

From the Editor

Presented to our readers – as the last in this year – December issue of Statistics in Transition new series, contains a set of twelve articles written by thirty-four authors from nine countries (in the order of appearance): USA, Nigeria, Thailand, India, Nepal, Morocco, Iran, Poland, and Ukraine.

The geographical diversity of this issue is accompanied by its thematic richness. From the volume's opening article – an invited paper – demonstrating practical solutions to the increasingly common problem of choosing between probability and nonprobability sampling, through selected issues of parameter estimation and distribution analysis in the subsequent papers, to specific approaches applied in different studies, including those based on conference presentations.

Taking advantage of the opportunity that this is the last issue of this year's issue, I would like to personally, and on behalf of the entire editorial team and our advisory bodies – the Editorial Board and Associate Editors – express deep gratitude to all our stakeholders – authors and readers, but above all, peer-reviewers. The knowledge and commitment of the reviewers helped our authors improve their work while contributing significantly to our efforts to ensure that articles published in SiTns meet the highest quality standards; hence the list of their names is included in the Acknowledgments at the end of this issue.

Invited paper

The opening text in this volume, as the invited paper, by **Terrance D. Savitsky**, **Matthew R. Williams**, **Julie Gershunskaya**, and **Vladislav Beresovsky**, entitled *Methods for combining probability and nonprobability samples under unknown overlaps* presents innovative perspective to this complex issue, including a novel approach that derives the propensity score for the observed sample as a function of inclusion probabilities for the reference and convenience samples as the main result. The presented approach allows specification of a likelihood directly for the observed sample as opposed to the approximate or pseudo likelihood. The authors construct a Bayesian hierarchical formulation that simultaneously estimates sample propensity scores and the convenience sample inclusion probabilities. A Monte Carlo simulation study to compare the likelihood based results with the pseudo likelihood based approaches considered in the literature was used.

Research articles

Umar K. Abdullahi, Fidelis I. Ugwuowo, and Nuanpan Lawson in the article *Power ratio cum median-based ratio estimator of finite population mean with known population median* propose the power ratio cum median-based ratio estimator of the finite population mean, which is a function of two ratio estimators in the form of an average. The estimator assumes the population to be homogeneous and skewed, while the properties (i.e. the bias and the Mean Squared Error – MSE) were derived alongside its asymptotically optimum MSE. The efficiency of the developed estimator jointly with its efficiency conditions by comparing it to selected estimators described in the literature were demonstrated. Empirically, a real-life data set from the literature and a simulation study from two skewed distributions (Gamma and Weibull) were used to examine the efficiency gain. Results from both the real-life dataset and the simulation study show efficiency gain for the proposed estimator that incorporates median of study variable (while for the other estimators the result show efficiency loss).

The paper by **Pitambar Das, Garib Nath Singh, and Arnab Bandyopadhyay**, *Ratio estimation of two population means in two-phase stratified random sampling under a scrambled response situation*, describes the development of an effective two-phase stratified random sampling estimation procedure in a scrambled response situation. Two different exponential, regression-type estimators were formed separately for different structures of two-phase stratified sampling schemes. The authors have studied the properties of the suggested strategy. The performance of the proposed strategy has been demonstrated through numerical evidence based on a data set of a natural population and a population generated through simulation studies. Taking into consideration the encouraging findings, suitable recommendations for survey statisticians are prepared for the application of the proposed strategy in real-life conditions.

The next paper, by **Oumaima Oullada, Mohamed Ben Ali, Ahmed Adri, and Said Rifai**, entitled *Model for measuring the impact of good pharmacovigilance practices of COVID-19 patients on hcp reactivity: Morocco case study* presents a conceptual model used to evaluate how the improvement of good pharmacovigilance practices of COVID-19 patients influences the reactivity of the healthcare personnel (HCP) in the Draa Tafilalet region in Morocco. The empirical study is based on a survey submitted to a sample of a total of 180 HCP and on the application of latent variable structural modelling through the partial least squares (PLS) method focusing on the reliability and validity of the proposed model. The study shows that the improvement of good pharmacovigilance practices impacts positively the reactivity of HCP in terms of adverse drug reactions (ADRs) reporting. The reliability of the measurement was > 0.7 , which allowed testing the internal and external validity of the conceptual model; several hypotheses were validated against two invalid derivative hypotheses.

Zahra Khoshkhoo Amiri's and **S.M.T.K. MirMostafaei's** article, *Analysis for the xgamma distribution based on record values and inter-record times with application to prediction of rainfall and COVID-19 records*, discusses the problem of classical and Bayesian estimation of the unknown parameter of the xgamma distribution based on record values and inter-record times. The problem of Bayesian prediction of future record values based on record values and interrecord times was also discussed. A new lifetime distribution, called "xgamma distribution", which can be used as an alternative to other lifetime distributions, like the exponential one, was introduced. A simulation study has been performed to compare the performance of the proposed estimators and the approximate Bayes predictors, complemented by two real data sets related to rainfall and COVID-19 records.

Magdalena Szymkowiak, Anasuya Roychowdhury, Satya Kr. Misra, Rajib Lochan Giri, and Subarna Bhattacharjee present the paper *A study of a survival data using kernel estimates of hazard rate and aging intensity functions*. The authors primarily focus on Aging Intensity (AI) and Hazard Rate (HR) functions estimated using four different kernels. They apply them to a case study of patients with primary malignant tumors of sternum with the right-censored data. It turned out that kernel estimates of HR and AI functions for patients with high grade tumor (HGT) are higher than for patients with low grade tumor (LGT), as expected. The authors believe that their study opens up a new direction for applying AI and HR functions in analyzing health and engineering related problems.

In the paper *Investigating the factors of blockchain technology influencing food retail supply chain management: a study using TOE framework*, **Manish Mohan Baral, Venkataiah Chittipaka, Surya Kant Pal, Subhodeep Mukherjee, and Hari Shankar Shyam** discuss the factors affecting blockchain adoption in the food retail supply chain and create awareness among retail managers for its adoption in their operations. A structured literature review was conducted to identify the TOE factors used in the research. TOE factors were used in many previous studies on technology adoption, like RFID, IoT, cloud computing, intelligent agent technology, and many more. With these factors, a questionnaire was developed for the survey. The questionnaires were sent to retail stores across India through online mode. The results were analysed using EFA and SEM techniques. The findings shown that TOE factors contribute to blockchain adoption by keeping the intention to adopt the technology as a mediating variable.

Rostyslav Klochko's and **Olena Piskunova's** article *Marketing segmentation of banks' corporate clients based on data mining technique* aims to segment a bank's corporate client base and develop a pricing strategy for each of the groups that have been singled out in the process. The study sample consisted of 4,500 corporate clients

of a Ukrainian bank who were active users of euro accounts. The k-means data mining algorithm was used to develop marketing segments. The optimal number of clusters was determined by weighing the results of calculating 26 indices from the NbClust package and the bank's business requirements. The study found that clusters 1st and 2nd were a concentration of unprofitable customers for whom an introduction of a service fee was urgently needed. Marketing segments 3 and 4 were customers who did not record net losses but with whom it was deemed necessary to work to improve their profitability. The remaining two segments were 'healthy' users of euro accounts. With regard to these customers, it was recommended no additional service fees should be imposed.

Other articles

XXXX Multivariate Statistical Analysis 2022, Lodz, Poland. Conference Papers

Grzegorz Kończak and **Katarzyna Stąpor** in the paper ***Changepoint detection with the use of the RESPERM method – a Monte Carlo study*** use RESPERM (residuals permutation-based method) as a single changepoint detection method based on regression residuals permutation, which can be applied to many physiological situations where the regression slope can change suddenly at a given point. The article presents the results of a Monte Carlo study on the properties of the RESPERM method for single changepoint detection in a linear regression model. The proposed method was compared with a well-known segmented method for detection breakpoint in linear models. In the simulation study six variants of noise were considered from normal, uniform and two variants of beta distributions together with two cases of equal and unequal variances. Three levels of variance in the distribution of random errors were taken into account: minor, major and dominant errors. The simulations were performed for different locations of changepoint in time series. The Monte Carlo study showed that when the input data are very noisy, the RESPERM method outperforms the segmented approach in terms of variance, and in the case of bias, the results of the two methods are comparable.

*XXXI Scientific Conference of the Classification
and Data Analysis Section (SKAD 2022)*

Adam Korczyński's paper entitled ***Bayesian predictive probability design – theory and practical example in a prospective study*** provides theoretical background and the practical perspective, pointing out the statistical properties but also technical aspects in conducting a trial with predictive design. Also, sensitivity of the design to the choice of prior distribution was considered. The Bayesian predictive design allows to draw conclusions on the prognosis given the actual results. Their theoretical properties are appealing as a tool for detecting the treatment sig. The practical application has shown

the usefulness of the approach from the perspective of the timing of the decision. This accords with argument for adaptive design allowing for reducing the overall sample size, cost of the study, drug development time length. The final decision would still require larger sample size, although the Bayesian design seems to have a supportive role.

Research Communicates and Letters

In the Research Communicates and Letters section an article by **Wojciech Wójciak** analyses *Another solution for some optimum allocation problem*. The study derives optimality conditions for the optimum sample allocation problem in stratified sampling, formulated as the determination of the fixed strata sample sizes that minimize the total cost of the survey, under the assumed level of variance of the stratified π estimator of the population total (or mean) and one-sided upper bounds imposed on sample sizes in strata. In this context, the author presumes that the variance function is of some generic form that, in particular, covers the case of the simple random sampling without replacement design in strata. The optimality conditions mentioned above are derived from the Karush-Kuhn-Tucker conditions, and the study formulates the LRNA in such a way that it also provides the solution to the classical optimum allocation problem of minimization of the estimator's variance under a fixed total cost (under one-sided lower bounds imposed on sample sizes in strata). In such a case, the LRNA can be considered as a counterparty to the popular recursive Neyman allocation used to solve the classical problem of sample allocation with added one-sided upper bounds.

Włodzimierz Okrasa

Editor

© Włodzimierz Okrasa. Article available under the CC BY-SA 4.0 licence

