

1.1: Terminology - Individuals/Population/Variables/Samples

One-Variable Statistics: Basics

Terminology: Individuals/Population/Variables/Samples

Oddly enough, it is often a lack of clarity about *who* [or *what*] *you are looking at* which makes a lie out of statistics. Here are the terms, then, to keep straight:

The units which are the objects of a statistical study are called the **individuals** in that study, while the collection of all such individuals is called the **population** of the study.

Note that while the term “individuals” sounds like it is talking about people, the individuals in a study could be things, even abstract things like events.

Example 1.1.2. The individuals in a study about a democratic election might be *the voters*. But if you are going to make an accurate prediction of who will win the election, it is important to be more precise about what exactly is the population of all of those individuals [voters] that you intend to study, but it *all eligible voters, all registered voters, the people who actually voted, etc.*

Example 1.1.3. If you want to study if a coin is “fair” or not, you would flip it repeatedly. The individuals would then be *flips of that coin*, and the population might be something like *all the flips ever done in the past and all that will every be done in the future*. These individuals are quite abstract, and in fact it is impossible ever to get your hands on all of them (the ones in the future, for example).

Example 1.1.4. Suppose we’re interested in studying whether doing more homework helps students do better in their studies. So shouldn’t the individuals be the students? Well, which students? How about we look only at college students. Which college students? OK, how about students at 4-year colleges and universities in the United States, over the last five years – after all, things might be different in other countries and other historical periods.

Wait, a particular student might sometimes do a lot of homework and sometimes do very little. And what exactly does “do better in their studies” mean? So maybe we should look at each student in each class they take, then we can look at the homework they did for that class and the success they had in it.

Therefore, the individuals in this study would be *individual experiences that students in US 4-year colleges and universities had in the last five years*, and population of the study would essentially be the collection of all the names on all class rosters of courses in the last five years at all US 4-year colleges and universities.

When doing an actual scientific study, we are usually not interested so much in the individuals themselves, but rather in

A **variable** in a statistical study is the answer of a question the researcher is asking about each individual. There are two types:

- A **categorical variable** is one whose values have a finite number of possibilities.
- A **quantitative variable** is one whose values are numbers (so, potentially an infinite number of possibilities).

The variable is something which (as the name says) *varies*, in the sense that it can have a different value for each individual in the population (although that is not necessary).

Example 1.1.6 In Example 1.1.2, the variable most likely would be *who they voted for*, a categorical variable with only possible values “Mickey Mouse” or “Daffy Duck” (or whoever the names on the ballot were).

Example 1.1.7 In Example 1.1.3, the variable most likely would be *what face of the coin was facing up after the flip*, a categorical variable with values “heads” and “tails.”

Example 1.1.8 There are several variables we might use in Example 1.1.4. One might be *how many homework problems did the student do in that course*. Another could be *how many hours total did the student spend doing homework over that whole semester, for that course*. Both of those would be quantitative variables.

A categorical variable for the same population would be *what letter grade did the student get in the course*, which has possible values **A**, **A-**, **B+**, ..., **D-**, **F**.

In many [most?] interesting studies, the population is too large for it to be practical to go observe the values of some interesting variable. Sometimes it is not just impractical, but actually impossible – think of the example we gave of all the flips of the coin, even in the ones in the future. So instead, we often work with

A **sample** is a subset of a population under study.

Often we use the variable N to indicate the size of a whole population and the variable n for the size of a sample; as we have said, usually $n < N$.

Later we shall discuss how to pick a good sample, and how much we can learn about a population from looking at the values of a variable of interest only for the individuals in a sample. For the rest of this chapter, however, let's just consider what to do with these sample values.

This page titled [1.1: Terminology - Individuals/Population/Variables/Samples](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jonathan A. Poritz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.