

STATISTICAL THINKING AND DESIGN

Best Practices for Collecting Data

- The sample needs to be large (usually >30 is a good target) and chosen randomly.
 - A sample is a data set that is a part of a larger population of interest. A good sample should be representative of the larger whole. For example, 100 randomly selected students would be a sample representative of all students in a particular school district, whereas 100 students taken from a single school or single teacher would not.
- The sample should be as representative of the population as possible.
 - A population is a data set that includes all elements of a defined group. For example, all students in a particular school district.

Common Survey Biases

- **Household bias:** If you only sample one member of any given household.
- **Non-response bias:** When only a small percentage of people respond to your survey, it's unclear which part of the population is responding.
- **Quota sampling bias:** When pollsters are given free choice in picking respondents. The sample should be random.
- **Response bias:** When people lie in their response to be perceived in a certain way (e.g., dieters don't want to admit they cheated on their diet). Proper survey question wording can avoid this.
- **Selection bias:** Choosing only a particular subset of the population. A famous example is when *Literary Digest* incorrectly predicted Alfred Landon would win the election over Franklin D. Roosevelt in 1936 because the magazine only polled people with phones and cars (in that day, those items generally belonged to wealthy citizens).
- **Undercoverage bias:** If you survey people by calling their home phones, you may miss a lot of subjects who only carry cell phones. This results in inadequate representation for your survey.
- **Voluntary response bias:** These are samples based on subjects who offer to participate. It's biased because people with stronger opinions are usually the ones who volunteer to participate in surveys.
- **Wording bias:** Non-neutral or leading questions may create answers that are under-representative of the population.

Experimental Design Techniques

- **Placebo:** One group gets a fake treatment to see if there is a significant difference between its results and the results of the group that receives the actual drug.
- **Double blind:** Neither the researcher nor the participants know who is getting what treatment.
- **Blocking:** Separating a large study group into subgroups based on a characteristic that may confound results.
- **Interviewer bias:** Certain traits of the interviewer may make the participant uncomfortable, causing them to answer questions dishonestly.