

November 20, 2019 Statistics Sweden Workshop

Abstract

“Survey-assisted modeling: Past, present and future statistical integration of the sample survey method with administrative and other big data systems”

Contemporary advances in large scale data acquisition, compilation and analysis have many statisticians and scientists questioning what the future role of traditional surveys will be (Groves, 2011). This workshop aims to demonstrate that, while the role of surveys may be changing, to meet the statistical challenges of the future our survey designs and methods must remain in a long term partnership with existing and emerging analytic methods for large scale administrative and other big data resources. Relative to the past where population data from Censuses and administrative data served to improve the efficiency of survey designs or calibrate estimates based on complex survey measurements, future sample surveys will increasingly be designed to “assist” our statistical approaches to modeling that rely heavily on administrative or other big data inputs. Sample surveys’ “assistant” role may be to collect information on unmeasured relationships in statistical models that depend on administrative data, to monitor and calibrate model-generated estimates or otherwise address the sources of error in our big data models of social, economic and health-related processes and outcomes.

The mutual dependency of sample surveys and external sources of population data is by no means new. Common statistical techniques such as post-enumeration surveys following Census data collections, small area estimation (Rao and Molina, 2015), combining of satellite imagery and ground-based survey observations to address agricultural and environmental estimation and modeling problems are a few common examples. Case studies from the author’s experience in economic, business dynamics and retirement studies conducted over the past 40 years as well as new work in the fields of genetics, mental health and neuroscience will be used to demonstrate this trend toward big data driving survey design, measurement and analysis. Statistical methods for integrating big data and survey data will be introduced and illustrated using these case studies.

Looking then to the future, the workshop will explore new, emerging fields of statistical and analytical inquiry that will certainly employ “survey-assisted modeling” and discuss the implications of these future trends for those of us who will be practicing sample survey design and survey statistics/methodology in the coming decades.