![Tilastokeskus logo]

Data Collection

Merja Kemppainen   14.8.2017

Quality measurement and quality management frameworks:

Ensuring the data collection quality of enterprise and household surveys – an overview and new ways

Group of experts on Business Registers, Paris 27- 29 September 2017
Merja Kemppainen (Statistics Finland)
merja.kemppainen@stat.fi

Abstract:

Statistics Finland collects data with inquiries and interviews from enterprises and individuals, when the necessary data cannot be obtained from the existing registers or elsewhere. According to the ESS quality assurance framework QAF, data collection should be routinely monitored and revised as required. A procedure to follow up nonresponse is required. The Data Collection Department at Statistics Finland annually prepares a report on response rate trends and the quality management of direct data collection. These reports – in even years, enterprises, in odd years, households – are an assessment of the statistical survey situation, a short overview of data quality reports, a review of the fieldwork efforts, possible administrative changes in data collection like the reform of the Statistics Act, and a review of new potential data collection tools and innovations. In this paper we focus on a general overview of data collection quality and those fieldwork efforts which are made to ensure quality in both enterprise and household surveys. There were 10 direct official household and about 50 direct enterprise surveys at Statistics Finland in 2016. In the previous household nonresponse review of Statistics Finland (2014) we listed 20 ideas and in the enterprise nonresponse report of Statistics Finland (2016) we listed 22 initiatives that can help in improving the response rate and data collection quality. Additionally, we present two new ways of controlling the data collection process. These are response rate visualization in a survey like the labour force survey and the mobile data collection possibility in the survey on goods transport by road. In mobile data collection, Statistics Finland produced a new mobile phone friendly questionnaire for lorry-drivers. The production of reliable statistics requires a wide array of data collection efforts and management. The quality of statistical data collection is a combination of the above-mentioned efforts to monitor and improve the data collection process and tools.

50 enterprise surveys with 150,000 responses

Statistics Finland collects data with inquiries and interviews from enterprises and individuals, when the necessary data cannot be obtained from existing registers or elsewhere. According to the ESS quality assurance framework QAF, data collection should be routinely monitored and revised as required. A procedure to follow up nonresponse is required. In this article we focus on a general overview of data collection quality and the practical fieldwork efforts made to ensure quality in both enterprise and household surveys in the data collection phase GSBPM 4 at Statistics Finland.

The Data Collection Department at Statistics Finland annually prepares a special report on response rate trends and the quality management of direct data collection. These reports – in even years, enterprises, in odd years, households – are an assessment of the official statistical data collection situation, a short

Data Collection

Merja Kemppainen   14.8.2017

overview of data quality reports, a review of the fieldwork efforts, possible changes in data collection and a review of new potential data collection tools and innovations.

There were 10 direct official household and about 50 direct enterprise surveys at Statistics Finland in 2015. About 150,000 accepted responses from enterprises and 140,000 households supplied information to Statistics Finland in 2016. Work has been carried out to promote the use of internet questionnaires both in household and especially in enterprise surveys. Statistics Finland's aim is to implement mixed mode data collection in all social surveys from 2017 onwards. Using Internet questionnaires alongside telephone interviews will be the outcome of numerous projects and extensive development.

Among enterprise surveys, 86 per cent of the respondents answered by web-questionnaires whereas the remaining enterprises responded using other means, such as paper questionnaires, phone interviews and automatic reporting during 2015.

The mean weighted nonresponse rate in these business surveys at Statistics Finland was 22 per cent in 2015. The amount of eligible responses were between 18 (Fuels in manufacturing) and 21,435 (The sales inquiry). There were 11 enterprise surveys where the response rate was 100 per cent.

Among the largest business surveys the nonresponse rate in the monthly Sales inquiry was 8 per cent (n=12 x 2,000), one per cent (n =12 x 885) in the monthly inquiry on volume index of industrial output, 26 per cent (n= 6,900) in the annual financial statements inquiry for enterprises, 44 per cent (n=18,000) in the business register inquiry for single-establishment enterprises, 22 per cent (n=2,500) in the business register inquiry for multi-establishment enterprises.

The largest nonresponse rates of enterprise surveys at Statistics Finland were found in the statistics on agriculture (44 per cent), the data collection on occupations (48 per cent) and in business register inquiry for single-establishment enterprises (44 per cent). Enterprises' obligation to provide data in these surveys are based on the Statistics Act.

A general view of nonresponse in enterprise surveys is not so clear. In some cases there has been a slight trend of increasing nonresponse. For example, in the annual financial statements inquiry for enterprises the nonresponse rate increased from 12 per cent to 26 per cent between 2010 and 2015 (collection year). On the other hand, in many business surveys the rate of response is stable. In household surveys, the nonresponse trend has been steadily growing. Over the past 15 years, response rates have declined in household surveys by about one per cent per year (e.g. Labor Force Survey).

Quality management in statistical household and enterprise surveys

Recently, Statistics Finland produced two reports describing nonresponse in its business and household surveys and listed ideas on how to decrease it [1, 2].

Quality management in household data collection

In the household nonresponse review of Statistics Finland (2014) we listed 20 ideas to increase response rate and data collection quality (GSBPM 4). They were:

1. Profiling respondents (targeted communication with respondents, responsive design).

2. Deep analysis of reasons for nonresponse (extensive background information about respondents, how to motivate those who refuse).

3. Analysing survey paradata (best fieldwork practices, e.g. sample unit transfers from one interviewer to another, timing and setting the pace between the survey mode).

4. Interviewer and respondent feedback from earlier rounds (summary of problems and hints).

5. Solving problem situations rapidly (e.g. IT problems).

6. Easy to log in mixed mode and web surveys (sign-in, short hyperlink, passwords, e-identification).

7. Re-organise survey mode (face-to-face, telephone, web and mixed-mode).

8. Survey communication plan (for Internet, social media and printed media).

9. Taking holiday periods into account.

10. Reducing respondent burden by developing and testing survey questions and concepts.

11. Renewal of the advance letters (harmonisation, modernisation, avoiding officialese).

12. Motivation letter for those who refuse to answer (targeted letters).

13. Making a survey leaflet (emphasising the usefulness of the survey).

14. Respondent incentive or a lottery (does it do more harm than good?).

15. Train and motivate interviewers.

16. Response rate goal for the whole survey and single interviewer (field work monitoring).

17. Having a break in contacting respondents during the field work period (between contacts).

18. Separate and power-operative nonresponse work (nonresponse groups and pairs, selected interviewers).

19. Help desk services (none, office hours or 24/7).

20. Add more available survey languages (advance letter, questionnaire, interviewer language, distance interpretation).

Quality management in enterprise data collection

In the enterprise nonresponse report of Statistics Finland (2016) we similarly listed 22 initiatives that can help in improving the response rate and data collection quality (GSBPM 4).

1. Expand a mandate to collect data for statistics (e.g. big data, renewing the Statistics Act).

2. Put penalty payment or administrative fixed penalty into operation for refusals.

Tilastokeskus

Data Collection

Merja Kemppainen   14.8.2017

3. Concentration of the data collection know-how of experts in the field.

4. Harmonisation of different surveys.

5. More automatization in business surveys.

6. XBRL-reporting (eXtensible Business Reporting Language) instead of direct data collection.

7. Unification of business surveys (one survey instead of two or more).

8. Special communication work with large enterprises (18 consolidated corporations in Finland).

9. Collate nonresponse cases in the informant register and analyse them.

10. Developing and improving survey instruments and control.

11. Centralising electronic data collection (administration).

12. Coordination of sampling (decrease response burden for one enterprise).

13. Take advantage of the statistical business data system (so called YTY enterprise system and IT architecture).

14. Follow-up the effective field work and call-back (automatization of follow-up).

15. Centralising follow-up calls.

16. Special identification system and data exchange layer for answering surveys (e-identification model and easy log in).

17. Data collection service for enterprises (informing enterprises in advance about future data collections, their timing, responding becomes easier).

18. Planning and testing questions and questionnaires (SurveyLaboratory).

19. Device-independent data terminal equipment (tablet, smartphone, laptop, PC).

20. Survey communication plan for data collection (before field work and during it, feedback from and especially to respondents).

21. Better advance letters, help-desk numbers, targeted advance letters and advertising.

22. Quality marks in data collection materials (to identify official statistics).

Shared and specific features

The lists above are not inclusive. The lists include both existing practices and new ideas, which are yet to be implemented. Can one find shared and specific features in quality management in these ten household and 50 enterprise statistical surveys?

One difference is in the use of background variables. Quality descriptions differ somewhat between household and enterprise surveys. In household surveys, it is more common to exploit and report more closely and deeply the background information of the collected data. Enterprise survey quality reports are more non-specific. Follow-up of the effective field work is more detailed in the household surveys than in business surveys in the case of Statistics Finland. The household survey application for field work follow-up is more complete than in the enterprise data collection system. Instead, the advantage of the statistical data

system and the data architecture is more sophisticated in the case of business surveys.

The mandate to collect data for statistics differs between household and enterprise surveys in Finland. In household surveys, responding is voluntary and in business surveys participation is obligatory. There are no plans to change this configuration. Instead, a renewal of the Finnish Statistics Act is currently underway. The interest is to expand its data collection authorities in order to access new data sources. Reshaping Statistics Finland's data collection web pages has been valuable both for household and enterprise respondents.

New ways of controlling the data collection process

In this chapter we introduce two new ways of controlling the data collection. These are response rate visualisation [1, 3] in a survey like the labour force survey and the mobile data collection [4, 5, 6,] possibility in the survey on goods transport by road.

Response rate visualisation in the labour force survey

The quality of statistical data is a combination of numerous efforts like design activities, set up collection, checking and cleaning data and other statistical production processes. This includes evaluation and monitoring of the quality of the statistical process. Response rate visualisation is one potential evaluation tool to monitor data collection quality. There are three different tools for response rate visualisation: PxWeb tables and graphics, Verti maps and Statistics Explorer animations.

It is possible to visualise nonresponse by different time periods, for example by month, quarter and year. Background variables are age, gender, education and region. Net nonresponse is divided into reasons of nonresponse: refusal, failing to reach and other reasons. These can be studied by background variables as well as net nonresponse, gross response and over coverage.

The visualisation tools are not to be used just this one time but continually once the macros are generated in the survey production process. The animations of the application are based on statistical data stored in the PxWeb database. The data comes directly from the data processing system. In Statistics Explorer, the database allows the user powerful and versatile options to further process, crop and save the data. The survey nonresponse data can be viewed as maps, scatter plots, histograms, and time graphs or data tables. The variable data are available from 2008, so the data can also be viewed through an animated time series.

In this data collection visualisation application you can make response rate and paradata tables from the PxWeb database, make graphics, maps, animations and save or send them by email as response rate 'stories'. Slow changes in data collection quality are hard to detect.

This application helps discover what is going on in data collection quality changes and to react when needed. Studying the structure and development of nonresponse will be more important in future. Things such as nonresponse rate (also in small subsets), over-coverage and item nonresponse rate for the most important variables need to be studied and reported further.

Figure 1. Response rate visualisation by PxWeb tables and graphics, Verti maps and Statistics Explorer animations in the labour force survey 2008 to 2015.

Px-Web tables and graphics, Verti maps and Statistics Explorer animations

E.g. net response rate by occupation 2008-2015

Response rate maps, cross tabs and bar charts
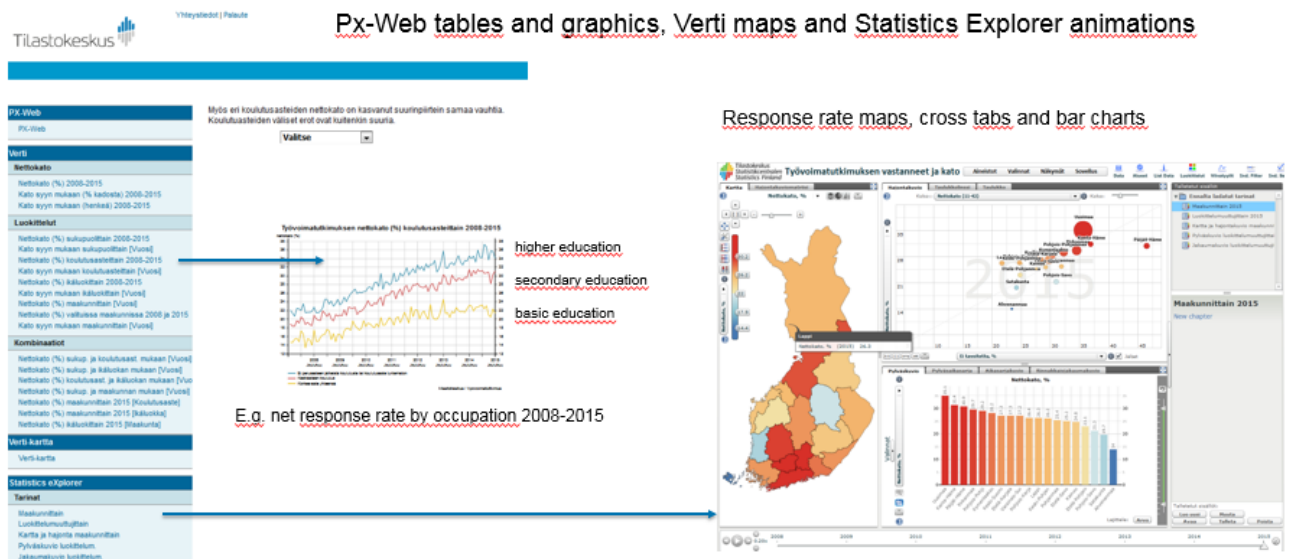
Mobile data collection in the survey on goods transport by road

In the collection of data for statistics on goods transport by road, data are requested from owners of lorries (both companies and private owners) on the lorries and on their use in Finland. In this survey, Statistics Finland produced a new mobile phone friendly questionnaire for drivers. So the drivers can answer when they are, for example, on break. The screen is responsive for mobile, tablet, laptop or desktop. The questionnaire notifies, for example, if there are missing or false answers. Only required and relevant questions are shown. After these changes, the response rate improved, although there was no more paper questionnaire available. Moving to a mobile data collection was cost saving. There was less expenses for printing and postal charges and less human resources were needed for manual entry. The statistics production process is now faster, the data is in the database faster and reminders are sent faster. However we noticed that small businesses are slow in the adoption of new technology. More information is needed on that.

Figure 2: Mobile friendly questionnaire in the survey on goods transport by road (in Finnish)

Data Collection

Merja Kemppainen   14.8.2017
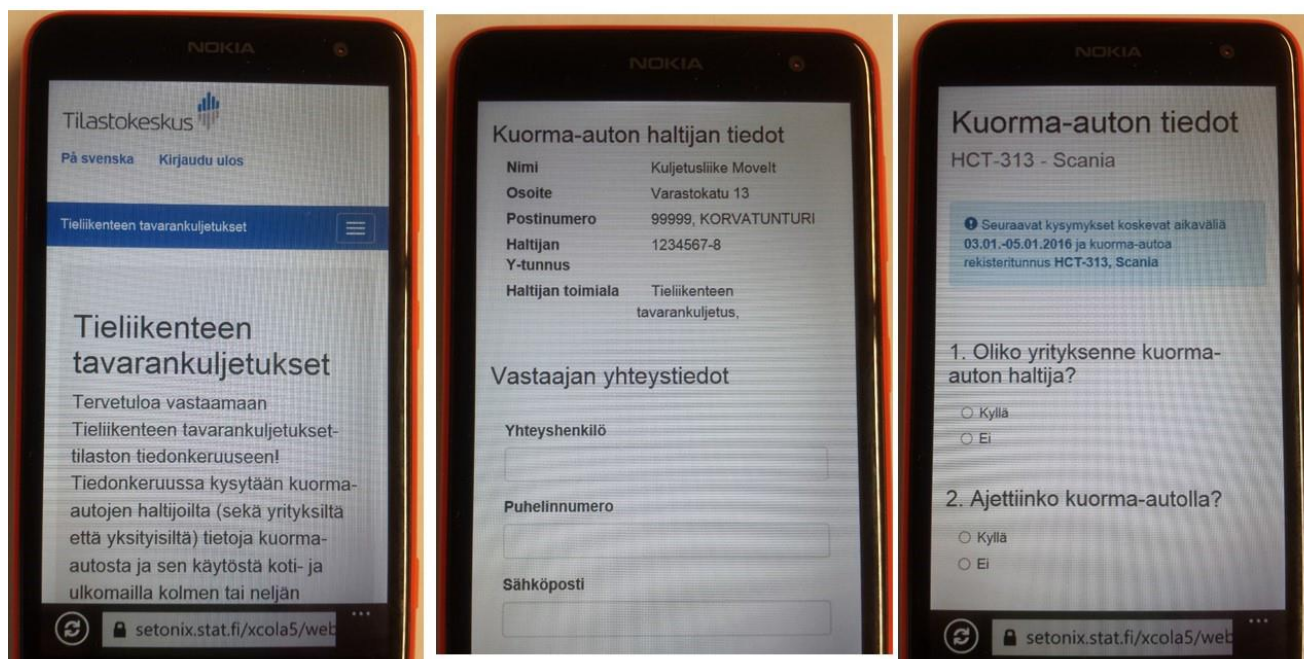
Conclusion

Over the past 15 years, response rates have declined in household surveys by about one per cent per year (e.g. LFS). In some surveys, the trend has been sharper. In some cases, there has been a trend of increasing nonresponse also in enterprise surveys but the trend is not as sharp as in household data collection. In many cases, enterprise survey nonresponse has remained unchanged. The production of reliable statistics requires a wide array of data collection efforts and management. Statistical data collection quality is a combination of the above-mentioned efforts to monitor and improve the data collection process.

VII. References

[1] Virtanen, Vesa & Harjunmaa Hanne (2016): Yritys vastaajana – yritystiedonkeruiden kato vuonna 2015. Enterprise as a respondent – Response rate trends in enterprise surveys in 2015. Internal report.

[2] Lehtinen Milla & Virtanen Vesa (2014): Annual Nonresponse Review 2013 – Individuals and Households. Statistics Finland. Data Collection.

[3] Huuhko Kim & Virtanen Vesa (2014). Statistical visualization of response rates. Data Collection Seminar 2014-10-15. Statistics Finland.

[4] Simonaho Mikko (2016). Data collection of Road freight transport. Presentation at Nonresponse in Enterprise Surveys at Statistics Finland -seminar.

[5] Koukkula Reeta. Developing the data collection of Road freight transport. Nordic meeting in the contact network for transport statistics. 19.11.2015.

[6] Rouhunkoski Jussi & Godenhjelm Petri. The Relationship between Response Burden and the Mode of Data Collection in Business Surveys: Experiences from the Finnish Road Freight Survey. The Fifth International Conference on Establishment Surveys in Geneva.

[7] Virtanen Vesa – Sarén Nina (2016). Non-response Review 2016: Individuals and Households. Internal report. Statistics Finland.