

6.3: Data Collection

Learning Objectives

- Describe how a variable such as height should be recorded
- Choose a good response scale for a questionnaire

Most statistical analyses require that your data be in numerical rather than verbal form (you can't punch letters into your calculator). Therefore, data collected in verbal form must be coded so that it is represented by numbers. To illustrate, consider the data in Table 6.3.1.

Table 6.3.1: *Example Data*

Student Name	Hair Color	Gender	Major	Height	Computer Experience
Norma	Brown	Female	Psychology	5'4"	Lots
Amber	Blonde	Female	Social Science	5'7"	Very little
Paul	Blonde	Male	History	6'1"	Moderate
Christopher	Black	Male	Biology	5'10"	Lots
Sonya	Brown	Female	Psychology	5'4"	Little

Can you conduct statistical analyses on the above data or must you re-code it in some way? For example, how would you go about computing the average height of the 5 students. You cannot enter students' heights in their current form into a statistical program -- the computer would probably give you an error message because it does not understand notation such as 5'4". One solution is to change all the numbers to inches. So, 5'4" becomes $(5 \times 12) + 4 = 64$, and 6'1" becomes $(6 \times 12) + 1 = 73$, and so forth. In this way, you are converting height in feet and inches to simply height in inches. From there, it is very easy to ask a statistical program to calculate the mean height in inches for the 5 students.

You may ask, "Why not simply ask subjects to write their height in inches in the first place?" Well, the number one rule of data collection is to ask for information in such a way as it will be most accurately reported. Most people know their height in feet and inches and cannot quickly and accurately convert it into inches "on the fly." So, in order to preserve data accuracy, it is best for researchers to make the necessary conversions.

Let's take another example. Suppose you wanted to calculate the mean amount of computer experience for the five students shown in Table 6.3.1. One way would be to convert the verbal descriptions to numbers as shown in Table 6.3.2. Thus, "Very Little" would be converted to "1" and "Little" would be converted to "2."

Table 6.3.2: *Conversion of verbal descriptions to numbers.*

1	2	3	4	5
Very Little	Little	Moderate	Lots	Very Lots

Example 6.3.1: How much information should I record?

Say you are volunteering at a track meet at your college, and your job is to record each runner's time as they pass the finish line for each race. Their times are shown in large red numbers on a digital clock with eight digits to the right of the decimal point, and you are told to record the entire number in your tablet. Thinking eight decimal places is a bit excessive, you only record runners' times to one decimal place. The track meet begins, and runner number one finishes with a time of 22.93219780seconds. You dutifully record her time in your tablet, but only to one decimal place, that is 22.9. Race number two finishes and you record 32.7 for the winning runner. The fastest time in Race number three is 25.6. Race number four winning time is 22.9, Race number five is.... But wait! You suddenly realize your mistake; you now have a tie between runner one and runner four for the title of Fastest Overall Runner! You should have recorded more information from the digital clock - that information is now lost, and you cannot go back in time and record running times to more decimal places.

The point is that you should think very carefully about the scales and specificity of information needed in your research before you begin collecting data. If you believe you might need additional information later but are not sure, measure it; you can always decide to not use some of the data, or “collapse” your data down to lower scales if you wish, but you cannot expand your data set to include more information after the fact. In this example, you probably would not need to record eight digits to the right of the decimal point. But recording only one decimal digit is clearly too few.

Example 6.3.2

Pretend for a moment that you are teaching five children in middle school (yikes!), and you are trying to convince them that they must study more in order to earn better grades. To prove your point, you decide to collect actual data from their recent math exams, and, toward this end, you develop a questionnaire to measure their study time and subsequent grades. You might develop a questionnaire which looks like the following:

1. Please write your name: _____
2. Please indicate how much you studied for this math exam:
a lot.....moderate.....little
3. Please circle the grade you received on the math exam: *A B C D F*

Given the above questionnaire, your obtained data might look like the following:

Name	Amount Studied	Grade
John	Little	C
Sally	Moderate	B
Alexander	Lots	A
Linda	Moderate	A
Thomas	Little	B

Eyeballing the data, it seems as if the children who studied more received better grades, but it’s difficult to tell. “Little,” “lots,” and “B,” are imprecise, qualitative terms. You could get more precise information by asking specifically how many hours they studied and their exact score on the exam. The data then might look as follows:

Name	Hours studied	% Correct
John	5	71
Sally	9	83
Alexander	13	97
Linda	12	91
Thomas	7	85

Of course, this assumes the students would know how many hours they studied. Rather than trust the students’ memories, you might ask them to keep a log of their study time as they study.

Contributors and Attributions

- Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/>). Project Leader: David M. Lane, Rice University.
- Heidi Zeimer

This page titled [6.3: Data Collection](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [David Lane](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.