

From the Editor

Hereby, we present to our readers the last issue of Statistics in Transition new series in 2022, containing 12 articles on various topics and nature. As usual, articles are grouped into three conventionally defined categories: research papers, other articles and Communicates and Letters. The authors of these articles come from many countries: Czech Republic, USA, Egypt, Poland, Iran, Nigeria, India, Sri Lanka, Botswana, and Thailand.

As the last publication of the year, this issue gives us the opportunity to express our appreciation and thanks to all our contributors: authors, reviewers and all the participants of the editorial process. We consider it a great achievement to have such a large and growing community of internationally renowned experts among our collaborators and journal's stakeholders. I would like to express my special gratitude to the reviewers – also on behalf of the authors of the published articles – as their comments and suggestions had a positive impact on the overall quality of the submitted papers.

On behalf of the Editorial Board, Associate Editors and the journal's readers I sincerely thank all our partners and patrons.

Research articles

The first paper by **Richard Hindls**, **Lubos Marek** and **Stanislava Hronová** entitled *Changes in the structure of household disposable income in selected countries as a reflection of the crises after 2000* shows how the relationship between the shares of households' wages and final consumption expenditure in their gross disposable income has developed over the past 20 years. The presented analysis uses publicly available national accounts data for 30 countries for the period of 2000–2019. The studied indicators include the proportion of households' wages and salaries, and final consumption expenditure in their gross disposable income. The analysis of the newly constructed measure t has shown a decrease (i.e. an approach to the origin of the coordinates in the spatial map of the 30 countries) of these proportions in the years of financial crisis and economic recession and, on the contrary, an increase (i.e. a move away from the origin of the coordinates of the spatial map) of the examined proportions in the years of prosperity (economic growth). To confirm this assumption, along with the substantive reasoning, the authors have also used the original measure t , which not only quantifies these statements sensitively, but also defines the intensity of the

phenomenon (the degree of approach or departure from the origin of the coordinates). The aggregate analysis is then applicable without any limitation in terms of the number of countries (or entire territories) and years studied – the procedure can be applied, for example, to groups of countries according to their economic development, their geopolitical demarcation, etc.

Yang Liu and **Balgobin Nandram** in their article *Sampling methods for the concentration parameter and discrete baseline of the Dirichlet process* start with observation that there are many models in the current statistical literature for making inferences based on samples selected from a finite population. The authors review the current sampling methods for the concentration parameter, which use the continuous baseline distribution, and compare three different methods: the adaptive rejection method, the mixture of gammas method and the grid method. A new method based on the ratio of uniforms, and a discrete baseline approach to the DP prior and sample the unobserved responses from the finite population both using a Polya urn scheme and a multinomial distribution were proposed. The discrete baseline approach to a Phytophthora data set was applied. The ratio of uniforms is more accurate and it is faster considering the computational time. The authors have corrected the true number of distinct values in the sample by introducing a latent variable that indicates from which urn a new observation comes. Due to using this approach, the authors could give a more accurate estimation of the finite population mean when the observations are discrete.

The next paper *Parameter estimation of exponentiated exponential distribution under selective ranked set sampling* prepared by **Amal S. Hassan**, **Rasha S. Elshaarawy** and **Heba F. Nagy** describe the PRSS (Partial Ranked Set Sampling) method, which allows flexibility for the experimenter in selecting the sample when it is either difficult to rank the units within each set with full confidence or when experimental units are not available. The authors introduce and define the density and likelihood function for a random variable under the PRSS scheme. The suggested ranked schemes include the PRSS, RSS, neoteric RSS (NRSS), and extreme RSS (ERSS). An intensive simulation study was conducted to compare and explore the behaviour of the proposed estimators. The study demonstrated that the maximum likelihood estimators via PRSS, NRSS, ERSS, and RSS schemes are more efficient than the corresponding estimators under SRS. Also, PRSS is not the best method compared to the other ranked schemes, but it is important in some cases when selecting the sample.

Piotr Sulewski and **Magdalena Szymkowiak** present *The Weibull lifetime model with randomised failure-free time*. They indicated that treating failure-free time in the three-parameter Weibull distribution not a constant, but as a random variable, makes the resulting distribution much more flexible at the expense of only one additional

parameter. Four compound Weibull distributions with the location parameter having Uniform, Weibull, Gamma and Normal distribution were defined. Using these proposed models the analysis of three real lifetime data sets was performed. The received results showed that the new models fit better the data under consideration than the standard three-parameter Weibull distribution. However, anyone who will decide to use any of the proposed compound Weibull distributions in data analysis has to be equipped with a powerful computational environment – Excel, Mathcad, Mathematica, Matlab, Scilab, etc.

In the paper ***Robustness of randomisation tests as alternative analysis methods for repeated measures design*** Abimibola Victoria Oladugba, Ajali John Obasi and Oluchukwu Chukwuemeka Asogwa discuss the problem of using randomisation tests (R -tests) which are regularly proposed as an alternative method of hypothesis testing in a situation when assumptions of classical statistical methods are violated in data analysis. The authors describe the robustness in terms of the type-I-error and the power of the R -test, which were evaluated and compared with that of the F -test in the analysis of a single factor repeated measures design. The Monte Carlo approach was used in the simulation study. The results showed that when the data were normal, the R -test was approximately as sensitive and robust as the F -test, while being more sensitive than the F -test when data had skewed distributions. When the sphericity assumption was met, both the R -test and the F -test were approximately equally sensitive, whereas the R -test was more sensitive and robust than the F -test when the sphericity assumption was not met

Narendra Singh Thakur's and Diwakar Shukla's paper ***Missing data estimation by the technique of chaining in the survey sampling*** pointed out that the sample surveys are often affected by missing observations and non-response caused by the respondents' refusal or unwillingness to provide the requested information, or due to their memory failure. In order to substitute the missing data, a procedure called imputation is applied, which uses the available data as a tool for the replacement of the missing values. Two auxiliary variables create a chain which is used to substitute the missing part of the sample. The authors present the application of the chain-type factor estimator as a means of source imputation for the non-response units in an incomplete sample. The proposed strategies were found to be more efficient and bias-controllable than similar estimation procedures described in the relevant literature. These techniques could also be made nearly unbiased in relation to other selected parametric values. The findings are supported by a numerical study involving the use of a data set, proving that the proposed techniques outperform other similar ones.

The article ***Zero-modified Poisson-Modification of Quasi Lindley distribution and its application*** by Ramajeyam Tharshan and Pushpakanthie Wijekoon presents the

Poisson-Modification of Quasi Lindley (PMQL) distribution as a newly introduced mixed Poisson distribution for over-dispersed count data. The authors introduce the Zero-modified PMQL (ZMPMQL) distribution as an alternative to the PMQL distribution in order to accommodate zero inflation/deflation. The method of obtaining the ZMPMQL distribution jointly with some of its important properties, namely the probability mass and distribution functions, mean, variance, index of dispersion, and quantile function are presented. The maximum likelihood (ML) estimation method is used for the unknown parameter estimation, and the simulation study is conducted in order to evaluate the asymptotic theory of the ML estimation method and to show the superiority of the ML method over the method of moments estimation. The applicability of the introduced distribution is illustrated by using a real-world data set. In order to estimate its unknown parameters, the authors derived its log-likelihood function and score functions, which showed that the maximum likelihood estimation method is a suitable method to estimate its unknown parameters via a Monte Carlo simulation study. The results revealed its superiority over some other existing mixed Poisson and zero-modified mixed Poisson distributions.

Wilfred Molefe in the paper *Optimal allocation for equal probability two-stage design* examines the optimal designs when it is not feasible for every cluster to be represented in a sample as in stratified design, by assuming equal probability two-stage sampling where clusters are small areas. The paper develops allocation methods for two-stage sample surveys where small-area estimates are a priority. The author seeks efficient allocations where the aim is to minimize the linear combination of the mean squared errors of composite small area estimators and of an estimator of the overall mean. Several alternatives, including the area-only stratified design, are found to perform nearly as well as the optimal allocation but with better practical properties. Designs are evaluated numerically using Switzerland canton data as well as Botswana administrative districts data. This optimal design is less clustered than the usual classical two-stage optimal sample size \bar{n}_{cl} when more priority is given to larger clusters ($q > 0$). The area-only stratified optimum and the area-only simple two-stage optimum should always be the best designs in minimizing the objective function but they are not when there is equal priority for each cluster, that is when $q = 0$. These two designs have undesirable properties of allocating zero or even negative sample sizes to smaller clusters.

The next paper, by **Arisa Jiratampradab**, **Thidaporn Supapakorn** and **Jiraphan Suntornchost**, presents *Comparison of confidence intervals for variance components in unbalanced one-way random effects model* to study and compare the methods for constructing confidence intervals for variance components in an unbalanced one-way random effects model. The methods are based on a classical exact, generalised pivotal quantity, a fiducial inference and a fiducial generalised pivotal quantity.

The comparison of criteria involves the empirical coverage probability that is maintained at the nominal confidence level of 0.95 and the shortest average length of the confidence interval. The simulation results show that the method based on the generalised pivotal quantity and the fiducial inference perform very well in terms of both the empirical coverage probability and the average length of the confidence interval. The classical exact method performs well in some situations, while the fiducial generalised pivotal quantity performs well in a very unbalanced design. Therefore, the method based on the generalised pivotal quantity is recommended for all situations.

Arvind Pandey, David D. Hanagal and Shikhar Tyagi focus on *Generalized Lindley shared additive frailty regression model for bivariate survival data*. Frailty models are the possible choice to counter the problem of the unobserved heterogeneity in individual risks of disease and death. Based on earlier studies, shared frailty models can be utilised in the analysis of bivariate data related to survival times (e.g. matched pairs experiments, twin or family data). It was assumed that frailty acts additively to the hazard rate. A new class of shared frailty models based on generalised Lindley distribution is established. By assuming generalised Weibull and generalised log-logistic baseline distributions, the authors propose a new class of shared frailty models based on the additive hazard rate. The parameters in these frailty models and the use of the Bayesian paradigm of the Markov Chain Monte Carlo (MCMC) technique were estimated, and model selection criteria were applied for the comparison of models. The kidney infection data allowed to conclude that the best model was analysed. To fit the proposed model the hybrid M-H algorithm was applied.

Other articles

The XXXIX Multivariate Statistical Analysis 2021, Lodz, Poland. Conference Papers

In the paper by **Morteza Nazifi and Hamid Fadishei** *Supsim: a Python package and a web-based JavaScript tool to address the theoretical complexities in two-predictor suppression situations* the authors show that two-predictor suppression situations continue to produce uninterpretable conditions in linear regression. Their study introduces two different versions of software called suppression simulator (Supsim): a) the command-line Python package, and b) the web-based JavaScript tool, both of which are able to simulate numerous random two-predictor models (RTMs). Such a comparison suggests that the basic mathematical concepts of two-predictor suppression situations need to be reconsidered with regard to the important issue of the statistical control function.

The study depicts a clear picture of the performance of the statistical control function in different suppression and non-suppression situations, and provides a mathematical proof indicating that the statistical control function does not work

correctly in suppression situations. The study also introduces an algorithm that can generate numerous simulated data sets showing all different kinds of suppression and non-suppression situations known so far, and therefore they help resolve the theoretical complexities related to two-predictor suppression situations by expanding the previous knowledge in this field.

Research Communicates and Letters

Barbara Wodecka and **Michał Stachura** discuss ***k -th record estimator of the scale parameter of the α -stable distribution***. The authors present an estimation technique that involves the k -th record theory. Several theoretical properties of the introduced scale parameter estimators are demonstrated. With the use of Monte Carlo methods, a comparative analysis is performed between the approach based on k -th records and approaches based on Hill's and Pickands' estimators. The research indicates several advantages of the k -th record approach over its other counterparts, especially when dealing with incomplete information about the underlying sample. It is also remarked that the insights, specified in the paper, should be perceived essentially as the advantages of the ' k -th record' approach over the other ones presented, since Berred's estimator, and the scale parameter estimator based on it, may be employed in cases of incomplete information about an underlying sample. The authors show that, on the one hand, this incompleteness may be very useful if an analysed data base must stay undisclosed, even for a researcher/statistician working on it, or more, the data are only partially recorded (i.e. record values of a proper order or several orders). On the other hand, if an analysed data base is absolutely fulfilled and disclosed, the ' k -th record' approach opens up opportunities to make use of permutation methods in order to make repeated estimation that leads to much more precise results.

Włodzimierz Okrasa

Editor

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

