Count regression models that simultaneously model heaping propensity via scaled distributions

James W. Hardin*
University of South Carolina, Columbia, SC – jhardin@sc.edu

Tammy H. Cummings, Alexander C. McLain, James R. Hussey, Kevin J. Bennett
University of South Carolina, Columbia, SC

Gina M. Wingood
Emory University, Atlanta, GA

We present an approach to modeling heaped count data outcomes. Such non-negative count outcomes illustrate a preference in respondents to report counts that are multiples of particular numbers. For example, when asked to report a frequency of behavioral outcomes (e.g., unprotected sexual encounters, illicit drug use) over the course of 6 months, the behavior of some respondents will be to consider the frequency over the course of one month and then report 6 times the monthly recall. Other respondents will faithfully report the number of episodes over the entire 6-month period. The final sample is then comprised of a mixture of two distributions. If we assume that associations of covariates are at the same rate for the two distributions of reporting behaviors, then log-linked count regression models for counts at their intended time periods differ only by an exposure variable. We illustrate an approach that combines these models while simultaneously modeling the likelihood of a given reporting behavior. Finally, we compare our approach to naïve models and to models that treat reported counts on heaping intervals as if they are interval censored.

Keywords: Count data; heaping; mixture model