

Machine Learning for Inverse Probability Weighting in the American Community Survey

Declining response rates and data collection interruptions are resulting in missing data complexity that traditional missing data techniques used in Census Bureau survey processing may not flexibly capture. At the same time, availability and linkability of administrative records, third party, and previous census/survey data has improved allowing for more informative response propensity models. These developments lend themselves to the study of data-driven enhancements on inverse probability weighting (IPW) methods to adjust for unit nonresponse. We study and compare the use of traditional statistical models and machine learning algorithms applied to complex survey data for model-based IPW nonresponse adjustment using auxiliary sources with multiple years of American Community Survey data. We share various measures for model comparisons, application-specific tuning parameter selection, and visualizations of geographically-differentiated results.

Darcy Morris is a Research Mathematical Statistician in the Center for Statistical Research and Methodology at the U.S. Census Bureau. Dr. Morris' research interests include missing data methods for probability and nonprobability data, categorical data analysis, and multivariate distributions with applications in a variety of economic, demographic, and social topics. She received her PhD in Statistics from Cornell University and a Master's in Statistics from George Washington University, where she is currently a Professional Lecturer in the Data Science Program.