Different Types of Data: Further Discussion and Examples

Being able to identify the type of data you have is fundamental both to navigating your way through our book and to selecting a good graphical representation of that data. So, to help you raise your confidence on this, we present more discussion on this issue here than we had space for in the book. This is organized around three fundamental questions you might ask about the type of data you have.

# Are my data quantitative or qualitative?

Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many, how much, or how often). Such data is measured or counted.

Qualitative data, on the other hand, are evaluations of ‘types’ and may be represented by a name, symbol, or a number code. Qualitative data does not involve measuring or counting; rather individuals are categorized. Hence, qualitative data is sometimes called ‘categorical’ data. This implies that the category of interest can take one of a fixed number of levels, and each individual you assess is generally assigned to one of those levels only.

Let us explore an example. Imagine that, at the start of a biostatistics laboratory course, the instructor said ‘Right, there are 76 of us in the lab, let’s collect some data relevant to health and fitness that we can use to explore some data analysis techniques. There is some measuring equipment scattered around the room for you to use.’

Anything that you measure using these various pieces of equipment will be quantitative: this might include height, weight, blood pressure, time taken to run 50m, and lung tidal volume. Anything that you count will also be quantitative—like resting heart rate (beats per minute), or the number of sit-ups that can be performed in a minute. If we ask each student their age as well, then this is also quantitative data because it is counted. This takes a bit of thinking about because if I asked you your age, you would not count back how many birthdays you have had—you would just know your age. But you only know your age because you have been keeping count of the advancing years throughout your life.

Now let’s turn to qualitative data. If we asked people questions about themselves to which the answer would be one from a range of alternative categories, then that will be qualitative data. For example, we might ask ‘Which of the following is the best description of your eye colour: blue, brown, hazel, or green?’ In all likelihood, most members of the class would be comfortably assigning themselves to one of those categories. We could take the same approach to collecting data on sex, hair colour, and blood group. The key feature of this type of data is that that subjects that are evaluated for a given qualitative trait are not subjected to measurement or counting, but rather are assigned to one from an array of categories. The number of categories can be relatively small (e.g. sex) or large (e.g. nationality).

Whether a trait is qualitative or quantitative is not just an intrinsic property of that trait, but also of how we evaluate it. If we measure people’s height, then that will result in quantitative data, but if we ask them ‘Considering your ethnicity, sex, and age, do you consider yourself shorter than average in stature, about average, or taller than average?’ then the data we collect will be qualitative.

As another example, we could assay blood samples for measles antibodies (suggestive of prior immunization or exposure to the disease). If we simply score people as either having appreciable levels of circulating antibodies or not, then this is qualitative data, but if we generate an estimated count of the number of antibodies present per ml of blood, then this is quantitative data.

# My data are quantitative: Are they continuous or discrete?

The question here is whether conceptually your answer could get more and more precise if you measured the variable with ever more accurate equipment. Weight is a good example of this. A person’s weight might be recorded as 80.1kg on their bathroom scales, but on the more precise scales at their local GP surgery, their weight might be recorded as 80.12kg. You could imagine that a set of scales exists in the physics department of their local university that will give a more precise value yet: 80.117kg. It is not impossible to imagine that even more sensitive scales exist, giving ever greater precision. Thus, weight measured this way is continuous.

Whereas if after some point no greater level of precision is possible, then the trait is discrete. The answer to the question ‘How many siblings do you have?’ is quantitative and discrete.

The observed number of sit-ups you perform in a single observed 60-second session is a discrete number (e.g. 10). However, the mean number of sit-ups that you can perform in a 60-second session might be continuous because here we can get an ever more refined answer the more sessions we decide to take the mean value from. At the risk of blowing your mind, the median number of sit-ups you can perform in a 60-second interval remains a discrete variable regardless of how many replicate 60-second trials we consider. This is because the median number has to be the number completed either in a given trial or the number that sits in the middle between two trials—the median cannot be recorded from any scale more precise than that.

The answer to the question ‘How old are you?’ might appear to be something that would yield continuous data—you could imagine notionally starting a stopwatch at the moment when you emerged fully from your mother and stopping it at the end of the question. You could further imagine that this stopwatch could be made more and more precise. However, this idea is completely notional—when I ask you the question ‘How old are you?’ you understand the social convention that I am really asking ‘How many complete years is it since you were born?’—so the question really does yield discrete data.

# My data are qualitative: Are they ordinal or nominal?

This question can be rephrased to ‘Are some orderings of the categories more intuitive than others?’ If the answer is yes, then the data is ordinal; otherwise, it is nominal.

If you were presenting the eye colour data, then the ordering in your table ‘brown, hazel, green, blue’ is not any more natural or intuitive than ‘green, hazel, brown, blue’—so this is nominal data. The same is true of sex or nationality. Whereas if we were presenting data from students’ responses to the question ‘What dress size is most of your clothing?’ then the order ‘8 or below, 10, 12, 14, 16, 18, 20, 22 or above’ is much more natural than ‘22 or above, 8 or below, 14, 12, 20, 10, 18, 16’. Thus, this data is ordinal.

Dress size is interesting because the labels used for the categories are numerical—but they are just labels, and you do not count anything in order to work out your dress size. Similarly, there is no simple measurement you could take (like circumference at the waist) that you could use to convert reliably to dress size.

If we collected data on the nationalities of a sample of people, then it would be natural when you were presenting the data to not have the countries involved listed in a completely random order—you might order them alphabetically, or you might group them so that countries from the same continent appear close to each other. But that is an issue of our convenience in accessing the data rather than a natural ordering. So we maintain that nationality data is nominal rather than ordinal. If in doubt, take three of the categories at random and ask yourself if there is natural ordering to them. We do not think ‘France, Italy, Peru’ is intrinsically more natural than ‘Italy, Peru, France’ whereas for dress sizes ‘8, 12, 16’ does seem a more natural ordering to us than ‘12, 8, 16’.