

## FROM THE EDITOR

In the current issue of *Statistics in Transition new series* four articles are devoted to the problems of estimation and sampling while in the next four ('research articles') a variety of different topics are discussed.

In the first paper, entitled *Estimating sensitive population proportion using a combination of binomial and hypergeometric randomized responses by direct and inverse mechanism*, **Kajal Dihidar** and **Manjima Bhattacharya** follow the approach discussed in some earlier papers (Chaudhuri and Dihidar 2014, and Dihidar 2016) towards estimating the sensitive population proportion using a combination of binomial and hypergeometric randomized responses by direct and inverse mechanism. Along with the traditional simple random sampling, with and without replacement, here they consider sampling of respondents by unequal probabilities. They start with an observation that individuals in a sample survey may (for various reasons) prefer not to confide to the interviewer the correct answers to certain potentially sensitive questions such as the illegal use of drugs, illegal earning, or incidence of acts of domestic violence, etc. The resulting evasive answer bias is ordinarily difficult to assess. The use of a randomized response method for estimating the proportion of individuals possessing those sensitive attributes can potentially eliminate the bias, as the authors demonstrate with numerical illustration, comparing a study of the relative efficiencies of the direct and an inverse mechanism.

**Ene-Margit Titt's** paper, *Residency Testing. Estimating the true population size of Estonia*, discusses the idea of using the residency index as a tool created for estimating the under- and over-coverage of population census and calculation of proper population size. For this aim the concept of a sign of life – a binary variable depending on register  $i$ , person  $j$  and year  $k$  has been introduced showing if the person was active in the register in a given year. The weighted sum of signs of life indicates the probability that the person belongs to the set of residents in a given year. To improve the stability of the index a linear combination of the previous value of the index and the sum of signs of life is used. Compared with other types of statistical models, the advantage of this methodology is the possibility of using a large number of different registers, which may have both positive and negative impact on the residency status. Some obstacles to using such an index are also mentioned – in some situations, the residency indexes cannot be used, for instance, for describing interior migration and population in small areas.

In the next article, *Efficient Family of Ratio-Type Estimators for Mean Estimation in Successive Sampling Using Auxiliary Information on Both Occasions*, **Beevi Nazeema T., Chandran C.**, propose an efficient family of ratio-type estimators using one auxiliary variable for the estimation of the current population mean under successive sampling scheme following methodology originally studied by Ray and Sahai (1980) under simple random sampling using one auxiliary variable for estimation of the population mean. They employ these estimators in successive sampling, with usual ratio estimator being identified as a particular case of the suggested estimators. The proposed family of estimators at optimum condition is compared with the simple mean per unit estimator, Cochran (1977) estimator and existing other members of the family. Among the results, the authors indicate that a smaller fresh sample at current occasion is required if a highly positively correlated auxiliary character is available, what also may reduce the cost of the survey.

Paper by **Alkaya Aylin, Ayhan Öztes H., Esin Alptekin**, *Sequential Data Weighting Procedures for Combined Ratio Estimators in Complex Sample Surveys*, is devoted to the issue of sample surveys weighting procedure to increase the quality of estimates. While there are many types of situations calling for such procedure, the authors point to unequal probability of selection, compensation for nonresponse, and post-stratification as to the most important reasons of weighting. The authors propose a sequential data weighting procedure for the estimators of combined ratio mean in complex sample surveys and general variance estimation for the population ratio mean. They illustrate the utility of the proposed estimator using data from Turkish Demographic and Health Survey 2003, and showing that that auxiliary information on weights can considerably improve the efficiency of the estimates of post-stratification.

The research articles set starts with the paper by **Wojciech Łukaszonek's** paper, *A Multidimensional and Dynamised Classification of Polish Provinces Based on Selected Features of Higher Education*, which addresses the issue of distribution of the unprecedented, fivefold increase in the number of students and the number of higher educational (HE) institutions in Poland over the so-called, post-1989, transition period. Given that the distribution was not uniform in any respect (space or time), the regional differentiation between country provinces is analysed with special attention being paid to the years 2002–2013. The applied procedure uses new statistical methods applicable to a space of double multivariate data. The covariance matrix used to construct principal components is structured as a Kronecker product. The results led to the identification of six groups of provinces, including two consisting of a single province – Masovian and Lesser Poland – which contain the biggest and the highest-ranked HE institutions in Poland (the University of Warsaw and Jagiellonian University).

The next paper, by **Shanker Rama, Shukla Kamlesh Kumar, Mishra Amarendra**, A *Three-Parameter Weighted Lindley Distribution and its Applications to Model Survival Time* treats about a semiparametric additive risks regression model for analysing middle-censored lifetime data arising from an unknown population. The authors propose a dual approach to estimating the regression parameters and the unknown baseline survival function: the first method uses the martingale-based theory, and the other one uses an iterative procedure. Results of simulation studies are reported to assess the finite sample behaviour of the estimators, followed by illustration of the utility of the model with a real life data set.

**Pinar Gunar Karadeniz's and Ilker Ercan's** paper *Examining Tests for Comparing of Survival Curves with Right Censored Data* addresses the problem faced in survival analysis – in estimating the survival probability of a population and comparing the survival experiences of different groups – given that data obtained from survival studies contains frequently censored observations. The authors examine several tests (Logrank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto and tests belonging to Fleming-Harrington test family with  $(p, q)$  values;  $(1, 0)$ ,  $(0.5, 0.5)$ ,  $(1, 1)$ ,  $(0, 1)$  ve  $(0.5, 2)$  are examined by means of Type I error rate obtained from a simulation study). As a result of the simulation study, Type I error rate of Logrank test is equal or close to the nominal value. The authors conclude that in the situation when survival data were generated from lognormal and inverse Gaussian distribution, Type I error rate of Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto and Fleming-Harrington  $(1,0)$  tests were close to the nominal value.

**Piyush Kant Rai, Pareek Sarla, and Joshi Hemlata**, employ SAE methodology in their article *Met and Unmet Need for Contraception: Small Area Estimation for Rajasthan State of India*, focused on a policy important, yet sensitive issue of access to contraception. Using data for 187 towns (of Rajasthan state of India) and the data from the District Level Household Survey (DLHS) 2002-04, and of the Census 2001 of India they estimate the proportion of women having met and unmet need (spacing and limiting fertility) of family planning. They employ Generalized Linear Mixed Model with logit-link function given the binomial nature of variables. The authors believe that the results of their analysis is of relevance to designing and implementing better policies and programmes in this area of possible state intervention.

**Włodzimierz Okrasa**

Editor