What will this chapter tell me?

Over the last couple of chapters we saw that I had gone from a child having dreams and aspirations of being a rock star, to becoming a living (barely) statistical test. A more dramatic demonstration of my complete failure to achieve my life's ambitions I can scarcely imagine. Having devoted far too much of my life to statistics, it was time to unlock the latent rock star once more. The second edition of this book had left me in desperate need of some therapy, so at the age of 29 I decided to learn to play the drums (there's a joke in there somewhere about it being the perfect instrument for a failed musician, but really they're much harder to play than people think). A couple of years later I had a call from an old friend of mine, Doug, who used to be in a band that my old band Scansion used to play with a lot: 'Remember the last time I saw you we talked about you coming and having a jam with us?' I had absolutely no recollection whatsoever of him saying this, so I responded 'Yes'. 'Well, how about it then?' he said. 'OK,' I said, 'you arrange it and I'll bring my guitar.' 'No, you whelk,' he said, 'we want you to drum. Can you learn some of the songs on the CD I gave you last year?' I'd played his band's CD and I liked it, but their songs were ridiculously fast and there was no way on earth that I could play them. 'Sure, no problem', I lied. I spent the next two weeks trying to become a much better drummer than I was, playing along to this CD as if my life depended on it. I'd love to report that when the rehearsal came I astounded them with my brilliance, but I didn't. I did, however, nearly have a heart attack and herniate everything in my body that it's possible to herniate. Still, we had another rehearsal, and then another and, 7 years down the line, we're still having them. The main difference now is that I play the songs at a speed that makes their old drummer sound like a sedated snail (www.myspace.com/fracturepattern). It's curious that I started off playing guitar (which I can still play, incidentally), and then I chose drums. Within famous bands, there are always assumptions about the personalities of different musicians: the singers are egocentric, guitarists are perceived to be cool, bassists introverted and happy to blend into the background, and drummers are supposed to be crazy hedonists, autistic (enjoying counting helps) or both. I'm definitely more autistic than hedonistic. If we wanted to test what personality characteristics predict the instrument you choose to play, then we'd have a categorical outcome (type of instrument) with several categories (drums, guitar, bass, singing, keyboard, tuba, etc.) and continuous predictors (neuroticism, extroversion, etc.). We've looked at how we can quantify associations between purely categorical variables, but if we have continuous predictors too then surely there's no model on earth that can handle that kind of complexity; should we just go to the pub and have a good time instead? Actually, we can do **logistic regression** – bugger!

Background to logistic regression

In the last chapter we started to look at how we fit models of the relationships between categorical variables. We have also seen throughout the book how we can use categorical variables to predict continuous outcomes. However, we haven't looked at the reverse process: predicting categorical outcomes from continuous or categorical predictors. In a nutshell, logistic regression is multiple regression but with an outcome variable that is categorical and predictor variables that are continuous or categorical. In its simplest form, this means that we can predict which of two categories a person is likely to belong to given certain other information. A trivial example is to look at which variables predict whether a person is male or female. We might measure laziness, pig-headedness, alcohol consumption and daily flatulence. Using logistic regression, we might find that all of these variables predict the gender of the person. More important, the model we build will enable us to predict whether a new person is likely to be male or female based on these variables. So, if we picked a random person and discovered that they scored highly on laziness, pig-headedness, alcohol consumption and flatulence, then our model might tell us that, based on this information, this person is likely to be male. Logistic regression can have life-saving applications. In medical research it is used to generate models from which predictions can be made about the likelihood that a tumour is cancerous or benign (for example). A database of patients is used to establish which variables are influential in predicting the malignancy of a tumour. These variables can then be measured for a new patient and their values placed in a logistic regression model, from which a probability of malignancy could be estimated. If the probability value of the tumour being malignant is low then the doctor may decide not to carry out expensive and painful surgery that in all likelihood is unnecessary. We might not face such life-threatening decisions, but logistic regression can nevertheless be a very useful tool. When we are trying to predict membership of only two categorical outcomes the analysis is known as **binary logistic regression**, but when we want to predict membership of more than two categories we use **multinomial (or polychotomous) logistic regression**.