

## 1.3: Research Questions, Types of Statistical Studies, and Stating Reasonable Conclusions

---

In the last section, we discussed the statistical analysis process. We begin the process by asking a question that can be answered by collecting data. Understanding the type of research question that is being asked helps us to know how we collect data in the next step.

### Research Questions and Types of Statistical Studies

In a **statistical study**, a **population** is a set of all people or objects that share certain characteristics. A **sample** is a subset of the population used in the study. **Subjects** are the individuals or objects in the sample. Subjects are often people, but could be animals, plants, or things. **Variables** are the characteristics of the subjects we study. For example, a variable could be hair color, age, salary, etc. In a previous lesson, we examined the relationship between a person's birth month and chronotype. The population was all adults. The sample was the 30 randomly selected adults. The variables were sets of personality traits and birth date groups.

- A research question about a population.
  - A research question about the causal relationship between two variables.
- 
1. Throughout this course, (a) we will learn how to make estimates about a population, (b) we will test claims about a population, (c) we will compare two populations, and (d) investigate a relationship between two variables (using means, proportions, or standard deviations) by asking **research questions about a population**. State which scenario (a-d) the following questions connect with.
    - Is there a relationship between the number of hours a full-time student works at a job and their GPA?
    - What is the average amount of hours spent studying community college students complete per week?
    - What proportion of community college students work full-time?
    - Do the majority (more than 50%) of community college students work full-time?
    - Do community college students who study more than 36 hours per week have a higher average GPA than those who do not?
    - Does the average amount of hours spent studying by community college students exceed 36 hours per week?

In all of the questions above, the researchers only *observe* subjects in a sample to learn about a population's characteristics. They do not control any of the variables. We call this study an **observational study**.

---

2. We will also examine **research questions about a causal relationship between variables**. For each of the following example questions, circle the word or words that suggest causation.

- a. Does studying more hours for a college class improve test grades?
- b. Does caffeine reduce the number of migraines (long-lasting headaches) for women?
- c. Do violent video games increase crime in the US?

To answer these questions, we investigate how one variable responds as another variable is manipulated or changed. An **explanatory variable** is the (input) variable being manipulated. A **response variable** is the (output) variable used to measure the impact from manipulation of the explanatory variable. An **experiment** involves a manipulation or change to the explanatory variable.

---

3. For the following questions, determine if researchers should conduct an observational study, or an experiment. Justify your choice.

- a. What is the average time it takes to recover from heart surgery?

- b. Do vehicle emissions cause climate change?

- c. What is the approval rating for the governor of California?

- d. Does more regular attendance in high school improve college success?

- e. Is race associated with the maternal mortality rate?

4. Read the following statistical study and answer the questions that follow:

We are interested in learning whether getting more sleep improves one's emotional state (emotional score out of 10 points determined by a professional assessment). We want to see if there is a difference between the emotional score of adults that sleep for 4 hours every night for a week and the emotional score of adults that sleep for 8 hours every night for a week. To investigate this question, we use 100 adult volunteers. The emotional score of each subject will be measured at the beginning of the study. 50 of the volunteers will participate in a sleep program where they are limited to 4 hours of sleep every night for a week. The other 50 participants can sleep for 8 hours every night for a week. At the end of a week, the emotional score will be measured again.

a. What is the research question?

b. Is the question about a population or a causal relationship between two variables?

c. Is this an observational study or an experiment?

1. If this is an observational study, what is the population?

2. If this is an experiment, what are the explanatory and response variables?

d. We need to divide the group of 100 volunteers into two groups of 50 so that there is a fair comparison between the 4 hour and 8 hour sleep groups. What would be a way to create two groups that have similar volunteers?

There are two types of reasonable conclusions that can be drawn from a study.

We may conclude that there is a causal relationship between two variables. This conclusion arises from an experiment when a significant change in the response variable was caused by the manipulation of the explanatory variable. In order to conclude causation, we must make sure that we create experimental groups that are similar. The best way to achieve this is through random assignment.

5. Read the following statistical study and answer the questions that follow:

a. Is this an observational study or an experiment? Justify your answer.

c. Is it reasonable to generalize the conclusion to the population? In other words, is it reasonable to conclude that the score for essays written in cursive was higher than for essays that were not written in cursive, on average? Explain your answer.

6. Read the following statistical study and answer the questions that follow:

Imagine that a psychologist is interested in finding out if listening to classical music has an effect on one's ability to recall material that has been read. The psychologist recruits volunteer students who say they like to study while listening to music. She randomly assigns them into two groups. Each group is told to read a famous poem. One group reads the poem in silence, and the other group reads the poem while they listen to classical music. After reading the poem, they take a brief assessment that asks the students to recall information about the poem. The psychologist concludes that students who listen to classical music while they read score lower than students who read in silence, on average.

- Is this an observational study or an experiment? Justify your answer.
- What is one possible reason for why the students who listened to classical music scored lower than those who did not?
- The psychologist found that the difference was so large that it was unlikely due to chance variation alone. Is it reasonable to conclude that listening to classical music caused students in the sample to score lower on the assessment? Explain.
- Is it reasonable to generalize this conclusion to the population? Why or why not?