What will this chapter tell me?

Although none of us can know the future, predicting it is so important that organisms are hard wired to learn about predictable events in their environment. We saw in the previous chapter that I received a guitar for Christmas when I was 8. My first foray into public performance was a weekly talent show at a holiday camp called ‘Holimarine’ in Wales (it doesn’t exist any more because I am old and this was 1981). I sang a Chuck Berry song called ‘My ding-a-ling’[1] and to my absolute amazement I won the competition.[2] Suddenly other 8-year-olds across the land (well, a ballroom in Wales) worshipped me (I made lots of friends after the competition). I had tasted success, it tasted like praline chocolate, and so I wanted to enter the competition in the second week of our holiday. To ensure success, I needed to know why I had won in the first week. One way to do this would have been to collect data and to use these data to predict people’s evaluations of children’s performances in the contest from certain variables: the age of the performer, what type of performance they gave (singing, telling a joke, magic tricks), and perhaps how cute they looked. A regression analysis on these data would enable us to predict the future (success in next week’s competition) based on values of the predictor variables. If, for example, singing was an important factor in getting a good audience evaluation, then I could sing again the following week; however, if jokers tended to do better then I could switch to a comedy routine. When I was 8 I wasn’t the sad geek that I am today, so I didn’t know about regression analysis (nor did I wish to know); however, my dad thought that success was due to the winning combination of a cherub-looking 8-year-old singing songs that can be interpreted in a filthy way. He wrote a song for me to sing about the keyboard player in the Holimarine Band ‘messing about with his organ’. He said ‘take this song, son, and steal the show’ … and that’s what I did: I came first again. There’s no accounting for taste.

[1] It appears that even then I had a passion for lowering the tone of things that should be taken seriously.

[2] I have a very grainy video of this performance recorded by my dad’s friend on a video camera the size of a medium-sized dog that had to be accompanied at all times by a ‘battery pack’ the size and weight of a tank (see Oditi’s Lantern).

The simple linear model

In the previous chapter we started getting down to the nitty-gritty of the linear model that we've been discussing since way back in Chapter 2. We saw that if we wanted to look at the relationship between two variables we could use the model in equation (2.3):

$$\text{outcome}_i = (bX_i) + \text{error}_i$$

In this model, *b* is the correlation coefficient (more often denoted as *r*) and it is a standardized measure. However, we can also work with an unstandardized version of *b*, but in doing so we need to add something to the model:

$$\text{outcome}_i = (b_0 + b_1 X_i) + \text{error}_i$$
$$y_i = (b_0 + b_1 X_i) + \varepsilon_i$$

The important thing to note is that this equation keeps the fundamental idea that an outcome for a person can be predicted from a model (the stuff in brackets) and some error associated with that prediction (ε_i). We are still predicting an outcome variable (y_i) from a predictor variable (X_i) and a parameter, b_1, associated with the predictor variable that quantifies the relationship it has with the outcome variable. This model differs from that of a correlation only in that it uses an *unstandardized* measure of the relationship (*b*) and consequently we need to include a parameter that tells us the value of the outcome when the predictor is zero.[3] This parameter is b_0.

Focus on the model itself for a minute. Does it seem familiar? Let's imagine that instead of b_0 we use the letter *c*, and instead of b_1 we use the letter *m*. Let's also ignore the error term for the moment. We could predict our outcome as follows:

$$\text{outcome}_i = mx + c$$

Or if you're American, Canadian or Australian let's use the letter *b* instead of *c*:

$$\text{outcome}_i = mx + b$$

Perhaps you're French, Dutch or Brazilian, in which case let's use *a* instead of *m*:

$$\text{outcome}_i = ax + b$$

Do any of these look familiar to you? If not, there are two explanations: (1) you didn't pay enough attention at school, or (2) you're Latvian, Greek, Italian, Swedish, Romanian, Finnish or Russian – to avoid this section being even more tedious, I used only the three main international differences in the equation above. The different forms of the equation make an important point: the symbols or letters

[3] In case you're interested, by standardizing *b*, as we do when we compute a correlation coefficient, we're estimating *b* for standardized versions of the predictor and outcome variables (i.e., versions of these variables that have a mean of 0 and standard deviation of 1). In this situation b_0 drops out of the equation because it is the value of the outcome when the predictor is 0, and when the predictor and outcome are standardized then when the predictor is 0, the outcome (and hence b_0) will be 0 also.

we use in an equation don't necessarily change it.[4] Whether we write $mx + c$ or $b_1X + b_0$ doesn't really matter, what matters is what the symbols represent. So, what do the symbols represent?

Hopefully, some of you recognized this model as 'the equation of a straight line'. I have talked throughout this book about fitting 'linear models', and linear simply means 'straight line'. So, it should come as no surprise that the equation we use is the one that describes a straight line. Any straight line can be defined by two things: (1) the slope (or gradient) of the line (usually denoted by b_1); and (2) the point at which the line crosses the vertical axis of the graph (known as the *intercept* of the line, b_0). These parameters b_1 and b_0 are known as the **regression coefficients** and will crop up time and time again in this book, where you may see them referred to generally as *b* (without any subscript) or ***bn*** (meaning the *b* associated with variable *n*). A particular line (i.e., model) will have a specific intercept and gradient.

Figure 8.2 shows a set of lines that have the same intercept but different gradients. For these three models, b_0 will be the same in each but the values of b_1 will differ in each model.

Figure 8.2 also shows models that have the same gradients (b_1 is the same in each model) but different intercepts (the b_0 is different in each model). I've mentioned already that b_1 quantifies the relationship between the predictor variable and the outcome, and Figure 8.2 illustrates this point. In Chapter 6 we saw how relationships can be either positive or negative (and I don't mean whether or not you and your partner argue all the time). A model with a positive b_1 describes a positive relationship, whereas a line with a negative b_1 describes a negative relationship. Looking at Figure 8.2 (left), the red line describes a positive relationship whereas the green line describes a negative relationship. As such, we can use a linear model (i.e., a straight line) to summarize the relationship between two variables: the gradient (b_1) tells us what the model looks like (its shape) and the intercept (b_0) tells us where the model is (its location in geometric space).

This is all quite abstract, so let's look at an example. Imagine that I was interested in predicting physical and downloaded album sales (outcome) from the amount of money spent advertising that album (predictor). We could summarize this relationship using a linear model by replacing the names of our variables into equation (8.1):

$$\begin{aligned} y_i &= b_0 + b_1X_i + \varepsilon_i \\ \text{album sales}_i &= b_0 + b_1\text{advertising budget}_i + \varepsilon_i \end{aligned} \tag{8.2}$$

Once we have estimated the values of the *b*s we would be able to make a prediction about album sales by replacing 'advertising' with a number representing how much we wanted to spend advertising an album. For example, imagine that b_0 turned out to be 50 and b_1 turned out to be 100. Our model would be:

$$\text{album sales}_i = 50 + (100 \times \text{advertising budget}_i) + \varepsilon_i \tag{8.3}$$

Note that I have replaced the betas with their numeric values. Now, we can make a prediction. Imagine we wanted to spend £5 on advertising, we can replace the variable 'advertising budget' with this value and solve the equation to discover how many album sales we will get:

$$\begin{aligned} \text{album sales}_i &= 50 + (100 \times 5) + \varepsilon_i \\ &= 550 + \varepsilon_i \end{aligned}$$

---

[4] For example, you'll sometimes see equation (8.1) written as $Y_i = (b_0 + b_1X_i) + e_i$. The only difference is that this equation has bs in it instead of *b*s. Both versions are the same thing, they just use different letters to represent the coefficients.

So, based on our model we can predict that if we spend £5 on advertising, we'll sell 550 albums. I've left the error term in there to remind you that this prediction will probably not be perfectly accurate. This value of 550 album sales is known as a **predicted value**.