Split Questionnaire Design for Massive Surveys

Feray Adiguzel*
Erasmus University, Rotterdam, Netherlands adiguzel@ese.eur.nl

Michel Wedel
University of Maryland, College Park, USA mwedel@rhsmith.umd.edu

Companies are conducting more and longer surveys than ever before. Massive questionnaires take more time; induce the use of undesired response styles; increase respondent fatigue and boredom; and result in more nonresponse, item nonresponse, and early break-off. Therefore, in practice, researchers frequently resort splitting questionnaires into parts (or splits) and administering each part to a randomly selected group of respondents. The problem then is how to split the questionnaires efficiently. As an alternative to the heuristic methods that are currently used to split questionnaires, this presentation will discuss a methodology to design the split questionnaire in a way that minimizes information loss. At the first stage, the modified Fedorov algorithm using the Kullback–Leibler distance as a design criterion will be applied to find the optimal splits using estimates from a first wave or pilot study. This design criterion is based on a general mixed data model that accommodates continuous, rank-ordered, and discrete measurement scales. The optimal construction of the split questionnaire design is easy and fast. At the second stage, Markov chain Monte Carlo procedures will be used to impute missing values that result from the design. Split questionnaire designs are generated by selecting either entire blocks of questions (between-block design) or sets of questions in each block (within-block design). The efficiency of split questionnaires generated with the proposed method is compared to multiple matrix sampling designs, incomplete block designs, and a heuristic procedure, using synthetic and empirical Web survey data. At last, a field study indicates that as a result of reduced respondent burden, the quality of data using split questionnaire designs improves.

Key Words: questionnaire design, missing data, Kullback–Leibler distance, modified Fedorov algorithm