Inference in Structural Equation Modeling Using Samples With Unequal Probabilities of Selection and Misspecified Models

Sharon Christ
slchrist@email.unc.edu
Odum Institute USA
University of North Carolina at Chapel Hill
06 Manning Hall, CB #3355
Chapel Hill, NC 27599-3355
USA

Abstract

Model-based inference relies on the assumption that the correct population model has been specified. When population models are correctly specified, parameter estimates are unbiased regardless of the sampling mechanism. However, population models that are misspecified results in misspecification bias in parameter estimates. Misspecification bias is the same regardless of the sampling mechanism when the population model holds for all members of the population (homogenous models). When the population model is different for subgroups of the population and those subgroups are sampled with unequal probabilities of selection, design based estimation using weighted estimators provides unbiased population average parameters for a single population model. Model-based inference using samples with unequal probabilities of selection may bias second order statistics such as standard errors regardless of the proper specification of the population model. Therefore, while parameter estimates are unbiased, hypothesis testing and confidence intervals may be improper. This paper is an analysis of the effects of unequal probabilities of selection on improper inference under both correct and incorrect model specification in structural equation modeling (SEM). Analytics for the effects of unequal

selection on standard error estimates for some SEM parameters are given. Also, a Monte Carlo simulation is presented whereby a SEM population model is used to generate a simulated finite population. Probabilities proportional to size (PPS) samples are drawn from the finite population with varying degrees of unequal probabilities of selection. And correct and incorrect models are fit to the PPS samples using weighted and unweighted estimation. Standard errors, mean square error and confidence interval coverage are compared across the weighted and unweighted analyses. Results vary depending on the degree of selection, the particular model variables related to selection, and the type of model misspecification.