Workshop

Coordinated Sampling for Business Surveys

Date: Friday, 1st March 2019, 9:00 – 16:30
Venue: Statistics Netherlands, The Hague (NL)

ENBES
EUROPEAN NETWORK FOR BETTER ESTABLISHMENT STATISTICS

Report

The workshop was organised by ENBES and hosted by Statistics Netherlands. The workshop consisted of five presentations (see programme) and one short contribution, alternated with two discussion sessions. In the first presentation Alina Matei (University of Neuchâtel) gave an overview of sample coordination methods. Also criteria for measuring the effectiveness of sample coordination were discussed and the application of sample coordination to spatially balanced sampling. In the other presentations the national implementations of sample coordination were discussed by the statistical offices of France, Sweden, Switzerland, United Kingdom and The Netherlands. The discussion sessions were moderated by Paul Smith (University of Southampton).

The sample coordination systems that were presented are all based on the use of permanent random numbers and are simultaneously used for sample selection and the spread of survey burden (negative coordination). The sampling systems however differ from each other in a number of aspects:

- some systems require that different surveys share the same sampling strata (basic stratification), whereas others do not,
- some systems only support (stratified) simple random sampling (SRS) while others also include other sampling designs such as probability proportional to size (PPS) sampling or cluster sampling,
- some systems support a fixed sample size, while others support fixed exact inclusion probabilities, but at the expense of a random sample size (Poisson sampling),
- in some systems both positive and negative sample coordination is supported, while others only support negative coordination,
- some systems take into account a weight that expresses the (relative) response burden of a survey while others do not; this weight can be survey-dependent or stratum-dependent,
- systems differ in whether and how mergers and splits in terms of the propagation and computation of response burden is treated,
- some systems use the complete history of survey burden to compute the samples that are drawn and have a computational cost which increases with the number of (time) instances that the surveys are drawn, while other just use a ‘single’ burden value which is computationally much faster.

In discussion it was stressed that for an institute which is considering to start with a sample coordination system, it is important to think carefully about the main goals that the system aims to achieve. Once a system has been started, it is usually difficult to make changes because that may imply a shift in the units that are involved in the sample. Two drivers were mentioned for introducing a sample coordination system. Firstly, a more evenly spread of the response burden for businesses. Secondly, it was mentioned that when businesses are required to provide information in more surveys the measurement errors tend to increase.

Negative sample coordination is generally used to spread the survey burden evenly over the population, where the spread can be both over time and the surveys. In the strata with smaller sampling fractions one might aim to have a certain maximum number of subsequent periods that a business stays in a panel survey, a maximum number of surveys per business, or one might aim for a sabbatical period during which a business does not receive any survey. With the currently developed systems one cannot give enterprises guarantees on the maximum number of surveys or the length of holiday periods and when businesses grow (and move into a larger size class) the lengths of these periods become smaller.

We exchanged different ideas concerning the evaluation of sample coordination systems. Evaluation measures are:

- the expected value of the overlap between surveys and its variance (see presentation of Alina Matei),
- a measure for the distribution of the response burden within the population, perhaps a Gini-type index,
- a measure of 'closeness' between the optimum spread of the surveys given the requirements (on positive and negative coordination and the number of surveys per sampling unit, holiday periods and so on) and its realisation. This third measure will require simulation studies to find the optimum.

A new development which was presented was a method for coordination of spatially balanced sampling. Businesses or firms (establishments) have an address. One may aim to achieve an even, rather than clustered, spread over a region. Possible applications are estimation of crop yields, where the even spread is a means to account for varying soil and weather conditions and thus yield more accurate results. Another application is estimation of soil use or intensity of use of solar panels from satellite images. Estimations based on samples from Big Data offer the opportunity to have attention for measurement error correction.

We also discussed that the burden can be expressed in different ways. For instance relative to the balance total, relative to turnover, relative to the number of employees. The burden of a survey may also depend on the complexity and statistical importance of a business.

Possible future topics for sample coordination systems are:

- Inclusion of new types of sampling designs for single surveys along with SRS and PPS samples into a coordination system. Examples are:
- sampling designs accounting for different unit types (from enterprise groups to firms),
- adaptive designs (e.g. an adaptive design for non-response follow up)
- spatially balanced samples.

Further, development of systems that are flexible in types of sampling design, in the design of the strata, that include explicit measures for survey burden and are computationally fast.

Finally, the participants expressed the importance of making software available that can easily be shared, such as packages in R. The sampling algorithms of the Dutch coordination system are already implemented in the R-package SBS.

Arnout van Delden
Paul Smith
Marc Smeets