Note # 1: What is Statistics?

Problem 1. Let's start by thinking of the basic principles of statistics and the reason why it is important for research.

a. What are the steps of solving a statistical problem?

b. What is the point of collecting data on a small group of individuals? Why wouldn't we collect data on our entire population?

Solution:

a. Design, description/summarization, probability, inference.

b. It is often not feasible to collect data on the entire population. A sample allows us to learn about the population.

Problem 2. In the previous question, we talked about the purpose of statistics, which leads into descriptive statistics and inferential statistics.

a. What is the difference between descriptive statistics and inferential statistics? Think about the group of individuals that are generally described with each.

b. What is an example of a time that you would need to use descriptive statistics?

c. What is an example of a time when you would need to use inferential statistics?

Solution:

a. Descriptive statistics summarize some groups (sample, population). inferential statistics (statistical inference) make a prediction about a population based on the sample.

b. Take a survey of my entire class, ask them how many pets they have >> make a graph, calculate the average (mean).

c. Use data in part b to estimate the average number of pets for all college students.

Problem 3. It has been stated that getting a flu vaccine can decrease your chances of getting the flu and that it can also help prevent people from having a severe case of the flu. Anna Grace wants to determine if getting a flu vaccine helps reduce the length of time it takes someone to recover from the flu, if they end up getting it. She takes a random sample of 150 individuals who still got the flu, in spite of getting the flu vaccine. The first person she surveys says they recovered from the flu in 4.5 days. The second person she surveys reports that they recovered from the flu in 3 days. The average recovery time for patients in this sample was 4.2 days.

a. What is the data?
b. What is the variable?
c. What is the population?
d. What is the parameter (describe in words)?
e. What is the value of the parameter?
f. What is the sample?
g. What is the statistic (describe in words)?
h. What is the value of the statistic?
i. In order to answer Anna Grace's research question, what other piece of information would we need?

Solution:

a. What is the data? Individual responses; e.g.: 4.5 days, 3 days, … etc.
b. What is the variable? The variable here is the length of time it takes to recover from the flu.
c. What is the population? The population is all people who got the flu after having the vaccine.
d. What is the parameter (describe in words)? Average recovery time of all people who got the flu after having the vaccine.
e. What is the value of the parameter? Unknown (μ=unknown)
f. What is the sample? 150 selected individuals.
g. What is the statistic (describe in words)? Average recovery time of 150 sampled individuals.
h. What is the value of the statistics? $\bar{x} = 4.2$
i. In order to answer Anna Grace's research question, what other piece of information would we need? Average recovery time of people who didn't get the flu shot, but got flu.

Problem 4. Now let's go through an exercise that will help us explore the idea of sampling variability. Remember, not every sample is going to look exactly like the population, which means not every statistic is going to be a perfect estimate of the parameter. Let's try to use a sample to estimate the average number of children in a family living in the United States.

a. Think of 5 people you know well. These people should not have the same parents. You are welcome to include yourself as one of the 5 people. How many children are in each person's family?

b. What is the average number of children in your sample?

c. The average number of children in families living in the United States in 2019 is 1.93. How does this compare with the value in your sample?

d. Let's look at the samples that some of our classmates got. How do they compare with each other?

Solution:

a. My sample

No.	Name	# of children
1.	Mohamed	3
2.	Adam	2
3.	Sara	2
4.	Alex	1
5.	Will	2

b. What is the average number of children in your sample? the average number of children in my sample is:

$$\bar{x} = \frac{3 + 2 + 2 + 1 + 2}{5} = 2\ children$$

c. My sample average (x-bar=2) very close to 1.93 but not exactly the same.

d. Ask your classmates, and write down the average that you obtained.

Problem 5. Look at the scenarios below and describe what type of variable is being measured. Is it numerical or categorical? If it is numerical, is it continuous or discrete? If it is categorical, is it nominal or ordinal?

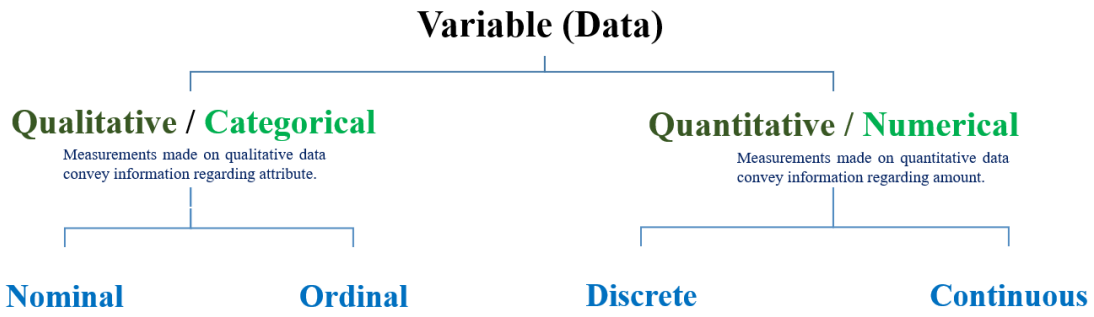
a. A professor is taking a survey of their students. One question she asks each student is what section they are in.

b. A student at Texas A&M is curious and wants to know how many classes other students in his major are taking.

c. A restaurant sends you a survey to fill out after you have ordered take out from them. They ask you to rate the quality of your food on a scale of 1 to 5, where 1 represents awful and 5 represents wonderful.

d. A high school teacher is curious to see how long students spent on an exam. Students had a 65-minute time limit on the exam and could turn it in at any time.

Solution:



a. Categorical - Nominal.

b. Numerical - Discrete.

c. Categorical - Ordinal.

d. Numerical - Continuous.