What will this chapter tell me?

We discovered in the previous chapter that I wrote a book. This book. There are a lot of good things about writing books. The main benefit is that your parents are impressed. They're not *that* impressed, because they think that a good book sells as many copies as *Harry Potter* and that people should queue outside bookshops for the latest enthralling instalment of *Discovering statistics* …. My parents are, consequently, quite baffled about how this book is seen as successful, yet I don't get invited to dinner by the Queen. Nevertheless, given that my family don't really understand what I do, books are tangible proof that I do *something*. The size of this book and the fact it has equations in it is an added bonus because it makes me look cleverer than I actually am. However, there is a price to pay, which is immeasurable mental anguish. In England we don't talk about our emotions, because we fear that if they get out into the open, civilization as we know it will collapse, so I definitely will not mention that the writing process for the second edition was so stressful that I came within one of Fuzzy's whiskers of a total meltdown. It took me two years to recover, just in time to start thinking about the third edition. Still, it was worth it because the feedback suggests that some people found the book vaguely useful. Of course, the publishers focus less on the book's helpfulness and more on sales figures and comparisons with other books. They have databases that have sales figures for this book and its competitors in different 'markets' (you are not a person, you are a 'consumer', and you don't live in a country, you live in a 'market'), and they gibber and twitch at their consoles creating pink frequency distributions (with 3-D effects) of these values. The data they get are frequency data (the number of books sold in a certain period of time). Therefore, if they wanted to compare sales of this book to its competitors, in different countries, they would need to read this chapter because it's all about analysing data for which we know only the frequency with which events occur. Of course, they won't read this chapter, but they should …

Analysing categorical data

So far we have looked at fitting models with categorical predictor variables, but always predicting a continuous outcome variable. Sometimes, however, we want to predict categorical outcome variables. In other words, we want to predict into which category an entity falls. For example, we might want to predict whether someone is pregnant or not, for which political party a person voted, whether a tumour is benign or malignant, whether a sports team will win, lose or draw. In all of these cases, an entity can fall into only one category, for example a woman can be pregnant or not; she can't be 'a bit pregnant'. The next two chapters deal with statistical models for categorical outcomes. We'll begin with some basic models of associations between categorical variables, then look at predicting categorical outcomes from categorical predictors, then in the next chapter we'll move on to look at predicting categorical outcomes from both categorical and continuous predictor variables.