

# contents

<i>Preface</i> .....	ix
<i>Acknowledgements</i> .....	xiii
<b>Part 1 Thinking about research</b>	
Chapter 1 The problem with people: variation and hypothesis testing .....	5
Chapter 2 Who will you study? samples and populations.....	23
Chapter 3 How will you describe your findings?.....	45
Chapter 4 What have you found? inferential statistics .....	87
<b>Part 2 Doing research</b>	
Chapter 5 Designing a study: running for the bus.....	99
Chapter 6 Designing a study: the best way to manage people.....	115
Chapter 7 Practical: it's a matter of taste .....	137
Chapter 8 Practical: abnormal cells on a slide – how people search for things.....	151
Chapter 9 Practical: constructing memories .....	161
Chapter 10 Practical: I'm trying to concentrate!.....	169
Chapter 11 Practical: what do you think? attitude research .....	181
Chapter 12 Practical: something's not right .....	193
Chapter 13 Practical: too many cooks or not enough heads? .....	199
Chapter 14 Practical: head size and intelligence .....	209

**Part 3 The really useful section**

Chapter 15	Think critically: don't believe everything you read .....	219
Chapter 16	Sample write-ups.....	225
<i>Glossary</i>	.....	231
<i>Index</i>	.....	253

# 1 part You are already thinking about research

## thinking about research

Right at this moment you are reading the first sentence of a book on how to do research with people. You can't be doing this by accident. We definitely remember asking our publishers to write something on the cover to say that this book is all about doing research. You chose to ignore this warning, so you must already be thinking about research – well done!

Okay, so we're being a bit silly here: it's the first book we've written together and we're a little giddy with the excitement. Bear with us, we'll have calmed down in a few paragraphs.

Actually, though, in amongst all this nonsense we are trying to make quite a serious point: you might well think that reading this book is the first step you have ever taken towards learning how to do research with people, but the truth is you have been thinking about how to study people all of your life. Pretty much from the moment you were born, you instinctively experimented with your caregivers to learn how they would react to various things you did – how else would you have learnt what is and is not acceptable in your culture? Today, every time you ask a question like 'I wonder why some people are more popular than others' or 'I wonder why people seem to get more aggressive when they drive a car,' you are setting up questions about people which could be addressed through systematic research. Indeed, if you put your mind to it, we are sure you would get at least part-way towards answering those questions right now.

And that is the point we wish to make: we are not about to teach you a completely new skill here. Instead, we are going to make you better at doing something you already know about, at least in part. Obviously we hope to teach you a lot of new information in this book. But we will also spend time taking rules and concepts that you already know in a fairly subconscious, intuitive sort of way, and making that knowledge more explicit. When you are properly aware of the ideas you already have about how research should be carried out, you can more easily use this knowledge when you conduct research yourself. You will also be better equipped to use other people's research: you will be able to spot mistakes in the data you come across, to know what questions you need to ask when you hear about research in the news, and to know which studies you can trust and which you cannot trust when making decisions about your own life.

Let us prove to you that you already know quite a lot about doing research with people, even if you have never thought about it before. As it happens we are both amateur pharmacologists and have worked away in our basements to produce a new drug called Holtodol. Holtodol not only makes you a great deal healthier and improves your memory, but also makes you more attractive to whichever gender it is you like to attract. Sounds good, doesn't it? So if we were to offer you some Holtodol, would you take it?

How about this: would you take the drug if we had tested it on two people, both 21-year-old men, for a week, and neither experienced any side-effects?

Would you take the drug if we had tested it on 10,000 people, with a mixture of men and women of various ages, for 10 years, and none of them experienced any side-effects?

We strongly suspect you would feel happier taking Holtodol in the second situation. The reason for this is that you already know, at some level, that although the finding is exactly the same – no side-effects – this is much more convincing when it comes from a large, mixed sample of people over a long period of time than when it comes from just two people over a few days.

Let us give you another example, which illustrates some of the issues that arise when we deal with people's feelings. There are two hospitals near where we live. Each claims its employees have the best working conditions in the area. Obviously they can't both be right, so we decide to put it to the test.

We take a member of staff from each hospital. Specifically, from Hospital A we take Professor Margaret Pritt, the 56-year-old Head of Ophthalmology, and from hospital B we take David Fish, a 19-year-old car park attendant. We ask each of them how happy they are at work. David Fish says he is 'really happy' whereas Margaret Pritt says she is 'fairly happy'. We therefore conclude that Hospital B treats its employees better, and suggest that Hospital A should change its working practices to be more like Hospital B.

Is this a fair conclusion? Should we really use this research to change policies throughout Hospital A? We suspect that right now you are violently shaking your head, and quite possibly wondering if you still have the receipt for this book so you can return it.

The thing is, we spoke to only one person in each hospital – there's every possibility we have chosen completely inappropriate people for our comparison. You can clearly see that the people we interviewed had very different jobs: one was the head of a clinical unit with decades of employment history and lots of responsibilities; her experience of work is likely to be totally different to that of a worker not long out of school. And it isn't just the differences in their jobs that might be a problem: David and Margaret are different genders, and are at completely different stages in their lives – middle-aged workers like Margaret could well be juggling their job with family responsibilities, which might affect their enjoyment of work. On the other hand, a younger worker like David might be affected by having less money and job security.

And all this is just the tip of the iceberg. There are a huge number of other problems with the study we have just described. For example, simply asking people 'how happy' they are is too vague and subjective: it would have been better to ask several questions, each using more objective terms, such as 'How often do you find you simply cannot

face going to work?'. And is 'happiness' even the right thing to be measuring in a study of working conditions anyway?

So we are confident that when you read our description of the hospital study, you immediately felt there was something wrong with it, even if you could not put your finger on what exactly it was. This shows that, even when addressing a fairly complex question, you already have some knowledge of what is good and bad research. In the rest of this book we will be developing this knowledge and adding to it. What we will not be doing is teaching you something completely new.

## What we are going to cover

The rest of Part 1 is going to lay some foundations before we get to the serious business of guiding you through how to plan and conduct research projects with people. We will start out by explaining the logic of how research works in Chapter 1, then Chapter 2 will look at the question of who will take part in your studies. Chapters 3 and 4 are an introduction to what you might do with the data you collect when carrying out research: Chapter 3 looks at how you can describe data and Chapter 4 introduces the tests that are used to see whether what you see with the people you test can be generalized to other people.



# chapter 1

## the problem with people: variation and hypothesis testing

From time to time, everybody who carries out research with people for a living wishes they worked with something a little simpler and more predictable. Sleep-deprived exploding panthers, for example. The problem is that, whilst people can be fascinating and amazing things, they are also maddeningly inconsistent. A physicist studying pieces of copper doesn't come in to work one day to find their blocks of metal are suddenly behaving wildly because it is windy outside (this happens all the time when studying small children). A biologist working with sheep doesn't have to worry about them having hangovers if they are tested early on Monday morning (as happens with some students). An engineer hanging weights on a model bridge to see how it flexes doesn't need to make allowances for the bridge getting better at its task with practice (as you would).

The point is that when we are dealing with people, with their memories and social lives and moods and beliefs, we have to be aware that every observation we make is the end product of a million different influences. Go and ask somebody what is their favourite colour. Done that? Okay, now think about the answer you received. Of course, the person you spoke to might simply have told you their favourite

colour. But what if you then learnt they had the colour vision deficit known as protanopia, and so were not be able to tell red and green apart? This gives their answer a rather different complexion, doesn't it? And anyway, how can you be sure the answer you received is really that person's favourite colour, and not just a colour other people have told them they should like (as when five-year-old girls say 'pink')? Perhaps the person's favourite colour was orange but they told you 'black' because they wanted to appear moody and interesting. They might have said the colour of your shirt, because they thought your question was stupid and just wanted to get rid of you. If they know you, they might have said *your* favourite colour because they wanted to reinforce your friendship. They might have said 'mauve', simply because saying the word 'mauve' makes you smile (try it). Or they might have been tediously predictable and said 'blue'.

The point is, all you have from that person is the name of a colour. Is it their favourite colour, as you requested? It might be, but there are so many possible influences on their response that you just can't know that with any certainty. This means you can't reach safe conclusions.

*Ian:* I've just spoken to Anna and I now know her favourite colour is sea-green.

*Nigel:* Weird, because *I've* just spoken to Anna and I very much think you'll find her favourite colour is lilac.

*Ian:* Sea-green.

*Nigel:* Lilac!

*Ian:* SEA-GREEN!

And so on, until the fisticuffs begin. When this sort of thing happens, how can we ever find the correct answer?

One approach might be to anticipate all the various influences that could affect a person's behaviour and deal with them in advance. So you could try to discover somebody's real favourite colour by asking them a question like: 'Disregarding your recent experiences, and what you think I might be expecting you to say, and what your mother told you was her favourite colour, and what you feel Society thinks you ought to say, and any colour you might be looking at right now (unless it actually is your favourite colour, in which case don't disregard it), and the colour of your bedroom walls, and the colours of your toys when you were a child, and the magnificent mouthfeel of the word 'mauve'<sup>1</sup> ... disregarding all that, what is your favourite colour?' You could try that, but you'd still fail to get a reliable answer

---

<sup>1</sup> The colour mauve has an amazing history, and is pretty much responsible for the modern world as we know it. Seriously. We really recommend reading Garfield, S. (2000). *Mauve: How One Man Invented a Colour that Changed the World*. Faber & Faber.

because you didn't realize that the person you are speaking to is a member of the Himba tribe from Namibia and so uses the same word to mean both 'dark red' and 'dark blue'.<sup>2</sup>

This is all by way of saying that tremendous amounts of variation and complexity are the norm when studying people. People differ from one another, and individuals differ over time, and so you cannot just measure things simply as you would with blocks of copper or model bridges. Nor can you hope to deal with this issue by anticipating all the factors which might affect people's responses, as we just attempted to, because, no matter how much you try to anticipate the influences on people's behaviour, people will always find a way to surprise you. Luckily, you do not need to do this, as procedures and techniques have evolved over the years to cope with people and their tricky ways. These procedures and techniques are what this book is all about.

Most of the techniques we will cover involve testing groups of people rather than individuals. This helps you look past any individual quirks to reveal information about people in general (remember how you were happier with a drug tested on 10,000 people rather than two?). These techniques also tend to rely on using statistical analyses at the end of your data collection, to assess how well your choice of groups has worked. This last bit scares many people, but we promise we will make it very easy for you by making sure you understand what the statistics are for, rather than the intimate details of how they are calculated.

What we hope most of all is that by showing you the difficulties inherent in studying people, and how these can be overcome through good planning and design, human research can be a thing of beauty. And we mean that seriously. You will know when you have really grasped the issues involved in human testing by your emotional responses: reading about an ingeniously designed study which conclusively answers a difficult question will give you a rush of excitement, like hearing a great piece of music; seeing a badly designed study will make your nose wrinkle in disgust. The first time you find yourself squirming as you hear somebody describe a study in which they neglected to use a control group, take comfort from the fact you have at last developed the instincts of a great human researcher.

## Good. Can I start doing research now?

Not just yet. Before you can do any research you need to make sure you know what question you are trying to answer, and this can take

---

<sup>2</sup> For more on the Himba and their descriptions of colour, see Adelson, R. (2005). Hues and Views. *APA Monitor on Psychology*, 36, 26.

more than one form. First, there is always a basic, straightforward question you are trying to answer, known as your **research question**. ‘Do radar operators work better when they drink coffee?’ and ‘Do environmental issues affect people’s decisions about how they travel?’ are both examples of research questions.

It is very simple to decide what your research question is. Let’s assume for a moment you come from a business management background and are interested in whether people who work for small companies feel more loyalty to their employers than people who work for large companies. In this case, your research question would simply be something like ‘Do people who work for small companies feel more loyalty to their employers than people who work for large companies?’ Alternatively, if you come from a medical background and are interested in whether people with gout feel better after taking Holtodol, your research question would be something like ‘Do people with gout feel better after taking Holtodol?’

The research question is not a particularly difficult concept to grasp, we hope you will agree. It is simply a direct question you would like to have answered, and it often takes the form of a yes/no question (although it can also take other forms, such as ‘What proportion of Holtodol users are aware they are treating their gout with a fictitious drug?’).

It is a very good idea always to be clear about what your research question is, as this can really help keep your work organized. We have supervised quite a lot of students over the years who have come to us feeling overwhelmed by their research projects – a very common experience when people first have to design and conduct a substantial study on their own, especially in the Dark Ages before this book existed. In particular, we have heard many students say things like ‘I’ve got so much material and I don’t know what is relevant and what is not!’ In almost every case, their problem was that, although they generally knew what they wanted to study, they did not have a clear research question. The solution was for them to take a step backwards (not literally – our offices aren’t big enough for people to walk around in) and to write down, in a single sentence, the question they wanted to answer with their study. Once the research question was totally clear, and preferably written down somewhere prominent where it could be consulted regularly, problems of organization disappeared. Is the paragraph you’ve just written necessary for your report? Look at the research question – it’ll probably tell you.

But, although a clear research question is terribly useful for organizing yourself, it is sadly not enough on its own for you to conduct research. For reasons we will explain in a moment, you cannot actually test your research question, no matter how hard you try. Instead, you must use the research question to generate something else, something which *will* let you do some testing: a **hypothesis**.

**pedant point** The process of carrying out research can go by various names: you might refer to conducting a 'study', an 'analysis', an 'experiment', a 'project', an 'investigation' and so on. Most of these terms are interchangeable: there is no difference between saying 'I am carrying out an investigation of reindeer antlers' and 'I am carrying out a study on reindeer antlers.'

The one term that is reserved, and not synonymous with the rest, is 'experiment'.

This is because an 'experiment' is a very specific type

of study/investigation/project in which you make a change to some aspect of the world, whilst holding everything else constant, and look at the effects of that change.

It would be an experiment if you took a large number of people who had never ridden a unicycle, gave half of them unicycle lessons and half of them no unicycle lessons, and then looked at whether the group which had the lessons were better unicyclists. Subject to a few assumptions which we will explore later, this is a genuine experiment as the *only* thing that varied between the two groups of people is that one group had lessons and the other group did not – everything else was constant. If you found there was a difference

## EXPERIMENTS, STUDIES AND INVESTIGATIONS

in unicycling ability between the two groups of people at the end of the procedure, you would be able to know with some certainty that it was caused by the lessons, as in every other sense the two groups were the same. This is an experiment, and it lets you make strong conclusions about cause and effect.

If, on the other hand, you asked some people how they felt about unicycling, this would be a *study* of unicycling, but it would not be an experiment. It would not be an experiment because you did not manipulate, or change, anything. It still would not be an experiment if you compared two groups

of people – say a group of circus performers with a group of police officers – on their unicycling abilities. This is still not a true experiment because *you* have not caused the difference between the groups; it is simply a naturally occurring difference.<sup>3</sup>

It is quite common for people to mix up their terms and say 'experiment' when really they mean 'study', and to be frank we can't rule out the possibility that we'll do it before this book is finished. But we always try to reserve the word 'experiment' for situations where there is a true experimental manipulation and to use the word 'study' when we are talking in general terms. It is good practice to get used to doing this too.

## Finding robust findings

Research with people generally takes one of two forms. First, research is sometimes purely exploratory. If you are the first person to investigate a new area, you generally just dive in and, using little more than educated guesses to guide you, start collecting data.

Imagine you were to discover tomorrow that trees can talk. This would be a totally new discovery. Your first action wouldn't be to go indoors and construct a careful research programme before doing anything else. No, the first thing you would do is ask a tree some questions! You would try to build some sort of understanding about

<sup>3</sup> To be pedantic about our pedantry, this sort of study can be called a *naturalistic experiment*.

your new discovery as quickly as possible in order that you could lay the foundations for later, more systematic research. This is exploratory research.

But most research is not like this. Instead, most research studies are in established areas. Some of this research tries to find answers to specific research questions like ‘is poverty linked to crime?’ and ‘does eating saturated fat increase the risk of coronary disease?’ Other research goes further to look at the mechanisms behind phenomena. For example, once you’ve found there is a relationship between poverty and crime, you might ask the question ‘exactly *how* does poverty influence crime?’ For now, we will focus on situations where you want to answer a particular research question rather than explore a brand new phenomenon.

When you carry out research to answer a specific research question, you really want to produce findings that are **robust**. This simply means that when you reach a conclusion at the end of your work, it is trustworthy. With a robust conclusion, nobody can come along afterwards and show that your conclusion is unreliable.

A lack of robustness was the problem we encountered when asking people about their favourite colours. After you asked someone their favourite colour, your conclusion was not robust because it was entirely possible for somebody to come along and contradict you, or to show that your method was faulty. If this happened, and somebody pointed out that your conclusion was unreliable, what response would you have other than insults and name-calling? That’s the problem with conclusions that are not robust – they lead to undignified science.

## Seeing into the future

So the conclusions you reach when you do research should be robust. But this alone is not enough, because you also want your conclusions to be *general statements about the world*. In other words, you want your conclusions to apply to new situations or people, not just the situation or the people you studied. For example, let’s say you carry out a study to see whether a certain food additive affects children’s behaviour. At the end of this study you will almost certainly want to be able to say something about how this food additive affects children in general, probably including children who have not yet been born. It would be pretty pointless if you could only make statements about the small specific group of kids that you studied.

What we are suggesting, then, is that *doing research is an attempt to predict the future*. We research things so that we can understand them, and when we understand something we can make statements

about what is likely to happen in the future and why it is likely to happen. Consider this: what would happen if you set fire to this book right now?<sup>4</sup> This is something you have never done before, but you can almost certainly predict what would happen if you did it. The reason you can predict what would happen is that you have a good understanding of the various factors involved – fire, paper, the place you are reading, and so on. If you knew nothing about fire and how it behaves, you would not be able to make this prediction about something that has not yet happened.

This idea that research is an attempt to predict the future can be seen in all fields. If we are testing a new drug to see if it is safe, we are doing this because we want to know what will happen when people use it after our study is over – this is predicting the future. If we are looking at whether a new management technique improves workers' productivity, we want to know if it will work for people and in organizations we have not specifically tested it on – this is predicting the future.

Of course, humans have been trying to predict the future for thousands of years. In ancient Rome there was a type of priest known as an Augur, whose job it was to look at how birds were flying through the air and to use this information to decide whether the emperor's plans would be successful or not. (The 'birds' usually said the inbred sword-wielding psychopath was a genius and his plans couldn't possibly go wrong.)

Now you probably don't need us to tell you that if your attempts to predict the future are based around watching blackbirds, you will struggle to produce robust conclusions about what will happen tomorrow. But in a more subtle way, researchers had the same difficulty reaching reliable conclusions about people right up into the twentieth century. They would identify a research question and make observations to test it, only to find their conclusions were questioned later and there was nothing they could say to prove their findings were valid.

For example, a researcher might study a group of households and say 'I have seen evidence of greater intelligence in children who eat lots of fish.' But if another researcher came along afterwards and said 'I too looked at children who ate lots of fish and found no evidence of their being more intelligent,' the two researchers had no good way to establish who was correct. Each had a different idea about what would happen in the future to children who ate a lot of fish. As often as not the winner of the debate would be the one with the most forceful personality or, frequently, the biggest beard.

In the 1920s, the statistician Ronald Fisher, himself no stranger to facial hair, looked for a way to make research more robust, to deal with this problem. In doing so he gave us a new method to guide our

---

<sup>4</sup> Do not set fire to this book!

studies. His solution has come to be called **null hypothesis testing**, and it can be a slightly tricky concept to understand the first time you encounter it. However, the logic behind null hypothesis testing has transformed research, giving scientists a method for doing studies which produce robust findings unlikely to be questioned. As such, an understanding of how it works is important, so let us explain it by guiding you through a quick study. And, not lacking ambition, we are going to do our study on quite a big question: does gravity work?

## Null hypothesis testing: gravity on trial

Pick up an object – nothing too fragile or valuable, please! – hold it up in the air and then let it go.

Did it fall down?

If not, kindly close this book and resume the demonstration when you have returned from orbit. If the object did fall down, it looks as though gravity is working. Excellent.

But the thing is, as researchers it is not enough for us to say ‘it looks as though...’ After all, from here it *looks as though* the sun rotates around the earth and it *looks as though* cows prevent fields from floating away by standing on them. Instead, we want to be *certain* about things.<sup>5</sup> So looking at the world and saying ‘it looks as though...’ is an easy way to make mistakes, and, in research, mistakes matter. Those of you reading this book towards a qualification will get bad grades if you carry out poor research; those of you who will do research for a living could damage your professional reputations. And those of you in biomedical fields could easily one day carry out research where a mistake would kill people. So let’s get away from ‘it looks as though...’ and let’s try to be more certain: pick up your object and drop it again.

Assuming it fell down once more, we now have two falls in a row and are getting a picture of how gravity works: objects that are not supported fall until they rest on something. But here is the key question: how many times would you have to drop your object and see it fall before you were *certain* this is how gravity works? Please decide on an answer to this question before reading further.

Unless you decided ‘infinity’ or something similar, you were wrong. Let us show you why.

---

<sup>5</sup> Or rather, as we will see later, we want to know *exactly how uncertain* we are about them.

If you pick any number – a billion, say – and drop the object that number of times, you will still not be *certain* about what gravity will do in the future. After dropping the object a billion times, you still cannot *know*, with 100% certainty, what will happen the next time. If we were to tell you that gravity makes objects fall down almost every time you drop them, but that just occasionally, instead of falling down, objects turn into winged badgers and fly away, well... you cannot *know* we are wrong.

You could, of course, test our suggestion by dropping the object once more, but the problem is still there: you have now shown the object falls the first billion-and-one times it is dropped, but you still cannot be *certain* what will happen the one-billion-and-second time it is dropped, because you have not yet seen it happen. Just as ‘tomorrow’ never arrives, ‘the next time it is dropped’ will never arrive for your object. Every time you test your object’s behaviour again, a new, unseen ‘next time’ immediately comes into existence and you can do nothing to prove we are wrong if we continue to insist that at some point your object will turn into a badger. You do not have the ability to predict the future with certainty, as you would like.

## Enough badgers – could we possibly have an example involving some pens in a box now, please?

But of course. We have a box with two pens inside – the two pens are identical in size, shape and weight, but one of them is blue and the other is red. We hope you can imagine somebody reaching into this box three times without looking and picking the red pen on each occasion (obviously they put the pen back after each attempt). Picking the red pen three times in a row wouldn’t be a particularly unusual event – it would happen, typically, to one out of every eight people who reached into the box three times.

If you can imagine somebody reaching into the box three times and picking the red pen on each occasion, can you also imagine somebody reaching into the box four times and getting the red pen every time? If that is harder to picture, think of it this way: it is exactly the same event you imagined a moment ago, followed by the person reaching into the box one more time and getting the red pen again. Getting the red pen five times in a row is exactly the same as that, just with one more red pen at the end. And so on.

We realize we are now getting dangerously close to the sort of tedious statistical thought-experiment we didn’t want in our lovely

user-friendly book,<sup>6</sup> so let's get to the point: although we haven't met them all, we are reliably informed that there are now several billion people in the world. If a billion people each reached into our box 20 times, we would expect nearly 1,000 of them to select the red pen 20 times in a row. Selecting the red pen 20 times in a row, when there is also a blue pen inside the box, is something that *could* happen. That is what we are saying.

Now imagine we offered the box to you. We tell you, 'This box has a red pen and a blue pen inside.' How would you check our claim? You reach into the box and feel that there are indeed two pens inside. But, because you know we are nasty academics with a strange sense of humour, you immediately worry that we are tricking you and that both the pens are the same colour! You need some way of testing our claim that there are both a red pen and a blue pen in the box. You notice we have deviously designed the box so that you cannot see inside and so the only option open to you is to pull a pen out. You do so: it is, unsurprisingly, red.

Having seen a red pen come out of the box, you replace it and carry on testing our claim, pulling a pen out of the box over and over, until you have done it 20 times. The pen is red on every occasion. Now, we claimed there were a red pen and a blue pen in the box. What could you say *for certain* about our claim at the end of your experiment? You would agree with the first part of our claim: there is definitely a red pen in there – you have a frankly tedious amount of evidence for this. But what about the second part of our claim, that there is also a blue pen? Can you say it is true? No, because you haven't seen it. Can you say it is false? Again no, because it might be in there and you have just been unlucky in not picking it on any of your attempts – you might happen to be one of those 1,000 people out of every billion who get the red pen 20 times in a row. The fact is, you can say *nothing* definite about whether or not the blue pen exists at this stage. Sure, you can say you *believe* there is no blue pen in the box, but you have no way of knowing *for certain* that it will not appear the next time you reach into the box.

This principle is really important when we do research with people. Let's return to the idea of testing a new drug to decide if it is safe. In fact, let's say you are testing the specific research question 'Does Holtodol cause side-effects in users?' You find a willing volunteer, give them a big dose of Holtodol and find they are fine afterwards. Do you therefore conclude 'No, Holtodol does not cause side-effects'?

We suspect that after testing just one person you would not be happy coming to this conclusion and would want to see more data. Good for you!

---

<sup>6</sup> Consider yourself lucky we're not yet talking about tossing coins.

So let's say you give the drug to 1,000 people – or even 10,000 – and still find no side-effects. Can you put your hand on your heart and say the answer is 'No, there are no side effects, Holtodol is 100% safe, it never produces side-effects and I would be happy for my mother to take it right now'?

We hope by now you would not be able to say this without some feelings of doubt, even if they are only small ones. We hope you see that, no matter how many people you test, there is always a possibility, however slight, that Holtodol is dangerous to some people and that you have, through a quirk of chance, met only people who are unaffected by it. The next person might be different. And just as tomorrow never arrives, nor does the 'next person'.

But this isn't satisfactory at all: it looks as though we can test an idea over and over and over but still never know anything for certain! And this is just what Fisher realized: as long as we go on testing the idea that we will see something happen, but not seeing it, there is always a 'next time' or a 'next person' which could change everything.

## Null hypothesis testing

Fisher's ingenious solution to this dilemma was subtly to change the questions we ask when we do research. Realizing that we can test ideas over and over without ever reaching a definite conclusion, he suggested we test the opposite idea instead.

**Original idea:** on earth, gravity makes unsupported objects fall

**New idea:** on earth, gravity *does not* make unsupported objects fall

So instead of predicting that something happens, we predict that something *does not* happen. This probably sounds fairly trivial, but it is really a big difference. Ask yourself this: how many times would you have to drop your object to test this new idea that gravity *does not* make things fall? Again, decide on an answer before reading further.

The answer is almost certainly 'one'. Go on, try it: drop the object again. If it fell down then you have conclusively, absolutely, indubitably and forever demolished the idea that objects do not fall when you let them go, and you have done this with only a single test. When we test the idea that an event will *not* happen, research becomes much easier. If you give an event a fair opportunity to happen and it does not, then your claim that it doesn't happen is probably true; if the event does happen, even once, you know for certain that your claim is false. Either way it's good.

And this is the basis of all research. We start with an idea we are interested in, which as we have seen is called the research question.

**Step 1. Research question:** Do dogs eat bananas?

Based on this, we derive an **experimental hypothesis**, which is a specific prediction that something will happen.

**Step 2. Experimental hypothesis:** We will see a dog eat a banana.

But this is not entirely suitable for research. If we see a dog eat a banana than we have an answer to our research question. But if we go our whole lives watching dogs and never see one eat a banana, we learn nothing for certain. For all we know, dogs eat bananas a lot, but only when nobody is watching. Or perhaps they eat bananas only in the dark, or perhaps it is only dogs in Chad who eat them. Or perhaps they do eat bananas, but only a different variety of banana than those we are using in our study. We therefore don't bother with this idea and come up with something else to test, as Fisher suggested. We test the opposite of the experimental hypothesis, the idea that nothing is happening...

**Step 3. Null hypothesis:** We will not see a dog eat a banana.

You only need to see one dog eat a banana to know for sure that this null hypothesis is false. This then gives you an answer to your research question (yes, dogs eat bananas). If, on the other hand, despite your giving dogs ample opportunity to eat a banana in front of you, this does not happen, your null hypothesis is probably correct and you again have an answer to your research question (no, dogs seem not to eat bananas).

**pedant point**

During our last bit of pedantry we said the word 'experiment' had a very specific meaning and could only be used to describe studies in which we manipulate one aspect of the world and look at the effects of doing so, as a way of establishing cause and effect.

It is therefore a little confusing that we use the term 'experimental hypothesis' to describe

any specific prediction, even if the study we are conducting isn't a true experiment. This is a fair point, and in our defence we have to say that these are not our terms. If it helps, you might want to use the more technical terms, which are ' $H_1$ ' for the experimental

**EXPERIMENTAL HYPOTHESES**

hypothesis and ' $H_0$ ' for the null hypothesis. Or you might favour the term 'alternative hypothesis', which you will find used in some sources. The problem with the term 'alternative hypothesis' is that, although it makes a lot of sense if you are approaching hypothesis testing from a statistician's perspective, it doesn't

make much sense if you are learning about hypothesis testing from the applied approach we are using here: as we have shown, this hypothesis is always developed *before* the null hypothesis, so it seems odd to

call it the alternative.

Oh, and whilst we're here being pedantic, can we point out that it is 'one hypothesis' and 'two hypotheses'? You already knew that? Excellent.

## pedant point

Here in Britain where, being British, we live, people use American English spellings when discussing computers and British English spellings for everything else. So when we Brits come across a word like *disk*, we feel a warm glow of smugness from knowing it *must* refer to part of a computer as it would be spelt *disc* otherwise. We can only imagine the hilarity that regularly ensues in Delaware when people ask for computer peripherals and get given brake parts. Or something.

Unfortunately, we don't find this sort of helpful clarity with the word 'billion'. Just to clear this up, there

are two meanings of the word 'billion'. The US version is 1,000,000,000, or a thousand millions, and the British version is 1,000,000,000,000, or a million millions. Sadly the million-millions sense of the word is rapidly disappearing, and so we have reluctantly used 'billion' in the US sense in this book.

Our reluctance stems not from a dislike of US English, as any nation capable of inventing the phrase 'no-brainer' is all right with us. Rather, we feel the international shift to the US meaning of 'billion' is unfortunate because we had a perfectly good word already for one thousand millions, which was 'milliard'. This is a truly magnificent word and we are sorry to see it die.

## BILLIONS

# Null hypothesis testing made as non-threatening as possible through the use of pasta sauces

(If you feel you have already understood null hypothesis testing, feel free to snort derisively right now and jump straight to the final section below, entitled 'A last point'.)

At this stage you would be forgiven for thinking this has all been really quite complicated – and it's only chapter 1! Don't worry: null hypotheses are the most difficult thing in this whole book and we definitely do not expect you to have grasped every aspect of this in a single reading – we are expecting that you will come back and read the material above a few more times as you work through the rest of the book. And it is only when you start to design studies of your own that it will start to become second nature to you.

Having said this, before we move on we would like to try and cement the concept in your mind just a little bit more. To do this, we will talk you through the sorts of thought processes we would use when we select the hypothesis for a study. This should illustrate what goes on with research questions and hypothesis testing when planning research for real.

*Nigel:* I'm cooking pasta tonight.

*Ian:* Nice.

*Nigel:* Yes, so I'm off into town to get some tomatoes for the sauce.

*Ian:* Why would you do that? Tomatoes are vile!

- Nigel:* You philistine! Tomatoes rock. Tomatoes rock the hardest.
- Ian:* Please tell me you're joking.
- Nigel:* You're just weird about vegetables.
- Ian:* Fruits.
- Nigel:* You know what I mean.
- Ian:* There's no way you should be making a tomato sauce for your pasta when you could make a cheese sauce instead. Cheese sauce is far better. Why else is cheese sauce so much more popular? That is a fact, my friend: a cold, hard fact.
- Nigel:* I don't –
- Ian:* Faaaaaaaaact.
- Nigel:* That is so *not* a fact! You just made it up. If anything, there's no difference. It just happens I'm having tomato sauce tonight, but they're both equally good sauces and they're both equally popular.<sup>7</sup>
- Ian:* Oh yes? Well let's settle this – with science!
- Nigel:* Right. So what's our research question? I suppose it's going to be something like 'Is one sauce more popular than the other?'
- Ian:* Yes, which means the hypothesis we'll test will be something like 'One sauce will be chosen more often than the other'.
- Nigel:* Right, let's get to work.
- Ian:* No, hold on: we're not there yet. We can't test that hypothesis. Think about it: suppose we speak to forty people and the majority prefer tomato sauce, that doesn't answer our research question. For all we know, if we tested forty more people the majority of those would prefer cheese and cancel out our original finding.
- Nigel:* Fine, so we won't test forty people – we'll test a hundred if it makes you happy. That way we've got both lots of people you've just described and a few more as well.
- Ian:* No, that's still no good. Even if we test one hundred people, that still won't tell us whether the sauces are equally popular or not. For all we know the next hundred might be totally different to that hundred, and we've no way of knowing that without testing them as well.
- Nigel:* So we do that. In fact, let's test a million people!
- Ian:* Still no good. Even if we ignore the fact it's practically impossible to test that many people, it still doesn't let us make a decision about our hypothesis. If we test a million people, we still can't be *sure* one sauce is more popular than the other. The next million could be different. The only way we could ever know for sure would be to test every person on the planet all at the same time. This is getting us nowhere.
- Nigel:* So what do we do?
- Ian:* We need to stop trying to test that experimental hypothesis. What we need to do is use it to create a null hypothesis. Instead of testing the idea 'One

---

<sup>7</sup> You're probably thinking this is all invented for the sake of your education, but tragically we have this conversation about once a month.

sauce will be chosen more often than the other' we need to test the idea that there's nothing happening: we test the idea there is no preference. We use a null hypothesis like 'There will be no difference in how often cheese sauce and tomato sauce are chosen'.

*Nigel:* Yes, that'll work. We test a reasonable number of people, which will give any preference a chance to show itself. If we still don't see a preference then we'll accept our null hypothesis that there isn't one. But if we do see a preference we'll reject our null hypothesis and accept our other hypothesis, which was 'One sauce will be chosen more than the other'.

*Ian:* Either way, we can settle on one of the hypotheses as a conclusion.

*Nigel:* Hoorah!

You might remember that at the top of this section we told people who were already happy with the concept of null hypothesis testing to skip on to the next section. So whilst those insufferable know-it-alls aren't looking, we would like to share a snippet of information with you more assiduous readers, which is this: because the null hypothesis is always simply suggesting *you will not find an effect*, it is very common for people not to bother even stating it when they write about their research. So if you see somebody say 'We tested the hypothesis that soldiers run faster than civilians,' what they mean is they tested the null hypothesis that there was no difference in running speed between soldiers and civilians. If you see somebody say 'My hypothesis was that the longer people spent studying, the better their exam score would be,' what they are telling you is that they tested the null hypothesis that there was no relationship between study time and exam score. Even if somebody doesn't mention it, they *did* test a null hypothesis. It's just that sometimes you have to work out what it was for yourself.

## A last point

Null hypothesis testing is a really important aspect of how we carry out research. It provides a method which most researchers agree allows us to obtain robust findings. We've guided you through some examples to show where this procedure came from and how it is used, but, as often happens in life, this is a concept which you can grasp in more than one way, and different people prefer different explanations. So to finish we would like briefly to sum up null hypothesis testing in a slightly different manner. If our explanation so far hasn't worked for you, perhaps this will make it clear.

Let's consider a question for which you have no evidence one way or the other. We are working together in a room right now, and we would like to ask you this: 'Is there a cat in this room with us?' What would you say, given that you cannot see us? There are two

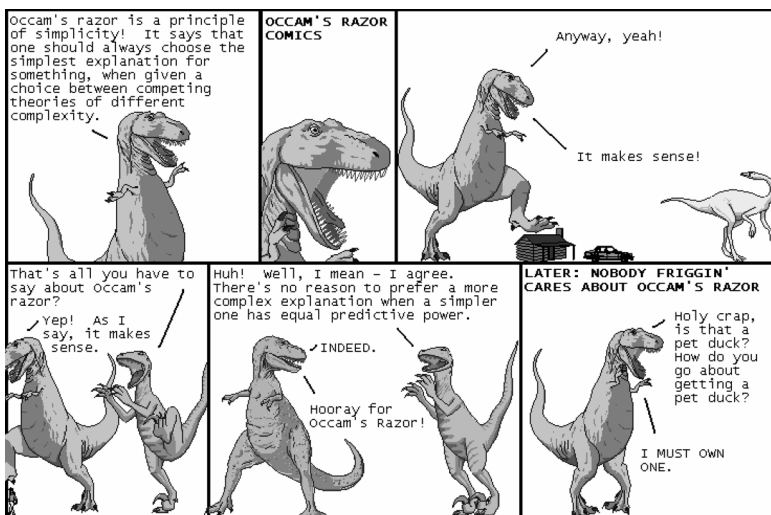
possibilities here: either there is a cat or there is not. Without any evidence one way or the other, which would you choose?

Many people in this situation, having to choose one situation over the other with no evidence at all, would choose the simplest option. Keeping things simple is generally a really good idea – something we'll say a lot more about soon. In this case, it is much simpler to assume there is no cat. A room, when it is built, does not have a cat in it. It is not the natural state of a room to contain a cat and extra effort has to be made to get a cat into a room, so it is clearly simpler to assume there is no cat. (You may have heard this described as 'Occam's Razor'. This is the idea that, when you have to choose between options and have no reason to favour any one option, you should always choose the simplest option. The principle gets its name from a mediaeval English monk called William of Occam who first described it.)

In research, the null hypothesis always states that nothing is going on – there is *no relationship* between two measures; giving people a new drug *won't* affect them; there is *no cat* in the room. The null hypothesis, then, always represents the simplest explanation. We need a good reason to abandon it in favour of a more complex explanation.

As we will see in later chapters, when we study a research question we collect evidence. It is only when that evidence clearly points towards a more complex explanation that we abandon our nice simple null hypothesis. It is for this reason that we focus on the null hypothesis, and describe research as testing it – deciding whether to keep it or reject it. See Box 1.1 for a little more on the types of hypotheses you might choose.

(There was no cat in the room, by the way.)



### Box 1.1 One- and two-tailed hypotheses

You will almost certainly encounter the terms ‘one-tailed hypothesis’ and ‘two-tailed hypothesis’. These are slightly odd names, and refer to the shape of certain graphs which we won’t go into here. However, their meaning is very simple.

*A two-tailed hypothesis predicts that there will be an effect in your data*

*A one-tailed hypothesis predicts that there will be an effect in your data, and says what that effect will be*

So, for example, consider these two hypotheses:

1. Women will drive better than men
2. There will be a difference in how men and women drive

The first hypothesis is more specific: it predicts an effect and says specifically what that effect will be; the second hypothesis simply predicts an effect. We call the first a one-tailed hypothesis and we call the second a two-tailed hypothesis. It is called ‘two-tailed’ because the effect can go either way – women could be better drivers than men or men could be better drivers than women.

## Test yourself

Here are some experimental hypotheses. For each, decide if it is one- or two-tailed. Remember: a two-tailed hypothesis predicts there will be an effect and a one-tailed hypothesis predicts exactly what the effect will be. The answers are just below.

1. People from Chile will be less patriotic than people from Argentina.
2. Eating sugar will lead to changes in people’s moods.
3. People who swim regularly will have more supple limbs than people who run regularly.
4. Boys and girls attending school have different numbers of friends.

Now here are some research questions. For each, what would be the experimental hypothesis and what would be the null hypothesis if you were carrying out a study? Also, for each experimental hypothesis you produce, is it one- or two-tailed?

1. Do left-handed people write at a different speed from right-handed people?
2. Does eating cheese in the evening give people nightmares?
3. Does learning to ride a motorcycle make people more alert to road dangers?

## Answers

1. One-tailed: we predicted a difference and said exactly what it would be.
2. Two-tailed: we predicted a difference but didn't say which direction it would go in – sugar could improve or impair mood.
3. One-tailed.
4. Two-tailed.

## Research questions

For each of these, your wording might be slightly different to ours, but make sure you have said the same general thing we have. In particular, it is quite possible that you have generated one-tailed hypotheses where we used two-tailed, and vice-versa.

1. Experimental hypothesis: There will be a difference in how fast left- and right-handed people write. Null hypothesis, which will be tested: There will be no difference in how fast left- and right-handed people write. Here we've produced a two-tailed experimental hypothesis – we've predicted a difference but not said which way round it will be.
2. Experimental hypothesis: People who eat cheese in the evening will experience more nightmares than people who do not eat cheese in the evening. Null hypothesis, which will be tested: There will be no difference in how many nightmares will be experienced by people who eat cheese in the evening and people who do not eat cheese in the evening. Here we've produced a one-tailed hypothesis – we've not only predicted a difference but we've also predicted exactly which group will experience more nightmares.
3. There are actually two approaches to this one, depending on whether you design a study comparing motorcyclists with non-motorcyclists, or whether you design a study which teaches people to ride a motorcycle. The fact you can study the same question in more than one way is a theme we'll say a lot more about.
  - a. Experimental hypothesis: Motorcyclists will perform better in a road danger test than equivalent non-motorcyclists. Null hypothesis: There will be no difference between motorcyclists and non-motorcyclists in the road danger test. Here we've used a one-tailed hypothesis.
  - b. Experimental hypothesis: Learning to ride a motorcycle will increase people's scores on a road danger test. Null hypothesis: There will be no change in people's scores on a road danger test when they learn to ride a motorcycle. Here we've used a one-tailed hypothesis – the two-tailed version would predict a change in test scores instead of an increase.

# Index

- 15/40 rule, 85  
 40/15 rule, *see* 15/40 rule
- Abelson, R.P., 194, 198  
 accidental sampling, *see* samples, opportunity  
 alpha, 92–3, 178, 231  
 alternative hypothesis, 16, 231  
   *see also* experimental hypothesis  
 alternative medicine, 129  
 alternatives, 141  
 analysis of variance, 82, 112, 205, 231  
 arithmetic mean, 55  
   *see also* mean  
 assumptions, underpinning  
   statistical tests, 75, 78, 81, 231  
 attitudes, 181–91  
 Augurs, 11  
 averages, *see* central tendency, measures of
- bar charts, 68–70, 71, 110, 177, 205  
 Beaman, C.P., 170, 180  
 bell curve, *see* normal distributions  
 beta, 232  
 biases, *see* result bias; samples, bias in  
 Bigler, E.D., 216  
 billions, 17  
 bimodal data, 59, 232  
 binomial tests, 146–7, 227, 232  
 biology, 26  
 Black, J.B., 194, 198  
 blind testing, 138–9, 232–3  
 Bonferroni corrections, 233  
 box-and-whisker plots, 70  
 Bransford, J.D., 162, 168  
 business management, ix, 8, 115–35, 200
- Cajochen, C., 63  
 case studies, 47, 233  
 categorical data, 47–8, 49–50, 53, 59, 68–9, 165, 233  
 causal relationships, 215, 233  
   *see also* experimental method  
 ceiling effects, 233  
 census, 24  
 central tendency, measures of, 54, 55–60, 233  
   *see also* mean; median; mode  
 chi-square tests, 149, 233  
 cluster sampling, *see* samples  
 cohorts, 233–4  
 communication studies, *see* media and communication studies  
 conclusions, 10  
 conditions, *see* independent variable  
 confidence intervals, 234  
 confidentiality, 134  
 confounding variables, 102, 124, 138–40, 172, 206, 234  
   *see also* experimenter effects; extraneous variables  
 consent, *see* informed consent  
 contingency tables, 165  
 continuous data, 47–8, 53, 68, 234  
 control, *see* control conditions; experimental control  
 control conditions, 120, 129  
 convenience sampling, *see* samples, opportunity  
 correlation, 50, 197–8, 212–15, 219–20, 235  
   *see also* Pearson's *r*; Spearman's rho  
 counterbalancing, 104–6, 122, 142–3, 151, 155, 173, 235  
 covariance, 235  
 criterion variables, 235–6
- cube root transformations, *see* power transformations  
 curvilinear relationships, 236
- data, 236  
 data transformations, *see* transformations for skewed data  
 debriefings, 134, 236  
 deception, in research, 134  
 Delaware, 17  
 demographics, 39–41, 189  
 dependent variables, 101–2, 119, 141, 153, 171, 203, 236  
 descriptive statistics, *see* statistics, descriptive  
 design, *see* study design  
 dichotomous data, 48, 236  
 discourse analysis, 47  
 discrete variables, *see* categorical data  
 dispersion, measures of, 54, 60–5, 145, 236  
   *see also* interpercentile range; interquartile range; range; standard deviation; variance  
 distributions, 236  
 double-blind procedures, *see* blind testing  
 dummy coding, 48  
 dyslexia, 23
- economics, ix, 37  
 education ix, 63, 185–91  
 effect sizes, 236–7  
 empiricism, 237  
 errors, Type I and Type II, 237  
 eta-squared, 237  
 ethical issues, 34, 42, 131–4, 142, 196, 237–8  
 ethnography, 238

- experimental control, 137–40, 147  
 experimental hypothesis, 16,  
     18, 238  
 experimental manipulation, *see*  
     experimental method  
 experimental method, 9, 16, 140,  
     147–8, 167, 238  
 experiment design, *see* study design  
 experimenter effects, 139, 211,  
     238–9  
 experiments, *see* experimental  
     method  
 exploratory research, 9–10  
 extraneous variables, 102–4, 138,  
     153, 172, 239  
     *see also* confounding variables  
 eyeballing data, 110
- familywise error, 239  
 fatigue, 104, 173, 195  
 Fidell, L., 37  
 Fisher, R., 11–12  
 floor effects, 239  
 Follett, M.P., 116–17
- Galambos, J.W., 194, 198  
 Galton, F., 209  
 Garfield, S., 6  
 Gauss, K., 74  
 Gaussian distributions, *see* normal  
     distributions  
 Gazzaniga, M.S., 216  
 generalizing, from samples to  
     populations, 10–11, 23–9,  
     32, 37, 88, 130, 146–7, 152,  
     201–3  
 geometric mean, 55  
 Germany, postcodes as nominal  
     data, 49  
 GIGO, 46  
 Goulandris, N., 23  
 grab sampling, *see* samples,  
     opportunity  
 graphs, 66–73, 110  
     *see also* bar charts; histograms;  
     pie charts; scatterplots  
 Graw, P., 63  
 Green, R.L., 216  
 Greer, B., 59
- H<sub>0</sub>, 16  
 H<sub>1</sub>, 16  
 Himba tribe of Namibia, 7  
 histograms, 67–8  
 Holtodol, 2, 8, 14–15, 27
- hypotheses, 8, 18, 19, 21, 97–8, 100,  
     108, 110, 119, 239, 243  
     *see also* experimental hypothesis;  
     null hypothesis testing
- illustrating data, *see* graphs  
 independent-samples designs,  
     103–4, 121, 123–6, 140,  
     154–6, 163, 172–3, 190, 201  
 independent variables, 101–2, 119,  
     138, 141, 153–4, 171, 174–5,  
     203, 234, 239  
 individual differences, 103–5,  
     173, 206  
 inferential research, 25, 88, 239  
 inferential statistics, *see* statistics,  
     inferential  
 informed consent, 132, 240  
 intelligence, 209–10  
 interactions, in analysis of  
     variance, 240  
 intercepts, in regression, 157–8  
 interpercentile range, 62–3  
 interpretative phenomenological  
     analysis, 47  
 interquartile range, 62–3, 70, 187–8  
 interval data, 51–3, 60, 240  
     interviewer bias, *see* experimenter  
     effects  
 IQ, 209, 240–1
- Johnson, M.K., 162, 168  
 Jones, D.M., 170, 180  
 judgement sampling, *see* samples
- Kräuchi, K., 63  
 Knoblauch, V., 63  
 Kruskal-Wallis test, 244
- levels, of independent variable, 101,  
     120, 241  
 levels of measurement, 48–53,  
     60, 241  
 Likert, R., 182  
 Likert scales, 52, 182–3, 241  
 linearity, in correlation, 214, 241  
 linear scales, 82  
 line graphs, 71–2  
 literature searches, 135  
 Loftus, W.C., 216  
 logarithmic transformations, 82  
 logistic regression, 165
- management, *see* business  
     management
- Mann-Whitney test, 190  
 matched-pairs sampling, 41–2,  
     103–4, 241  
     *see also* samples  
 mauve, 6  
 mean, 55–6, 109, 176, 241  
 measures of central tendency, *see*  
     central tendency, measures of  
 measures of dispersion, *see*  
     dispersion, measures of  
 media and communication  
     studies, 133  
 median, 56–9, 70, 187, 241  
 medicine, ix, 26, 33, 57, 91, 151  
 memory, 161–2, 166–8  
 mental chronometry, 151  
 meta analysis, 242  
 milliard, 17  
 mode, 59, 80–1, 187, 242  
 monotonic relationships, 242  
 Mulhern, G., 59  
 mundane realism, 242  
 mutual exclusivity, 146
- N (number of data points), 55  
 naturalistic experiments, 9  
 nominal data, 48–50, 53, 60, 242  
 normal distributions, 73–82, 242  
 North, R., xiii  
*n*th name sampling, *see* samples,  
     systematic  
 null hypothesis testing, 12–20, 87,  
     90–1, 93, 146, 178, 242–3
- Occam's Razor, 20  
 omnibus tests, 243  
 one-tailed hypothesis, *see*  
     hypotheses  
 operationalizing measures, 76,  
     119–20, 175, 179  
 opinion polls, 38  
 opportunity sampling, *see* samples,  
     opportunity  
 order effects, 104, 106, 123, 142,  
     155, 243  
 ordinal data, 50–1, 53, 60, 183, 243  
 outcome variables, *see* criterion  
     variables  
 outliers, 30, 39, 55–7, 59, 61–2, 85,  
     111, 130, 220
- p*, 88–96, 112–13, 147, 205, 213, 244  
 pain research, ethics of, 132  
 parametric statistics, 244  
 participants, *see* samples

- Pearson's  $r$ , 212  
 percentages, 144  
 percentiles, 62–3, 244  
 pharmacology, ix  
 pictographs, 66–7  
 pie charts, 70, 72, 144–5, 244  
 pilot studies, 100, 106, 108, 244  
 placebos, 139, 244–5  
 political polls, 38–9  
 political studies, ix  
 populations, 25–7, 28, 37, 38, 88, 92, 245  
   *see also* generalizing, from samples to populations  
 positive reinforcement, 117  
 power, statistical, 245  
   *see also* power analysis  
 power analysis, 31, 195, 245  
 power transformations, 83–4, 86  
 practice effects, 104, 123, 173  
 predictor variables, 246  
 probability, 89–91, 246  
 professional bodies, research governance, 133–4  
 proportions, 69–70, 72, 144–5  
 protanopia, 6  
 pseudo-random numbers, 35  
 psychology, ix, 26, 37  
  
 quota sampling, *see* samples, quota  
  
 random effects, 246  
 randomization, 106, 155, 247  
 randomized control trials, 35–6, 247  
 random sampling, *see* samples, random  
 range, 61–3, 70, 247  
   *see also* interpercentile range; interquartile range  
 rating scales, 51, 182  
   *see also* Likert scales; Thurstone scales  
 ratio data, 51–3, 60, 247  
 raw data, 54  
 reciprocal transformations, 84  
 regression, 157, 247  
 regression fallacy, 247–8  
 regression to the mean, 248  
 reliability, 240, 248, 251  
 repeated-measures designs, 104–5, 121–3, 126, 140, 154, 163, 173–4, 190, 248  
 replication, 248  
 reports, writing, 225–30  
 research questions, 8, 18, 140, 248  
  
 residuals, in regression, 85  
 result bias, 163  
 robust research, 9–12, 19, 147–8  
 Rutledge, J.N., 216  
  
 samples, 25, 27–43, 88, 142, 152, 188, 248–9  
   bias in, 33–4, 36, 43, 174, 249  
   cluster, 34, 36  
   desired characteristics, 27–30  
   difficult-to-reach, 42–3  
   judgement, 40–1  
   matched-pairs, 41–2  
   opportunity, 36–7, 142, 152, 188, 196, 210, 243  
   quota, 38–40, 246  
   random, 33–6, 210, 246  
   size of, 28–31, 85, 103, 188  
   snowball, 42–3, 249–50  
   stratified, 38–40, 210, 250  
   student, validity of, 37  
   systematic, 36  
   techniques for selecting, 33–43  
 sampling, *see* samples  
 scatterplots, 72–3, 156–7, 197, 212  
 Schultz, R., 216  
 scripts, mental, 193–4  
 sensitivity, 249  
 signal detection theory, 197–8  
 significance, 92–6, 119, 249  
 skewed distributions, 57, 79–86, 249  
 Skinner, B., 117  
 slopes, in regression, 157  
 social constructionism, 47  
 sociology, ix, 37, 133  
 snowball sampling, *see* samples, snowball  
 Snowling, M., 23  
 Spearman's rho, 212, 250  
 square root transformation, *see* power transformation  
 standard deviation, 63–5, 75, 111, 176, 250  
 standard error of the mean, 250  
 standardized instructions, 143, 156, 175, 203, 250  
 statistical power, *see* power, statistical  
 statistical significance, *see* significance  
 statistical tests  
   analysis of variance, 82, 112, 205, 231  
   assumptions of, 75, 78, 81, 231–2  
   binomial test, 146–7, 227, 232  
   chi-square test, 149, 233  
   correlation, 50, 197–8, 212–15, 219–20, 235  
   eta-squared, 237  
   Kruskall-Wallis test, 244  
   logistic regression, 165  
   Mann-Whitney test, 190  
   Pearson's  $r$ , 212  
   regression, 157  
   Spearman's rho, 212, 250  
    $t$  test, 82, 177, 190, 251  
   Wilcoxon test, 177, 190  
 statistics, 31, 45–86, 87–96, 100, 119, 129–30, 165  
   descriptive, 45–86, 87, 109, 176  
   inferential, 87–96, 112, 239  
   *see also* mean; median; mode; significance; standard deviation; statistical tests; variance  
 Stevens, S., 48  
 stratified sampling, *see* samples, stratified  
 study design, 97, 99–108, 115–27, 137–42, 152–5, 162–3, 172–3, 201–2  
   *see also* independent-samples designs; repeated-measures designs  
 Stukel, T.A., 216  
 summative research, 24–5, 27, 88, 250  
 survival rates, 57  
 systematic observation, 250  
 systematic sampling ( $n$ th name sampling), 36  
   *see also* samples  
  
 Tabachnik, B., 37  
 temperatures, 51–2  
 Thurstone, L.L., 183, 189  
 Thurstone scales, 183–8  
 tied values, when calculating median, 58–9  
 Tramo, M.J., 209, 216  
 transformations for skewed data, 82–6  
 Treisman, A., 159  
 $t$  test, 82, 177, 190, 251  
 twin studies, 42  
 two-alternative forced-choice method, 141

- two-tailed hypothesis, *see* hypotheses
- Type I errors, *see* errors
- Type II errors, *see* errors
- Ugurel, S., 57
- validity, 239, 245–6
- variables, 251  
*see also* dependent  
variables; independent  
variables
- variance, 63, 65, 251
- Walker, I., 23
- Weaver, J.B., 216
- Wilcoxon test, 177, 190
- Willerman, L., 209, 216
- William of Occam, 20
- Wirz-Justice, A., 63
- Z-scores, 251