STATISTICS IN TRANSITION new series, March 2019
Vol. 20, No. 1, pp. 1−4

FROM THE EDITOR

This issue of the *Statistics in Transition new series* is slightly modified in strictly editorial terms. Namely, we resign from traditionally maintained distinction between papers grouped, respectively, under the headings 'sample and estimation issues' and 'research papers'. From now on, they will be combined into one section, as original research papers. The other sections remain unchanged. In this way we are getting a possibility to place the paper we consider of a special interest to the readers at the forefront. In order to ensure the appropriate paper is obtained we are arranging - with the help of the distinguished members of our Editorial Board and Associate Editors panel – for an invited paper to be submitted by renewed scholars and leading experts. The paper by Danny Pfeffermann et al., which inaugurates this issue, provides an example of this new approach.

This issue contains a set of nine papers, characterized briefly below.

Danny Pfeffermann's, Dano Ben-Hur's and Olivia Blum's paper ***Planning the next Census for Israel*** is devoted to currently hot topic of designing census while dealing with challenges beyond those considered 'standard' in methodology of such a type of research. They are posed by the fact that despite having fairly accurate population register at the national level (consisting of about 9 million persons), Israel has much less accurate the register for small geographical (statistical) areas, with an average area enumeration error of about 13%. In order to correct the errors at the area level the following three-step procedure is employed: (i) draw a sample from an enhanced register to obtain initial direct sample estimates for the number of persons residing in each area on “census day”; (ii) fit the Fay-Herriot model to the direct estimates in an attempt to improve their accuracy; (iii) compute a final census estimate for each statistical area as a linear combination of the estimate obtained in step (ii) and the register figure. The authors also consider a procedure to deal with not missing at random (NMAR) nonresponse (in step i) - the proposed procedures are illustrated using data from the 2008 Census in Israel.

The paper ***Imputation of missing values by using raw moments*** by **Muhammed Umair Sohail, Javid Shabbir** and **Fariha Sohil** presents a method of imputation that has been found to be a cheaper procedure from a cost point of view in a situation when the sample data have missing values. The authors' study shows the improvement of the performance of imputation procedures by utilizing the raw moments of the auxiliary information rather than their ranks, especially, when the ranking of the auxiliary variable is expensive or difficult to achieve. Equations for bias and mean squared error are obtained by large sample-based approximation. Through the numerical and simulation studies it can be easily understood that the proposed method of imputation can outperform their counterparts.

Kumari Priyanka's and **Pidugu Trisandhya's** paper ***Modelling sensitive issues on successive waves*** addresses the problem of estimation of population mean of a sensitive character using non-sensitive auxiliary variable at current wave

in two wave successive sampling. A general class of estimator is proposed and studied under randomized and scrambled response model. Many existing estimators have been modified to work for sensitive population mean estimation. The modified estimators became the members of the proposed general class of estimators. The detail properties of all the estimators have been discussed. Their behaviour under randomized and scrambled response techniques have been elaborated. Numerical illustrations including simulation have been accompanied to judge the performance of different estimators. Finally, suitable recommendations are forwarded.

In the next article, ***Nonrandomized response model for complex survey designs*** by **Raghunath Arnab, Dahud Kehinde Shangodoyin** and **Antonio Arcos** discuss Warner's randomized response (RR) model, which is used to collect sensitive information for a broad range of surveys but possesses several limitations – such as lack of reproducibility and higher costs; and it is not feasible for mail questionnaires. To overcome such difficulties, nonrandomized response (NRR) surveys have been proposed. The proposed NRR surveys are limited to simple random sampling with replacement (SRSWR) design. In this paper, NRR procedures are extended to complex survey designs in a unified set-up, which is applicable to any sampling design and wider classes of estimators. Existing results for NRR can be derived from the proposed method as special cases.

Y. S. Ramakrishnaiah, Manish Trivedi and **Konda Satish** in the paper ***On the smoothed parametric estimation of mixing proportion under fixed design regression model*** re-examine the estimator proposed by Boes (1966) – James (1978), herein called BJ estimator, which was constructed for estimating mixing proportion in a mixed model based on independent and identically distributed (i.i.d.) random samples. They also propose a completely new (smoothed) estimator for mixing proportion based on independent and not identically distributed (non-i.i.d.) random samples. The proposed estimator is nonparametric in true sense based on known "kernel function" as described in the introduction. The following results of the smoothed estimator have been checked under the non-i.i.d. set-up: (i) its small sample behaviour as compared with the unsmoothed version (BJ estimator) based on their mean square errors by using Monte-Carlo simulation, and established percentage gain in precision of smoothed estimator over its unsmoothed version measured in terms of their mean square error, (ii) its large sample properties such as almost surely (a.s.) convergence and asymptotic normality of these estimators. These results are completely new in the literature not only under the case of i.i.d., but can be generalised to non-i.i.d. design as well.

The next article, ***The odd generalized exponential log-logistic distribution group acceptance sampling plan*** by **Devireddy Charana Udaya Sivakumar, Rosaiah Kanaparthi, Gadde Srinivasa Rao,** and **Kruthiventi Kalyani** presents a group acceptance sampling plan (GASP) being developed when the lifetime of the items follows odd generalized exponential log-logistic distribution (OGELLD), and the multiple number of items as a group can be tested simultaneously in a tester. The design parameters such as the minimum group size and the acceptance number are derived under specified the consumer's risk and the test termination time. The operating characteristic (OC) function values are calculated (intended) according to various quality levels, and the minimum ratios of the true average life to the specified average life at the specified producer's risk are derived. The

methodology is illustrated using real data on health and survival times of guinea pigs in days, and results show that the proposed methodology performs well as compared with existing sampling plans.

Soma Dhar, Lipi. B. Mahanta and **Kishore. K. Das** in the paper ***Formulation of the simple Markovian model using fractional calculus approach and its application to analysis of queue behaviour of severe patients*** introduce a fractional order of a simple Markovian model where the arrival rate of the patient is Poisson, i.e. independent of the patient size. Fraction is obtained by replacing the first order time derivative in the difference differential equations, which govern the probability law of the process with the Mittag-Leffler function. The probability distribution of the number N(t) of patients suffering from severe disease at an arbitrary time t is derived. Also, the authors obtain the mean size (number) of the patients suffering from severe disease waiting for service at any given time t, in the form of En 0:5;0:5(t), for different fractional values of server activity status, n = 1;0:95;0:90 and for arrival rates a = b = 0:5. A numerical example is also evaluated and analysed by using the simple Markovian model with the help of simulation techniques.

In the article ***An application of functional data analysis to local damage detection*** by **Jacek Leśkow** and **Maria Skupień** vibration signals sampled with a high frequency are used as a basic source of information about machine behaviour. Few minutes of signal observations easily translate into several millions of data points to be processed with the purpose of the damage detection. Big dimensionality of data sets creates serious difficulties with detection of frequencies specific for a particular local damage. In view of that, traditional spectral analysis tools like spectrograms should be improved to efficiently identify the frequency bands where the impulsivity is most marked (the so-called informative frequency bands or IFB). The authors propose the functional approach known in modern time series analysis to overcome these difficulties. The data sets are treated as collections of random functions to apply techniques of the functional data analysis (FDA). In effect, massive data sets can be represented by few real-valued functions and corresponding parameters, which are the eigen-functions and eigen-values of the covariance operator describing the signal. Also, a new technique based on the bootstrap resampling is proposed to choose the optimal dimension in representing big data sets under processing. Using real data generated by a gearbox and a wheel bearings, it is demonstrated how these techniques work in practice.

Sukanya Intarapak's and **Thidaporn Supapakorn's** paper, ***An alternative matrix transformation to the F test statistic for clustered data*****,** discusses the problem of regression analysis of clustered data when the error of cluster data violates the independence assumption. Consequently, the OLS based test statistic leads to incorrect inferences. To overcome this shortcoming, the transformation is required to apply to the observations. The authors propose an alternative matrix transformation that adjusts the intra-cluster correlation with Householder matrix and apply it to the F test statistic based on GLS (generalized least squares) procedures for the regression coefficients hypothesis. By Monte Carlo simulations of the balanced and unbalanced data, it is found that the F test statistic based on the GLS procedures, with Adjusted Householder transformation, performs well in terms of the type I error rate and power of the test.

In the last article (in the *other articles* section), ***Survival regression models for single event and competing risks based on pseudo-observations*** by **Ewa Wycinka** and **Tomasz Jurkiewicz,** a survival data problem associated with the presence of censored observations is discussed. If no censoring occurs in the data, standard statistical models could be employed to analyse them. Pseudo-observations can replace censored observations and thereby allow standard statistical models to be used. Authors apply a pseudo-observation approach to single-event and competing-risks analysis. In the empirical part of the study, the use of regression models based on pseudo-observations in credit-risk assessment was investigated. Default, defined as a delay in payment, was considered to be the event of interest, while prepayment of credit was treated as a possible competing risk. Credits that neither default nor are prepaid during the follow-up were censored observations. Typical application characteristics of the credit and creditor were the covariates in the regression model. In a sample of retail credits provided by a Polish financial institution, regression models based on pseudo-observations were built for the single-event and competing-risks approaches. Estimates and discriminatory power of these models were compared to the Cox PH and Fine-Gray models.

Włodzimierz Okrasa
Editor