South Carolina Researchers Develop Multiple Imputation Model to Deal with Missing Values in Bounded Variables

Drs. Marco Geraci and Alexander McLain, faculty members of the Epidemiology and Biostatistics Department at the University of South Carolina’s Arnold School of Public Health, have published a study developing a new multiple imputation approach for bounded variables with missing values. The study was funded by an Advanced Support for Innovative Research Excellence (ASPIRE) grant from the Office of the Vice President for Research. They published their resulting paper in Psychometrika.

Missing data are a common issue in statistical analyses. Multiple imputation is a technique that has been applied in countless research studies and has a strong theoretical basis. There are methodological challenges in multiple imputation when the variable with missing values is bounded. Bounded variables are constrained to take values only on a limited range and are ubiquitous in all fields of research, including education, psychology, and public health. Examples of these variables include measurements of length, weight and volume, psychometric scales, clinical scores, survey questionnaire items, and school grades.

Most of the statistical literature on multiple imputation has focused on unbounded variables, with mostly ad hoc remedies for variables with bounded support. These approaches can be unsatisfactory when applied to bounded variables as they can produce misleading inferences.

In this paper, the authors propose a flexible quantile-based imputation model suitable for distributions defined over singly or doubly bounded intervals. Proper support of the imputed values is ensured by applying a family of transformations with singly or doubly bounded range.

Simulation studies demonstrate that the authors’ method is able to deal with skewness, bimodality, and heteroscedasticity and has superior properties as compared to competing approaches, such as log-normal imputation and predictive mean matching. Drs. Geraci and McLain demonstrate the application of the proposed imputation procedure by analyzing data on mathematical development scores in 5-year-old children from the Millennium Cohort Study, United Kingdom, and data on depression scores from a psychiatric study of the clinical effects of imipramine in depressed inpatients.