

Latent Class Analysis in Survey Methodology

Frauke Kreuter

Joint Program in Survey Methodology

Latent class analysis is as a powerful tool to estimate the components of total survey error: coverage and nonresponse error as well as measurement error in survey questions. The use of LCA in this context is different from most applications in Psychology or other Social Sciences. While in those fields it is more common to collect multiple measures of constructs of interest, federal data collection efforts prioritize large sample sizes and high response rates over multiple measures. In addition, many of the variables in government surveys have true scores that (with enough time and the appropriate record data) could be measured without error. When multiple measures and external records are not available, latent class analysis can (with some identifying restrictions) nevertheless help to estimate measurement error.

In this talk I will present three studies which set out to examine the effectiveness of LCA to this end. In the first study we use data from a survey of University of Maryland alumni together with alumni records, to evaluate the usefulness of the LCA technique for question evaluation. Our results indicate that the based on the LCA models the same qualitative conclusions would be drawn about the item quality but the quantitative estimates are weaker compared to the record check study. In the second study we compared four different survey question evaluation methods -- expert reviews, cognitive interviews, quantitative measures of reliability and validity, and error rates from latent class models. We examine the performance of each evaluation method in identifying deliberately flawed questions. In the third study LCA models are applied to data from the National Survey of Family Growth (NSFG), focusing on a pair of similar items about abortion that are administered under different modes of data collection. To evaluate these items, we made assumptions about the error rates within various subgroups and compared those estimates to external data on the prevalence of abortion. Here the models yielded relatively low error rates, supporting an argument by Spencer that LCA models may consistently underestimate the actual error rates.