



Introduction to Latent Class Analysis with Applications

Paul P. Biemer*

RTI International, RTP, USA - ppb@rti.org

Latent class analysis refers to a statistical framework for modeling categorical data along with their classification errors. The general latent class model postulates that the true value of a variable is unobservable (latent) while a survey response may constitute a single indicator of this latent variable. The LC model parameters include the target population proportions for a categorical variable to be estimated in the survey and the probabilities of misclassification probabilities (for e.g., false positive and false negative, for dichotomous response variables) for measuring the variable. Survey item reliability and construct validity as well as estimator bias can be defined and interpreted in this context. One advantage of viewing survey classification error model as a LCM is the availability of general software for estimating the model parameters including the error components. However, the assumptions of the traditional LCM can be somewhat restrictive. An even more general model can be obtained by viewing the LCM as a type of log linear model with latent variables. In doing so, a wide range of error structures and error evaluation designs can easily be discussed and analyzed using log-linear modeling notation and methods.

This presentation will provide a general introduction to latent class analysis describing what it is, the types of research questions for which it is ideally suited to answer and some cautions to regard in its applications. LCA for both cross-sectional and panel survey analysis will be discussed. Methods for analyzing complex surveys designs with missing data will also be covered. Several illustrative applications of LCA to survey data will be briefly described.

Keywords: measurement errors, survey data quality, total survey error, log linear models with