

## SC212, Spring 2022 — HW1

Matthew Bass

### SECTION 1.1 .

- 1.6** For the situations described in Exercises 1.1 to 1.6 : (a) What are the cases? (b) What is the variable and is it quantitative or categorical?

Record whether or not the literacy rate is over 75% for each country in the world.

- a.* The cases (samples) are each country in world.
- b.* The variable is literacy rate which is quantitative but if you are recording whether or not each country's literacy rate is over 75% or not that is a categorical (binary in this case) variable.

- 1.8:** In Exercises 1.7 to 1.10, a relationship between two variables is described. In each case, we can think of one variable as helping to explain the other. Identify the explanatory variable and the response variable.

Amount of fertilizer used and the yield of a crop

The explanatory variable is the amount of fertilizer used and the response variable is the yield of crop

#### **1.20 How Fast Do Homing Pigeons Go?**

Homing pigeons have an amazing ability to find their way home over extremely long distances. How fast do they go on these trips? In the 2019 Midwest Classic, held in Topeka, Kansas, the fastest bird went 1676 YPM (yards per minute), which is about 56 miles

per hour. <sup>15</sup> The top seven finishers included three female pigeons (Hens) and four male pigeons (Cocks). Their speeds, in YPM, are given in Table 1.4.

**Table 1.4** Speed of homing pigeons, in yards per minute

Hens:	1676, 1452, 1449
Cocks:	1458, 1435, 1418, 1413

- a. How many cases are there in this dataset? How many variables are there and what are they? Is each variable categorical or quantitative?

There are 7 cases in the dataset and there are 2 different variables. The first variable is the gender of the pigeon which is a categorical variable recording if the pigeon is a male (Cock) or female (Hen). The second variable recorded is the homing pigeons flights speed which is quantitative and recorded in yards per minute.

- b. Display the information as a dataset with cases as rows and variables as columns.

	Gender	YPM
Pigeon 1	F	1676
Pigeon 2	M	1458
Pigeon 3	F	1452
Pigeon 4	M	1435
Pigeon 5	F	1449
Pigeon 6	M	1418
Pigeon 7	M	1413

### 1.24 Pennsylvania High School Seniors

The data in **PASeniors** shows results for a sample of 457 high-school seniors in the state of Pennsylvania, selected at random from all students who participated in the Census at Schools project <sup>17</sup> between 2010 and 2019 . Each of the questions below relate to information in this dataset. Determine whether the answer to each question gives a value for a quantitative variable, a categorical variable, or is not a value for a variable for this dataset.

- a.* What mode of transportation do you use to get to school?  
Categorical Variable
- b.* Do you have any allergies?  
Categorical Variable
- c.* What proportion of students in this sample are vegetarians?  
Not a value for a variable but a stat from a variable that could be taken.
- d.* How many hours did you spend last week working at a paid job?  
Quantitative Variable
- e.* What is the difference between typical hours of sleep you get on school nights and non-school nights?  
Quantitative Variable
- f.* What is the maximum time (in minutes) that a student in this sample needs to get to school?  
Not a value for a variable but a stat from a variable that could be taken.
- g.* If you could have a super power would you choose invisibility, telepathy, super strength, ability to fly, or ability to freeze time?  
Not a value from the dataset.

## SECTION 1.2 .

**1.36** In Exercises 1.34 to 1.37, state whether the data are best described as a population or a sample.

The US Department of Transportation announces that of the 250 million registered passenger vehicles in the US, 2.1% are electro-gas hybrids.

## Population Data

- 1.38** In Exercises 1.38 to 1.41, describe the sample and describe a reasonable population.

A sociologist conducting a survey at a mall interviews 120 people about their cell phone use.

The sample is the 120 people that go to the mall that are surveyed and a reasonable population would be that it represents a general makeup of the type of people (race, socioeconomic, ect..) at the mall where the survey has taken place.

- 1.44** To investigate interest across all residents of the US in a new type of ice skate, a random sample of 1500 people in Minnesota are asked about their interest in the product.

- a. The sample is the 1500 residents of Minnesota that were surveyed.
- b. The population of interest for the survey was the entire population of the United States
- a. A population we can generalize from the surveys sample is for the entire population of Minnesota, not the entire United States

- 1.50** In Exercises 1.46 to 1.51, state whether or not the sampling method described produces a random sample from the given population.

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.

This method does not provide a random sample from the given population

- 1.56** In Exercises 1.52 to 1.56, indicate whether we should trust the results of the study. Is the method of data collection biased? If it is, explain why.

Send an email to a random sample of students at a university asking them to reply to the question: "Do you think this university should fund an ultimate frisbee team?" A small number of students reply. Use the replies to estimate the proportion of all students at the university who support this use of funds.

This method of data collection is biased (a kind of sampling bias). In particular, this is likely to lead to response bias where those that respond to the survey might only be those that feel strongly about the topic (could lean mostly positive or negative, not sure on exactly how people will react). What can be known for sure though is that there will be a response bias.

**1.60 Climate Change** In June 2018, a poll asked a random sample of 1000 US adults whether global warming will be a serious problem for the United States.<sup>32</sup> The results show that 51% think global warming will be a very serious problem, 27% think it will be a somewhat serious problem, and 21% think it will not be a serious problem.

- a. What is the sample? What is the intended population?

The sample is the 1000 US adults that were surveyed. The intended population is the whole United States adult population.

- b. Is it reasonable to generalize this result and estimate that 21% of US adults think that global warming will not be a serious problem for the United States?

It is reasonable.

**1.66 Teaching Ability** In a sample survey of professors at the University of Nebraska, 94% of them described themselves as "above average" teachers.

- a. What is the sample? What is the population?

The sample is the professors at the University of Nebraska that answered surveyed. The population is all of the professors at the University of Nebraska.

- b. Based on the information provided, can we conclude that the study suffers from sampling bias?

There is not enough information in my opinion to know whether the study suffers from sampling bias or not. However I am assuming that the survey was sent to all the professors at the University of Nebraska and therefore this study will most likely suffer from response bias because it is only teachers that feel confident in their teaching that responded to the survey.

- c. Is 94% a good estimate for the percentage of above-average teachers at the University of Nebraska? If not, why not?

I would say that 94 % is not a good estimate for the percentage of above-average teachers at the University of Nebraska because as I stated in my last answer it is likely this study suffers from response bias so only teachers that feel confident about their teaching ability are going to respond. This means that this percentage is most likely going to be biased much higher than the true percentage of all teachers at the University of Nebraska

## SECTION 1.3 .

- 1.80** In Exercises 1.76 to 1.81, we give a headline that recently appeared online or in print. State whether the claim is one of association and causation, association only, or neither association nor causation.  
Cat owners tend to be more educated than dog owners.

Association only

- 1.86** Exercises 1.82 to 1.87 describe an association between two variables. Give a confounding variable that may help to account for this association.

Air pollution is higher in places with a higher proportion of paved ground relative to grassy ground.

The confounding variable can be the number of cars and overall industrial and human activity in that area because in places with

a higher proportion of paved ground relative to grassy ground, there is bound to be more vehicular and other kinds of pollution producing activities resulting in higher air pollution in those areas as well.

- 1.88** In Exercises 1.88 to 1.91, we describe data collection methods to answer a question of interest. Are we describing an experiment or an observational study?

To examine whether eating brown rice affects metabolism, we ask a random sample of people whether they eat brown rice and we also measure their metabolism rate.

This is an observational study.

- 1.90** To examine whether planting trees reduces air pollution, we find a sample of city blocks with similar levels of air pollution and we then plant trees in half of the blocks in the sample. After waiting an appropriate amount of time, we measure air pollution levels.

This is an experiment.

- 1.92** Exercises 1.92 to 1.94 refer to questions of interest asked in Section 1.1 in which we describe data collection methods. Indicate whether the data come from an experiment or an observational study.

”Is there a sprinting gene?” Introduced in Example 1.5 on page 9.

This data came from an observational study.

- 1.98 How to Debate a Science Denier** Science deniers oppose robust and valid results of scientific inquiry. A recent study <sup>61</sup> investigates the most effective ways to debunk scientific misinformation. Science advocates can respond to misinformation using topic rebuttal (providing scientific facts on the subject) or technique rebuttal (explaining more generally the false techniques used by science deniers). In the study, 1773 participants first listened to a science denier and were then randomly assigned to one of four conditions: no rebuttal, topic rebuttal, technique rebuttal, or

both topic and technique rebuttal. Participants' attitudes toward the science topic were measured and recorded three times: before participation, after the science denier, and after the rebuttal. Results indicate the importance of having a rebuttal: participants on average were influenced by the science denier, but the influence was mitigated by having any rebuttal. The study further showed that topic or technique rebuttal worked equally well, and having both provided no additional benefit.

- a. QUESTION ON CASES MEANING What are the cases in this study?

The cases are the 1773 participants in the study.

- b. What is the explanatory variable? Is it categorical or quantitative?

The explanatory variable is what happens after the person hears the science denier and it is a categorical variable because there are 4 options as to what happens to the person. They hear no rebuttal, topic rebuttal, technique rebuttal, or both topic and technique rebuttal.

- c. Is this an observational study or an experiment?

This is an experiment

- d. How many rows and how many columns will the dataset have, if we use cases as rows and variables as columns?

The dataset will have 1773 rows and 2 columns

**1.110 Late Night Eating It** is well-known that lack of sleep impairs concentration and alertness, and this might be due partly to late night food consumption. A study <sup>74</sup> took 44 people aged 21 to 50 and gave them unlimited access to food and drink during the day, but allowed them only 4 hours of sleep per night for three consecutive nights. On the fourth night, all participants again had to stay up until 4 am, but this time participants were randomized into two groups; one group was only given access to water from 10 pm until their bedtime at 4am, while the other group still



had unlimited access to food and drink for all hours. The group forced to fast from 10pm on performed significantly better on tests of reaction time and had fewer attention lapses

- a. What are the explanatory and response variables?

The explanatory variable is what group the person was placed in on the fourth day (the fasting group or not) and the response variables are their performance on tests of reaction time and their number of attention lapses.

- b. Is this an observational study or a randomized experiment?

This is a randomized experiment.

- c. Can we conclude that eating late at night worsens some of the typical effects of sleep deprivation (reaction time and attention lapses)?

Yes we can.

- d. Are there likely to be confounding variables? Why or why not? **GO BACK TO**

There are no likely to be many because all the participants are held within a controlled environment. However reaction time still be some confounding variables.

**1.118 Split the Mill?** When the time comes for a group of people eating together at a restaurant to pay their bill, sometimes they might agree to split the costs equally and other times will pay individually. If this decision were made in advance, would it affect what they order? Suppose that you'd like to do an experiment to address this question. The variables you will record are the type of *payment* (split or individual), *sex* of each person, number of *items* ordered, and the *cost* of each person's order. Identify which of these variables should be treated as explanatory and which as response. For each explanatory variable, indicate whether or not it should be randomly assigned.

*Payment.* This is an explanatory variable that should be randomly as-

signed because it is core to the question of the experiment of if they decide ahead of time how they are going to pay (split or individual) and how it effects what they order.

*Sex.* This is an explanatory variable that can be randomly sampled (in a sense of which participants are in the group , which also effects the gender make up of each groups). However obviously you can not just randomly assign a person a gender.

*Items.* This is a response variable.

*Cost.* This is a response variable.

## Unit A Synthesis .

We have seen that random sampling allows us to generalize to a broader population and that random assignment to groups allows us to conclude causation. A study may include either one of these, or both, or neither. For each of the questions in Exercises A.1 to A.6:

- a. Does the study appear to use random sampling? (Yes or No)
- b. Does the study appear to use random assignment? (Yes or No)

**A.4** At a large university, 100 students are randomly selected to take part in a marketing study. Half of these students will be randomly assigned to view one advertising campaign while the other half will watch a different advertising campaign.

- a. Yes it appears to be using random sampling
- b. Yes it appears to be using random assignment

**A.6** In a study examining the reaction of mice to alcohol, all 50 mice in a lab will be randomly assigned to either have alcohol mixed in with their water or to have non-alcoholic liquid of similar taste and caloric content mixed in with their water.

- a. No it appears to not be using random sampling
- b. Yes it appears to be using random assignment