time to the occurrence of a certain event of interest (such as time to death or time to disease progression in clinical research). Common methods for time-to-event analysis require time to be measured on a continuous scale. In practice, however, event times may be intrinsically discrete or exact event times may not have been recorded so that only time intervals in which the events occurred are known. Such data, called discrete time-to-event data, can be represented by a set of binary variables. Therefore, popular approaches for binary regression, including adaptations of logistic regression and classification trees, can be utilized for modeling discrete event times.

Here we present a novel regression approach for discrete time-to-event analysis that combines the advantages of parametric (e.g. linear) and tree-structured models. We introduce an algorithm that inherently performs variable selection and facilitates

the inclusion of non-linear effects and interactions in a common additive predictor function. The performance of the proposed method is assessed in a simulation study, and its application is illustrated by analyzing data collected at the University Hospital Bonn. Specifically, data of patients with acute odontogenic infection and lymphatic filariasis are considered. The proposed method is shown to be flexible and competitive with parametric approaches, including a linear model applying the Least Average Shrinkage and Selection Operator (LASSO), in terms of predictive performance and the ability to detect informative variables.

SEHA2

203 Blinded sample size re-estimation in a paired diagnostic study

Maria Stark, Antonia Zapf

University Medical Center Hamburg-Eppendorf, Germany

In a paired confirmatory diagnostic accuracy study, a new experimental test is compared within the same patients to an already existing comparator test. The gold standard defines the true disease status. If feasible and ethical, regulatory agencies prefer this study design to an unpaired design [1]. The initial sample size calculation is based on assumptions about, among others, the prevalence of the disease and the proportion of discordant test results between the experimental and the comparator test [2].

To adjust these assumptions during the study period, an adaptive design for a paired confirmatory diagnostic accuracy study is introduced. This adaptive design is used to re-estimate the prevalence and the proportion of discordant test results to finally re-calculate the sample size.

An example and a simulation study illustrate the adaptive design. The type I error rate, the power and the sample size of the adaptive design are compared to those of a fixed design. Both designs hold the type I error rate. The adaptive design reaches the advertised power. The fixed design can either be over- or underpowered depending on a possibly wrong assumption regarding the sample size calculation.

The adaptive design compensates inefficiencies of the sample size calculation and therefore it supports to reach the desired study aim.

References

- [1] Committee for Medicinal Products for Human Use, Guideline on clinical evaluation of diagnostic agents. Available at: <https://www.ema.europa.eu>. 2009, 1-19.
- [2] O. S. Miettinen, The matched pairs design in the case of all-or-none responses. 24 2 1968, 339-352.

SPMDI

204 PAC-Bayes training for sparse neural networks

Maximilian Steffen, Mathias Trabs

Karlsruhe Institute of Technology (KIT), Germany

Increasing computational power and storage capacity have made high-dimensional datasets accessible to many areas of research such as medicine, natural and social sciences. While classical statistical methods are not compatible with high-dimensional data, especially due to the curse of dimensionality, machine learning methods have been successfully applied to regression problems in practice. On the theoretical level, a popular way to circumvent the curse of dimensionality is the concept of sparsity. In a nonparametric regression model we develop a PAC-Bayesian approach to train a ReLU-neural network with sparse weights and shifts. Our main result is a non-asymptotic oracle inequality which in particular shows that the estimation method adapts to the regularity of the regression function.

MS2

205 Trusted Smart Surveys: Solutions for the European Statistical System - Objectives and Main Challenges

Shari Stehrenberg, Alexandra Ils

Federal Statistical Office Germany (Destatis), Germany

(Trusted) Smart Surveys offer opportunities to improve the quality of survey data and reduce the response burden by combining traditional survey data collection methods with non-traditional and new forms of data collection. However, Smart Surveys may run the risk of not being very successful in terms of adoption and acceptance by the respondents that are data providers and users of smart devices, due to concerns about privacy. Mitigating this risk, a carefully designed and cleverly communicated Trusted Smart Surveys system has better chances to be adopted and accepted. To employ the potential of (Trusted) Smart Surveys, both methodology and architecture of statistical processes need to be modified and extended. The ESSnet Smart Surveys project aims to investigate the development and dissemination of (Trusted) Smart Surveys by conceptualizing a European platform to share Trusted Smart Survey solutions and components.

Smart Surveys encompass (inter)active data explicitly provided by the respondent as well as passive data collected in the background by sensors in smart devices (e.g. smartphones, tablets, activity trackers). Trusted Smart Surveys refer to an augmentation of the Smart Survey concept by technological solutions aimed at increasing the degree of trustworthiness, hence promoting public acceptance and participation.

In this presentation, we will provide a general overview of the objectives of the ESSnet Trusted Smart Surveys project. We will address the main obstacles of Trusted Smart Surveys as solutions for the ESS and give an outlook on future work related to Trusted Smart Surveys.

OAS2

206 TrainBayes – A training study for medical and law students to improve Bayesian reasoning

Nicole Steib¹, Karin Binder², Büchter Theresa³

1 – University of Regensburg, Germany; 2 – Ludwig-Maximilians-Universität Munich, Germany; 3 – University of Kassel, Germany

Bayesian reasoning is often associated with misconceptions, which in the past have led to incorrect diagnoses and treatments in medicine and to erroneous court decisions in the field of law. Two strategies have proven helpful in such Bayesian tasks: 1) Provide statistical information in the format of natural frequencies instead of probabilities (Gigerenzer & Hoffrage, 1995) and 2) Visualizations, e.g., double trees or unit squares (Binder, Krauss & Wiesner, 2020; Eichler, Böcherer-Linder & Vogel, 2020). There is even a comprehensive meta-analysis (McDowell & Jacobs, 2017) that identifies person-related and task-related parameters that have an impact on the correctness of Bayesian judgements.

Although Bayesian reasoning is so important in disciplines such as medicine or law and can lead to fatal errors, there has been little empirical evidence of training studies using the successful strategies natural frequencies and visualizations. In the lecture we present the project TrainBayes (http://www.bayesianreasoning.de/br_trainbayes.html), which compares two optimal trainings with corresponding control groups: Both optimal trainings work with natural frequencies but different visualizations: the first group uses double trees, the second group unit squares. The trainings do not only focus on the participants' ability to calculate the correct solution, but also on the influence of parameter changes in Bayesian tasks. We present the trainings, their objectives and first insights into the results.

SLSE2

207 Functional additive models on manifolds of planar shapes and forms

Almond Stöcker, Sonja Greven

Humboldt-Universität zu Berlin, Germany

Defining shape and form as equivalence classes under translation, rotation and - for shapes - also scale, we extend generalized additive regression to models for the shape/form of planar curves and/or landmark configurations. The model respects the resulting quotient space geometry of the response, employing the squared geodesic distance as loss function and a geodesic response function to map the additive predictor to the shape/form space. For fitting the model, we propose a Riemannian L2-Boosting algorithm well suited for a potentially large number of possibly parameter-intensive model terms, which also yields automated model selection. We provide novel intuitively interpretable visualizations for (even non-linear) covariate effects in the shape/form space via suitable tensor-product factorization. The usefulness of the proposed framework is illustrated in an analysis of 1) astragalus shapes of wild and domesticated sheep and 2) cell forms generated in a biophysical

model, as well as 3) in a realistic simulation study with response shapes and forms motivated from a dataset on bottle outlines.

ARM1

208 Boosting Bivariate Structured Additive Distributional Regression Models

Annika Strömer¹, Nadja Klein², Christian Staerk¹, Andreas Mayr¹

1 – Department of Medical Biometrics, Informatics and Epidemiology, University of Bonn, Germany; 2 – Emmy Noether Research Group in Statistics and Data Science, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

We develop a model-based boosting approach for bivariate distributional regression models within the framework of generalized additive models for location, scale, and shape. The underlying statistical regression model enables the simultaneous modeling of all distributional parameters of the parametric distribution of a bivariate response conditional on explanatory variables, while being applicable for potentially high-dimensional data settings. Moreover, we can incorporate data-driven variable selection in the algorithm taking different types of effects into account. As a special merit of our approach, we can thus model the association between two continuous or discrete outcomes through the relevant covariates. For illustration, we analyze genotype data from the UKbiobank with a focus on the potential polygenic dependency between high cholesterol and chronic ischaemic heart disease. We employ a bivariate logit model where not only the marginal probabilities but also the odds ratio is modeled through clinical variables and high-dimensional genetic information. The method shows promising results in terms of further insights into the relationship between high cholesterol and chronic ischemic heart disease.

ARM3

209 Parametric post-processing of dual-resolution precipitation ensemble forecasts

Marianna Szabó, Sándor Baran

University of Debrecen, Hungary

The European Centre for Medium-Range Weather Forecasts (ECMWF) produces operational ensemble-based predictions that describe the range of possible scenarios and their likelihood of occurrence. According to its strategic plans till 2025, ECMWF wants to improve the resolution of ensemble forecasts, which requires a substantial increase of computational resources. One direction of studies is to determine the optimal combination of high and low-resolution ensemble forecasts on a fixed computational cost providing the best predictive performance. In the case of temperature and precipitation accumulation previous results [1,2] confirm the exi-

stence of an optimal mixture both for raw and post-processed forecasts, where for temperature a parametric, whereas for precipitation non-parametric methods were applied.

Compared with the latter approaches, parametric post-processing of precipitation often provides an improved forecast skill combined with a lower computational cost. In this study, a censored shifted gamma (CSG) distribution is applied for modeling precipitation accumulation, where parameters of the predictive distribution are estimated using an artificial neural network. The predictive performance of this method for the different resolution mixtures is compared with the corresponding raw forecasts and the classical CSG based non-homogeneous regression.

References

[1] Baran, S., Leutbecher, M., Szabó, M., Ben Bouallégue, Z. (2019) Statistical post-processing of dual-resolution ensemble forecasts. *Q. J. R. Meteorol. Soc.* 145(721), 1705-1720

[2] Gascón, E., Lavers, D., Hamill, T. M., Richardson, D. S., Ben Bouallégue, Z., Leutbecher, M., Pappenberger, F. (2019) Statistical post-processing of dual-resolution ensemble precipitation forecasts across Europe. *Q. J. R. Meteorol. Soc.* 145(724), 3218-3235

SAES2

210 Truncated generalized extreme value distribution based EMOS model for calibration of wind speed ensemble forecasts

Patricia Szokol¹, Sándor Baran¹, Marianna Szabó²

1 – University of Debrecen/Faculty of Informatics, Hungary; 2 – Doctoral School of Informatics, University of Debrecen, Hungary

In recent years, ensemble weather forecasting have become a routine at all major weather prediction centres. These forecasts are obtained from multiple runs of numerical weather prediction models with different initial conditions or model parametrizations. However, ensemble forecasts can often be underdispersive and also biased, so some kind of post-processing is needed to account for these deficiencies. One of the most popular state of the art statistical post-processing techniques is the ensemble model output statistics (EMOS), which provides a full predictive distribution of the studied weather quantity.

We propose a novel EMOS model for calibrating wind speed ensemble forecasts, where the predictive distribution is a generalized extreme value (GEV) distribution left truncated at zero (TGEV). The truncation corrects the disadvantage of the GEV distribution based EMOS models of occasionally predicting negative wind speed values, without affecting its favorable properties. The new model is tested on four data sets of wind speed ensemble forecasts provided by three different ensemble prediction systems, covering various geographical domains and time periods. The forecast skill of the TGEV EMOS model is compared with the predictive perfor-

mance of the truncated normal, log-normal and GEV methods and the raw and climatological forecasts as well. The results verify the advantageous properties of the novel TGEV EMOS approach.

SAES1

211 A Bayesian Inference for Intensity Based Load-Sharing Models with Damage Accumulation

Sophie Tchanyou Ganme, Katja Ickstadt

Technische Universität Dortmund, Germany

In a load-sharing system, where the load of the failed components is equally redistributed among the remaining existing components, we consider two models of intensity functions to account for damage accumulation. The first model incorporates the Basquin link function built as a product of the intensity function without damage accumulation and the damage history function of the system until the considered time. The second model is built in a similar way but considers another load-sharing system and it also takes a Weibull function as the link function. We assume non-informative priors and the estimation as well as the posterior computation of the parameters are performed with the use of Bayesian methods such as Markov Chain Monte Carlo methods. A compact simulation study is carried out to assess the behavior of the estimation methods. The performance of the proposed models are relatively considerable. This was illustrated by applying said models to a dataset of experiments on failures of prestressed wire strands.

BS1

212 scVIDE: Single Cell Variational Inference for Designing Experiments

Martin Treppner^{1,2}, Harald Binder^{1,2}

1 – Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany; 2 – Freiburg Center for Data Analysis and Modeling, University of Freiburg, 79104, Freiburg, Germany

To investigate the complexity arising from single-cell transcriptomics data, researchers often resort to deep generative models (DGMs). Using DGMs, the inherent biological signals of gene expression data, such as pathways or gene programs, can be encoded into lower-dimensional latent representations. However, the number of cells that are necessary to learn these latent representations with adequate statistical power is usually unknown.

We propose single-cell variational inference for designing experiments (scVIDE), which is based on a previously introduced variational autoencoder framework. Using a permutation technique to construct a null distribution for the likelihood of the model, we can estimate the statistical power for a given single-cell RNA-sequencing

(scRNA-seq) data set.

We demonstrate the performance of our model using a number of data sets with different complexity stemming from various sequencing protocols. We show how scVIDE can be used to determine the statistical power of scRNA-seq studies and also use the generative properties of the model to generate synthetic data. Using the synthetic cells, we also estimate the statistical power for planning larger, future experiments. To address the increasing data set sizes in transcriptomics studies and the corresponding research questions, we also illustrate the use of scVIDE to estimate the statistical power for sub-clustering. For this purpose, we use data from the human kidney cell atlas and evaluate the power for sub-clustering of the immune cells contained therein.

BSB2

213 Data, parameters and models: Some insights in statistical modeling

Francis Tuerlinckx

KU Leuven, Belgium

In this presentation I will talk about three important insights that I have recently acquired regarding modeling and statistical analysis: (1) It is not about the parameters but about the data; (2) it is not about these data but other data; (3) it is not only about this model but also about other models. I will illustrate these insights with examples of recent work carried out by us or others.

LVM

214 Statistical properties of a prevalence estimator for chronic diseases based on a differential equation: Simulation study in the illness-death-model

Sabrina Tulka¹, Annika Hoyer², Ralph Brinks¹

1 – Lehrstuhl für Medizinische Biometrie und Epidemiologie, Universität Witten/Herdecke, Germany; 2 – Biostatistik und Medizinische Biometrie, Medizinische Fakultät OWL, Universität Bielefeld, Germany

We examine the bias and coverage of the prevalence (p) estimator determined by the ordinary differential equation (ODE) $\frac{dp}{da} = (1 - p)[i - p(m_1 - m_0)]$, where m_0 , m_1 and i are the mortality rates of the non-diseased and diseased and the incidence rate in the time-heterogeneous illness death model (IDM) for chronic diseases with time-scale a .

The solution of the ODE [1] is compared with empirical prevalences from simulated populations transiting through the IDM motivated by diabetes [1]. We use discrete

event simulation [2] for different population sizes (n) and estimate prevalences at specific ages.

We assess bias by comparing empirical estimates from the simulation with the solution of the ODE. Coverage of the 95%-Wald confidence intervals is calculated, which is a reasonable approximation as the prevalence is asymptotically normally distributed (using an argument from [3]).

Bias and coverage are examined by 5000 simulation runs for population sizes $n=10, 50, 100, 200$ and 500 and ages from 30 to 90 (step size 5 years). Prevalences obtained from solving the ODE range from 0.62% (age 30) to 35% (age 80). For small population sizes, coverage is at least 90% and reaches up to 98% . For populations of 100 or more, coverage is close to 95% with a range of 94% to 98% . Good coverage can be explained the central limit theorem applied to the numbers of diseased and non-diseased individuals and their sum.

References

[1] <https://doi.org/10.1155/2018/5091096>

[2] <https://doi.org/10.1371/journal.pone.0106043>

[3] <https://doi.org/10.2307/2342070>

SME1

215 Over-optimistic evaluation and reporting of novel cluster algorithms: An illustrative study

Theresa Ullmann¹, Anna Beer², Maximilian Hünemörder³, Thomas Seidl², Anne-Laure Boulesteix¹

1 – Institute for Medical Information Processing, Biometry, and Epidemiology, LMU Munich, Germany; 2 – Institute for Informatics, LMU Munich, Germany; 3 – Department of Computer Science, CAU Kiel, Germany

When researchers publish new cluster algorithms, they usually demonstrate the strengths of their novel approaches by comparing the algorithms' performance with existing competitors. However, such studies are likely to be optimistically biased towards the new algorithms, as the authors have a vested interest in presenting their method as favorably as possible in order to increase their chances of getting published. Therefore, the superior performance of newly introduced cluster algorithms is over-optimistic and might not be confirmed in independent benchmark studies performed by neutral and unbiased authors. This problem is known among many researchers, but so far, the different mechanisms leading to over-optimism in cluster algorithm evaluation have never been systematically studied and discussed. Researchers are thus often not aware of the full extent of the problem. We present an illustrative study to illuminate the mechanisms by which authors - consciously or unconsciously - paint their cluster algorithm's performance in an over-optimistic light. Using the recently published cluster algorithm Rock as an example, we demonstrate how optimization of the used data sets or data characteristics, of the algorithm's parameters and of the choice of the competing cluster algorithms leads

to Rock's performance appearing better than it actually is. Our study is thus a cautionary tale that illustrates how easy it can be for researchers to claim apparent superiority of a new cluster algorithm. This illuminates the vital importance of strategies for avoiding the problems of over-optimism (such as, e.g., neutral benchmark studies), which we also discuss in our presentation.

CaC

216 Random-effects models for quantifying heterogeneity in infectious disease transmission involving two types of contacts

Steffen Unkel^{1,2}, **Steven Abrams**^{3,4}, **Andreas Wienke**⁵, **Niel Hens**^{3,6}

1 – Department of Medical Statistics, University Medical Center Göttingen, Germany; 2 – School of Life Sciences, University of Siegen, Germany; 3 – I-BioStat, Data Science Institute, Hasselt University, Diepenbeek, Belgium; 4 – Global Health Institute, Department of Family Medicine and Population Health, University of Antwerp, Belgium; 5 – Profile Area Clinical Studies & Biostatistics and Institute of Medical Epidemiology, Biostatistics and Computer Science, Martin-Luther-University Halle-Wittenberg, Germany; 6 – Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Belgium

Epstein-Barr virus (EBV) and Herpes simplex virus type 1 (HSV1) are two members of the family of human herpesviruses. Both infectious agents share the main route of transmission, which is oropharyngeal via an exchange of saliva. Two types of contacts are believed to be involved in transmission: general person-to-person contacts, which peak in childhood, and intimate contacts through kissing, which occur after puberty and peak in teenage years. Hence, modelling the transmission of EBV and HSV1 is a rather complex task. Inspired by the work of Farrington and Whitaker (2005), we propose tailor made time-invariant and time-varying random-effects (frailty) models for the hazard rates, with the aim of quantifying individual heterogeneity in paired case I interval-censored data on EBV and HSV1 infections. The data arose from a serological survey undertaken in the UK in 1994 (Morris et al. 2002). The results could be applied to provide better estimates of key epidemiological parameters, such as the basic reproduction number, R_0 , which is underestimated in case heterogeneity is present in the data but ignored in models for infectious disease transmission.

References

- [1] Farrington C. P. and Whitaker H. J. (2005): Contact surface models for infectious diseases: estimation from serologic survey data, *Journal of the American Statistical Association*, Vol. 100, No. 470, pp. 370-379.
- [2] Morris M.C., Edmunds W.J., Hesketh L.M., Vyse A.J., Miller E., Morgan-Capner P. and Brown D.W.G. (2002): Sero-epidemiological patterns of Epstein-Barr and herpes simplex (HSV-1 and HSV-2) viruses in England and Wales, *Journal of Me-*

217 Introducing Regularisation to Generalised Joint Regression Modelling with an Application to Football

Hendrik van der Wurp, Andreas Groll

TU Dortmund University, Germany

In this work, we propose an extension of the versatile joint regression framework for bivariate count responses of the R package GJRM by Marra and Radice by incorporating an (adaptive) LASSO-type penalty. The underlying estimation algorithm is based on a quadratic approximation of the penalty. The method enables variable selection and the corresponding estimates guarantee shrinkage and sparsity. Hence, this approach is particularly useful in high-dimensional count response settings. The proposal's empirical performance is investigated in a simulation study and an application on FIFA World Cup football data.

ARM2

218 Benchmarking in cluster analysis: Preview of a white paper

Iven Van Mechelen

KU Leuven and the IFCS Task Force on Benchmarking, Belgium

To achieve scientific progress in terms of building a cumulative body of knowledge, careful attention to benchmarking is of the utmost importance. This means that proposals of new methods of data pre-processing, new data-analytic techniques, and new methods of output post-processing, should be extensively and carefully compared with existing alternatives. To date, benchmarking and recommendations for benchmarking have been frequently seen in the context of supervised learning. Yet, unfortunately, there has been a dearth of guidelines for benchmarking in an unsupervised setting, with the area of clustering as an important subdomain. To address this problem, a Task Force within the International Federation of Classification Societies (IFCS) is currently preparing a white paper on benchmarking in cluster analysis. In this white paper discussion is given to general and fundamental conceptual underpinnings of benchmarking in the field of cluster analysis by means of simulated as well as empirical data. Subsequently, the specifics of how to address benchmarking questions in clustering are dealt with, and foundational recommendations are made. In the present talk I will offer a sneak preview of this white paper.

CaC

219 A parametric additive hazard model for time-to-event analysis

Dina Voeltz^{1,2}, Amelie Forkel², Anke Schwandt³, Oliver Kuss⁴, Annika Hoyer¹

1 – Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany; 2 – Department of Statistics, Ludwig-Maximilians-University Munich, Germany; 3 – Nuvisan GmbH, Neu-Ulm, Germany; 4 – Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Germany

Non- and semi-parametric models for analysing time-to-event outcomes focussing on estimating hazard ratios continuously fall in disgrace in terms of interpretation, technical implementation, and flexibility. In recent years, hazard ratios have been criticized for instance for their misleading interpretation as relative risks and their non-collapsibility. An approach that does not suffer from these disadvantages is the additive hazard model [1,2]. However, this model is rarely used in practice because it assumes a semi-parametric additive hazard. This renders computation and interpretation complicated, e.g., effect estimates are necessarily time-dependent and can only be given via graphs. As a remedy, we propose a new parametric additive hazard model which allows to communicate results on the original time rather than on the hazard scale. Being an essentially parametric model, it has survival, hazard and density functions directly available. Parameter estimation is straightforward by maximizing the log likelihood function. As an exemplary application, we illustrate the model for different parametric distributional assumptions using data from the HALLUCA study [3]. We report the resulting parameter estimates and survival curves which fit well with routinely used Kaplan-Meier curves.

References

- [1] Aalen OO. A linear regression model for the analysis of life times. *Stat Med* 1989;8:907-925.
- [2] Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika* 1994;81:61-71.
- [3] Bollmann A, Blankenburg T, Haerting J, Kuss O, Schütte W, Dunst J, Neef H. HALLUCA study. Survival of patients in clinical stages I-IIIb of non-small-cell lung cancer treated with radiation therapy alone. *Strahlenther Onkol.* 2004;180(8):488-96.

SEHA1

220 Flexible joint models for multivariate longitudinal and time-to-event data using a functional principal components representation of shared random effects

Alexander Volkman¹, Nikolaus Umlauf², Sonja Greven¹

1 – Humboldt-Universität zu Berlin, Germany; 2 – Universität Innsbruck, Austria

The joint modeling of multiple longitudinal biomarkers together with a time-to-event outcome is a challenging modeling task. The high dimensionality of generalized linear mixed effects models is often too computationally restrictive for shared parameter joint models, especially when the subject-specific marker trajectories follow highly nonlinear courses. We therefore propose a parsimonious representation of the shared random effects using an approximation with multivariate functional principal components. This allows better scalability as the dimension of the random effects does not directly increase with the number of markers, only with the number of principal component basis functions used in the approximation of the random effects. The functional principal component representation allows to estimate highly flexible subject-specific random trajectories as no parametric assumptions for the random effects are needed. The modeled subject-specific deviations from the mean can then be distinctly different for each biomarker. We build on the framework of flexible Bayesian additive joint models (R-package ‘bamlss’), which also supports nonlinear estimates of covariate effects via Bayesian P-splines. To include the functional principal components as a parsimonious basis in the estimation of the joint model, we propose to estimate them in a preliminary step. There, univariate functional principal component analyses for each marker can be combined to estimate the multivariate functional principal components. We apply the joint model to a study on acute cellular rejection after pediatric liver transplantation with a high number of longitudinal biomarkers. The appropriateness of our approach is illustrated by a simulation study and we compare it to competing methods.

ARM2

221 Power of Logrank and RMST tests for non-constant hazard ratio

Reinhard Vonthein, Christina Hölsken

Universität zu Lübeck, Germany

Background: Recent trials revealed increasing hazard ratios (HR) in oncology[1] and decreasing HR in cardiology[2]. One-dimensional estimands have to be interpreted conditional on the time frame, then.

Purpose: To compare the power of the logrank test to that of the RMST test in case of non-constant HR.

Methods: A mixture of a Weibull and an exponential distribution was fitted to published results as a starting value. Simulations with typical trial sizes and time fra-

mes across the 4-dimensional parameter space were selected to have realistic effects. In one application, two distribution functions were chosen so that the cumulative incidence function of a blinded interim analysis was half-way between them.

Results: A simple, but tentative, rule emerged for the scenarios studied: If HR is increasing, the RMST test has higher power than the logrank-test. If HR is decreasing, the logrank-test has higher power than the RMST test. A shiny app for pointwise simulations is available from the author.

Conclusion: The above hypothesis should be proven analytically. The absolute effect measure RMST should be reported along with the relative measure HR with a common time frame.

References

- [1] Alexander, BM. et al. (2018) Hazards of Hazard Ratios — Deviations from Model Assumptions in Immunotherapy. *N Engl J Med* 378:1158-1159
- [2] Uno, H. et al. (2015). Alternatives to hazard ratios for comparing efficacy or safety of therapies in noninferiority studies. *Ann Intern Med* 163(2):127-134.

SPMDI

222 Using methods to identify influential points in high-dimensional data

Shuo Wang, Edwin Kipruto, Willi Sauerbrei

Faculty of Medicine and Medical Center - University of Freiburg, Germany

The increasing availability of high-dimensional data (HDD) prompted the development of high-dimensional regression methods. The corresponding regression diagnostic technique still seems to be underdeveloped. Most of the classical methods proposed in low-dimensional data (LDD) can hardly be directly applied in the high-dimensional setting because of conceptual differences in regression methods. A couple of specific methods to detect influential points (IPs) in HDD have been proposed; however, it seems that they are hardly used in practice. We will present several methods to identify IPs in HDD, and apply them in several publicly available datasets. Boosting is used as the major regression method, with lasso also used in some parts. We consider two types of tuning parameters, namely, fixed and adaptive tuning parameters. These IPs methods are mainly based on two strategies: data transformation and case deletion, with the latter commonly referring to leave-K-out. The transformation method was initially proposed to handle extreme points (Boulesteix and Sauerbrei, 2011. <https://doi.org/10.1002/bimj.201000189>), and here we will compare the regression models for the transformed and untransformed data. These methods assess the potential effects of cases on selected models (De Bin, 2017. <https://doi.org/10.1016/j.csda.2017.07.001>), prediction (Rajaratnam, 2019. <https://doi.org/10.1080/10618600.2019.1598869>), and marginal correlations (Zhao, 2013. <https://doi.org/10.1214/13-AOS1165>). We will also illustrate other techniques proposed that may be helpful in IPs detection in HDD, such as feature rankings, hierarchical clustering and dendrogram. Our results are

helpful in understanding some important issues of IP detection in HDD.

SHDA2

223 Innovative design and analysis approaches for master protocols

James Wason

Newcastle University, United Kingdom

New statistical methods and trial designs have a big role to play in improving clinical trials. They can lead to more efficiency, better evidence and better outcomes of patients enrolled on the trial. This has been highlighted by the COVID-19 pandemic, where innovative methods such as adaptive platform trials have played a major role in improving treatment.

In this talk, I will discuss several areas of research about using an innovative class of trial design called master protocols. Master protocols allows combining several separate but related clinical trials together. This provides substantial operational and statistical advantages. However, it also introduces several statistical issues that new methods are required for addressing.

My talk will concentrate on how master protocol approaches pioneered in oncology, such as basket and umbrella trials, may be made useful in other areas, especially chronic inflammatory diseases.

I will finish by considering other areas of methodology research that are required to enable maximum utility of the master protocol approach.

SPMDI

224 Efficient multivariate inference in general factorial diagnostic studies

Maximilian Wechsung, Konietschke Frank

Charité-Universitätsmedizin Berlin, Germany

We propose new simultaneous inference methods for diagnostic trials with elaborate factorial designs. Instead of the commonly used total area under the receiver operating characteristic curve, we focus on partial areas under the curve as parameters of interest. We construct a nonparametric multiple contrast test for these parameters and show that it asymptotically controls the family-wise type one error rate. Finite sample properties of this test are investigated in a series of computer experiments. We provide empirical and theoretical evidence supporting the conjecture that statistical inference about partial areas under the curve is more efficient than inference about the total area under the curve.

RNS1

225 Robust Detection for Change-Points in Functional Time Series based on Spatial Signs and Bootstrap

Lea Wegner, Martin Wendler

Otto-von-Guericke Universität Magdeburg, Germany

One main strategy in changepoint detection for time series is to project the data on a finite-dimensional space with techniques such as functional principal components. In contrast, there are recent proposals to base the statistical tests on the full functional information, typically modeled as Hilbert-space-valued time series. Up to now, test statistics for changepoint detection in functional time series are based on sample means and outliers can influence the test result.

Generalizing the Wilcoxon statistic, we have constructed a new functional version of a two-sample U-statistic with a bounded antisymmetric kernel. We will present limit theorems for U-statistics with values in Hilbert spaces and deduce the asymptotic distribution of our changepoint statistic. Because of the boundedness of the kernel, the statistic is indeed robust against outliers.

Since this class of test statistics does not rely on dimension reduction, the limit distribution provides an infinite-dimensional covariance operator as a parameter, which is difficult to estimate. Because of this, we propose a new variant of the dependent wild bootstrap adapted to U-statistics in Hilbert spaces.

RNS2

226 Classification of atmospheric circulation patterns using a smoothed deep learning approach

Maximilian Weigert¹, Magdalena Mittermeier², David Rügamer³, Ralf Ludwig², Helmut Küchenhoff¹

1 – Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany; 2 – Department of Geography, LMU Munich, Germany; 3 – Department of Statistics, LMU Munich, Germany

European weather is largely driven by the constellations of large-scale high- and low-pressure systems categorized into atmospheric circulation patterns. One of the most relevant records of circulation patterns is the Hess and Brezowsky synoptic catalog providing a categorization into 29 classes. Still today, daily air pressure systems are manually assigned to one of these classes by experts. For large climate simulation ensembles such as the one provided by the ClimEx project (<https://www.climex-project.org/>), however, an automated classification method is required to efficiently apply this classification scheme.

The field of deep learning offers new possibilities for this classification task taking into account the underlying spatio-temporal data structure. Based on a combination of the catalog of labeled circulation patterns and atmospheric variables, we propose a smoothed convolutional neural network approach to objectively classify the 29 circulation patterns. We address various data-specific challenges inherent to

circulation pattern classification and also present in other climate patterns such as subjective labels and unambiguous transition periods. An ablation study compares the performance of networks containing different types of smoothing.

For model training, we use the publicly available ERA-20C reanalysis data with daily information on atmospheric variables over 111 years on a spatial resolution of $5^\circ \times 5^\circ$ degrees. The obtained model is applied to the 50-member ClimEx ensemble data to analyze the occurrence of specific circulation patterns over time within a given climate scenario. This way, our work can help to identify drivers of extreme events in climate simulations and unveil the impact of climate change on these drivers.

AIML1

227 Data Competence Network – Data Literacy Education an der Technischen Universität Dortmund

Henrike Weinert, Katja Ickstadt, Henrik Müller

Technische Universität Dortmund, Germany

Die fortschreitende Datafizierung verändert den gesellschaftlichen und beruflichen Alltag rapide, der versierte Umgang mit Daten wird zur Basiskompetenz, so dass Data-Literacy-Kompetenzen in sämtlichen Disziplinen und Bereichen benötigt werden. Die Technische Universität Dortmund hat sich das Ziel gesetzt, ihre 34.500 Studierenden in rund 80 Studiengängen darauf vorzubereiten, die gesellschaftliche Entwicklung mitgestalten, den wissenschaftlichen Fortschritt vorantreiben und neue Geschäftsmodelle entwickeln zu können. Die TU Dortmund verfügt über einen einzigartigen Kompetenzenmix mit starken naturwissenschaftlich-technische Disziplinen, die einen fächerübergreifenden Fokus auf Datenwissenschaften pflegen, sowie gesellschafts- und geisteswissenschaftliche Disziplinen, also traditionell datenfernen Fachkulturen. Im Rahmen des Projekts Data Competence Network (DaCoNet) wird ein Konzept entwickelt, um beide Seiten stärker miteinander zu vernetzen und die reichhaltige Data-Science-Expertise der TU Dortmund auch anderen Fächern zugänglich zu machen, die mit der fortschreitenden Digitalisierung konfrontiert sind.

In diesem Vortrag stellen wir das DaCoNet-Konzept vor, das auf einem dreistufigen Zielsystem basiert: Sensibilisieren, Qualifizieren, Multiplizieren. Unser Ansatz zur Vermittlung von Grundverständnis für Daten und Datenverarbeitung nutzt sowohl neu geschaffene Kurse als auch bestehende Kurse für Statistik oder Data Science, die auf die Gesamtsicht auf den Wertschöpfungsprozess aus Daten im Sinne von Data Literacy angepasst werden. Wir gehen auf die Erfolgsfaktoren und Herausforderungen ein, die sich insbesondere aus dem heterogenen Teilnehmerfeld ergeben. Wichtig sind dabei reale Datenanwendungen aus einem breiten Spektrum möglichst vieler Disziplinen, um einerseits die Bedürfnisse der verschiedenen Fächer zu berücksichtigen, andererseits die Interdisziplinarität und die Zusammenarbeit zwischen verschiedenen Fachbereichen, sowohl auf Seiten der Lehrenden als auch der Studierenden zu stärken.

SLSE1

228 Truncating the exponential with a uniform distribution

Rafael Weißbach¹, **Dominik Wied**²

1 – University of Rostock, Germany; 2 – University of Cologne, Germany

or a sample of Exponentially distributed durations we aim at point estimation and a confidence interval for its parameter. A duration is only observed if it has ended within a certain time interval, determined by a Uniform distribution. Hence, the data is a truncated empirical process that we can approximate by a Poisson process when only a small portion of the sample is observed, as is the case for our applications. We derive the likelihood from standard arguments for point processes, acknowledging the size of the latent sample as the second parameter, and derive the maximum likelihood estimator for both. Consistency and asymptotic normality of the estimator for the Exponential parameter are derived from standard results on M-estimation. We compare the design with a simple random sample assumption for the observed durations. Theoretically, the derivative of the log-likelihood is less steep in the truncation-design for small parameter values, indicating a larger computational effort for root finding and a larger standard error. In applications from the social and economic sciences and in simulations, we indeed, find a moderately increased standard error when acknowledging truncation.

Reference

[1] Weißbach, R., Wied, D. Truncating the exponential with a uniform distribution. Stat Papers (2021). <https://doi.org/10.1007/s00362-021-01272-x>

SEHA4

229 Weighted Change-Point Tests based on 2-Sample-U-Statistics

Martin Wendler, **Herold Dehling**, **Kata Vuk**

Otto-von-Guericke-Universität Magdeburg, Germany

Two-sample-U-statistics with bounded, antisymmetric kernels can be used to construct robust change-point tests, which are not sensitive to outliers. We use a weighted combination of such statistics to improve the power for early or late changes in a time series. We show the convergence of the test statistic to an extreme value distribution and illustrate the performance with some simulation results. Furthermore, we consider so called epidemic changes and use convergence in the space of Hölder-continuous functions to derive the limit.

RNS2

230 The role of estimands in machine learning algorithm evaluation

Max Westphal, Rieke Alpers

Fraunhofer Institute for Digital Medicine MEVIS, Germany

The goal of algorithm evaluation in machine learning (ML) is the quantification of the generalizability of a ML algorithm or model. As the apparent training performance is not indicative of the targeted out-of-sample performance, data splitting is required to obtain realistic evaluation results. Various procedures have been proposed for that matter, e.g. different (grouped, nested) cross-validation variants. For ML practitioners, it is however often not clear which of the available approaches is most suitable for their specific problem.

In this work we connect this issue to the ongoing estimand discussion in statistics. We performed a literature review to summarize common solutions and pitfalls in the context of clinical risk prediction modelling and decision support. Moreover, we conducted several numerical experiments with real and simulated data to investigate which methods are (un)suited for an unbiased estimation of different types of generalization performance, e.g. reproducibility and regional or temporal transportability.

Our results illustrate that the validity of all (internal, internal-external, external [1]) evaluation attempts relies on assumptions regarding the (planned) utilization context, i.e. when, where and under which conditions a ML algorithm is implemented. However, such assumptions are rarely stated explicitly and likely often violated. We derived a structuring taxonomy and guiding principles to enable ML practitioners to identify and describe the most relevant estimand(s) and an appropriate methodological approach for their evaluation experiments.

Reference

[1] Steyerberg, E. W., & Harrell Jr, F. E. (2016). Prediction models need appropriate internal, internal-external, and external validation. *Journal of clinical epidemiology*, 69, 245.

AIML2

231 Bayesian meta-analysis for exact and interval censored binomial outcomes

Manuel Wiesenfarth^{1,2}, Jasmin Wandel³

1 – Cogitars GmbH, Germany; 2 – German Cancer Research Center (DKFZ); 3 – Berner Fachhochschule, Switzerland

Hierarchical models are routinely used in the context of meta-analyses with binomial outcomes. However, frequently in some strata the number of successes/events may only be known to fall in some interval while strata sizes are known. Thus, only

bounds for success rates may be observed for some strata. Such interval censored binomial outcomes may for example occur in surveys on side effects of a drug with categories “yes”, “no” and “unsure” and which may in addition contain missing values. Here, the proportion of patients actually experiencing a side effect is unknown and the proportion of positive answers only provides a lower bound (Pruszyński, 2010).

We implement Bayesian hierarchical logistic and beta-binomial models for exact and interval-censored binomial outcomes which to our knowledge have not been studied before. It is shown that apparently vague priors in the logistic regression version may incorporate undesirable information.

Simulation studies are presented as well as an application in the context of health technology assessments (HTA) where the event of interest is presence of an adverse event (AE) of any grade. Here, some studies may only report the number of severe AEs, thus only providing a lower bound for the number of any grade AEs.

We show that properly accounting for interval censoring may contribute limited information to the overall mean of the hierarchical model, but may improve interval estimates and reveal important information which would be hidden otherwise.

Reference

[1] Pruszyński (2010). Bayesian models for discrete censored sampling and dose finding (Doctoral dissertation).

MA2

232 Distinguishing between breaks in the mean and breaks in persistence under long memory

Simon Wingert, Mwasi Paza Mboya, Philipp Sibbertsen

Leibniz University Hannover, Germany

A procedure to discriminate between stationarity, a break in the mean and a break in persistence in a time series that may exhibit long memory is introduced. The asymptotic properties of test statistics based on the CUSUM statistic are studied. In a Monte Carlo study we further analyze the finite sample properties of the procedure. An application to inflation rates shows the potential of our procedure for future research.

TSA2

233 Using LASSO regression to estimate the population-level impact of pneumococcal conjugate vaccines

Anabelle Wong^{1,2}, Sarah Kramer¹, Marco Piccininni^{2,3}, Jessica L. Rohmann^{2,3}, Tobias Kurth², Sylvie Escolano⁴, Ulrike Grittner^{5,6}, Matthieu

Domenech de Cellès¹

1 – Max Planck Institute for Infection Biology, Berlin, Germany; 2 – Institute of Public Health, Charité – Universitätsmedizin Berlin, Germany; 3 – Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Germany; 4 – Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Disease, Inserm U1181 (B2PHI), UVSQ, University Paris Saclay, Institut Pasteur, Paris, France; 5 – Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, Germany; 6 – Berlin Institute of Health, Charité – Universitätsmedizin Berlin, Germany

The pneumococcal conjugate vaccines (PCVs) protect against diseases caused by *Streptococcus pneumoniae*, such as meningitis, bacteremia, and pneumonia. It is challenging to estimate their population-level impact due to the lack of a perfect control population and the subtleness of signals when the endpoint – like all-cause pneumonia – is non-specific. Here we present a new method for PCV impact estimation – using LASSO regression to predict the counterfactual outcome for vaccine impact inference. First, we used a simulation study to test the performance of LASSO regression and established methods including the synthetic control (SC) approach. We found that LASSO achieved accurate and precise estimation, even in complex simulation scenarios where causal variables were removed. Then we applied LASSO to published real-world data from Chile, Ecuador, Mexico, and the US, and found that it yielded similar estimates of vaccine impact to SC. Overall, the performance of LASSO was comparable to SC. Our method is easy to implement and to interpret and can therefore be applied to study the impact of PCVs and of other vaccines.

SME3

234 The Evolution of Statistical Process Monitoring

William H. Woodall

Virginia Tech, United States of America

This presentation provides a relatively non-technical overview of the development of statistical process monitoring methods from their introduction by Walter Shewhart in the Industry 2.0 era to the challenges now faced in Industry 4.0. Monitoring methods have been adapted and extended over time to reflect increases in the amount of available data, changes in data structure and characteristics, and increases in computing power. It is argued that it is increasingly the case in practice that monitoring methods should be desensitized, counter to the prevailing research quest for steadily increasing sensitivity in detecting process changes. A relatively simple cumulative sum (CUSUM) method is described for tuning detection performance. The discussion also includes some comments on quality management and some important historical milestones.

P4

235 Meta-analysis of mean values with unreported variances: estimation bias due to variance heteroscedasticity when common joint variance is assumed

Justyna Wünnstel, Peter Schlattmann

Universitätsklinikum Jena, Germany

Introduction: Sangnawakij et al (2019, 2020) proposed new estimators for the mean effect and between-study variance in the meta-analyses of the mean difference between independent groups, when study specific variances are unknown and homoscedasticity has to be assumed.

Methods: Sangnawakij's estimators for the mean difference were investigated in a simulation study under heteroscedasticity. Study specific variances σ^2 were generated from Γ -distributions with different shape and scale values (1,2; 2,2; 6,1). Study-specific effects D_i were generated from normal distributions with means $\mu = 0, 2$ and 5. The heterogeneity variance τ^2 was set to 0, 4, 6 and 10. Sample sizes for the respective meta-analyses were $k = 5, 10, 30$ and 50 studies. The performance of the estimators $\hat{\mu}$ and $\hat{\tau}^2$ was evaluated in terms of bias, mean squared error (MSE) and coverage probability.

Results: Sangnawakij's estimator $\hat{\mu}$ had a well estimated mean value, very small bias and MSE in the fixed effects (FE) and random effects (RE) model. Coverage probability for $\hat{\mu}$ was 71.17% for the FE model and 69.74% for the RE model. Sangnawakij's estimator for τ^2 showed considerable MSE but small bias.

References

- [1] Sangnawakij, P., Böhning, D., Niwitpong, S. A., Adams, S., Stanton, M., & Holling, H. (2019). Meta-analysis without study-specific variance information: Heterogeneity case. *Statistical methods in medical research*, 28(1), 196-210
- [2] Sangnawakij, P., & Böhning, D. (2020). A simple test for the difference of means in meta-analysis when study-specific variances are unreported. *Journal of Statistical Computation and Simulation*, 90(14), 2524-2536.

MA1

236 Bayesian Learning in a General Class of Multivariate Nonlinear Dynamic Panel Data Models

Ilya Zarubin

Cologne Graduate School (CGS) - University of Cologne, Germany - Institute of Econometrics and Statistics

We propose a unified Bayesian modelling and inference approach in a general class of multivariate dynamic nonlinear panel (longitudinal) data models comprising various multivariate count and fractional response distributions.

The proposed state space approach is combined with structured additive distributional regression to model the parameters of complex multivariate distributions as functions of stochastic latent state processes as well as structured additive predictors. The latter are additive compositions of different types of covariate effects, e.g. non-linear effects of continuous covariates, random effects, or spatial effects. Inference is realized by a generic, computationally efficient variant of Particle Markov Chain Monte Carlo methods (Andrieu et al. 2010) termed Particle Gibbs (Lindsten et al. 2014). Particle Gibbs utilizes Sequential Monte Carlo to approximate the high dimensional multivariate latent state process and Markov Chain Monte Carlo for posterior parameter inference.

We investigate the performance of several SMC procedures in high dimensions, including the standard Bootstrap Particle Filter, the auxiliary Particle Filter, the iterated auxiliary Particle Filter and Particle Efficient Importance Sampling. Two applications illustrate our approach in practice: a multivariate compositional count model of COVID-19 pandemic data and a multivariate fractional model for the energy mix in the United States.

References

- [1] Particle Markov chain Monte Carlo methods (Christophe Andrieu, Arnaud Doucet, Roman Holenstein – Journal of the Royal Statistical Society: Series B)
- [2] Particle Gibbs with ancestor sampling (F. Lindsten, Michael Jordan, Thomas Schoen – Journal of Machine Learning Research, 2014)

BS2

237 Bayesian Boosting for Simultaneous Estimation and Selection of Fixed and Random Effects in High-Dimensional Mixed Models

Boyao Zhang, Colin Griesbach, Elisabeth Bergherr

Georg-August-Universität Göttingen, Germany

election of relevant fixed and random effects without prior choices made from possibly insufficient theory is important in mixed models. Inference with current boosting techniques suffers from biased estimates of random effects and the inflexibility of random effects selection. We propose a new inference method “BayesBoostcombining Bayesian methods and gradient boosting, which performs estimation and selection of fixed and random effects in mixed models simultaneously. The method introduces a novel selection strategy for random effects, which allows for computationally fast selection of random slopes even in high dimensional data structures. Additionally, the new method not only overcomes the shortcomings of Bayesian inference in giving precise and unambiguous guidelines for the selection of covariates by benefiting from boosting techniques, but also provides Bayesian ways to construct estimators for the precision of the parameters such as variance or credible interval, which is not available in conventional boosting frameworks. The effectiveness of the new

approach can be observed via simulation and in a real world data example.

References

- [1] Tutz, G. and Groll, A. (2010). Generalized linear mixed models based on boosting. In *Statistical Modelling and Regression Structures*, pages 197-215. Springer.
- [2] Griesbach, C., Säfken, B., and Waldmann, E. (2021). Gradient boosting for linear mixed models. *The International Journal of Biostatistics*, 17(2):317-329.

BS1

238 Simulation-based Design Optimization for Statistical Power: Utilizing Machine Learning

Felix Zimmer¹, Rudolf Debelak^{1,2}

1 – University of Zurich, Switzerland; 2 – University of Leipzig, Germany

Planning an adequately powered research design increasingly goes beyond determining an appropriate sample size. More sophisticated scenarios require the simultaneous tuning of various parameters of the design and can only be tackled using Monte Carlo simulation. In addition to the desired statistical power, we want to ensure the optimality of the solutions with respect to a cost metric, such as the financial cost of a study. We introduce a surrogate modeling method to optimize both power and cost with the help of machine learning. Applications include finding design parameters that imply a desired power at minimum cost or, alternatively, maximum power given a cost threshold. As surrogate models, which guide the search process, we use Gaussian process regression and support vector regression. We demonstrate the performance of the method in an extensive simulation study using various hypothesis test scenarios with single and multidimensional design parameters using scenarios from classical statistics, multilevel modeling, and item response theory.

SBES

239 Opportunities, Challenges, and Limitations of Social Media Recruitment for Rare and Hard-to-Reach Populations

Zaza Zindel, Simon Kühne

Bielefeld University, Germany

A growing portion of the global population is active on social media platforms. This opens up new opportunities for survey research. A rather novel approach uses online social media platforms like Facebook or Instagram to recruit web survey participants. These platforms are a promising resource to promote surveys and recruit participants, especially if the survey targets a rare population. They provide relatively inexpensive access to a large number of potential respondents and often

offer a wealth of additional information to identify and address otherwise hard-to-reach populations. However, this recruitment presents a number of challenges, including under-coverage and self-selection, fraud and fake interviews, and issues with weighting survey data to provide unbiased estimates. This talk will provide insights into the opportunities and hurdles that social media platforms offer for survey research. Two social media samples of rare populations will be presented and discussed. Furthermore, by comparing a social media sample with a simultaneously collected, probability-based, face-to-face sample, the opportunities, challenges, and limitations of social media recruitment will be evaluated compared to traditional sampling methods.

SM1

Index

- Aaron, Catherine, 35
Abbott, Sam, 53
Abrams, Steven, 169
Adrion, Christine, 44
Afach, Sivem, 70
Albrecht, Wiebke, 96
Aleksandrov, Boris, 70
Allignol, Arthur, 74
Alpers, Rieke, 178
Amelia, Nicholas, 60
an der Heiden, Matthias, 53
Anatolyev, Stanislav, 35
Apfel, Nicolas, 35, 36
Atmaca, Saide, 36
Avraam, Demetris, 157
Axenfeld, Julian B., 37
- Bach, Paul, 38
Bach, Philipp, 38, 160
Balestra, Chiara, 39
Banerjee, Soumya, 157
Baran, Ágnes, 40
Baran, Sándor, 40, 164, 165
Bauer, Alexander, 40
Baumeister, Marlène, 41
Baumer, Benjamin, 60
Becher, Heiko, 145
Becker, Marc, 42
Beckmann, Lars, 42
Beer, Anna, 168
Beering, Carina, 43
Behning, Charlotte C., 44
Beisemann, Marie, 45
Benda, Norbert, 91
Bender, Andreas, 110
Bender, Ralf, 113, 114, 142
Benner, Axel, 68
Berens, Florian, 45, 46
Berger, Moritz, 160
Bergherr, Elisabeth, 78, 134, 182
Berthold, Frank, 61
Beyersmann, Jan, 74, 143
- Bibinger, Markus, 47
Biehler, Rolf, 73
Biella, Marco, 134
Binder, Harald, 79, 85, 166
Binder, Karin, 163
Binder, Nadine, 47, 48
Bischl, Bernd, 42, 110
Bishop, Tom R. P., 157
Bjermo, Jonas, 49
Blesch, Kristin, 49
Blocher, Hannah, 50
Bodory, Hugo, 51
Böhning, Dankmar, 88
Boland, Leonie, 147
Bolko, Anine E., 51
Borgen, Nicolai Topstad, 82
Borsch, Marvin, 52
Bosch, Philipp, 155
Boulesteix, Anne-Laure, 52, 116, 168
Bracher, Johannes, 53
Brinks, Ralph, 89, 167
Broderick, Tamara, 54
Bruch, Christian, 37, 144
Brunotte, Guy-Niklas, 54
Buczak, Philip, 55
Budszuhn, Timo, 55
Büchter, Theresa, 163
Bürkner, Paul, 56
Burgard, Jan Pablo, 56, 85
Burton, Paul, 157
Butts, Carter Tribbley, 57
- Carcy, Salomé, 125
Carlan, Manuel, 134
Carollo, Angela, 57
Carpenter, James, 58
Castro-Prado, Fernando, 59
Cederbaum, Jona, 79, 87
Çetinkaya-Rundel, Mine, 60
Chaimani, Anna, 70
Chen, Sina, 155
Chernozhukov, Victor, 38

Christensen, Kim, 51
 Classe, Franz Leon, 61
 Collins, Gary S., 136
 Cornesse, Carina, 156

 Danzer, Moritz Fabian, 61
 de Cellès, Matthieu Domenech, 180
 Debelak, Rudolf, 183
 Deffner, Veronika, 62
 Dehling, Herold, 77, 177
 Demetrescu, Matei, 63
 Dette, Holger, 47, 63
 Di Stefano, Roberta, 120
 Dicken, Volker, 92
 Didelez, Vanessa, 64
 Dietz, Sabrina, 74
 Dimitriadis, Timo, 64
 Ditzhaus, Marc, 41, 65
 Djeudeu, Dany, 66
 Dobler, Dennis, 65
 Doebler, Philipp, 45, 55
 Doehler, Sebastian, 66
 Dormuth, Ina, 67
 Duda, Julia Christin, 68
 Dürre, Alexander, 77

 Edelmann, Dominic, 59, 68
 Eilers, Paul H. C., 57, 117
 Engel, Joachim, 69
 Engelke, Sebastian, 141
 Erfurth, Kerstin, 138
 Escolano, Sylvie, 180
 Evrenoglou, Theodoros, 70

 Fackle Fornius, Ellinor, 49
 Faldum, Andreas, 61
 Farbmacher, Helmut, 35
 Fassò, Alessandro, 117
 Faymonville, Maxime, 70
 Felderer, Barbara, 71, 144, 156
 Fernández-Fontelo, Amanda, 72
 Fissler, Tobias, 72
 Fleckenstein, Monika, 44
 Fleischer, Yannik, 73
 Fokianos, Konstantinos, 115
 Foraita, Ronja, 64
 Forkel, Amelie, 171
 Frank, Sandra, 74
 Freise, Fritjof, 74
 Fridljand, Daniel, 75

 Fried, Roland, 55, 77, 115
 Friede, Tim, 91, 113, 129, 142
 Friedrich, Sarah, 143
 Fritz, Cornelius, 76
 Fryzlewicz, Piotr, 76
 Funk, Sebastian, 53

 Gampe, Jutta, 57
 Geisler, Benjamin, 92
 Genuneit, Jon, 65
 George, Ann Cathrice, 123
 Gerharz, Alexander, 149
 Gertheiss, Jan, 89, 156
 Gneiting, Tilmann, 53
 Gnettner, Felix, 77
 Goeman, Jelle J., 59
 Goerdts, Lukas, 44
 Görz, Sheila, 77
 Greven, Sonja, 72, 163, 172
 Griesbach, Colin, 78, 182
 Grittner, Ulrike, 180
 Groll, Andreas, 149, 170
 Groß, Jürgen, 123
 Gross, Marcus, 138
 Grothe, Oliver, 78
 Günther, Felix, 53
 Gutzeit, Maurilio, 79

 Hackenberg, Maren, 79
 Hadam, Sandra, 80
 Häussler, Tim, 145
 Hahn, Bianka, 147
 Hailer, Davide, 53
 Halbleib, Roxana, 97
 Hans, Nicolai, 81
 Hardin, Johanna, 60
 Hashemi, Reza, 94
 Hastie, Trevor John, 81
 Haupt, Andreas, 82
 Heinze, Georg, 83
 Hengstler, Jan G., 96
 Hennig, Christian, 83
 Henninger, Felix, 72
 Hens, Niel, 169
 Hepp, Tobias, 84
 Hernan, Miguel, 84
 Herrmann, Carolin, 67
 Hess, Moritz, 85
 Heyder, Stefan, 53, 85, 90
 Hilbert, Sven, 86

Hobert, Sebastian, 45, 46
 Höhle, Michael, 87
 Hölsken, Christina, 172
 Hofacker, Angelina, 87
 Hoffmann, Clara, 98
 Hoffmann, Sabine, 52
 Hoga, Yannick, 64
 Holling, Heinz, 74, 88, 127
 Holovchak, Anastasiia, 89
 Holz, Frank G., 44
 Hornung, Roman, 116
 Hoshiyar, Aisouda, 89
 Hosseinkouchak, Mehdi, 63
 Hotz, Thomas, 53, 85, 90
 Hoyer, Annika, 89, 167, 171
 Huber, Cynthia, 91
 Huber, Martin, 35, 51
 Huebner, Marianne, 145
 Hünemörder, Maximilian, 168
 Hütter, Mandy, 134
 Huisman, Mark, 105
 Hurley, Catherine, 91

Ickstadt, Katja, 66, 166, 176
 Ils, Alexandra, 162
 Ismaël, Castillo, 59

Jackson, Dan, 93
 Jacqmin-Gadda, Helene, 94
 Jäckle, Annette, 92
 Jäckle, Sonja, 92
 Jahn, Malte, 94
 Jalal, Hawre, 40
 Jansen, Christoph, 50
 Jansen, Katrin, 88
 Janssen, Arnold, 65
 Jentsch, Carsten, 70, 140
 Jeske, Daniel R., 114
 Josse, Julie, 95

Kappenberg, Franziska, 96
 Katzfuss, Matthias, 96
 Kazak, Ekaterina, 97
 Kern, Christoph, 61
 Kertel, Maximilian, 97
 Kiers, Henk A. L., 89
 Kieser, Meinhard, 129
 Kieslich, Pascal J., 72
 Kim, Hee-Young, 94
 Kipruto, Edwin, 173

Kiraly, Franz, 110
 Kirch, Claudia, 77, 98
 Kirschning, Thomas, 147
 Klaassen, Sven, 160
 Klein, Nadja, 38, 81, 98–100, 103, 108, 164
 Klein, Philipp, 98
 Kleinemeier, Jana, 100
 Klinkhammer, Hannah, 100
 Klokman, Sebastiaan, 134
 Kneib, Thomas, 101, 134
 Knop, Eric Samuel, 102
 Knüppel, Malte, 102
 Kock, Lucas, 103
 Koenen, Niklas, 103
 König, Inke R., 108
 König, Jochem, 148
 Körtner, John, 155
 Köster, Frank, 104
 Kolb, Jan-Philipp, 144
 Konietzschke, Frank, 143, 174
 Koschate, Jessica, 125
 Kramer, Sarah, 180
 Krause, Linda, 105
 Krause, Robert Wilhelm, 105
 Krawitz, Peter, 100
 Krennmair, Patrick, 106
 Kress, Sara, 111
 Kreuter, Frauke, 72
 Kreutz, Clemens, 106, 135
 Kröker, Ilja, 56
 Krüger, Fabian, 102
 Krüger, Tyll, 85
 Küchenhoff, Helmut, 53, 175
 Kück, Jannis, 71, 159, 160
 Kühne, Simon, 183
 Kunz, Cornelia Ursula, 107, 108, 153
 Kunzmann, Peter, 153
 Kurth, Tobias, 180
 Kurz, Malte, 38
 Kuss, Oliver, 171

Laabs, Björn-Hergen, 108
 Laffers, Lukas, 51
 Lakatos, Mária, 109
 Lang, Michel, 42, 110
 Lang, Moritz N., 110
 Langen, Henrika, 35
 Langrock, Roland, 120, 121
 Lau, Michael, 111

- Lee, Kuang-Yao, 158
 Less, Vivien, 112
 Leucht, Anne, 112
 Li, Bing, 158
 Liang, Xiaoran, 36
 Liebl, Dominik, 113
 Lilienthal, Jona, 55, 113, 142
 Lindner, Holger A., 147
 Lindner, Moritz, 44
 Lorenz, Fabian, 135
 Lubich, Bradley, 114
 Ludwig, Ralf, 175
 Luo, Ye, 159
- Maiworm, Matthias, 114
 Maj, Carlo, 39, 100
 Maletz, Steffen, 115
 Malinovskaya, Anna, 115
 Mandl, Maximilian Michael, 116
 Maranzano, Paolo, 117
 Martignon, Laura, 69
 Marx, Brian D., 117
 Mavridis, Dimitris, 70, 154
 Maxand, Simone, 118
 Mayer, Alexander, 52, 118
 Mayr, Andreas, 39, 78, 81, 100, 159, 164
 Mboya, Mwasi Paza, 179
 McConville, Kelly, 119
 McNamara, Amelia, 60
 Meier, Jonas, 119
 Meißner, Tobias Waldemar, 55
 Mellace, Giovanni, 120
 Mews, Sina, 120
 Meyer, Marco, 98
 Michels, Rouven, 121
 Mies, Fabian, 121
 Mikolov, Tomáš, 122
 Miller, Frank, 49, 122
 Mittermeier, Magdalena, 175
 Moebus, Susanne, 66
 Möllenhoff, Kathrin, 47
 Möller, Annette, 123
 Müller, Christian L., 125
 Müller, Emmanuel, 39
 Müller, Henrik, 176
 Murphy, Susan A., 123
- Nießl, Christina, 116
 Nieto-Reyes, Alicia, 77
- Nikolakopoulos, Stavros, 154
 Nikolakopoulou, Adriani, 126
 Noack, Claudia, 124
 Nott, David J., 103
 Nowak, Wolfgang, 56
 Ntzoufras, Ioannis, 154
 Nugent, Rebecca, 124
- Ötting, Marius, 120
 Oladyshkin, Sergey, 56
 Olma, Tomasz, 124
 Ostner, Johannes, 125
 Otto, Philipp, 117
 Otto-Sobotka, Fabian, 125
 Ozga, Ann-Kathrin, 150
- Paetz, Friederike, 151
 Pakkanen, Mikko S., 51
 Papakonstantinou, Theodoros, 126
 Papakonstantinou, Thodoris, 157
 Paparoditis, Efsthathios, 112
 Parsamaram, Parisa, 127
 Pauly, Markus, 41, 55, 65, 67, 97, 149
 Perme, Maja Pohar, 130
 Pfau, Maximilian, 44
 Pfeffermann, Danny, 127
 Pfeifer, Gregor, 139
 Pham, Vi Thanh, 128
 Pilz, Maximilian, 129
 Placzek, Marius, 129
 Podworny, Susanne, 73
 Pohle, Marc-Olive, 102
 Pohle, Marc-Oliver, 72
 Pohlmeier, Winfried, 97
 Putter, Hein, 57
- Rademacher, Daniel, 112
 Radermacher, Walter J., 130
 Rahmenführer, Jörg, 68, 96, 133, 140, 152
 Rappl, Anja, 134
 Rauch, Geraldine, 67
 Rauh, Johannes, 79
 Rebholz, Tobias R., 134
 Refisch, Lukas, 135
 Reimherr, Matthew, 113
 Reinecke, Jost, 120
 Rendtel, Ulrich, 138
 Renò, Roberto, 138
 Reutter, Mirjam, 139

Riebler, Andrea, 139
 Rieger, Jonas, 78, 140
 Riley, Richard D., 136
 Robinson, Mark, 140
 Rodemann, Julian Martin, 141
 Rodrigues, Paulo, 63
 Röger, Elias, 92
 Röttger, Frank, 141
 Röver, Christian, 113, 142
 Rohmann, Jessica L., 180
 Roquain, Etienne, 66
 Rothe, Christoph, 124
 Rubarth, Kerstin, 143
 Rucker, Gerta, 126
 Rügamer, David, 175
 Rühl, Jasmin, 143
 Rundel, Colin, 60

Säfken, Benjamin, 101
 Sand, Matthias, 144
 Sándor, Baran, 109
 Sapatinas, Theofanis, 112
 Sauerbrei, Willi, 145, 173
 Schaller, Jannik, 146
 Schaurer, Ines, 144
 Schefzik, Roman, 147
 Schenk, Alina, 147
 Schepers, Markus, 148
 Scheuer, Dominik, 85
 Schienle, Melanie, 53
 Schikowski, Tamara, 111
 Schlattmann, Peter, 181
 Schlosser, Anika, 48
 Schmid, Lena, 149
 Schmid, Matthias, 147, 160
 Schmid, Matthias C., 44
 Schmid, Timo, 106, 138
 Schmidt, Patrick, 149
 Schmidt, Rene, 61
 Schmitt, Sabrina, 150
 Schmitz-Valckenberg, Steffen, 44
 Schneider-Lindner, Verena, 147
 Schollmeyer, Georg, 50, 151
 Schorning, Kirsten, 152
 Schürmann, Christoph, 114
 Schürmeyer, Leonie, 152
 Schüssler, Julian, 153
 Schüürhuis, Stephen, 153
 Schultz, Carsten, 151
 Schulz, Anke, 42

Schumacher, Jakob, 92
 Schumacher, Martin, 48
 Schwabe, Rainer, 74, 127
 Schwandt, Anke, 171
 Schwarzer, Guido, 126
 Schweitzer, Michael, 66
 Schwender, Holger, 111
 Seidl, Thomas, 168
 Seifert, Caroline, 105
 Seiler, Stephan, 154
 Seitidis, Georgios, 154
 Seitz, Sarem, 84
 Selb, Peter, 153, 155
 Selk, Leonie, 156
 Sibbertsen, Philipp, 112, 179
 Sigle, August, 47
 Silber, Henning, 156
 Silbersdorff, Alexander, 101
 Skipk, Guido, 42
 Skipka, Guido, 114
 Skolkova, Alena, 157
 Smith, Michael Stanley, 99
 Soelvsten, Mikkel Mikkel, 35
 Sofack, Ghislain, 157
 Solea, Eftychia, 158
 Sonabend, Raphael, 110
 Speller, Jan, 159
 Spindler, Martin, 38, 71, 159, 160
 Spuck, Nikolai, 160
 Staerk, Christian, 100, 159, 164
 Stark, Maria, 161
 Stauffer, Reto, 110
 Steffen, Maximilian, 162
 Stehrenberg, Shari, 162
 Steib, Nicole, 163
 Steland, Ansgar, 121
 Stöcker, Almond, 163
 Strauch, Konstantin, 148
 Strömer, Annika, 164
 Strohmaier, Kristina, 139
 Sturtz, Sibylle, 113, 142
 Syliqi, Diella, 53
 Szabó, Marianna, 164, 165
 Szokol, Patricia, 165

Tchanyou Ganme, Sophie, 166
 Terzer, Tobias, 61, 68
 Thiel, Manfred, 147
 Trabs, Mathias, 162
 Treppner, Martin, 166

Tsirpitz, Renata Eirini, 122
 Tuerlinckx, Francis, 167
 Tulka, Sabrina, 167

Ullmann, Theresa, 116, 168
 Ullrich, Alexander, 53
 Umlauf, Nikolaus, 172
 Unkel, Steffen, 169

van Aert, Robbie C. M., 93
 van de Kassteele, Jan, 53
 van der Wurp, Hendrik, 170
 Van Mechelen, Iven, 170
 Veliyev, Bezirgen, 51
 Voeltz, Dina, 171
 Volkmann, Alexander, 172
 Vonthein, Reinhard, 172
 Vuk, Kata, 177

Wandel, Jasmin, 178
 Wang, Shuo, 173
 Wang, Zigan, 159
 Wason, James, 174
 Watson, David S., 49
 Wechsung, Maximilian, 174
 Wegner, Lea, 175
 Weigelt, Sarah, 55
 Weigert, Maximilian, 40, 53, 175
 Weinert, Henrike, 176
 Weiß, Christian, 70, 94
 Weißbach, Rafael, 177
 Wendler, Martin, 175, 177
 Westenberger, Ana, 108
 Westphal, Max, 92, 178
 Weyandt, Kai, 144
 White, Ian R., 70
 Wiborg, Øyvind Nicolay, 82
 Wied, Dominik, 52, 108, 118, 177
 Wienke, Andreas, 169
 Wiesenfarth, Manuel, 178
 Wigmann, Claudia, 111
 Wingert, Simon, 179
 Wisskott, Ruben, 89
 Witte, Janine, 64
 Wolf, Christof, 37
 Wolfram, Daniel, 53
 Wong, Anabelle, 180
 WongPiccininni, Marco, 180
 Woodall, William H., 180
 Wright, Marvin N., 49, 103

Wünstel, Justyna, 181
 Würz, Nora, 106

Yao, Weixin, 114
 Yaqine, Houda, 120

Zapf, Antonia, 105, 161
 Zarubin, Ilya, 181
 Zeileis, Achim, 110
 Zhang, Boyao, 182
 Zierk, Jakob, 84
 Zieschang, Tania, 125
 Zimmer, Felix, 183
 Zindel, Zaza, 183
 Zöllner, Daniela, 126, 157
 Zwick, Markus, 87
 Zwiernik, Piotr, 141



www.dagstat2022.de

www.dagstat.de

European Network for Business and Industrial Statistics – Deutsche Sektion •
Statistisches Bundesamt • AG Statistische Methoden in der Epidemiologie der
Deutschen Gesellschaft für Epidemiologie e.V. • Deutsche Statistische Gesellschaft •
Sektion Methoden der Politikwissenschaft der Deutschen Vereinigung für
Politikwissenschaft • DMV-Fachgruppe Stochastik e.V. der Deutschen Mathematiker-
Vereinigung • Gesellschaft für Klassifikation e.V. • Deutsche Gesellschaft für
Medizinische Informatik, Biometrie und Epidemiologie e.V. • Deutsche Region der
Internationalen Biometrischen Gesellschaft • Fachgruppe Methoden und Evaluation
der Deutschen Gesellschaft für Psychologie • Sektion Methoden der Empirischen
Sozialforschung der Deutschen Gesellschaft für Soziologie • Verband Deutscher
Städtestatistiker • Verein zur Förderung des schulischen Stochastikunterrichts e.V. •
Ausschuss Ökonometrie im Verein für Sozialpolitik e.V.