

Statistical Society of Slovenia

International Conference

22nd

Applied Statistics

2026

September 21-23, 2026

Koper, Slovenia

Book of Abstracts

International Conference
APPLIED STATISTICS 2026
Program and Abstracts

September 21–23, 2026
Koper, Slovenia

<i>Edited by</i>	Andrej Kastrin and Lara Lusa
<i>Published by</i>	Statistical Society of Slovenia Litostrojska cesta 54 1000 Ljubljana, Slovenia
<i>Publication year</i>	2026
<i>Publication format</i>	PDF
<i>Electronic version:</i>	https://as.mf.uni-lj.si/as2026book.pdf

PROGRAM and ABSTRACTS

PROGRAM

Day	Time	Room 1	Room 2	
Monday	8.30–15.00	Registration		
	9.00–9.30	Opening		
	9.30–10.20	Invited Lecture		
	10.20–10.50	Break		
	10.50–11.50	Complex Networks	Mathematical Statistics	
	11.50–12.00	Break		
	12.00–13.00	Data Science	Artificial Intelligence 1	
	13.00–14.30	Lunch		
	14.30–15.50	Social Sciences and Humanities 1	Measurement and Modeling	
	15.50–16.10	Break		
	16.10–17.30	Social Sciences and Humanities 2	Statistical Literacy	
	18.00–19.30	Sightseeing		
	19.45–21.00	Reception		
	Tuesday	8.30–12.00	Registration	
		9.00–9.50	Invited Lecture	
9.50–10.20		Break		
10.20–11.40		Biostatistics 1	Artificial Intelligence 2	
11.40–11.50		Break		
11.50–13.10		Biostatistics 2	Official Statistics	
13.10–14.40		Lunch		
14.40–16.00		Statistical Applications	Poster Session	
16.00–16.05		Closing		
Wednesday	9.00–12.00	Workshop		

8.30–15.00 **Registration**

9.00–9.30 **Opening**
Room 1

9.30–10.20 **Keynote**
Room 1

Chair: Anuška Ferligoj

1. **Parametric sparse models for distributional data**
Paula Brito

10.20–10.50 **Break**

10.50–12.20 **Data Science 1**
Room 1

Chair: Paula Brito

1. **Validation of Adolescents' EMA Profiles through Classification of Symbolic Distributional Data**
Jasminka Dobša, Simona Korenjak-Černe, Miranda Novak, Maja Buhin Pandur and Lucija Šutić Barbarić
2. **Oscillation detection: Ten data-driven approaches**
Ian Marsh
3. **Practical limitations of using the linear regression correction of Bland Altman analysis**
Maša Kušar
4. **FDI and Portfolio Investments in Ukraine: A Statistical Comparison of Dynamic Characteristics (1994–2026)**
Vadym Maslii

10.50–12.20 **Official Statistics**
Room 2

Chair: Irena Ograjenšek

1. **Nowcasting the Total Market Production Index for Slovenia**
Sašo Polanec, Tjaša Bartolj, Aleša Lotrič Dolinar and Mojca Bavdaž
2. **Noise Addition as a Prospective Solution to Over-suppression**
Žan Korošec
3. **Event History Analysis of Response Behaviour: A Comparative Study Across Respondent Groups in Three Independent Socio-Demographic Surveys in Italy**
Marija Mamolo, Loredana Di Consiglio and Raffaella Succi
4. **Estimation of Monthly Unemployment Rate in Slovenia**
Zarja Fabjan
5. **A Modelling Approach to Estimating Average Daily Tourist Spending in Slovenia**
Primož Bogner

12.20–12.30 **Break**

12.30–13.30 **Mathematical Statistics**
Room 1

Chair: Mihael Perman

1. **How contagious is happiness, Bayesian spatial modeling of life satisfaction with individual covariates**
Marek Brabec, Dana Hamplova, Marek Maly and Patricia Martinkova
2. **Stability and Precision of Entropy-Based Aggregation Methods in Ill-Conditioned Regression: A Simulation Study**
Ana Helena Tavares, Pedro Macedo and Maria Costa
3. **A common-shape Bayesian nonparametric model for multivariate mean and covariance change points**
Gordon Ross
4. **The cdf of BCF: Posterior contraction rate and Bernstein-von Mises theorem for Bayesian causal forests**
Andrej Srakar
5. **Change-Point Detection in Innovation Distributions of Circular Regression Models**
Bojana Milošević and Dunja Prokić

12.30–13.30 **Social Sciences and Humanities 1**
Room 2

Chair: Gregor Sočan

1. **Social Support Networks of Older Adults Before, During, and After the COVID-19 Pandemic**
Marjan Cugmas and Anuška Ferligoj
2. **The Role of Firms in Wage Inequality: Evidence from Linked Employer-Employee Data for Croatia and Slovenia**
Tjaša Bartolj, Barbara Kalar, Nataša Kump, Ria Ivandić, Velibor Mačkić and Filip Novokmet
3. **Using a Delphi study to assess consensus and content validity in a cybersecurity competence framework for older adults**
Anže Mihelič, Kaja Prislan Mihelič and Ajda Šulc
4. **Psychosocial Risks in the Workplace: A Thematic Analysis of Instagram Posts**
Vanja Erčulj and Monika Martinc
5. **Assessing the Quality of Mixed Methods Research in Intervention Studies: Model Development, Empirical Application, and Integration Patterns**
Gizela Kopač and Damjan Černe

13.30–14.30 **Lunch**

14.30–16.30 **Complex Networks**
Room 1

Chair: Marjan Cugmas

1. **Biblium in Orange: Visual-Programming Bibliometric Analysis**
Lan Umek

2. **Generalized Newman's normalization**

Vladimir Batagelj

3. **Controlling or normalizing? Blockmodeling valued networks with unequal relational capacities**

Aleš Žiberna and Fabio Ashtar Telarico

4. **The Recipe Similarity Network: A Statistical Framework for Extracting Structural Information from Culinary Databases**

Michele Bellingeri, Axel Bidon-Chanal Badia, Marta Vila Rigat, Roberto Alfieri, Massimiliano Turchetto and Davide Cassi

5. **Stochastic Modelling of Scientific Collaboration and Knowledge Production**

Luka Kronegger

14.30–16.00 **Data Science 2**

Room 2

Chair: Ana Zalokar

1. **Multi-Criteria and Sequential Strategies for Data-Driven Design of Experiments**

Matteo Borrotti

2. **Registry Data to Analysis-Ready Data Through Structured Initial Data Analysis**

Maja Jurtela, Lara Lusa, Tina Žagar, Nika Bric, Mojca Birk and Vesna Zadnik

3. **Machine Learning and Regression-Based Prediction of Low Birth Weight Using Maternal Nutritional Factors, Biomarkers, and Obstetric Factors: A Multicenter Hospital-Based Study in Addis Ababa, Ethiopia**

Jiregna Olani Kedida, Thierry Bossy, Corrado Lanera and Dario Gregori

4. **Performance Comparison for Privacy-Preserving Federated Learning Frameworks**

Shinto Pulickal Thomas, Thierry Bossy, Dario Gregori and Corrado Lanera

5. **Beyond Calibration and Resolution: A New Decomposition of Posterior Cross-Entropy for Evaluating Likelihood Ratio Systems**

Giulia Cereda, Fabio Corradi and Cecilia Viscardi

16.00–16.45 **Break**

16.45–18.00 **Social Sciences and Humanities 2**

Room 1

Chair: Luka Kronegger

1. **From Online Support to Bystander Intervention: A Mixed-Methods Perspective on Intimate Partner Violence in Slovenia**

Vanja Erčulj

2. **Monte Carlo Evaluation of Methods for Estimating Regression Coefficients Involving Latent Variables**

Katja Škoda

16.45–18.00 **Measurement and Modeling**

Room 2

Chair: Aleš Žiberna

1. Resolution, not frequency: a discrimination index of self-assessment and achievement in computer-based testing

Danilo Strugarevic, Nikola Gligorijevic, Vladimir Cabric, Rade Rakic and Dejan Andjelkovic

2. Measurement invariance, predictive invariance and comparability of individual measurements

Gregor Sočan

3. Detecting Asymmetric Persistence in Stationary Time Series: Reconstruction of a Modified Langevin Model from Natural Data

Zbigniew Czechowski

4. One Population, Twenty Studies: A Simulation-Based Framework for Exploring Sources of Heterogeneity in Biomedical Research

Antonina Spalińska, Wojciech Glinkowski and Bartłomiej Michalak

19.00–19.45 **Piano Concert**

20.00–21.30 **Reception**

8.30–12.00 **Registration**

9.00–9.50 **Keynote**
Room 1

Chair: Rok Blagus

1. **The Firth correction – a recap and new developments**
Georg Heinze

9.50–10.20 **Break**

10.20–11.50 **Biostatistics 1**
Room 1

Chair: Georg Heinze

1. **AUGlmmTMB: an R package for non-boundary estimation in binomial mixed effects models**
Rok Blagus and Tina Košuta
2. **Net Survival Analysis in Thymic Malignancies: Preliminary Results from a SEER Cohort**
Aqsa Ali, Gloria Brigiari, Ileana Baldi, Giulia Lorenzoni, Dario Gregori and Aslihan Şentürk Acar
3. **Challenges in Analysis of Longitudinal Observational Study: A Case of Osseointegration of Dental Implants**
Nataša Kejžar, Maša Kušar and Sebastjan Perko
4. **From Bioinformatics Quality Control to Statistical Modelling: A Structured Initial Data Analysis Checklist for Human High-Dimensional Genotype Data**
Cosimo Grazzini, Stefania Morelli, Elena Pilli, Giulia Cereda and Lara Lusa

10.20–11.50 **Artificial Intelligence 1**
Room 2

Chair: Irena Ograjenšek

1. **Beyond Traditional Methods: Exploring Machine Learning for Imputation in Official Statistics**
Kaja Malešič and Matej Divjak
2. **Mathematical Modelling in Modern Science and Control: Impact on the Continuity of the Philosophy of Science, Major Disturbances in Engineering, and the Mystification and Masking of Knowledge**
Robert Baždarić
3. **Correlation-Aware Bayesian Aggregation of Large Language Models with Identifiability and Calibrated, Distribution-Free Uncertainty**
Irem Tanriverdi and Ceylan Yozgatligil
4. **On the Performance of Language Models in Structural Time-Series Classification**
Yağmur Türkmen, Kerem Recep Gür, Pınar Cemre Yazıcı, Pelin Erkaya, Berke Baytak, Ismail Guzel, Pınar Karagoz and Ceylan Yozgatligil

11.50–12.00 **Break**

12.00–13.30 Biostatistics 2
Room 1

Chair: Nataša Kejžar

1. **Robustness and Reproducibility of Prognostic Factors in Lung Cancer: A Multi-verse Analysis Based on SEER Data**
Jiregna Olani Kedida, Stefania Lando, Gloria Brigiani, Mariavittoria Vescovo and Dario Gregori
2. **Safety assessment of CO₂ insufflation during PEG insertion: a non-inferiority study with a twist**
Nuša Cesar, Matjaž Homan and Bor Vratinar
3. **Lead time estimation in cancer screening programmes: extension of the MOCCI method**
Ema Požek and Bor Vratinar
4. **A Data-Augmentation Framework for Lq-Penalized Linear Mixed-Effects Models**
Rok Blagus and Lan Gerdej

12.00–13.30 Artificial Intelligence 2
Room 2

Chair: Luka Kronegger

1. **Artificial intelligence in survey research and beyond**
Irena Ograjenšek and Matej Pičulin
2. **Distributional Feature Importance for Correlated Predictors**
Burcu Koca Guler and Ceylan Yozgatligil
3. **A comparison of mixed-effects models and machine learning methods for longitudinal prediction**
Mariša Ratajec and Rok Blagus
4. **Perceptual Aspects of Uncertainty Visualization**
Bernarda Petek, David Nabergoj and Erik Štrumbelj

13.30–14.40 Lunch

14.40–16.10 Statistical Applications 1
Room 1

Chair: Maja Pohar Perme

1. **The impact of parallelism assumption in risk prediction models for ordinal outcomes**
Hana Šinkovec
2. **A Comparison of Causal Inference Methods for Survival Analysis Using a Simulation Study**
Petra Brlogar and Damjan Manevski
3. **Would the Lady Taste the Ratios? Concentration-Based and Compositional Analysis of Tea HPLC Data**
Patrik Prša and Damijana Kastelec
4. **Working with real-life data: Challenges and opportunities**
Dragan Tristan, Patricija Jaklič, Manca Prošek and Irena Ograjenšek

5. Joint Distribution of Correlated Exponential Random Variables via a Sum-of-Products Representation

Fatemeh Khaledian, Reza Asvadi, Tad Matsumoto and Mohd Azri Bin Mohd Izhar

14.40–16.10 Statistical Applications 2

Room 2

Chair: Nataša Kejžar

1. DBGenz: Efficient Integration of Multivariate Normal and t Probabilities

Jakob Peterlin and Nataša Kejžar

2. How to divide by zero? The many existing solutions in logistic regression and a proposed new one

Gaj Vidmar and Jakob Peterlin

3. Zero-Shot Classification of Water Quality Conditions from Multi-Resolution Satellite Imagery Using Vision-Language Foundation Models for Environmental Monitoring

Muhammad Zaid Qamar, Mohammed Ajaoud, Cristiano Ciccarelli and Massimiliano Lega

4. Comparative analysis of co-occurrence-based and embedding-based word networks in biomedical literature

Tom Rupnik Medjedovič and Andrej Kastrin

16.10–16.15 Closing

Room 1

9.00–12.00 **Workshop**
Room 1

Chair: Lara Lusa

1. **The weight of evidence: Likelihood ratios in forensic statistics**
Giulia Cereda

ABSTRACTS

Keynote

Parametric sparse models for distributional data

Paula Brito

University of Porto, Porto, Portugal

mpbrito@fep.up.pt

In classical multivariate statistics and machine learning, data are typically organized in a tabular format, where each row corresponds to an individual unit and each column records a single value for a given variable. However, this representation becomes inadequate when the data inherently involve variability. This situation arises when the units of analysis are not individual entities but rather abstract concepts—such as diseases instead of specific patients—or groups formed on the basis of shared characteristics. In such cases, for each descriptive variable, the variability observed within each concept or group should be taken into account, rather than relying solely on central tendencies (e.g., means, medians, or modes), in order to preserve potentially relevant information. Symbolic Data Analysis offers a framework for representing and analyzing such complex data, enabling aggregation at various levels of detail while retaining the associated variability. New types of variables have been introduced, where observations take the form of sets, intervals, or distributions over a given domain. In this work, we focus on numerical data described by empirical distributions. We propose parametric models based on representing each distribution using a central statistic and the logarithms of inter-quantile ranges for a selected set of quantiles. Multivariate normal distributions are assumed for the full set of indicators, considering alternative sparse structures for the covariance matrix. Interval-valued data is a particular case within this framework. The proposed model enables multivariate parametric analysis of distributional data, including analysis of variance, discriminant analysis, and model-based clustering. Applications to real-world data illustrate the relevance and usefulness of the approach.

Data Science 1

Validation of Adolescents' EMA Profiles through Classification of Symbolic Distributional Data

Jasminka Dobša, Simona Korenjak-Černe, Miranda Novak, Maja
Buhin Pandur and Lucija Šutić Barbarić

University of Ljubljana, School of Economics and Business, Ljubljana, Slovenia

simona.cerne@ef.uni-lj.si

The aim of this study is to propose an adaptation of the k -nearest neighbors (k -NN) algorithm for the classification of symbolic distributional data (Verde & Irpino, 2022) based on the L_2 Wasserstein distance. Symbolic Data Analysis (SDA) is a statistical framework designed to analyze aggregated data while preserving their internal variability. The proposed classification algorithm is applied to adolescents' profiles obtained from Ecological Momentary Assessment (EMA) data. EMA consisted of short questionnaires administered seven times per day over one week, resulting in up to 49 observations per participant for each variable. The study was conducted on a sample of 132 secondary school students in Croatia. Profiles are first identified by clustering symbolic distributional data, where each respondent is represented as a symbolic object summarizing the distribution of their repeated EMA responses. Composite variables are derived based on factor analysis reflecting confidence, attentiveness, positive and negative emotions related to friends, and overall positive and negative affect. The resulting profiles are then interpreted using additional survey data. Representing respondents as symbolic objects preserves within-person variability in self-evaluations across the observation period. The adapted k -NN algorithm is then used to assess the recognizability of the identified profiles. High classification accuracy would indicate that the profiles correspond to distinct and stable patterns that can be reliably recognized by machine learning methods, providing additional evidence for their validity. Verde, R., Irpino, A. (2022). Partitional and Hierarchical Clustering of Distributional Data Using the Wasserstein Distance. In Brito, P. & Dias, S. (Eds.), *Analysis of Distributional Data* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370545>

Oscillation detection: Ten data-driven approaches

Ian Marsh

RISE, The Research Organisations of Sweden, AB, Kista, Sweden

ian.marsh@ri.se

This paper addresses industrial *stability*. Machines can oscillate in often unpredictable, subtle manners, that may have consequences, some serious. In manufacturing, oscillations can have repercussions on a part being machined, the machine performing the task, or both, incurring time and financial costs. Machines that operate in a circular motion can also oscillate, slightly off center rotation, or dirt/grim in the bearings can produce oscillations. Another example is knocking or nudging a piece of equipment from a fixed operating point, which oscillates until reaching a stable position again. The oscillations can be periodic or aperiodic. Industrial signal measurement and processing are increasingly being deployed in control performance assessment, particularly in oscillation monitoring applications. Therefore, the goal of this paper is to find a reliable method for the detection of oscillations in industrial data. We had access of 27 labelled univariate time series, derived from an industrial process. 7 series contained oscillations, 20 did not. We present simple to complex models of oscillation detection. We use a number of different techniques, and quantify which of the 27 possible correct identifications were correct. We also explain how generate additional sequences with similar statistical properties, for further training where needed. We also compare and contrast our work too others in this field, especially where the same data was used. Finally, we conclude, relatively simple methods can correctly identify time series with oscillations.

Practical limitations of using the linear regression correction of Bland Altman analysis

Maša Kušar

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

masa.kusar@mf.uni-lj.si

Background: Regression-based Bland-Altman limits of agreement (LoA) are recommended when the bias between methods is associated with the measurement magnitude. The theoretical correction uses the pairwise average of the two methods, but in practice only the measurements with one method are available. **Objective:** To quantify the bias introduced when regression-based LoA are implemented using only the new method value. **Methods:** We performed a simulation with $B = 10000$ replicates per condition and sample size $n = 1000$. We estimated different levels of true-value variability, the random error of the new method, and the sizes of the proportional error of both methods. For each condition, we estimated (i) theoretical coverage using LoA evaluated at the pairwise average and (ii) practical coverage using LoA evaluated at the new method only. **Results:** Using only the measured value with the new method can lead to a significant reduction in the empirical coverage of the nominal 95% limits of agreement. However, this issue is mostly present in situations that would make one of the measurement methods clearly inferior, thereby making interchangeable use impossible (and a direct comparison moot). On the other hand, in most scenarios where interchangeable use would be realistic, the change in empirical coverage is minimal and can likely be safely ignored. **Conclusion:** Substituting the value of the new method for the pairwise average can materially change the LoA coverage in specific circumstances. When limited to practically relevant scenarios, this reduction in coverage is, by and large, negligible.

FDI and Portfolio Investments in Ukraine: A Statistical Comparison of Dynamic Characteristics (1994–2026)

Vadym Maslii

West Ukrainian National University, Department of Applied Mathematics, Ternopil, Ukraine

vadummasly1974@gmail.com

Foreign direct investment (FDI) and portfolio investment represent two major forms of international capital inflows. Despite their importance for recipient economies, the literature provides conflicting evidence regarding their dynamic characteristics. While FDI is often considered more resilient to economic disturbances, empirical findings on the differences between FDI and portfolio investment remain inconclusive. This study aims to statistically compare FDI and portfolio investments received by Ukraine over the entire period of its independence. The central hypothesis is that these forms of international capital exhibit distinct dynamic characteristics reflected in their volatility, trend behaviour, and responses to systemic shocks. The empirical analysis is based on quarterly balance-of-payments data covering the period from the first quarter of 1994 to the first quarter of 2026. The methodological framework combines descriptive statistics, measures of variation, variance comparison procedures, stationarity testing, and structural break analysis. The focus is placed on identifying differences in the statistical properties of the series rather than comparing the absolute volumes of capital inflows. Ukraine is examined as a long-term case of a capital-recipient economy for which comparable statistical data on both FDI and portfolio investments are available over more than three decades of economic, political, and geopolitical transformations. Preliminary expectations suggest that portfolio investments may demonstrate greater short-term variability and stronger sensitivity to systemic shocks than FDI. The findings are expected to contribute to a better understanding of the statistical behaviour of different forms of international capital and their role in economies exposed to prolonged periods of structural transformation and instability.

Official Statistics

Nowcasting the Total Market Production Index for Slovenia

Sašo Polanec, Tjaša Bartolj, Aleša Lotrič Dolinar and
Mojca Bavdaž

University of Ljubljana, School of Economics and Business, Ljubljana, Slovenia

mojca.bavdaz@ef.uni-lj.si

The timely availability of accurate short-term economic statistics is essential for effective policy- and decision-making. While some partial indicators of economic activity are available with a delay of up to a month, two key indicators, produced by the Statistical Office of the Republic of Slovenia (SURS) each have a shortcoming: GDP growth rates are only published quarterly, and the monthly Total Market Production Index (TMPI) is currently released with a two-month delay. The research funded by SURS and ARIS (V5-2346) aimed at developing a methodological framework for nowcasting short-term developments in Slovenia's market economy. The empirical analysis evaluates a range of nowcasting models for the aggregate TMPI and its four sectoral components: industry, construction, trade and services. We accounted for the need to work with data of mixed frequencies and with time series starting and ending at different points in time. The modelling strategy combines publicly available macroeconomic, sectoral and financial indicators for Slovenia and countries related to Slovenia, with aggregated microdata available to SURS, while also assessing the practical limitations of alternative big-data sources for regular statistical production. The results suggest that a disaggregated nowcasting framework, supported by carefully selected auxiliary indicators and regularly monitored model performance, can provide a robust basis for earlier assessment of short-term economic developments. The proposed approach offers practical guidance for the gradual integration of early estimates into official statistics, while highlighting the importance of transparent communication of model uncertainty in real-time economic monitoring.

Noise Addition as a Prospective Solution to Over-suppression

Žan Korošec

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

zan.korosec@gov.si

Disclosure control represents a considerable challenge for National Statistical Institutes (NSIs), particularly in countries that are relatively small and therefore tend to have a high concentration of economic activity or industry (i.e., spatial agglomeration and market saturation). To exacerbate the problem, research design often incorporates tables with multiple explanatory variables featuring intricate hierarchical classifications and detailed breakdowns, whereas it should also be noted that sets of such tables may be interlinked, necessitating harmonised protection. Conventional suppression-based methods are habitually employed in such instances; however, the combination of already sparse cell counts and cascading secondary suppression can render large portions of published tables inaccessible to users, thereby severely compromising data utility. To surmount this impediment the methodology section at the Statistical Office of the Republic of Slovenia began exploring perturbative methods, specifically the noise addition approach proposed by Evans, Zayatz and Slanta (1998) in which random noise drawn from a selected distribution is added to microdata prior to tabulation. As the attribution is localised at the record level, additivity is inherently preserved, while a permanent noise factor also ensures equal protection across all generated tables. To further enhance the procedure, we incorporated the balancing mechanism recommended by Massell and Funk (2007), which optimises the amount of noise assigned to otherwise non-sensitive cells by alternating the direction of the noise, so that the perturbed cumulative sum is closer to the original value. Admittedly, the trial-and-error phase with calibration cannot be averted when determining the acceptable noise magnitude nor can balancing be made universal as it requires a representative sub-table that is specific enough to differentiate between record classes but still enables transfer to other tables. Nevertheless, our findings suggest that noise addition offers a viable path to disclosure control, retaining utility and not compromising confidentiality.

Event History Analysis of Response Behaviour: A Comparative Study Across Respondent Groups in Three Independent Socio-Demographic Surveys in Italy

Marija Mamolo, Loredana Di Consiglio and Raffaella Succi

Istat, Trieste, Italy

mamolo@istat.it

Non-response is a major source of bias in sample surveys, affecting the quality and representativeness of statistical estimates. Understanding the process of response is essential for the adoption of strategies that can enhance the participation in the survey avoiding sample selectivity. This includes not only the propensity to respond, but also its timing. Only a few studies have overcome the cross-sectional perspective and investigated the response patterns by analysing the time-to-response during the fieldwork (Becker, 2024). We apply techniques of event history analysis to study the response process in three Italian socio-demographic sample surveys carried out by Istat in 2024: Aspects of Daily Life, Citizens and Leisure Time, and Families and Social Subjects. The surveys use a sequential mixed-mode design with CAWI (Computer-Assisted Web Interview) and CAPI (Computer-Assisted Personal Interview). We focus on three groups of respondents: households with all members aged 70 and more, households with all members of foreign citizenship and other households. The Kaplan-Meier non-parametric model shows a different timing of response between the three groups. During the CAWI phase households with only foreign members are less responsive, while during the CAPI phase the response is triggered in all the groups. In the piecewise constant exponential model, the time intervals represent the phases of the sequential mixed-mode design. The model with household and contextual variables and their time-interactions reveals the importance of the dynamic perspective: significant differences in the response behaviour of the three groups and in the effect of the covariates are shown across the fieldwork phases. Time plays an important role in the analysis of the process of response. The results call for the adoption of differentiated fieldwork strategies to hamper the non-response, not only with respect to the household and contextual profile of the sample units but also with regard to the fieldwork phases.

Estimation of Monthly Unemployment Rate in Slovenia

Zarja Fabjan

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

zarja.fabjan@gmail.com

Monthly unemployment rate (MUR) is one of the key European economic indicators. It is used for the formulation and monitoring of the European Union's policies for growth and job creation. In Slovenia, the Statistical Office (SURS) produces MUR in accordance with Commission Implementing Regulation (EU) 2019/2241. Since data on survey-based unemployment are collected only quarterly, monthly estimates must be derived to meet the requirements. Since March 2021, SURS has applied the multiplicative Denton benchmarking method, implemented in JDemetra+ Benchmarking plug-in, to estimate MUR. For additional information, monthly data on registered unemployed persons and monthly data on persons in employment from the statistical survey Persons in Employment have been used as indicator series. However, a quality assessment published by Eurostat in 2025 indicate that, while Slovenia meets recommended thresholds for volatility, it significantly exceeds acceptable limits for revisions. This has motivated the exploration of alternative estimation options. Several methods are evaluated using the data from May 2022 to January 2026. Some of them are: (1) the current Denton method, (2) a hybrid approach combining forecasting up to the end of the quarter combined with the Denton method, (3) the Cholette method, and (4) regression-based temporal disaggregation techniques. The choice of methods is additionally informed by ESS guidelines on temporal disaggregation, benchmarking and reconciliation published in 2018. The comparison is based on seasonally adjusted total unemployment rates, using four quality indicators recommended by Eurostat; two evaluating volatility and two revisions. The results show that all tested methods perform well with respect to volatility measures, but none meet the recommended standards for revision indicators. Among the approaches, the hybrid method that combines forecasting with the Denton method shows the most favorable results in terms of reducing revisions, suggesting it as a promising direction for further methodological work.

A Modelling Approach to Estimating Average Daily Tourist Spending in Slovenia

Primož Bogner

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

pb07213@student.uni-lj.si

Our aim is to present the statistical approach for estimating annual average daily spending of foreign tourists in Slovenia, focusing on months not directly observed in the TU-TURISTI survey. Using data from the 2021-2022 survey cycle, the study aims to improve existing estimation methods used by the Statistical Office of the Republic of Slovenia. The main challenge is that survey data are collected for only selected months within each season, requiring estimation for the remaining months. The approach assumes that tourist behaviour is relatively consistent within a season, allowing observed months to be used for predicting missing ones. Data are obtained from survey responses and administrative registers, then merged, filtered, and weighted to approximate the full population. Data preparation includes handling repeated individuals across months and constructing variables such as country groups, municipality types, and hotel categories. Missing groups, particularly children, are incorporated by replicating and adjusting adult responses, though this introduces simplifying assumptions. Tourist expenditure is calculated as the sum of accommodation, prepayments, daily spending and transportation cost, expressed per person per night. Due to strong skewness, transformations and robust modelling methods are applied. Several models are evaluated: a linear model, a linear mixed model, a generalized linear model, and XGBoost. Performance is assessed using 20-fold cross validation and metrics such as RMSE and MAE. While XGBoost shows strong overall accuracy, it performs less reliably for smaller subgroups. The generalized linear model with a gamma distribution and log link is selected as the most suitable. Final estimates for missing months are obtained through poststratification and validated using bootstrap sampling, producing generally accurate results.

Mathematical Statistics

How contagious is happiness, Bayesian spatial modeling of life satisfaction with individual covariates

Marek Brabec, Dana Hamplova, Marek Maly and Patricia Martinkova

The Czech Academy of Sciences, Institute of Computer Science, Praha, Czechia

mbrabec@cs.cas.cz

We will present a Bayesian modeling framework for modeling spatial areal data with individual-specific covariates. The data format is somewhat non-standard in that: (i) the spatial aggregation is at two levels (e.g., country and NUTS3 regions within country) and (ii) most of the covariates are collected individually with the aim to correct for subject idiosyncracies and hence areal unit aggregation is not appropriate. To this end, we develop a full-rank spatial model, starting with common autocorrelation-related parameter and subsequently expanding to allow for country-pair-specific parameters allowing to explore details of the similarities among neighboring regions in more detail. We illustrate the modeling on happiness and life satisfaction scores from the ESS study exploring the spatial contagion of these variables as a phenomenon interesting in broader sociological context. The work has been supported by the DigiWell project Z.02.01.01/00/22_008/0004583.

Stability and Precision of Entropy-Based Aggregation Methods in Ill-Conditioned Regression: A Simulation Study

Ana Helena Tavares, Pedro Macedo and Maria Costa

University of Aveiro, Aveiro, Portugal

ahtavares@ua.pt

Ill-conditioned regression models frequently lead to unstable estimation and unreliable inference, particularly in settings involving heterogeneous subgroups and strong collinearity. This presentation aims to evaluate the performance of entropy-based aggregation strategies as robust alternatives to classical estimation procedures under such conditions. The main focus is on comparing three aggregation schemes, namely, bagging, magging, and neagging, combined with ordinary least squares (OLS), generalized maximum entropy (GME), and weighted generalized maximum entropy (W-GME) estimators. A Monte Carlo simulation study comprising 48 scenarios is used to assess estimator stability and precision under varying degrees of collinearity, subgroup structure, parameter support, and error distributions. The results show that while predictive performance is broadly comparable across methods, substantial differences emerge in inferential stability. OLS-based aggregation deteriorates under severe ill-conditioning, whereas entropy-based estimators consistently provide more stable and precise inference. In particular, magging combined with W-GME yields the most robust performance across high-collinearity regimes. For conference participants, this presentation highlights how aggregation choices can critically affect inferential reliability even when predictive accuracy appears unchanged. The findings provide practical guidance for selecting robust estimation strategies in complex, high-dimensional regression settings.

A common-shape Bayesian nonparametric model for multivariate mean and covariance change points

Gordon Ross

University Of Edinburgh, School of Mathematics, Edinburgh, United Kingdom

gordon.ross@ed.ac.uk

Many changepoint methods sit at one of two extremes. Parametric models give clear changes in means and covariances, but can mistake skewness, heavy tails or multimodality for extra changepoints. Fully nonparametric Bayesian models avoid a fixed distributional form, but usually learn a separate distribution in each segment, wasting information when regimes differ mainly by mean and covariance. We propose a common-shape Bayesian nonparametric changepoint model for this intermediate case. Each segment has its own mean vector and covariance matrix, while the underlying distributional shape is shared across segments after standardisation. The shared shape is modelled using a centred Dirichlet process mixture. In Chinese-restaurant-process terms, all standardised observations share one restaurant, with segment parameters explaining how that common shape is shifted and stretched over time. In several dimensions, standardisation is part of the model: different whitening conventions can change the shape being shared. We therefore use symmetric square-root standardisation within segments and only centre-and-scale preprocessing globally. Simulations focus on short segments, variance and covariance changes, and comparisons with independent nonparametric, Gaussian and frequentist changepoint methods.

The cdf of BCF: Posterior contraction rate and Bernstein-von Mises theorem for Bayesian causal forests

Andrej Srakar

Jožef Stefan Institute, Ljubljana, Slovenia

andrej.srakar@ijs.si

Bayesian causal inference is an important area in causality research (Cinelli et al., 2025; Wang et al., 2025). Standard Bayesian prior for causal inference is Bayesian additive regression trees (BART, see Chipman et al., 2010). Theoretical properties of BART have been subject of important recent research (Ročková and van der Pas, 2017; Ročková, 2019; Jeong and Ročková, 2022). In their recent contribution on the topic Jeong and Ročková propose a new class of sparse piecewise heterogeneous anisotropic Hölder functions and derive their minimax optimality properties. We extend their work to Bayesian Causal Forests (BCF), proposed by Hahn, Murray and Carvalho (Hahn et al., 2020). BCF is a modular Bayesian approach, based on regression decomposition into prognostic and treatment effect part of the outcome regression specification. In our article we accomodate for the modular nature of BCF by extending Laplace approximation for cut posteriors used in previous work by Pompe and Jacob (2021) to the sparse piecewise heterogeneous anisotropic Hölder functions framework accounting for the mathematical nature of the BART prior. This leads to the asymptotics using piecewise polyhedral multifunctions (see e.g., Meng et al., 2025). Posterior contraction rate and an asymptotic Bernstein-von Mises theorem for the BCF approach are derived. It seems feasible to generalize our approach and findings also to other types of both nonparametric BART priors (e.g., generalized BART, spatial BART, RDD BART) as well as BCF (e.g., shrinkage BCF), for example using piecewise harmonic functions in Besov and Sobolev metric spaces. Usages of findings are demonstrated for several standard causal models: regression discontinuity design, maximal partially directed acyclic graph, and causal mediation. Our article delves into a still underexplored and highly important area of asymptotics for Bayesian nonparametric priors, and we discuss also extensions to other BNP approaches outside of the BART and causal inference framework.

Change-Point Detection in Innovation Distributions of Circular Regression Models

Bojana Milošević and Dunja Prokić

University of Belgrade, Faculty of Mathematics, Belgrade, Serbia

bojana.milosevic@matf.bg.ac.rs

We study change-point detection in the innovation distribution of circular regression models. The goal is to identify distributional changes in the regression residuals while the regression function remains unchanged. To this end, we construct residual-based change-point statistics using empirical circular characteristic functions. The proposed approach is capable of detecting a wide range of changes in the innovation distribution. Numerical experiments illustrate its finite-sample performance under several circular data-generating mechanisms. Potential applications include environmental, biological, and engineering studies involving directional measurements.

Social Sciences and Humanities 1

Social Support Networks of Older Adults Before, During, and After the COVID-19 Pandemic

Marjan Cugmas and Anuška Ferligoj

University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia

marjan.cugmas@fdv.uni-lj.si

Social support is strongly associated with physical and psychological health. For older adults, embeddedness in social support networks and local communities is especially important, as it helps them remain active, socially integrated, and supported in everyday life. Those without any social support are among the most vulnerable. In Slovenia, the share of older adults reporting no informal social support rose from around 1% in 2002 to 7.8% in 2020, coinciding with the first year of the COVID-19 pandemic. This presentation examines the prevalence and characteristics of social support network types among older adults in Slovenia between 2002 and 2025, with particular attention to post-pandemic recovery. Ego-centred network data from surveys of adults aged 65 and over, conducted in 2002, 2020, and 2025, were analysed using clustering for symbolic data and correspondence analysis. Network types were identified based on the number and characteristics of support providers, including relationship type, spatial proximity, and frequency of contact across emotional, instrumental, and socialising support. The identified network types closely correspond to Wenger's typology. Trends in their prevalence suggest that some network types were more strongly affected by the COVID-19 pandemic, while others reflect longer-term demographic and social changes. The findings indicate a partial recovery of informal social support after the pandemic: by 2025, the share of older adults with no social support had declined to 4.4%. Nevertheless, inequalities persist, with men showing a particularly higher likelihood of social isolation.

The Role of Firms in Wage Inequality: Evidence from Linked Employer-Employee Data for Croatia and Slovenia

Tjaša Bartolj, Barbara Kalar, Nataša Kump, Ria Ivandić, Velibor Mačkić and Filip Novokmet

Institute for Economic Research, Ljubljana, Slovenia

bartoljt@ier.si

Wage inequality has risen sharply since the 1980s, with particularly pronounced increases in former socialist economies. While the existing literature largely attributes this to skill-biased technological change or shifting labour market institutions, the contribution of firms has received little systematic attention — despite long-standing evidence that equally skilled workers can earn substantially different wages depending on where they work. We investigate the role of firms in shaping wage inequality in Croatia and Slovenia from the 1970s to the present. Two economies that transitioned from self-managing socialism to market systems in the early 1990s offer a unique historical context, allowing us to trace the firm contribution to inequality from the late socialist period through transition and into the present — and to examine how structural change reshapes wage-setting at the firm level. Our primary contribution is the construction of comprehensive linked employer-employee longitudinal datasets for both countries, drawing on individual-level social security records with universal coverage of workers and firms, supplemented by firm-level financial statements and individual tax records. Using these datasets, we produce consistent estimates of national wage inequality and decompose it into within-firm and between-firm components over the study period. We then estimate the Abowd, Kramarz, and Margolis (1999) model to separate the between-firm component into firm-specific wage premiums and worker sorting effects — assessing the extent to which rising inequality reflects firms paying differently for the same skills, a signature of employer wage-setting power, versus changes in worker composition across firms. This paper presents first results from this ongoing research program, with initial estimates for Slovenia and Croatia covering the post-transition period, and discusses key methodological challenges in constructing the datasets and estimating the AKM model in transition economy contexts.

Using a Delphi study to assess consensus and content validity in a cybersecurity competence framework for older adults

Anže Mihelič, Kaja Prislan Mihelič and Ajda Šulc

University of Maribor, Faculty of Criminal Justice and Security, Ljubljana, Slovenia
ajda.sulc@um.si

Cybersecurity competence frameworks are often used to guide education, prevention, and policy interventions, yet existing frameworks are rarely tailored to the needs, risks, and everyday digital practices of older adults. Our paper will demonstrate how a Delphi study can be used as a structured methodological approach for assessing expert consensus about relevance and clarity of a newly developed cybersecurity competence framework for older adults and how this translates to content validity assessment. The initial framework was developed on the basis of a literature review and an extensive analysis of existing competence frameworks. It was then submitted to a panel of experts and relevant institutions for evaluation through a Delphi study. In total, 133 institutions and individual experts were invited to participate, with around one-third of the invited responding positively. Experts were asked to assess the relevance of each proposed competence for older adults, as well as the clarity and comprehensibility of the competence descriptions. The presentation will focus on the methodological design of the Delphi study, including expert recruitment, item-level assessment, and criteria for evaluating consensus. We will also show how expert ratings and qualitative comments can be used to improve the competence descriptions and provide evidence for content validity—that the framework covers relevant and meaningful aspects of cybersecurity competence for older adults. With that, the presentation will present an applied example of using Delphi methodology in the development and validation of a competence framework.

Psychosocial Risks in the Workplace: A Thematic Analysis of Instagram Posts

Vanja Erčulj and Monika Martinc

University of Maribor, Faculty of Criminal Justice and Security, Ljubljana, Slovenia

monika.martinc@gmail.com

Psychosocial risks in the workplace encompass aspects of work design, organization, and social context that may adversely affect employees' mental health and well-being. These risks may manifest differently across standard and non-standard forms of employment. At the same time, social media platforms provide a growing source of user-generated data that can offer valuable insights into employees' work-related experiences. This study aimed to identify and compare psychosocial risks reported by employees in standard and non-standard employment using data collected from Instagram. A dataset of 1,254 publicly available posts published between 2019 and 2025 on a Slovenian Instagram profile dedicated to employment-related issues was analyzed. The study employed a qualitative phenomenological approach and thematic analysis following Braun and Clarke's framework. The analysis identified several recurring themes, with pay, interpersonal relationships, workload, and work organization emerging as the most frequently reported psychosocial risk factors. While psychosocial risks were present across both employment types, notable differences were observed. Posts related to non-standard employment more frequently emphasized job insecurity and economic uncertainty, whereas posts related to standard employment more often highlighted workload pressures, organizational issues, and time demands. The findings contribute to understanding how employees experience and communicate psychosocial risks and demonstrate the value of social media data as a complementary source for workplace research. The study also illustrates the potential of thematic analysis of large-scale social media content for identifying emerging workplace issues and informing occupational health policies.

Assessing the Quality of Mixed Methods Research in Intervention Studies: Model Development, Empirical Application, and Integration Patterns

Gizela Kopač and Damjan Černe

Independent researcher, Ljubljana, Slovenia

gizela.kopac@gmail.com

Mixed methods research is increasingly used in intervention studies, yet its methodological quality remains uneven. This contribution presents a conceptual-theoretical model for assessing the quality of mixed methods research in intervention studies. The model is structured into three dimensions—input, process, and output—and comprises 22 quality criteria operationalized through 70 binary indicators. It covers the conceptual foundations and design of mixed methods research, quantitative and qualitative components, integration across different stages of the research process, interpretation and inference, and intervention-specific methodological elements. The indicators were weighted according to their relative occurrence in the methodological literature and normalized to produce an overall quality score. The model was empirically applied to 439 mixed methods intervention studies published between 2007 and 2023. Overall weighted quality scores ranged from 0.20 to 0.81, with a median of approximately 0.47. The analysis also revealed considerable variation in the presence of integration indicators. Integration was identified in 11.6% of studies at the sampling stage and in 16.2% at the analysis stage, while mixed methods interpretation was identified in 43.7%. The integration process was described in 36.4% of studies, an explicit integration strategy in 13.9%, and meta-inferences in 10.2%. As an extension of the original analysis, selected integration indicators are examined jointly within the input–process–output structure of the model. Their configurations within individual studies are analysed to identify patterns of integration and potential discontinuities across the research process. The study thus extends indicator-level assessment by examining how integration elements occur together within individual studies, providing a more process-oriented understanding of the quality of mixed methods research in intervention studies.

Complex Networks

Biblum in Orange: Visual-programming bibliometric analysis (invited talk)

Lan Umek

University of Ljubljana, Faculty of Public Administration, Ljubljana, Slovenia

lan.umek@fu.uni-lj.si

Bibliometric software is today largely constrained by two limitations: programming skills (Python or R libraries) are typically required, or the software is provided as a standalone application that cannot easily be combined with other analytical tools. Biblum in Orange, an open-source add-on through which bibliometric and scientometric analysis is embedded into Orange—a widely used visual-programming data-mining suite—is presented. Built on the Biblum library, the add-on allows analytical pipelines to be assembled by drag-and-drop, so that data can be imported from Scopus, Web of Science, or OpenAlex, statistically analysed, and interactively visualized without any code being written. Its distinctive contribution is comparative group analysis, in which a corpus is partitioned into user-defined subgroups and associations between these groups and bibliometric entities (authors, keywords, sources, countries) are quantified through chi-square testing and correspondence analysis. Because it is embedded in Orange, hybrid workflows are enabled in which bibliometrics is combined with machine learning, text mining, and network analysis. Live workflows will be demonstrated on a sample corpus, and benefits for teaching and reproducibility will be discussed.

Generalized Newman's normalization

Vladimir Batagelj

Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia • University of Primorska, Andrej Marušič Institute, Koper, Slovenia • University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia

vladimir.batagelj@fmf.uni-lj.si

Newman's normalization of two-mode networks is generalized from binary networks to weighted networks.

Controlling or normalizing? Blockmodeling valued networks with unequal relational capacities

Aleš Žiberna and Fabio Ashtar Telaarico

University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia
ales.ziberna@fdv.uni-lj.si

We address the problem of blockmodeling valued networks in which units have different relational capacities, where the relational capacities should not influence the similarity of units, which should instead depend on relative preferences. A similar problem has been addressed in several works (Matveeva et al., 2023; Nordlund, 2016, 2020; Nordlund & Žiberna, 2020), where the usual solutions involve some form of normalization (or similar methods), with the exception of the correlation-based approach by Nordlund. Another approach is to use something equivalent to degree correction in valued networks (Peixoto, 2018). The presentation will focus on the results of simulations “exploring, under different data-generating mechanisms, which of these strategies provides better recovery of the original partition.

The Recipe Similarity Network: A Statistical Framework for Extracting Structural Information from Culinary Databases

Michele Bellingeri, Axel Bidon-Chanal Badia, Marta Vila Rigat,
Roberto Alfieri, Massimiliano Turchetto and Davide Cassi

University of Parma, Department of Mathematical, Physical and Computer Sciences, Parma, Italy

We present a network-based statistical framework for the quantitative analysis of recipe collections, integrating intersection graph theory with complex network science. The approach is applied to three Catalan cookbooks — two traditional and one haute cuisine — comprising 779 recipes, providing a standardised and expert-curated dataset that avoids inconsistencies typical of user-generated repositories. Each cookbook is encoded as a Recipe Similarity Network (RSN), an undirected weighted graph in which nodes represent recipes and edges connect recipes sharing at least one ingredient. Edge weights are defined by a novel normalised similarity measure accounting for ingredient set size, reducing the size bias inherent in raw intersection counts and improving upon metrics such as the Jaccard index. We investigate how ingredient substitution and edge weight normalisation affect network topology and node centrality. Both factors substantially alter the identification of structurally central recipes. In the traditional corpus, *Allioli* emerges as the most central recipe, while in the haute cuisine network the most central recipe is *Becada amb brioix del seu salmís*. Structural analysis reveals that all RSNs approach complete-graph topology at low similarity thresholds. As the threshold increases — equivalent to a bond percolation process — the networks fragment, with the haute cuisine network disintegrating faster, quantitatively confirming that haute cuisine recipes are more specialised and ingredient-heterogeneous. A community detection algorithm combining progressive link removal with maximal clique identification uncovers tightly ingredient-overlapping recipe clusters. A 7-clique identified in the traditional corpus provides a principled foundation for ingredient-based recipe recommendation and guided culinary innovation. The framework is transferable to any domain where entities are described by overlapping sets.

Stochastic Modelling of Scientific Collaboration and Knowledge Production

Luka Kronegger

University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia

luka.kronegger@fdv.uni-lj.si

This study was conducted as part of a larger research project examining the role of mentorship in scientific collaboration and knowledge production (https://www.cmi-fdv.si/projekti/arrs/?page_id=1605). We analyse the dynamics of research interests and mentoring relations among Slovenian researchers using bibliographic and administrative data. Research topics are identified from publication records using BERTopic, and changes in bipartite author–topic networks are modelled using stochastic actor-oriented models (SAOMs). The analysis is performed separately for six scientific fields and includes mentoring ties, scientific age, gender, discipline, and publication activity. The results indicate topic-based clustering beyond disciplinary affiliation, consistent with the concept of invisible colleges, as well as cumulative advantage in the adoption of research topics. Younger researchers are more likely to enter new research areas, while mentoring is associated with the transmission of research orientations, particularly in natural sciences and engineering.

Data Science 2

Multi-Criteria and Sequential Strategies for Data-Driven Design of Experiments

Matteo Borrotti

University of Milano-Bicocca, Milan, Italy

matteo.borrotti@unimib.it

Design of Experiments (DoE) has traditionally provided principled statistical tools for obtaining informative data with a limited number of experimental runs, quantifying uncertainty, and supporting decision-making. However, many contemporary applications in engineering, biomedicine, and the social sciences involve complex systems that are high-dimensional, costly to evaluate, noisy, multi-objective, and subject to operational constraints. These features motivate an evolution of classical DoE toward adaptive, data-driven, and computationally scalable strategies. This contribution discusses a unified statistical perspective linking optimal experimental design with modern machine learning and sequential learning. First, we focus on multi-criteria optimal design, where several design criteria must be optimized simultaneously and the solution is naturally represented by a Pareto set rather than by a single optimum. Building on multi-objective design algorithms for multi-stratum experiments, we discuss how decision rules such as the utopia-point criterion and TOPSIS technique can convert non-dominated solutions into actionable compromise designs. This is particularly relevant in industrial and applied settings, where statistical efficiency, prediction accuracy, and operational feasibility must often be balanced. Second, we consider sequential design in high-dimensional spaces. Bayesian optimization and active learning provide useful computational principles for adaptive experimentation, but standard Gaussian-process-based approaches may become computationally demanding or unstable in high-dimensional or mixed-variable settings. We therefore discuss scalable alternatives based on local-global surrogate modeling, local quadratic approximations, and adaptive sparsity control, aimed at selecting new experimental runs by balancing exploration, exploitation, and dimensionality reduction. The presentation emphasizes DoE as a statistical bridge between classical optimal design, sequential learning, Bayesian optimization, and machine learning. Through methodological discussion and illustrative examples, we argue that combining design-quality criteria with data-driven models can improve the reliability, interpretability, and practical usability of adaptive experimental procedures.

Registry Data to Analysis-Ready Data Through Structured Initial Data Analysis

Maja Jurtela, Lara Lusa, Tina Žagar, Nika Bric, Mojca Birk and
Vesna Zadnik

Institute of Oncology Ljubljana, Ljubljana, Slovenia

maja.jurtela@gmail.com

Much statistical work takes place after data collection but before final statistical analysis or modelling: the analysis population is defined, eligibility criteria are applied, variables are constructed, and data limitations are identified. These steps often determine what the final analysis can validly estimate, yet they are frequently documented only in code or informal notes. This issue is particularly evident in cancer registries, where routinely collected extensive and complex data are used for various purposes – routine reporting of cancer indicators, submissions to international databases, internal analyses, and research by external users. Existing initial data analysis (IDA) frameworks provide a relevant foundation, but they do not address repeated extraction from a live database, extensive validation procedures, or changing coding rules and classification systems in cancer registries. We propose a structured IDA framework for cancer registries, adapted from existing IDA principles for observational single-study datasets. It is organised into metadata, cleaning, screening, and reporting stages, and links all outputs through a dataset version identifier. This separates rule definition from rule execution, making it possible to trace how eligibility criteria and derivations produced the final dataset. The resulting outputs include both a final analysis-ready dataset and an accompanying report that enables users to understand how the dataset was prepared, and what its key properties are, before using it for statistical analyses.

Machine Learning and Regression-Based Prediction of Low Birth Weight Using Maternal Nutritional Factors, Biomarkers, and Obstetric Factors: A Multicenter Hospital-Based Study in Addis Ababa, Ethiopia

Jiregna Olani Kedida, Thierry Bossy, Corrado Lanera and Dario Gregori

University of Padova, Department of Cardiothoracic Vascular Sciences and Public Health, Padova, Italy
jiregnaolani.kedida@studenti.unipd.it

Low birth weight (LBW) remains a major contributor to neonatal morbidity and mortality in low-resource settings. Early identification of pregnancies at risk of LBW is essential for improving maternal and neonatal outcomes. This study aimed to evaluate machine-learning (ML) and regression-based approaches for predicting LBW using maternal nutritional factors, biomarkers, and obstetric factors in multicenter hospital data from Addis Ababa, Ethiopia. Of 410 imported newborn records collected from multiple hospitals in Addis Ababa, 284 eligible mother–newborn pairs with observed birth outcomes were included in the final analysis. LBW was defined as birth weight <2500g. Repeated stratified cross-validation with out-of-fold predictions was used to evaluate Least Absolute Shrinkage and Selection Operator (LASSO), elastic net, random forest, and Extreme Gradient Boosting (XGBoost) models. Model performance was assessed using area under the receiver operating characteristic curve (AUC), precision–recall area under the curve (PR-AUC), calibration, sensitivity, specificity, balanced accuracy, and decision-curve analysis. The prevalence of LBW ranged from 7.6% to 12.5% across maternal and obstetric subgroups. Elastic net achieved the highest discrimination performance (AUC=0.68), followed by random forest (AUC=0.67), LASSO (AUC=0.66), and XGBoost (AUC=0.64). PR-AUC values ranged from 0.25 to 0.27 across models. XGBoost demonstrated the highest sensitivity (39%) and balanced accuracy (62%), whereas LASSO and elastic net showed superior calibration and positive net clinical benefit across clinically relevant risk thresholds. Gestational age at delivery was the most influential predictor across machine-learning models, followed by maternal body mass index, serum folate, zinc-adjusted biomarker levels, maternal income, and maternal age. These findings demonstrate the potential utility of machine-learning methods for identifying clinically relevant predictors of LBW and supporting early risk stratification in hospital-based maternal care settings.

Performance Comparison for Privacy-Preserving Federated Learning Frameworks

Shinto Pulickal Thomas, Thierry Bossy, Dario Gregori and Corrado Lanera

University of Padua, Unit of Biostatistics, Epidemiology and Public Health, Padova, Italy

shintopulickal.thomas@ubep.unipd.it

Machine learning deployment in healthcare increasingly requires both strong privacy guarantees and efficient distributed computation. Multicenter studies present particular challenges due to stringent data protection requirements, institutional constraints on data sharing, and heterogeneous data sources. These considerations motivate the development and evaluation of scalable privacy-preserving machine learning frameworks suitable for real-world collaborative research. In this study, we conducted a comparative evaluation of two privacy-preserving distributed learning frameworks using the COSTAMI dataset as a benchmark. Performance was assessed with respect to computational efficiency, memory consumption, predictive accuracy, and privacy-related characteristics. To ensure methodological consistency, identical data preprocessing procedures, model architectures, hyperparameter settings, and optimization strategies were employed across all experiments. Model updates were aggregated using the Federated Averaging algorithm, while local optimization was performed using stochastic gradient descent. Across experiments, both frameworks achieved comparable predictive performance for acute myocardial infarction cost prediction. In the absence of Differential Privacy, the models attained a mean absolute error(MAE) of 597.27 euros for a mean observed cost of 3034 ± 2880 euros. While computational efficiency differed between the evaluated frameworks, explained variance remained comparable. Building on these findings, ongoing work focuses on secure federated learning with Differential Privacy. Current experiments are evaluating the privacy-utility trade-off across a range of privacy budgets (ϵ , δ) to establish formal privacy guarantees while maintaining predictive performance. This work will provide a comprehensive assessment of privacy-preserving federated learning in realistic multicenter healthcare settings and support the deployment of secure, scalable, and regulation-compliant machine learning workflows.

Beyond Calibration and Resolution: A New Decomposition of Posterior Cross-Entropy for Evaluating Likelihood Ratio Systems

Giulia Cereda, Fabio Corradi and Cecilia Viscardi

University of Florence, Department of Statistics, Computer Science, Applications »G. Parenti«, Florence, Italy

giulia.cereda@unifi.it

Likelihood ratios (LRs) are widely used in forensic science and other decision-making contexts to quantify the strength of evidence. When multiple methods to evaluate LRs are available, a principled comparison requires evaluation criteria that account not only for discrimination performance but also for calibration and information loss. Posterior cross-entropy (PCE), rooted in Bayesian decision theory and information theory, provides a natural framework for this purpose. In this work, we propose a novel decomposition of posterior cross-entropy that extends existing formulations based on reliability and resolution. The new decomposition expresses the expected logarithmic cost as the sum of three components: reliability loss, posterior uncertainty, and a new quantity that we term coarseness, quantifying the information lost through the data reduction induced by the summary statistics employed by a method. The proposed framework is applied to the evaluation of several likelihood-ratio methods developed for the forensic rare type match problem involving Y-STR haplotypes. These methods differ substantially in the way they compress information, ranging from simple count-based approaches to models exploiting the full allelic structure of genetic profiles. Through extensive simulation experiments, we show how the proposed decomposition provides insight into the trade-off between calibration, discrimination, and information reduction, revealing methodological differences that are not apparent from standard performance measures alone. The results suggest that coarseness represents an additional and practically relevant dimension in the assessment of LR systems. More generally, the proposed framework is applicable beyond forensic genetics and offers a new perspective for evaluating probabilistic prediction systems in statistics, data science, and artificial intelligence.

Social Sciences and Humanities 2

From Online Support to Bystander Intervention: A Mixed-Methods Perspective on Intimate Partner Violence in Slovenia

Vanja Erčulj

University of Maribor, Faculty of Criminal Justice and Security, Ljubljana, Slovenia

vanja.erculj@um.si

Intimate partner violence (IPV) remains a largely hidden phenomenon, strongly shaped by social support, public attitudes, and willingness to intervene. This contribution integrates findings from two complementary studies conducted in Slovenia using qualitative and quantitative approaches to provide a broader understanding of victims' help-seeking and societal responses to IPV. The qualitative study explored online support groups as sources of social support for women experiencing IPV. Previous research has shown that victims of IPV rarely seek help from formal institutions, which highlights the potential importance of informal and online support environments. The findings show that women in the investigated community mostly seek informational support, spend little time actively contributing to online discussions, and rarely become engaged members of supportive online groups. Most posts described psychological violence perpetrated by an intimate partner, while emotional support was more frequently sought by victims of sexual violence. Posts describing experiences of violence received more reactions and support than posts explicitly asking for help, suggesting that indirect disclosure may generate greater engagement within online communities. The quantitative study examined bystanders' attitudes toward IPV and their willingness to intervene. Results showed that tolerance of IPV significantly predicts passivity and lower readiness to intervene or report violence. Important differences emerged by gender and place of residence, with men and respondents from rural areas expressing more tolerant attitudes toward certain forms of IPV. Overall, the findings emphasize the crucial role of attitudes in shaping intervention behavior and underline the importance of strengthening both informal support mechanisms and public intolerance toward IPV.

Monte Carlo Evaluation of Methods for Estimating Regression Coefficients Involving Latent Variables

Katja Škoda

University of Ljubljana, Ljubljana, Slovenia

skoda.katja@gmail.com

Structural equation modeling (SEM) is the standard approach for estimating regression relationships between latent variables because it simultaneously models measurement error and structural relations. However, alternative approaches are often used in applied research due to their simplicity or because SEM may encounter convergence problems, particularly in small samples. Common alternatives include regression based on estimated latent variables, which can produce biased regression coefficient estimates due to measurement error. A recently proposed alternative is Structural-After-Measurement (SAM), a two-step approach that accounts for measurement error while potentially reducing nonconvergence and improper solutions. Despite its potential advantages, SAM remains relatively unexplored, particularly for ordinal indicators commonly used in social science research. This study presents a Monte Carlo simulation comparing four approaches for estimating regression coefficients involving latent variables: regression using standardized mean scores, regression using Empirical Bayes Modal (EBM) factor scores, SAM, and SEM. The simulation systematically varies sample size, factor loadings, number of indicators, indicator scale type (continuous vs. ordinal), indicator distribution, and the role of the latent variable in the regression model. Performance is evaluated in terms of the bias, and root mean squared error (RMSE) of the standardized regression coefficient, as well as the frequency of nonconvergence and improper solutions. Preliminary results suggest that estimation accuracy is primarily affected by sample size, factor loadings, and the number of indicators. Overall, SEM and SAM outperform regression approaches based on estimated latent variables, producing less biased and more accurate estimates of the standardized regression coefficient.

Measurement and Modeling

Resolution, not frequency: a discrimination index of self-assessment and achievement in computer-based testing

Danilo Strugarevic, Nikola Gligorijevic, Vladimir Cabric, Rade Rakic and Dejan Andjelkovic

Alfa University, Faculty of Information Technologies, Belgrade, Serbia

strugarevic@avmss.edu.rs

Computer-based testing produces a final score that captures answer correctness but does not separate knowledge-based responses from successful guessing. A respondent's "I know / I don't know" judgement is a candidate complementary signal, but in raw form it is ambiguous. We ask which property of self-assessment relates to achievement: how often a respondent claims knowledge, or how well those claims separate correct from incorrect answers. We analysed 748 multiple-choice responses from 25 respondents on 30 items, scored with guessing-corrected (negative) marking from -1 (incorrect) through 0 (omitted) to +1 (fully correct), with partial credit and a binary verbal self-assessment per item. Crossing self-assessment with outcome produced 33.7% paradoxical instances (lucky guesses and confident errors), so self-assessment alone is an insufficient validity signal. As instances are nested within respondents and items, individual-difference analyses were aggregated to the respondent level. Following the calibration–resolution distinction in the metacognition literature, we examined three respondent-level indices: claim frequency (proportion of "I know"; a response-bias measure), concordance accuracy (the agreement rate between self-assessment and outcome), and a resolution index we term the personal GAP, the difference in mean score between self-declared known and unknown items. Claim frequency ($r = 0.08$, ns) and concordance accuracy ($r = -0.30$, ns) were unrelated to mean score, whereas the personal GAP was positively associated with achievement ($r = 0.42$, $p = 0.04$). Because the GAP shares components with mean score, we interpret this association as explanatory rather than predictive. Achievement is tracked not by how often knowledge is claimed but by how sharply those claims discriminate success from failure. We position the personal GAP as an interpretable resolution index for response-reliability analysis, and are extending it with per-respondent confidence slopes in mixed-effects logistic models, rank-based measures, and out-of-sample validation.

Measurement invariance, predictive invariance and comparability of individual measurements

Gregor Sočan

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia

gregor.socan@ff.uni-lj.si

The importance of establishing measurement invariance in the development of measurement instruments has been widely acknowledged across the social and behavioural sciences. However, the psychometric literature has largely neglected the question of under what conditions individuals' test scores can be meaningfully compared. It is commonly assumed—or at least implicitly suggested—that scalar invariance is a sufficient condition for score comparability. I argue that this assumption is incorrect. I introduce a new concept, person-comparison invariance (PCI), which refers to the condition under which individuals from different groups who obtain the same observed score have the same expected level of the latent trait. I derive the conditions under which PCI holds and show that scalar invariance alone is insufficient. Specifically, when scalar invariance holds, invariance of both latent means and omega coefficients is additionally required for PCI to be satisfied. Furthermore, PCI can be shown to constitute a highly restrictive special case of Millsap's predictive invariance. I present a simple graphical diagnostic for evaluating PCI, as well as a structural equation modelling (SEM)-based testing procedure suitable for practical applications. Finally, I provide empirical illustrations demonstrating the magnitude of bias in score comparisons under conditions commonly encountered in practice.

Detecting Asymmetric Persistence in Stationary Time Series: Reconstruction of a Modified Langevin Model from Natural Data

Zbigniew Czechowski

Polish Academy of Sciences, Institute of Geophysics, Warsaw, Poland

zczech@igf.edu.pl

The problem of asymmetric cross-correlations was first identified in financial data analysis, where stronger cross-correlations between financial assets were observed during market downturns than during upward movements. In natural nonstationary time series, asymmetric autocorrelations have been reported and a modified Detrended Fluctuation Analysis (mDFA) method for estimating pairs of Hurst exponents was proposed. However, this approach is not applicable to stationary time series, where distinct trend episodes cannot be reliably identified. We propose a comprehensive framework that enables both the estimation of asymmetric persistence parameters and the reconstruction of a nonlinear stochastic model of time series. The model is formulated as a modified discrete Langevin equation with reduced and finite-range memory. The modification consists of introducing an additional stochastic factor into the diffusion term. This factor depends on a sequence of the previous m increment signs and determines the sign of the entire diffusion term. We therefore interpret this reduced memory as a persistence. Persistence asymmetry can be viewed as a directed stochastic force competing with the deterministic force generated by the potential given by the drift term. As a consequence, the principal activity of the process is shifted away from the potential minimum. For the special case $m = 1$, we derive the associated Fokker–Planck equation analytically and obtain explicit corrections (induced by asymmetric persistence) to the standard reconstruction procedure for Markovian Langevin equation. Persistence parameters are estimated by comparing two histograms of joint distribution functions corresponding to positive and negative preceding increments. The proposed reconstruction procedure has been successfully validated using synthetic time series generated by the modified Langevin model with various drift and diffusion functions and different persistence parameters. An application to a geophysical time series demonstrates that the reconstructed model accurately captures the dynamics of the investigated phenomenon.

One Population, Twenty Studies: A Simulation-Based Framework for Exploring Sources of Heterogeneity in Biomedical Research

Antonina Spalińska, Wojciech Glinkowski and Bartłomiej Michalak

Cardinal Stefan Wyszyński University, Faculty of Medicine, Warsaw, Poland

antonina.slubowska@gmail.com

Scientific findings often differ across studies because researchers make different decisions regarding outcome definitions, follow-up periods, covariate adjustment, and analytical strategies. Demonstrating the impact of these decisions is challenging because the underlying truth is typically unknown. We developed a simulation-based framework that allows these sources of variability to be examined in a controlled setting while preserving the realism of clinical research. Two clinically realistic observational research scenarios were designed for medical students: the association between rehabilitation timing and outcomes after stroke, and the prognostic value of admission C-reactive protein for complications in pneumonia. Student groups developed study protocols, selected outcomes, proposed covariates, and prepared statistical analysis plans. Based on these specifications, two synthetic populations of 10,000 patients were generated using literature-informed clinical relationships and plausible confounding structures. Twenty groups selected one of the two topics. Independent random samples of 200 patients were drawn for each group from the corresponding synthetic population. Students analyzed the data in jamovi using methods covered in the course, including descriptive statistics, tests of association, and generalized linear regression models. Preliminary results demonstrate substantial heterogeneity despite common research questions and shared underlying populations. Groups selected different outcome measures, follow-up periods, predictor sets, and adjustment strategies, leading to diverse analyses and interpretations of the same clinical problem. Because the synthetic data-generating mechanisms were known, these analyses can be compared with the corresponding population-level relationships, allowing direct examination of how design choices, analytical decisions, and sampling variability shape scientific conclusions. The presentation will discuss the construction of clinically plausible synthetic populations, examples of heterogeneity arising from reasonable methodological choices, and meta-analytic synthesis of group findings. The framework offers a practical approach for teaching statistical reasoning, observational study design, and evidence synthesis, while providing

a controlled environment for studying variability in biomedical research.

Keynote

The Firth correction – a recap and new developments

Georg Heinze

Medical University of Vienna, Vienna, Austria

georg.heinze@meduniwien.ac.at

Maximum likelihood estimates of statistical model parameters such as regression coefficients in generalized linear and related models are known to be asymptotically unbiased, but they can exhibit small-sample bias which in its extreme form may correspond to non-existent (quasi-infinite) estimates. This situation has been termed complete or quasi-complete separation (Albert and Anderson, 1984) for binary outcome models, or monotone likelihood (because the likelihood cannot be maximized by a finite regression coefficient) in Cox regression models. The Firth correction, originally developed by David Firth (1993) to reduce the small sample bias of maximum likelihood estimates, has been suggested as an assumption-lean solution to the problem of separation, providing finite estimates which clearly improve the properties of their maximum likelihood counterparts (Heinze and Schemper, 2002). We will provide an overview over the phases of development, evaluation and application of the Firth correction, covering its application to logistic regression but also to other outcome models. We will also compare the Firth correction to other proposed solutions to separation such as log-F priors or Bayesian approaches (Greenland and Mansournia, 2015; Gelman et al, 2008). Recently, progress has been made in several directions, which cover improving the calibration of Firth-corrected models (Puhr et al, 2017), modifying the strength of the penalty (Elgmati et al, 2015), and applying a Firth-type penalty to clustered data (Geroldinger et al, 2022; Sterzinger and Kosmidis, 2023). We will provide an overview over these developments, highlighting new areas of application.

Biostatistics 1

AUGglmmTMB: an R package for non-boundary estimation in binomial mixed effects models

Rok Blagus and Tina Košuta

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

rok.blagus@mf.uni-lj.si

Binomial mixed-effects models are frequently used to analyse binomial outcomes in clinical, epidemiological, ecological, and other applied research settings with clustered observations. In practice, however, maximum likelihood estimation may fail, leading to infinite or extreme fixed-effect estimates, singular random-effects covariance structures, non-computable standard errors, and unreliable inference. We present AUGglmmTMB, an R package for stable estimation of binomial mixed-effects models in such settings. The package implements penalized likelihood methods based on extensions of Firth-type penalties for fixed effects and inverse-Wishart-type penalties for random-effects covariance matrices. The penalized estimator is implemented through data augmentation, allowing the workflow to remain compatible with glmmTMB. AUGglmmTMB supports several implementations of the penalized estimator, as illustrated during the talk, and provides different options for specifying penalty parameters, including a novel fully data-driven procedure that avoids the need for prior information. The functionality of the package is illustrated using data on European passerine birds. In this example, standard maximum likelihood estimation leads to extreme fixed-effect estimates due to separation, a boundary random-effects correlation, and unavailable standard errors. In contrast, the penalized approach yields finite estimates, regularizes the covariance structure, and restores standard-error computation. Overall, AUGglmmTMB provides a practical and accessible framework for fitting binomial mixed-effects models in settings where standard maximum likelihood estimation is unstable or infeasible.

Net Survival Analysis in Thymic Malignancies: Preliminary Results from a SEER Cohort

Aqsa Ali, Gloria Brigiari, Ileana Baldi, Giulia Lorenzoni, Dario
Gregori and Aslihan Şentürk Acar

University of Padua, Padua, Italy

aqsa.ali@studenti.unipd.it

Thymic malignancies are rare diverse group of tumors that occur due to dysfunction in thymic cells and are known to have a poor prognosis. Understanding long-term survival patterns among disease stages is important for prognosis and future treatment planning of the patients. Objective: To conduct a preliminary relative survival analysis of thymic malignancies using SEER data and state-level life tables. A preliminary survival analysis is conducted using SEER-derived thymic malignancy data linked with state-specific life tables. Expected survival is estimated using demographic and geographic matching variables including age, sex, race/ethnicity, calendar year, and US state. Net survival is estimated using the Ederer II and Pohar-Perme estimators, while relative survival ratios are calculated using Ederer I and Hakulinen methods to compare commonly used relative survival approaches. Survival estimates are evaluated at 1, 3, 5, and 10 years, with additional exploratory analyses by SEER historic stage. The cohort contains predominantly older adults (median age 52 years). Using the Ederer II method, relative survival is estimated 91.0% (95% CI: 89.1–92.8), 80.3% (95% CI: 77.7–83.0), 75.6% (95% CI: 72.7–78.6), and 65.2% (95% CI: 61.8–68.8) at 1, 3, 5, and 10 years, respectively; Pohar-Perme estimates are slightly lower. Stage specific relative survival, estimated using the Pohar Perme method, showed a clear gradient (localized>regional>distant). Age ($HR = 1.04$, 95% CI: 1.03–1.05) and male sex ($HR = 1.24$, 95% CI: 1.06–1.45) were highly associated with worse survival, while distant-stage disease was associated with higher excess mortality in the additive hazards model ($p < 0.001$). Conditional survival analyses indicated improved survival prospects among patients who had survived several years after diagnosis. Findings suggest that stage at diagnosis plays a major role in survival and indicate stable relative survival estimates among patients with thymic malignancies. Relative survival methods using state-based life tables suggests practical framework for studying rare cancers and may support future prognostic investigations.

Challenges in Analysis of Longitudinal Observational Study: A Case of Osseointegration of Dental Implants

Nataša Kejžar, Maša Kušar and Sebastjan Perko

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

natasa.kejzar@mf.uni-lj.si

A retrospective–prospective cohort study was conducted using data from patients treated with dental implants at a dental clinic. Osseointegration was the primary outcome, assessed through implant stability measurements. Participants received either no adjunctive treatment or one or both of the following therapies intended to promote tissue healing and bone regeneration: pulsed electromagnetic field therapy (PEMF) and intravenous ozone/oxygen auto-hemotherapy. The outcome was measured at baseline and subsequently at monthly follow-up visits for up to six months after implantation. Additional covariates included patient age, sex, implant location (upper vs. lower jaw), and implant type. As multiple implants could be placed within the same patient, the data exhibited a hierarchical longitudinal structure. Initial data analysis for longitudinal data was performed to identify data set particularities and any patterns of missingness. Subsequently, several longitudinal modelling approaches were applied to evaluate changes in implant stability over time across treatment groups. To account for both the repeated measurements and the clustering of implants within patients, linear mixed-effects models were employed. Particular attention was devoted to modelling the within-subject correlation structure over time. While the mixed model for repeated measures (MMRM), considered the regulatory standard for longitudinal data analysis, provides a flexible framework for modelling temporal correlation, incorporating the additional hierarchical level arising from multiple implants per patient is less straightforward. We discuss alternative modelling strategies for addressing these challenges and present further methodological considerations when the outcome is defined as change from baseline in implant stability rather than absolute stability measurements.

From Bioinformatics Quality Control to Statistical Modelling: A Structured Initial Data Analysis Checklist for Human High-Dimensional Genotype Data

Cosimo Grazzini, Stefania Morelli, Elena Pilli, Giulia Cereda and
Lara Lusa

University of Florence, Department of Statistics, Computer Science, Applications »G. Parenti«, Florence, Italy

cosimo.grazzini@unifi.it

Modern studies of complex human traits and populations increasingly rely on large-scale Single-Nucleotide Polymorphism (SNP) datasets for association testing, heritability estimation, population structure analysis, and polygenic risk prediction. Although such datasets typically undergo extensive bioinformatic preprocessing and quality control to transform raw sequencing or genotyping outputs into a structured analytic dataset, the data ultimately received by the statistician may still contain residual technical artefacts, biological structure, or undocumented consequences of preprocessing decisions. More broadly, a careful understanding of the characteristics of the analytic dataset is essential for appropriate modelling, reliable analysis, consistent interpretation of findings, and reproducibility. Building on existing IDA checklists for cross-sectional and longitudinal studies, we propose an IDA checklist tailored to human high-dimensional SNP data. The checklist is not intended to repeat or replace the upstream bioinformatic workflow, but to provide a structured statistical phase for documenting, verifying, describing, and understanding the analytic dataset before formal modelling. The framework distinguishes between verification of documented SNP- and individual-level quality-control criteria and exploratory IDA procedures aimed at characterising the structure, consistency, and potential limitations of the final dataset. The work has two complementary aims. First, we review how SNP datasets are inspected and processed after upstream bioinformatic preprocessing in methodological and applied studies, interpreting these practices within the IDA framework. Second, we translate these findings into a practical checklist, supported by a glossary of relevant biological concepts and reproducible code fragments for both specialist and non-specialist users of SNP data. By bridging bioinformatics quality control and statistical modelling, the proposed framework aims to improve transparency, reproducibility, and communication in analyses of human high-dimensional SNP data, particularly in studies where preprocessing decisions may influence downstream association, prediction, or population-genetic inference.

Artificial Intelligence 1

Beyond Traditional Methods: Exploring Machine Learning for Imputation in Official Statistics

Kaja Malešič and Matej Divjak

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

kaja.malesic@gov.si

Imputation methods, as an integral part of the statistical data editing process, are applied in official statistics to replace missing and erroneous values with estimated ones. Preliminary findings indicate that certain machine learning (ML) methods may enhance predictive performance. In addition, they may reduce the need for manual intervention compared to traditional approaches, as ML models can automatically learn complex relationships between variables directly from data, reducing reliance on manually defined rules. However, processes in official statistics must rely on methodologically sound and transparent procedures supported by measurable quality standards; therefore, thorough testing and evaluation of new techniques are essential. The ML algorithms were tested using data on earnings of employed persons in Slovenia. In addition to earnings-related variables, the dataset includes contextual variables related to legal entities and individuals. A simulation-based approach was applied in which missing values were artificially introduced into the target variable, regular earnings per hour, under different missing data mechanisms (MCAR, MAR, and MNAR) and at varying missingness rates (1%, 5%, 10%, and 20%). Initially, several ML algorithms were evaluated. The algorithm providing the most favourable balance between predictive performance and computational efficiency was selected for further experimentation. The predicted values were then compared with the original (true) values using standard performance metrics. The results were additionally benchmarked against well-established traditional imputation methods, allowing for a direct comparison of the relative performance of the approaches. Particular attention was given to the assessment of ML-based methods in terms of predictive performance, robustness, stability, transparency, and interpretability. Furthermore, the potential of ML-based approaches was examined with respect to operational efficiency, code parameterization, required manual intervention, traceability, reusability, and reproducibility, which are important prerequisites for implementation in a statistical production environment.

Mathematical modelling in modern science and control: Impact on the continuity of the philosophy of science, major disturbances in engineering, and the mystification and masking of knowledge (invited talk)

Robert Baždarić

University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

robert.bazdaric@pfri.uniri.hr

By creating the cyber-physical world, we must assume the responsibility of the creator, not merely that of the user. In the philosophy of science, our scientific contributions act as constructive elements in affirming social and human goals. Historically accumulated experiences, first as system creators and then as users, generate substantial knowledge, properly encoded in words and symbols as an extract of our learning processes. This presentation will discuss and reveal major processes as constructive parts of a multilayered structure that must be explicitly visible in our future scientific contributions. These are understandable only if explained in the standardised language of truth-finding: mathematics. As we variously form, develop, and work with or ultimately for artificial intelligence (AI) and/or statistics, these should not remain behind the iron bars of excessive computation and memorisation. Unfortunately, for some time, working for AI has generally been the main activity. As defined activities will be very difficult to distinguish, augmenting knowledge and separating it from learning may help. Our past scientific endeavours in building cyber-physical systems constitute crucial knowledge for maintaining continuity in progressive knowledge production. Applied mathematics, especially control theory, whether traditional or contemporary, should be widely understandable. It must continuously correct itself and provide substantial knowledge for interaction with the process, and most importantly with other scientific disciplines. Sufficient abstraction is widely developed in mathematics and, unlike social life, which lags behind technological development, with mathematics it is quite the opposite. Now, all scientific disciplines must provide mathematical expressions in predicate logic that are clear for readers from any field of science. Our models are repeatable knowledge and ontic truth generators only during testing. It is time to go beyond the fractal of everything but nothing: the interdisciplinary conversation as a consensus model for truth in the time domain.

Correlation-Aware Bayesian Aggregation of Large Language Models with Identifiability and Calibrated, Distribution-Free Uncertainty

Irem Tanriverdi and Ceylan Yozgatligil

Middle East Technical University, Ankara, Türkiye

iremt@metu.edu.tr

Large language models (LLMs) are often asked to interpret the same text, such as its meaning, tone, or intent. Five such models may agree unanimously on only one case in five, so merging them reliably is the real problem. Sharing training data and alignment methods, they tend to err together and to be overconfident. Classical aggregators such as Dawid-Skene and MACE assume independent errors, an assumption that fails for correlated LLMs. We develop a hierarchical Bayesian framework that combines several correlated LLMs into one calibrated decision with finite-sample, distribution-free error control. The framework (i) learns how reliable each model is for each class from a small labelled set, (ii) discounts agreement between models that share failure modes, so five lookalike votes do not count as five independent ones, (iii) adds a context-sensitive pragmatic layer whose strength is provably identifiable from the data rather than hand-tuned, with a credible interval excluding zero, (iv) calibrates its probabilities and uses conformal prediction to control the error rate on automatically classified cases, and (v) treats model disagreement as signal, not noise. We test it on speaker intent in television dialogue. About 10,000 utterances were annotated by five LLMs, 5,000 labelled independently by two people (Cohen's $\kappa = 0.50$ on intent). It automatically classifies 61% of cases at a certified error below 10%, halving the human effort, whereas no single model certifies any. It also outperforms the best single model (accuracy 0.728 to 0.787, $p < 0.001$) and, while fully interpretable, matches or surpasses a non-interpretable benchmark. Finally, the framework produces a calibrated prediction of how often two independent people would disagree on the label (calibration error 0.085), and extends these gains to figurative meaning. These results support treating LLM ensembles as correlated multi-agent systems rather than independent annotators.

On the Performance of Language Models in Structural Time-Series Classification

Yağmur Türkmen, Kerem Recep Gür, Pinar Cemre Yazıcı, Pelin Erkaya, Berke Baytak, Ismail Guzel, Pinar Karagoz and Ceylan Yozgatligil

Middle East Technical University, Ankara, Türkiye

yagmur.turkmen@metu.edu.tr

Accurately identifying structural characteristics of time series, including trends, structural breaks, volatility regimes, and anomalies, is fundamental to reliable forecasting, anomaly detection, and statistical inference across domains such as finance, healthcare, and climate science. Classical approaches often rely on stationarity assumptions that are violated in practice, and standard statistical tests yield inconsistent results when confronted with complex real-world patterns. Recent large language models (LLMs) and multimodal models capable of processing graphical input have shown promise for structural feature identification; however, the factors governing their classification performance remain poorly understood. This study presents a systematic evaluation of open-source LLMs and multimodal models on hierarchical time-series classification using BETISE, a synthetic benchmark of univariate series labeled with structural properties. The controlled generative design of BETISE enables unambiguous ground-truth assignment, a key advantage over real-world data where structural labels are often uncertain or contested. We define classification tasks across multiple hierarchy levels, from broad category discrimination (stationarity, deterministic trends, stochastic trends, structural breaks, and anomalies) to fine-grained subtype identification within the trend, break, and anomaly families. Performance is assessed using accuracy alongside macro-averaged precision, recall, and F1-score to ensure balanced evaluation across both frequent and rare classes. We systematically investigate four factors: (1) model scale, compared across varying parameter counts; (2) input modality, contrasting a text-only LLM receiving numerical encodings against a multimodal model receiving plotted series as images; (3) labeled example availability, varying the number of in-context examples from zero to many; and (4) supervised adaptation via parameter-efficient fine-tuning. Results show that supervised fine-tuning consistently yields the largest performance gains across all tasks, while graphical input provides the greatest benefit at coarser levels of the classification hierarchy. These findings offer empirically grounded guidance for practitioners seeking to apply language and multimodal models to structural

time-series reasoning.

Biostatistics 2

Robustness and Reproducibility of Prognostic Factors in Lung Cancer: A Multiverse Analysis Based on SEER Data

Jiregna Olani Kedida, Stefania Lando, Gloria Brigiani,
Mariavittoria Vescovo and Dario Gregori

University of Padova, Department of Cardiothoracic Vascular Sciences and Public Health, Padova, Italy
jiregnaolani.kedida@studenti.unipd.it

Accurate lung cancer staging is essential for prognosis and treatment, but conclusions can depend on analytic choices. Using SEER lung cancer data, we applied multiverse analysis with Cox proportional hazards model for overall survival. Of 1,013,305 records, clinician-defined criteria yielded 538,982 patients. We varied classes of predictors, covariate sets, and interaction terms (about 1,800 specifications) and used reproducible seeding for parallel reliability. Effect estimates varied across specifications, showing prognostic associations robust to analytic choices. Robustness used effect-size stability: median hazard ratio (HR) and the 2.5th–97.5th range of HRs across specifications; $HR > 1$ indicates higher hazard of death. For stage; compared with stage I, stage II had HR 1.57 (1.21–1.82), stage III 2.52 (1.21–4.35), and stage IV 4.84 (2.18–9.88). Other predictors: tumor nodes 0.84 (0.05–4.31), tumor grade 0.54 (0.03–3.45), tumor size 0.82 (0.15–4.84), chemotherapy 0.65 (0.07–3.69), radiation 0.85 (0.07–4.57), histology 0.65 (0.06–2.95), pleura invasion 0.59 (0.09–2.83), liver metastasis 0.56 (0.06–3.85), brain metastasis 0.56 (0.06–3.48), lung metastasis 0.63 (0.07–3.48).

Safety assessment of CO_2 insufflation during PEG insertion: a non-inferiority study with a twist

Nuša Cesar, Matjaž Homan and Bor Vratinar

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

bor.vratinar@mf.uni-lj.si

Clinical research questions are often formulated as simple group comparisons, even when the actual clinical decision is more individualized. We present a case study from a randomized study comparing CO_2 and air insufflation in children undergoing percutaneous endoscopic gastrostomy. The initial question was whether CO_2 insufflation was as safe as air. The analysis showed that CO_2 insufflation led to higher post-procedure blood CO_2 values than air after adjustment for baseline CO_2 . However, this answer was clinically unsatisfactory: a small increase in CO_2 was expected and not necessarily harmful, provided that post-procedure values remained below a clinically relevant safety threshold. This motivated a reformulation of the problem. Instead of asking whether CO_2 was safe on average, we asked for which patients CO_2 insufflation could be considered acceptably safe. The target estimand became the probability that post-procedure CO_2 exceeds a predefined safety threshold, conditional on baseline CO_2 and treatment group. We present and compare two approaches for estimating this probability. The first approach models threshold exceedance directly using logistic regression, with baseline CO_2 as a predictor. The second approach models the continuous post-procedure CO_2 measurement using linear regression and then derives a probability-of-exceedance curve from the fitted mean and estimated residual variance, under a normality assumption. We compare the two approaches across a range of simulation scenarios and then apply them to the empirical CO_2 insufflation study.

Lead time estimation in cancer screening programmes: extension of the MOCCI method

Ema Požek and Bor Vratnar

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

ema.pozek@mf.uni-lj.si

Breast cancer screening programmes aim to detect the disease at an early phase by regularly examining asymptomatic women for signs of cancer via mammography. Early detection can broaden treatment options and improve chances of recovery, whereas women who do not attend screening are typically diagnosed only when symptoms appear. Naïve survival comparisons between women with screen-detected cancers and women not included in the programme are biased. To address this bias and quantify the benefits of screening, two concepts are key: lead time and overdiagnosis. In a counterfactual framework, each woman is considered in both potential worlds: one in which she is invited to screening and one in which she is not. Lead time is defined as the difference between the times of diagnosis in these two worlds. Overdiagnosis refers to cancers detected by screening that would not have been diagnosed in the absence of invitation, for example when a non-progressive tumour is detected at screening but would never have progressed to a symptomatically detectable phase. The MOCCI (Minimizing the Observed and Counterfactual Cancer Incidence) method was developed to jointly estimate lead time and overdiagnosis by comparing cancer incidence in invited and non-invited groups and identifying the lead time distribution which best explains the observed incidence differences using maximum likelihood estimation principles. We extend the MOCCI method by incorporating tumour size at diagnosis through a biological tumour growth model. Tumour size, alongside age and calendar time at diagnosis, is introduced as an additional modelling dimension to enable estimation of the latent process parameters (e.g. tumour growth rate). Within this framework, lead time and overdiagnosis are derived from the estimated latent process. We outline the extended estimation procedure and present a simulation study assessing the validity and robustness of the proposed approach.

A Data-Augmentation Framework for L_q -Penalized Linear Mixed-Effects Models

Rok Blagus and Lan Gerdej

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

lan.gerdej@mf.uni-lj.si

Penalized estimation in linear mixed-effects models (LMMs) is increasingly important in modern biomedical studies, where high-dimensional molecular measurements are often observed together with longitudinal, repeated, or clustered outcomes. Penalization based on L_q penalties, with $0 \leq q \leq 2$, provides a flexible framework for regularization and variable selection. However, implementations of L_q -penalized estimators remain limited, particularly in high-dimensional settings. Motivated by the analysis of a riboflavin gene-expression dataset, we develop a simple and flexible framework for L_q penalization of fixed effects in LMMs. Our approach uses data augmentation: by introducing appropriately constructed pseudo-observations, the penalty is incorporated directly into the likelihood, allowing estimation with standard maximum likelihood software, such as ‘glmmTMB’ and ‘nlme’. This data augmentation yields an exact implementation of L_2 penalization and, through an iterative adaptive ridge procedure, provides stable approximations to L_q penalties for $0 \leq q < 2$, including L_1 and L_0 . We apply the method to the riboflavin dataset and investigate how different choices of q affect variable selection and prediction. In particular, we compare L_2 , L_1 , and L_0 -type penalization within our framework, and compare L_1 variable selection results against existing L_1 -penalization software, including ‘HighDimMixedModels.jl’ and ‘lmmlasso’. Selected genes are further examined using biological annotation and pathway resources such as Gene Ontology and KEGG, allowing us to assess whether different approaches recover biologically meaningful signals. Overall, we provide a practical framework for variable selection and prediction in mixed models with high-dimensional fixed effects.

Artificial Intelligence 2

Artificial intelligence in survey research and beyond

Irena Ograjensek and Matej Pičulin

University of Ljubljana, School of Economics and Business, Ljubljana, Slovenia

irena.ograjensek@ef.uni-lj.si

AI and LLM tools are here to stay and from the onset they have been challenging both the stakeholders in the survey research process and the survey process workflows. This presentation aims to address these challenges from various viewpoints, ranging from theory, practice, ethics, regulation, to available toolkits. In the presentation, experiences from related methodological fields (such as for example qualitative research) are also addressed, with the goal of contributing towards shaping the research agenda in a very complex domain.

Distributional Feature Importance for Correlated Predictors

Burcu Koca Guler and Ceylan Yozgatligil

Middle East Technical University, Department of Statistics, Ankara, Turkey

bkoca@metu.edu.tr

In the literature, there exists a rapidly increasing trend in the use of machine learning and deep learning models thanks to their high predictive power and their ability to handle non-linearity and complex structures. However, because of their high number of dimensions, the inner workings of the models are obscured, making the model decisions mysterious. Model explanation can take several forms such as rule-logic extraction and counterfactual explanations, however explainability through feature importance is the most widely adopted strategy. SHAP and impurity-based importance are examples of the primary eXplainable AI (XAI) methods targeting feature importance. However, they suffer from the high level of relation between the variables, which is called multicollinearity, and these methods distribute marginal contributions among correlated variables. Hence, in this study, we explore a framework that focuses on feature importance under multicollinearity while mainly drawing on optimal transport theory, specifically the Wasserstein distance together with structural information among the variables. Rather than relying on marginal effects alone, the framework quantifies the feature importance based on distributional shift in the model's behavior. Preliminary findings suggest that this integration is a promising alternative for explainability and interpretability, especially in the presence of multicollinearity.

A comparison of mixed-effects models and machine learning methods for longitudinal prediction

Mariša Ratajec and Rok Blagus

University of Ljubljana, Ljubljana, Slovenia

ratajec.marisa@gmail.com

In medical studies, patients are often measured repeatedly over time. Their laboratory values and clinical measurements are known up to a certain point, and sometimes their condition's future development is of interest. Predicting a patient's future values is a challenge, as measurements from the same person are correlated, and some variables that change over time remain unknown. Mixed-effects models are commonly used for this type of data because they account for repeated measurements and individual differences. However, using them for prediction is tricky, especially when future values of time-varying variables are unavailable. Machine-learning methods can sidestep this problem by using the observed patient history as a set of input features. We explored whether a similar approach could be applied to mixed-effects models and how they compare with random forest and XGBoost. The models were evaluated in a simulation study under different longitudinal data structures and training sample sizes. The results show that mixed-effects models can perform similarly to machine-learning methods in some simulation scenarios. The choice of method depends on the data and what we want from the model. If the data structure is clear, the sample is small, there are few variables, and interpretability is important, mixed-effects models can be a competitive alternative to machine-learning methods for prediction. However, if the data are more complex, with many interactions, unusual patterns, variables and enough training data, random forest or XGBoost may be a better choice.

Perceptual Aspects of Uncertainty Visualization

Bernarda Petek, David Nabergoj and Erik Štrumbelj

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

bernarda.petek@fri.uni-lj.si

In our previous research, we developed a general method for visualizing uncertainty in statistical graphics by treating the visualization itself as a statistical functional. The method raised many interesting questions for future work, both theoretical and practical. From a practical perspective, uncertainty visualizations are meant to communicate uncertainty, so it is important to understand how people perceive them. Two perceptual questions stand out. First, uncertainty visualizations may contain visible gaps because of their sample-based nature. Although these gaps do not affect the theoretical guarantees, it remains unclear whether users interpret them as meaningful features or simply as visual noise. The second question concerns the fuzzy appearance of uncertain text and digits. It remains unclear whether people recognize this fuzziness as uncertainty and whether it discourages them from reading unwarranted precision into the values. In this presentation, we will briefly introduce these perceptual open questions and present our ongoing empirical work toward addressing them.

Statistical Applications 1

The impact of parallelism assumption in risk prediction models for ordinal outcomes

Hana Šinkovec

University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia

hana.sinkovec@gmail.com

Background: Ordinal outcomes, such as disease severity or diagnostic ratings, are increasingly used in medical research as they retain more information than binary outcomes. A key challenge in risk prediction is balancing interpretability and flexibility. While standard ordinal regression models rely on the parallelism assumption, more flexible approaches such as multinomial regression require estimation of many additional parameters. Based on simulation studies, multinomial regression has been recommended for risk prediction despite potential overfitting. Methods: We reanalyzed data from more than 250,000 participants without diabetes, prediabetes, or diabetes from the 2015 Behavioral Risk Factor Surveillance System. Several modelling approaches were compared, including standard ordinal models, multinomial, and stereotype models. Performance of those models was assessed on the full dataset and under reduced sample size scenarios generated through random subsampling to evaluate predictive performance, robustness to violations of model assumptions, and interpretability. Results: Despite violations of the parallelism assumption, all modelling approaches achieved similar predictive performance in the large study population. In reduced sample size scenarios, multinomial regression showed slightly higher predictive accuracy but also greater variability in calibration performance. This variability was reduced when variable selection through backward elimination was applied. Conclusions: Our results in this large real-world application suggest that apparent violations of ordinal model assumptions do not necessarily translate into inferior predictive performance and that simpler ordinal models often remain adequate. Further empirical evaluations across diverse clinical settings are needed to clarify when increased model flexibility provides meaningful practical benefits. Such evidence is especially important given the current lack of guidance on minimum sample size requirements for developing reliable risk prediction models with ordinal outcomes.

A Comparison of Causal Inference Methods for Survival Analysis Using a Simulation Study

Petra Brlogar and Damjan Manevski

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

damjan.manevski@mf.uni-lj.si

Survival analysis concerns the analysis of time-to-event data and is widely used in medical research. However, when such data arise from observational studies, estimating causal effects of treatments or exposures is complicated by confounding. While classical survival analysis methods, such as the Kaplan–Meier estimator and the Cox proportional hazards model, are useful for describing associations and survival patterns, they do not by themselves provide a framework for valid causal inference in the presence of confounding. Several causal inference methods have therefore been adapted to the survival setting, including the G-formula, inverse probability of treatment weighting (IPTW), and augmented inverse probability of treatment weighting (AIPW). In this work, we compare these approaches through a simulation study, with particular emphasis on their robustness to different data-generating mechanisms and model misspecification. We first replicate a recent simulation study and then extend it by varying the strength of covariate effects, the distribution of treatment assignment, and the correctness of the models used for the treatment and outcome mechanisms. The performance of the causal methods is evaluated alongside the classical Kaplan–Meier estimator. The results show that methods accounting for confounding generally produce less biased estimates than the classical Kaplan–Meier estimator. However, all methods are sensitive in terms of bias in cases when the corresponding model is misspecified. Weighting-based methods also show increased variability when treatment groups are highly imbalanced, particularly for rare treatments. During the presentation we will provide a thorough illustration of the performance of the methods in various settings. The results show that methods accounting for confounding generally produce less biased estimates than the classical Kaplan–Meier estimator. However, all causal methods are sensitive to model misspecification, which can lead to substantial bias. Weighting-based methods also show increased variability when treatment groups are highly imbalanced, particularly when the treatment is rare.

Would the Lady Taste the Ratios? Concentration-Based and Compositional Analysis of Tea HPLC Data

Patrik Prša and Damijana Kastelec

University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia

patrik.prsa@bf.uni-lj.si

High-performance liquid chromatography (HPLC) is commonly used to characterize samples through chemical profiles, which are usually analyzed with standard multivariate methods. However, when compound abundances are normalized to a common total, the data become compositional. In this setting, each compound is understood relative to all others, and a change in one compound necessarily affects the relative proportions of the remaining components. This distinction is important because concentration-based analyses retain information on absolute or standardized concentrations, including total compound abundance, whereas compositional data analysis (CoDA) focuses on log-ratios between compounds. In this study, published HPLC data from tea samples were reanalyzed using both concentration-based multivariate statistics and CoDA. The approaches were compared in terms of sample type separation, explained variability, and interpretability. Centered log-ratios (CLR), analyzed through log-ratio analysis (LRA), were used as the geometrically coherent CoDA approach, expressing each compound relative to the geometric mean of all compounds. Additive log-ratios (ALR) were used as an interpretable representation, comparing each compound with a selected denominator. Caffeine was chosen as the ALR reference because it best preserved the overall log-ratio geometry, as assessed by Procrustes correlation. Concentration-based PCA reproduced the reported low-dimensional structure, with the first two principal components explaining 77.3% of total variability. PCA of ALR-transformed data explained 89.3% of variation in the first two dimensions, while LRA of CLR-transformed data explained 82.3% of total log-ratio variability. Both approaches separated strongly processed teas from less oxidized tea types, but the interpretation differed. Standard multivariate methods captured differences in absolute compound concentrations, whereas CoDA highlighted changes in relative compound composition. ALR transformation with caffeine as reference provided an interpretable representation of the log-ratio patterns underlying the compositional separation. The comparison shows that CoDA can complement standard workflows and improve interpretation when relative composition is central.

Working with real-life data: Challenges and opportunities

Dragan Tristan, Patricija Jaklič, Manca Prošek and Irena
Ograjensek

University of Ljubljana, Ljubljana, Slovenia

irena.ograjensek@ef.uni-lj.si

This presentation addresses challenges and opportunities of working with real-life data. In this framework we focus: (i) on workflow procedures which need to be applied to facilitate any sound statistical analysis (with special emphasis on data quality dimensions and testing); (ii) on our practical experiences studying the day-to-day data interaction realities of selected (mostly Slovenian) business entities (with special emphasis on »win« and »fail« cases). We apply SWOT analysis to show how in practice analytical opportunities might get thwarted by internal weaknesses and what kind of rewards the business entities which are courageous enough to explore these opportunities might gain.

Joint Distribution of Correlated Exponential Random Variables via a Sum-of-Products Representation

Fatemeh Khaledian, Reza Asvadi, Tad Matsumoto and Mohd Azri Bin Mohd Izhar

Shahid Beheshti University, Tajrish, Iran

fa.khaledian@alumni.sbu.ac.ir

We derive a new closed-form joint probability density function (PDF) of correlated exponential random variables. Such distributions arise naturally in reliability engineering, queueing theory, and wireless communications. Our primary motivation is the accurate statistical modeling of instantaneous signal-to-noise ratio (SNR) fluctuations in non-line-of-sight wireless channels with correlated Rayleigh fading. For dimensions $N \geq 3$, obtaining an explicit joint PDF under a general Gaussian correlation model is analytically challenging and typically requires multidimensional numerical integration. We overcome this difficulty by deriving an exact sum-of-products representation through a multinomial expansion of the Gaussian correlation kernel, which decomposes the joint distribution into a weighted sum of products of independent constituent distributions. These distributions admit a natural interpretation in diversity combining, where signals received through multiple branches, undergoing independent or correlated fading, are jointly processed to enhance reliability. Leveraging the proposed joint PDF representation, we analyze the theoretical limits of two diversity combining strategies in terms of system outage probability. The first is Maximum Ratio Combining (MRC), weighting each branch signal by its instantaneous SNR before combining. The second is Combining After Decoding (CAD), decoding each branch independently before packet-level combining. Under identical channel statistics and correlation structures, we compare outage probability of the two schemes. The analysis shows that CAD can achieve lower outage probability than conventional MRC for a practically relevant range of correlation levels, highlighting a trade-off between decoding complexity and effective diversity gain. The presentation includes derivation of the multivariate PDF under a Gaussian correlation model, closed-form expressions for MRC outage probability, and analytical bounds on CAD outage probability from information-theoretic capacity regions. These results address: “Is MRC the best achievable combining strategy?” Applications to cell-free wireless systems employing Distributed Multiple-Input Multiple-Output/Multiple Access Channel (D-MIMO/MAC) architectures with packet combining, a promising enabling technology for sixth-generation (6G) networks, are also discussed.

Statistical Applications 2

DBGenz: Efficient Integration of Multivariate Normal and t Probabilities

Jakob Peterlin and Nataša Kejžar

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

jakob.peterlin@mf.uni-lj.si

DBGenz is a R/Julia package for efficient numerical integration of multivariate normal and multivariate t random vectors. The package implements a novel doubly blocked version of the Genz separation-of-variables algorithm, designed to exploit modern parallel computer architectures more effectively than existing implementations. By combining algorithmic improvements with cache-friendly blocking and parallel execution, DBGenz can achieve speedups of 100× or more compared with existing R packages, depending on the dimensionality of the problem and the hardware used. The package supports integration over rectangular regions for both nonsingular and singular positive semidefinite covariance matrices, making it applicable to a broad range of statistical problems in which degeneracy, high correlation, or rank deficiency may arise. This is particularly important in multiple testing, simultaneous inference, Gaussian process modelling, spatial statistics, mixed models, Bayesian computation, and the evaluation of probabilities for correlated test statistics. Multivariate t integration further enables applications involving finite-sample inference, robust modelling, and procedures based on estimated covariance matrices. DBGenz aims to provide a fast, reliable, and accessible computational backend for researchers and practitioners who need accurate multivariate probability calculations in moderate- to high-dimensional settings. By substantially reducing computation time, the package makes it feasible to use methods based on multivariate normal and t probabilities in settings where existing tools may be prohibitively slow.

How to divide by zero? The many existing solutions in logistic regression and a proposed new one

Gaj Vidmar and Jakob Peterlin

University Rehabilitation Institute, Ljubljana, Slovenia

gaj.vidmar@ir-rs.si

The problem of division by zero occurs in logistic regression when one or more of the independent variables can perfectly predict the binary outcome. This is known as the problem of complete separation (or quasi-complete separation, if there is some overlap in the case of a numerical predictor, or there is only one zero in the contingency table in the case of a binary predictor). In that case, the usual maximum-likelihood algorithm pushes the coefficient estimates towards infinity. A wide range of approaches and methods have been developed to solve this problem, which will be briefly overviewed. They include Firth's small-sample bias correction, exact logistic regression, hidden logistic regression, regularised regression (ridge, lasso and elastic net), Bayesian logistic regression, ordinary linear regression, empirical logit, Anscombe's transformation, Laplace correction, empirical Bayes, and Lidstone and Dirichlet smoothing. A new possibility for grouped data will be proposed, which is a variant of so-called dynamic adjustment that takes into account the proportion of zeros and ones in the outcome for a given predictor value (or combination of predictors). The results of extensive simulations will be presented.

Zero-Shot Classification of Water Quality Conditions from Multi-Resolution Satellite Imagery Using Vision-Language Foundation Models for Environmental Monitoring

Muhammad Zaid Qamar, Mohammed Ajaoud, Cristiano Ciccarelli and Massimiliano Lega

University of Naples Parthenope, Napoli, Italy

muhammadzaid.qamar001@studenti.uniparthenope.it

Supervised water quality classifiers trained on satellite imagery break down outside their training domain, and collecting labeled data to retrain them is expensive and slow. Vision-language models like Contrastive Language-Image Pre-training (CLIP) avoid this, they classify images from natural language prompts, with no labeled data required. Whether this actually works on satellite imagery for environmental monitoring is the question this study addresses. We benchmark four CLIP-based models on water quality classification over Lake Trasimeno, Italy, a well-monitored eutrophic lake with documented algal bloom events, using Sentinel-2 (10m, 379 patches) and PlanetScope SuperDove (3m, 2,658 patches) imagery. The 3m vs. 10m resolution contrast makes this a direct test of how spatial resolution affects zero-shot classification performance. Three conditions are classified: clean water, eutrophication, and algal bloom. To our knowledge, this is the first study to benchmark zero-shot vision-language models across multiple satellite sensors for water quality classification. Performance is compared using overall accuracy, macro-averaged F1, and McNemar's test for pairwise model comparison. CLIP ViT-B/32 and ViT-L/14, both Vision Transformer (ViT) architectures, along with CLIP-RSICD, fine-tuned on the Remote Sensing Image Captioning Dataset (RSICD), all fall below random chance (7–16%), actively misclassifying spectral features. RemoteCLIP reaches 48.5% on Sentinel-2 and 55.7% on PlanetScope with false-color composites, statistically superior to all others ($p < 0.0001$). Domain-specific pretraining matters more than model size: RemoteCLIP substantially outperforms CLIP-RSICD despite identical architectures. PlanetScope outperforms Sentinel-2 across all models, and false-color composites beat standard RGB (red-green-blue) imagery consistently on both sensors. The research group's established remote sensing pipelines for Italian lake systems, combined with long-term environmental monitoring field expertise, made reliable cross-sensor comparison and independent ground-truth validation possible. The results show which models, sensors, and input configurations can be trusted for operational water quality monitoring, and which cannot.

Comparative analysis of co-occurrence-based and embedding-based word networks in biomedical literature

Tom Rupnik Medjedovič and Andrej Kastrin

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

tom.rupnik.medjedovic@kclj.si

Uncovering semantic structure in scientific literature is essential for automated text mining, yet the relative performance of frequency-based co-occurrence networks and dense word embeddings remains an open question. This work-in-progress study systematically investigates how semantic clusters differ when derived from traditional lexical co-occurrence models versus geometric word embedding spaces. Titles and abstracts are retrieved from PubMed using targeted queries. For co-word analysis, the text is lemmatised and a co-occurrence matrix is constructed with association-strength normalisation; for embedding analysis, raw token sequences are used to train Word2Vec and FastText models. Two clustering strategies are evaluated on the embeddings: direct k-means partitioning of the vector space, and graph-based community detection on a cosine-similarity network. Cluster quality is assessed with internal validation metrics: Silhouette Score, Davies–Bouldin Index, Calinski–Harabasz Index, and Dunn Index. The complete pipeline — from retrieval to clustering and evaluation — has been implemented and tested on an initial corpus, producing comparable validation scores across all models. Preliminary metrics reveal distinct patterns in cluster compactness and separation, pointing to practical trade-offs between frequency-based semantic maps and embedding spaces when grouping related terminology. Preliminary findings suggest that the choice of word representation (i.e., raw frequency statistics versus dense vector similarity) directly shapes the resulting semantic clusters, with each approach capturing different aspects of term relatedness. Internal validation metrics provide a principled basis for comparison, complemented by planned expert review of cluster content. This clarifies the strengths and limitations of both paradigms for organising biomedical terminology.

Workshop

The weight of evidence: Likelihood ratios in forensic statistics

Giulia Cereda

University of Florence, Department of Statistics, Computer Science, Applications »G. Parenti«, Florence, Italy

giulia.cereda@unifi.it

How should we quantify the strength of forensic evidence? While statistical methods are widely used in forensic science, their interpretation is not always straightforward, and different modelling choices can lead to substantially different conclusions. This workshop presents likelihood ratios (LR) as a coherent framework for evaluating the weight of evidence, and discusses their role within forensic statistics. The focus is on practical aspects: how likelihood ratios are constructed, how they should be interpreted, and how they behave under different modelling assumptions. Using examples primarily drawn from forensic DNA analysis, we illustrate how key choices—such as the definition of the relevant population, the treatment of uncertainty, and the handling of dependence—directly affect the evaluation of evidence. These examples highlight an important point: the strength of the evidence is not an intrinsic property of the data, but depends on the underlying statistical model. We also discuss common pitfalls in the interpretation of probabilistic evidence and show how intuitive reasoning can be misleading in this context. Participants will learn how to critically assess likelihood ratio calculations and to recognise when conclusions are sensitive to modelling assumptions. Finally, we briefly introduce Bayesian networks as a tool for handling more complex scenarios involving multiple and dependent sources of evidence. The workshop aims to provide a clear and practically grounded understanding of likelihood ratio-based reasoning, equipping participants to engage critically with forensic statistical evidence.

INDEX

Index

A

Ajaoud, M, 80
Alfieri, R, 41
Ali, A, 57
Andjelkovic, D, 50
Ashtar Telaarico, F, 40
Asvadi, R, 77
Azri Bin Mohd Izhar, M, 77

B

Baldi, I, 57
Bartolj, T, 23, 34
Batagelj, V, 39
Bavdaž, M, 23
Baytak, B, 63
Baždarić, R, 61
Bellingeri, M, 41
Bidon-Chanal Badia, A, 41
Birk, M, 44
Blagus, R, 56, 68, 71
Bogner, P, 27
Borrotti, M, 43
Bossy, T, 45, 46
Brabec, M, 28
Bric, N, 44
Brigiani, G, 65
Brigiari, G, 57
Brito, P, 18
Brlogar, P, 74
Buhin Pandur, M, 19

C

Cabric, V, 50
Cassi, D, 41
Cemre Yazıcı, P, 63
Cereda, G, 47, 59, 82
Cesar, N, 66

Ciccarelli, C, 80
Corradi, F, 47
Costa, M, 29
Cugmas, M, 33
Czechowski, Z, 52

Č

Černe, D, 37

D

Di Consiglio, L, 25
Divjak, M, 60
Dobša, J, 19

E

Erčulj, V, 36, 48
Erkaya, P, 63

F

Fabjan, Z, 26
Ferligoj, A, 33

G

Gerdej, L, 68
Gligorijevic, N, 50
Glinkowski, W, 53
Grazzini, C, 59
Gregori, D, 45, 46, 57, 65
Guzel, I, 63

H

Hamplova, D, 28
Heinze, G, 55
Helena Tavares, A, 29
Homan, M, 66

I

Ivandić, R, 34

J

Jaklič, P, 76

Jurtela, M, 44

K

Kalar, B, 34

Karagoz, P, 63

Kastelec, D, 75

Kastrin, A, 81

Kejžar, N, 58, 78

Khaledian, F, 77

Koca Guler, B, 70

Kopač, G, 37

Korenjak-Černe, S, 19

Korošec, Ž, 24

Košuta, T, 56

Kronegger, L, 42

Kump, N, 34

Kušar, M, 21, 58

L

Lando, S, 65

Lanera, C, 45, 46

Lega, M, 80

Lorenzoni, G, 57

Lotrič Dolinar, A, 23

Lusa, L, 44, 59

M

Macedo, P, 29

Mačkić, V, 34

Malešič, K, 60

Maly, M, 28

Mamolo, M, 25

Manevski, D, 74

Marsh, I, 20

Martinc, M, 36

Martinkova, P, 28

Maslii, V, 22

Matsumoto, T, 77

Michalak, B, 53

Mihelič, A, 35

Milošević, B, 32

Morelli, S, 59

N

Nabergoj, D, 72

Novak, M, 19

Novokmet, F, 34

O

Ograjenšek, I, 69, 76

Olani Kedida, J, 45, 65

P

Perko, S, 58

Petek, B, 72

Peterlin, J, 78, 79

Pičulin, M, 69

Pilli, E, 59

Polanec, S, 23

Požek, E, 67

Prislan Mihelič, K, 35

Prokić, D, 32

Prošek, M, 76

Prša, P, 75

Pulickal Thomas, S, 46

R

Rakic, R, 50

Ratajec, M, 71

Recep Gür, K, 63

Ross, G, 30

Rupnik Medjedovič, T, 81

S

Şentürk Acar, A, 57

Sočan, G, 51

Spalińska, A, 53

Srakar, A, 31

Strugarevic, D, 50

Succi, R, 25

Š

Šinkovec, H, 73

Škoda, K, 49

Štrumbelj, E, 72

Šulc, A, 35

Šutić Barbarić, L, 19

T

Tanriverdi, I, 62

Tristan, D, 76

Turchetto, M, 41

Türkmen, Y, 63

U

Umek, L, 38

V

Vescovo, M, 65

Vidmar, G, 79

Vila Rigat, M, 41

Viscardi, C, 47

Vratanar, B, 66, 67

Y

Yozgatligil, C, 62, 63, 70

Z

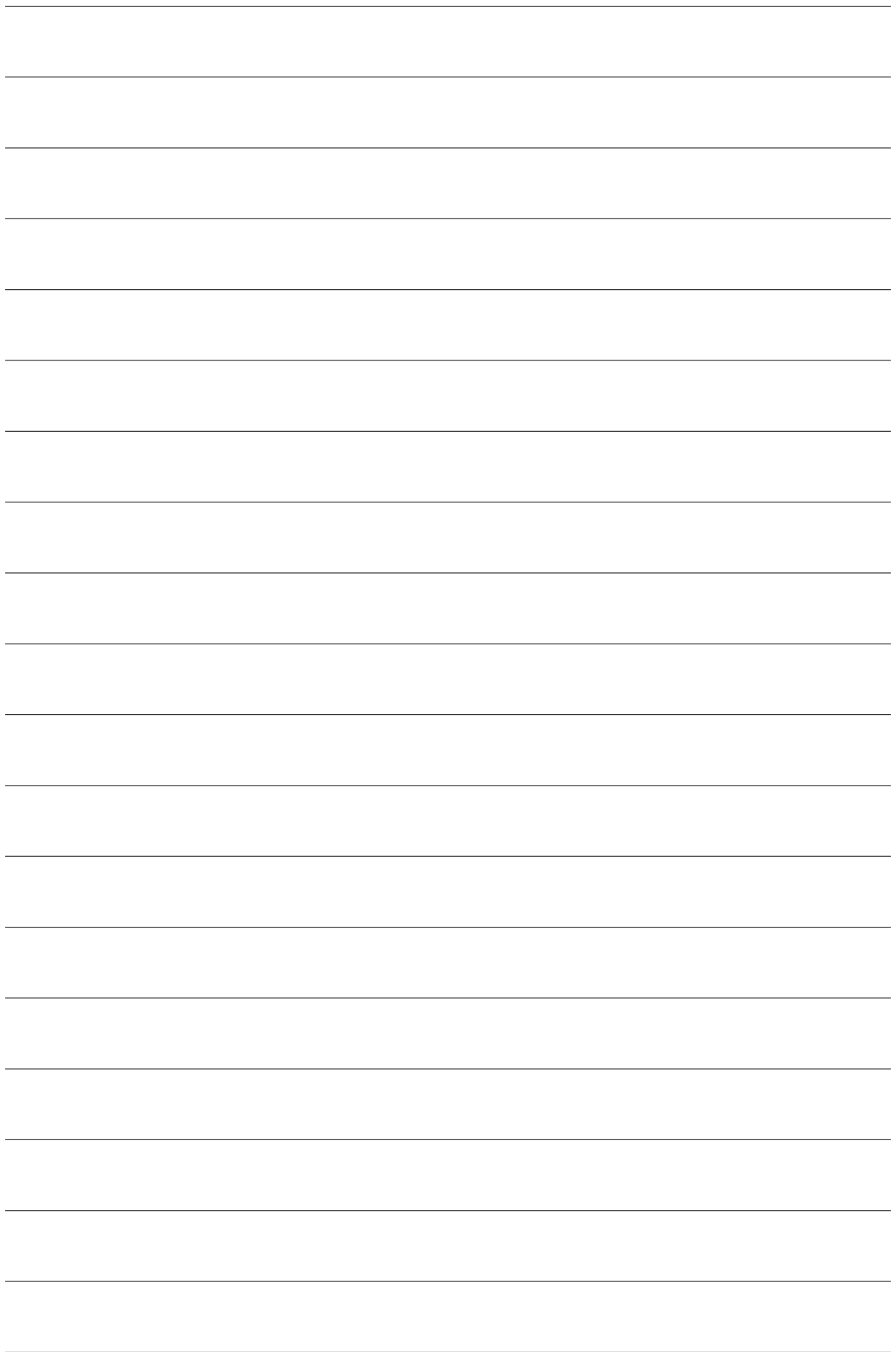
Zadnik, V, 44

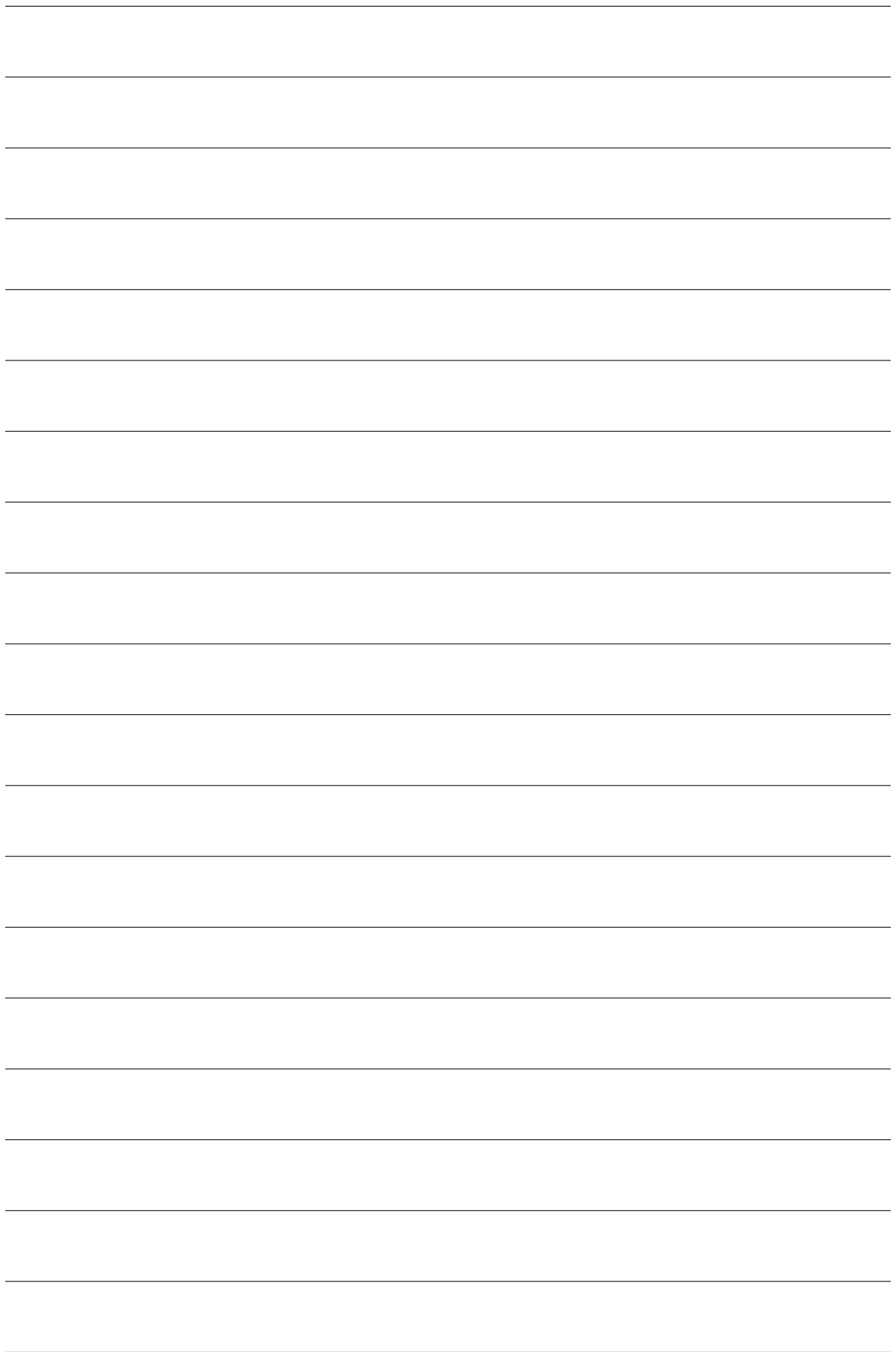
Zaid Qamar, M, 80

Ž

Žagar, T, 44

Žiberna, A, 40





```
// Backpropagation algorithm
```

```
#include <stdio.h>
#include <afx.h>
#include <math.h>
#include <stdlib.h>
```

```
#define InputN 64    // No. of neurons in the input layer
#define HN 25        // No. of neurons in the hidden layer
#define OutN 64      // No. of neurons in the output layer
#define datanum 500  // No. of training samples
```

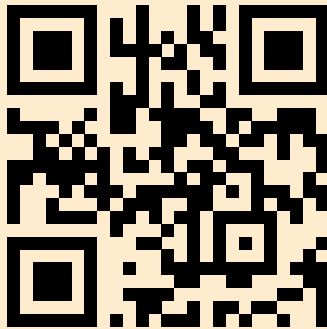
```
void main(){
    double sigmoid(double);
    CString result = "";
    char buffer[200];
    double x_out[InputN]; // Input layer
    double hn_out[HN];    // Hidden layer
    double y_out[OutN];   // Output layer
    double y[OutN];       // Expected output layer
    double w[InputN][HN]; // Weights from input layer to hidden layer
    double v[HN][OutN];   // Weights from hidden layer to output layer

    double deltaw[InputN][HN];
    double deltav[HN][OutN];

    double hn_delta[HN]; // Delta of hidden layer
    double y_delta[OutN]; // Delta of output layer
    double error;
    double errlimit = 0.001;
    double alpha = 0.1, beta = 0.1;
    int loop = 0;
    int times = 50000;
    int i, j, m;
    double max, min;
    double sumtemp;
    double errtemp;

    // Training set
    struct{
        double input[InputN];
        double teach[OutN];
    }data[datanum];

    // Generate data samples
```



<https://as.mf.uni-lj.si>