Methodology 2

CHAPTER OUTLINE

Learning Objectives

By the end of this chapter you should appreciate that:

- scientific psychologists follow strict methods when they conduct research;
- there is a significant difference between reliability and validity;
- there are different types of psychological data (e.g. behavioural, self-report and physiological);
- statistics are vital to psychological research;
- experiments help us to tackle the question of causation;
- there is an important difference between causation and correlation;
- ethical research practice is essential.

INTRODUCTION

'Psychology is the scientific study of behaviour.' 'Psychology is the science of the mind.' 'Psychology is an approach to understanding behaviour that uses scientific methods.' If you study psychology long enough you are likely to hear statements like these.

Why are words like 'science' and 'method' so important to psychologists? Why do they continually talk about ways of doing things, rather than just doing them? Surely anyone can be a psychologist just by being interested in the mind or in behaviour?

The answers to these questions are actually pretty simple. Whether they admit it or not, *everybody* is interested in the mind and behaviour. And just about everybody believes that he or she is an expert when it comes to understanding their own mental life and thoughts. This is understandable. After all, even if all the experts who have contributed to this book were watching you read this chapter, for all their expertise, they would not know as well as you do what is going through your mind.

But being interested in the mind or claiming that you are an expert is not enough to make you a psychologist. What is more, there are lots of other experts, such as philosophers and anthropologists, who are interested in the study of mind and behaviour. What sets psychologists apart from these other experts is their training in *psychological methods*. Many of these methods also appear in other scientific disciplines, partly because psychological methods are derived from general scientific methods. Not all psychologists agree entirely with all the methods, but they all understand them and know how to use the ones relevant to their own work. It is this common training that, more than anything else, makes us psychologists.

Of course, psychologists have methods for doing many different things. The methods we discuss in this chapter are those used for doing research. In other words, they are methods for finding out the answers to questions about the mind and behaviour.

You know from your own experience that different people often come up with different answers to the same question. What is the tallest mountain in the world? You might say Mount Everest, but your answer is only correct if you measure the height from sea level, as we generally do. If you measure from the centre of the Earth, then some mountains in South America are actually taller (because the Earth is not quite round). So the answer to the question depends on how you interpret it. This does not necessarily mean there is no such thing as truth or that everything is arbitrary, but simply that the methods used to reach a conclusion need to be understood fully.

It can be much harder to answer interesting psychological questions than to measure the height of a mountain. There are plenty of psychological questions that have been asked many times and have not been definitively answered. For this reason, it is all the more important that we understand the ways that different psychologists try to answer the questions they are interested in. Then, if we find an unusual answer, we can decide whether it is in some sense worth taking seriously if we first understand the methods that have been used to reach it.

In the nineteenth century, before the science of psychology was established, a lot of people interested in the mind and behaviour studied phrenology. Phrenologists believed that behaviour could be understood by studying the bumps on people's heads. A particular bump in a particular place was supposed to be associated with particular types of behaviour. We now know that this is nonsense.

Nonsense or not, if the last hundred or so years of research in experimental psychology had not taken place, we would have no systematic way of knowing that this is nonsense (see chapter 14). To this day people put forward bad, sometimes harmful, ideas related to important tasks such as treating psychological disorders, assessing employees' ability to do a job, or teaching children to read. Psychological methods allow us to assess these ideas. This means that we can move beyond just arguing about whether an idea is good, and evaluate the evidence with reference to some shared and pre-determined criteria.

This is a vast topic. Even if every page of this massive text were devoted to research methods and statistics, most psychologists would agree that there was still more material that you needed to absorb to become an expert.

The aim of this chapter is to provide some signposts that will help to guide you. It will not enable you to deal with every statistical and methodological question that you might face, but you will learn how to develop and set about answering your own questions.

SOME FUNDAMENTALS

Research methods and statistics

We study research methods and statistics in order to benefit from the science of psychology – to qualify as a psychologist, to use psychological knowledge in other fields, or simply to apply psychology to issues that crop up in everyday life.

Imagine that your employer requires you to sit an intelligence test before you can be considered for higher duties in your organization. You sit the test and are told that your IQ score is 110. What does this mean? Your immediate future may depend on this number, but you can only understand it if you know something about intelligence testing (see chapter 13) and standard scores.

Similarly, newspapers and other sources are full of reports relating to research on psychological issues. Many draw alarming conclusions – often because they mistakenly assume that correlation is the same as causation (see below). A good knowledge of psychological research methods allows you to avoid making the mistakes that journalists, politicians and many others make because they lack the necessary scientific understanding.

Another consideration is that, in most countries, in order to become a psychologist you need to be not just an informed consumer of psychological research but also a *producer* of it. In other words, you need to conduct a piece (or several pieces) of research. Imagine you were employed by a school to determine whether its students are more or less intelligent than students at another school, or in an average school. You would not be able to answer this question properly without conducting a well-planned piece of psychological research.

Even if you do not go on to further study, the ability to conduct, analyse and evaluate psychological research is a very marketable skill that is central to a large number of occupations, such as marketing, management or policy making.

Finally, the intellectual challenges explored in this chapter can be stimulating and interesting in themselves. They are not inconsiderable, but if you can master them you will be better equipped to understand psychology as a whole. This is not because studying methodology and statistics is an end itself (though it can be), but because it is a tool that allows you to get more (personally, intellectually and scientifically) from doing psychology.

Carrying out quality research

Psychological research enables us to find out more about human behaviour and the mental processes that underpin it. We also need to be sure that our answers are correct.

Suppose we are interested in whether 'absence makes the heart grow fonder'. Is it enough simply to look around, make informal observations and come to a conclusion we feel comfortable with? In one sense it is, and, as naturally inquisitive people, we do this sort of thing all the time as a means of forming our own opinions. But this approach inevitably leads different people to different conclusions – because we each focus on different information and have different experiences, different agendas. So some people think absence makes the heart grow fonder while others think the very opposite, that 'absence leads the heart to wander'.

To know which is correct, when each is correct and, more importantly, why, we need to act as scientists, not lay-scientists. Using the *scientific method* differentiates psychology from other disciplines that address similar questions. The scientific method is a set of procedures for acquiring and testing knowledge through systematic observation or experimentation.

scientific method a procedure for acquiring and evaluating knowledge through systematic observation or experimentation

Reliability and validity

The most prized qualities of psychological research are *reliability* and *validity*. Put simply, reliability relates to our confidence that a given finding can be replicated – and is not just a 'freak' or chance occurrence. Reliability in psychological research has much the same meaning in relation to a car. A reliable car is one that nearly always works. A reliable finding can nearly always be reproduced.

reliability the extent to which a given finding will be consistently reproduced on other occasions

validity the extent to which a given study investigates what it purports to investigate

Validity relates to our confidence that a given finding shows what we believe it to show. A valid car is a genuine car (a car that does what it is meant to do). A valid finding is a genuine finding (a finding that is what it purports to be – i.e. one that enhances your understanding in the manner indicated).

Imagine we carry out a study in which we send someone to an exotic overseas location on a number of occasions to see whether this makes them think more favourably of their partner. Let us assume that it does. If it does so repeatedly, then it is a reliable finding. However, the study does not necessarily show that absence makes the heart grow fonder, but may instead demonstrate that exotic holidays make people feel better about their lives in a general way – including about their partners. The finding is therefore almost certainly not valid in a specific sense.

Disputed validity is one of the most common and thorny problems in psychological research. Disputes arise when findings which purport to show one thing are *reinterpreted* to suggest that they actually show something completely different. The fact that the validity of research is often questioned is no bad thing. Indeed, this form of analysis and debate is central to psychological research and to one's skill as a researcher.

Other qualities of good research

As well as being valid and reliable, psychological research needs to be public, cumulative and parsimonious.

To become public, research must be published in a reputable scholarly journal. Sometimes, though rarely, it is translated into popular writing, as was the work of Freud, Pavlov, Piaget and Milgram. The likelihood of a piece of psychological research being adopted for popular publication can depend on such things as topicality, shock value or political trends, and its impact may be transitory. In contrast, the criteria for publication in scientific journals are much more clearly laid out, and they provide an enduring record of the key findings that emerge from a particular piece (or programme) of research.

Cumulative research builds on and extends existing knowledge and theory. It is not enough just to collect information in a haphazard or random fashion. Instead, research should build on previous insights in a given area. Newton expressed this idea clearly when he observed: 'if I have been able to see further than others it is because I have stood on the shoulders of giants'. Generally speaking, a piece of psychological research does not have value in isolation, but by virtue of extending or challenging other work in the field.

The cumulative nature of research is often revealed through literature reviews. These are research papers (normally published in reputable scientific journals) that discuss the results of multiple studies by different researchers. In some cases these reviews involve statistical analyses combining the results of many studies. This process is called *meta-analysis*.

meta-analysis a quantitative method for combining results across a number of studies by first converting the findings of each study into a metric for comparison

Parsimonious research develops explanations of findings that are as simple, economical and efficient as possible. In explaining the results in a given field, psychologists therefore attempt to account for as many different findings as possible using the smallest number of principles. For example, it may be that person A performs better than person B on a test of memory because A was more alert as a consequence of being tested at a different time of day. Or A might have ingested a psychoactive agent before testing took place, whereas B had not. By controlling for the possible influences of time of day, ingested substances and so on, we are left with the most parsimonious explanation for why A and B differ in their level of memory performance.

The role of theory in psychology

Science does not progress simply through the accumulation of independent facts. These facts have to be integrated in terms of theoretical explanations (*theories*). Theories are statements of why, not just what. They are capable of:

1. accounting for multiple facts, and
2. predicting what might happen in novel situations.

theory a coherent framework used to make sense of, and integrate, a number of empirical findings

The purpose of most psychological research is to test such predictions in the form of *hypotheses* – i.e. statements of cause and effect that are derived from a given *theory* and tested by research. So theories generally precede experimentation, not vice versa.

For example, the statement that absence makes the heart grow fonder does not provide a theoretical framework, but the following statement is distinctly more theory-based: 'separation from an object causes us to exaggerate an object's qualities (whether good or bad) because memory distorts reality'. This is because this statement attempts to explain and not just describe the relationship between separation and emotion. Moreover, having made this statement, we can test it by generating hypotheses and doing appropriate research. One *hypothesis* might be that people with memory disorders will make less extreme judgements of absent loved ones than people without such disorders.

hypothesis a statement about the causal relationship between particular phenomena (i.e. A causes B), usually derived from a particular theoretical framework, which is designed to be tested via research investigation

DESIGNING EXPERIMENTS IN PSYCHOLOGY

What can we measure?

Something that differentiates psychology from other sciences is that the things in which we are interested – mental states and processes – can never be directly observed or measured. You cannot touch or see a mood, a thought, a disposition, a memory or an attitude. You can only observe things that are associated with these phenomena.

While this problem does occur in other sciences (such as astronomy), it can often be overcome through technological development (e.g. a better telescope). Psychology has made significant advances too (e.g. measuring skin conductance and brain blood flow), but these techniques still only allow psychologists to study the outcomes of mental activity, or things that are associated with it – never the activity itself.

Psychologists have developed three main types of measure to help them examine mental processes and states:

1 Behavioural measures These involve observation of particular forms of behaviour in order to make inferences about the psychological phenomena that caused or contributed to them. For example, developmental psychologists (see chapter 9) might observe which toys are approached or avoided by children in a play situation. On the basis of such observations, they might plausibly infer that decisions to approach a toy are determined by the toy's colourfulness.

2 Self-report measures These involve asking people about their thoughts, feelings or reaction to a particular question. Provided that it is possible for the participants to reflect consciously on the relevant thoughts or behaviours, their responses can be used either to supplement other behavioural measures or as data in themselves. So a researcher could ask a six-year-old (but clearly not a six-month-old) 'Which toys do you like?' or 'Did you pick that toy because it was brightly coloured?'

3 Physiological measures These involve measuring things that are believed to be associated with particular forms of mental activity. For example, heart rate or galvanic skin response (GSR – a measure of the electrical conductivity of the skin) can serve as measures of anxiety or arousal. In our developmental example, researchers might look at children's heart rate to see whether they become more excited when particular toys are presented or taken away.

Decisions about which of the above measures to use will be dictated by a number of factors. Many of these are practical and will be linked to other methodological choices. For example, self-report measures are relatively cheap and easy to administer, and so lend themselves to survey-based research examining large numbers of people in naturalistic settings. On the other hand, physiological measures can be difficult and expensive to obtain, so they are normally used only in experimental research with very few participants. However, decisions about which measures to use are guided as much by the particular question a researcher wants to address as by practical considerations.

experimental method a research method in which one or more independent variables are systematically manipulated and all other potentially influential variables are controlled (i.e. kept constant), in order to assess the impact of manipulated (independent) variables on relevant outcome (dependent) variables

quasi-experimental method embodies the same features as the experimental method but does not involve the random assignment of participants to experimental conditions

A rundown on research methods

Psychological research involves four main methods: the (true) *experimental method*, the *quasi-experimental method*, the *survey method* (sometimes

survey method the systematic collection of information about different variables in order to investigate the relationship between them

case study method research method that involves a single participant or small group of participants who are typically studied quite intensively

called the correlational method), and the *case study method*.

The experimental method

manipulation the process of systematically varying an independent variable across different experimental conditions (sometimes referred to as the experimental treatment or intervention)

experimental control the method of ensuring that the groups being studied are the same except for the manipulation or treatment under investigation

One very common research method is to manipulate one or more variables and to examine the effect of this *manipulation* on an outcome variable. To do this, the researcher examines participants' responses in the presence and the absence of the manipulation. *Experimental control* is used to make the different situations identical in every respect except for the presence or absence of the manipulation. Experiments can involve different people in each situation or the same people in different situations. People who take part in experiments are called participants, but if you read older research papers they are generally referred to as subjects.

Here is an example. To test the effect of a new training method (a manipulation) on memory, we might take 100 people and expose half of them to the new method. For reasons we will discuss in more detail below, we would assign participants to the two groups on a random basis (e.g. by the toss of a coin). We will call the first group the *experimental group*, as it is subjected to a relevant experimental *treatment*. The other half of our participants would not be exposed to the new training method. As they receive no experimental treatment, they are referred to as a *control group* (also discussed in more detail below). After administering the treatment, we would measure the performance of the two groups on a memory task and then compare the results.

experimental group participants in an experiment who are exposed to a particular level of a relevant manipulation or treatment (as distinct from a control group)

treatment the experimental manipulation of the independent variable

control group participants in an experiment who are not subjected to the treatment of interest (as distinct from the experimental group)

condition a situation in a research study in which participants are all treated the same way

between-subjects design a research study involving a systematic manipulation of an independent variable with different participants being exposed to different levels of that variable

The various levels of treatment in an experiment (including the control) are referred to as *conditions*. This experiment has two conditions and a *between-subjects design* (because the design involves making comparisons between different participants in different conditions). Note, however, that the same question could also have been addressed in a *within-subjects design*, which would involve comparing the memory performance of the same people with and without the new training method. The two basic designs have different strengths and weaknesses, which we will discuss below in relation to issues of experimental control.

within-subjects design a research design in which the same participants are exposed to different levels of the independent variable

The different conditions in the experiment make up the *independent variable* (or IV), sometimes called the treatment variable. A variable is simply something that changes or varies (is not constant). In true experiments, the independent variable is systematically manipulated or varied by the experimenter. Experiments can (and typically do) have more than one independent variable.

independent variable the treatment variable manipulated in an experiment, or the causal variable believed to be responsible for particular effects or outcomes

Experiments also involve at least one *dependent variable* (or DV). This is an outcome or measurement variable, and it is this variable that the experimenters are interested in observing and which provides them with data. In our last example, the dependent variable is the level of memory performance. Use the initial letter 'd' to remember the link between the dependent variable and the data it provides.

dependent variable the variable in which a researcher is interested in monitoring effects or outcomes

Control is the basis of experimental design. It involves making different conditions identical in every respect except the treatment (i.e. the independent variable).

In a between-subjects experiment, this is achieved by a process of *random assignment* of participants to the different conditions. For example, people should be assigned at random (e.g. on the basis of coin tossing), rather than putting, say, the first 50 people in one condition and the second 50 in another. This practice rules out the possibility that there are systematic differences in, say, intelligence, personality or age between the groups.

random assignment the process of assigning participants to study conditions on a strictly unsystematic basis

If there is a difference in results obtained from measuring the dependent variable for each group, and we have equated the groups in every respect by means of random assignment, we can infer that the difference must be due to our manipulation of the independent variable.

The quasi-experimental method

In quasi-experimental studies the independent variable is not (or cannot be) manipulated as such, and so assignment to experimental groups cannot be random. The fact that no manipulation

Pioneer

Donald Thomas Campbell (1916–96) trained as a social psychologist. He was a master methodologist and is best known for devising the method of quasi-experimentation, a statistics-based approach that allows replication of the effects of true randomization, which is often impossible in the study of human behaviour. Campbell also supported use of qualitative methods, according to the goals and context of the study. He promoted the concept of triangulation – that every method has its limitations, and multiple methods are usually needed to tackle important research questions.

occurs interferes dramatically with our ability to make conclusive causal inferences. Examples of independent variables that cannot be manipulated by an experimenter include gender and age. Obviously experimenters cannot change the gender or age of participants, but they can compare the responses of groups of people with different ages or of different genders.

Compared to the experimental method, there is no real control over the independent variable, so we cannot conclude that it is necessarily responsible for any change in the dependent variable. On this basis, as we will see, the quasi-experimental method actually has more in common with survey methodology than with the experimental method. It has all the weaknesses of the experimental method, but it lacks the main strength. In practice, it is often conducted in conjunction with the experimental method. For example, in our learning study we might compare the effect of the new training method on both men and women.

The survey (or correlational) method

The survey method is commonly used to identify the naturally occurring patterning of variables in the 'real world' rather than to explain those patterns (though often people want to put an explanatory gloss on them).

So to examine whether absence makes the heart grow fonder we could conduct a survey to see if people who are separated from their partners because of travelling away from home (group A) say more positive things about their partners than people who never travel away from home without their partners (group B). This might be an interesting exercise, but the validity of any causal statements made on the basis of such findings would be very limited.

For example, if we found from our survey that group A said more positive things about their partners when they were travelling than group B, it would be impossible to demonstrate conclusively that absence was the cause of the difference between groups A and B. In other words, while our survey could show us that absence is associated with a fonder heart, it could not conclusively show that absence actually causes the heart to grow fonder. It is quite possible (odd as it may sound) that the sorts of people who travel away from home without their partners are simply those that like their partners more (so fondness makes the heart go absent). Or perhaps both fondness and absence are caused by something else – for example, social class (i.e. being wealthy makes people both fond and absent).

In large part, then, surveys rely on methodologies that identify relationships between variables but do not allow us to make conclusive causal inferences.

The case study method

Most of the above methods are used for studies involving large numbers of participants. But what if only a few are available? How, for example, would you do research if you were interested in the reading difficulties of people with particular forms of brain damage? To investigate questions like this, researchers often resort to the case study method, which involves intensive analysis of a very small sample. This has particular problems (often with reliability), but some of the most famous studies in psychology have used this method – in particular the work of Freud (see chapter 14).

Taking a qualitative approach

When researchers report and comment on behaviour, without attempting to quantify it, they are using a *qualitative research method*. This involves attempts to understand behaviour by doing more than merely converting evidence into numbers.

Qualitative methods can include coding, grouping and collecting observations without assigning actual numbers to the observation. So a qualitative analysis of the speed of animals might result in the statement that the cheetah is a fast land animal, and a quantitative analysis might involve comparing the maximum speed of animals over (say) 20 metres. To take an example of human behaviour, you probably take a qualitative approach to the friendliness of the people you meet. In other words, you probably judge people as relatively friendly or unfriendly, but you would be unlikely to come up with a number that expresses their friendliness quotient.

Qualitative techniques are sometimes used in the initial stages of quantitative research programmes to complement the quantitative techniques, but they are also used by psychologists who challenge conventional approaches to psychological research. This may be because they believe that the conventional methods are inadequate for addressing the richness and complexity of human behaviour. In turn, many mainstream psychologists are critical of qualitative methods. (For further discussion of qualitative methods, see Haslam & McGarty (2003).)

Experiment versus survey

One common, but mistaken, belief is that the difference between surveys and experiments is a question of location, with surveys being conducted in the community and experiments in the laboratory. This is often the case, but not always. Experiments can be conducted outside laboratories, and surveys can be conducted in them.

The main differences between experiments and surveys relate to the sorts of questions that each can answer. As we suggested earlier, experiments tend to be concerned with establishing causal relationships between variables, and they achieve this by randomly assigning participants to different treatment conditions. Surveys, on the other hand, tend to be concerned with measuring naturally occurring and enduring relationships between variables. Researchers who use surveys usually want to generalize from the sample data they obtain to a wider population. They do this by using the sample to estimate the characteristics of the population they are interested in.

Why choose to carry out a survey rather than an experiment? Two reasons: sometimes we are only interested in observing relationships, and sometimes manipulations simply are not possible. This reasoning is not restricted to psychology. Astronomers or geologists rarely conduct experiments, simply because it is often impossible to manipulate the independent variables of interest (e.g. the position of certain stars or the gravitational force of a planet). Instead they rely largely on the same logic of controlled observation that underpins psychological surveys. But this does not mean that astronomy or geology are unscientific.

Surveys can also allow researchers to eliminate some causal links. If there is no relationship (at least in the survey environment) between variables, this allows us to conclude that one does not cause the other. For example, if no relationship is found between age and intelligence, then it is impossible for intelligence to cause age, or vice versa (bearing in mind that a relationship could be concealed by a third, or background, variable).

Which is the best method to use?

This is a very complex issue and depends on many factors, not least practical ones – including the amount of time, money and expertise that a researcher has. However, as a general principle, it is worth emphasising that no one method is universally superior. Part of any research psychologist's role is to make judgements about the appropriateness of a method for investigating the issues at hand. Being a good researcher is not a question of whether you do experiments or surveys: it is more a matter of when and how you do them.

In view of the potential limitations of any one method, many researchers consider using multiple research methods to explore the same issue in many different ways. This is the process of triangulation. If consistent results are obtained from a variety of different methods (perhaps from a quantitative experiment, a survey and qualitative case studies), this will tend to justify greater confidence in the findings. For this reason, the need to make methodological choices should be seen as an asset for researchers, rather than a basis for arguments about who has the best methods. The challenge researchers face is to exploit that asset appropriately.

Deciding what to manipulate

In selecting an independent variable for any piece of research, we must first decide what we are interested in. For example, we might be interested in whether attributional style (the way people explain events) affects people's responses to failure. We might hypothesize that people who tend to blame themselves for failure (i.e. those who internalize failure) are more likely to become depressed than people who blame their failure on other things (i.e. who externalize failure).

So the central theoretical variable – the focus of our interest – is the participants' attributional style. But, how can we manipulate this for the purposes of our experiment? Clearly we cannot open up people's heads and turn a dial that says 'attributional style' to maximum or minimum.

To get around such obstacles, psychologists usually manipulate the theoretical variable indirectly. They do this by identifying an independent variable that they believe will have a specific impact upon a given mental process, and then check that this is the case.

In our example, the researchers may expose participants to failure (e.g. in a test) and then ask some of them to answer questions like 'Can you explain why you did so much worse than everyone else?' – questions that encourage the participants to reflect on their own contribution to their performance (i.e. to internalize). They may then ask other participants questions like 'Do you think the fact that you were not allowed to revise for the test affected your performance?' – questions that encourage them to reflect on the contribution of other factors to their performance (i.e. to externalize).

To be sure that this manipulation has had the desired effect on the theoretical variable, the researchers may then want to perform a *manipulation check*. For example, in the case given above, the researchers might measure whether the 'internalizing' question produces greater agreement with a measure such as: 'How much do you think you were responsible for the test outcome?'

manipulation check a procedure that checks the manipulation of the independent variable has been successful in changing the causal variable the experimenter wants to manipulate

Note also the significant ethical issues relating to this study. The experimental manipulation could have the effect of making some participants more depressed – indeed, that is the hypothesized outcome in the condition where participants are encouraged to internalize their failure. We discuss ethical issues later in this chapter.

Deciding what to measure

As with the selection of IVs, the selection of dependent variables is often complicated by practical constraints. For example, if we are investigating the impact of alcohol consumption on road fatalities, we may manipulate the independent variable straightforwardly (by getting experimental groups to consume different quantities of alcohol). But it would be irresponsible (and illegal) to then get the participants to drive down a busy street so that we can count how many pedestrians they knock down!

To get round this, we may ask the high alcohol group to consume only a few beverages. But there are two problems with this.

Everyday Psychology

Testing the effectiveness of therapy for depression

'I'm feeling depressed.' Most likely you have heard someone say this. But such statements should not be confused with clinical depression, a disorder that produces greater impairment in everyday functioning than many physical health problems (e.g. hypertension, arthritis, diabetes; see chapter 15). You have approximately a 15 per cent chance of experiencing clinical depression in your lifetime. Should you be unfortunate enough to experience a depressive disorder you would surely want to get treatment for it that is effective. Cognitive behaviour therapy and non-directive counselling are common treatments for people with depressive symptoms (see chapter 16). But are these treatments any more effective than usual general practitioner (GP) care?

You might turn to psychological research for an answer. But how can psychologists evaluate the effectiveness of cognitive behaviour therapy, non-directive counselling and usual GP care in treating depression? An effective test requires the use of an experimental design in which patients are randomly assigned to treatment groups. This is necessary to allow any effects to be attributed to treatment type, rather than any other variable that might lead a participant to choose one treatment over another.

Ward et al. (2000) followed this procedure and allocated patients to one of three treatment groups: two psychological treatments (non-directive counselling and cognitive behaviour therapy) and one control condition (usual GP care). They measured the patients' level of depression before treatment began, at four months and at 12 months following the completion of the treatment.

What did they find? At four months, patients in both the psychological treatment groups (non-directive counselling and cognitive behaviour therapy) had significantly lower depression scores than patients in the control condition (usual GP care). There was no significant difference between the effectiveness of the two psychological treatments.

We can conclude from this experimental test (known as a 'clinical trial' when treatments are being tested; see chapter 16) that the two psychological treatments for depression are effective, at least in the short term. Ideally, however, you would want a treatment that produces lasting results, especially in light of the fact that depression tends to be both chronic and recurrent (see chapter 15). But when Ward et al. examined depressive symptoms at 12 months following treatment there was no significant difference between any of the three groups. In this study, then, the psychological treatments for depression were shown to be effective in the short term but not in the long term. Happily, there are other studies that document the longer-term effectiveness of cognitive behaviour therapy as a treatment for depression (see chapter 16).

Ward, E., King, M., Lloyd, M. et al., 2000, 'Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness', *British Medical Journal*, 321, 1383–8.

First, alcohol may only affect driving behaviour when more than a few beverages are consumed. Second, our dependent variable (number of pedestrians killed) will not be sufficiently sensitive to detect the independent variable's impact. In other words, we may have good reason to think that alcohol could impair driving performance, but the degree of impairment may not (fortunately!) be so profound as to cause a detectable increase in the number of deaths caused.

To deal with this, we therefore have to select dependent variables that are both relevant to the outcome we have in mind and sensitive to the independent variable. In the case of drink-driving, we may look at participants' reaction time, because we believe that this is a critical determinant in driving safety and is likely to be a sensitive enough variable to detect an impairment in driving performance due to alcohol. We can then design and carry out a study in the laboratory, measuring the impact of alcohol consumption on reaction time.

In our attributional style example, too, it is unlikely that our manipulation of the independent variable will have a dramatic impact on the participants' depression. So if our dependent variable was the number of participants who need to be treated by a clinical psychologist, our experiment is very unlikely to uncover any effects. To get around this problem, we could administer a depression inventory, in which we ask the participants a battery of questions (e.g. 'Are you self-confident?', 'Do you feel hopeless about the future?') in order to measure their susceptibility to depression. We could then test our hypothesis by seeing whether scores on the depression inventory revealed a higher susceptibility to depression among participants who had been encouraged to make internal attributions.

Different ways of measuring

The psychologist S.S. Stevens developed a famous distinction between forms of data that psychologists can deal with. The four types he came up with are nominal, ordinal, interval and ratio measures.

Pioneer

Stanley Smith Stevens (1906–73) made significant contributions to several areas of psychology. He was an expert on the psychophysics of hearing and was interested in measurement and experimental psychology. Stevens set out to redefine psychological measurement by changing the perspective from that of inventing operations (the physical view) to that of classifying scales (a mathematical view). He also discovered that methods such as 'just noticeable differences', rating scale categories and paired comparisons produce only ordinal scales. Stevens' most oustanding contribution was his successful argument that there are different kinds of scales of measurement, being the first to define and discuss nominal, ordinal, interval and ratio scales.

Nominal measures

The data collected in this way are in the form of names, which can be categorized but cannot be compared numerically in any way. Examples include genders, countries and personality types.

Ordinal measures

These can be ranked in some meaningful way. Examples are the placings obtained by competitors in a race or an ordered set of categories (e.g. low stress, moderate stress and high stress).

Interval measures

Numerical measures without a true zero point are called interval measures, and cannot be used to form ratios. An example is temperature. The zero point has been arbitrarily chosen to be the freezing point of water rather than absolute zero (where there is no temperature), and it is simply not true that 40 degrees Celsius is twice as hot as 20 degrees Celsius. Similarly, it would not make sense to say that someone who responded with a '6' on the attribution scale above was twice as much of an externalizer as someone who responded with a '3'.

Ratio measures

Full numerical measures with a true zero point are ratio measures. Psychologists frequently assume that scores obtained from psychological measurement can be treated as ratio measures. But this assumption is not always justified.

Producing trustworthy results

Internal validity

We can be confident about the results of psychological research when the methods are valid. An experiment is said to have *internal validity* when we are confident that the results have occurred for the reasons we have hypothesized, and we can rule out alternative explanations of them.

internal validity the extent to which the effect of an independent (manipulated) variable on a dependent (outcome) variable is interpreted correctly

These alternative explanations (or threats to internal validity) can involve an experimental *confound* – an unintended manipulation of an independent variable. The risk of confounds can be reduced by better experimental design.

confound an unintended or accidental manipulation of an independent variable that threatens the validity of an experiment

Suppose we conduct a study to look at the effect of crowding on psychological distress by putting 50 people in a crowded room and 50 people in an open field. Having found that the people in the room get more distressed, we may want to conclude that crowding causes distress. But the participants' body temperature (generated by having a lot of people in one room) may represent a confound in the study: it may be the heat, not the crowding, that produces the effects on the dependent variable. The experiment could be redesigned to control for the effects of this confound by using air-conditioning to keep the temperature the same in both conditions.

External validity

A study has a high level of *external validity* when there are no reasons to doubt that the effects obtained would occur again outside the research setting. We might, for example, question a study's external validity if participants responded in a particular way because they knew that they were taking part in a psychological experiment. They might inadvertently behave in a way that either confirms or undermines what they believe to be the researcher's hypothesis. In experiments we usually try to deal with this specific potential problem by not telling experimental participants about the hypotheses that we are investigating until after the experiment has finished.

external validity the extent to which a research finding can be generalized to other situations

STATISTICS IN PSYCHOLOGY

Samples and populations – sorting out the jargon

You will often hear psychologists talking about samples and populations in relation to statistical analysis of research. What do they mean by these terms?

A population is a set of people, things or events that we are interested in because we wish to draw some conclusion about them. The population could consist of all people, or all people with schizophrenia, or all right-handed people, or even just a single person.

A sample is a set selected from the population of interest and used to make an inference about the population as a whole. This kind of inference is called a *generalization*. A sample would normally be a group of people selected from a larger group, but it could also be a sample of behaviour from one person, or even a sample of neurons from a region of the brain (see chapter 3).

generalization related to the concept of external validity, this is the process of making statements about the general population on the basis of research

If we wish to generalize to a population, we need to make sure that the sample is truly representative of the population as a whole. This means that the sample should be similar to the population in terms of relevant characteristics. For example, if we are doing research on the human visual system, then members of our sample group need to have eyesight that is similar to the rest of the human population (as opposed to being, for example, noticeably worse). The easiest and fastest way to achieve this is to draw a *random sample* (of a reasonable size) from the population.

random sample a sample of participants in which each has the same chance of being included, ensured by using random participant selection methods (e.g. drawing lots)

Describing numerical results

descriptive statistics numerical statements about the properties of data, such as the mean or standard deviation

Two key properties, referred to as *descriptive statistics*, come into play when we describe a set of data – or the results of our research. These are the *central tendency* (what we usually call the average) and the amount of *dispersion* – or variation.

central tendency measures of the 'average' (most commonly the mean, median and mode), which tell us what constitutes a typical value

dispersion measures of dispersion (most commonly range, standard deviation and variance) describe the distance of separate records or data points from each other

Imagine a choreographer selecting a group of dancers for a performance supporting a lead dancer who has already been cast. The choreographer wants the supporting cast to be pretty much the same height as the lead dancer and also pretty much the same height as each other. So the choreographer is interested in the average height (which would need to be about the same as the lead dancer's height) and the dispersion, or variation, in height (which would need to be close to zero).

There are a number of ways in which the choreographer – or the psychologist – can measure central tendency (average) and dispersion.

Measures of central tendency

Measures of central tendency give us a typical value for our data. Clearly, 'typical' can mean different things. It could mean:

- the average value;
- the value associated with the most typical person; or
- the most common value.

In fact, all three are used by researchers to describe central tendency, giving us the following measures:

- The *mean* is the average value (response) calculated by summing all the values and dividing the total by the number of values.
- The *median* is the value with an equal number of values above and below it. So, if all values are ranked from 1 to N, the median is the $((N+1)/2)$th value if N is odd. If N is even, the median is the mean of the two middle values.
- The *mode* is the value that occurs most frequently in a given data set.

mean the sum of all the scores divided by the total number of scores

median the middle score of a ranked array – equal to the $((N+1)/2)$th value, where N is the number of scores in the data set

mode the most commonly occurring score in a set of data

Measures of dispersion

We might also want to describe the typical distance of responses from one another – that is, how tightly they are clustered around the central point. This is typically established using one of two measures. The first and probably most obvious is the *range* of responses – the difference between the maximum and minimum values. But in fact the most commonly used measure of dispersion is *standard deviation* (SD). This is equal to the square root of the sum of the squares of all the differences (deviations) between each score and the mean, divided by the number of scores (in fact, the number of scores minus one if we want a population estimate, as we usually do). If this sounds complex, do not be too concerned: scientific calculators allow you to compute standard deviations very easily. The square of the standard deviation is called the *variance*.

standard deviation the square root of the sum of the squares of all the differences (deviations) between each score and the mean, divided by the number of scores (or the number of scores minus 1 for a population estimate)

variance the mean of the sum of squared differences between a set of scores and the mean of that set of scores; the square of the standard deviation

Research close-up 1

A survey on psychiatric disorders

The research issue

Until the mid 1980s, research into the prevalence of psychiatric disorders, such as affective (mood) disorders, relied on institutional records. Lubin et al. (1988) set out to investigate the relationships between affect and demographic and physical health variables in a representative population sample.

Design and procedure

The Revised Multiple Affect Adjective Check List (MAACL-R) was administered to 1,543 adults throughout the United States. This sample was designed to produce an approximation of the adult civilian population at the time. The MAACL-R provided measures of five traits: anxiety, depression, hostility, positive affect (optimistic mood state) and sensation seeking. In addition, participants were interviewed in order to elicit demographic information and subjective impressions of physical health.

Results and implications

The sensation-seeking scale was not found to be internally reliable and so was not used in further analyses. The researchers analysed the data primarily using *t*-tests, analysis of variance and correlation.

Statistical tests showed that females scored significantly higher than males on measures of anxiety, depression and positive affect. Correlational analyses revealed significant relationships between measures of affect and subjective measures of physical health. Positive affect had a highly significant positive correlation with physical health ($r = .30$), while anxiety ($r = -.10$), depression ($r = -.20$) and hostility ($r = -.09$) were all significantly negatively correlated with physical health (though the correlations were relatively small).

Note, however, that this study does not allow us to conclude that feeling healthy causes one to be happier, less anxious, less depressed and less hostile, or that being happy (and not anxious, depressed or hostile) causes one to be physically healthy.

Lubin, B., Zuckerman, M., Breytspraak, L.M., Bull, N.C., Gumbhir, A.K., & Rinck, C.M., 1988, 'Affects, demographic variables, and health', *Journal of Clinical Psychology*, 44, 131–41.

Compared to the range alone, standard deviation tells us a lot about a distribution of scores, particularly if they are *normally distributed* – a feature we discuss further below. If this is the case, we know, for example, that about 68 per cent of all values will fall within 1 SD of the mean, 95 per cent fall within 2 SDs and 99 per cent fall within 3 SDs. For reasons that will become clear in later chapters (e.g. chapter 13), this sort of information is very useful.

How can we confidently generalize our results?

Although psychologists often spend a lot of time studying the behaviour of samples, most of the time they want to generalize their results to say something about a whole population – often called the underlying population. Knowing how ten particular people are going to vote in an election may be interesting in itself, but it is even more interesting if it tells us who is likely to win the next election.

But how can we make inferences of this sort confidently? By using *inferential statistics* we can make statements about underlying populations based on detailed knowledge of the sample we study and the nature of random processes. The key point here is that, while random processes are (as the name tells us) random, in the long run they are highly predictable. Not convinced? Toss a coin. Clearly, there is no way that we can confidently predict whether it is going to come down heads or tails. But if we were to toss the coin fifty times, we could predict, reasonably accurately, that we would get around twenty-five heads. The more tosses we make, the more certain we can be that around about 50 per cent of the tosses will come up heads (and it is this certainty that makes the business of running casinos very profitable).

inferential statistics numerical techniques used to estimate the probability that purely random sampling from an experimental population of interest can yield a sample such as the one obtained in the research study

Of course, psychologists do not usually study coin tosses, but exactly the same principles apply to things they do study. For example, the mean IQ is 100 (with an SD of 15), so we know that if we study a large number of people, about 50 per cent will have an IQ greater than 100. So if we get data from 100 people (e.g. a class of psychology students) and find that all of them have IQs greater than 100, we can infer with some confidence that there is something psychologically 'special' about this sample.

Our inference will take the form of a statement to the effect that the pattern we observe in our sample is 'unlikely to have arisen as a result of randomly selecting (sampling) people from the population'. In this case, we know this is true, because we know that psychology students are not selected randomly from the population but are selected on the basis of their performance on tests related to IQ. But even if we did not know this, we would be led by the evidence to make an inference of this kind.

Inferential statistics allow researchers to quantify the probability that the findings are caused by random influences rather than a 'real' effect or process. We do this by comparing the distribution obtained in an empirical investigation with the distribution suggested by statistical theory – in this case the *normal distribution*. We then make predictions about what the distributions would look like if certain assumptions (regarding the lack of any real effect on the data) were true. If the actual distribution looks very different from the one we expect, then we become more confident that those assumptions are wrong, and there is in fact a real effect or process operating.

normal distribution the symmetrical, bell-shaped spread of scores obtained when scores on a variable are randomly distributed around a mean

For example, the distribution of the mean IQ score of groups of people drawn from the population tends to have a particular shape. This is what we mean by the normal distribution – see figure 2.1. If a particular set of data does not look as though it fits the (expected) normal distribution, then we would start to wonder if the data really can be assumed to have been drawn at random from the population in question. So if you drew a sample of 100 people from a population and found that their mean IQ was 110, you can be fairly sure that they were not randomly drawn from a population with a mean of 100. Indeed, the normal distribution shows us that the likelihood of an event as extreme as this is less than one in a thousand.

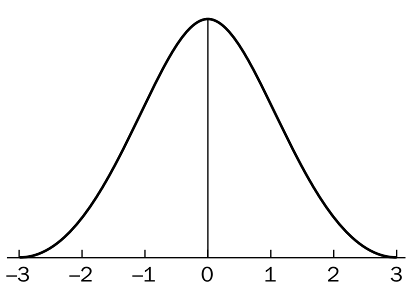

Figure 2.1

The shape of the normal distribution.

Finding out if our results are remarkable

Is there something going on here?

When we use inferential statistics, we might be in a position to make exact probability statements (as in the coin tossing example), but more usually we have to use a test statistic.

Two things influence our judgement about whether a given observation is in any sense remarkable:

1. the information that something is 'going on'; and
2. the amount of random error in our observations.

In the IQ example, information comes from the fact that scores are above the mean, and random error relates to variation in the scores of individual people in the sample.

For this reason, the statistics we normally use in psychology contain both an information term and an error term, and express one as a ratio of the other. So the test statistic will yield a high value (suggesting that something remarkable is going on) when there is relatively more information than error, and a low value (suggesting that nothing remarkable is going on) when there is more error than information.

Imagine we gave an IQ test to a class of 30 children and obtained a mean IQ of 120. How do we find out the statistical likelihood that the class mean differs reliably from the expected population mean? In other words, are we dealing here with a class of 'smart kids', whose performance has been enhanced above the expected level by some factor or combination of factors? Or is this difference from the population mean of 100 simply due to random variation, such as you might observe if you tossed a coin 30 times, and it came up heads 20 times?

A statistical principl known as the *law of large numbers* tells us that uncertainty is reduced by taking many measurements of the same thing (e.g. making 50 coin tosses rather than one). It means, for example, that although around 9 per cent of the population have IQs over 120, far fewer than 9 per cent of classes of 30 randomly selected students will have a mean IQ over 120. This statistical knowledge makes us more confident that if we *do* find such a class, this is highly unlikely to be a chance event. It tells us instead that these children are performing considerably higher than might be expected.

law of large numbers the idea that the average outcomes of random processes are more stable and predictable with large samples than with small samples

We can summarize the process here as one of deciding where the sample mean lies in relation to the population mean. If there is a very low chance of sampling that mean from the population we conclude that the sample is probably not drawn from that population but instead belongs to another population. Perhaps more intelligent students were assigned to this class by the school authorities, or perhaps they came from an area where education funding was especially good. In short, we cannot be sure *what the explanation is*, but we can be relatively sure that *there* *is something to be explained* and this is the purpose of conducting statistical tests.

Judging when two conditions in an experiment are sufficiently different to be worth interpreting

Think back to our 'memory training study', in which one group of participants in an experimental condition experience a new training method and another group in a control condition do not, then both groups take a memory test. Common sense tells us that we are likely to get two sets of memory scores – one for the experimental condition, one for the control – with different means.

But how do we decide whether the difference is big enough to be meaningful? This is where inferential statistics come into play. Appropriate statistical procedures allow us to decide how likely it is that this difference could occur by chance alone. If that likelihood is sufficiently low (typically less than 1 in 20 or 5 per cent), we would reject the *null hypothesis* (expressed as H_0) that there is no difference between the means and that the manipulation of the independent variable has had no effect. Instead we would conclude that the manipulation of the IV has had a significant impact on the dependent variable – that is, that training does indeed improve memory. This process is typically referred to as *significance testing*, and this is one of the main approaches to statistical inference. While statistical tests can never tell us whether our results are due to chance, they can guide us in judging whether chance is a plausible explanation.

null hypothesis the hypothesis that the research reveals no effect

significance testing the process of deciding whether research findings are more plausibly due to chance (H_0) or due to real effects (H_1)

How does significance testing work in this case – that is, when comparing two means? In essence it comes down to the difference between the means relative to the variation around those means and the number of responses on which the means are based. The statistics that we calculate for comparing means are called *t* and *F* statistics. A large *t* or *F* statistic means there is a small probability that a difference as big as the one we have obtained could have occurred by randomly selecting two groups from the same population (i.e. it is not likely that the difference is due to chance). If that probability is sufficiently small, we conclude that there probably is a real difference between the means – in other words, that the difference is statistically significant.

Judging whether two variables are related

A lot of what we have discussed so far relates to comparisons between means, which is typically what we do when we use experimental methodology. But in a range of other research situations we are interested in assessing the relationship between two variables. For example, how is height related to weight? How is stress related to heart disease?

This type of question can be asked in experiments (what is the relationship between the amount of training and memory?), but is more typically addressed in surveys, where the researcher has multiple values of each variable. Suppose we are working on the concept of attraction, which occurs at many levels. We might have data recording both people's attraction to their partners and the amount of time they have spent apart from them, our interest lying in whether higher levels of attraction are associated with higher levels of time spent apart, or whether high levels of attraction are associated with lower levels of time spent apart, or whether there is no clear relationship between the two variables. This type of data is described as *bivariate*, as opposed to *univariate*.

bivariate the relationship or association between two variables ('variate' is another word for variable)

univariate relating to a single variable

One useful way to set about answering this type of question is to draw a scatterplot – a two-dimensional graph displaying each pair of observations (each participant's attraction to their partner and the time spent apart).

Figure 2.2 shows an obvious relationship between attraction and time apart: the higher one is, the higher the other is. We describe this as a positive correlation. A negative correlation would be obtained when one value decreases as the other increases.

Note that the stronger the relationship, the less scattered the various points are from a straight line, and the more confidently we can estimate or predict one variable on the basis of the other. In this example, it becomes easier to estimate from someone's

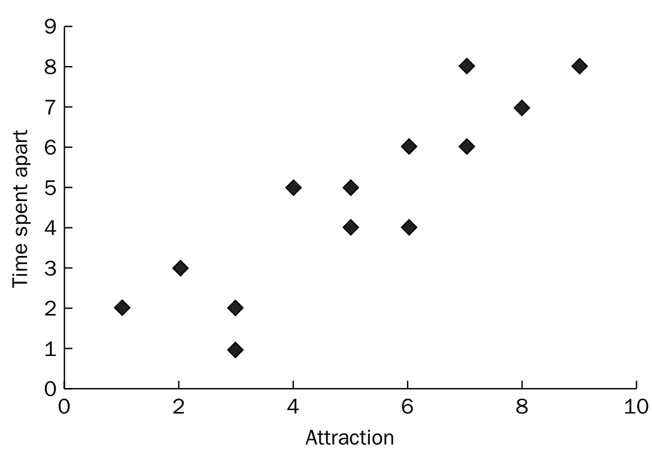

Figure 2.2

A graph (called a scatterplot) of the relationship between attraction and time spent apart. Note: these are not real data.

Research close-up 2

An experiment on group pressure

The research issue

Group pressure and conformity have been researched extensively in social psychology. One of the most influential studies is Asch's (1951) experiment involving a line judgement task. Asch wanted to investigate whether participants would choose an incorrect answer in order to conform to the group, and whether conformity would increase as the size of the group increased.

Design and procedure

Participants were shown two cards. One card showed one line, while the other card showed three. The participant's task was to indicate which of the three lines on the second card was the same length as the line on the first card.

Each participant was assigned to one of seven group size conditions, ranging from completing the task alone to completing the task with 16 other group members. The other group members were confederates and were trained to give the same wrong answer on 12 out of the 18 trials. Using this experimental design, Asch was able to test his null hypothesis, which was that the confederates would not affect the participants' responses.

Results and implications

Asch rejected his null hypothesis because, on average, participants chose the wrong line more often when there were confederates present than when they completed the task alone.

Furthermore, the mean number of errors increased as the group size increased. This led Asch to conclude that there was probably pressure to conform to a group's opinions and decisions, and that this pressure was likely to increase as groups became larger. Nonetheless, the results of this study were not subjected to statistical testing, so there is uncertainty as to the conclusions that can be drawn from this particular study.

Since then, however, other studies researching group pressure have found similar results that have been proved statistically significant. (See chapter 18 for more on this.)

Asch, S.E., 1951, 'Effects of group pressure on the modification and distortion of judgements' in H. Guetzkow (ed.), *Groups, Leadership and Men*, Pittsburgh: Carnegie.

attraction how much time they have spent apart from their partner, or to estimate level of attraction from the time spent apart.

Understanding correlation

Correlation does not imply causation

A mistake that is made by researchers more often than it ought to be is to assume that, because two variables are highly correlated, one is responsible for variation in the other. *Always remember that correlation does not imply causation.*

Suppose we conduct a study that reveals a strong positive correlation between the consumption of alcohol and aggressiveness. On this basis it cannot be concluded that alcohol causes aggressiveness. You could equally argue that aggressiveness causes people to drink more, or the relationship may be the product of a third factor, such as upbringing. Perhaps having hostile parents leads people to be aggressive and also to drink more. It is therefore possible that upbringing encourages alcohol consumption *and* aggressiveness, without each having a direct effect on the other.

There are many real-life examples of spurious correlations that have arisen from the influence of a third factor. For example, when researchers found that there was a high correlation between the presence of 'spongy tar' in children's playgrounds and the incidence of polio, they misguidedly inferred that 'spongy tar' caused polio. As a result, some schools went to great expense to get rid of it. In fact, both spongy tar and polio were both linked to a third factor: excessively high temperature. So it was this that needed to be controlled, not the type of tar in the playground.

This inability to draw strict causal inferences (and the associated temptation to do so) is by far the most serious problem associated with both correlational and survey methodology.

The measurement of correlation

Correlations are usually measured in terms of *correlation coefficients*. The most common of these is the Pearson product–moment correlation,

correlation coefficient a measure of the degree of correspondence or association between two variables that are being studied

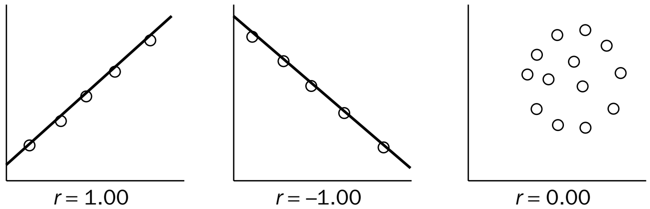

Figure 2.3

Examples of positive, negative and zero correlations.

Pearson's *r* the commonly used name for Pearson's product-moment correlation coefficient

or *Pearson's r*. The value of *r* indicates how strong a correlation is and can vary from −1.00 to +1.00.

As with *t*-tests, computation of Pearson's *r* involves going through a series of standard steps. These allow us to establish whether high scores on one variable are associated with high scores on the other, and if low scores on one variable are associated with low scores on the other.

As figure 2.3 illustrates, an *r*-value of +1.00 indicates a perfect positive correlation, and an *r*-value of −1.00 indicates a perfect negative correlation. In both these cases, the value of one variable can be predicted precisely for any value of the other variable. An *r*-value of 0.00 indicates there is no relationship between the variables at all.

Pioneer

Karl Pearson (1857–1936) graduated from Cambridge University in 1879 but spent most of his career at University College, London. His book *The Grammar of Science* (1892) was remarkable in that it anticipated some of the ideas of relativity theory. Pearson then became interested in developing mathematical methods for studying the processes of heredity and evolution. He was a major contributor to statistics, building on past techniques and developing new concepts and theories. He defined the term 'standard deviation' in 1893. Pearson's other important contributions include the method of moments, the Pearson system of curves, correlation and the chi-squared test. He was the first Galton Professor of Eugenics at University College, London, holding the chair from 1911 to 1933.

FINAL THOUGHTS

The role of ethics in research

Psychology is a science, and science is part of society. It follows that psychological scientists must work within limits imposed by society, including the standards that society sets for behaviour.

Psychological researchers are bound by research ethics – a code, or set of rules, that tells them which sorts of behaviour are acceptable when conducting research. These rules relate primarily to avoiding the risk of harm to research participants.

One important feature of ethical research is *informed consent*. The participants (or their guardians if they are children) must have the research procedures explained to them so that they can make an informed choice about whether they wish to participate. Any risks of harm to participants must be minimised, and if they cannot be eliminated, they must be justified.

informed consent the ethical principle that research participants should be told enough about a piece of research to be able to decide whether they wish to participate

Imagine some clinical psychologists develop a new form of therapy to treat a mental illness. Rather than simply using the therapy in their practice, they must first decide how to evaluate the treatment. Suppose that, in reality, the treatment has a slight risk of causing harm to participants. Before the researchers can test the effectiveness of the treatment, they must be confident that the potential benefits heavily outweigh any potential harm.

Where research involves animals, their treatment must be humane and meet the standards of animal welfare.

Major psychological societies, such as the American Psychological Association and the British Psychological Society, maintain web links that provide details of their ethical codes, and all researchers need to be familiar with these.

Want to know more about research?

As we noted at the outset, this chapter can really only scratch the surface of the field of research methods in psychology. If you want to know more, there are a number of very good books on research methods in psychology. Many of them cover statistics in great detail, others cover methodology and a third selection cover both methodology and statistics.

However your training in psychology develops, your most important aspiration should be to become an informed and critical user of its methods. It is only by confidently exploring the limits and strengths of methodology that we are able to extend the limits and build on the strengths of psychological knowledge.

Summary

- Research, as well as being important in its own right, underpins every aspect of involvement in psychology from introductory studies right through to professional practice.
- It is important therefore that research is done effectively. This means ensuring that research achieves valid findings through reliable and reproducible methods that involve testing and developing explanations (theories) of these findings.
- Psychologists have various methods at their disposal, and being a good researcher means choosing the appropriate method for the question in hand. The appropriateness of a method also depends on how ethical it is.
- Experiments have substantial advantages: by using random assignment to equate conditions in all respects except for the manipulation, they allow us to explain any results in terms of two competing explanations – chance, and the effects of the experimental treatment.
- Statistical tests are particularly useful for helping us to decide between these two explanations. For example, if we find differences in the mean scores of two different conditions that are much larger than we would expect, it is unlikely that both conditions are identical (i.e. belong to the same statistical population).
- Statistics can also help us to judge whether it is plausible that two variables are related. Such relationships are called correlations, and there are many ways to measure them. In considering any correlation, it is always crucial to remember that correlation does not imply causation. Just because two variables are related, we cannot conclude that the first variable causes the second, even if this seems plausible. It is also possible that the second causes the first, or that some third factor causes both.

REVISION QUESTIONS

1. Does it matter whether psychology is a science? If so, why?
2. What is problematic about findings that are reliable but not valid?
3. What do you think are the relative merits of behavioural, self-report and physiological measures of psychological processes?
4. How do experiments help us to deal with the problem that correlation does not imply causation?
5. What are the main descriptive statistics?
6. Why do we carry out statistical tests?
7. Can statistical tests tell us whether results are due to chance?
8. Are positive correlations stronger than negative correlations?
9. What purpose do research ethics serve?

Further Reading

BPS Code of Conduct, Ethical Principles & Guidelines (Nov. 2000). London: British Psychological Society. www.bps.org.uk/documents/Code.pdf
Similar guidelines have been developed by the American Psychological Association (http://www.apa.org/ethics/code2002.html), the Australian Psychological Society (http://www.psychsociety.com.au/aps/ethics/default.asp) and the New Zealand Psychological Society (www.psychology.org.nz/psychinnz/2002%20Code%20Cover.pdf).

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally.
The classic short text on research design that all psychologists should read (and read again).

Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologist*, 49, 997–1003.
An interesting discussion of some of the problems with the significance testing approach that we have outlined here.

Haslam, S.A., & McGarty, C. (2003). *Research Methods and Statistics in Psychology*. London: Sage.
This is the book on which this chapter was based. It involves a far more detailed treatment of the material (especially statistical tests) than is possible here.

Howell, D.C. (1999). *Fundamental Statistics for the Behavioral Sciences*. Belmont, CA: Duxbury.
An accessible and widely used introduction to statistics written by a psychologist.

Leong, T.L., & Austin, J.T. (1996). *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants*. Thousand Oaks, CA: Sage.
A useful guide on a whole range of research matters from research design to publication.

Contributing authors:
Craig McGarty & Alex Haslam