Chapter 26

Surveys Without Questions

Jelke Bethlehem
Statistics Netherlands

Our society experiences an ever-increasing demand for statistical information. One way to collect such information is by means of a survey, in which questions are asked about the current or past situation of the objects, historic events, habits, knowledge or behavior. Long years of practice have shown that surveys work. Nevertheless, there are practical problems.

Maybe one of the most important problems of survey sampling is nonresponse, which may result in invalid estimates of population characteristics. Another problem is the survey questionnaire. It is a poor measuring instrument compared to measuring devices used in physics. Many things can go wrong in asking questions about opinions, behavior and historic events, and these results in measurement errors. And, of course, surveys are time-consuming and expensive. This all raises the question: can we collect data without asking questions?

Are there alternative ways of getting information? One approach could be to use data available in registers and other administrative sources. This chapter describes some of possibilities and difficulties of the use of such register data. Particularly, quality issues are important. It also describes some of the pitfalls of combining register data from various sources, and of combining register data with survey data.

GLOSSARY OF KEY CONCEPTS

Administrative register. A register that is primarily used for administrative purposes, that is, a register containing information on objects that is required for administrative or other governmental action concerning individual objects.

Key variable / Identification variable, Variables that appear in different data sets, and that are used to link a record of an object in one data set to a record of the same object in another data set.

Mass imputation. A form of imputation in which a large amount of missing values for individuals are replaced by synthetic values, computed using nonmissing information for these objects.

Measurement error. An error that occurs when the respondent does not understand the question, or does not want to give the true answer, or if the interviewer makes an error in recording the answer. Also, interview effects, question wording effects, and memory effects belong to this group of errors. A measurement error causes a difference between the true value and the value processed in the survey.

Nonresponse. The phenomenon that individuals in the selected sample do not provide the requested information, or that the provided information is useless.

Nonsampling errors. Errors that even occur if the whole population is investigated. Nonsampling errors are errors made during the process of recording the answers to the questions.

Primary data analysis. Statistical analysis of a data set that has specifically been collected for the study at hand.

Register. A register is a collection of data on a well-defined group of objects. For each individual object, the register contains the values of the same well-defined set of variables. These variables describe the state of the objects at a specific moment in time. A register has facilities to update the information about objects contained in it.

Sampling errors. Errors introduced by the sampling design. They are due to the fact that estimates are based on a sample and not on a complete enumeration of the population. The sample is selected by means of a random selection procedure. Every new selection of a sample will result in different elements, and thus in a different value of the estimator.

Secondary data analysis. Statistical analysis of a data set that has been collected by others for other purposes.

Survey. A study that collects planned information from a sample of individuals in order to estimate particular population characteristics.

Synthetic matching. A form of matching records from two data sets. Records are grouped using the values of a set of identification variables. Within groups, records from both data sets are combined randomly.

Weighting adjustment/Adjustment weighting The process of assigning weights to observed individuals in a survey. The weights are computed such that the weighted distribution of certain auxiliary variables is identical to the population distribution of these variables.