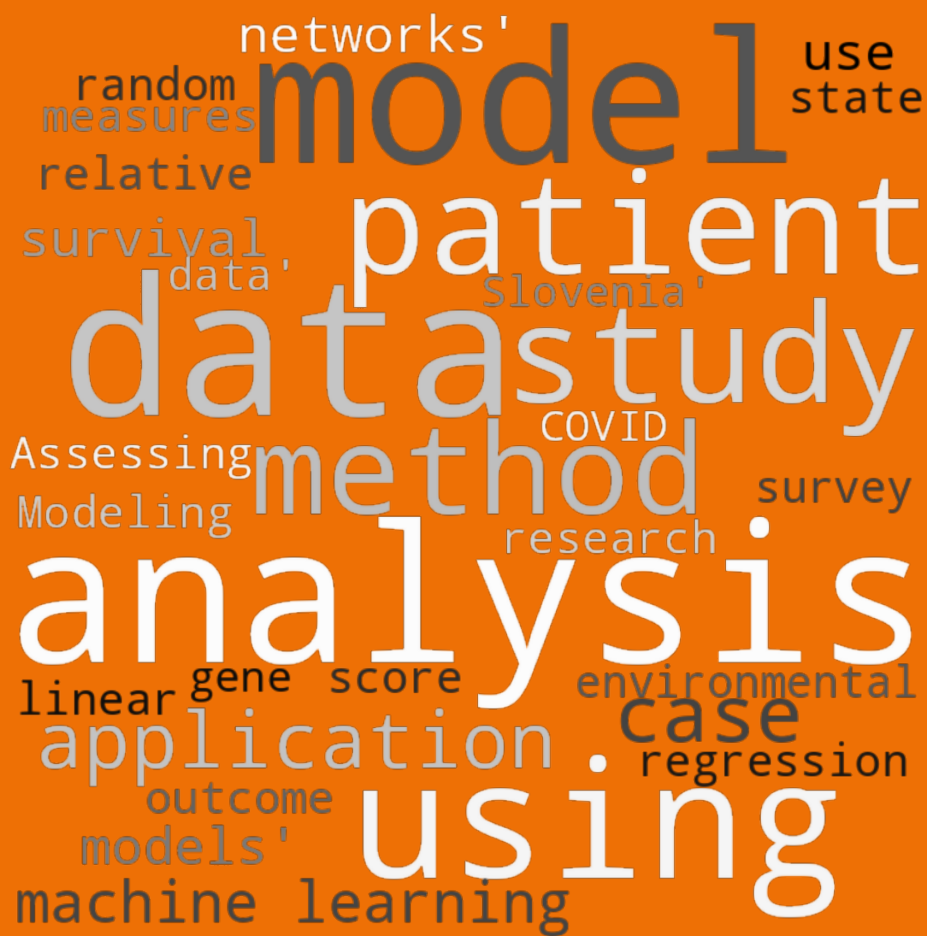


International Conference  
**APPLIED STATISTICS 2024**  
Program and Abstracts



September 23–25, 2024  
Koper, Slovenia

International Conference  
**APPLIED STATISTICS 2024**  
Program and Abstracts

September 23–25, 2024  
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>	2024
<i>Publication format</i>	PDF
<i>Electronic version:</i>	<a href="https://as.mf.uni-lj.si/uploads/pdf/as2024book.pdf">https://as.mf.uni-lj.si/uploads/pdf/as2024book.pdf</a>



## ***PROGRAM and ABSTRACTS***



***PROGRAM***



		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 Reception	
	19.45 – 21.00		
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	

08.30–15.00     **Registration**

09.00–09.30     **Opening**  
Room 1

*Chair: Lara Lusa*

09.30–10.20     **Invited Lecture**  
Room 1

*Chair: Anuška Ferligoj*

1. **Hierarchical multilevel analysis of network dynamics**  
Tom Snijders

10.20–10.50     **Break**

10.50–11.50     **Complex Networks**  
Room 1

*Chair: Tom Snijders*

1. **OpenAlex2Pajek: An R-library for creating bibliometric networks**  
Vladimir Batagelj
2. **Exploratory analysis of topic co-occurrence in the web counselling service**  
Marjan Cugmas

10.50–11.50     **Mathematical Statistics**  
Room 2

*Chair: Mihael Perman*

1. **Semiparametric modeling of educational performance characteristics**  
Marek Brabec
2. **Spectral CLTs with long memory for large language models**  
Andrej Srakar
3. **Using normal cones to solve non-linear differential equations for imprecise Markov chains**  
Damjan Škulj

11.50–12.00     **Break**

12.00–13.00     **Data Science**  
Room 1

*Chair: Vladimir Batagelj*

1. **Infrastructure innovations and their impact on electricity markets volatility: A case study of Sicily and Sardinia**  
Marina Bertolini, Francesco Lisi and Pierdomenico Dutillo
2. **Informational and topological analysis of 2015–2023 Sentinel 1 time series for detecting vegetation pest diseases**  
Luciano Telesca, Nicodemo Abate, Michele Lovallo and Rosa Lasaponara
3. **Discrete-event simulation to evaluate the best location of infrastructure to implement drones in last-mile logistics**  
Caterina Caramuta, Alessia Grosso and Giovanni Longo

12.00–13.00     **Artificial Intelligence 1**  
Room 2

*Chair: Damjan Škulj*

1. **Application of federated learning: A comparative study on cost profiling of patients after acute myocardial Infarction using federated learning**  
Shinto Pulickal Thomas, Jiregna Olani Kedida, Corrado Lanera and Dario Gregori
2. **Recognition of dietary activity via commercial smartwatch orientation sensor analysis and generative data augmentation**  
Mohammad Junayed Bhuyan, Ileana Baldi, Dario Gregori, Corrado Lanera and Luca Vedovelli
3. **Ethical considerations of using artificial intelligence tools in higher education**  
Vanja Erčulj and Ajda Šulc

13.00–14.30     **Lunch**

14.30–15.50     **Social Sciences and Humanities 1**  
Room 1

*Chair: Marjan Cugmas*

1. **Collecting visual data with web surveys: The case of bedroom flooring**  
Ana Slavec
2. **Addressing statistical literacy challenges through data repositories**  
Irena Vipavc Brvar
3. **Attitudes toward the use of artificial intelligence in medicine**  
Vanja Erčulj, Jaro Lajovic, Jonas Miklavčič, Stjepan Štivić, Eva Škrlep and Štefan Grosek
4. **The perception of professional autonomy of nurses in Slovenia**  
Rebeka Lekše and Ana Slavec

14.30–15.50     **Measurement and Modeling**  
Room 2

*Chair: Irena Ograjenšek*

1. **Validity as the fundamental measurement quality in quantitative, qualitative and mixed methods empirical research**  
Joca Zurc, Valentin Bucik, Tina Kogovšek and Anuška Ferligoj
2. **Efficient study design to estimate population means with multiple measurement instruments**  
Michal Bitan, Malka Gorfine, Laura Rosen and David M. Steinberg
3. **Spatial noise in panel regressions**  
Giovanni Millo
4. **Sample size planning in measurement invariance research**  
Gregor Sočan

15.50–16.10     **Break**

16.10–17.30      **Social Sciences and Humanities 2**

Room 1

*Chair: Gregor Sočan*

1. **A Bayesian model for age at death with unequal adult mortality and cohort effects**  
Matteo Dimai and Marek Brabec
2. **Effect sizes for detection of differential item functioning**  
Michaela Cichrová, Adéla Hladká and Patrícia Martinková
3. **SemSex: A new semi-automated tool for assessing sex education representation in Slovene curricula for evaluation and policy enhancement**  
Tim Prezelj, Timotej Knez and Slavko Žitnik
4. **Methodological competencies and attitudes towards artificial intelligence among students in pedagogical study programmes**  
Joca Zurc

16.10–18.00      **Statistical Literacy**

Room 2

*Chair: Vanja Erčulj*

1. **International statistical literacy project: Past, present, and future**  
Irena Ograjenšek
2. **Statistical literacy in early childhood**  
Ana Zalokar and Lara Lusa
3. **Advancing statistical literacy among adults through interactive play**  
Ana Slavec
4. **Statistics in picture books: Finding opportunities for enhancing statistical literacy in early childhood**  
Lara Lusa

18.00–19.30      **Sightseeing**

19.45–21.00      **Reception**

08.30–12.00      **Registration**

09.00–09.50      **Invited Lecture**  
Room 1

*Chair: Maja Pohar Perme*

1. **Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1**  
Hein Putter

09.50–10.20      **Break**

10.20–11.40      **Biostatistics 1**  
Room 1

*Chair: Hein Putter*

1. **A while-alive estimator for recurrent events**  
Maja Pohar Perme
2. **Predictors with limit of detection: Modelling event-free survival in patients with PAD**  
Nataša Kejžar, Aleš Blinc, Anja Boc, Vinko Boc, Mojca Božič Mijovski, Kevin Pelicon and Klemen Petek
3. **Leveraging cancer incidence for lead time estimation in cancer screening programmes**  
Bor Vratinar and Maja Pohar Perme
4. **Goodness of fit for linear mixed models with multiple levels of random effects**  
Jakob Peterlin

10.20–11.40      **Artificial Intelligence 2**  
Room 2

*Chair: Nada Lavrač*

1. **Machine learning at scale: From logistic regression to multi-tower approaches**  
Blaž Škrlj
2. **Enhancing clinical study analyses through federated learning: A systematic review using generative pretrained transformer and large language models**  
Jiregna Olani Kedida, Dario Gregori, Corrado Lanera and Honoria Ocagli
3. **Stochastic blockmodels of the dynamic global-trade network: A triangulation study**  
Fabio Ashtar Telarico and Aleš Žiberna
4. **Generative artificial intelligence: A new frontier in academic writing**  
Sara Blatnik, Hana Brezar, Lovrenc Cvijič and Andrej Kastrin

11.40–11.50      **Break**

11.50–13.10      **Biostatistics 2**

Room 1

*Chair: Janez Stare*

1. **Ranking of Olympic-style weightlifting performances adjusted for body mass and sex**  
Marianne Huebner
2. **Visualizing a marker's degrees of necessity and of sufficiency in the predictive-ness curve**  
Andreas Gleiss
3. **Regression modelling in an extended multi-state model using relative survival**  
Damjan Manevski, Hein Putter, Eva A. S. Koster, Johanneke E. A. Portielje, Maja Pohar Perme and Liesbeth C. De Wreede
4. **Mapping the landscape of physiotherapy research: A scientometric study**  
Alja Nike Kastrin and Janez Žibert

11.50–13.10      **Official Statistics**

Room 2

*Chair: Mojca Bavdaž*

1. **Nowcasting monthly indicators in official statistics**  
Aleša Lotrič Dolinar, Mojca Bavdaž, Tjaša Bartolj, Jerneja Kos, Barbara Mörec, Darja Peljhan, Sašo Polanec, Tjaša Redek and Jelena Zorić
2. **Daily tourism flow in Slovenia and Slovenia Interactively: The usage of PowerBI for almost real-time data visualizations**  
Teja Zupančič and Igor Štefelin
3. **Data quality in official statistics and in the Statistical Office of the Republic of Slovenia**  
Domen Repnik
4. **Enhancing statistical data collection through respondent-centric approaches**  
Ema Mišić
5. **Campaign on the 80th anniversary of the Statistical Office of the Republic of Slovenia**  
Irena Svetin and Anže Jesenovec

13.10–14.40      **Lunch**

14.40–16.00      **Statistical Applications**

Room 1

*Chair: Blaž Škrlj*

1. **Enabling topic-modeling for specific domains via domain-adaptation of LLMs**  
Boshko Koloski and Senja Pollak
2. **Statistical learning and music perception: Analyzing children's songs through the cognitive computational model IDyOM**  
Lorena Mihelač
3. **On estimating standard deviation from average moving range**  
Gaj Vidmar

14.40–16.00     **Poster Session**  
Room 2

*Chair: Nataša Kejžar*

1. **Developing a prediction model to estimate the number of slaughtered bovine animals**  
Nikolina Rizanovska
2. **Impacts of assumption violations on accuracy in test equating: Insights from a simulation study**  
Michaela Vařejková, Patrícia Martinková and Eva Potužníková
3. **Forecasting lung cancer incidence using an age-period-cohort model**  
Maja Jurtela, Tina Žagar and Vesna Zadnik
4. **Paired t-test and mixed models**  
Urška Komatar and Tinkara Perme
5. **Comparing tests for normality: A simulation study**  
Lan Gerdej and Tanja Gošnjak

16.00–16.05     **Closing**  
Room 1

9.00–12.00      **Workshop**  
Room 2

*Chair: Lara Lusa*

1. **Time-dependent effects and time-dependent covariates in survival analysis**  
Hein Putter





## ***ABSTRACTS***

## Invited Lecture

### Hierarchical multilevel analysis of network dynamics

Tom Snijders

University of Groningen / University of Oxford, Groningen, The Netherlands  
tom.snijders@nuffield.ox.ac.uk

Multilevel longitudinal network data sets are starting to be available more and more. This offers new possibilities for generalization of results of network analysis to populations of networks, and requires new methods of analysis. A multilevel version of the Stochastic Actor-Oriented Model (SAOM) was developed by Koskinen and Snijders (2023), and some experience with its use is presented here. In this hierarchical model, there are two sets of parameters: parameters at the highest level, the population of groups, which may be called the population parameters; and parameters at the group (or network) level, determining the network dynamics in each group, modelled by a SAOM. The distribution of the group-level parameters conditional on the population parameters is assumed to be multivariate normal. A fully Bayesian approach is followed in which the population-level parameters are treated as parameters with a prior distribution, and estimation is done by Markov chain Monte Carlo (MCMC) methods. To obtain good convergence of the MCMC algorithm, it is necessary to let only some of the group-level parameters vary across groups and keep the rest fixed, like in Hierarchical Linear Models for regular multilevel analysis. The analysis of each network then borrows strength from the data for the other networks, much like in the Hierarchical Linear Model. This method allows estimation of SAOMs for smaller data sets than is possible for single groups. An issue for the practical application is the necessity for the user to specify a prior distribution. Some examples are presented of the sensitivity of the estimation results to the specification of the prior.

## Complex Networks

### **OpenAlex2Pajek: An R-library for creating bibliometric networks**

Vladimir Batagelj

Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia

University of Primorska, Andrej Marušič Institute, Koper, Slovenia

`vladimir.batagelj@fmf.uni-lj.si`

OpenAlex (<https://docs.openalex.org/>) is a fully open catalog of the global research system. It's named after the ancient Library of Alexandria and was made by the nonprofit OurResearch. OpenAlex launched in January 2022 with a free API and data snapshot. It is a free alternative to commercial bibliographic services such as Web of Science and Scopus. Through its API it provides programming access to bibliographic data and enables complex analyses and the development of higher-order bibliographic services. We will present OpenAlex2Pajek - an R-library for creating bibliometric networks on selected topics from the OpenAlex database. We will show how to get bibliographic networks for some typical cases (a topic determined by keywords, selected institutions, sequence of networks for selected years, etc.) and the co-authorship network between world countries. We will also present some analyses of these networks. OpenAlex2Pajek is available at <https://github.com/bavla/OpenAlex>.

## **Exploratory analysis of topic co-occurrence in the web counselling service**

Marjan Cugmas

University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia  
marjan.cugmas@fdv.uni-lj.si

The online counselling service This is me is the largest counselling service for young people in Slovenia, with more than 57,000 questions answered since its establishment in 2001 at the National Institute of Public Health. Part of the daily work of the editorial office is to assign a topic to each question received. The most common topics of the questions relate to relationships and mental health (about half of all questions) and to body and physical health (about a quarter of all questions). While the structure of questions by topic, age, gender and other user characteristics is well known, less is known about the occurrence of topics. Understanding the topic co-occurrence is crucial as it can provide deeper insights into the complex issues adolescents face. These insights can help to develop more comprehensive and effective support strategies that address multiple issues simultaneously, rather than treating them in isolation. In addition, emerging trends can be identified to assist the service in adapting and improving other activities within the prevention programme. Therefore, the results of the applied exploratory analysis are presented using the social network methodology. Here, the nodes represent the topics and the links between the topics represent the amount of co-occurrence. Since the co-occurrence depends on the frequency of occurrence of each topic, the links are normalised using different normalizations. Special attention is paid to the most severe topics (e.g., suicide, anxiety, depression, self-injurious behaviour, and others) and other selected topics.

## Mathematical Statistics

### **Semiparametric modeling of educational performance characteristics**

Marek Brabec

Czech Academy of Sciences, Institute of Computer Science, Praha, Czechia  
mbrabec@cs.cas.cz

We will describe a flexible semiparametric approach to sequential modeling educational performance of individual students during exam period. In particular, we use complexity penalized splines motivated from expert considerations. The course of a student performance for several mathematically-oriented subjects is modelled via conditional probability decomposition following the ordering given by prerequisites utilizing both the information about the past performance and background individual-specific but time-invariant variables. We show two possible strategies to model identification based on frequentist and Bayesian paradigms. The modular nature of the model allows not only to construct predictions but also to get formalized insights into how the student performance gap propagates. Then we illustrate their performance on real data obtained from a current TACR (Technological Agency of the Czech Republic) project focused on development of an online dynamic system for student support.

# Spectral CLTs with long memory for large language models

Andrej Srakar

Institute for Economic Research, Ljubljana, Slovenia  
andrej.srakar@ier.si

Since the pioneering works from the 1980s by Breuer, Dobrushin, Major, Rosenblatt, Taqqu and others, central and noncentral limit theorems for  $Y_t$  have been constantly refined, extended and applied to an increasing number of diverse situations. In recent years, fourth moment theorem CLTs, quantitative CLTs, Breuer-Major and Dobrushin-Major CLTs, de Jong CLTs, functional CLTs and others have been developed. Recently, Maini and Nourdin (2024) extended this to spectral central limit theorems valid for additive functionals of isotropic and stationary Gaussian fields. Their work uses Malliavin-Stein method and Fourier analysis techniques to situations where  $Y_t$  admits Gaussian fluctuations in a long memory context. In another recent article, Wang et al. (2023) augmented existing language models with long-term memory. Namely, existing large language models (LLMs) can only afford fix-sized inputs due to the input length limit, preventing them from utilizing rich long-context information from past inputs. They proposed a framework of Language Models Augmented with Long-Term Memory, which enables LLMs to memorize long history. In our article we develop spectral central limit theorems in a context of augmented large language models of Wang and coauthors. Our analysis is put in a mean-field analysis context to derive appropriate limiting theorems in the usual two part scheme: a nonlinear partial differential and linear stochastic partial differential equation, to take into account the mean field limit and CLT. Analysis is set in a stochastic Ising model interacting particle systems perspective to account for the Transformer structure of the LLM (Geshkovski et al., 2024). We present applications on datasets from finance and medical imaging. In conclusion we discuss possible Bayesian extensions, as well as implications for statistical estimation and inference in a natural language processing context.

# Using normal cones to solve non-linear differential equations for imprecise Markov chains

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia  
damjan.skulj@fdv.uni-lj.si

Imprecise Markov chains involve imprecise probability distributions – typically represented as convex sets of distributions – in the modeling of stochastic processes. The Kolmogorov differential equation, traditionally used to calculate distributions in continuous-time Markov chains, transforms into a non-linear differential equation in this context. Standard linear algebraic methods are insufficient for finding its solutions, necessitating the use of convex optimization, which increases computational complexity. Existing approaches often rely on time interval discretization, requiring a large number of grid points for optimization to achieve sufficient accuracy. We propose a new, more efficient method that significantly reduces the number of optimization steps needed to reach the desired accuracy. This method leverages the Lipschitz continuity of the solutions with respect to time, ensuring that optimal solutions at nearby time points are likely associated with the same extreme point. Consequently, the solution vector remains within the same normal cone of the corresponding extreme point, transforming the optimization problem into a linear one and enabling faster computations. Preliminary tests demonstrate that our method significantly outperforms existing methods in most cases.



## Data Science

### **Infrastructure innovations and their impact on electricity markets volatility: A case study of Sicily and Sardinia**

Marina Bertolini, Francesco Lisi and Pierdomenico Dutillo

University of Padua, Department of Statistical Sciences, Padua, Italy  
marina.bertolini@unipd.it

The outcomes of the electricity markets—in terms of both prices and volumes—are strongly influenced by the infrastructures serving the market reference area. In Italy, the geographical market zones present peculiar characteristics deriving from the level of interconnections with other zones, installed energy mix and local consumption. This is particularly true for the two major islands, Sicily and Sardinia, which have limited interconnections with the mainland. In 2016, the Sorgente-Rizziconi cable was completed, enhancing the connection between Sicily and the mainland: according to the operators, this event had an impact on day-ahead prices, especially on their level. Currently, a new cable—the so called Tyrrhenian link—is in the planning phase to connect the other major island, Sardinia, with the rest of the market through Sicily. The construction of this new connection can be considered as an innovation in the electricity grid. Through the analysis of day-ahead prices in the Sicilian market and their breakdown into different components, the paper aims to study the effect that the electricity grid innovation for Sicily has had on the volatility of zonal electricity prices. Additionally, the study also aims at predicting and estimating the implications of new innovations on the grid, that are foreseen for both the Sardinian and the Sicilian markets. Working on islanded markets is particularly relevant as it allows higher control on the market environment and to draw conclusions on how infrastructure impacts on the volatility of electricity prices. With proper adaptations, results might be generalized to other markets. Conference participants can benefit from the presentation both for the methodological part and for the possibility of transferring the results to other study contexts.

# **Informational and topological analysis of 2015–2023 Sentinel 1 time series for detecting vegetation pest diseases**

Luciano Telesca, Nicodemo Abate, Michele Lovallo and Rosa Lasaponara

National Research Council, Institute of Methodologies for Environmental Analysis, Tito, Italy  
luciano.telesca@cnr.it

The goal of this study was to evaluate the effectiveness of informational and topological statistical methods applied to satellite time series for detecting vegetation features caused by pest-induced diseases. In particular, we analyzed Sentinel 1 (S1) Normalized Difference Vegetation Index time series from 2015 to 2023. The selected pilot site was Castel Porziano, located in a peri-urban area of Rome, affected by *Toumeyella parvicornis* (TP), a parasite that has been severely impacting the Pinus trees of that area in the recent years. For comparison, we also selected Follonica, located on the western side of Tuscany and characterized by the same type of vegetation as Castel Porziano, but is healthy. For both sites, we extracted 150 pixels to cover their respective areas. The analysis employed two methods: 1) the Fisher-Shannon (FS) method, that allows to characterize the time dynamics of complex time series by means of the Fisher information measure (FIM) and the Shannon entropy power (SEP); 2) the Visibility Graph (VG) analysis, which converts time series into networks, whose nodes represent the series values linked between each other by their reciprocal »visibility«. Both the methods were used to discriminate between healthy and TP-infected pixels. Our results highlight the effectiveness of S1 time series analysed by FS and VG methods in diagnosing plant deterioration due to pests. This work was supported by the project »Coelum: Spies of Climate change and tools for mitigating the effects: EO and AI based methodological approach for Urban Park Management«.

## **Discrete-event simulation to evaluate the best location of infrastructure to implement drones in last-mile logistics**

Caterina Caramuta, Alessia Grosso and Giovanni Longo

University of Trieste, Trieste, Italy  
alessia.grosso@phd.units.it

The focus on new technologies in the field of freight transport and last-mile logistics is constantly growing. This is due to the significant increase in demand for transportation, mainly caused by the growth of e-commerce in recent years, which has consequently impacted transportation systems. Among the emerging technologies to mitigate these impacts there are drones. The latter are used alone or in combination with other means of transport, such as trucks, for last-mile deliveries and promise significant environmental, economic, transport and social benefits over more traditional modes of transport. However, the implementation of such systems requires a series of infrastructures, such as intermediate depots or parking areas, to be integrated into the transport network. These infrastructures must be strategically located to minimise delivery time, costs and environmental impacts, thus providing an efficient and accessible service to users. In this work, a discrete-event model was developed to determine the best location for the necessary infrastructure. As an input to the model, in addition to the infrastructure and transport mode aspects, it was necessary to estimate the transport demand in the area. To this end, a statistical analysis was applied to the socio-economic data of the study area. The model was then used to run several simulations that, depending on the transport demand of the considered scenario, indicate the preferred positioning of depots in the transport network. Finally, the analysis of the results of the various simulations carried out made it possible to identify the best location overall, taking into account the stochastic variation in demand. This methodology was then applied to a case study in the mountainous area of north-eastern Italy, in particular in the Friuli Venezia Giulia region.

## Artificial Intelligence 1

### **Application of federated learning: A comparative study on cost profiling of patients after acute myocardial Infarction using federated learning**

Shinto Pulickal Thomas, Jiregna Olani Kedida, Corrado Lanera  
and Dario Gregori

University of Padua, Padua, Italy

shintopulickal.thomas@ubep.unipd.it

Acute myocardial infarction (AMI) presents a significant healthcare challenge, with early discharge strategies being explored to reduce hospital stays and costs without compromising clinical outcomes. The COSTAMI study showed the cost-effectiveness of early discharge with stress echocardiography compared to usual care. However, efficient health care planning can also be accomplished by predictive modelling techniques implemented on cost data. This study aims to develop a predictive model in this context using Federated Learning (FL), a distributed machine learning technique that allows multiple institutions to collaborate without sharing raw data. The primary objective of this study is to develop a predictive model for cost of care of patients with AMI using federated learning and to compare the results with those obtained in the original classical analyses of the COSTAMI study. The model will be trained in an isolated environment for each centre in the COSTAMI, thereby simulating a privacy-preserving environment while enabling the creation of a shared model. The FL approach facilitates the creation of a shared model, enabling health care organization decisions based on comprehensive datasets from all isolated environments. The model's predictions are validated against a test set and a validation split from the training data. The prediction model is developed with input and approval from principal stakeholders and theoretical experts, ensuring robustness and reducing the computational burden on resource-constrained devices. The integrated approach is expected to perform in FL supported systems and software. The anticipated outcomes suggest that federated learning could improve prediction accuracy in cost-profiling at patient level, while operating in isolation across centres. With a better grasp of cost-effectiveness, supporting the potential integration of advanced privacy measures such as differential privacy and homomorphic encryption, this approach may help in adopting evidence-based healthcare management.

# **Recognition of dietary activity via commercial smartwatch orientation sensor analysis and generative data augmentation**

Mohammad Junayed Bhuyan, Ileana Baldi, Dario Gregori,  
Corrado Lanera and Luca Vedovelli

University of Padua, Padua, Italy

mohammadjunayed.bhuyan@studenti.unipd.it

Smartwatches equipped with a range of sensors have emerged as promising tools for monitoring and improving individuals' health and lifestyle. This study focuses on the application of commercial wristwatch sensor (accelerometers and gyroscopes) data for detecting eating episodes, aiming to provide a non-intrusive, continuous monitoring of dietary behavior. Twenty healthy participants were equipped with smartwatches (Garmin® Fenix® 5 watches), during four meals under semi-naturalistic settings. All meals were recorded on video, synchronized with sensors, and eating behavior was manually annotated by two reviewers. Raw sensors data were transformed to obtain a rectangular dataset. Rigorous machine learning methods were used to classify eating activity. More than 130k observations were obtained and we applied machine learning models to classify eating activity and hyperparameter tuning was applied on the best models. We finally applied data augmentation techniques using Generative Pre-trained Transformer (GPT) and Generative Adversarial Network (GAN) to address challenges related to variations in eating behaviors. Results showed a good performance of tree-based classifiers (balanced accuracy  $\approx 0.80$ ). Accelerometer features were identified as the most important. We demonstrated the efficacy of smartwatch sensors in monitoring dietary behaviors. The utilization of tree-based models, coupled with GPT-based augmentation, highlights a promising avenue for enhancing the accuracy of eating episode detection. This study underscores the potential of wearable technology in providing a non-intrusive and continuous approach to dietary monitoring, albeit with room for improvement in specificity metrics.

# **Ethical considerations of using artificial intelligence tools in higher education**

Vanja Erčulj and Ajda Šulc

University of Maribor, Faculty of Criminal Justice and Security, Ljubljana, Slovenia  
vanja.erculj@um.si

Artificial intelligence (AI) tools are becoming widely used in higher education, amongst others for grades prediction. Also, students use AI during their study to enhance the understanding of study topics and preparation of study papers. With usage of AI several ethical dilemmas arise. Four major considerations that are emphasized in the scientific literature are fairness, accountability, transparency and ethics—commonly referred as “FATE”. Fairness is concerned with bias in prediction when using AI tools, accountability with responsibility for erroneous decisions of AI tools developers and users, transparency with understanding how AI tools work and ethics as an umbrella term encompassing ethical issues concerning the data privacy and Internet user’s autonomy. A pilot study was conducted on students with the objective to explore their understanding of FATE when using AI tools. An online survey was carried out including 178 undergraduate university students. The analysis showed that majority of students already used AI tools for study purposes, although current use is still rare. The results indicate that students are aware of ethical issues when using AI tools and that they are concerned for the privacy of data they share online.

## **Social Sciences and Humanities 1**

### **Collecting visual data with web surveys: The case of bedroom flooring**

Ana Slavec

University of Primorska, Koper, Slovenia

InnoRenew CoE, Izola, Slovenia

ana.slavec@famnit.upr.si

Flooring materials have distinct properties that significantly influence the comfort and well-being of residents. Natural materials, in particular, can contribute to better indoor air quality. Moreover, their lower carbon footprint in both production and disposal makes them more environmentally friendly compared to synthetic alternatives. However, official data often lacks specific information about the materials used for flooring and other household furnishings. In August 2023, we invited members of an online market research panel in Slovenia to participate in a questionnaire on residential renovation and furnishing, which garnered responses from 1009 participants. The survey included a question about the type of flooring material used in bedrooms, accompanied by a sub-question to gauge the respondents' confidence in their answer. A week later, participants were invited to complete a follow-up survey that required them to upload a photo of their bedroom floor, resulting in 320 photos collected. For cases where uploads did not occur, we investigated the reasons why. Additionally, all respondents were asked if they would be willing to provide other types of visual and supplementary data. In this study, we explore the willingness of respondents to provide visual data and how it varies across different socio-demographic characteristics. Furthermore, we assess the quality of the visual inputs in terms of image authenticity and their potential to complement and enrich survey data. Finally, drawing on our experiences, we address key methodological challenges and offer recommendations for future research aimed at collecting visual data in surveys.

# Addressing statistical literacy challenges through data repositories

Irena Vipavc Brvar

University of Ljubljana, Faculty of Social Sciences, Slovenian Social Science Data Archives, Ljubljana, Slovenia

[irena.vipavc@fdv.uni-lj.si](mailto:irena.vipavc@fdv.uni-lj.si)

Data repositories provide a secure and structured environment for storing and sharing research outputs, ensuring that valuable data is preserved and accessible for future use. With growing emphasis on Open Science, including the funders as European Commission and national research bodies, researchers are increasingly encouraged to make their data openly available. As a result, more data, accompanied by structured metadata, controlled vocabularies and survey instruments, are to researchers as well as the broader public through various repositories. The role of data archives is crucial in ensuring that data is well-described and trustworthy, allowing researchers to trust published materials. This presentation will focus on key elements of data discovery, showcasing the materials stored in these repositories and highlighting how they can be effectively utilized in future research endeavours as well as teaching. By improving data discoverability and usability, data repositories play a pivotal role in advancing scientific progress and supporting the Open Science principles. Additionally, platforms such as the Atlas of European Values provide accessible data for increasing statistical literacy and offer key visualizations on policy-relevant themes, making them valuable tools for both researchers and policymakers. These platforms contribute to bridging the gap between data availability and its practical application in research, teaching, and policy development.



## **Attitudes toward the use of artificial intelligence in medicine**

Vanja Erčulj, Jaro Lajovic, Jonas Miklavčič, Stjepan Štivić, Eva Škrlep and Štefan Grosek

University of Maribor, Faculty of Criminal Justice and Security, Ljubljana, Slovenia  
vanja.erculj@um.si

Artificial intelligence (AI) is used in preclinical, clinical and research work in various branches of medicine. This raises various ethical issues for users. Martinho and colleagues identified four key perspectives of researchers on the use of AI in medicine. The aim of this research was to investigate the attitudes and perspectives on the use of AI in medicine of Slovenian researchers working in the medical field. For this purpose, an online survey was carried out among researchers and lecturers working at the University Clinical Centre Ljubljana and Maribor. With the permission of Martinho and colleagues, we localised the questionnaire, pilot-tested it and built the final version. The web link was sent to 536 researchers and educators, from which 165 decided to participate in the study. The findings show a high importance of ethics and clear rules in the use of AI in medicine. Through principal component analysis, 6 components were extracted, indicating perspectives, that are of key importance to the participants. AI technology can be helpful for healthcare professionals, but they also need to understand how it works. The use of AI must not expose patients to additional risks and must be used in accordance with the strictest bioethical principles. The perspectives are related to those identified by Martinho and colleagues in their study on researchers from three different countries working in different medical fields.

# **The perception of professional autonomy of nurses in Slovenia**

Rebeka Lekše and Ana Slavec

University of Primorska, Faculty of Education, Koper, Slovenia

rebeka.lekse@fvz.upr.si

Nurses' professional autonomy is key to high quality care as it allows nurses to make decisions based on their expertise and experience. We conducted a survey among the general population and a sample of employed nurses in Slovenia. The results provide an insight into the complexity of factors influencing nurses' autonomy in the Slovenian healthcare system and allow us to analyse the different views of the general population and employed nurses about their professional role. The studies shows that the general population recognises the key role of nurses in the healthcare system, but often does not fully understand their competence and autonomy in decision-making. In contrast, nurses self-identify as autonomous professionals who act on the basis of their knowledge and experience. The clear discrepancy between the public's perception and nurses' self-perception highlights the need to sensitise the public to the complexity of nurses' role and their key role in care. As the concept of professional autonomy is central to many professions, education has a crucial role to play in promoting and reinforcing this concept.

## Measurement and Modeling

### **Validity as the fundamental measurement quality in quantitative, qualitative and mixed methods empirical research**

Joca Zurc, Valentin Bucik, Tina Kogovšek and Anuška Ferligoj

University of Maribor, Faculty of Arts, Maribor, Slovenia

joca.zurc@um.si

Validity, as the assessment of the relationship between a theoretical construct and a measured variable, is, along with reliability, the most important quality dimension of any social science research, regardless of its methodological orientation. Nevertheless, there are only a few attempts in the literature to define validity criteria in empirical research. Quantitative and qualitative researchers endeavour to define valid measurements but do not agree on the common underlying concepts or terminology. Our study aimed to explore the criteria for assessing validity in social science research within three different methodological orientations—quantitative, qualitative and mixed. The study was based on a systematic review of previous research published in indexed social science methodological journals such as the Journal of Mixed Methods Research, International Journal of Social Research Methodology, Quality & Quantity, and International Journal of Multiple Research Approaches, as well as international databases such as Google Scholar, Web of Science, WebSurvey Methodology, PsycINFO, and PsycARTICLES. Meta-synthesis, a conceptual summary of the results of several comparable studies on a particular research problem, was used to analyse the studies that met the inclusion criteria. Eighteen studies were included in the analysis, of which 77.8% addressed the validity of one methodological approach. However, the remaining studies focused on analysing the validity of two or three methodological approaches. The meta-synthesis revealed three central themes: (i) the change in the concept of validity over time in all three methodological approaches, (ii) different approaches and methods for assessing validity in quantitative and qualitative research, and (iii) the development of new specific criteria for assessing validity in mixed methods research. One-third of the research was devoted to the development of new criteria for assessing validity. Further methodological research is needed to develop new criteria for assessing validity in empirical social science research.

## **Efficient study design to estimate population means with multiple measurement instruments**

Michal Bitan, Malka Gorfine, Laura Rosen and David M. Steinberg

College of Management, Rishon LeZion, Israel  
michalbi2@colman.ac.il

**Background:** Many studies assessing exposure to substances like tobacco smoke or dietary nutrients rely on both direct biomarker measurements (e.g. urinary cotinine, sodium levels) as well as indirect self-report measures. Combining these measurement approaches can provide a more accurate picture than using a single type of measure alone. **Methods:** Building on previous measurement error models, we develop a framework for efficiently designing studies with both direct and indirect exposure measurements. Key parameters factoring into the optimal design include the relative costs of the different measurements, the measurement error variances, and the regression relationship between direct and indirect measures. Our approach optimizes the sample size, proportion receiving direct measurements, and number of replicates to minimize the variance of the estimated population mean exposure for a given budget. We provide open-source software to implement the methods. **Results:** When applying the methods to conditions that arise in real-world studies, we found that the optimal design was fairly insensitive to inaccurate specification of the input parameters at the planning stage. The optimal allocation of resources between study arms depends on whether direct measurements are taken on all participants or just a subset. If budget constraints limit direct measurements to a subset, more resources should be allocated to the arm with higher population variance for the indirect measures. The number of replicates was driven by cost and error variance ratios, with additional replicates of cheaper indirect measures improving precision when direct measures were infeasible for all participants. Overall, selective use of costlier direct measures combined with replicating cheaper self-reports enabled identifying efficient designs across scenarios. **Conclusion:** Accounting for measurement properties through combined direct and indirect measures enables more powerful and cost-effective study designs than relying on a single measurement approach. The outlined framework allows researchers to optimize their design given project-specific constraints and measurement characteristics.

## Spatial noise in panel regressions

Giovanni Millo

University of Trieste, Trieste, Italy  
giovanni.millo@deams.units.it

Following the recent contribution of Kelly, we address the bias in estimated standard errors arising from variables sharing a common spatial structure. To illustrate the issue, we fit independently-generated, stationary but spatially correlated variables ("spatial noise") in a panel data regression, showing false positive rates of up to 50% in standard significance diagnostics. This is a well-known result in (cross-sectional) spatial econometrics since Bivand (1980). Smith and Lee (2012) present a recent Monte Carlo analysis assessing the magnitude of the bias. Spatial econometric models, in particular the spatial error model (SEM), solve the problem by modelling the spatial dependence. The "true" spatial data generating process, nevertheless, is rarely known. Panel data allow for a number of robust estimation techniques to alleviate the issue without assuming a known pattern for the spatial dependence. We assess how far these tools can go in controlling the empirical size of the significance test. We also evaluate the empirical performance of parametric solutions to the problem, including estimating spatial econometric models with a misspecified spatial weights matrix. References: Bivand, R. (1980). A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. *Quaestiones Geographicae*, 6, 5-10.; Kelly, M. (2019). The standard errors of persistence. Working paper.; Smith, T. E., & Lee, K. L. (2012). The effects of spatial autoregressive dependencies on inference in ordinary least squares: a geometric approach. *Journal of geographical systems*, 14, 91-124.

# Sample size planning in measurement invariance research

Gregor Sočan

University of Ljubljana, Ljubljana, Slovenia

gregor.socan@ff.uni-lj.si

The study of measurement invariance (or measurement bias) has become a standard part of the evaluation of tests, scales, and questionnaires, especially in cross-cultural settings. At the same time, power analysis is a critical component of any statistical hypothesis testing process. However, power analysis requires standardized measures of effect size, a topic that has received surprisingly little attention in the field of measurement invariance. As a result, the literature provides limited guidance on the statistical power of measurement invariance tests and related sample size requirements. Recently, however, some general suggestions for the interpretation of certain effect size indicators have emerged. The purpose of this study was to empirically estimate (by simulation) the power to test the measurement invariance hypothesis at various effect size levels and to examine how factors such as the type of effect size measure (signed vs. unsigned), the scaling of the latent variable, and the magnitude of the factor loadings influence it. I will discuss the determinants of statistical power, clarify some counterintuitive findings, and provide tentative guidelines for applied research.

## Social Sciences and Humanities 2

### **A Bayesian model for age at death with unequal adult mortality and cohort effects**

Matteo Dimai and Marek Brabec

University of Trieste, Trieste, Italy  
matteo.dimai@phd.units.it

The conventional approach to modeling age at death through parametric models often employs a three-component mixture as a standard. However, recognizing the need to more accurately reflect the diverse trends in old age mortality, this study introduces a five-component Bayesian mortality model. This model is designed to address data exhibiting cohort effects and transitions in the old age mortality component, which moves from a skew-normal distribution to a mix of Gaussian and skew-normal distributions as the period progresses. The proposed model comprises a point probability for infant mortality, a uniform distribution for background mortality, a Gaussian distribution for premature mortality, and two components for old age mortality (Gaussian and skew-normal). Applied to data from 10 Italian regions spanning 1974–2022, including cohort effects for cohorts born between 1913 and 1947, the model demonstrates a stable interpretation of mortality components over time. In contrast to simpler models, it reveals a significant shift in old age mortality from a skew-normal to a combined Gaussian and skew-normal distribution around the millennium's turn. This study not only sheds light on the evolution of old age mortality in Italy, highlighting a non-uniform shift and compression of mortality across the population, but also lays the groundwork for future forecasting models by maintaining a consistent parameter interpretation throughout the period.

## Effect sizes for detection of differential item functioning

Michaela Cichrová, Adéla Hladká and Patrícia Martinková

Institute of Computer Science, Czech Academy of Sciences and Faculty of Mathematics and Physics,  
Charles University, Prague, Czechia

cichrova@cs.cas.cz

An analysis of multi-item measurements in educational or psychological assessments includes testing for Differential Item Functioning (DIF), that is, identifying items that function differently for distinct population groups, indicating potential bias. The presence of DIF can be tested using various statistical methods, such as those based on contingency tables, logistic regression, and item response theory models. For items with statistically significant DIF, it is important to assess the practical significance by quantifying the magnitude of DIF with an appropriate effect size measure. This is necessary because even differences with no practical impact may be statistically significant. In this work, we review existing DIF effect size measures and the cut-off values used to classify the effect size magnitudes for the Mantel-Haenszel test, SIBTEST, and model-based approaches. We conduct a simulation study to investigate the properties of these effect size measures and their existing classification guidelines. Based on the simulation study, we suggest updating some of the values. Additionally, we suggest appropriate cut-off values for effect size measures based on the area between the item characteristic curves, taking those based on the Mantel-Haenszel approach as a reference. Furthermore, we investigate the newly proposed cut-off values through an additional simulation study. We propose solutions to observed inconsistencies and issues, focusing on the practical implementation using the R software.



## **SemSex: A new semi-automated tool for assessing sex education representation in Slovene curricula for evaluation and policy enhancement**

Tim Prezelj, Timotej Knez and Slavko Žitnik

University of Ljubljana, Ljubljana, Slovenia

tim.prezelj@gmail.com

Ensuring comprehensive representation of critical topics in educational curricula is essential for a balanced education. This paper introduces an automated approach for evaluating the coverage of sex education topics in Slovene curricula. We developed a detailed ontology for sex education, enriched with connections to the Wikidata knowledge graph, creating a robust analytical framework based on internationally recognized UNESCO and WHO sex education standards. Our methodology includes constructing sentence-level classifiers to detect and categorize references to sex education concepts within curriculum documents. The models successfully identified and classified relevant concepts, demonstrating the effectiveness of our approach. Additionally, we created a pipeline that transforms curriculum documents into structured knowledge graphs, enabling systematic analysis and visualization of the content. This research not only highlights the representation of sex education in Slovene curricula but also offers a generalizable methodology for (semi)automated curriculum analysis across various educational domains. Moreover, the approach has broader applications, such as examining perspectives on different topics by comparing Wikipedia articles in different languages to the English version, among other possibilities. This work represents a significant step toward using automated tools for curriculum evaluation and policy improvement, providing valuable insights for educators, policymakers, and researchers seeking to enhance educational content and its alignment with international standards. Furthermore, this approach holds broader potential for exploring diverse perspectives on various topics by comparing Wikipedia articles in different languages with their English counterparts, among other applications.

# **Methodological competencies and attitudes towards artificial intelligence among students in pedagogical study programmes**

Joca Zurc

University of Maribor, Faculty of Arts, Maribor, Slovenia

joca.zurc@um.si

Teachers' attitudes towards artificial intelligence play a decisive role in its acceptance and use in the classroom. Therefore, the recommendations emphasise that the higher education of pedagogical students should include the development of digital competencies in at least one stand-alone course as well as transversal competencies in other courses of their curriculum. Teaching research methodology, which was initially associated with the use of advanced information and communication technologies for data analysis since the beginning of teachers' higher education, provides an excellent opportunity to promote a positive attitudes towards artificial intelligence. The aim of our research was to empirically investigate pedagogical students' views on quantitative, qualitative and mixed methods competencies and assess their attitudes towards artificial intelligence. The study was conducted on a sample of 478 students (81.6% female, mean age = 23 years) taking a course in pedagogical research methodology. The data was collected via an anonymous online questionnaire via the 1KA platform and analysed using a t-test for independent samples, Pearson correlation and hierarchical cluster analysis. The results showed that most respondents identified themselves as novice researchers (88%) with stronger competencies in quantitative research methodology (77%). 81% of respondents rated the use of ICT and artificial intelligence systems in methodology courses as applicable, with appropriate difficulty and support from teachers. Students were categorised into three cluster groups based on their attitudes towards artificial intelligence: Supporters, Sceptics and Different thinkers. As many as 63.9% of all respondents were classified as sceptics of artificial intelligence. However, familiarity with the concept of artificial intelligence correlates with the perception of competence to use it in the classroom ( $r > 0.60$ ). In conclusion, this study shows which interactive activities using ICT and artificial intelligence promote students' autonomy and active learning and contribute to a well-established knowledge of research methodology.

## Statistical Literacy

### International statistical literacy project: Past, present, and future

Irena Ograjenšek

University of Ljubljana, Ljubljana, Slovenia

irena.ograjensek@ef.uni-lj.si

The International Statistical Literacy Project (ISLP) was initiated in 1994 as the World Numeracy Program of the International Statistical Institute (ISI), with the goal of spreading quantitative skills around the world, especially in developing countries and among the young, but also among adults in all walks of life. In 2000, the ISLP was placed under the umbrella of the International Association for Statistical Education, which is the education section of the ISI. The ISLP provides an online repository of international resources and news; international activities to promote the resources and the individuals and institutions behind them; as well as outreach activities to increase awareness of how important statistical literacy is in the 21st century. To this end, among other activities, the ISLP has run several international poster competitions targeted at elementary school pupils, as well as high school and university students since 2007. In the ISLP's jubilee (30th) year of existence, like in the past, the posters submitted for the competition should reflect or illustrate usage analysis, interpretation and communication of statistics or statistical information. Following the National Competitions, the winners of the International Competition will be determined. They will be announced, and their posters displayed, at the 65th World Statistics Congress of the ISI in The Hague in October 2025. Further to that, every second year, the Best Cooperative Project Award in Statistical Literacy is presented in recognition of an outstanding, innovative and influential statistical literacy project that affects a broad segment of the general public, and is fruit of cooperation among different types of institutions. In the framework of this presentation, more details on the past, present, and future ISLP activities will be provided, and the questions if, and if yes, how, (Slovenian) statistical community could become engaged in the ISLP activities (now and/or in the future), addressed.

## Statistical literacy in early childhood

Ana Zalokar and Lara Lusa

University of Primorska, Koper, Slovenia  
ana.zalokar@famnit.upr.si

Statistical literacy for children is important, as it equips them with valuable skills that are increasingly relevant in our data-driven world. Familiarity with statistical concepts allows children to better adapt to data analysis tools and technologies as they progress through their education and future careers. In this contribution we will present how basic statistical concepts can be introduced to children using games and examples, and how children can acquire data sense through play, presenting some activities that can be used in classrooms or workshops devoted to statistics. The activities are suitable for children involved in the first three years of primary education. The goal of the activities is to cover all the stages of problem-solving circle, which includes the specification of a statistical question, data collection, data analysis and the interpretation of the results. The activities aim at teaching that data can be used to answer different types of questions, that data can be collected through surveys or measurements (that involve different types of variable), and that their summarization and description is helpful to answer the addressed question. The activities also stress the importance of the presentation and interpretation of the results. We will especially focus on topic Data collection, where children can be involved in simple data collection activities, where they can physically join in the gathering the data. Other topics as Measurement and Size, where measurement concepts like length, height, and weight are used for understanding the concepts of more and less; Observation skills, where the children should be encouraged to notice and talk about patterns, shapes, and sizes in their environment and Real-World Applications, where real-world applications of statistics in a child-friendly way can be discussed, will also be incorporated in the presentation, while they are all key aspects to consider when promoting statistical literacy among children.

# **Advancing statistical literacy among adults through interactive play**

Ana Slavec

University of Primorska, Koper, Slovenia  
ana.slavec@famnit.upr.si

In today's data-driven society, understanding statistical information is essential, as many decisions rely on it. In my contribution, I will present the fundamental principles of statistical literacy, a vital skill for interpreting and comprehending data, especially for professionals in fields like science. This year, I volunteered to design a 90-minute workshop aimed at improving participants' statistical literacy for the European Researchers' Night, an annual event focused on raising awareness of the importance of research and innovation, promoting careers in science and enhancing public engagement with researchers. To prepare, I reviewed literature on enhancing statistical literacy, with a focus on the importance of interactive and engaging learning. Based on exploring different approaches to make statistics more accessible through play, I present a concept of a game that will be used at the workshop and discuss how it addresses various aspects of statistical literacy.

## **Statistics in picture books: Finding opportunities for enhancing statistical literacy in early childhood**

Lara Lusa

University of Primorska, Koper, Slovenia

lara.lusa@famnit.uni-lj.si

As information becomes more abundant and ubiquitous, statistical literacy, the ability to understand, interpret, and critically evaluate statistical information, becomes a valuable skill even in early childhood. Early promotion of statistical literacy can be achieved by activities like counting, sorting, and comparing, introducing visual representations of data and data summaries, linking statistical concepts to everyday experiences, using games and storytelling activities with statistical concepts woven into the narratives. The significance of picture books in early childhood education is well-established, and their potential to serve as effective tools for introducing mathematical concepts is well-known. The mathematical contents can be conveyed explicitly with the objective to teach, or implicitly, either purposefully or unintentionally. Using picture books to enhance statistical literacy has received limited attention and identifying suitable high-quality picture books and providing guidance on how to use them, remains a challenging task. This contribution aims to describe the current presence of statistics in picture books and our methodology for evaluation. We centred our focus on the problem-solving process (question formulation, data, analysis and interpretation), and devised a questionnaire to assess the presence of its elements in picture books. The preliminary results are based on the analysis of more than one hundred high-quality picture books, included in the most recent lists of Children's Notable Books curated by the American Library Association. The results reveal that explicit and implicit statistical content in picture books is retrievable in less than 10% of the analysed books. Additional examples of statistical contents are identified in books not included in the target list. To support educators, we offer a curated list of books suitable for developing teaching material, complete with examples and usage guidelines. This initiative aligns with the Statistical Society of the Republic of Slovenia's commitment to enhancing statistical literacy in schools.

## Invited Lecture

# **Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1**

Hein Putter

Leiden University Medical Center, Leiden, The Netherlands  
h.putter@lumc.nl

The Fine-Gray subdistribution hazard model has become the default method to estimate the incidence of outcomes over time in the presence of competing risks. This model is attractive because it directly relates covariates to the cumulative incidence function (CIF) of the event of interest. An alternative is to combine the different cause-specific hazard functions to obtain the different CIFs. A limitation of the subdistribution hazard approach is that the sum of the cause-specific CIFs can exceed 1 (100%) for some covariate patterns. Using data on 9479 patients hospitalized with acute myocardial infarction, we estimated the cumulative incidence of both cardiovascular death and non-cardiovascular death for each patient. We found that when using subdistribution hazard models, approximately 5% of subjects had an estimated risk of 5-year all-cause death (obtained by combining the two cause-specific CIFs obtained from subdistribution hazard models) that exceeded 1. This phenomenon was avoided by using the two cause-specific hazard models. We provide a proof that the sum of predictions exceeds 1 is a fundamental problem with the Fine-Gray subdistribution hazard model. Care should be taken when using the Fine-Gray subdistribution hazard model in situations with wide risk distributions or a high cumulative incidence, and if one is interested in the risk of failure from each of the different event types.

## Biostatistics 1

### A while-alive estimator for recurrent events

Maja Pohar Perme

University of Ljubljana, Ljubljana, Slovenia

maja.pohar@mf.uni-lj.si

Chronic diseases are often characterized through the repeated occurrence of unfavorable events, for example hospitalizations, but the risk of death may also not be negligible. This implies that we are observing two processes and need to define a clinically interpretable estimand of interest. In this talk we focus on the estimand that has been referred to as  $R_2$ , discuss its interpretation and study its properties. We then turn to the estimation of  $R_2$ , propose a non-parametric estimator based on inverse probability weighting and present its properties. As the last step, we consider regression with  $R_2$  as the outcome. In the wish to directly model  $R_2$ , two general approaches may be of use—pseudo-observations or inverse probability weighting. A simulation study was performed to evaluate the two approaches in a wide range of simulation scenarios and both their properties and their problems are described in the talk.



## **Predictors with limit of detection: Modelling event-free survival in patients with PAD**

Nataša Kejžar, Aleš Blinc, Anja Boc, Vinko Boc, Mojca Božič  
Mijovski, Kevin Pelicon and Klemen Petek

University of Ljubljana, Ljubljana, Slovenia

[natasa.kejzar@mf.uni-lj.si](mailto:natasa.kejzar@mf.uni-lj.si)

Laboratory instruments for measuring biomarkers often have limits. I.e., they cannot precisely report or detect values below a certain value, calling it a limit of detection (LoD). There is no standard statistical way of treating such biomarkers as predictors in statistical analyses. We explore different possibilities and present them on data of patients with peripheral arterial disease (PAD) admitted for endovascular revascularisation of the lower limb arteries when modelling all-cause death combined with major ischemic events within one year. The aim in the data set was to estimate the prognostic value of two biomarkers: high-sensitive cardiac troponin I (hs-cTnI) and ischemia-modified albumin (IMA). The hazard for composite event in the PAD data set of 487 patients from 2020–2022 was fitted by multiple Cox proportional hazards model (PH), where population's already-known predictor variables were accounted for. Biomarkers with LoD were added as predictors to this model and the LoD issue was considered in different ways: (1) values under LoD were changed to LoD/2, (2) treating biomarker as categorical (below/above LoD), (3) recoding variable in two (numerical and dichotomous) and (4) using joint modelling (Tobit regression for estimating the biomarker and Cox PH model for estimating the outcome). In the data set, 65% of the patients had hs-cTnI values below the LoD and 23% had IMA values below the LoD. Biomarker values differed between patients with and without composite event when not adjusting for other predictors. Multiple Cox PH models showed marginal significance for both biomarkers across all different LoD treatments. Different simulation settings should be explored to determine, when simple treatment of variable with LoD (treating it as categorical or non-measurable values as LoD/2) is not sufficient.

# **Leveraging cancer incidence for lead time estimation in cancer screening programmes**

Bor Vratanar and Maja Pohar Perme

University of Ljubljana, Ljubljana, Slovenia

bor.vratana

---

In cancer screening programmes, participants are regularly screened every few years using blood tests, urine tests, or medical imaging to detect cancer at an earlier time, when it is presumed to be more curable. Without screening, cancer would likely progress undetected until symptoms appear. The interval between early detection and the eventual onset of symptoms, had screening not been conducted, is known as lead time. Understanding lead time is essential for better planning and timing of treatment interventions. Estimating lead time is challenging because it is a hypothetical random variable that can only be inferred indirectly. In our study, we introduce a novel method for estimating lead time, using a data source previously untapped for this purpose—cancer incidence. We hypothesize that earlier detection of cancer due to screening should result in observable shift in cancer incidence rates, stratified by age and year of diagnosis. Our method leverages this information, and estimates lead time using a maximum likelihood estimator. In principle, the user specifies the distribution of lead time, and the method finds the parameters that best fit the observed shift in cancer incidence. Our approach is flexible, allowing for the inclusion of additional covariates and accounting for overdiagnosis. The data required for this method are routinely available from cancer registries and provided by population tables, making it easy for implementation. We validated our method through simulations and applied it to data from the Slovenian breast cancer screening programme, demonstrating its effectiveness and utility.

# Goodness of fit for linear mixed models with multiple levels of random effects

Jakob Peterlin

University of Ljubljana, Ljubljana, Slovenia

[jakob.peterlin@mf.uni-lj.si](mailto:jakob.peterlin@mf.uni-lj.si)

Linear mixed models (LMMs) are invaluable for analyzing data from longitudinal studies, particularly in biostatistics, because they can model both the mean and covariance structures of the data. Typically, LMMs incorporate a single level of random effects, allowing data to be divided into mutually independent segments. However, more complex covariance structures may necessitate LMMs with multiple levels of random effects, complicating the assessment of the fit of an LMM. In my presentation, I will discuss the challenges of evaluating the goodness of fit for LMMs with multiple levels of random effects. I will focus primarily on the practical aspects of this analysis, introducing a package that I have developed in the Julia programming language designed to test the goodness of fit of LMMs with multiple levels of random effects. This package can also be used from R, thus bridging the gap between these programming environments. Additionally, I will share insights from the development process, particularly the difficulties encountered while adapting the package to handle LMMs with multiple crossed levels and large numbers of groups within these levels, where the assumption of data sparsity becomes less tenable. I will also briefly discuss my experience of creating a tool primarily in Julia for R users.

## Artificial Intelligence 2

### **Machine learning at scale: From logistic regression to multi-tower approaches**

Blaž Škrlj

Jožef Stefan Institute, Ljubljana, Slovenia  
blaz.skrlj@ijs.si

Online Machine Learning is a lively research field that joins highly optimized, systems-level approaches with few/single-pass learning. Performant online learning systems are the backbone of many contemporary recommender systems, enabling large-scale retail, search and advertising. In this presentation, we will overview some of the key algorithms commonly deployed in practice, from simpler linear classifiers to interaction-aware factorization machines and, more recently, Cross-network-based multi-tower architectures that excel at handling multiple modalities at once (e.g., user-item space and text space). We will conclude the presentation with an overview of a real-life example of how we scaled CPU-based Deep Field-aware Factorization Machines to hundreds of millions of predictions per second.

# **Enhancing clinical study analyses through federated learning: A systematic review using generative pretrained transformer and large language models**

Jiregna Olani Kedida, Dario Gregori, Corrado Lanera and  
Honorina Ocagli

University of Padua, Padua, Italy

jiregnaolani.kedida@studenti.unipd.it

**Background:** Federated Learning (FL) offers an innovative approach to clinical research, enabling detailed data analysis from a multicenter while ensuring patient data privacy. Evaluating the federability of algorithms, especially those profoundly and traditionally linked to biomedical research, poses critical challenges. **Objective:** With the support of OpenAI's GPT-4o large language model (LLM), this systematic review (SR) aims to analyze the latest advancements in FL methodologies for clinical studies. **Material and methods:** Embase, PubMed, Scopus, Web of Science, and ArXiv were accessed to retrieve the articles to screen. Each database's data is processed to extract individual citation records using specific formatting patterns. We used GPT-4o with designed query prompts to automate and refine data extraction, synthesis, and interpretation, making systematic reviews more efficient and accurate. Metrics such as sensitivity, specificity, positive and negative predictive values (PPV and NPV), workload saving, and others will be employed to evaluate the performances of GPT-4o in a sample of records. **Results:** We gathered data from the major medical databases: Embase(n=71), PubMed (n=78), Scopus (n=179), and Web of Science (n=179), compiling a comprehensive dataset of 507 records. The results after title/abstract screening will be evaluated for full-text screening. All the retrieved articles will be processed by GPT-4o and a statistician. The following information will be extracted: different algorithms and methodologies employed in FL, their applications across various domains, the performance metrics, data privacy concerns, communication overhead, and computational efficiency. **Discussion:** FL significantly enhances data security and analytical capabilities, effectively addressing key challenges of traditional centralized data analysis. Overall, FL integration in clinical research promises improved accuracy and efficiency, leading to better research outcomes and patient care, while ensuring data confidentiality. **Keywords:** Federated Learning, Clinical Studies, Systematic Review, GPT, Large Language Models, Data Privacy, Methodological Advances

# **Stochastic blockmodels of the dynamic global-trade network: A triangulation study**

Fabio Ashtar Telarico and Aleš Žiberna

University of Ljubljana, Ljubljana, Slovenia  
telaricof@fdv.uni-lj.si

Relational data with a prominent temporal dimension are providing researchers with new ways to model network evolution, particularly through \*stochastic blockmodeling\* (SBM). However, inferring patterns in dynamic networks using SBM presents challenges, especially due to the different implementations and conceptualizations that the literature offers and which have been applied mostly to binary networks. Not least because many networks in social and economic sciences, such as the global trade network (GTN), are inherently valued/weighted. Existing SBMs were designed to model networks like food webs and interpersonal relationships, which exhibit features like transitivity and the rich-get-richer effect, but the GTN operates under different principles. The theoretical assumptions and mechanisms driving GTN's evolution raise questions about the effectiveness of current SBMs in modeling its changes over time, as edges' incidence and their value may follow different processes. To help fill the gap, this study compares various implementations of SBMs for dynamic networks on the GTN using data from the UNCTAD database. The SBMs are evaluated based on two competing views of international trade: New Trade Theory, which predicts trade concentration between similarly productive countries, and World-System Analysis/Dependency Theory, which highlights the unequal nature of international trade. By examining the aggregated trade network and disaggregated trade flows, the study aims to identify whether SBMs can capture the clustering of units based on local tie-generating mechanisms or global mechanisms underlying the centre-periphery structure of unequal exchange. The expectation is that different tie-weighting methods will significantly influence the clustering outcomes, revealing the complex nature of modeling the GTN with SBMs.

## **Generative artificial intelligence: A new frontier in academic writing**

Sara Blatnik, Hana Brezar, Lovrenc Cvijič and Andrej Kastrin

University of Ljubljana, Ljubljana, Slovenia

andrej.kastrin@mf.uni-lj.si

Recent breakthroughs in generative artificial intelligence (AI) have spurred the adoption of new technologies in academic writing. We conducted a systematic review of scientific publications, focusing on the use of generative AI tools in the preparation of academic papers in biomedicine. Our analysis encompassed 35 original research and review articles, leading us to identify three groups of AI tools that, when used judiciously, (i) streamline the retrieval of relevant scientific literature, (ii) facilitate the drafting, creation, and completion of scientific texts, and (iii) enable the rapid extraction and summarization of data and information from scientific papers.

## Biostatistics 2

### **Ranking of Olympic-style weightlifting performances adjusted for body mass and sex**

Marianne Huebner

Michigan State University, East Lansing, MI, United States

huebner@msu.edu

Men typically outperform women in Olympic-style weightlifting, and heavier athletes tend to outperform those that are lighter. While there have been efforts to standardize weightlifting performances across body mass to compare athletes, existing models do not address the sex gap. Utilizing results from the International Weightlifting Federation (IWF) world championships and Olympics 2017-2023 a unified scale for male and female athletes is established across the range of body masses. The resulting scaled scores show stability over time in IWF competition results from 2000 to 2023, ensuring the validity of the model for ranking performances of mixed sex teams into the future. This is joint work with Tim Cole, UK.



## Visualizing a marker's degrees of necessity and of sufficiency in the predictiveness curve

Andreas Gleiss

Medical University of Vienna, Vienna, Austria  
andreas.gleiss@meduniwien.ac.at

A cause, represented by certain values of prognostic factors, is considered necessary for an event  $D$  if, without the cause,  $D$  cannot develop. It is considered sufficient for  $D$  if it is unavoidable in the presence of the cause. The degree to which a factor is necessary (DN) or sufficient (DS) for  $D$  ranges from zero to one. These have been proposed as generalizations of attributable risk based on simple functions of unconditional and conditional event probabilities (Gleiss & Schemper, StatMed, 2019). Predictiveness curves have been proposed as graphical means to integrate the predictiveness of a marker with its performance as a classifier (Pepe et al, AmJEp, 2008). They show the risk for an event, as derived by a model with one or more predictors, depending on risk percentiles that represent the predictors' distribution in the underlying population. At a certain risk percentile ( $p_0$ ) the predictiveness curve crosses the horizontal line that represents the estimate of the event's prevalence  $P(D)$  and thus unconditional risk prediction without any predictors. We show that DN equals the area AN between  $P(D)$  and the predictiveness curve for percentiles below  $p_0$  relative to the total area below  $P(D)$ , which is equal to  $P(D) \times p_0$ . On the other hand, DS is equal to the area AS between  $P(D)$  and the predictiveness curve for percentiles above  $p_0$  relative to the total area above  $P(D)$ , which is equal to  $(1 - P(D)) \times (1 - p_0)$ . In addition, we show that the proportion of explained variation, a common measure of predictiveness and relative importance of prognostic factors, is also closely connected to the graphical areas AN and AS. Using a prostate cancer dataset, we demonstrate that the predictiveness curve extended by these new interpretations of areas provides a comprehensive evaluation of markers or sets of markers for prediction.

## **Regression modelling in an extended multi-state model using relative survival**

Damjan Manevski, Hein Putter, Eva A. S. Koster, Johanneke E. A. Portielje, Maja Pohar Perme and Liesbeth C. De Wreede

University of Ljubljana, Ljubljana, Slovenia

damjan.manevski@mf.uni-lj.si

In survival analysis, multi-state models have become a staple tool for considering intermediate events, apart from death as the main event of interest. As for many diseases the number of deaths not related to the disease in question is non-negligible, our interest lies in distinguishing between disease-specific (excess) and other (population) mortality occurring after intermediate events. The field of relative survival deals with this distinction when the cause of death information is not provided. In this work, we are interested in applying the relative survival approach in the multi-state model framework for estimating covariate effects and calculating subject-specific predictions. In a recent paper<sup>1</sup>, we have proposed an extended multi-state model using relative survival in which suitable definitions of transition hazards and probabilities are introduced as well as their non-parametric estimators. In this work, we explore regression modelling in the extended model. We expand the Cox-type multiplicative model which is commonly used in multi-state models. An existing implementation for modelling the disease-specific hazards based on the EM algorithm<sup>2</sup> is upgraded for multi-state models where delayed entry has to be considered in the intermediate states. However, as issues arise in the case when the estimated overall mortality is smaller than the background mortality, we consider an alternative - the Aalen additive hazards model which can deal with this issue. A simulation study compares the two approaches in various data settings, especially when considering delicate factors such as delayed entry and small excess death rates. The usage of these models is illustrated using Dutch breast cancer data. The R code needed for applying the implemented work is shown using the packages `mstate` and `relsurv`.

# Mapping the landscape of physiotherapy research: A scientometric study

Alja Nike Kastrin and Janez Žibert

University of Ljubljana, Ljubljana, Slovenia

nike.kastrin@gmail.com

This study presents a comprehensive scientometric analysis of research publications in the field of physiotherapy, aiming to map the landscape of scholarly activity over the last decades. We used information from Web of Science, Scopus, and MEDLINE/PubMed to create a corpus of 18493 bibliographic records in the field of physiotherapy. We then examined key metrics such as publication trends, citation patterns, authorship networks, and the geographical distribution of research output. The findings reveal significant growth in physiotherapy research, with a notable increase in collaboration among authors across different countries. The analysis identifies core research areas, influential authors, and high-impact journals that have shaped the field. The most frequently discussed research topics focus on low back pain, stroke, assessment in physiotherapy, and evidence-based practice. This work provides a comprehensive overview of the development and current state of physiotherapy research, which is critical for future research. Further work should include articles from the PEDro database. According to the available research evidence, the present work is the first comprehensive and methodologically complex bibliometric review of publications in the field of physiotherapy.

## Official Statistics

### Nowcasting monthly indicators in official statistics

Aleša Lotrič Dolinar, Mojca Bavdaž, Tjaša Bartolj, Jerneja Kos,  
Barbara Mörec, Darja Peljhan, Sašo Polanec, Tjaša Redek and  
Jelena Zorić

University of Ljubljana, Ljubljana, Slovenia

alesa.lotric.dolinar@ef.uni-lj.si

Monthly official indicators are typically reported with lags of up to two months, which often proves to be late, especially in turbulent economy environment. How can reliable estimates be provided at a shorter lag, i.e. nowcasted? We present a brief literature overview of nowcasting methodology, and current picture regarding nowcasting implementation across European national statistical institutes (and beyond). Besides, taking availability of the relevant data sources into account, we suggest methodological approach to be used to produce timelier, but still accurate estimates of the total production index of the market economy, published monthly by SURS.

## **Daily tourism flow in Slovenia and Slovenia Interactively: The usage of PowerBI for almost real-time data visualizations**

Teja Zupančič and Igor Štefelin

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia  
teja.zupancic@gov.si

This year the Statistical Office of the Republic of Slovenia (SURS) launched a new interactive tool called »Tourism Daily«, which shows daily data on tourist arrivals and overnight stays in accommodation establishments in Slovenia with only a three-day delay. With increasingly turbulent and rapidly changing economy environment, decision-makers seek most up-to-date and real time data to be able to actively react to changes in tourism flows and to adjust planning in time. The presented product supplements SURS's regular monthly statistics on tourism flows in Slovenia and enables new possibilities for monitoring and better planning of tourism flows in Slovenia as it shows the data in almost real time. The data are shown in Microsoft Power BI, which SURS defined as the most suitable tool for displaying data in a new, different, and interactive way. In Addition, Power BI uses Microsoft AZURE data cloud, which is defined as one of the safest data clouds (enables safe uploads of unpublished data). The data visualization shows tourist arrivals and overnight stays by country of guests' origin, by 12 statistical regions and 47 municipalities (only those that meet the principle of statistical confidentiality) and by types of accommodation establishments from two administrative sources, the Register of Accommodation Establishments and eTurizem. These two data sources enable full data coverage and are also key sources for conducting the regular monthly accommodation statistics. Another application we developed in Power BI is Slovenia Interactively, which presents the most important current economic, social, and environmental data about Slovenia through interactive charts on a dashboard. The application pulls data via API calls and transforms multidimensional cubes using Power Query into a suitable format to present in interactive visualizations (graphs).

# **Data quality in official statistics and in the Statistical Office of the Republic of Slovenia**

Domen Repnik

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

domen.repnik@gov.si

For every producer of statistics it is important that users trust the data and the products that they produce and disseminate. This topic is especially important for the producers of official statistics. In this presentation, we will introduce quality reporting in official statistics and in the Statistical Office of the Republic of Slovenia (SURS). The first part will describe the beginnings of the quality assurance system in official European statistics. We will take a closer look at how they were developing and at their origins in other organisations. We will compare principles from that period, how they developed and how new principles were added. The next part will talk about quality assurance at SURS. We will introduce the legal framework and how it is connected with general principles of quality management, the European Statistics Code of Practice, the Fundamental Principles of Official Statistics, and basic principles. The last part will focus on data quality and quality reporting at SURS from the practical point of view. In the past the quality of statistical data used to be dealt with mostly in connection with data accuracy in the narrow sense. Recently, quality has been more oriented towards different quality dimensions: relevance, accuracy of estimates, timeliness and punctuality of publication, accessibility and clarity of information, comparability of statistics and coherence of results. We will take a closer look at what each covers, where we can find them and how they developed through time, future plans and why they are important, especially in the time of crisis.

## **Enhancing statistical data collection through respondent-centric approaches**

Ema Mišić

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia  
ema.misic@gov.si

Statistical data collection is a fundamental process that involves gathering, organizing, and analysing numerical information to describe and understand various phenomena. Equally crucial is the concern for data providers, to provide them with user-friendly, secure ways for submitting data and to ensure reasonable burden. The Slovenian Statistical Office (SURS) faces the increasing challenge of collecting accurate and timely statistical data, particularly from individuals and households. This challenge is exacerbated by societal changes, technological advancements, digitalisation, and economic crises. To efficiently address these issues, SURS has started with implementation of various strategies: (i) Efficient data collection methods: SURS has adopted modern technologies to streamline data collection processes. Mix-mode data collection, online surveys, data linking, web scraping, and the utilization of administrative data are some of the solutions used. (ii) Respondent burden assessment and mitigation: SURS aims to minimize respondent burden by pre-testing questionnaires and implementing strategies like shorter or pre-filled questionnaires, simplified question formats, and clear instructions. (iii) Enhanced communication and transparency: SURS uses tailored invitation letters, plain language, and visual aids to improve understanding and participation. In addition, SURS has established a dedicated help desk to provide assistance to respondents and businesses with any questions or concerns related to data collection. (iv) Building partnerships and collaboration: SURS establishes partnerships and cooperation with various stakeholders for the exchange of administrative data sources, greater awareness and support for the establishment of databases for multiple purposes, thereby preventing duplication of data collection. (v) New ways and approaches: SURS explores different ways to encourage participation, such as offering rewards, piloting smart surveys, and profiling respondents. NEWLINE While these strategies have shown positive results, SURS recognizes the need for continuous improvement to adapt to evolving challenges and maintain effective data collection. Future plans include refining data collection methods, expanding outreach initiatives, and exploring innovative technologies.

## Campaign on the 80th anniversary of the Statistical Office of the Republic of Slovenia

Irena Svetin and Anže Jesenovec

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

irena.svetin@gov.si

The Statistical Office is celebrating its 80th anniversary this year, and on this occasion, we have prepared several activities. The occasional publication was printed in a smaller edition besides being available in digital form on the website. For promotional material, we made socks in a retro style, and a diary that is equipped with infographics and thus a useful gift that is also informative at the same time. The third part of the promotional material was a candy box. Promotional material was distributed in gift bags specially made for this purpose and equipped with thank-you notes. The central event of the anniversary was an open-air exhibition of infographics, a set of 22 posters on which we presented some of the stories written in the language of official statistics about Slovenia and the lives of its residents. The conceptual starting point for the design of the visual identity of the campaign consists of two parts. On the one hand, it reflects the reason for the event itself, i.e. the anniversary or the celebration, while the other part communicates the primary theme of the statistical office, i.e. the data, or in this case, their visualization. The basic building block of visible messages thus became the auxiliary grid—an element that is usually pushed into the background in visualizations, but here we deliberately enhanced it and used it as one of the main connecting elements. On this grid, coloured circles appear on the promotional materials, symbolizing the dots in a scatterplot, and at the same time, with their colourfulness and overlap, they are associated with paper confetti from birthday parties, thus linking to the previously mentioned starting point of the celebration.



## Statistical Applications

### Enabling topic-modeling for specific domains via domain-adaptation of LLMs

Boshko Koloski and Senja Pollak

Jožef Stefan Institute, Ljubljana, Slovenia

boshko.koloski@ijs.si

In the context of rapidly increasing scientific publications, researchers face challenges in staying updated with advancements in their respective fields. This paper addresses these challenges by presenting advancements in topic modeling using state-of-the-art language models. The AHAM methodology is introduced, enhancing the BERTopic framework for domain-specific adaptation to improve scientific text analysis. Leveraging the LLaMa2 model, the approach incorporates one-shot learning, with domain experts assisting in crafting prompts to generate topic definitions. The methodology employs language generation and translation scores to assess inter-topic similarity, aiming to reduce outliers and overlap between topic definitions. Validated on a corpus of scientific papers, AHAM has proven effective in revealing novel insights across various research areas, demonstrating its utility in domains like computational creativity and hate-speech literature.

# **Statistical learning and music perception: Analyzing children's songs through the cognitive computational model IDyOM**

Lorena Mihelač

Faculty of Information Studies, Novo mesto, Slovenia

lorena.mihelac@sc-nm.si

The presentation explores the complexities involved in modeling human perception of music through computational methods, raising several fundamental questions. It examines whether a computer can accurately model human music perception, and if so, to what extent it can achieve precision and similarity to human cognition. Additionally, the presentation considers whether such models offer distinct advantages and whether they can address certain challenges that traditional methods cannot resolve. The presentation is organized around three primary topics: (i) the investigation of (ir)regularities in the musical structure of children's songs, (ii) the examination of music segmentation in children across different age groups, and (iii) the analysis of (dis)similarities in children's songs both between and within 22 European countries. A key focus is on implementing the cognitive computational model IDyOM, which employs statistical learning and probabilistic prediction to acquire and process internal representations of musical pieces and styles. This model shows how computational approaches can provide new insights into the study of music perception, particularly in the context of children's music.

## On estimating standard deviation from average moving range

Gaj Vidmar

University Rehabilitation Institute Republic of Slovenia, Ljubljana, Slovenia

gaj.vidmar@signir-rs.si

The estimation of standard deviation from average moving range (AMR) is fundamental to the field of statistical process control (SPC), but it is scarcely known and never applied in other fields of statistics. The AMR approach is first briefly theoretically reviewed, including its relation to the Studentised range distribution. Next, some interactive spreadsheet simulations are presented, which are of substantial as well as pedagogical interest. With normally distributed data, the AMR approach outperforms the usual (a.k.a. corrected) estimate (i.e., square root of unbiased variance estimator) in very small samples (of size up to 10). With uniformly distributed data, the AMR approach is superior only in samples of size up to 5, and increasingly inferior with larger samples. This is also the case with a bathtub-shaped (e.g., arcsine) distribution, whereby the change occurs at sample size 4. With highly skewed data, the AMR approach outperforms the usual estimate with sample size 2 to 4, while from 5 onwards they are practically identical. As expected, the approximate unbiased estimate (i.e., using division by  $n - 1.5$  instead of  $n - 1$ ) performs better than the usual one from sample size 4 onwards, and becomes practically identical with  $n > 10$ . The simulations also highlight the less-known fact that the AMR-based estimate is preferable with grouped data in the sense of a process with an oscillating mean, which yields a high positive autocorrelation. In that case, the AMR-based estimate is close to the within-group mean-square as defined in ANOVA and, unlike the usual estimate (which is much larger), leads to sensible control limits in the control chart for individual measurements (I chart). Conversely, with extreme negative autocorrelation, the AMR-based estimate is much larger than the usual one, which leads to even wider control limits on the I chart.

## Poster Session

### Developing a prediction model to estimate the number of slaughtered bovine animals

Nikolina Rizanovska

Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia  
nikolina.rizanovska@gov.si

The Central Register of Bovine Animals (slo: CRG) contains information about every bovine animal in Slovenia. Therefore, it is used as the primary administrative source for calculating many statistics. However, one of the statistics - purpose of breeding (whether an animal is intended for slaughter or for breeding) should be reported in advance. This statistics is currently required by the current Livestock and meat regulation 1165/2008 and from 2025 on will be required by SAIO legislation (Regulation (EU) 2022/2379). Until now, we have been gathering this information from the CRG in a combination with the December livestock survey (telephone data collection – CATI). The exact number of the slaughtered animals and animals for breeding is known seven months after the report day. The general objective of the project was to develop a prediction model for estimating the number of bovine animals for slaughter/not for slaughter solely by using administrative data sources (CRG). Our aim is to reduce the burden for the respondents since the CATI collection method would no longer be needed for cattle-related data. Using the survival analyses, we formed a more detailed picture about the bovine animals registered at the CRG. Furthermore, we created eight different models and compared those using validation (Nagelkerke's determination coefficient -  $R^2$  and AU) and calibration (intercept and slope) parameters. In the final model, we used logistic regression and included five explanatory variables (age, breed, sex, farm type, cohesion region) that were conceptually important for determining the number of animals for slaughter and not for slaughter. The model was trained on the data from the previous year. Data obtained from the models are more accurate than the current method and for this reason our plan is that when the new SAIO legislation enters into force, we will use only administrative sources and models.

## **Impacts of assumption violations on accuracy in test equating: Insights from a simulation study**

Michaela Vařejková, Patrícia Martinková and Eva Potužníková

Institute of Computer Science, Czech Academy of Sciences and Faculty of Mathematics and Physics,  
Charles University, Prague, Czechia

varejkova@cs.cas.cz

In educational testing, comparing individuals' test scores often requires test-equating methods, especially when different test forms are used for different groups. Traditional approaches use an anchor test, where a common set of items is administered to all groups. When this is not feasible, alternative methods, such as the non-equivalent group design with covariates (NEC design), must be applied. This involves using covariates like age, school grades, or other test scores, which are highly correlated with the primary test scores, to adjust for differences in ability between groups. The NEC design relies on the assumption that the conditional distribution of test scores, given the covariates, is the same across groups. However, this assumption may be violated if the covariates' relationship with test scores varies between groups. One situation where this problem can occur is if one of the covariates is itself measured by multiple test forms. Our study investigates the impact of such violations on the accuracy of resulting equated scores. We conducted a simulation study with two non-equivalent groups differing in ability, using covariates including school type, educational status (as binary variables), and another test score (continuous variable). The simulations reflected real-world scenarios from the Czech Republic's national matura exam. The results showed minimal bias in equated scores when the assumption of the same conditional distribution was held, especially with large sample sizes. However, notable bias emerged when the assumption was violated, particularly due to different test difficulties for continuous covariate. To address this bias, we proposed an additional equating step: equating the covariate itself before incorporating it into the primary equating process. This adjustment aimed to make the covariate scores comparable across groups, increasing the odds of satisfying the assumption of the same conditional distribution. Results from our simulation show the suggested repeated covariate-equating approach improves equating score accuracy.

## Forecasting lung cancer incidence using an age-period-cohort model

Maja Jurtela, Tina Žagar and Vesna Zadnik

Institute of Oncology, Ljubljana, Slovenia

mjurtela@onko-i.si

The age-period-cohort (APC) model is a generalised linear model that estimates the effects of aging, time period, and birth cohort on disease incidence. Because the three components are completely linearly dependent, we apply constraints that allow us to quantify each effect separately. We used the APC model to forecast lung cancer rates in Slovenia for the next 20 years, stratified by age and gender. Cancer incidence data were sourced from the Slovenian Cancer Registry, spanning the period from 1961 to 2020. Historical data on smoking prevalence, a principal risk factor responsible for 80-85% of lung cancer cases, came from nine Slovene public opinion surveys conducted by the Public Opinion and Mass Communication Research Centre from 1978 to 2000 and from six CINDI studies conducted by the National Institute of Public Health from 2001 to 2020. The data were grouped into 5-year categories across age, period and cohort effects, separated by gender. In our APC model, we used the Poisson distribution to account for the count data of cancer cases, normalised by population exposure using offsets, and we introduced drift to address the inherent non-identifiability issue. Cross-validation was performed to validate the predictive accuracy of the model, and the best-fitting model was selected based on the lowest AIC/BIC criteria and used to forecast future incidence. We quantified the relationship between smoking prevalence, interpolated by cubic spline, and lung cancer incidence for 5-year periods using a linear regression model while taking into account a 20-year delay until lung cancer development. Future incidence was multiplicatively adjusted based on the resulting estimates. We present the results and discuss the challenges related to model fitting and interpretation.

## Paired t-test and mixed models

Urška Komatar and Tinkara Perme

University of Ljubljana, Ljubljana, Slovenia  
urskak2000@gmail.com

In our study, we explore the equivalence and comparative performance of a paired t-test and linear mixed models (LMMs) in statistical analysis. While the equivalence holds under the assumption of balanced, paired data with complete cases, their performance differs in scenarios with highly correlated data and missing values. A linear mixed model, a paired t-test, and an independent t-test were compared based on their assumptions, test size, and power. In order to test different examples, we generated data with varying sample sizes and different percentages of missing data, introducing incomplete cases, to highlight the advantages of LMMs over paired t-tests in handling missing data. While the paired t-test requires complete cases, LMMs can incorporate all available data, maintaining higher statistical power and making it a preferred choice for data with a considerable number of missing values. Another impact considered in the study is the correlation between paired measurements. Specifically, in cases with high correlation (e.g.,  $\rho = 0.8$ ), the LMM and a paired t-test demonstrate significantly higher power compared to the independent t-test. In the extreme scenario where one value is deleted from each pair, we are left with two independent samples, leading to equivalent performance between the linear mixed model and the independent t-test. Based on our conclusions, we propose a user's guide for determining when each of the tests in question is the most appropriate or equivalent to the other. This guide takes into account the level of correlation, the presence of missing data, and the sample size, providing a practical framework for researchers to choose the most suitable statistical test for their data.

## Comparing tests for normality: A simulation study

Lan Gerdej and Tanja Gošnjak

University of Ljubljana, Ljubljana, Slovenia

lg12410@student.uni-lj.si, lan.gerdej@gmail.com

In our paper, we explore the application and properties of the Kolmogorov-Smirnov (KS) test for assessing normality, highlighting its limitations when parameters are not predefined. Additionally, we examine and compare alternatives that don't require predefined parameters for testing normality, such as the Lilliefors test and the Shapiro-Wilks (S-W) test. The KS test is a non-parametric method for the equality of continuous, one-dimensional probability distributions. It is used to determine whether a sample originates from a population with a specific distribution, making it unsuitable for normality tests without known parameters. In our further analysis, we compare the Lilliefors test—an adaptation of the KS test—with the S-W test, both specifically tailored for testing normality. Unlike the KS test, the Lilliefors test does not require predefined distribution parameters under the  $H_0$ ; it estimates these parameters from the sample and assumes them under  $H_0$ . This test effectively fits the sample's empirical cumulative distribution function to the theoretical distribution under  $H_0$ , with adjusted critical values that reject the null hypothesis at lower deviations. Similarly, the S-W test also doesn't require predefined distribution parameters but, while evaluating the same null hypothesis, employs a more computationally demanding method. Based on simulations, both the Lilliefors and S-W tests exhibit similar properties regarding test size for samples smaller than 1000. The power comparisons indicate that the S-W test virtually always has greater power than the Lilliefors test. Therefore, we strongly recommend using the S-W test for assessing normality due to its higher power. The difference in power is less pronounced in very small samples, where both tests have low power, as well as in very large samples and distributions that deviate significantly from normality, where both tests have high power.



## Workshop

### **Time-dependent effects and time-dependent covariates in survival analysis**

Hein Putter

Leiden University Medical Center, Leiden, The Netherlands  
h.putter@lumc.nl

The Cox model is the dominant model to quantify the effect of covariates on time-to-event outcomes. It operates under the proportional hazards assumption, which acknowledges the time-dependent behaviour of the hazard, but assumes that the ratio of two hazards for different values of the predictor is constant over time. We will discuss mechanisms that could cause violation of the proportional hazards assumption, leading to time-dependent effects of the predictor, and discuss ways of dealing with possible violations. Seemingly similar, but quite different, is the situation where the value of the predictor can change over time. We will discuss commonly made mistakes in this setting, like immortal time bias, and correct ways of dealing with this situation. The workshop will include illustrations in R, and opportunities for the participants to try out some of the proposed methods in R or another statistical package of their choice.

## ***INDEX***

# Index

## A

A. S. Koster, E, 57  
Abate, N, 25  
Ashtar Telaarico, F, 53

## B

Baldi, I, 28  
Bartolj, T, 59  
Batagelj, V, 19  
Bavdaž, M, 59  
Bertolini, M, 24  
Bitan, M, 35  
Blatnik, S, 54  
Blinc, A, 48  
Boc, A, 48  
Boc, V, 48  
Božič Mijovski, M, 48  
Brabec, M, 21, 38  
Brezar, H, 54  
Bucik, V, 34

## C

C. De Wreede, L, 57  
Caramuta, C, 26  
Cichrová, M, 39  
Cugmas, M, 20  
Cvijič, L, 54

## D

Dimai, M, 38  
Duttilo, P, 24

## E

E. A. Portielje, J, 57  
Erčulj, V, 29, 32

## F

Ferligoj, A, 34

## G

Gerdej, L, 71  
Gleiss, A, 56  
Gorfine, M, 35  
Gošnjak, T, 71  
Gregori, D, 27, 28, 52  
Grosek, Š, 32  
Grosso, A, 26

## H

Hladká, A, 39  
Huebner, M, 55

## J

Jesenovec, A, 63  
Junayed Bhuyan, M, 28  
Jurtela, M, 69

## K

Kastrin, A, 54  
Kastrin, AN, 58  
Kejžar, N, 48  
Knez, T, 40  
Kogovšek, T, 34  
Koloski, B, 64  
Komatar, U, 70  
Kos, J, 59

## L

Lajovic, J, 32  
Lanera, C, 27, 28, 52  
Lasaponara, R, 25  
Lekše, R, 33  
Lisi, F, 24  
Longo, G, 26  
Lotrič Dolinar, A, 59  
Lovallo, M, 25  
Lusa, L, 43, 45

## **M**

M. Steinberg, D, 35  
Manevski, D, 57  
Martinková, P, 39, 68  
Mihelač, L, 65  
Miklavčič, J, 32  
Millo, G, 36  
Mišić, E, 62  
Mörec, B, 59

## **O**

Ocagli, H, 52  
Ograjenšek, I, 42  
Olani Kedida, J, 27, 52

## **P**

Pelicon, K, 48  
Peljhan, D, 59  
Perme, T, 70  
Petek, K, 48  
Peterlin, J, 50  
Pohar Perme, M, 47, 49, 57  
Polanec, S, 59  
Pollak, S, 64  
Potužniková, E, 68  
Prezelj, T, 40  
Pulickal Thomas, S, 27  
Putter, H, 46, 57, 72

## **R**

Redek, T, 59  
Repnik, D, 61  
Rizanovska, N, 67  
Rosen, L, 35

## **S**

Slavec, A, 30, 33, 44

Snijders, T, 18

Sočan, G, 37  
Srakar, A, 22  
Svetin, I, 63

## **Š**

Škrlep, E, 32  
Škrlj, B, 51  
Škulj, D, 23  
Štefelin, I, 60  
Štivić, S, 32  
Šulc, A, 29

## **T**

Telesca, L, 25

## **V**

Vařejková, M, 68  
Vedovelli, L, 28  
Vidmar, G, 66  
Vipavc Brvar, I, 31  
Vratanar, B, 49

## **Z**

Zadnik, V, 69  
Zalokar, A, 43  
Zorić, J, 59  
Zupančič, T, 60  
Zurc, J, 34, 41

## **Ž**

Žagar, T, 69  
Žiberna, A, 53  
Žibert, J, 58  
Žitnik, S, 40

