

THE DEPARTMENT OF EDUCATIONAL PSYCHOLOGY'S RESEARCH METHODS,  
MEASUREMENT, & EVALUATION (RMME) PROGRAMS AND THE DEPARTMENT OF  
STATISTICS AT THE UNIVERSITY OF CONNECTICUT PRESENT:

# MODELING COARSENEDED CATEGORICAL VARIABLES: TECHNIQUES AND SOFTWARE

**DR. JOSEPH L. SCHAFFER, U.S. CENSUS BUREAU**

Coarsened data can express intermediate states of knowledge between fully observed and fully missing. For example, when classifying survey respondents by cigarette smoking behavior as 1=never smoked, 2=former smoker, or 3=current smoker, we may encounter some who reported having smoked in the past but whose current activity is unknown (either 2 or 3, but not 1). Software for categorical data modeling typically provides codes for missing values but lacks convenient ways to convey states of partial knowledge. A new R package *cvam*: Coarsened Variable Modeling, extends R's implementation of categorical variables (factors) and fits log-linear and latent-class models to incomplete datasets containing coarsened and missing values. Methods include maximum likelihood estimation using an expectation-maximization algorithm, approximate Bayesian and Bayesian inference via Markov chain Monte Carlo. Functions are also provided for comparing models, predicting missing values, creating multiple imputations, and generating partially or fully synthetic data. In the first major application of this software, data from the U.S. Decennial Census and administrative records were combined to predict citizenship status for 309 million residents of the United States.



Dr. Joseph L. Schaffer is Senior Mathematical Statistician for Analytic Modeling in the Research and Methodology Directorate of the U.S. Census Bureau. Before joining the Census Bureau in 2012, he served as Associate Professor of Statistics at The Pennsylvania State University. He received his Ph.D. in Statistics from Harvard University in 1992. His areas of research include statistical methods for missing values, longitudinal data, statistical computing and software development, causal inference, latent class analysis, and combining information from surveys, censuses and administrative records. He has served on numerous advisory panels for the federal statistical system and is a Fellow of the American Statistical Association.

## **Colloquium Access Information:**

Friday, 3/24/2023, 11am ET  
<https://tinyurl.com/rmme-Schafer>  
Meeting # 2621 299 3126  
Password: RMMESTAT

Join by video system: Dial 26212993126@uconn-cmr.webex.com. You can also dial 173.243.2.68 and enter your meeting number.

Join by phone: +1-415-655-0002 US Toll  
Access code: 262 129 93126