Lies, Damned Lies and…

In virtually all of the studies discussed in this book, the authors interpreted their data using some kind of statistical analysis. In general, I have spared you the nitty-gritty details of these analyses, because you can understand the studies perfectly well without them. But in this section, I would like to take your hand and lead you very gently through the logic and methods of basic statistics. We will even do a couple of statistical analyses together. Don't worry, you won't have to do any maths; the hardest thing you will have to do is type some numbers into a spreadsheet.

Let's start out with a very simple – if sexist - hypothesis: Men will be able to name more Premiership football managers [Americans readers might prefer to substitute Major League Baseball coaches] than women. In order to test this hypothesis, we get 10 men and 10 women and ask them to name as many managers as possible. Let's say the men can name an average (mean) of 10.5 managers, whilst the women can name an average (mean) of 8.5 managers. This means we can say that our prediction was correct and go home, right?

Wrong. The problem is that if you take a group of 20 participants (wearing bin bags over their whole bodies so that nobody can tell their gender) and split them into two groups of 10, it's pretty unlikely that the average number of managers that each group can name will be *identical*. Most likely, the average number of managers named by each group will be different, if pretty similar; for example 9.8 for one group and 9.2 for another. Now, here comes the crucial part. In our original example, one group (the men) named an average of 10.5 managers, whilst another group (the women) named an average of 8.5. **So, how do we know that the difference between the two groups is due to the fact that one is male and one is female, as opposed to the fact that if you divide people into two groups at random, one will *just so happen* to score a bit higher than the other?** The answer is that we don't. And this is why we need statistical testing.

So, simplifying only a little bit, what the statistical test does is this: First, it looks to see how big the difference between the two groups is (in our case, the difference the groups is 2.0 because the men and women named 10.5 and 8.5 managers respectively). Then it tells us how likely it is that we would get a difference between groups that is this big *by pure chance* if we just created the groups at random (for example using the bin-bags over the body method). As for exactly *how* it works this out, frankly who know and who cares? All you need to know is that if the chances of a difference of this size arising by pure fluke are very small, then we conclude that the difference was *not* due to pure fluke, and instead to the fact that one group consisted of men, and the other of women.

How small is "very small"? Well, when you run a statistical test (as we will do a minute) it gives you a number – called the *p* value – that tells you exactly how likely it is that a difference between the two groups that is as big as the one that you've got could have turned up by chance alone. The easiest way to understand this number is to multiply it by 100, to convert it into a percentage. If the resulting number is less than 5% - i.e., there a less than a 5% chance that a between-groups difference of this size could have arisen by fluke, we conclude that the difference did not arise by fluke. In technical terms, we say the difference was "statistically significant".

Statistical testing, like many things, is much easier to understand if you actually do it. Ideally, you should actually go out and find 10 men and 10 women and ask each of them to name as many Premier League managers as possible. If you can't be bothered, you can use my made-up scores below.

Men	Managers named	Women	Managers named
Participant 1	11	Participant 11	7
Participant 2	4	Participant 12	7
Participant 3	17	Participant 13	6
Participant 4	15	Participant 14	13
Participant 5	8	Participant 15	9
Participant 6	9	Participant 16	8
Participant 7	20	Participant 17	10
Participant 8	10	Participant 18	8
Participant 9	4	Participant 19	9
Participant 10	7	Participant 20	8

Type these results (either your real ones or my fictitious ones) into the spreadsheet **Statistics1.xls** which you can download from the companion website (www.psychoddities.com). The spreadsheet will run a statistical test, and tell you whether any difference between the groups meets the criteria for *statistical significance* outlined above (i.e.., if there's less than a 5% chance that a difference this big could have arisen by pure fluke if the groups were formed at random).

If you used my made-up scores, you will notice that although the women and men score a mean of 8.5 and 10.5 managers respectively, this difference is *not* statistically significant. At 28%, the probability of getting a difference this big

between two groups purely by chance – whilst not *that* high – is too high for us to be able to say with confidence that this isn't happening here.

By the way, this spreadsheet isn't specific to comparing men and women on the number of managers they can name; you can use it to compare any two groups on anything at all. Simply replace "men" (C6) and "women" (E6) with the names of the two groups that you want to compare (perhaps "northerners" and "southerners" or "young people" and "old people") and "number of Premiership managers named" (B1) with whatever you want to compare them on (perhaps "number of pies eaten per week" or "number of own teeth still remaining").

Finally, it's important to be aware that this idea of testing whether apparently-meaningful differences between groups could have arisen by chance alone – which is done as a matter of course in all sciences – has not entered mainstream public consciousness. Newspapers are *always* reporting as meaningful tiny differences between groups that are *extremely* likely to have arisen by chance alone. Worse still, I have seen newspaper articles in which a researcher's claim that – for example – a miniscule increase in crime rates is "not statistically significant" is portrayed as the special pleading of a boffin who is determined to hide the truth. In fact, the truth is that if an "increase" in crime rates is not statistically significant, there's no reason to think that crime rates have actually changed at all.

So, yes, there are lies, damned lies and statistics. But the damnedest lies are told not when statistical tests are used, but when they are ignored.

Dr Ben Ambridge
University of Liverpool

www.Psychoddities.com
www.BenAmbridge.com