

STATISTICS IN TRANSITION-new series, December 2012
Vol. 13, No. 3, pp. 607—610

REPORT ON SURVEY SAMPLING AND SMALL AREA STATISTICS SESSIONS DURING THE CONGRESS OF POLISH STATISTICS IN POZNAŃ, 18–20 APRIL 2012

Survey sampling is a field of statistics with a long time tradition in Poland starting from Jerzy Sława Neyman's papers on stratified random sampling. The considerations were continued, among others, by R. Zasępa and Z. Pawłowski. Nowadays, research of many Polish survey statisticians in Poland include studies on small area estimation. This is the reason why four sessions on both survey sampling and small area estimation were organized during the Congress of Polish Statistics. During the sessions fourteen papers in English were presented, including five invited lectures given by Ray Chambers from Wollongong University, Jean-Claude Deville, Lorenzo Fattorini from University of Siena, Malay Ghosh from University of Florida and Li-Chun Zhang from Statistics Norway.

Ray Chambers presented a paper prepared together with Gunky Kim entitled *Regression analysis using data obtained by probability linking of multiple data sources*. In the presentation authors described how the framework for the inference using probability-linked data can be extended to define methods for efficient bias-corrected regression analysis when three registers are linked, or when a sample is linked to two separate registers, enabling longitudinal analysis. Their development allows for correlated linkage errors as well as errors arising when not all records can be linked. It can also be extended when more than three data sets are linked. The authors presented, among other things, the results of simulation study on accuracy of regression parameters.

Jean-Claude Deville together with Daniel Bonnéry and Guillaume Chauvet prepared a presentation entitled *Neyman type optimality for marginal quota sampling*. They were seeking for a similar to Neyman optimization procedure leading to inclusion probabilities specific for each cell of the table. A simple model based method was used but the assumptions of the model were strong. When the sampling is near from maximal entropy, there exists a quite simple approximation of the variance which is model free. The authors tried to optimize this approximation. A natural iterative procedure was used and it was shown to converge to this optimum.

Lorenzo Fattorini presented a lecture on *Design-based inference on ecological diversity*. He considered three main issues of ecological diversity analysis. Firstly, he studied the problem of measuring ecological diversity by means of suitable

indexes. Secondly, the problem of estimating these indexes was considered when biological populations were sampled by means of sampling schemes actually adopted by biologists. Thirdly, the problem of comparing and ordering biological communities with respect to their diversity was analyzed.

The presentation of Malay Ghosh entitled *Finite population sampling: a model-design synthesis* was devoted to the problem of the choice of the approach in the survey sampling between the design-based and the model-based one. While the basic conceptual disagreement between the two approaches could not be resolved, from an operational point of view, the author found an agreement between the two. He derived some general results which provided this agreement exactly or asymptotically and illustrated the results with several examples.

Li-Chun Zhang presented a paper *Micro calibration for data integration*. In the paper he explained how micro calibration could be achieved for categorical as well as continuous data. Some possibilities of using micro calibrated data were discussed, such as for the purpose of small area estimation.

In their presentation Konrad Furmańczyk and Stanisław Jaworski from Medical University of Warsaw and Warsaw University of Life Sciences entitled *Change point detection in a sequence of independent observations* proposed a new method where the change was identified when some given threshold was exceeded. The method was based on minimax decision rule.

Sara Franceschi from Università di Firenze presented a paper prepared together with Lorenzo Fattorini from Università di Siena and Daniela Maffei from Università di Firenze entitled *Design-based treatment of unit nonresponse by the calibration approach*. They considered the non-response calibration weighting in a complete design-based framework. Approximate expressions of design-based bias and variance of the calibration estimator were derived, design-based consistency was investigated and some estimators of the sampling variance were proposed. The results of the simulation study demonstrated how the reliability of the procedure was mainly determined by the capability of selecting auxiliary variables in such a way that their relationship with the interest variable was similar for both the respondent and non-respondent sub-populations.

Wojciech Gamrot from University of Economics in Katowice prepared a presentation *On empirical inclusion probabilities*. The author studied the problem when unknown first order inclusion probabilities were unknown and had to be replaced by estimates obtained in a simulation study. Such estimates are called empirical inclusion probabilities. In the presentation the number of sample replications was analyzed which should have been drawn to ensure desired accuracy of the population total estimates. What is more, an attempt was made to review known solutions to this problem and to improve them for a certain sampling scheme.

Tomasz Klimanek from both University of Economics in Poznań and Statistical Office in Poznań presented a paper *Using indirect estimation with*

spatial autocorrelation in social surveys in Poland with an application of indirect estimation to estimate some characteristics of labour market in the population of people aged 15 and over at the level of NUTS4 in Wielkopolska region in 2008. What is more, he compared the precision of the direct estimator with those of the EBLUP estimator and the SEBLUP estimator (which takes into account spatial correlation).

Tomasz Józefowski from Statistical Office in Poznań presented a paper *Using a SPREE estimator to estimate the number of unemployed across subregions*. He used SPREE estimator to adjust values in the cells of an estimated contingency table to the totals obtained by means of survey sampling. The example of estimating the number of unemployed at the level of subregions of Wielkopolska province using data on registered unemployment and LFS was presented.

Marcin Szymkowiak from the University of Economics in Poznań presented a paper *Construction of calibration estimators of total for different distance measures*. The main goal of the paper was to present the construction of calibration estimators for the total for different types of distance measures. Its empirical part, based on a simulation study using real data, provided a comparison of different types of distance measures that could be used in the construction of calibration weights in the case of non-response.

Imbi Traat from University of Tartu in a paper *Domain Estimators Calibrated on Reference Survey* studied the situation when population totals, or totals of larger domains, for certain variables were estimated in one survey called the reference survey. Estimates based on the same variables but at a more detailed domain level were obtained from a second (later) survey, called the present survey. Consistency was required for the present survey estimates in the sense that domain totals for common study variables had to sum up to the corresponding estimated totals in the reference survey (or - in special case - known totals from registers). To solve the problem the author used calibration framework but in a more general setting.

Janusz L. Wywiał from University of Economics in Katowice presented a paper on *Estimation of population mean on the basis of a simple sample ordered by auxiliary variable*.

He considered estimation of the population average in a finite population by means of sampling strategies dependent on an order statistic of an auxiliary variable. The inclusion probabilities were dependent on the probability function of the order statistic of the auxiliary variable highly correlated with a variable of interest. The simple estimator of the population mean (total) of the variable under study was proposed. Its basic parameters were derived and the limit distribution of the estimator was also considered. Finally, on the basis of the simulation analysis, the accuracy of the estimator was compared with the precision of the well known estimators including the simple regression one.

Tomasz Żądło from University of Economics in Katowice in his paper *On prediction of totals for spatially correlated domains* studied small area estimation problems for longitudinal data. He proposed some special case of the General Linear Mixed Model including spatial and temporal correlation and assumed that the population could change in time and domains affiliation of population elements could change in time. Based on the assumption the author derived the empirical best linear predictor and the estimator of its MSE. Theoretical considerations were supported by simulation study.

Prepared by:

Janusz L. Wywiał, Tomasz Żądło