

Causal Modeling Using Behavior Risk Factor Surveillance System Data

Yiyang Fan

Behavioral Risk Factor Surveillance System BRFSS Dataset

BRFSS is the system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.

Now collects data in all 50 states. It has around 45% overall response rates (47.9% in 2020) and completes more than 400,000 adult interviews each year.

Probability sample of all households with telephone in each state. Most state sample disproportionately from strata that correspond to sub-state regions.

Weights for each individual are included. The design weight accounts for the probability of selection and adjusts for nonresponse bias and non-coverage errors.

BRFSS Dataset and Proposed research question

Totally 279 columns (feature) and 401,958 rows (sample) for 2020 data.

Columns include demographics, health status, health care access, disability, tobacco use, alcohol consumption, immunization, exercise, sleep, chronic health conditions, cancer, hepatitis, cognitive decline, diabetes, etc.

Is smoking and drinking causally associated with hepatitis?

Any other interesting research questions are very welcome