ONE RANDOM SAMPLE AS A TEST PLATFORM IN SEARCH FOR MORE ACCURATE ESTIMATES THROUGH SUB-SAMPLES SELECTED BY A MODEL-BASED ALLOCATION

Mauno Keto
University of Jyväskylä, Finland, mauno.j.keto@student.jyu.fi

Erkki Pahkinen
University of Jyväskylä, Finland, pahkinen@maths.jyu.fi

A typical overall sample size in Finnish nationwide surveys is between 1 500 and 3 000. The survey requires labor and financial resources. Owing to various reasons (refusals, difficult reachability, measurement problems etc.), the number of contacts to potential respondents in opinion polls may be many times as large as the desired sample size. Similar problems are possible also in technical surveys. A survey with overall sample size 1 500 may cost as much as 30 000 €. Therefore, the reduction of the sample size should be considered, but without losing reliability in the estimated parameters.

Estimates for the variables of interest are generally needed both on population and on subpopulation level. For this reason, the sample allocation should be planned carefully. All relevant information serving sample allocation is seldom utilized. According to public announcements given by market research companies, quota sampling using proportional allocation based on gender and age group combinations is commonly implemented.

We study the improvement of area and population estimates obtained from one random sample selected from real data by proportional allocation. We simulate subsamples with lower sample size from the original sample according to an allocation, which is based on multi-objective Pareto optimization and which uses a model-based area total estimator. Our model is the unit-level linear mixed model. We use three types of area total estimators: two design-based estimators Horvitz-Thompson and model-assisted GREG and the model-based EBLUP estimator. We assess the statistical properties (accuracy and bias) of the allocations and estimators.

Key words: cost reduction, subsample, model and estimator, conditioning, multi-objective optimization, performance.

References

Battese, G. E., Harter, R. M., and Fuller, W. A. 1988. An error component Model for Prediction of County Crop Areas using Survey and Satellite Date. Journal of the American Statistical Association 83, 28-36.

Friedrich, U., Münnich, R., and Rupp, M. 2018. Multivariate Optimal Allocation with Box-Constraints. Austrian Journal of Statistics 47, 33–52.

Keto, M., Hakanen, J., and Pahkinen, E. 2018. Register data in sample allocations for small-area estimation. An International Journal of Mathematical Demography 25, 184-214. DOI: https://doi.org/10.1080/08898480.2018.1437318.

Keto, M. and Pahkinen, E. 2017. On overall sampling plan for small area estimation. Statistical Journal of the IAOS 33, 727-740. DOI: 10.3233/SJI-170370.

Lehtonen, R. and Veijanen, A. 2009. Design-Based Methods of Estimation for Domains and Small Areas. In Handbook of Statistics, Vol. 29B, 219-249. New York: Elsevier.

Miettinen, K. 1999. Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston.

Rao, J. N. K. and Molina, I. 2015. Small Area Estimation (2nd Edition). Hoboken, NJ: John Wiley & Sons, Inc.