UCLA Statistics

STATISTICS 10
Introduction to Statistical Reasoning

Bingling Wang
Ph.D., Lecturer, Department of Statistics

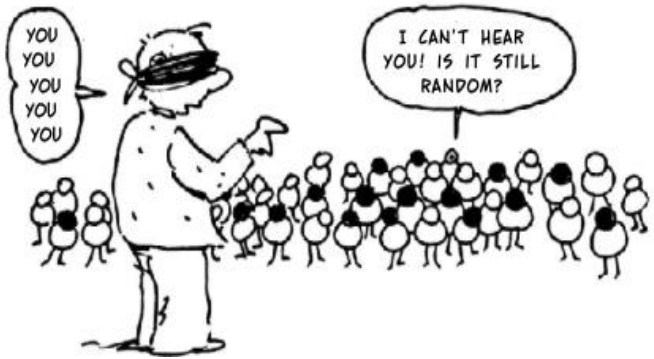
PRODUCING DATA

Producing Data

Two stages

1. **Sampling**
   - For the conclusions to be valid, we must have a **Representative sample** of the population.
   - A sample that is not representative is said to be **biased**.
   - One way to collect a sample that is representative is **Simple Random sampling** (**SRS**).
2. **Study design**
   - Studies should be designed to discover what we want to know about the variables of interest for the individuals in the sample.
   - Anecdotes (stories based one's experience) are not useful for making cause-and-effect conclusions about population.

Bias



A sample is **biased** if it is not representative because of the systematic under- or over-estimation of the values of the variable of interest.

Selection bias

- Undercoverage
- Volunteer bias
- Nonresponse bias

Example: A senator conducted a poll in her state by calling 100 people whose names were randomly sampled from the phone book (mobile phones and unlisted numbers aren't in phone books).

The senator's office called those numbers until they got a response from all 100 people chosen.

Simple Random Sampling

One way that should (but is not guaranteed) give a representative sample is

Simple random sampling (SRS)

A procedure for sampling from a population in which

(a) Every single object of the population is equally likely to be chosen

(b) Every possible sample has an equal chance of being selected.

Steps

1. Define the **target population**
2. Start with a **sampling frame**: a list of every objects in the population
3. Decide on the **sample size**
4. With the sampling frame, **randomly** draw a sample of objects with specified size **without replacement (SRSWOR)**

   -- no object can be repeated in a sample

Example

Suppose you want to determine the musical preferences of all students at your university based on a sample of students.

Here are some examples of possible ways to pursue this problem.

1. Post a music-lovers' survey on the university's pop music forum, asking students to vote for their favorite type of music.

2. Stand outside the Fine Arts Building and ask students around to respond to your question about musical preference.

3. Obtain a student directory with email addresses of all the university's students and send the music poll to every 10th name on the list.

4. Obtain a student directory with email addresses of all the university's students, and randomly sample some addresses and send your music poll to these sampled students.

Study Design

I. Observational study

- Subjects are not assigned to the treatment group or control group by the researchers.
- The subjects in the study are simply observed.
- Researcher do not interfere with the subjects other than taking measurements.

II. Sample survey

A particular type of observational study in which individuals report variables' values themselves, frequently by giving their opinions.

We need to be careful with cause-and-effect conclusion from observational studies alone because of potential confounding variables.

Study Design

III. Experimental Study

- The researcher actively manipulates the treatment variable.
- Subjects are assigned to the control/treatment group.
- At least one treatment variable to manipulate and one outcome variable to measure.
- The outcome variable is measured and compared between groups receiving different treatments.

It is possible to answer questions about population-level causal effects with experiments.

- Well-designed and well-executed controlled experiments are most important ways for answering questions about causality.
- Controlled experiment requires at least a treatment group and a control group.

Exercise

DOES DRINKING COFFEE AFFECT PEOPLE'S SLEEPING HABIT?

A study took random sample of adults and asked them about their coffee drinking and sleeping habits. The data showed that people who drank coffee daily were more likely to go to sleep later than those who didn't.

Identify:

- Treatment
- Outcome
- Observational study or experimental study?
- Can you conclude that drinking coffee causes people to sleep late?

Controlled Experiments

Key features:

1. Sample size

Large sample size ensures that we observe the full range of variability in the objects we study.

2. Random assignment

- Subjects are assigned to the control or treatment group by a **randomization procedure**
- Control for **confounding variables**.
- Balances out differences to make the groups **comparable**. Bias may occur for non-randomized assignment and the results are influenced in one particular direction.
- An experiment where subjects are randomly assigned is called a **randomized experiment**.

Controlled Experiments

Key features:

3. **Blinding --** Knowing what treatment was assigned can bias the study.
   - Researchers should be blind to the assignment
   - Participants should be blind to the assignment

When neither the researchers nor the participants know whether the participants are in the treatment or the comparison group, it is a **double-blinded** study.

The double-blind format helps prevent the bias that can result if one group acts differently from the other because they know they are being treated differently, or because the researchers treat the groups differently or evaluate them differently because of what the researchers hope or expect.

Controlled Experiments

Key features:

4. Placebos

- Anything that seems to be a "real" treatment,
  - e.g., harmless pill, a shot, or some other type of "fake" treatment.
- It is important that the comparison group receive attention similar to what the treatment group receives, so that both groups feel they are being treated the same by the researchers.
- This format controls for possible differences between groups that occur simply because some subjects are more likely than others to expect their treatment to be effective.

Causal Inference

Experimental study:

- In general, with a well-designed experiment we have a better chance of establishing causation than with an observational study.
- However, experiments are subject to certain pitfalls, and there are many situations in which an experiment is not an option.

Observational study:

- Because of the existence of a virtually unlimited number of potential confounding variables, we can never be 100% certain of a claim of causation based on an observational study.
- However, A well-designed observational study may still provide fairly convincing evidence of causation under the right circumstances.

Counterfactual causal inference is one the most important statistical ideas of the past 50 years!

(Paper by Andrew Gelman, Aki Vehtari, 2020)

Exercise

CAN BRAIN GAMES IMPROVE INTELLIGENCE SKILLS?

Brain-training video games, such as Nintendo's Brain Age, claim to improve basic intelligence skills, such as memory. A study published in the journal Nature investigated whether playing such games can actually boost intelligence (Owen et al. 2010).

The researchers explain that 11,430 people logged onto a webpage and were randomly assigned to one of three groups. Group 1 completed six training tasks that emphasized "reasoning, planning and problem-solving." Group 2 completed games that emphasized a broader range of cognitive skills. Group 3 was a control group and didn't play any of these games; instead, members were prompted to answer "obscure" questions. At the end of six weeks, the participants were compared on several different measures of thinking skills.

Results: The control group did just as well as the treatment group.

Which features of a well-designed controlled experiment does this study have?

Which features are missing?