

HIGHLIGHTS OF THE WSC2021 IN SURVEY STATISTICS

D. Krapavickaitė

Vilnius Gediminas Technical University, Lithuania
e-mail: danute.krapavickaite@vilniustech.lt

Abstract

The 63rd World Statistics Congress was held on July 11-16, 2021, virtually. There were about 1600 participants from 104 countries, 770 authors submitted their abstracts/papers/posters. The number of presentations was higher than the number of papers.

Scientific program included 18 IASS-sponsored invited sessions, among them two special invited sessions. There were more presentations devoted to the survey statistics. The most popular survey statistics topics discussed in the WSC:

- population census,
- imputation of missing data [6],
- machine learning (classification example of the random forest usage and section on nonprobability sampling included in [12])
- population size estimation.

The main motif of the WSC presentations in survey statistics – data integration:

- macrointegration: time series of temporary employment - combining quarterly sample survey (LFS) and monthly employment register (ER) data;
- nonprobability samples; survey using Facebook sample data [5, 3];
- data integration by combining probability sample and nonprobability sample data, probability sample survey data and big data for finite population inference [1,2,4,7-11];
- small area estimation in the case of probability sample and non-probability data set [2, 10].

The facebook data-based sample survey carried out by a big team of statisticians from the Carnegie Mellon University (CMU), University of Maryland (UMD) and Facebook, US, will be discussed in more detail.

References

1. Beaumont, J.-F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, **46**, 1-28.
2. Beaumont, J.-F., J. N. K. Rao. (2021) Pitfalls of making inferences from nonprobability samples: Can data integration through probability samples provide remedies? *The Survey Statistician*, **83**, 11-22.
3. Bradley, V. C., S. Kuriwaki, M. Isakov, D. Sejdinovic, X.-L. Meng, S. Flaxman. (2021) Are We There Yet? Big Data Significantly Overestimates COVID-19 Vaccination in the US. *MedRxiv, The preprint server for health sciences* (has not been peer-reviewed). <https://www.medrxiv.org/content/10.1101/2021.06.10.21258694v1>. <https://doi.org/10.1101/2021.06.10.21258694>.
4. Hill C. A., P. Biemer, T. D. Buskirk, L. Japac, A. Kirchner, S. Kolenikov, L. E. Lyberg (editors). (2020) *Big Data Meets Survey Science: A Collection of Innovative Methods*. ISBN: 978-1-118-97632-6.

5. Kreuter F. et al. (2020). Partnering with Facebook on a university-based rapid turn-around global survey. *Survey Research Methods*. **14**(2), pp. 159-163. <https://doi.org/10.18148/srm/2020.v14i2.7761>.
6. Lee D., J. K. Kim (2020). Semiparametric imputation using conditional Gaussian mixture models under item nonresponse. *Biometrika*. <https://doi.org/10.1111/biom.13410>.
7. Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox and the 2016 US presidential election. *Annals of Applied Statistics*, **12**, 685-726.
8. Tam S.-M., Holmberg A. (2020). New Data Sources for Official Statistics – A Game Changer for Survey Statisticians? *The Survey Statistician*, **81**, 21-35.
9. Tam, S.-M. (2015). A statistical framework for analyzing Big Data. *The Survey Statistician* **72**, 36-51.
10. Rao J. N. K. (2020) On Making Valid Inferences by Integrating Data from Surveys and Other Sources, *Sankhyā B: The Indian Journal of Statistics*. <https://doi.org/10.1007/s13571-020-00227-w>.
11. Yang S., J. K. Kim. (2020) Statistical data integration in survey sampling: a review. *Japanese Journal of Statistics and Data Science*, **3**(1), 625–650. <https://doi.org/10.1007/s42081-020-00093-w>.
12. Valliant, R., Dever, Jill A., Kreuter, F. (2018) *Tools for Designing and Weighting Survey Samples*, Springer. ISBN 978-3-319-93632-1.