

Math 247: Introduction To Data (Sections 1.1, 1.2)

Introductory Example (page 3 in textbook): Cell phones and cancer

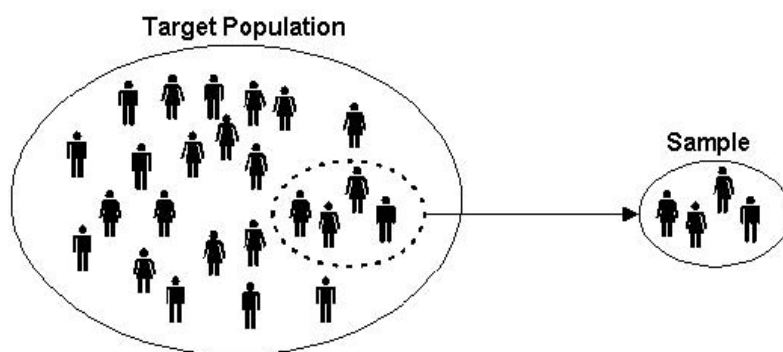
<https://www.ncbi.nlm.nih.gov/pubmed/19513546>

(Target) Population: The entire group that we want to know something about.

Sample: A subset of a population group, containing the “units” (individuals) who are actually observed and whose data (the sample data) is included in our analysis.

A parameter: A number (mean, proportion, standard deviation, etc.) that describes a population. Parameters are **unknown** unless we have **census** data.

A statistic: A number (mean, proportion, standard deviation, etc.) that describes a sample. This is also called an “estimator”. Sample statistics are **known** because we have **sample** data.



Parameters:

Statistics:

Mean
(average)

Proportion
(percentage)

Standard
Deviation

Statistical Inference: When we use the data from a sample to determine something about a population.

The features of an inferential study:

Research question:

Population (of Interest):

Sample:

Variables:

Types of Variables

Categorical/Qualitative variables:

Coding:

Numerical/Quantitative variables:

Ordinal Data:

Identify the research question, the variable(s) of interest and whether the variable(s) is/are categorical or numerical. the population of interest, the sample, and the sample size.

”Online Students Give Instructors Higher Marks If They Think Instructors Are Men”

<https://news.ncsu.edu/2014/12/macnell-gender-2014/>

“A new study shows that college students in online courses give better evaluations to instructors they think are men – even when the instructor is actually a woman.”

“To address whether students judge female instructors differently than male instructors, the researchers evaluated a group of 43 students in an online course. The students were divided into four discussion groups of 8 to 12 students each. A female instructor led two of the groups, while a male instructor led the other two. At the end of the course, students were asked to rate the discussion group instructors [on a scale of 1 to 5] on the traits of professionalism, fairness, respectfulness, giving praise, enthusiasm and promptness.”

The research question:

the variable(s) of interest and type (categorical or numerical):

the population of interest or implied population:

the sample and the sample size:

Section 1.4: Collecting Data to Understand Causality (Controlled Experiments vs. Observational Studies)

Statistical studies are often about a link (association) between variables. These links are often expressed as being “causal” meaning one thing caused another thing to happen or not happen. (Also called “cause-and-effect”.)

Anecdotes and anecdotal evidence: An anecdote is basically a story someone tells about an experience.

Example: “I drank a bottle of kombucha every day during the fall semester and didn’t get a cold.”

If someone told you this, what would you think about kombucha and colds?

Independent Variable (Variable of Interest):

Dependent Variable:

Inferential Statistical Studies: Using data from samples, you can investigate whether there actually is a “statistically significant” link between two variables.

There are two major categories of Inferential Statistical Studies:

OBSERVATIONAL STUDIES

EXPERIMENTS

It’s critical to understand the difference between these two types of studies and to know what conclusions can be drawn from them, particularly whether the study establishes cause-and-effect, or only shows that there’s a link (an association).

Observational Study (passive): A study where the researcher did not manipulate the variable of interest.

Example: “I found 50 students who drank kombucha regularly during fall and found none of them got a cold during that time.”

As compared to what?

We need a comparison group!

“I found 50 students who drank kombucha regularly during fall semester and 50 students who didn’t drink kombucha during fall semester and found that the kombucha drinkers didn’t get a cold and the non-kombucha drinkers, on average got 2 colds.”

Is there a link (association) between drinking kombucha and getting a cold, based on this data?

Does this study show that drinking kombucha helps prevent colds? Why or why not?

Observational Study (with groups): This is a study in which the “assignment” to groups is not made by the investigator. The subjects fall into a category by natural circumstances (by their own choice, in the kombucha example) that are not manipulated by the investigator.

Confounders (lurking variables): This is another variable that affects the outcome. There is some link between the independent variable, the confounder, and the dependent variable.

What are some potential confounders that could explain the differences between the groups in our observational kombucha study?

An observational study does NOT show cause-and-effect between the independent and dependent variables due to the possibility of confounding!
--

Experiment (active): An experiment is a study where the researcher actively manipulates the variable of interest.

A bad experiment: “I randomly chose 50 students, had them drink kombucha every day in fall, and found that none of them got colds.”

As compared to what?

Controlled Experiment (Randomized Controlled Trial, RCT): This is an experiment with groups where the researcher randomly assigns the experimental units to the groups.

Features of a well-designed controlled experiment:

Random Assignment (Randomization): The researcher takes the original sample from the population then randomly assigns them (by essentially picking names from a hat) to the experimental group (treatment group) and the control group (comparison group).

Blinding: The study is designed so that the participants don't know whether or not they have received the treatment.

Placebo: The comparison group receives attention similar to what the treatment group receives so that both groups believe they are receiving the treatment.

Double-blind Experiment: Neither the investigator nor the subjects know who is in which treatment group.

Vocabulary for Variables in Experiments:

Treatment Variable: This is the independent variable. It is the variable of interest, the one that the researcher is testing to see if it causes something to happen or not happen.

Response (Outcome) Variable: This is the dependent variable. This is the “effect” the researcher is looking for, to see whether this effect is “caused” by the treatment variable.

Let's redesign the Kombucha Study so that it's a controlled experiment. We'll use 100 students.

Research Question:

Experiment Design:

A controlled experiment CAN show establish cause-and-effect between the independent and dependent variables since randomization (random assignment to groups) should distribute potential confounders evenly into the groups.

Revisiting Example 1: "Online Students Give Instructors Higher Marks If They Think Instructors Are Men"

"A new study shows that college students in online courses give better evaluations to instructors they think are men – even when the instructor is actually a woman."

To address whether students judge female instructors differently than male instructors, the researchers evaluated a group of 43 students in an online course. The students were randomly assigned to one of four discussion groups. A female instructor led two of the groups, while a male instructor led the other two. The instructors each led one group as their given gender then led their other group pretending to be the other gender. At the end of the course, students were asked to rate the discussion group instructors [on a scale of 1 to 5] on the traits of professionalism, fairness, respectfulness, giving praise, enthusiasm and promptness. The study found that students, on average, gave the "female" instructor lower ratings, whether the instructor was a woman or pretending to be a woman.

What is the research question?

What is the treatment?

What is the response?

Is this study an experiment or an observational study? Explain.

If this is an experiment, was it blind? Explain.

Was it double-blind? Explain.

Can we conclude that perceived instructor gender affected how students rated that instructor?