What will this chapter tell me?

Having failed to become a rock star, I went to university and eventually ended up doing a Ph.D. (in Psychology) at the University of Sussex. Like many postgraduates, I taught to survive. I was allocated to second-year undergraduate statistics. I was very shy at the time, and I didn't have a clue about statistics, so standing in front of a room full of strangers and talking to them about ANOVA was about as appealing as dislocating my knees and running a marathon. I obsessively prepared for my first session so that it would go well; I created handouts, I invented examples, I rehearsed what I would say. I went in terrified but knowing that if preparation was any predictor of success then I would be OK. About half way through one of the students rose majestically from her chair. An aura of bright white light surrounded her and she appeared to me as though walking through dry ice. I guessed that she had been chosen by her peers to impart a message of gratitude for the hours of preparation I had done and the skill with which I was unclouding their brains of statistical mysteries. She stopped inches away from me. She looked into my eyes and mine raced around the floor looking for the reassurance of my shoelaces. 'No one in this room has a rabbit[1] clue what you're going on about', she spat before storming out. Scales have not been invented yet to measure how much I wished I'd run the dislocated-knees marathon that morning. To this day I have intrusive thoughts about students in my lectures walking zombie-like towards the front of the lecture theatre chanting 'No one knows what you're going on about' before devouring my brain in a rabid feeding frenzy.

   The aftermath of this trauma is that I threw myself into trying to be the best teacher in the universe. I wrote detailed handouts and started using wacky examples. Based on these I was signed up by a publisher to write a book. This book. At the age of 23 I didn't realize that this was academic suicide (really, textbooks take a long time to write and they are not at all valued compared to research articles), and I also didn't realize the emotional pain I was about to inflict on myself. I soon discovered that writing a statistics book was like doing a factor analysis: in factor analysis we take a lot of information (variables) and SPSS effortlessly reduces this mass of confusion into a simple message (fewer variables). SPSS does this in a few seconds. Similarly, my younger self took a mass of information about statistics that I didn't understand and filtered it down into a simple message that I *could* understand: I became a living, breathing factor analysis … except that, unlike SPSS, it took me two years and some considerable effort.

---

[1] She didn't say 'rabbit', but she did say a word that describes what rabbits do a lot; it begins with an 'f' and the publishers think that it will offend you.

When to use factor analysis

In science we often need to measure things that cannot be measured directly (so-called latent variables). For example, management researchers might be interested in measuring 'burnout', which is when someone who has been working very hard on a project (a book, for example) for a prolonged period of time suddenly finds himself devoid of motivation, inspiration, and wants to repeatedly headbutt their computer, screaming 'please, Mike, unlock the door, let me out of the basement, I need to feel the soft warmth of sunlight on my skin'. You can't measure burnout directly: it has many facets. However, you can measure different aspects of burnout: you could get some idea of motivation, stress levels, whether the person has any new ideas and so on. Having done this, it would be helpful to know whether these facets reflect a single variable. Put another way, are these different measures driven by the same underlying variable?

This chapter explores factor analysis and principal component analysis (PCA) – techniques for identifying clusters of variables. These techniques have three main uses: (1) to understand the structure of a set of variables (e.g., Spearman and Thurstone used factor analysis to try to understand the structure of the latent variable 'intelligence'); (2) to construct a questionnaire to measure an underlying variable (e.g., you might design a questionnaire to measure burnout); and (3) to reduce a data set to a more manageable size while retaining as much of the original information as possible (e.g., factor analysis can be used to solve the problem of multicollinearity that we discovered in Chapter 8 by combining variables that are collinear).

There are numerous examples of the use of factor analysis in science. Most readers will be familiar with the extroversion–introversion and neuroticism traits measured by Eysenck (1953). Most other personality questionnaires are also based on factor analysis – notably Cattell's (1966a) 16 personality factors questionnaire – and these inventories are frequently used for recruiting purposes in industry (and even by some religious groups). Economists, for example, might also use factor analysis to see whether productivity, profits and workforce can be reduced down to an underlying dimension of company growth, and Jeremy Miles told me of a biochemist who used it to analyse urine samples.

Both factor analysis and PCA aim to reduce a set of variables into a smaller set of dimensions (called 'factors' in factor analysis and 'components' in PCA). To non-statisticians, like me, the differences between a component and a factor are difficult to conceptualize (they are both linear models), and the differences are hidden away in the maths behind the techniques.[2] However, there are important differences between the techniques, which I'll discuss in due course. Most of the practical issues are the same regardless of whether you do factor analysis or PCA, so once the theory is over you can apply any advice I give to either factor analysis or PCA.

---

[2] PCA is not the same as factor analysis. This doesn't stop idiots like me from discussing them as though they are. I tend to focus on the similarities between the techniques, which will reduce some statisticians (and psychologists) to tears. I'm banking on these people not needing to read this book, so I'll take my chances because I think it's easier for you if I give you a general sense of what the procedures do and not obsess too much about their differences. Once you have got the basics under your belt, feel free to obsess about their differences and complain to all of your friends about how awful the book by that imbecile Field is …