Chapter 5

Fundamentals of Statistics

Sampling

A researcher wants information on the incomes of everyone in Illinois. It is not practical to ask every individual in the state. Instead, the researcher gets the information from a group of 1200 individuals selected from the whole. This is called sampling.

Definitions

- The **population** in a statistical study is the *complete* set of people or things being studied.

- The **sample** is the subset of the population from which the raw data are actually obtained.

- **Population parameters** are specific characteristics of the population that a statistical study is designed to estimate.

- **Sample statistics** are numbers or observations that summarize the raw data.

Population and sample

population

Question: **What proportion of adults in the U.S. own dogs?**

parameter

(describes population)

sample

A random sample of 1000 adults in the U.S. are asked whether they own one or more dogs. 34% of them say yes.

statistic

(describes sample)

Definitions

- A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

- A statistical study suffers from **bias** if its design or conduct tends to favor certain results.

Sampling Techniques

- Simple Random Sample: choose individuals from the population so that every set of individuals of a given size has an equal chance to be the sample selected.

- Systematic Sampling: choose every kth member of the population.

- Stratified Sampling: partition the population into at least two strata and choose a sample from each stratum.

Identify the Sampling Method Used

- You are conducting a survey of students in a dormitory. You choose your sample by knocking on the door of every fifth room.

  Choosing every fifth room makes this a systematic sample. The sample may be representative, as long as students were randomly assigned to rooms.

- To survey opinions on a proposed new water line, a research firm randomly draws the addresses of 200 homeowners from a public list of all homeowners.

  The records presumably list all homeowners, so drawing randomly from this list produces a simple random sample. It has a good chance of being representative of the population.

Sampling Methods

Even with random sampling, **undercoverage** can occur when some groups in the population are left out of the process of choosing a sample.

To prevent undercoverage in a survey of citizen satisfaction with refuse pickup, 50 residents are randomly selected from each of the twenty-five wards in a city using lists of refuse service customers for each ward. This is an example of **stratified sampling**.

Customers at a supermarket are sampled to determine their opinion about a political issue.

Depending on the location of the market, there may be a bias according to age, economic class, education level, political affiliation, etc.

*This is an example of potential **selection bias** due to using a **convenience sample**.*

Voluntary Response Sample

A sample of people who choose themselves by responding to a general appeal is called a voluntary response sample.

Voluntary response is a likely source of *participation bias*, because people with strong opinions, especially negative opinions, are most likely to participate.

Example: Television viewers are invited to call an 800 number to report their position on a bill to increase state gasoline taxes.

Types of Statistical Study

- In an **observational study**, researchers observe or measure characteristics of the sample members but do not attempt to influence or modify these characteristics.

- In an **experiment**, researchers apply a treatment to some or all of the sample members and then look to see whether the treatment has any effects.

Example

Which is an experiment and which is an observational study?

(a) You ask a sample of smokers how many cigarettes they smoke daily, and measure their blood pressure.

(b) You select a sample of smokers and measure their blood pressure. Then you ask them to reduce their smoking by 5 cigarettes per day; after 3 months you recheck their blood pressure.

(a) is an observational study,

(b) is an experiment.

Experiment: a New Curriculum

The Bigfoot Mountain School District, concerned about poor mathematics preparation, adopts an ambitious new mathematics curriculum. After three years of the new curriculum, students completing sixth grade have an average achievement score 10% higher than they had before the treatment. Bigfoot Mountain pronounces the curriculum a success.

Problem: the experiment cannot distinguish the effects of the changes in parents and teachers from the effect of the new curriculum.

The new curriculum is confounded with the other changes that occurred at the same time. (confounding variables)
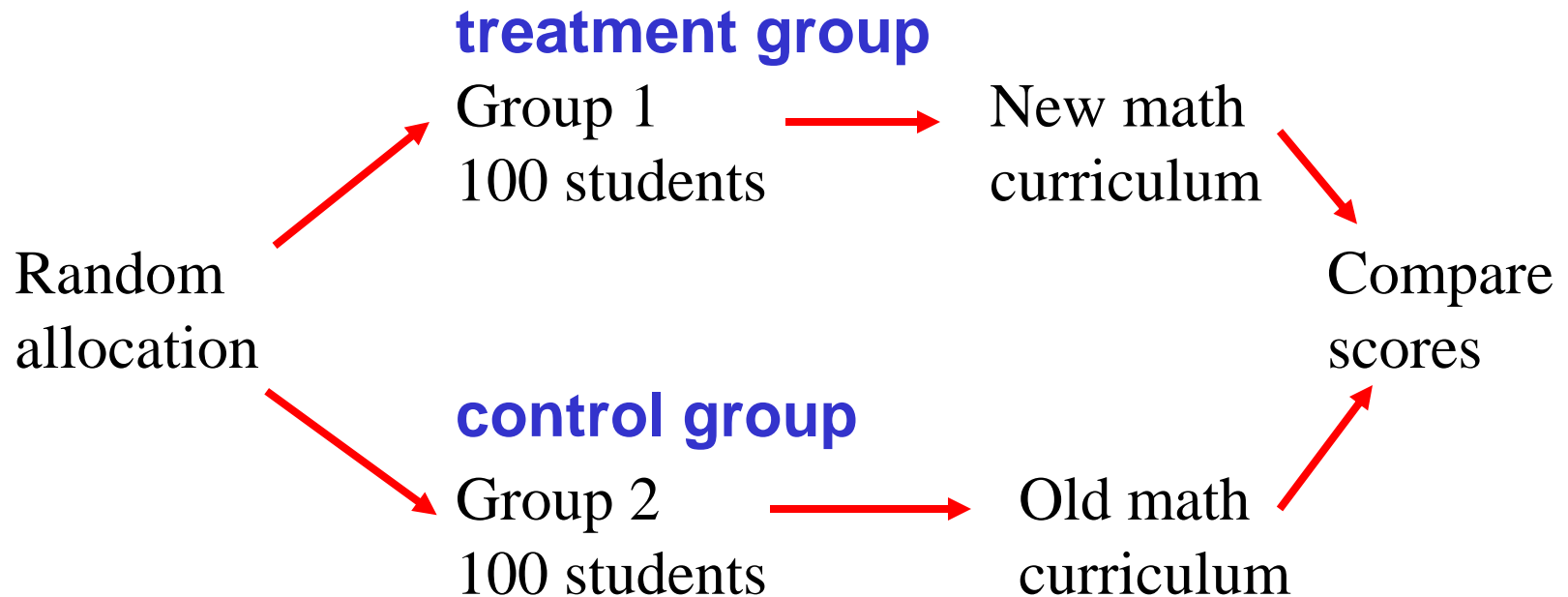
Treatment and Control Groups

- The **treatment group** in an experiment is the group of sample members who receive the treatment being tested.

- The **control group** in an experiment is the group of sample members who do *not* receive the treatment being tested.

  It is important for the treatment and control groups to be selected randomly and to be alike in all respects except for treatment.

A Randomized Comparative Experiment

treatment group

Group 1
100 students → New math curriculum

Random allocation

control group

Group 2
100 students → Old math curriculum

Compare scores

Use a randomized comparative experiment to evaluate a new Mathematics curriculum.

Placebos and the Placebo Effect

- A **placebo** lacks the active ingredients of a treatment being tested in a study, but is identical in appearance to the treatment. Thus, study participants cannot distinguish the placebo from the real treatment.

- The **placebo effect** refers to the situation in which patients improve simply because they believe they are receiving a useful treatment.

Blinding in Experiments

- An experiment is **single-blind** if the participants do not know whether they are members of the treatment group or members of the control group, but the experimenters do know.

- An experiment is **double-blind** if neither the participants nor the experimenters (people administering the treatment) know who belongs to the treatment group and who belongs to the control group.

Definitions

- A **case-control study** is an observational study that resembles an experiment because the sample naturally divides into two (or more) groups.

- The participants who engage in the behavior under study form the **cases**.

- The participants who do not engage in the behavior are the **controls**.