

1.3: What Can Statistics Do for Us?

There are three major things that we can do with statistics:

- *Describe*: The world is complex and we often need to describe it in a simplified way that we can understand.
- *Decide*: We often need to make decisions based on data, usually in the face of uncertainty.
- *Predict*: We often wish to make predictions about new situations based on our knowledge of previous situations.

Let's look at an example of these in action, centered on a question that many of us are interested in: How do we decide what's healthy to eat?

There are many different sources of guidance, from government dietary guidelines to diet books to bloggers.

Let's focus in on a specific question: Is saturated fat in our diet a bad thing?

One way that we might answer this question is common sense.

If we eat fat then it's going to turn straight into fat in our bodies, right?

And we have all seen photos of arteries clogged with fat, so eating fat is going to clog our arteries, right?

Another way that we might answer this question is by listening to authority figures. The Dietary Guidelines from the US Food and Drug Administration have as one of their Key Recommendations that "A healthy eating pattern limits saturated fats". You might hope that these guidelines would be based on good science, and in some cases they are, but as Nina Teicholz outlined in her book "Big Fat Surprise" (Teicholz 2014), this particular recommendation seems to be based more on the dogma of nutrition researchers than on actual evidence.

Finally, we might look at actual scientific research. Let's start by looking at a large study called the PURE study, which has examined diets and health outcomes (including death) in more than 135,000 people from 18 different countries. In one of the analyses of this dataset (published in *The Lancet* in 2017; Dehghan et al. (2017)), the PURE investigators reported an analysis of how intake of various classes of macronutrients (including saturated fats and carbohydrates) was related to the likelihood of dying during the time that people were followed. People were followed for a *median* of 7.4 years, meaning that half of the people in the study were followed for less and half were followed for more than 7.4 years. Figure 1.1 plots some of the data from the study (extracted from the paper), showing the relationship between the intake of both saturated fats and carbohydrates and the risk of dying from any cause.

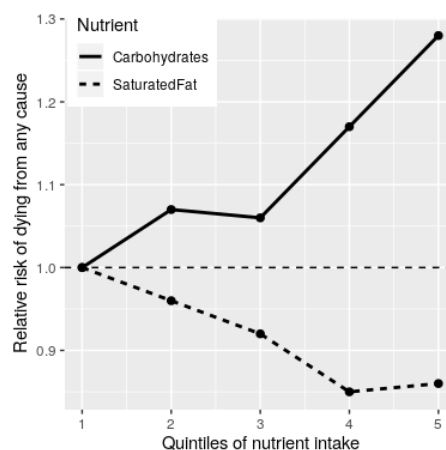


Figure 1.1: A plot of data from the PURE study, showing the relationship between death from any cause and the relative intake of saturated fats and carbohydrates.

This plot is based on ten numbers. To obtain these numbers, the researchers split the group of 135,335 study participants (which we call the "sample") into 5 groups ("quintiles") after ordering them in terms of their intake of either of the nutrients; the first quintile contains the 20% of people with the lowest intake, and the 5th quintile contains the 20% with the highest intake. The researchers then computed how often people in each of those groups died during the time they were being followed. The figure expresses this in terms of the *relative risk* of dying in comparison to the lowest quintile: If this number is greater than 1 it means that people in the group are *more* likely to die than are people in the lowest quintile, whereas if it's less than one it means that people in the group are *less* likely to die. The figure is pretty clear: People who ate more saturated fat were *less* likely to die during the study, with the lowest death rate seen for people who were in the fourth quintile (that is, who ate more fat than the lowest 60% but less than the top

20%). The opposite is seen for carbohydrates; the more carbs a person ate, the more likely they were to die during the study. This example shows how we can use statistics to *describe* a complex dataset in terms of a much simpler set of numbers; if we had to look at the data from each of the study participants at the same time, we would be overloaded with data and it would be hard to see the pattern that emerges when they are described more simply.

The numbers in Figure 1.1 seem to show that deaths decrease with saturated fat and increase with carbohydrate intake, but we also know that there is a lot of uncertainty in the data; there are some people who died early even though they ate a low-carb diet, and, similarly, some people who ate a ton of carbs but lived to a ripe old age. Given this variability, we want to *decide* whether the relationships that we see in the data are large enough that we wouldn't expect them to occur randomly if there was not truly a relationship between diet and longevity. Statistics provide us with the tools to make these kinds of decisions, and often people from the outside view this as *the* main purpose of statistics. But as we will see throughout the book, this need for black-and-white decisions based on fuzzy evidence has often led researchers astray.

Based on the data we would also like to make predictions about future outcomes. For example, a life insurance company might want to use data about a particular person's intake of fat and carbohydrate to predict how long they are likely to live. An important aspect of prediction is that it requires us to generalize from the data we already have to some other situation, often in the future; if our conclusions were limited to the specific people in the study at a particular time, then the study would not be very useful. In general, researchers must assume that their particular sample is representative of a larger *population*, which requires that they obtain the sample in a way that provides an unbiased picture of the population. For example, if the PURE study had recruited all of its participants from religious sects that practice vegetarianism, then we probably wouldn't want to generalize the results to people who follow different dietary standards.

This page titled [1.3: What Can Statistics Do for Us?](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.