MPSDS/JPSM Brownbag Seminar
November 16, 2022

Utility of Commercial Data for Sampling Population Subgroups: A Case of Health and Retirement Study

Sunghee Lee is a Research Associate Professor at Survey Research Center, University of Michigan. Her research focuses on sampling and measurement issues with hard-to-survey population subgroups as well as racial, ethnic, and linguistic minorities.

Chendi Zhao is a Research Assistant and first-year Ph.D. student in the Program in Survey and Data Science

Anqi Liu is a master's student in MPSDS at the University of Michigan. She works closely with Dr. Sunghee Lee on the Health and Retirement Study sampling.

Abstract

A standard approach for targeting population subgroups in household surveys is to sample general population and then to screen for eligible households. This becomes increasingly costly as the subgroup accounts for a small proportion of the population, which is the case for the Health and Retirement Study (HRS). HRS is a population-based longitudinal study of adults ages 50 and older in the U.S. and maintains its representativeness by adding a new age cohort every 6 years. In 2016, HRS targeted those born between 1960 and 1965 with an additional goal of oversampling racial/ethnic minorities. This group is less than 10% of the population. In order to increase the efficiency of screening, HRS had traditionally used probability proportionate size sampling in its area-probability sample with the age-eligible population size as a measure of size as well as stratification based on the race/ethnicity distribution of area sampling units. For 2016, HRS sampling additionally used stratification at the address level by enhancing the population of addresses in the sample areas with commercial data. This study examines the utility of commercial data for increasing efficiency with a focus on its availability and accuracy by analyzing a dataset that combines sampling frame data, screening data, main survey data as well as external data from the American Community Survey.