Handling Nonsampling Errors— Case Salo

Maria Valaste[1,2]

[1]University of Helsinki, e-mail: maria.valaste@helsinki.fi
[2]The Social Insurance Institution of Finland (KELA), email: maria.valaste@kela.fi

Abstract

In real life nonsampling errors are almost inevitable. This paper concentrates on nonsampling errors. A research project "Sudden structural change – case study of Nokia-city Salo" from the Social Insurance Institution of Finland is discussed.
Keywords: survey, nonsampling error, nonresponse error, coverage error, measurement error

1 Introduction

Salo is a middle size town in South-Western Finland approximately 50 kilometers from Turku and 100 kilometers from Helsinki. Until 2012 the assembly factor of Nokia mobile phone company was situated in Salo. It employed more than 4000 persons and was the largest private employer is the area.

In the summer 2012 the factory was closed down causing Salo to become an area of sudden structural change. It has already since 2009, when some of the major subcontractors of Nokia were transferred to Asia, received millions of euros in order to minimize the negative effects of sudden structural change.

One aim in the research project is to follow the inhabitants of Salo and their well-being for several (approximately 10) years in order find out how they cope with the sudden structural change and its effects (Ylikännö & Kehusmaa, 2015). First baseline survey was conducted in spring 2013 and the second follow up survey in spring 2015. Currently the third survey is planned.

As expected, the survey data was not complete. E.g. the both surveys consisted nonresponse. This paper will focus on nonsampling errors.

2 Nonsampling errors

In a perfect case the variable of interest is measured on every unit in the sample without error, so that errors in the estimates occur only because just part of the population is included in the sample. Such errors are referred to as sampling errors. (Thompson, 2012). In real life nonsampling errors may also arise.

Groves (1989); Alwin (1991, 2007); de Leeuw *et al.* (2008); Groves *et al.* (2009) specify four sources of error in surveys: coverage error, sampling error, nonresponse error and measurement error. Most important types of nonsampling errors are nonresponse, coverage errors and measurement errors (de Leeuw *et al.*, 2008). Lehtonen & Pahkinen (2004) also adds to this list processing errors.

Nonresponse error

Nonresponse error occurs when some of the sampled units do not respond and when these units differ from those who do and in a way relevant to the study (de Leeuw *et al.*, 2008). There are two types of nonresponse in surveys: unit nonresponse and item nonresponse. Unit nonresponse is the failure to obtain any information from an sample unit. Item-nonresponse refers to the failure to obtain information for one or more questions in a survey, given that the other questions are completed. (de Leeuw *et al.*, 2008).

The methodologies for handling unit non-response and item non-response can differ but in both cases the reasons for missing values has to be investigated. Usually indicator variable is created for unit response or item response and missingness rates and descriptive statistics are computed.

Statistical weighting can be used to make the sample resemble the population with respect to some characteristics. E.g. post-stratification is a basic calibration method for to reduce the bias due to unit non-response. In order to create post-stratification weights an auxiliary information for specified subgroups of the population is required. The weights of the sample units is adjusted to match the totals within the specified subgroups. The subgroups are called post-strata, and the statistical adjustment procedure is called post-stratification.

Coverage error

In surveys two types of coverage errors may exist: undercoverage and overcoverage errors. An undercoverage error arises when some population elements are not included in the sampling frame. An overcoverage error is present when a unit from the target population appears more then once in the sampling frame. A good coverage of the frame population can guarantee a low coverage errors.

Measurement error

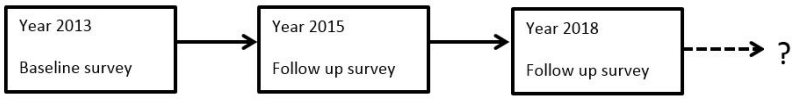
A measurement error is a lack of measurement precision due to weakness in the measurement instrument. Carefully planned and tested measurement instruments can reduce measurement errors.

3 Case Salo

First baseline survey was conducted in spring 2013 and the second follow up survey in spring 2015 (Figure 1). The data of the first survey study was gathered from the mailed questionnaire which was distributed in spring 2013. The questionnaires were distributed to everyone living in Salo and representing the following birth cohorts: 1961–1963, 1971–1973, 1981–1983, and 1991–1993. Of the study population, 2133 subjects completed and returned the questionnaire. The response rate was 29%. Subjects were asked to answer to the questions about their background, educational level and main type of activity, residency, willingness to relocate, use of services, health, social well-being and income. (Valaste, 2015)

The second follow-up survey in spring 2015 utilized the mailed questionnaire but also web survey. Target population was those who participated the first survey and also those who have moved to Salo after the first survey. 2287 subject completed the questionnaire. 1285 subjects participated in the baseline and follow up survey and 1002 subjects were new subjects. The response rate for the follow up survey was 29%.

Figure 1: Baseline survey and follow up surveys.



The frame population for both surveys was determined from the central population register. In baseline survey 51 respondents was not reached and 2 refused to answer. In the second survey 8 refused to answer.

In both surveys older cohorts were more active than the younger cohorts. Females responded more actively than males. Both baseline and follow up surveys included nonresponse. Post-stratification weights was constructed. Auxiliary information (gender and age group) were available and post-stratification weight was created for both surveys. Also a more sophisticated approaches was considered but unfortunately a limited information on the frame population was available.

4 Conclusion

Currently the third survey is planned. As earlier survey rounds, this also will have challenges especially nonresponse issues. How to improve the response rate? And in general, what is a lesson learned from the previous surveys?

References

Alwin, D. F. (1991). Research on survey quality. *Sociological Methods and Research* **20**, 3–29.

Alwin, D. F. (2007). *Margins of error: a study of reliability in survey measurement*. Wiley, New Jersey.

de Leeuw, E. D., Hox, J. J. & Dillman, D. A. (2008). The cornerstones of survey research. In de Leeuw E. D., H. J. J. & D. D. A., eds., *The international handbook of survey methodology*. Erlbaum/Taylor & Francis, New York/London.

Groves, R. M. (1989). *Survey errors and survey costs*. John Wiley & Sons, New York.

Groves, R. M., Jr., F. J. F., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology*. New Jersey: John Wiley & Sons.

Lehtonen, R. & Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. John Wiley & Sons, Chichester, 2nd edn.

Thompson, S. K. (2012). *Sampling*. New Jersey: John Wiley & Sons.

Valaste, M. (2015). Aineisto ja menetelmät. In M. Ylikännö & S. Kehusmaa, eds., *Muuttuva Salo. Kyselytutkimus äkillisen rakennemuutoksen alueen asukkaiden hyvinvoinnista*, chap. 3. Kela, Helsinki, pp. 13–16.

Ylikännö, M. & Kehusmaa, S. (2015). *Muuttuva Salo. Kyselytutkimus äkillisen rakennemuutoksen alueen asukkaiden hyvinvoinnista*. Kela.