

Analyses of web survey data

*Vesa Kuusela
Social Survey Unit
Statistics Finland*

The data capture process in a web survey differs from that of a postal survey and an interviewer administered survey. The differences have give rise to new types of errors in data and therefore they should be filtered out or otherwise be taken into account in the data analysis of the survey.

In an interviewer-administered survey the interviewer controls what is entered in the form and he or she also is able to assess the feasibility and reliability of the answers. Usually the obtained data will be ready for analysis after minor checking. Also unit non-response is coded adequately and may be processes using available methods. Item non-response remains (usually) in a moderate level.

In a postal survey, like in a web survey, the data capture process cannot be controlled. However, in a postal survey, the forms have to be returned to the institution or company conducting the survey and that requires some effort. On the other hand, sending a static web questionnaire is usually a very simple task and when using an interactive web questionnaire the data is stored all the time. Often in a postal survey the returned forms will be checked before data entry. The validity of the data is assessed (and invalid data filtered) in checking of the forms and partly also during data entry. Therefore, in a postal survey the data to be analysed is composed of answers, which can be considered given seriously. In a web survey one cannot be sure of that. For example, some respondents may have only browsed through the questionnaire and entered the answers only to reach the end of the form.

In a web survey, only partly filled forms, interruptions, is a new phenomenon which practically nonexistent in interviewer administered surveys and postal surveys, because interviews are interrupted very seldom and only partly filled mail questionnaires are not sent. Especially, in and interactive web questionnaire interruptions may become a problem.

The paper discusses different erroneous ways to fill in a web form and in which ways those could be discovered. Results of a data cleaning analysis of a web survey data will also be presented. The web survey was published in a newspaper with a general invitation and without an identification the respondents. The analysis shows that 10% - 25% of returned forms could (should) be disregarded, depending on the criteria for exclusion.