BADEN: Bayesian Adaptive Survey Design Network

State-of-the-Art

Statistics Sweden (Statistiska centralbyrån)

Peter Lundquist

March 25th, 2015

1. Background

Statistics Sweden is a Swedish government agency responsible for producing official statistics. The agency has approximately 1,400 employees. Data collection is carried out by the agency.

The main modes for data collection among individuals and households at Statistics Sweden are telephone (CATI) and mail. Face-to-face interviewing is seldom used and the use of web questionnaires is currently being investigated. Administrative data and data collection process data are used in all individual and household surveys in the estimation; to adjust for the nonresponse and during the data collection to contact the sampled objects. There are many similarities with the way Statistics Netherlands produce their statistics, but there are also differences.

Increasing nonresponse and increased costs for data collection, in particular in individual and household surveys, have been a problem for Statistics Sweden during recent years. Some surveys have a long history of high nonresponse rates, e.g. the Household Budget Survey with a response rate of 38 percent 2012, while other surveys have experienced a sharp increase in recent years. One example is the Labour Force Survey (LFS): in November 2014 the nonresponse was 38.8 percent; in November 2007 the corresponding figure was 18.9 percent. The yearly total sample for LFS is for the moment 362,000 which is about 50 percent of the total number of telephone interviews carried out by Statistics Sweden during one year.

The last decade has demonstrated a quick change in the use of cell phones. Almost all adult Swedish citizens below (say) 75 years have a cell phone today. During the same time period, the proportion of fixed telephones has decreased. In 2004, about 94 percent of the households in Sweden had a landline; in 2014 the corresponding figure was about 68 percent.

In the BADEN network, the intention is to focus on household surveys. We will in addition launch work on adaptive design for the business sector and we seek cooperation in this area too.

At Statistics Sweden there are currently two types of development projects in the data collection field: short-term and long-term projects. Where the short-term projects aim at finding strategies to improve the data quality and increase the response rate, the long-term projects focus on theoretical developments such as adaptive and/or responsive design. An important outcome of the network, from our point of view, is to enable the organizations to create a sustainable long-term development, integrated with the necessary short-term developments.

2. Current use of adaptive and responsive surveys

Several studies at Statistics Sweden have confirmed that data collection motivated principally by a desire to get the best possible ultimate response rate is inefficient. The studies suggest that scarce resources have been spent with little effect on the estimates and little improvement in representativity. With this background, a project on responsive design was launched in 2011, with the objective to explore procedures for obtaining a balanced set of respondents from a given probability sample.

The procedures considered rely on indicators that can be monitored during data collection. The identified indicators mirror different aspects of the nonresponse problem. A motivation for an increased and systematic use of auxiliary information in the data collection phase is that better control of the field work can be obtained. The data inflow can be continuously followed, and the emphasis of the analysis can be altered, for example by focusing on low-responding subgroups, identified with the aid of selected auxiliary variables.

So far, most experiments have been carried out in retrospective on previously collected survey data (see papers by Lundquist and Särndal in the last Section). The concepts proposed by the project are general in scope and can be applied to a variety of sample surveys. In the project, the 2009 Swedish Living Conditions Survey (LCS) was chosen to illustrate the use of the indicators. The studied indicators also used paradata from the Swedish CATI-system for examining the data collection in retrospect. The results in the first report confirmed earlier findings; the follow-up efforts did not produce the improvement that one would hope for.

The use of the indicators has also been applied to other surveys at Statistics Sweden, for inspection of the data collection and in the evaluation of embedded experiments.

The indicators are quite simple to translate to program code, but no official tool has been developed at Statistics Sweden. One draw-back is the lack of financial support. For the moment, the focus of Statistics Sweden is on improving the infrastructure for data collection and implementing mixed-mode designs.

3. Current research on adaptive survey design
I. Reduce imbalance and improve accuracy

This project consists of two subprojects. The projects investigate expected benefits at the estimation stage when the balance of the response set is improved during data collection.

The first subproject is joint work between Carl-Erik Särndal and researchers from University of Tartu in Estonia (Imbi Traat and Kaur Lumiste). Results show that the variance of the bias – measured as the deviation of the calibration estimator from the (unrealized) full-sample unbiased estimator – is a linearly decreasing function of the response imbalance, which is measured and controlled continuously over the data collection period. An attractive prospect is thus a lower risk of bias if one can manage the data collection so as to get low imbalance. The theoretical results are validated in a simulation study using data from an Estonian household survey.

In the second subproject, Peter Lundquist and Carl-Erik Särndal use empirical data from two Swedish surveys in simulation studies to illustrate the questions above. The evaluation will be based on recently developed indicators for nonresponse. The work will be presented at the ISI-conference in 2015.

II. A simulation technique to assess the behaviour of representativeness and balance indicators

Pär Karlsson investigates indicators in a simulation study. During the data collection process it is observed that the indicators numerically deviate from the ideal values. The question is how these

numbers should be interpreted. What is the range of negligible deviations? How will these indicators behave when the fraction of the sample with observed data increases? A method is presented that will assist in evaluating the behaviour of the indicators by simulation. The simulated orderings can be tailored to mimic assumptions similar to missing completely at random or missing at random, where the ordering of the queue depends on auxiliary variables.

III. Use of nonresponse indicators in the Swedish Turnover in the Service Sector (TSS)

The recently started project has three members: Annika Lindblom, Peter Lundquist and Carl-Erik Särndal. Increased nonresponse is also observed in the business sector, the increase is however not as drastic as in the individual and household sector. There are good opportunities to monitor and control the data collection efficiently by using somewhat different strategies for different subgroups of businesses (e.g. branches or sizes). Such differentiation is done today, but maybe not in a structured and coordinated manner and with the most efficient use of the information. Approaches to adaptive design for individual and household statistics could be used also on the business side. With the help of survey -and process data, new strategies for data collection can be developed with the objective to achieve a better balance in response and to increase the quality of the survey estimates. The project will consider work done at the US Census Bureau and Statistics Netherlands.

4.   A proposed research agenda for Bayesian Adaptive Survey Designs

At Statistics Sweden at least two senior statisticians have the possibility to take part in the BADEN-project, Pär Karlsson and Peter Lundquist. Other senior statisticians with close involvements in the activities below are Anton Johansson, Annika Lindblom, Carl-Erik Särndal and Sara Westling.

The agenda for the coming year is:

- Short course at Stockholm University, April 20th 2015
- Poster presentation at the JOS 30th-anniversary, June 2015
- Invited paper presented at ISI, July 2015
- Invited paper presented at JSM, August 2015
- Adaptive/responsive design project in business sector, April 2015-
- Collaboration between Särndal and researchers at Tartu university
- Additional research at Statistics Sweden depending on kick-off meeting in February, March 2015-

Related projects at Statistics Sweden, January-December 2015:

- Process developments in data collection: Indicators and telephone process data
- Experiment with text-messages in LFS
- Mixed mode studies

5.   Collaboration

From Statistics Sweden's point of view, the collaboration outside the network could include other organizations e.g. NSI's or universities. For the moment there are two potential partners for collaboration: one is between Carl-Erik Särndal and researchers at University of Tartu, this is a promising theoretical collaboration in the adaptive design area. The second is to include bachelor or master-students from Stockholm University in selected BADEN-studies.

6. References

- Axelson, M., Fjelkegård, L., Lundquist, P. and Werner, P. (2013). Nonresponse work at Statistics Sweden – examples from the past and plans for the future. In ISI proceedings, Hong Kong.
- Karlsson, P. (2015). A simulation technique to assess the behaviour of representativeness and balance indicators. Accepted for presentation at JOS 30th anniversary.
- Lundquist, P. and Särndal, C.E. (2012). Aspects of Responsive Design for the Swedish Living Conditions Survey, Research and Development – Methodology reports from Statistics Sweden, 2012:1, http://www.scb.se/statistik/_publikationer/OV9999_2012A01_BR_X103BR1201.pdf.
- Lundquist, P. and Särndal, C.E. (2013a). Aspects of responsive design. With applications to the Swedish Living Conditions Survey. *Journal of Official Statistics*, 29, 557-582.
- Lundquist, P., and Särndal, C.E. (2013b). Responsive design, Phase II – Features of the nonresponse and applications, Research and Development – Methodology reports from Statistics Sweden, 2013:1, http://www.scb.se/statistik/_publikationer/OV9999_2013A01_BR_X103BR1301.pdf.
- Schouten, B., Cobben, F., Lundquist, P. and Wagner, J. (2014). Theoretical and empirical evidence for balancing of survey response by design? Discussion paper 201415, Statistics Netherlands, The Hague, available at www.cbs.nl.
- Särndal, C.E. and Lundquist, P. (2014a). Accuracy in estimation with nonresponse: A function of degree of imbalance and degree of explanation. *Journal of Survey Statistics and Methodology,* 2, 361-387.
- Särndal, C.E. and Lundquist, P. (2014b). Balancing the response and adjusting estimates for nonresponse bias: Complementary activities. *Journal de la Société Française de Statistique*, 155(4), 28-50.
- Särndal, C.E., Traat, I. and Lumiste, K. (2014). Reducing the response imbalance: Is the accuracy of the survey estimates improved? (Manuscript sent to journal)