Week 2: Experiments

Professor Kathryn Jacobs

To Do

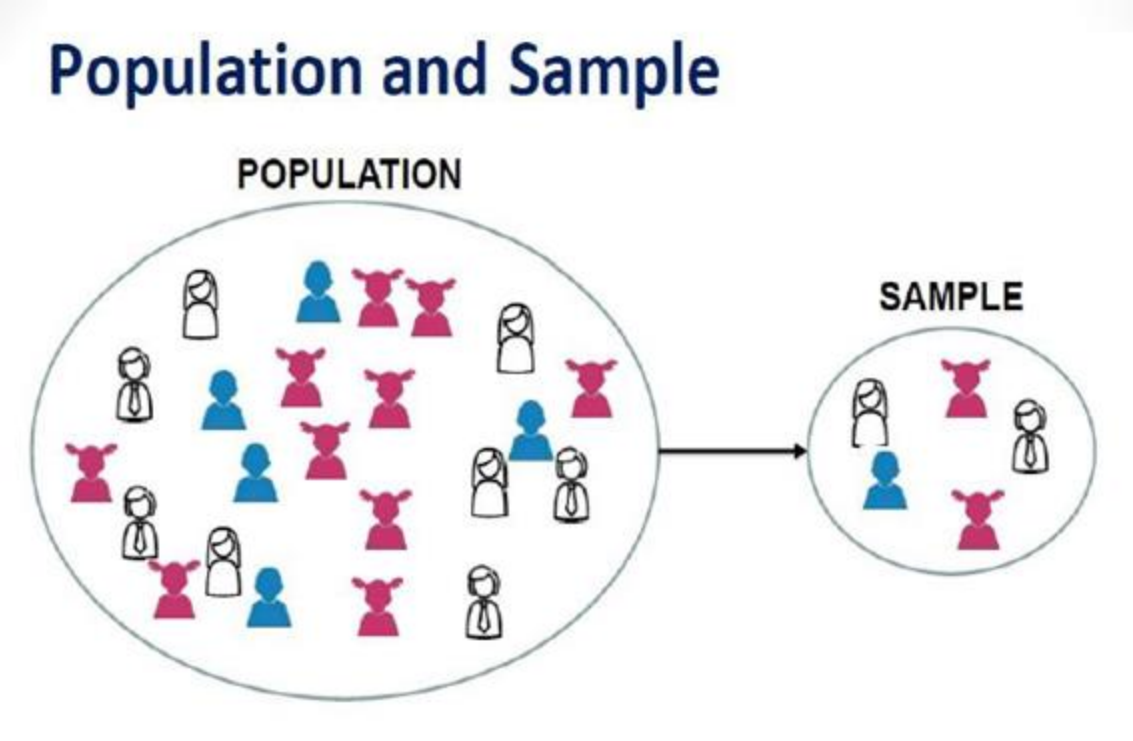
Review

- Population vs. Sample
- Sampling methods
- Sampling biases

Introduce

- Parameter vs. Statistic
- Experimental design

Population vs. Sample



Population vs. Sample: Practice

State whether or not the sampling method described produces a random sample from the given population.

(a) The population is all employees at a company. All employees are emailed a link to a survey.
(b) The population is adults between the ages of 18 and 22. A sample of 100 students is collected from a local university, created by randomly choosing from student ID numbers
(c) The population is all trees in a forest. We walk through the forest and pick out trees that appear to be representative of all the trees in the forest.

Population vs. Sample: Practice

State whether or not the sampling method described produces a random sample from the given population.

(a) The population is all employees at a company. All employees are emailed a link to a survey. No- this is not a sample
(b) The population is adults between the ages of 18 and 22. A sample of 100 students is collected from a local university, created by randomly choosing from student ID numbers. No - this sample does not match the population
(c) The population is all trees in a forest. We walk through the forest and pick out trees that appear to be representative of all the trees in the forest. No- this is representative, but not random

Parameter vs. Statistic

Population = parameter

Sample = statistic

Ex:

The average age of all Americans is 38.5 years

The average age of a random sample of 1000 residents of Minneapolis is 33.3 years

Sampling Methods

Simple random - every person has same chance of being chosen

Stratified - sort into groups, randomly sample from each group

Cluster - sort into clusters, randomly sample entire clusters

Multistage - cluster, then simple random from within chosen clusters

Systematic - every 4th/10th/50th

Sampling Methods: Practice

A statistics student who is curious about the relationship between the amount of time students spend on social networking sites and their performance at school decides to conduct a survey. Various research strategies for collecting data are described below. For each, name the **sampling method** proposed.

(a) He randomly samples 40 students from the study's population, gives them the survey, asks them to fill it out and bring it back the next day.

(b) He gives out the survey only to his friends, making sure each one of them fills out the survey.

(c) He posts a link to an online survey on Facebook and asks people to fill out the survey.

(d) He randomly samples 5 classes and asks a random sample of students from those classes to fill out the survey.

Sampling Bias

Non-response: some people won’t respond

Voluntary response: only those with strong feelings will respond

Convenience sample: whoever is convenient

Undercoverage: sampling at certain times or in certain places will miss specific people

Sampling Methods: Practice

A statistics student who is curious about the relationship between the amount of time students spend on social networking sites and their performance at school decides to conduct a survey. Various research strategies for collecting data are described below. For each, list any **possible bias**

(a) He randomly samples 40 students from the study's population, gives them the survey, asks them to fill it out and bring it back the next day.

(b) He gives out the survey only to his friends, making sure each one of them fills out the survey.

(c) He posts a link to an online survey on Facebook and asks people to fill out the survey.

(d) He randomly samples 5 classes and asks a random sample of students from those classes to fill out the survey.

Experimental Design

Principles of experimental design

1. Controlling
2. Randomization
3. Replication
4. Blocking

Experimental Design: Controlling

Compare experimental treatment to a control group

Ex: Group 1 gets new experimental painkiller, group 2 gets placebo

Control group must match experimental group as closely as possible

Experimental Design: Randomization

- Randomly sample from population when possible (this counts simple random sampling, but also cluster and stratified sampling)
- Random assignment to control vs. experimental group

BOTH are necessary to have a valid experiment

Experimental Design: Replication

Accomplished in 2 ways, either

1. Replicate using another study of identical (ish) design
2. Use a large enough sample size that the study alone is valid

Experimental Design: Blocking

Helps account for confounding variables

- Block study participants into groups (strata) based on variables [eg age groups, education status]
- Randomize participants within strata into control vs. experimental groups
- Like stratified sampling, only instead of just selecting people we're assigning them to groups

Blocking: Example

- An experimenter wants to test if energy gels improve performance in a race
  - Treatment: energy gel
  - Control: no energy gel

- She suspects that energy gels effect pros and amateurs differently, so she blocks pros and amateurs separately

Blocking: Example

Issue: We think energy gel affects pro and amateur athletes differently

Solution: Block

1. Separate athletes by pro vs. amateur status
2. Randomly assign pro athletes to either Energy Gel (treatment group) or No Energy Gel (control group)
3. Randomly assign amateur athletes to either Energy Gel (treatment group) or No Energy Gel (control group)
4. Test run times of treatment and control groups. Pros and amateurs are now equally represented in both groups

Practice: THINK ~2 mins

A researcher is interested in the effects of exercise on mental health, and she proposes the following study:

Use stratified random sampling to ensure representative proportions of 18-30, 31-40, and 41- 55 year-olds from a local gym. Next, randomly assign half the subjects from each age group to exercise twice a week and instruct the other half not to exercise.

b) What is the explanatory variable and the response variable?

(b) What are the treatment and control groups in this study?

(c) Does this study make use of blocking? If so, what is the blocking variable?

(d) Does this study make use of blinding?

(e) Can this study can be used to establish a causal relationship between exercise and mental health? Can the conclusions be generalized to the population at large?

Practice: PAIR ~2 mins

A researcher is interested in the effects of exercise on mental health, and she proposes the following study:

Use stratified random sampling to ensure representative proportions of 18-30, 31-40, and 41- 55 year-olds from a local gym. Next, randomly assign half the subjects from each age group to exercise twice a week and instruct the other half not to exercise.

b) What is the explanatory variable and the response variable?

(b) What are the treatment and control groups in this study?

(c) Does this study make use of blocking? If so, what is the blocking variable?

(d) Does this study make use of blinding?

(e) Can this study can be used to establish a causal relationship between exercise and mental health? Can the conclusions be generalized to the population at large?

Practice: SHARE ~5 mins

A researcher is interested in the effects of exercise on mental health, and she proposes the following study:

Use stratified random sampling to ensure representative proportions of 18-30, 31-40, and 41- 55 year-olds from a local gym. Next, randomly assign half the subjects from each age group to exercise twice a week and instruct the other half not to exercise.

b) What is the explanatory variable and the response variable?

(b) What are the treatment and control groups in this study?

(c) Does this study make use of blocking? If so, what is the blocking variable?

(d) Does this study make use of blinding?

(e) Can this study can be used to establish a causal relationship between exercise and mental health? Can the conclusions be generalized to the population at large?