

Chapter 8: Producing Data: Sampling Key Ideas

Two primary methods of data collection:

- **Observational study:** observing individuals and measuring variables of interest without attempting to influence responses.
 - **Purpose:** descriptive, study of associations
- **Experiment:** deliberately imposing some treatment on individuals in order to observe their response.
 - **Purpose:** investigate cause-and-effect relationships

Confounding is a difficult problem in establishing causation:

- Two variables (explanatory variables or lurking variables) are **confounded** when their effects on a response variable cannot be distinguished from each other.

Sampling methods

Key terms:

- **Population:** the entire group about which we wish to make inference.
- **Sample:** the part of the sampling frame that we actually examine in order to gather information.
- **Sampling design:** the precise procedure used to choose a sample from the population.

Sampling Methods

- **Non-probability sampling methods:** (see below)
 - **Voluntary response sample:** people who choose themselves by responding to a *general appeal*.
 - **Convenience sampling:** sampling the individuals easiest to reach.
- **Probability sampling:** each member of the population has a (theoretically) known chance of being selected for the sample.
 - **Simple random sample (SRS):** a sample of size n chosen from the sampling frame in such a way that every subset of size n has an equal chance of being chosen (e.g., draw n from a hat)
 - **Stratified random sample:** a probability sample obtained by first dividing the population into *strata* (identifiable homogenous groups) and then drawing separate SRSs from the strata. The separate SRSs are combined to form the full stratified sample.

Bias

- **Bias:** systematic error in the way the sample represents the population

Sources of Bias:

- **Undercoverage:** some groups of the population are omitted from the process of choosing the sample.
- **Nonresponse bias:** individuals can't be contacted or refuse to participate
- **Response bias:** behavior of respondent or interviewer causes bias
- **Wording bias:** wording of the question influences the response.