

New developments in randomized response

Peter G.M. van der Heijden and Ardo van den Hout

Abstract

Randomized response has been proposed in the sixties as a tool to collect sensitive information in surveys. The main idea is that people are not asked the sensitive question directly. Instead there is a device (such as dice, cards, spinners) that determines whether the sensitive question is to be answered. For example, in the forced response design two dice are thrown, and people are asked to answer 'yes' if the outcome is 2, 3 or 4, 'no' if the outcome is 11, 12, and answer honestly when the outcome is between 5 and 10. In this way every individual is 'safe' when he follows this procedure, yet it is possible to estimate the probability of 'yes' and 'no' to the sensitive question. Although interest has faded somewhat after much interest in the sixties and seventies, in the Netherlands the randomized response procedure has become a standard tool of application in research funded by the government into rule transgression. Examples are research into rule transgression in the Dutch tax law, individual rent subsidy, social security benefit and food administration. The government subsidizes our research into this area. This research deals with data collection and analysis. In this presentation we give an overview of research into data analysis.

From a statistical point of view an interesting question is how randomized response data can be adequately analyzed. We will show that randomized response data can be considered as a special case of misclassified data, where the probabilities of misclassification are known. There are also links with statistical disclosure control and data mining, where the data are perturbed after they have been collected for privacy protection. It turns out that an important class of research questions concerning randomized response data can be answered with a restricted form of the latent class model, because the sample is of a mixture of people who have to answer honestly and people who are forced to say 'yes' or 'no'. Other research questions involving quantitative explanatory variables can be answered using an adapted form of logistic regression. We give an overview of analytic methods, and discuss problems that come with them, such as boundary solutions, and what happens when there are individuals that do not follow the randomized response procedure.