Statistiska centralbyrån
Statistics Sweden

Advisory Scientific Board
Ingegerd Jansson
Suad Elezović

Notes
November 12, 2013

Meeting with the Advisory Scientific Board of Statistics Sweden November 12, 2013

Board members

Stefan Lundgren, Statistics Sweden, chair
Mats Wadman, Statistics Sweden, co-chair
Lilli Japec, Statistics Sweden, co-chair
Ingegerd Jansson, Statistics Sweden, secretary
Suad Elezović, Statistics Sweden, secretary
Professor Jan Björnstad, Statistics Norway
Professor Sune Karlsson, Örebro University
Professor Frauke Kreuter, University of Maryland
Professor Edith de Leeuw, Utrecht University
Professor Thomas Laitila, Statistics Sweden
Professor Xavier de Luna, Umeå University
Professor Lars Lyberg, Stockholm University
Professor Daniel Thornburn, Stockholm University

Other attendees

Martin Axelson, Statistics Sweden
Tiina Orusild, Statistics Sweden
Mikaela Järnbert, Statistics Sweden
Lina Fjelkegård, Statistics Sweden
Pär Karlsson, Statistics Sweden
Frida Videll, Statistics Sweden
Eva Elvers, Statistics Sweden
Frank Weideskog, Statistics Sweden
Krister Näsén, Statistics Sweden
Kalle Kristiansson, Statistics Sweden
Anette Björnram, Statistics Sweden
Annica Isaksson, Statistics Sweden

Stephanie Eckman, IAB

Current issues at Statistics Sweden

Speaker: Stefan Lundgren

Stefan Lundgren informed the Board about the most important activities at Statistics Sweden, such as:

- The Swedish government has proposed a bill with changes to the Official Statistics. Act following the review of SCB by Bengt Westerberg a year ago.

    - Former guidelines replaced by quality criteria specified in Code of Practice.
    - Legal requirement that municipalities have to give SCB access to all data we need for the official statistics and for European statistics.
- A new bill to the government will be sent regarding financial statistics.
- Requirement from researches to have better access to micro data. An investigation is going on and the result will be published next summer.
- Statistics Sweden experience big problems with non-response in the surveys, especially the Labour Force Survey and the costs are increasing. The Data collection department is monitored to find out how to improve the results.
- SCB is planning to outsource parts of the interviews to private companies.
- Information about the fact that Deputy Director General Mats Wadman is leaving Statistics Sweden December 1. The recruitment procedure will start as soon as possible.

Reply to recommendations

No recommendations and no answers

Big Data

Speakers: Ingegerd Jansson and Annica Isaksson
Discussant: Frauke Kreuter

Summary of presentation

'Big Data' has recently come forth as a new and hot buzzword in the business world. The possibilities of the enormous amounts of data created by modern technology such as Twitter, Facebook, GPS, mobile devices, etc. seem endless. The data are already out there, a gold mine ready to explore, all that is needed are the right tools to chop, dig, and refine the data into useful business information. The challenge was quickly taken by IT people, and tailored methods for processing, analysing, and presenting the data have started to develop: Big Data Analytics. The statistical community has also been alerted, and recent statistical conferences have included activities dedicated to various aspects of Big Data and Big Data Analytics[1]. Early investigations of Big Data use in official statistics breathe optimism. Dunne (2013) notes that in the current economic environment, the Central Statistics Office of Ireland is challenged with doing a lot more with less, and Big Data has the potential to deliver this. At Statistics Netherlands, Daas et al. (2013) claim that "when produced in a methodologically sound manner, official statistics based on Big Data can be cheaper, faster, and more detailed than the official statistics known to date". Said authors expect to see some Dutch official statistics based on Big Data in the coming years.
A core question for Statistics Sweden is what view we, as a producer of official statistics, should take on Big Data. At the opening session of NTTS in March 2013, Robert M. Groves gave a keynote address on Big Data in official statistics. He suggested three options for official statistics production: Ignore Big Data, destroy all official statistical systems and replace them with Big Data, or blend Big Data with traditional sources. In his opinion, the first two options are largely unacceptable. In this paper, we tentatively propose Groves' third choice, using Big Data to enhance or partly replace traditional data sources, as a possible course to take for Statistics Sweden. To go a step further and replace our current

productions systems in order to allow major parts of the official statistics production to rely solely on Big Data would, in our opinion, pose great challenges that require thorough investigation.
Additional ways of using experiences gained from Big Data Analytics might exist. For instance, results from Big Data Analytics performed by others can signal emerging needs for new official statistics, perhaps produced in new ways. They may also give new ideas on how to utilize the vast amounts of data already available at Statistics Sweden through traditional means of data collection. Thus irrespective of how Big Data is or will be used by Statistics Sweden in the near future, it will be of utter importance to actively follow and be prepared to utilize the knowledge gained by others in the development of Big Data Analytics.
As noted for instance by Dunne (2013), there is a need for National Statistical Institutes (NSIs) to position themselves to engage with Big Data. In this paper, we aim to address some of the challenges Statistics Sweden faces regarding Big Data, and provide a starting point for a Board discussion on how Statistics Sweden should proceed in this matter.

Questions to the Board
We are interested in how the Board looks upon current and potential uses of Big Data at Statistics Sweden. Do you think these kinds of data will have a role to play in the official statistics production, and if so, what role should it be? Also, how does the Board think Big Data should be defined in the context of official statistics?
We welcome the views of the Board on the various methodological challenges connected to the use of Big Data in official statistics production: the challenges mentioned in Section 2.2 and others that we might have failed to identify. Which issues do you think are most serious, do you think they are possible to handle, and do you think it is worthwhile to try?
Statistics Sweden will keep monitoring Big Data activities elsewhere, and hope to collabourate with other countries and within the EU. Does the Board have other suggestions on how Statistics Sweden should proceed? Some possible activities (proposed for instance in Dunne 2013) are pilot projects, exploration of methodological problems in partnership with universities, and other training activities that develop relevant competencies and methodologies. If you support these suggestions, what pilot projects does the Board think would be useful, what methodological problems in particular should be addressed, and which training activities would be relevant?

Discussion

Frauke discusses the main characteristics of Big Data: Organic or by product, massive, messy and irregular. She addresses the question whether the administrative data may be considered as Big Data. It is not useful for an NSI to include admin data in Big Data. Challenges identified by SCB are many: lack of knowledge about the data generating process, lack of influence over the data generating process, lack of control over stability of the data genesis, and lack of information within the data. There are some examples of Big Data use at SCB, e.g. Consumer Price Index- cash registers, Balance of Payments- data from credit card companies, SPOT satellite data and some more. A useful approach could be to investigate whether the respondent burden could be reduced by means of using an alternative data source. Furthermore, the statistical process control might be improved since the literature in this area is full of ideas of how to use Big Data. It is important to include in-house created process data.
Collaboration with other actors outside the official statistics could be beneficial. National Statistical Institutes could learn from the Big Data community about

handling of large files, careful sampling, tolerance for inexact measurements and real-time processing. On the other hand, the Big Data community could learn from the survey community about the total error framework, causation, transparency and metadata, data protection and data access. Frauke commended SCB for addressing the issue of Big Data.

Other issues raised during the discussion

- There is a risk that producers of statistics could be replaced by other actors.
- The Dutch (Netherlands Statistics?) developed statistics on vacancy on labour market, web-based and updated daily.
- Statistics Sweden has to be a part, otherwise other companies will take over.
- Big Data is just another data source that can be used. Not so dramatic as we usually perceive it.
- Big Data is an alternative to what we usually do.
- Possibility to use Big Data for generating hypothesis.
- Important to be involved.
- Useful to Consumer Price Index.
- New statistics on Labour Market that couldn't be done before.
- Use to show a modern image.
- Look for alternative data sources, e.g. to meet declining response rates.
- Consumer Price Index more efficient.
- Be creative, think of other data sources of information, as a complement or to replace.
- Complement to "old" sources.
- Think outside the box.
- Methodology is a big challenge.
- SCB need to look into what Big Data can be. Communication is a problem. SCB used to be a main player, now we got competition from others. Problem with non-response. How do SCB argue for our surveys with a big non-response? How do we proceed?
- Need of statisticians and other competences, such as computer scientists.
- Training is needed.
- Communication, more skills, graphical journalists.
- Uncertainty is not communicated well enough, the same goes for privacy preserving measures.
- Work more with analysis.
- SCB could improve by collabourate with other actors, teaming up with universities. SCB can do more in the area of Big Data.
- SCB need various skills and should cooperate with skilled people outside the organization. One field can learn from another field.

Conclusions from Ingegerd and Annica: Good to get a confirmation of how SCB works and that SCB is working in the right direction.

Measurement error, current issues at Statistics Sweden

Speakers: Lina Fjelkegård, Krister Näsén and Pär Karlsson
Discussant: Stephanie Eckman

Summary of presentation

Statistics Sweden presented the main results from two different projects that has been carried out at Statistics Sweden 2012/2013 within the field Measurement Error:

During 2013 there has been a project, *Measurement error project*, with an aim to survey methods for assessing measurement errors and assess their potential usefulness at Statistics Sweden. At the scientific board meeting we wish to discuss the identified methods and hope to find out if the projects has misinterpreted or left out any methods.

Statistics Sweden has done a study on measurement errors and their causes in the Swedish Labour Force Surveys (LFS). The method chosen for the study was gold standard re-interviews that were recorded in order to get the most accurate measurement. This method for estimating the measurement errors is quite costly and labour intensive and to perform a study like this with great frequency might not be feasible but the amount of information gained from this approach might justify its use.

The second part of the Measurement error project was to produce concrete examples of some methods as applied to surveys at Statistic Sweden. Markov Latent Class Models can assess measurement errors in a panel study, without the need for extra interventions such as re-interviews. As the Labour Force Survey had recently done a re-interview, there was an opportunity to do the Markov latent class analysis and compare the result with the results of the re-interview. Markov latent class analyses opens up the possibility to keep track of the measurement error process over time, in fact it would be a control chart. At the scientific board meeting we wish to present some of the results and discuss how this methods can be used at Statistics Sweden in the future.

Discussion

Stephanie Eckman discussed the issues raised in the three submitted papers about the measurement errors at Statistics Sweden.
The first issue was about the methods for assessing measurement error. One question was concerned with randomization in experiments- should it be done within or between interviewers? Furthermore, the way of assigning the cases before data collection was discussed and one recommendation was to implement assignment “on the fly” implying trust on data collectors and to assign only after agreement to participate. Record check is difficult because of mismatch between administrative data and survey questions. One solution is to use record check with experiments. Survey quality prediction is generally a difficult issue. It is necessary to estimate share of variance in answers due to interviewers. Also, there is a need to identify problematic interviewers for additional training, reassignment, separation etc.
The second topic was about implementation of gold standard re-interviews in LFS as a method to reduce measurement errors. Recommendation was to give more information about the implementation details: the response rates in wave 1 and wave 2, distribution of time between two interviews and how the interviewers were “hand-picked”. Stephanie mentioned that IAB (Institute for Employment Research in Germany, where she currently works) has data to do related analysis.
Concerning the paper about the Markov LCA application to LFS, Stephanie pointed out that the problem with differentiating between unemployment and not in labour force was still the problem. Re-interviews could tell us why but

Markov LCA could perhaps be used to evaluate new questions without repeating expensive re-interview procedure. Stephanie concluded that the presented papers were an important contribution to understanding measurement errors in important surveys. They are benefits to Statistics Sweden but also to a larger official statistics & survey community.

During the discussion several more issues were brought up:

- Question of money, costly with re-interviews.
- Seasonal patterns.
- Are newly unemployed people more willing to answer questions then people who have been unemployed for a long time?
- Misclassification is an explanation to big discrepancies.
- Compare different methods.
- Difficulties to explain errors.
- Younger people have difficulties to understand the difference between temporarily and permanently employed.
- Problem with monitoring the interviews, is it a bigger problem for the interviewer or the interviewee. Is there a legally problem regarding the working environment (?) with the monitoring? Discuss with other statistical institutions how they experienced these types of problems.
- False interviews exists – therefore supervised interviews are necessary.
- Methods investigated are mostly for household surveys and individuals so far. Editing, debriefing and process data are also useful for business statistics.
- Actions to quantify and prevent variability and bias, but what about actions to correct for it? Modelling necessary.