Stats 7 Lecture 1 practice

August 1, 2022

Class website

https://catalinamedina.github.io/uci-stats7-su22-website/

Note:

- First learning objective
- Extension on first lecture video quiz due date

Migraine study

- A migraine is a particularly painful type of headache, which patients sometimes wish to treat with acupuncture.
- To determine whether acupuncture relieves migraine pain, researchers conducted a randomized controlled study where 89 females diagnosed with migraine headaches were randomly assigned to one of two groups: treatment or control.
  - 43 patients in the treatment group received acupuncture that is specifically designed to treat migraines.
  - 46 patients in the control group received placebo acupuncture (needle insertion at non-acupoint locations).
  - 24 hours after patients received acupuncture, they were asked if they were pain free.

Migraine study: results

		Pain free		
		Yes	No	Total
Group	Treatment	10	33	43
	Control	2	44	46
	Total	12	77	89

- What percent of patients in the treatment group were pain free 24 hours after receiving acupuncture?
- 10/43 = 0.23 ➡ 23%
- What percent were pain free in the control group?
- 2/46 = 0.04 ➡ 4%.
- In which group did a higher percent of patients become pain free 24 hours after receiving acupuncture?
- A higher percentage of patients in the treatment group were pain free 24 hours after receiving acupuncture.

Migraine study: results (continued)

- Your findings so far might suggest that acupuncture is an effective treatment for migraines for all people who suffer from migraines. However, this is not the only possible conclusion that can be drawn based on your findings so far. What is one other possible explanation for the observed difference between the percentages of patients that are pain free 24 hours after receiving acupuncture in the two groups?
- It is possible that the observed difference between the two group percentages is due to chance.

Types of variables
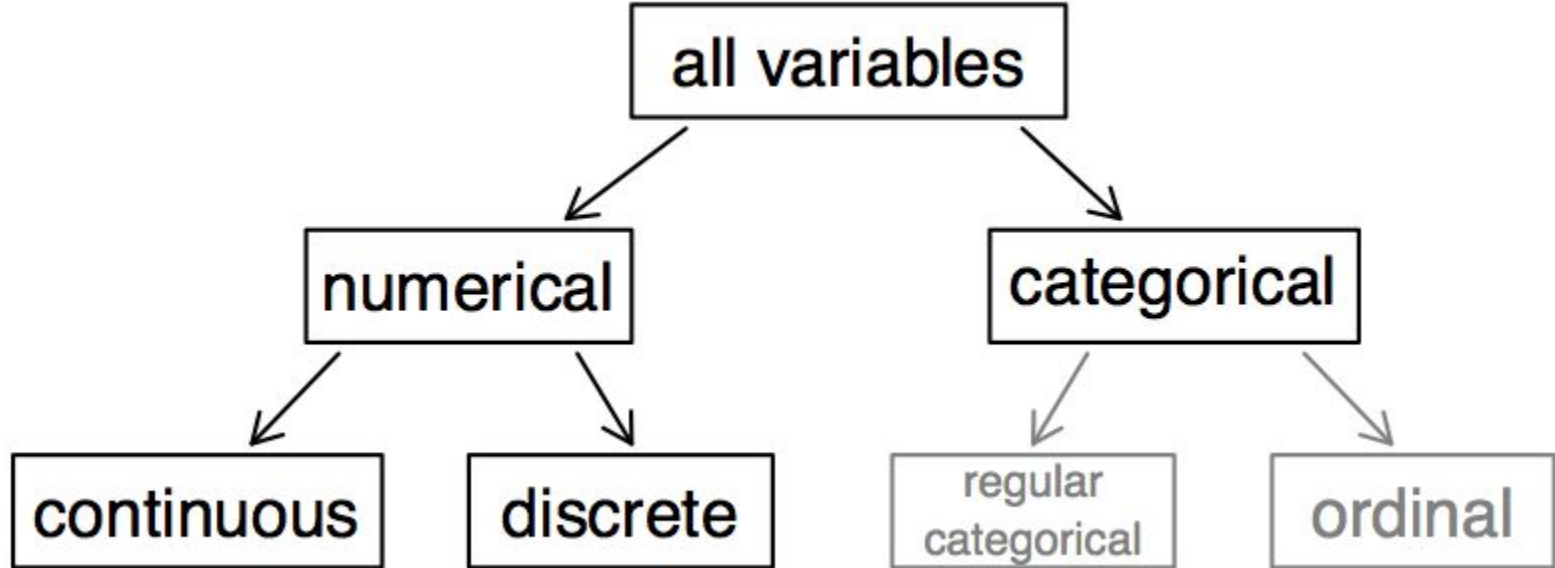
Air pollution and birth outcomes study

- Researchers collected data to examine the relationship between air pollutants and preterm births in Southern California.
- During the study air pollution levels were measured by air quality monitoring stations.
- Specifically, levels of carbon monoxide were recorded in parts per million, nitrogen dioxide and ozone in parts per hundred million, and coarse particulate matter (PM10) in µg/m3 .
- Length of gestation data were collected on 143,196 births between the years 1989 and 1993, and air pollution exposure during gestation was calculated for each birth.
- The analysis suggested that increased ambient PM10 and, to a lesser degree, CO concentrations may be associated with the occurrence of preterm births.

Air pollution and birth outcomes study

- Identify the main research question of the study.
- "Is there an association between air pollution exposure and preterm births?"
- Who are the subjects in this study, and how many are included?
- 143,196 births in Southern California between 1989 and 1993
- What are the variables in the study? Identify the type of each variable.
- Measurements of carbon monoxide, nitrogen dioxide, ozone, and particulate matter less than 10µg/m3 (PM10) collected at air-quality-monitoring stations as well as length of gestation. Continuous numerical variables.

Fisher's irises

- Sir Ronald Aylmer Fisher was an English statistician, evolutionary biologist, and geneticist who worked on a data set that contained sepal length and width, and petal length and width from three species of iris flowers (setosa, versicolor and virginica). There were 50 flowers from each species in the data set.

Fisher's irises

- How many cases were included in the data?
- 50×3 = 150
- How many numerical variables are included in the data? Indicate what they are, and if they are continuous or discrete.
- Four continuous numerical variables: sepal length, sepal width, petal length, and petal width
- How many categorical variables are included in the data, and what are they? List the corresponding levels (categories).
- One categorical variable, species, with three levels: setosa, versicolor, and virginica

General Social Survey question

- The General Social Survey asked the question, "After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?" to a random sample of 1,155 Americans. The average relaxing time was found to be 1.65 hours.
- Determine which of the following is an observation, a variable, a sample statistic (value calculated based on the observed sample), or a population parameter.
    - Number of hours spent relaxing after an average work day.
    - An American in the sample.
    - Average number of hours all Americans spend relaxing after an average work day.
    - 1.65

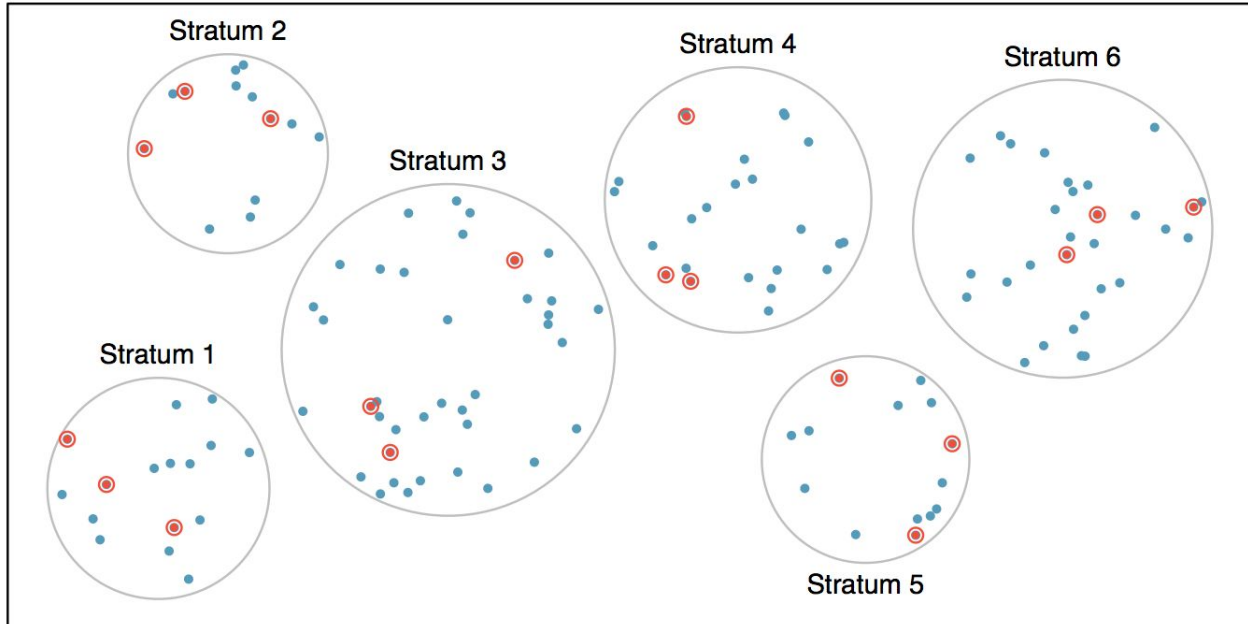   (Variable, observation, population parameter, sample statistic)

Simple Random Sample

Randomly select cases from the population, where there is no implied connection between the points that are selected.
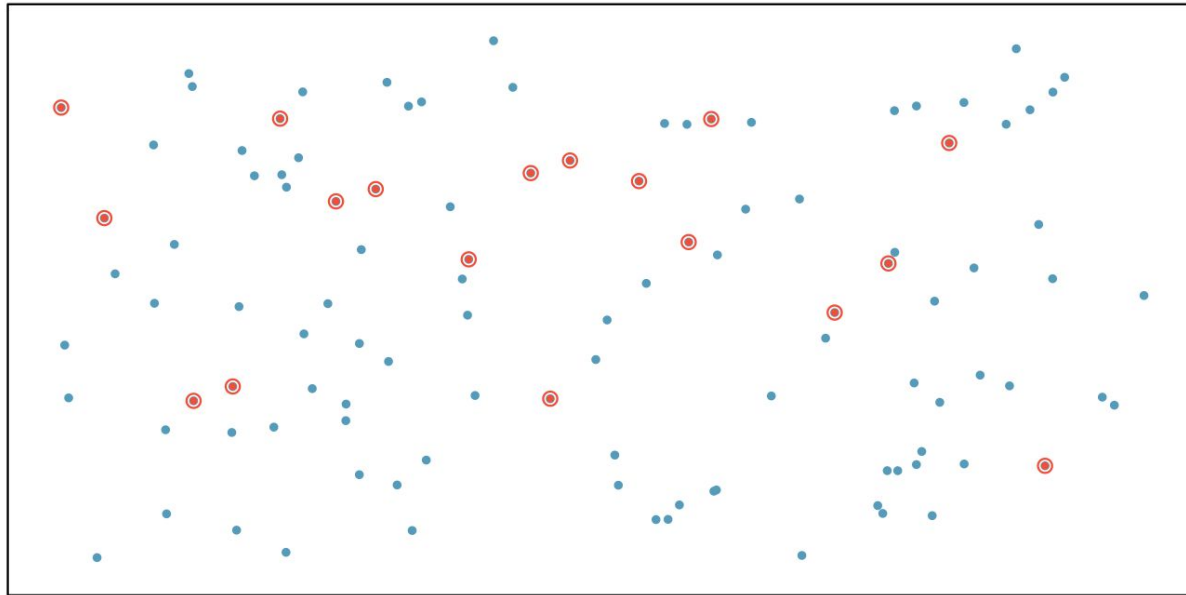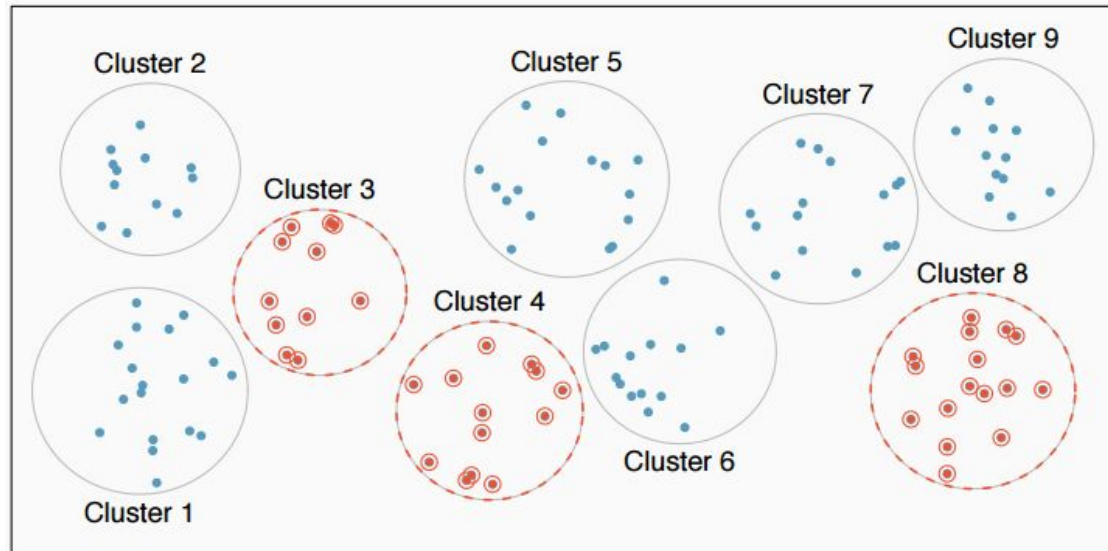
Stratified Sample

Strata are made up of similar observations. We take a simple random sample from <u>each</u> stratum.

Cluster Sample

Clusters are usually not made up of homogeneous observations. We take a simple random sample of clusters, and then sample all observations in that cluster. Usually preferred for economical reasons.

Multistage Sample

Clusters are usually not made up of homogeneous observations. We take a simple random sample of clusters, and then take a simple random sample of observations from the sampled clusters
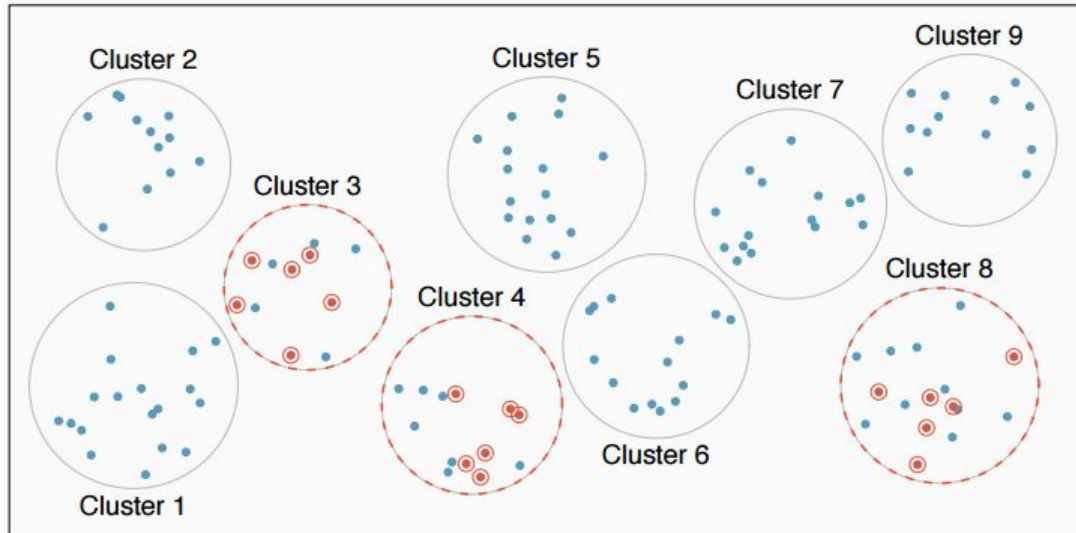
Housing proposal across dorms

- On a large college campus first-year students and sophomores live in dorms located on the eastern part of the campus and juniors and seniors live in dorms located on the western part of the campus.
- Suppose you want to collect student opinions on a new housing structure the college administration is proposing and you want to make sure your survey equally represents opinions from students from all years.
- What type of study is this?
- Suggest a sampling strategy for carrying out this study

Observational, stratified sampling

Internet use and life expectancy

The following scatterplot was created as part of a study evaluating the relationship between estimated life expectancy at birth (as of 2014) and percentage of internet users (as of 2009) in 208 countries for which such data were available.
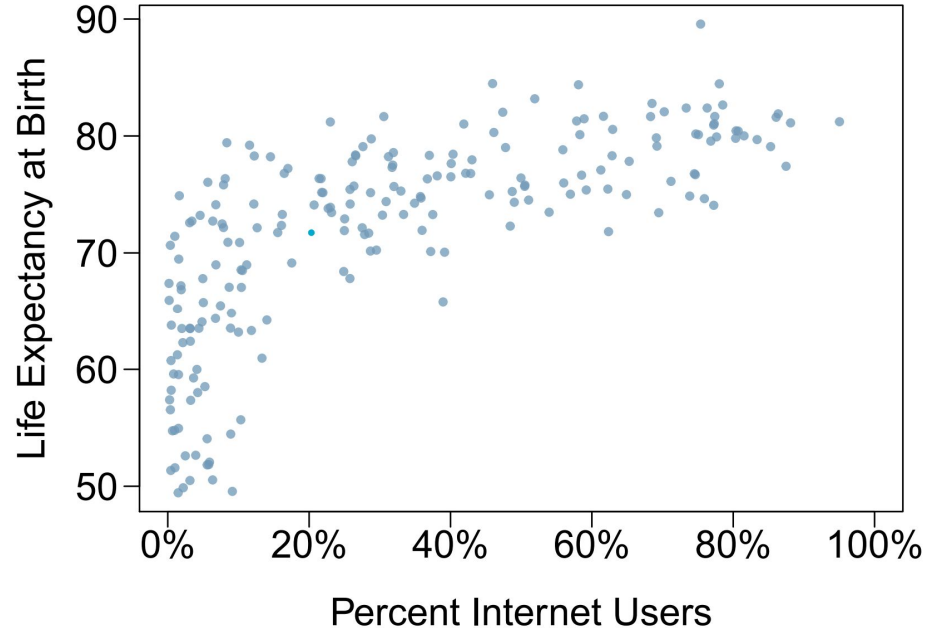
Internet use and life expectancy

- Describe the relationship between life expectancy and percentage of internet users.
- Positive nonlinear association
- What type of study is this?
- Observational study
- State a possible confounding variable that might explain this relationship and describe its potential effect.
- Wealth: countries with individuals who can widely afford the internet can probably also afford basic medical care.

Recall relationships between variables

Random Assignment vs. Random Sampling

	Random assignment	No random assignment	
ideal experiment			*most observational studies*
Random sampling	Causal conclusion, generalized to the whole population.	No causal conclusion, correlation statement generalized to the whole population.	Generalizability
No random sampling	Causal conclusion, only for the sample.	No causal conclusion, correlation statement only for the sample.	No generalizability
most experiments	Causation	Correlation	*bad observational studies*

Haters are gonna hate, study confirms

- A study published in the Journal of Personality and Social Psychology asked a group of 200 randomly sampled men and women to evaluate how they felt about various subjects, such as camping, health care, architecture, taxidermy, crossword puzzles, and Japan in order to measure their attitude towards mostly independent stimuli.
- Then, they presented the participants with information about a new product: a microwave oven.
- This microwave oven does not exist, but the participants didn't know this, and were given three positive and three negative fake reviews.

Haters are gonna hate, study confirms

- People who reacted positively to the subjects on the dispositional attitude measurement also tended to react positively to the microwave oven, and those who reacted negatively tended to react negatively to it.
- Researchers concluded that "some people tend to like things, whereas others tend to dislike things, and a more thorough understanding of this tendency will lead to a more thorough understanding of the psychology of attitudes."

Haters are gonna hate, study confirms

- What are the cases?
- The cases are 200 randomly sampled men and women.
- What is (are) the response variable(s) in this study?
- The response variable is attitude towards a fictional microwave oven
- What is (are) the explanatory variable(s) in this study?
- The explanatory variable is dispositional attitude

Haters are gonna hate, study confirms

- Does the study employ random sampling?
- Yes, the cases are sampled randomly
- Is this an observational study or an experiment? Explain your reasoning.
- This is an observational study since there is no random assignment to treatments.
- Can we establish a causal link between the explanatory and response variables?
- No, we cannot establish a causal link between the explanatory and response variables since the study is observational.
- Can the results of the study be generalized to the population at large?
- Yes, the results of the study can be generalized to the population at large since the sample is random

Flawed reasoning

- Identify the flaw(s) in reasoning in the following scenarios.
- Explain what the individuals in the study should have done differently if they wanted to make such strong conclusions.

Flawed reasoning

Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, "Do you find that your work schedule makes it difficult for you to spend time with your kids after school?" Of the parents who replied, 85% said "no". Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.

Non-responders may have a different response to this question, e.g. parents who returned the surveys likely don't have difficulty spending time with their children.

Flawed reasoning

A survey is conducted on a simple random sample of 1,000 women who recently gave birth, asking them about whether or not they smoked during pregnancy. A follow-up survey asking if the children have respiratory problems is conducted 3 years later. However, only 567 of these women are reached at the same address. The researcher reports that these 567 women are representative of all mothers.

It is unlikely that the women who were reached at the same address 3 years later are a random sample. These missing responders are probably renters (as opposed to homeowners) which means that they might be in a lower socio-economic status than the respondents

Flawed reasoning

An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

There is no control group in this study, this is an observational study, and there may be confounding variables, e.g. these people may go running because they are generally healthier and/or do other exercises.

Eat better, feel better?

- In a public health study on the effects of consumption of fruits and vegetables on psychological well-being in young adults, participants were randomly assigned to three groups: (1) dietas-usual, (2) an ecological momentary intervention involving text message reminders to increase their fruits and vegetable consumption plus a voucher to purchase them, or (3) a fruit and vegetable intervention in which participants were given two additional daily servings of fresh fruits and vegetables to consume on top of their normal diet.

Eat better, feel better?

- Participants were asked to take a nightly survey on their smartphones.
- Participants were student volunteers at the University of Otago, New Zealand.
- At the end of the 14-day study, only participants in the third group showed improvements to their psychological well-being across the 14-days relative to the other groups.

Eat better, feel better?

- What type of study is this?
- Randomized controlled experiment.
- Identify the explanatory and response variables.
- Explanatory: treatment group (categorical, with 3 levels). Response variable: Psychological wellbeing.
- Comment on whether the results of the study can be generalized to the population.
- No, because the participants were volunteers.
- Comment on whether the results of the study can be used to establish causal relationships.
- Yes, because it was an experiment

Eat better, feel better?

- A newspaper article reporting on the study states, "The results of this study provide proof that giving young adults fresh fruits and vegetables to eat can have psychological benefits, even over a brief period of time." How would you suggest revising this statement so that it can be supported by the study?
- The statement should say "evidence" instead of "proof".

Questions?

Credits

Examples adapted from OpenIntro Statistics (4th edition) by David Diez, Mine Cetinkaya-Rundel, and Christopher D Barr https://www.openintro.org/book/os/ protected under the Creative Commons License