

Homework 1 Solutions

Math 150

Enrique Treviño

1.2 Sinusitis and antibiotics. Researchers studying the effect of antibiotic treatment for acute sinusitis compared to symptomatic treatments randomly assigned 166 adults diagnosed with acute sinusitis to one of two groups: treatment or control. Study participants received either a 10-day course of amoxicillin (an antibiotic) or a placebo similar in appearance and taste. The placebo consisted of symptomatic treatments such as acetaminophen, nasal decongestants, etc. At the end of the 10-day period patients were asked if they experienced significant improvement in symptoms. The distribution of responses are summarized below.

		<i>Self-reported significant improvement in symptoms</i>		Total
		Yes	No	
<i>Group</i>	Treatment	66	19	85
	Control	65	16	81
	Total	131	35	166

- What percent of patients in the treatment group experienced a significant improvement in symptoms? What percent in the control group?
- Based on your findings in part (a), which treatment appears to be more effective for sinusitis?
- Do the data provide convincing evidence that there is a difference in the improvement rates of sinusitis symptoms? Or do you think that the observed difference might just be due to chance?

SOLUTION.

- For the treatment group is $66/85 \approx 0.776 = 77.6\%$. For the control group it's $65/81 \approx 0.802 = 80.2\%$.
- The control group.
- The difference is small, so it seems it can be due to chance.

1.4 Buteyko method, study components. The Buteyko method is a shallow breathing technique developed by Konstantin Buteyko, a Russian doctor, in 1952. Anecdotal evidence suggests that the Buteyko method can reduce asthma symptoms and improve quality of life. In a scientific study to determine the effectiveness of this method, researchers recruited 600 asthma patients aged 18-69 who relied on medication for asthma treatment. These patients were split into two research groups: one practiced the Buteyko method and the other did not. Patients were scored on quality of life, activity, asthma symptoms, and medication reduction on a scale from 0 to 10. On average, the participants in the Buteyko group experienced a significant reduction in asthma symptoms and an improvement in quality of life. Identify

- the cases,

- (b) the variables and their types, and
- (c) the main research question

in this study.

SOLUTION.

- (a) The cases are the 600 asthma patients.
- (b) One variable is whether they did Buteyko breathing or not? This variable is categorical. The other variables are “quality of life”, “activity”, “asthma symptoms”, and “medication reduction”. All of them are on a scale from 0 to 10, so they are numerical.
- (c) Does the Buteyko breathing method improve quality of life and reduce asthma symptoms?

1.8 Smoking habits of UK residents. A survey was conducted to study the smoking habits of UK residents. Below is a data matrix displaying a portion of the data collected in this survey. Note that “£” stands for British Pounds Sterling, “cig” stands for cigarettes, and “N/A” refers to a missing component of the data.

	sex	age	marital	grossIncome	smoke	amtWeekends	amtWeekdays
1	Female	42	Single	Under £2,600	Yes	12 cig/day	12 cig/day
2	Male	44	Single	£10,400 to £15,600	No	N/A	N/A
3	Male	53	Married	Above £36,400	Yes	6 cig/day	6 cig/day
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1691	Male	40	Single	£2,600 to £5,200	Yes	8 cig/day	8 cig/day

- (a) What does each row of the data matrix represent?
- (b) How many participants were included in the survey?
- (c) Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.

SOLUTION.

- (a) A UK resident.
- (b) 1691.
- (c) Sex is categorical, age is numerical (discrete), marital status is categorical, gross income is categorical (ordinal), whether a person smokes is categorical, amount of cigarettes per day on weekends is numerical (discrete), amount of cigarettes per day on weekdays is numerical (discrete).

1.11 Buteyko method, scope of inference. Exercise 1.4 introduces a study on using the Buteyko shallow breathing technique to reduce asthma symptoms and improve quality of life. As part of this study 600 asthma patients aged 18-69 who relied on medication for asthma treatment were recruited and randomly assigned to two groups: one practiced the Buteyko method and the other did not. Those in the Buteyko group experienced, on average, a significant reduction in asthma symptoms and an improvement in quality of life.

- (a) Identify the population of interest and the sample in this study.
- (b) Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

SOLUTION.

- (a) A population of interest is all asthma patients aged 18-69 who relied on medication for asthma treatment. The sample in this study is a sample of 600 asthma patients ages 18-69 who relied on medication for asthma treatment.
- (b) The results of the study cannot be generalized to the general population since the sample wasn't taken randomly, so it's unclear whether they are representative. The study, however, does suggest a causal relationship because it was a randomized experiment.

1.14 Cats on YouTube. Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 1000 videos for you. You find that 2% of these videos are cat videos. Determine which of the following is an observation, a variable, a sample statistic, or a population parameter.

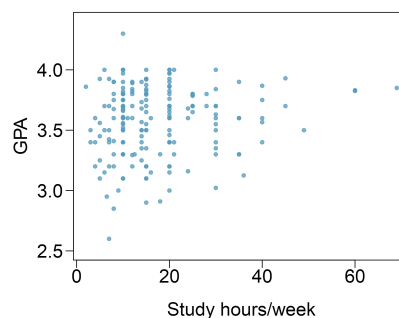
- (a) Percentage of all videos on YouTube that are cat videos.
- (b) 2%.
- (c) A video in your sample.
- (d) Whether or not a video is a cat video.

SOLUTION.

- (a) Population parameter.
- (b) Sample statistic.
- (c) An observation.
- (d) Variable.

1.15 GPA and study hours. A survey was conducted on 193 Duke University undergraduates who took an introductory statistics course in 2012. Among many other questions, this survey asked them about their GPA, which can range between 0 and 4 points, and the number of hours they spent studying per week. The scatterplot below displays the relationship between these two variables.

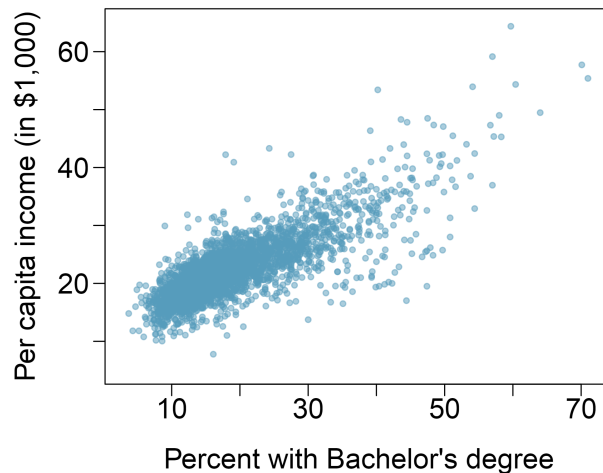
- (a) What is the explanatory variable and what is the response variable?
- (b) Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.
- (c) Is this an experiment or an observational study?
- (d) Can we conclude that studying longer hours leads to higher GPAs?



SOLUTION.

- (a) The explanatory variable is study hours per week and the response variable is the GPA.
- (b) There seems to be a positive association, that is, that the more study hours, the higher the GPA. Some unusual observations are: there is a person with a GPA above 4, but 4 should be the limit. There are two people that claimed to study more than 60 hours per week.
- (c) It is an observational study.
- (d) We can't because it is an observational study, we can just point to an association (or correlation).

1.16 Income and education in US counties. The scatterplot below shows the relationship between per capita income (in thousands of dollars) and percent of population with a bachelor's degree in 3,143 counties in the US in 2010.



- (a) What are the explanatory and response variables?
- (b) Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.
- (c) Can we conclude that having a bachelor's degree increases one's income?

SOLUTION.

- (a) Percentage of people with a bachelor's degree (in the county) is the explanatory variable. Per capita income (in the county) is the response variable.
- (b) There is a positive association. The higher the percentage of people with a bachelor's degree in a county, the higher the income per capita in the county.
- (c) We cannot because it is an observational study, so we cannot infer causation.

1.20 Stressed out, Part I. A study that surveyed a random sample of otherwise healthy high school students found that they are more likely to get muscle cramps when they are stressed. The study also noted that students drink more coffee and sleep less when they are stressed.

- (a) What type of study is this?

- (b) Can this study be used to conclude a causal relationship between increased stress and muscle cramps?
- (c) State possible confounding variables that might explain the observed relationship between increased stress and muscle cramps.

SOLUTION.

- (a) Observational study.
- (b) No, because it is an observational study.
- (c) Maybe the students get stressed around exam time. Since it is exam time, they study more, so they sleep less and drink more coffee to be able to study more. They also sit more because they study more, so they might get muscle cramps from the inactivity.

1.28 Reading the paper. Below are excerpts from two articles published in the *NY Times*:

- (a) An article titled *Risks: Smokers Found More Prone to Dementia* states the following:

“Researchers analyzed data from 23,123 health plan members who participated in a voluntary exam and health behavior survey from 1978 to 1985, when they were 50-60 years old. 23 years later, about 25% of the group had dementia, including 1,136 with Alzheimer’s disease and 416 with vascular dementia. After adjusting for other factors, the researchers concluded that pack-a-day smokers were 37% more likely than nonsmokers to develop dementia, and the risks went up with increased smoking; 44% for one to two packs a day; and twice the risk for more than two packs.”

Based on this study, can we conclude that smoking causes dementia later in life? Explain your reasoning.

- (b) Another article titled *The School Bully Is Sleepy* states the following:

“The University of Michigan study, collected survey data from parents on each child’s sleep habits and asked both parents and teachers to assess behavioral concerns. About a third of the students studied were identified by parents or teachers as having problems with disruptive behavior or bullying. The researchers found that children who had behavioral issues and those who were identified as bullies were twice as likely to have shown symptoms of sleep disorders.”

A friend of yours who read the article says, “The study shows that sleep disorders lead to bullying in school children.” Is this statement justified? If not, how best can you describe the conclusion that can be drawn from this study?

SOLUTION.

- (a) We cannot conclude that because it is an observational study. However we can see an association.
- (b) Once again, we cannot conclude it, since it is an observational study. One way to describe the conclusion is “there is a correlation between sleep disorders and bullying”.

1.30 Stressed out, Part II. In a study evaluating the relationship between stress and muscle cramps, half the subjects are randomly assigned to be exposed to increased stress by being placed into an elevator that falls rapidly and stops abruptly and the other half are left at no or baseline stress.

- (a) What type of study is this?
- (b) Can this study be used to conclude a causal relationship between increased stress and muscle cramps?

SOLUTION.

- (a) Experiment.
- (b) The study creates stress in a way that might affect muscle cramps. If an elevator falls rapidly and stops abruptly, that could affect muscle cramps more because of the physical action than by the stress it induced. I don't think this experiment would help answer whether there is a causal relationship between increases stress and muscle cramps.

1.31 Light and exam performance. A study is designed to test the effect of light level on exam performance of students. The researcher believes that light levels might have different effects on males and females, so wants to make sure both are equally represented in each treatment. The treatments are fluorescent overhead lighting, yellow overhead lighting, no overhead lighting (only desk lamps).

- (a) What is the response variable?
- (b) What is the explanatory variable? What are its levels?
- (c) What is the blocking variable? What are its levels?

SOLUTION.

- (a) Exam performance.
- (b) Light level. The levels are fluorescent, overhead lighting, yellow overhead lighting, no overhead lighting (only desk lamps).
- (c) The blocking variable is sex (male or female). The levels are male and female.

1.36 Exercise and mental health. A researcher is interested in the effects of exercise on mental health and he proposes the following study: Use stratified random sampling to ensure representative proportions of 18-30, 31-40 and 41- 55 year olds from the population. Next, randomly assign half the subjects from each age group to exercise twice a week, and instruct the rest not to exercise. Conduct a mental health exam at the beginning and at the end of the study, and compare the results.

- (a) What type of study is this?
- (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) Comment on whether or not the results of the study can be used to establish a causal relationship between exercise and mental health, and indicate whether or not the conclusions can be generalized to the population at large.
- (f) Suppose you are given the task of determining if this proposed study should get funding. Would you have any reservations about the study proposal?

SOLUTION.

- (a) Experiment.
- (b) The treatment group is the group that exercises 2 times a week, the control group is the one that does not exercise.
- (c) Yes. The blocking variable is age. The blocks are 18-30 year olds, 31-40 year olds 41-55 year olds.
- (d) No. The patients know if they are in the group that exercises or the one that doesn't.

- (e) One can establish a causal relationship and one can generalize to the population at large (because it came from random sampling). One possible issue is how much the patients used to exercise before the experiment. Maybe some active people were sent to the no-exercise group and viceversa.
- (f) As mentioned above, one reservation I would have is that it's unclear if the effect would come from having people outside their comfort zone or from the actual exercising. It's also a little morally dubious to force people not to exercise. A better idea would be to have an experiment where you split into one group that exercises once a week more than before and a control group that exercises the same as before.