What will this chapter tell me?

At the age of 15, I was on holiday with my friend Mark (the drummer) in Cornwall. I had a pretty decent mullet by this stage (nowadays I just wish I had enough hair to grow a mullet) and had acquired a respectable collection of heavy metal T-shirts from going to various gigs. We were walking along the cliff tops one evening at dusk reminiscing about our times in Andromeda. We came to the conclusion that the only thing we hadn't enjoyed about that band was Malcolm and that maybe we should reform it with a different guitarist.[1] As I was wondering who we could get to play guitar, Mark pointed out the blindingly obvious: I played guitar. So, when we got home Scansion was born.[2] As the singer, guitarist and songwriter I set about writing some songs. I moved away from writing about flies and set my sights on the pointlessness of existence, death, betrayal and so on. We had the dubious honour of being reviewed in the music magazine *Kerrang!* (in a live review they called us 'twee', which is really not what you want to be called if you're trying to make music so heavy that it ruptures the bowels of Satan). Our highlight, however, was playing a gig at the famous Marquee Club in London (this club has closed, not as a result of us playing there I hasten to add, but in its day it started the careers of people like Jimi Hendrix, the Who, Iron Maiden and Led Zeppelin).[3] This was the biggest gig of our career and it was essential that we played like we never had before. As it turned out, we did: I ran on stage, fell over and in the process de-tuned my guitar beyond recognition and broke the zip on my trousers. I spent the whole gig out of tune and spread-eagle to prevent my trousers falling down. Like I said, I'd never played like *that* before. We used to get quite obsessed with comparing how we played at different gigs. I didn't know about statistics then (happy days), but if I had I would have realized that we could rate ourselves and compare the mean ratings for different gigs; because we would always be the ones rating the gigs, this would be a repeated-measures design, so we would need a repeated-measures ANOVA to compare these means. That's what this chapter is about; hopefully it won't make our trousers fall down.

[1] I feel bad about saying this because Malcolm was a very nice guy and, to be honest, at that age (and some would argue beyond) I could be a bit of a cock.

[2] Scansion is a term for the rhythm of poetry. We got the name by searching through a dictionary until we found a word that we liked. Originally we didn't think it was 'metal' enough, and we decided that any self-respecting heavy metal band needed to have a big spiky 'X' in their name. So, for the first couple of years we spelt it 'Scanxion'. Like I said, I could be a bit of a cock back then.

[3] http://www.themarqueeclub.net

Introduction to repeated-measures designs

So far in this book, when looking at comparing means, we've concentrated on situations in which different entities contribute to different means; for example, different people take part in different experimental conditions. It doesn't have to be different people, it could be different plants, companies, plots of land, viral strains, goats or even different duck-billed platypuses (or whatever the plural is). I've completely ignored situations in which the same people (plants, goats, hamsters, seven-eyed green galactic leaders from space, or whatever) contribute to the different means. I've put it off long enough, and now I'm going to take you through what happens when we do ANOVA on repeated-measures data.

'Repeated measures' is a term used when the same entities participate in all conditions of an experiment or provide data at multiple time points. For example, you might test the effects of alcohol on enjoyment of a party. Some people can drink a lot of alcohol without really feeling the consequences, whereas others, like myself, have only to sniff a pint of lager and they start flapping around on the floor waving their arms and legs around shouting 'Look at me, I'm Andy, King of the lost world of the Haddocks'. Therefore, it is important to control for individual differences in tolerance to alcohol, and this can be achieved by testing the same people in all conditions of the experiment: participants could be given a questionnaire assessing their enjoyment of the party after they had consumed 1 pint, 2 pints, 3 pints and 4 pints of lager. There are lots of different ways to refer to this sort of design (Figure 14.2).

We saw in Chapter 1 that this type of design has several advantages; however, in Chapter 11 we saw that the accuracy of the *F*-test in ANOVA depends upon the assumption that scores in different conditions are independent (see Section 11.3). When repeated measures are used this assumption is violated: scores taken under different experimental conditions are likely to be related because they come from the same entities. As such, the conventional *F*-test will lack accuracy. The relationship between scores in different treatment conditions means that we have to make an additional assumption; put simplistically, we assume that the relationship between pairs of experimental conditions is similar (i.e., the level of dependence between experimental conditions is roughly equal). This assumption is called the assumption of **sphericity**, which, trust me, is a pain in the butt to pronounce when you're giving statistics lectures at 9 a.m. on a Monday.

The assumption of sphericity

The assumption of sphericity can be likened to the assumption of homogeneity of variance in between-groups ANOVA. Sphericity (denoted by ε and sometimes referred to as *circularity*) is a more general condition of **compound symmetry**. Compound symmetry holds true when both the variances across conditions are equal (this is the same as the homogeneity of variance assumption in between-groups designs) and the covariances between pairs of conditions are equal. So, we assume that the variation within experimental conditions is fairly similar and that no two conditions are any more dependent than any other two. Although compound symmetry has been shown to be a sufficient condition for ANOVA using repeated-measures data, it is not a necessary condition. Sphericity is a less restrictive form of compound symmetry and refers to the equality of variances of the *differences* between treatment levels. So, if you were to take each pair of treatment levels, and calculate the differences between each pair of scores, then it is necessary that these differences have approximately equal variances. As such, *you need at least three conditions for sphericity to be an issue.*