

2.1: Terminology - Explanatory/Response or Independent/Dependent

All of the discussion so far has been for studies which have a single variable. We may collect the values of this variable for a large population, or at least the largest sample we can afford to examine, and we may display the resulting data in a variety of graphical ways, and summarize it in a variety of numerical ways. But in the end all this work can only show a single characteristic of the individuals. If, instead, we want to study a *relationship*, we need to collect two (at least) variables and develop methods of descriptive statistics which show the relationships between the values of these variables.

Relationships in data require at least two variables. While more complex relationships can involve more, in this chapter we will start the project of understanding *bivariate data*, data where we make two observations for each individual, where we have exactly two variables.

If there is a relationship between the two variables we are studying, the most that we could hope for would be that that relationship is due to the fact that one of the variables *causes* the other. In this situation, we have special names for these variables

[def:explanatoryresponsevars] In a situation with bivariate data, if one variable can take on any value without (significant) constraint it is called the **independent variable**, while the second variable, whose value is (at least partially) controlled by the first, is called the **dependent variable**.

Since the value of the dependent variable depends upon the value of the independent variable, we could also say that it is explained by the independent variable. Therefore the independent variable is also called the **explanatory variable** and the dependent variable is then called the **response variable**.

Whenever we have bivariate data and we have made a choice of which variable will be the independent and which the dependent, we write x for the independent and y for the dependent variable.

[eg:depindepvars1] Suppose we have a large warehouse of many different boxes of products ready to ship to clients. Perhaps we have packed all the products in boxes which are perfect cubes, because they are stronger and it is easier to stack them efficiently. We could do a study where

- the *individuals* would be the boxes of product;
- the *population* would be all the boxes in our warehouse;
- the *independent variable* would be, for a particular box, the length of its side in *cm*;
- the *dependent variable* would be, for a particular box, the cost to the customer of buying that item, in US dollars.

We might think that the size *determines* the cost, at least approximately, because the larger boxes contain larger products into which went more raw materials and more labor, so the items would be more expensive. So, at least roughly, the size may be anything, it is a free or *independent* choice, while the cost is (approximately) determined by the size, so the cost is *dependent*. Otherwise said, the size *explains* and the cost is the *response*. Hence the choice of those variables.

[eg:depindepvars3] Suppose we have exactly the same scenario as above, but now we want to make the different choice where

- the *dependent variable* would be, for a particular box, the volume of that box.

There is one quite important difference between the two examples above: in one case (the cost), knowing the length of the side of a box give us a hint about how much it costs (bigger boxes cost more, smaller boxes cost less) but this knowledge is imperfect (sometimes a big box is cheap, sometimes a small box is expensive); while in the other case (the volume), knowing the length of the side of the box perfectly tells us the volume. In fact, there is a simple geometric formula that the volume V of a cube of side length s is given by $V = s^3$.

This motivates a last preliminary definition

[def:deterministic] We say that the relationship between two variables is **deterministic** if knowing the value of one variable completely determines the value of the other. If, instead, knowing one value does not completely determine the other, we say the variables have a **non-deterministic relationship**.

This page titled [2.1: Terminology - Explanatory/Response or Independent/Dependent](#) 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.