1.5: Variables

Learning Objectives

- Define and distinguish between qualitative and quantitative variables
- Define and distinguish between discrete and continuous variables

Introduction to Variables

Recall that **variables** are properties or characteristics of some event, object, or person that can take on different values. Just as there are different types of characteristics, there are different types of variables. Since some variable types do not admit certain computations and analyses, we must pay close attention to the variables and data that we analyze.

Qualitative and Quantitative Variables

At the most general level, we can understand variables as measuring a quality (such as hair color, eye color, religion, favorite movie, gender, and so on) or a quantity (such as height, weight, age, shoe size, temperature, and so on). As such, we have the classification of **qualitative and quantitative variables**, respectively. Intuitively, we might classify variables with numbers as quantitative and variables with characteristics as qualitative. Our intuitive grasp requires further clarification when we consider variables such as race placement $\{1^{st}, 2^{nd}, 3^{rd}, \dots\}$. Just because the values that a variable takes on can be represented or encoded with numbers does not make the variable quantitative. Doing so would render this variable classification meaningless; think about why. Having an inherent order associated with the values that a variable takes on also does not make it quantitative. A **variable is quantitative** if the arithmetic difference between any two values that it takes on is well-defined and informative. While we could subtract 1 from 3 to arrive at 2. We did not gain any additional information. The inherent order already told us that 3^{rd} is two spots lower than 1^{st}; as such, race placement is qualitative. Race finish-time, on the other hand, is a quantitative variable. The arithmetic difference gives us a meaningful measurement as to how much quicker or slower one racer was compared to another. When determining if a variable is quantitative or qualitative, consider if the gaps between possible numerical values add significant meaning.

Text Exercise 1.5.1

1. Most secondary and post-secondary schools classify their students as freshmen, sophomores, juniors, and seniors. Classify class rank as a qualitative or quantitative variable. Explain your reasoning.

Answer

While commonly labeled with $9, 10, 11,$ and 12 including the associated order in class rank, that is not sufficient to warrant a quantitative variable designation. There is no informative arithmetic difference between the different values of the variable; this makes class rank a qualitative variable.

2. Credit card companies typically assign a 16 digit number to an individual to help increase transaction security. Classify customer credit card number as a qualitative or quantitative variable. Explain your reasoning.

Answer

We must ask if the arithmetic difference between credit card numbers is meaningful and informative. Given the numerical encoding, we can compute the subtraction easily enough. Is that difference informative? The credit card number does not reflect the balance, the credit limit, the age of the account, or when the latest number was assigned. When there are two cardholders for a single account, each card holder has a different credit card number. It appears that there is no information gained in knowing the arithmetic difference. The numbers just serve as identifiers to an account; therefore, customer credit card number is qualitative.

3. We mark the passage of time with the month, day, and year. Classify the date as a qualitative or quantitative variable. Explain your reasoning.

Answer

At first glance, values like September 20, 2019 and October 25, 2021 look like they cannot be subtracted. We can count how many days lie between these two values (766 days), and hence the gap between values carries significant meaning. A well-defined arithmetic difference may be difficult write down (it can be done), but intuitively we understand that differences mean how much time has passed between the dates. This is informative. Thus, date is a quantitative variable.

Discrete and Continuous Variables

Quantitative variables can be further classified by the possible values they take on. As quantitative variables, we know that these values in their essence are numbers, but more can be told. Consider the number of people who are in attendance at a FHSU football game. The number fluctuates from game to game, but we know that the possible values are $0, 1, 2, 3, 4, 5, 6, 7, \ldots, 6362$ (Lewis Field's seating capacity). Now consider how long it takes for a football game to finish, denoted with the variable t. There must be four quarters, each with at least 15 minutes of game time and an intermission that is supposed to be at least 20 minutes. If we then consider the play clock, injuries, official reviews, timeouts, overtime, delays, and plays running past time, we would say that the time it takes for a football game to finish could be any number of minutes with 80 minutes as a minimum. For example, it could finish after 83.00000001 minutes, 92.475920 minutes, 92.4759202 minutes, or even $102.1340294478\ldots$ minutes. We could express the possible values as an interval ($\{t \geq 80\}$ or alternatively $[80, \infty)$). If we look at game attendance on the other hand, we could not express the possible values as an interval (one cannot have 7.5 people in attendance); for every pair of distinct possible values, there is a number between them that is not a possible value. Herein lies the distinction between discrete and continuous variables.

A quantitative variable is **continuous** if it can take on any numerical value in some interval of real numbers. A quantitative variable is **discrete** if it is not continuous. A common way to describe a discrete variable is to say that there are gaps between all the possible values that the variable takes on.

? Text Exercise 1.5.2

1. Consider the amount of U.S. currency stored in bank accounts. Classify this variable as a qualitative or quantitative variable. If quantitative, further classify the variable as discrete or continuous. Explain your reasoning.

Answer

The amount of money stored in bank accounts is quantitative because the differences between account balances indicate how much more or less one account has compared to another. The smallest unit of U.S. currency is the penny ($\$0.01$). This implies that there must be a gap between all possible values (one cannot have $\$5.7365$) and that the set of possible values does not contain an interval. Thus U.S. currency in a bank account is a discrete quantitative variable.

2. Consider the floors on an eighteen-story apartment building labeled in order $\{1^{st}, 2^{nd}, 3^{rd}, \ldots, 18^{th}\}$. Classify this variable as a qualitative or quantitative variable. If quantitative, further classify the variable as discrete or continuous. Explain your reasoning.

Answer

The arithmetic difference between two floors returns the same information as the inherent ordering. This variable is thus qualitative. As such, the variable does not get classified as discrete or continuous.

3. Consider the number of floors in apartment buildings. Classify this variable as a qualitative or quantitative variable. If quantitative, further classify the variable as discrete or continuous. Explain your reasoning.

Answer

The values that this variable can take on are $\{1, 2, 3, 4, \ldots\}$. The arithmetic difference between the number of floors indicates the difference in the number of floors between the different apartment buildings. It is therefore, a quantitative variable. The number of floors are counted using whole numbers; therefore, there are gaps between the possible values that the variable takes on. This makes the variable discrete.

4. Consider the heights of adult females. Classify this variable as a qualitative or quantitative variable. If quantitative, further classify the variable as discrete or continuous. Explain your reasoning.

Answer

The arithmetic difference between the heights of two adult females reveals the height disparity between them. Height of adult females is a quantitative variable. The heights could take on any value in an interval (given any two distinct height values there are other possible height values between them) and is therefore continuous.

The classification of discrete and continuous variables focuses on the possible values that a variable takes on. As you may have considered in the previous exercise, there can be a disparity between the values a variable may take on and our ability to measure them. If we based our classification on our ability to measure, every variable would be discrete. If we measured heights accurately enough, any number in some interval could potentially be observed. In principal, someone's height could be 72.65787652998736 inches, despite the fact that we would rarely take the effort to measure that accurately. Contrast this with measuring the number of people in a population; no matter how accurately we measure, we know that it is not possible to obtain a value of 800.78 or 110.2. Thus, the distinction between discrete and continuous variables lies in the possible values that could theoretically be, not the values that may be measured in practice. We are finite, limited beings, and that is good to keep at the forefront of our minds. Remember that we are seeking to understand the reality of the world around us and act accordingly.
