Why do we play sport? How does a good player become an elite player? How useful is exercise in combating childhood obesity? Research is vital to understanding all aspects of sport, exercise and health and it is only through research that questions such as these can be answered. Sport and exercise scientists are playing an ever more significant role within society, often leading the way with developments in sport performance and health sciences. However, to stay at the forefront of applied work and provide the highest quality of service to clients, you need to base your advice on a sound knowledge base. This can only be developed through high quality research.

Throughout this unit, you will develop a knowledge and understanding of some of the key issues that are associated with research methods. This will range from examining some different quality indicators such as validity and reliability, to looking at how you can make sure that you are meeting the necessary professional standards associated with research in sport and exercise sciences. After this, you will do more applied work examining how you can structure and organise research, and look at the different ways that you can collect and analyse data using both qualitative and quantitative methods.

Learning outcomes
After completing this unit you should:
1. know key issues in research methods for the sport and exercise sciences
2. know data collection techniques for the sport and exercise sciences
3. know qualitative data analysis techniques for the sport and exercise sciences
4. know quantitative data analysis techniques for the sport and exercise sciences.
## Assessment and grading criteria

This table shows you what you must do in order to achieve a pass, merit or distinction grade, and where you can find activities in this book to help you.

<table>
<thead>
<tr>
<th>To achieve a pass grade the evidence must show that you are able to:</th>
<th>To achieve a merit grade the evidence must show that, in addition to the pass criteria, you are able to:</th>
<th>To achieve a distinction grade the evidence must show that, in addition to the pass and merit criteria, you are able to:</th>
</tr>
</thead>
</table>
| **P1** describe qualitative and quantitative research  
See Assessment activity 4.1, page 6 | **M1** explain key issues that affect research in sport and exercise sciences  
See Assessment activity 4.1, page 6 | **D1** analyse key issues that affect research in sport and exercise sciences  
See Assessment activity 4.1, page 6 |
| **P2** identify key issues that affect research in sport and exercise sciences  
See Assessment activity 4.1, page 6 | | |
| **P3** outline the types, techniques, and classifications of data that are common in research in the sport and exercise sciences  
See Assessment activity 4.2, page 13 | | |
| **P4** describe two ethical and legal issues associated with research in sport and exercise sciences  
See Assessment activity 4.3, page 16 | **M2** explain the implications of not working both ethically and legally when conducting research in the sport and exercise sciences  
See Assessment activity 4.3, page 16 | **D2** analyse the implications of not working both ethically and legally when conducting research in the sport and exercise sciences  
See Assessment activity 4.3, page 16 |
| **P5** describe the three main stages of qualitative data analysis in the sport and exercise sciences  
See Assessment activity 4.4, page 19 | **M3** justify, for a selected research-based example, the most appropriate research design and techniques for qualitative data collection and data analysis  
See Assessment activity 4.4, page 19 | |
| **P6** describe two contrasting quantitative data analysis techniques used in the sport and exercise sciences  
See Assessment activity 4.5, page 32 | **M4** justify, for a selected research-based example, the most appropriate research design and techniques for quantitative data collection and data analysis  
See Assessment activity 4.5, page 32 | |
How you will be assessed

This unit will be assessed by internal assignments that will be designed and marked by the tutors at your centre. Your assessments could be in the form of:

- written reports
- oral presentations
- discussions
- posters.

Simon, 19-year-old trainee performance analyst

This unit has helped me to understand that there is more to research than just reading things. It's all about discovering things through your own investigations and there are lots of different ways of carrying out these investigations.

There were lots of practical learning opportunities throughout this unit where we could collect our own data and analyse it – so we got a feel for the overall process and it felt more like doing sport science rather than just being given lots of number crunching to do. I'd definitely recommend it to you!

I enjoyed looking at the different ways in which I can apply the work on statistics to a performance environment. When I was at school, I was always a bit scared of maths but the way I learned about doing statistics in this unit helped me to get over that fear and now I'm doing a job where using maths is a major part of what I do!

Over to you

1. Which areas of this unit are you looking forward to?
2. Which bits do you think you might find difficult?
3. What do you think you will need to do to prepare for this unit?
1. Key issues in research methods for the sport and exercise sciences

What is the role of research in modern sport and exercise science?
What do you think about when you hear the term ‘research’? For some people, it means people in white coats dropping things into petri dishes whereas for others it could mean big needles and complex machines. How do you think research is applicable to sport and exercise environments?

1.1 Research
Research is a systematic process of investigation and study carried out with the goal of advancing knowledge.

Before conducting research, you need to consider which type of research would be the most suitable to answer your research questions. There are two main types of research:

- quantitative
- qualitative.

Quantitative and qualitative research

Quantitative research is a formal, objective and systematic process in which numerical data is used to obtain information. It involves testing a hypothesis or trying to discover relationships. It is generally deductive research (this means that a scientist would start from a hypothesis and then begin observations to prove the hypothesis). It is designed to establish differences, relationships or causality (does one thing cause another?).

Qualitative research is generally subjective and involves words rather than numbers. It looks at feelings, opinions and emotions and is concerned with trying to explain why rather than what or how many. It tends to be inductive, which means a hypothesis can be developed through the research. It tries to explain differences, relationships or causality. Qualitative data can also produce quantitative data, for example, you may record how many people said that they like playing sport because they can spend time with their friends.

1.2 Key issues
Key issues influence the quality of your research. These are:

- validity
- reliability
- accuracy
- precision.
Key terms

**Quantitative research** – a formal, objective and systematic process in which numerical data is used to obtain information.

**Qualitative research** – a more subjective form of research that tries to explain differences, relationships or causality using non-numerical data such as words.

**Validity** – whether you are measuring what you are supposed to be measuring.

**Reliability** – the repeatability of a set of results.

**Accuracy** – how close a measurement is to the true value.

**Precision** – how fine or small a difference a measurement can detect.

Validity

Validity is essential in research because it relates to whether you are actually measuring what you planned to measure. There are different types of validity, but two key types are internal validity and external validity. Internal validity relates to whether the results of the study can be attributed to the different treatments in the study. This means that for your research to claim internal validity, you need to ensure that you have controlled everything that could affect the results of the study. External validity relates to whether or not the results of the study can be applied to the real world.

Reliability

Reliability relates to whether, if you carried out the research again, you would get the same or similar results. However, reliability can be claimed without results being correct. For example, if you always ask the wrong questions in research, you will always get the same wrong answers. This will mean the test is reliable because you have received the same wrong answers, even though they are not the ones you wanted.

In quantitative research, reliability can be one researcher conducting the same test on the same individual on a number of occasions, and getting the same or similar results. Alternatively, it can be different researchers conducting the same test on the same individual and getting the same or similar results. In qualitative research, reliability relates to the same researcher placing results into the same categories on different occasions, or different researchers placing results into the same or similar categories.

There are certain factors you should take into account that can affect reliability. For example:

- errors can happen when researchers don’t know how to use the equipment correctly
- the equipment may be poorly maintained
- the wrong type of equipment may be selected.

There are two types of reliability: inter-researcher reliability and test-retest reliability.

Inter-researcher reliability examines whether different researchers in the same situation would get the same (or similar) results. An example of when inter-researcher reliability is a problem is body composition assessment. When people are learning to use the skinfold calliper technique of assessing body composition, it is difficult to take accurate measurements from the correct sites. Researchers come up with different values. When this happens, you cannot claim to have achieved inter-researcher reliability.

Test-retest reliability relates to doing the same test on different occasions and getting the same (or similar) results. An example of a test-retest reliability issue in sport or exercise research is the measurement of heart rate. Heart rate can be affected by different factors, such as temperature, time of day, diet, sleep patterns, physical activity levels and alcohol. If you measured the heart rate on the same person at the same time of day, but on different days, you could get different measurements.

Accuracy

Accuracy relates to how close your measurement is to the ‘gold standard’, or what you are intending to measure. Imagine you are looking at the weight of a boxer before a fight. If the boxer has an actual weight of 100 kg and your weighing device shows he weighs 100.1 kg, you could say this is accurate. However, if the measuring device shows he weighs 103 kg, you would say this isn’t accurate as it isn’t close to his body weight.

Precision

When working in a research setting, any measurement you take will have some unpredictability. The degree of unpredictability relates to the amount of precision the tool selected for measurement has. Precision is related to the refinement of the measuring process.
It is concerned with how small a difference the measuring device can detect. Precision is closely related to repeatability/reliability. An easy way to get to grips with accuracy and precision is to think about target sports, such as archery. If you were to hit the bull’s eye on the archery board with all of your arrows, you would say that you had been both accurate and precise. However, if you missed the board completely in different directions with your arrows, you would say that you had been neither accurate nor precise. This is shown in the diagram below.

Figure 4.1: Can you explain the accuracy and precision shown on each archery target?

Assessment activity 4.1

You are applying for a job as a performance analyst within a national cricket organisation. As part of the application process, the head of sport science support for the organisation has provided you with a sample data set and asked you to produce a PowerPoint presentation that interprets the data set. The data set is below with the accompanying presentation guidance.

<table>
<thead>
<tr>
<th>Bowler</th>
<th>Speed gun results</th>
<th>Timing gate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75 mph 80 mph 77 mph</td>
<td>76.98 mph 77.02 mph 76.95 mph</td>
</tr>
<tr>
<td>2</td>
<td>81 mph 84 mph 83 mph</td>
<td>80.02 mph 80.05 mph 80.06 mph</td>
</tr>
</tbody>
</table>

1. Describe qualitative and quantitative research. Then say whether the table contains qualitative or quantitative data. **P1**
2. Name the different key issues that could have affected the quality of data in the data set. **P2**
3. Using the data set, explain the different key issues (for example, the bowling speeds of bowler X are more reliable because…). **M1**
4. Using the key issues, analyse which method of data collection has provided you with the best quality of data. **D1**

Grading tips
- To attain P1 include a description of qualitative and quantitative research in your PowerPoint.
- To attain P2 and M1 make sure that you look at all of the different key issues when analysing your data set.
- To attain M1 and D1 try to provide as much detail in your presentation as you can and make sure that you link your arguments to the data set.

Functional skills

If you produce a presentation that is fit for purpose, you could provide evidence of your ICT skills in communicating information.

PLTS

By asking questions to extend your thinking about how the key issues in research methods relate to the data set, you can develop your skills as a critical thinker.
2. Data collection techniques for the sport and exercise sciences

2.1 Types of data

Primary data
Primary data is data that you collect through questionnaires, interviews and observations which you use to investigate your research problem.

Secondary data
Secondary data is previously published data found in books, journals, government publications, websites and other forms of media. Secondary data is used to form rationales for your research and to support or counter-argue your research findings.

2.2 Classifications of data

Discrete data
Discrete data is a form of data where only separate, isolated or opposite values can be achieved (for example, male/female, win/lose, yes/no).

Nominal data
A nominal scale is where participants are put into categories and counted, for example, grouping basketball players under the team they play for. You will group the players in this way to count them, not necessarily to say that one group is better than another.

Ordinal data
Ordinal data is ranked data that gives no indication of the difference between levels. It allows you to say who is best and second best, but does not tell you the difference between the two. This type of data provides the researcher with a rank order, but does not give an exact value. For example, on a badminton ladder, the person at the top is assigned a rank of 1, the person second down is awarded a rank of 2, the third person is awarded a rank of 3, and so on. There is nothing to say, however, that the person at the top of the ladder is three times as good as the person in third place on the ladder.

Continuous
Continuous data is data that can have any numerical value with any number of decimal places. For example, lap times in a Moto GP race can be classed as continuous data because of the values they are given (1 minute, 35.37 seconds).

Interval
Interval data is based on a scale that has equal intervals of measurement with equal intervals between each score. For example, in a figure skating scoring scale there is the same difference between scoring 5 and 5.5 as there is between scoring 5.5 and 6.

Ratio
Ratio data has proportional equal units of measurement. Ratio scales range from zero upwards and cannot have negative scores. For example, if a rugby team scores 40 points, it is worth twice as much as their opponents who have scored 20 points.

2.3 Qualitative data collection techniques

Three main types of data collection are involved with qualitative research:

- interviews
- focus groups
- observations.

Each type of data collection method has its advantages and disadvantages which we will discuss in this section.

Interviews
An interview is a conversation with a purpose. There are four types of interview: structured, unstructured, semi-structured and focus groups (focus groups will be covered separately as they are a group-based interview whereas the others are individual interviews).
Advantages and disadvantages of interviews

No one type of interview is ideal – it should be matched to the situation or participant. Interviews are used in qualitative research because they are a useful way for researchers to understand the beliefs, opinions and emotions of participants. They are useful because the researcher gets a view of what the participant thinks in the participant’s own words. This gives the researcher a greater understanding of the meanings that the participant attaches to their experiences of events. Some of the advantages and disadvantages of interviews and focus groups are summarised in the table below. Discuss these with your friends to develop your understanding.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants can express their views in their own words.</td>
<td>They require more resources and are more time-consuming than using questionnaires.</td>
</tr>
<tr>
<td>Participants can provide information from their own perspective.</td>
<td>They tend to use small sample sizes as interviews are time-consuming.</td>
</tr>
<tr>
<td>Unexpected data may come out in the interview.</td>
<td>The participant can take the interview off in a number of directions.</td>
</tr>
<tr>
<td>Body language, tone and pitch of voice, and speed of speech, can be assessed.</td>
<td>Data analysis is more difficult and takes longer than using questionnaires.</td>
</tr>
<tr>
<td>The researcher can establish a rapport with the participant and investigate target groups.</td>
<td>The quality of the data is dependent on the quality of the questioning and quality of responses.</td>
</tr>
</tbody>
</table>

Table 4.1: Advantages and disadvantages of interviews and focus groups.
The listening part of an interview is as important as the speaking part. A good interviewer knows when to keep quiet and listen and when to speak. Don’t interrupt the participant when they are speaking as this can prevent them from wanting to answer further questions.

Focus groups
Focus groups are similar to interviews, but involve more than one participant. There are usually between six and 12 participants and the researcher acts as a discussion facilitator rather than an interviewer. In this context, your role as the researcher is to ensure that the focus group stays on topic and doesn’t wander. Focus groups are more effective if everyone has a say in the discussion. They can provide you with a better quality of data because the discussion gets deeper as the group develops ideas. They are a good way of finding out opinions and ideas.

Observations
Two main types of observation are used in qualitative research: participant and non-participant.

- **Participant observation** means that the researcher is actively involved in the topic they are researching. For example, if you were studying team cohesion in rugby, you could join a rugby team, to observe ‘from the inside’ and gain your own experiences of cohesion as a player. Data would then be recorded in the form of field notes, with you recording your own thoughts, feelings, opinions, emotions and experiences. This method is useful when trying to discover the more delicate aspects of group behaviour that are not easy to see from the outside.

- **Non-participant observation** involves the researcher observing ‘from the outside’. There is no interaction with the individuals or the activity being observed. For example, if you wanted to look at injuries during a basketball match, you could watch how many injuries happened, what types of injuries they were and record the numbers on a data recording sheet.

Below there is a summary of the advantages and disadvantages of observational methods.

### Recording observational data
Observation checklists and field notes are methods of recording data in observational research. Observation checklists are used more frequently in quantitative research, or in qualitative research for observing more simple forms of data. For example, if you were observing a young developing footballer while researching two-footedness in football, you could use an observation checklist similar to the one in Figure 4.2.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations can be ‘here and now’ rather than being dependent on recall.</td>
<td>There is potential for the researcher to misunderstand what they are seeing.</td>
</tr>
<tr>
<td>They can take place in natural settings rather than research settings.</td>
<td>It can be difficult to identify and record the correct type of data.</td>
</tr>
<tr>
<td>They allow for the identification of behaviours that may not be apparent to</td>
<td>The Hawthorne effect: if the person knows they are the subject of research,</td>
</tr>
<tr>
<td>the person and may not have been discovered through interviews.</td>
<td>they may act differently and could invalidate the whole project – the</td>
</tr>
<tr>
<td></td>
<td>researcher must be very careful exactly how they approach the people in the</td>
</tr>
<tr>
<td></td>
<td>observational research.</td>
</tr>
<tr>
<td>They allow for the identification of behaviours that the person may not</td>
<td></td>
</tr>
<tr>
<td>wish to disclose.</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.2: Advantages and disadvantages of observational methods.*
Field notes are more commonly used in qualitative data collection than in quantitative data collection. They allow the researcher to record their observations. They are more flexible than observation checklists and allow you to collect more complex data. Field notes should include:

- descriptive notes that tell you about the setting of the research, the participants that are the subject of research and how they behave in a particular setting
- detailed notes to help you remember certain details over time
- reflective notes which should contain the researcher’s thoughts, opinions, beliefs, evaluations and experiences and form an integral part of the data analysis.

2.4 Quantitative data collection techniques

Several data collection techniques can be used for quantitative data collection. You have covered how non-participant observation can be used in both qualitative and quantitative research – don’t forget about this technique when considering quantitative research. Other techniques used in quantitative research include questionnaires. The settings in which data will be collected are either field-based data collection settings or laboratory-based data collection settings.

Questionnaires

Questionnaires are used when you are trying to collect a large amount of data from large groups and when the data you want to collect is not in-depth. If you need to obtain more in-depth information, questionnaires would not be suitable alone. However, they could be effective if used alongside other qualitative methods of data collection (such as interviews). As with other data collection methods, questionnaires have advantages and disadvantages.
Unit 4  Research methods for sport and exercise sciences

Questionnaire design

If your questionnaire looks poorly organised and unprofessional, it may be thrown away, particularly if you decide to post your questionnaires to people. If it looks well organised and purposeful, you have a better chance of it being completed. The use of coloured paper, artistic designs, dotted lines and tick boxes all help, but ensure your design is geared towards the audience it is aimed at. For example, make it easy and simple to use for young children. When designing your questionnaire, remember that if it is more than one page long it is much less likely to be filled in – so keep it short.

Always consider why you are asking a question. This will stop you including unnecessary questions. The quality of your questionnaire will increase as its validity increases. Decide which format would be most appropriate for the question you want to ask. Should it be an open question or a closed question? When you start to design your questionnaire, you need to consider a number of factors including:

- what you want to find out
- your sample (this will affect how you write your questionnaire)
- the length and appearance of your questionnaire (when you design it, don’t make it too long or difficult to answer)
- how and when you are going to distribute your questionnaire. If you are going to distribute it by hand, wait for it to be completed rather than going away and returning later. Another way you can distribute your questionnaire is by post or email, but this reduces the chances of it being returned. Include a return address and a covering letter to explain why your questionnaire is being sent out
- how to analyse the results.

Remember that there are different types of questions: open and closed.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>They are people-friendly if the form is designed correctly.</td>
<td>Questions can be too complex if the form is designed incorrectly.</td>
</tr>
<tr>
<td>They are an opportunity to reduce participant bias.</td>
<td>There are control issues.</td>
</tr>
<tr>
<td>The participant can be anonymous.</td>
<td>There is no opportunity for probing questions.</td>
</tr>
<tr>
<td>The data is structured.</td>
<td>There is a potential for a low response rate.</td>
</tr>
<tr>
<td>They are usually accessible to most people.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Advantages and disadvantages of questionnaires.

- **Open questions** are used more in qualitative research than in quantitative research. They allow people to express ideas, opinions and sentiments in words. They are used when asking questions that could lead to complex or in-depth answers, or if you are unsure of what the answers to the question could be. Open questions can take longer to answer than closed questions, so make sure that you plan your research to account for this. An example of an open ended question could be ‘What are your thoughts on the promotion of sport for people with disabilities within the UK?’

- **Closed questions** are used when a specific response is required and answers involve ranking, scales or categories. These questions are used more in quantitative data collection as they generate numbers for you to analyse using statistical methods. The participants respond to answers that the researcher has included on the questionnaire. The responses are in less depth than those from open questions. An example of a closed question could be ‘Do you like playing rugby? Yes/No’.

**Remember**

When designing your questionnaire, the following are important:

- Make sure that the first questions are straightforward and ask for facts.
- Do not put questions at the start that require lengthy answers.
- Leave personal or potentially sensitive questions to the end.
- Group questions together when they follow a similar theme or topic.
- Do not ask leading questions.
- Do not include questions that ask for responses on two different topics.
- Keep your questions simple and clear.
- Use an appropriate structure to make it attractive to your audience.
- Use tick boxes to make it easy to fill in.
- Do not make the questionnaire too long.
Laboratory-based data collection

Laboratory-based data collection involves collecting data in an environment where all the conditions and variables are controlled, so that you are only measuring the variables in question. One advantage of laboratory-based data collection is that it has high levels of internal validity. You are controlling all your variables so you know that you are only measuring the aspect you mean to measure. One disadvantage of laboratory-based data collection is that it has low levels of ecological validity because the data is not collected in an environment that reflects the situation in which the activity is performed. Another disadvantage of laboratory-based data collection is that it normally requires the use of expensive or technical equipment to collect data, making it difficult to use this if you don’t have a lot of resources.

Field-based data collection

Field-based data is collected in the environment that simulates the one in which the sport is played. One of the key strengths of field-based data collection is that it mimics the performance environment so you can claim ecological validity when you are collecting data in this setting. Field-based data collection can be cheaper than laboratory-based collection, making it more accessible to people without lots of resources. However, one limitation is that you don’t control all the variables in this data collection setting, so it can be difficult to claim internal validity.

2.5 Research designs

A number of research designs are used within sport and exercise sciences. A research design is the overall structure of your research. Some of the common designs that you need to understand are experimental research, cross-sectional research, case study research, longitudinal research and comparative research.

Experimental

The aim of experimental research is to look at the effects of an independent variable on a dependent variable. To use this research design effectively, you need to understand the terms independent and dependent variable. The independent variable affects the dependent variable. For example, an athletics coach wants to find out if her lower back flexibility training is benefiting the athlete’s high jump performance. She has asked you to research the topic for her. As the coach wants to find out if flexibility affects performance, flexibility is the independent variable and performance is the dependent variable.

Cross-sectional

Cross-sectional research involves using a range of participants with different backgrounds, ages and genders from the overall population. For example, if you want to look at preferences for team sports or individual sports in people in the UK, cross-sectional research would be useful. This would allow you to obtain opinions from a range of people.

Case study: Taking it to the max!

Jamie is a sport scientist going through his supervised experience at the youth team of a professional football club. It is pre-season and he is trying to assess the VO_2 max of some of the players. As he has a number of methods available to him, Jamie has decided to use both field-based and laboratory-based data collection methods to try and find out which is most effective. To measure VO_2 max in a field-based way, Jamie has decided to use a multi-stage fitness test to get an indirect prediction of VO_2 max, and to measure it in a laboratory setting, Jamie has decided to use a gas analysis system.

1. Which method do you think will be the most accurate and why?
2. Which method do you think will be most valid and why?
3. If you were Jamie, which do you think you would want to use and why?
You would send your participants a survey-type questionnaire that allowed them to say which type of sport they preferred. Then you could produce some descriptive statistics for the results of the study (for example, 73 per cent of men prefer team sports, 20 per cent of men prefer individual sports and 7 per cent of men have no preference).

Case study
Case study research is where you investigate a particular phenomenon (e.g. an individual or team) over a long period of time. It takes into account the development of the area of investigation over time and the environment in which the research resides. For example, to investigate the psychological effects of injury at different stages of injury and recovery, a case study design would be suitable. It allows you to investigate one person over a period of time and at different times throughout the stages of injury. This means you can draw conclusions relating to that individual and suggest these conclusions as directions for future research on a larger scale.

Longitudinal
Longitudinal research involves measuring the same variables over a long period of time and requires greater resources than other types of research, so be careful when approaching this design. Longitudinal research is useful if you want to examine the developmental characteristics of a group; for example, to investigate factors associated with talent development in a particular sport, longitudinal research would be a good option. It allows you to focus completely on developmental issues over an extended period of time.

Comparative
In comparative research, the researcher compares two or more things with the aim of discovering something about one or all of them. For example, if you wanted to see if there were any similarities between boys’ and girls’ opinions on hooliganism in football, you could use a comparative design.

Assessment activity 4.2
You are working as a sport and exercise scientist and you have been asked to do some promotional work with less experienced sport and exercise scientists to try to increase their interest in research.

In order to do this, produce an information leaflet to introduce people to the different types and classifications of data and the different methods of data collection in sport and exercise sciences.

Grading tip
To attain P3 try to make your leaflet as interesting and informative as possible by including lots of examples to demonstrate your points.

Functional skills
By writing your leaflet on types of data and data collection methods you could provide evidence of your English skills in writing.

PLTS
If you explore the different types of data and data collection methods and make judgements about how they could be applied to different examples, you can develop your skills as an independent enquirer.
2.6 Ethical and legal issues

The British Association of Sport and Exercise Sciences (BASES)

BASES produced a code of conduct that governs how sport and exercise scientists work as practitioners and as researchers. The code of conduct outlines ethical and legal issues that are essential to safe research within the sport and exercise sciences.

- **Ethical clearance** – when conducting research, ensure that you are working ethically and legally. One of the first things you have to do before starting is to gain ethical clearance from an appropriate body. If you conduct any research as part of your course, ethical clearance will come from your tutor, college or school ethics committee. An ethics committee is an organisation that looks at your research proposal and says whether it is safe and ethical. It will confirm whether you can start work on your project.

- **Informed consent** – once you have gained ethical approval for your research project, you need to get informed consent from your participants. This is an ethical and legal requirement of research. It can be verbal but it is safer for both you and your participant if you obtain it in writing. An informed consent form consists of:
  - a description of the investigation
  - details of the procedure to be followed
  - details of any risks to the participant
  - details of the potential benefits of taking part in the research
  - a section that offers to answer any questions and confirms that any questions asked have been answered fully
  - an indication that the participant can withdraw at any time without being penalised
  - a section that explains that any information collected about the participant will remain confidential.

- **Confidentiality/data protection** – where confidentiality is concerned, you may disclose only information that is important to the study you are conducting. Any data you collect is protected under the terms of the Data Protection Act (1998). No data that makes the participants personally identifiable should be included in your research project. Data collected should be stored in a locked filing cabinet or on a password protected computer accessible only by you and your research supervisor.

- **Safety of the participants** – when conducting research, the key concern is the participants’ safety. The researcher must maintain the highest professional standards so as not to endanger participants or themselves. The researcher should treat all participants equally and only work within their own area of competence.

- **Acting with due regard for equality and impartiality** – to preserve the reputation of sport and exercise science, you must remain totally unbiased in your actions and practices when working within sport and exercise science. This means that you cannot let factors such as race, age or gender affect the way you work with clients. You must not exploit personal relationships for personal gain. Any decisions you make must be completely objective (based on facts rather than on opinions).

### Importance of ethical and legal issues

- **Ensure the welfare and safety of participants and the researcher** – the ethical and legal guidelines are in place to maintain the safety of the participants and the researcher.

- **Ensure that researchers only work within area of expertise** – if you think about the dangers that participants could be placed in through some areas of research, you will understand why the ethical and legal guidelines need to be in place so that researchers work only within their own areas of competence. For example, someone who has never used a syringe must not be allowed to take blood from participants.

- **Preserving and developing the reputation of the sport and exercise sciences** – the guidelines exist to promote excellence in sport and exercise sciences. This means that someone who works in a research capacity must maintain the reputation of sport and exercise sciences by following the procedures.

### Implications of not working ethically and legally

The implications of not working within the ethical and legal procedures are severe. The most obvious is that you are risking your participants’ welfare and safety. If you do this in a professional setting, restrictions can be placed on you.
Figure 4.3: An example of an informed consent form.

- **Tribunals** – if you do not work within ethical and legal guidelines, you could be subject to a tribunal leading to a fine, a written warning or not being allowed to conduct research again.

- **Legal or civil action** – if you work without due regard for ethical and legal issues within sport and exercise science, you are leaving yourself open to legal or civil action if something goes wrong. Ensure you follow the code of conduct when working in an applied setting, so you are not open to this type of action.

- **Measures to stop future research** – when working as a member of BASES, you are subject to the BASES disciplinary procedure if you do not work ethically and legally. You could be called before a tribunal where a group of BASES officers and the chair of BASES will review your particular case to discover if you have been working ethically and legally. From this tribunal, significant measures can be put in place to stop research temporarily or permanently, such as temporary suspension or permanent expulsion from BASES.
Assessment activity 4.3

You are working as a trainee sport scientist. You are preparing to do some research into the developmental experiences of female youth athletes. Your research will involve you interviewing and observing the female athletes over an extended period of time so your supervisor wants you to produce a presentation on the ethical and legal issues associated with the work you are doing. Prepare a 10-minute presentation that includes the following information:

1. A description of two ethical and legal issues associated with the research project you are going to be doing. P

2. An explanation of the implications of not working ethically and legally in a research setting. M

3. An analysis of the implications of not working ethically and legally in a research setting. D

Grading tips

- To attain P4, before starting your presentation, download and read a copy of the BASES Code of Conduct and decide what you think are the two most important ethical and legal issues for this project. Describe these two issues as the basis of your presentation.
- To attain M2, provide examples of things that you would class as not working ethically and legally in this setting and discuss what could happen to you if you were to do any of these things.
- To attain D2, for the same examples, think about, and then communicate, why you would be subject to different implications of not working ethically and legally. Support your arguments with specific sections of the BASES Code of Conduct.

PLTS

If you support your arguments relating to which ethical and legal issues are most strongly linked to the case study, you can develop your skills as an independent enquirer.

Functional skills

If you interpret the information in the BASES Code of Conduct for relevance against the case study, you could provide evidence of your English skills in reading.

3. Qualitative data analysis techniques for the sport and exercise sciences

Before analysing your qualitative data, you must prepare it for analysis. In the case of individual interviews and focus groups, you should transcribe the recorded interviews verbatim (word for word).

3.1 Stages of data analysis

After transcribing your data, you will go through three stages of qualitative data analysis:

1. data reduction
2. displaying data
3. drawing conclusions and verifying data.

Data reduction

Data reduction involves reducing large amounts of data into manageable chunks. The most common form of data analysis in the data reduction stage is coding. If you choose to conduct a qualitative research project, you need to code your data as part of your data analysis. Coding is when you organise raw data (sentences, phrases or words from your questionnaires or interviews) into categories. The categories are given a valid heading and must have a rule for inclusion.
Having a rule for inclusion helps to guide which data you place in each category. For example, if you were researching ‘factors affecting talent development in football’, you could have a category called ‘importance of parental tangible support’. Your rule for inclusion could be ‘statement made refers to concrete support given to player from parent (for example, the purchase of playing kit or transport to matches) being either a positive or negative influence on the player’s development’.

**Coding**

Coding techniques are simple when you get used to them. When coding data, all you are doing is breaking it down into smaller parts. You then put it back together in parts that relate to each other, before making sure that all your categories are valid. Coding involves the line-by-line analysis of data in minute detail and is used to generate categories.

Coding involves three stages: open coding, axial coding and selective coding.

- **Open coding** – in open coding, data is broken down and examined. Your aim is to identify all the key statements in the interviews that relate to the aims of your research and your research problem. After identifying the key statements, you can start to put the key points that relate to each other into categories, but you need to give each category a suitable heading to do this. When you start to organise your data under different categories, you have started the coding process.

- **Axial coding** – after the open coding stage, the next stage is to put the data back together. Part of this process means re-reading the data you have collected so you can make precise explanations about your area of interest. To do this, you need to refine the categories that you started to create during open coding. During this stage you may develop new categories. To allow you to refine your codes at this stage ask more questions about the categories (and the codes) you have created. Some questions you may consider are:
  - Can I relate certain codes together under a more general code?
  - Can I place codes into a particular order?
  - Can I identify any relationships between different codes?

- **Selective coding** is the final stage of coding. It involves aiming to finalise your categories (and codes) so that you can group them together. When you group them together, you will produce different diagrams to show how your categories link together. The key part of this is to select a main category, which will form the focal point of your diagram. You also need to look for data that contradicts previous research, rather than data that supports it. This helps you to make better arguments and draw more conclusions based on your data.

**Remember**

Sometimes you may not go all the way through to selective coding. In smaller research projects, and projects where you are not necessarily trying to produce a theory or a model, you may find that open and axial coding provide sufficient data analysis.

**Other techniques**

As well as being able to manually analyse your data, there are electronic packages that you can use to analyse data (although in the early stages of your research career it is unlikely that you will use these and it is better to manually analyse your data so that you get used to handling the data). Some of the different packages available include ATLAS/ti and NVivo 7.

**Take it further**

**Research electronic packages**

Using the Internet, research the different electronic packages that are available for qualitative data analysis and find out what their advantages and disadvantages are.

**Displaying data**

There are different ways to display your data. The way that you display it will affect the argument or point you are trying to make. The different types of diagram used are network diagrams, Venn diagrams, radial diagrams and cycle diagrams.
Network diagrams

Network diagrams show hierarchical relationships between different ideas. The example below shows that there are a number of benefits to the use of imagery (the top of the hierarchy or the most important part of information to take away) and that these benefits include increased self-efficacy, skill acquisition and injury rehabilitation.

Venn diagrams

Venn diagrams consist of two or more overlapping circles. They show how different topics relate to each other. In the example below, you can see how the different disciplines within the sport and exercise sciences interact to make up the overall discipline.

Radial diagrams

A radial diagram (also known as a spider diagram) illustrates a relationship where each item is linked to a central item. This diagram can be thought of as a simple organisation chart that starts from the centre rather than the top. A radial diagram flows from the outside in or the inside out.

Drawing conclusions and verifying data

Your conclusions must be valid and reliable. Two common techniques used to do this are triangulation and member checking.
**Triangulation** refers to using different data collection methods in the same study. For example, you could use interviews and questionnaires or you could use the same interviews with different types of participants (such as athletes and coaches). Alternatively, you could ask different researchers to collect data and independently draw conclusions before checking their findings with each other.

**Member checking** – during member checking, you complete your data analysis and draw conclusions relating to the aims of the study. You then show the analysis to the participants who took part in the research so that they can check that you have understood and communicated everything correctly. If they decide that your analysis is correct, you can claim that the data is valid.

**Activity: Analyse this**

Using the interview that you completed earlier in the unit, transcribe and analyse the data that you collected so that you can produce a summary of the findings. Use the different stages of data analysis to analyse your data.

---

**Assessment activity 4.4**

You are working as a sports therapist and you want to find out about how different athletes respond psychologically to chronic sports injuries that mean they cannot play or train for a long period of time. Ideally, you would like to study individual athletes, over an extended period of time, with a variety of age ranges and experiences.

Before you can do the research, you need to complete a discussion with your supervisor that includes the following information:

1. **The best research design to use and a justification of your choice.**
2. **The best data collection technique (or techniques) to use and a justification of your choice.**
3. **A description of the different stages of data analysis.**
4. **The best way of analysing your data and a justification of your choice.**

**Grading tips**

- To attain P5 describe the data reduction, displaying data, and drawing conclusions and verifying data stages.
- To attain M3 try to relate the research methods that you select to specific aspects of the project that you want to complete as this will allow you to make a stronger justification. If you think it is more appropriate to use a combination of different research methods, make sure that you justify each method.

---

**PLTS**

By presenting a persuasive case for your choices of research methods, you could develop your skills as a reflective learner.

**Functional skills**

By discussing your ideas for the most appropriate research methods, you could provide evidence of your English skills in speaking and listening.
### 4. Quantitative data analysis techniques for the sport and exercise sciences

#### 4.1 Organising data

There are different methods of organising your data during quantitative data analysis, each of which provides a good starting point to the appropriate research project. The methods include range, rank order distribution, simple frequency distribution and grouped frequency distribution.

**Range**

Range is the distance in numerical value from the highest to the lowest value collected. You calculate the range by subtracting the lowest value from the highest value.

**Rank order distribution**

Rank order distribution means placing your data into an ordered list from the lowest to the highest in a single column, ensuring you include all the scores. Rank order distribution is used when the number of participants is less than or equal to 20 (n ≤ 20).

**Simple frequency distribution**

Simple frequency distribution is used when the number of participants is greater than 20 (n > 20) and when the range is less than or equal to 20 (r ≤ 20). You use simple frequency distribution with a table that has two columns, one for raw data scores (X) and one for frequency scores (f). The frequency column is the number of times that particular score was achieved.

Here is an example of how to lay out your data. A basketball coach is looking at the number of free throws missed in each game over a season. He has 25 games to assess (n > 20) and the number of missed shots per game ranges from 1 to 7 (r ≤ 20), so simple frequency distribution is suitable. The data is set out as shown below.

<table>
<thead>
<tr>
<th>Number of missed shots (X)</th>
<th>Frequency (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

| n = 25                      |

**Grouped frequency distribution**

In quantitative research, you often work with ranges greater than 21 (r > 21) and with more than 20 participants (n > 20). This is when to use grouped frequency distribution. As with simple frequency distribution, the table has two columns: X and f – except this time the X column is for groups of scores and the f column is for frequency.

To keep your data on a single sheet of paper, you normally have between 10 and 20 groups of scores – the ideal number is 15. You need to decide on the interval size for each group, which is calculated using the formula \( i = \text{range} ÷ 15 \).

Here is an example: an athletics coach is looking at the times recorded (in seconds) of athletes who want to represent the college at 5000 m. She has 30 times to look at ranging from 900 seconds to 1094 seconds. Grouped frequency distribution is a suitable method because both \( r > 21 \) and \( n > 20 \). The interval size for each group is 13 seconds \( (r = 194 \text{ seconds}; 193 ÷ 15 = 12.93 \text{ seconds}, \text{which is rounded up to 13}) \). The data is shown in the table below.

<table>
<thead>
<tr>
<th>Time (X)</th>
<th>Frequency (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1082–1094</td>
<td>1</td>
</tr>
<tr>
<td>1068–1081</td>
<td>1</td>
</tr>
<tr>
<td>1054–1067</td>
<td>1</td>
</tr>
<tr>
<td>1040–1053</td>
<td>1</td>
</tr>
<tr>
<td>1026–1039</td>
<td>5</td>
</tr>
<tr>
<td>1012–1025</td>
<td>8</td>
</tr>
<tr>
<td>998–1011</td>
<td>3</td>
</tr>
<tr>
<td>984–997</td>
<td>2</td>
</tr>
<tr>
<td>970–983</td>
<td>2</td>
</tr>
<tr>
<td>956–969</td>
<td>1</td>
</tr>
<tr>
<td>942–955</td>
<td>1</td>
</tr>
<tr>
<td>928–941</td>
<td>1</td>
</tr>
<tr>
<td>914–927</td>
<td>1</td>
</tr>
<tr>
<td>900–913</td>
<td>2</td>
</tr>
<tr>
<td>( n = 30 )</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: Why is simple frequency distribution used here?

Table 4.5: Why is grouped frequency distribution used here?
Although using grouped frequency distribution is a useful way of organising large amounts of data, some information is lost through this process. Once scores have been placed into groups, it is impossible to know the individual values. For example, if you look at the 1012–1025 seconds row in the table above, it is only possible to identify that eight athletes fell within that range, but you won’t know what the individual times were.

### 4.2 Displaying data

There are different ways of displaying your data including graphs, histograms, bar charts and cumulative frequency graphs. However, before conducting statistical analysis you must understand distribution curves. There are three types: normal distribution curves, positively skewed curves and negatively skewed curves.

#### Take it further

**Look for examples**

Using books and the Internet, find examples of graphs, histograms, bar charts and cumulative frequency graphs and research when you may want to use each of these. You may find websites such as http://openlearn.open.ac.uk/ useful.

A normal distribution of data means that most of the examples in a set of data are close to the ‘average’, while a few examples are at one extreme or the other. Normal distribution graphs have these characteristics:

- the curve has a single peak
- it is bell-shaped
- the mean (average) lies at the centre of the distribution and the distribution is symmetrical around the mean
- the two ‘tails’ of the distribution extend indefinitely and never touch the x axis
- the shape of the distribution is determined by the mean and standard deviation (see page 22).

Not all sets of data have graphs that look this perfect. Some have relatively flat curves, others will be steeper. Sometimes the mean will lean a little bit to one side or the other. However, all normally distributed data will have something like this bell-shaped curve. Generally,

if you go right or left one standard deviation from the mean (the red area on the graph) you will include about 68 per cent of the scores in the distribution. Two standard deviations away from the mean (the red and green areas) account for about 95 per cent of the scores, whereas three standard deviations (the red, green and blue areas) account for about 99 per cent of the scores.

![Figure 4.8: What should you look for when trying to identify normal distribution?](image)

#### Remember

Standard deviation tells you how tightly all the various examples are clustered around the mean in a set of data. When the examples are tightly bunched together and the bell-shaped curve is steep, the standard deviation is small. When the examples are spread apart and the bell curve is flat, this tells you that you have a relatively large standard deviation.

#### Positively skewed curves and negatively skewed curves

If the shape of the curve is asymmetrical, your data is not distributed normally and is said to be positively or negatively skewed. Positively skewed means the longer tail of the curve points to the positive (higher) end of the scale and the scores are bunched to the left of the centre. Negatively skewed means the longer tail of the curve points to the negative (lower) end of the scale and the scores are bunched to the right of the centre.
4.3 Measures of central tendency and variability

Measures of central tendency are numbers that describe what is average or typical of the distribution. These measures include the mean, median and mode (measures of central tendency) and standard deviation (measure of variability).

Key terms

Mean – the measure of central tendency that is calculated by adding up all of the values and dividing the answer by the number of values. You may also know this term as the ‘average’.

Median – the middle value in a series of numbers.

Mode – the value that occurs most frequently.

Identification of outliers

Another important concept within central tendency and variability is that of outliers. Outliers are results that are radically different from what you would consider normal scores. Statistics that are drawn from data sets that contain outliers can be misleading.

Standard deviation

Standard deviation is a number that indicates how much each of the values in the distribution deviates from the mean (or centre) of the distribution. If the data points are all close to the mean, then the standard deviation is close to zero. If many data points are far from the mean, then the standard deviation is far from zero. If all the data values are equal, then the standard deviation is zero.

The formula for calculating standard deviation (sd) is as follows:

\[ sd = \sqrt{\frac{\sum (X - M)^2}{n - 1}} \]

where:

- \( sd \) = standard deviation
- \( \sum \) = sum of
- \( X \) = individual score
- \( M \) = mean
- \( n \) = number of participants

Here is how to calculate standard deviation:

- Calculate the mean.
- Subtract the mean from each subject’s score \((X - M)\).
- Square the answer \((X - M)^2\).
- Sum the squared scores \(\sum (X - M)^2\).
- Divide by the number of participants minus 1 \((n - 1)\).
- Take the square root of the answer.

4.4 Data analysis

Commonly, data analysis in quantitative research comes through the use of statistical tests. Statistics can seem frightening but they are more time-consuming than difficult, and you will have seen and used some of these methods before. There are two types of statistics: descriptive and inferential. You covered descriptive statistics earlier when you looked at central tendency and variability. Inferential statistics assess relationships or differences between data sets; and are further subdivided into two groups: parametric tests and non-parametric tests (see Figure 4.10).

Tests – selecting, types and explanations

Throughout this next section, you will look at a range of parametric and non-parametric tests, an explanation of each one and examples of different tests, and ways of selecting appropriate tests.

A good way to select your test is to use a decision tree like the one in Figure 4.11. If you follow the decision tree using the information available, you will find the test that you need to use. The process of using the decision tree is similar to planning a bus or a train journey – follow the line and find your stops to get to the destination!
Unit 4 Research methods for sport and exercise sciences

Quantitative data analysis

Inferential statistics
- Parametric tests
  - ANOVA
  - t-tests
  - MANOVA
    - Pearson product moment correlation coefficient (r)
- Non-parametric tests
  - Wilcoxon matched pairs signed ranks test
  - Chi squared
  - Mann Whitney U
  - Spearman rank-order correlation

Descriptive statistics
- Central tendency
  - Mean
  - Median
  - Mode
- Variability
  - Standard deviation

Key
- Tests of difference
- Tests of association

Figure 4.10: Why do you think there are so many statistical tests in sport and exercise sciences?

Key
- Nominal
  - Relationships between rankings
    - Differences between unrelated rankings (two groups)
      - Mann Whitney U
  - Differences among frequency counts
    - Chi squared
- Interval or ratio
  - Analyse mean differences
    - Pearson product moment correlation coefficient
- Ordinal
  - Relationships
    - Spearman rank-order correlation
  - Two separate groups of data
    - Dependent t-test
  - Two related groups of data
    - Independent t-test

Figure 4.11: Use the decision tree to find the appropriate statistical test for your work.
Parametric tests – t-tests

The most common t-tests are the dependent t-test (also known as the paired samples t-test) and the independent t-test.

When you complete your t-test and want to see if your result is significant or not, you need to know whether you are completing a one-tailed test or a two-tailed test.

Dependent t-test

The dependent (paired samples) t-test examines significant differences between two sets of related scores, such as whether the mean high jump scores of one group are different when measured pre- and post-training (see the example below). The test is calculated using the formula below:

\[ t = \frac{\sum D}{\sqrt{\frac{\sum D^2 - (\sum D)^2}{n(n-1)}}} \]

where:
- \( D \) = difference between before and after
- \( n \) = number of paired scores
- \( \sum \) = sum of

Instructions:
- Calculate your \( t \) value using the formula.
- Calculate your degree of freedom (df). For the dependent t-test, use the formula \( df = n - 1 \).
- Compare your \( t \) value to the table of critical values.
- After calculating your \( t \) value, compare your result to the critical value of \( t \) (use Table 4.7). Find your \( df \) value (in this case 9), then go across and see if your result is greater than or equal to the number in the column below the 0.05 level. If the value achieved for your \( t \)-test is equal to or greater than the number shown, your results are significant to that level. Note that if \( df > 120 \), use the infinity column at the end of Table 4.7 (\( \infty \)).

As you can see from the example below the \( t \) value calculated (5.09) is greater than the critical value of \( t \) (2.262) meaning your result is significant to the 0.05 level. This means that you can say that there is a significant difference between high jump scores pre- and post-training.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pre-training height (cm)</th>
<th>Post-training height (cm)</th>
<th>D (post-training minus pre-training)</th>
<th>( D^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>176</td>
<td>179</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>169</td>
<td>172</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>171</td>
<td>175</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>173</td>
<td>177</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>164</td>
<td>166</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>170</td>
<td>171</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>161</td>
<td>168</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>159</td>
<td>169</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>163</td>
<td>166</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>170</td>
<td>176</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>n = 10</td>
<td></td>
<td></td>
<td>( \sum D = 43 )</td>
<td>( \sum D^2 = 249 )</td>
</tr>
</tbody>
</table>

Table 4.6: Investigating the effects of a 12-week plyometric training programme on high jump performance.

Key terms

Parametric tests – statistical tests that use interval or ratio data. They assume that the data is drawn from a normal distribution and has the same variance.

Non-parametric tests – statistical tests that use ordinal or nominal data.

One-tailed test – a test that assumes one group will be better than the other, or at least no worse than the other. For example, girls will be better than boys.

Two-tailed test – a test that assumes there will be a difference between both groups, but doesn’t say which will be better. For example, there will be a difference between girls and boys.

Degree of freedom – used as a correction factor for bias and to limit the effects of outliers, and based on the number of participants you have.
### Level of significance for one-tailed test

<table>
<thead>
<tr>
<th>df</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
<th>.005</th>
<th>.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>636.619</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
<td>6.965</td>
<td>9.295</td>
<td>31.598</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>12.941</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>8.610</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>6.589</td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>5.405</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>4.781</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>4.587</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>4.437</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>4.318</td>
</tr>
<tr>
<td>13</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>4.221</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>4.140</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>4.073</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
<td>2.120</td>
<td>2.583</td>
<td>2.921</td>
<td>4.015</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.740</td>
<td>2.110</td>
<td>2.567</td>
<td>2.898</td>
<td>3.965</td>
</tr>
<tr>
<td>18</td>
<td>1.330</td>
<td>1.734</td>
<td>2.101</td>
<td>2.552</td>
<td>2.878</td>
<td>3.922</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.539</td>
<td>2.861</td>
<td>3.883</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.528</td>
<td>2.845</td>
<td>3.850</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.518</td>
<td>2.831</td>
<td>3.819</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td>2.819</td>
<td>3.792</td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.500</td>
<td>2.807</td>
<td>3.767</td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
<td>3.745</td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td>2.787</td>
<td>3.725</td>
</tr>
<tr>
<td>26</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
<td>3.707</td>
</tr>
<tr>
<td>27</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.690</td>
</tr>
<tr>
<td>28</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.674</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.462</td>
<td>2.756</td>
<td>3.659</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
<td>3.646</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
<td>3.551</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
<td>3.460</td>
</tr>
<tr>
<td>120</td>
<td>1.289</td>
<td>1.658</td>
<td>1.980</td>
<td>2.358</td>
<td>2.617</td>
<td>3.373</td>
</tr>
<tr>
<td>∞</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
<td>3.291</td>
</tr>
</tbody>
</table>

**Table 4.7: Critical values of t**

**Remember**

When calculating statistics, a good tip is to change all the letters in the formula into the numbers you will be working with, and then rewrite the equation using the numbers – it will seem much friendlier then!
Independent t-test
The independent t-test is the most frequently used t-test. It is used when you have two groups and are trying to discover if the mean scores of two groups can be considered to be significantly different. The independent t-test is suitable when the data you have collected is interval or ratio data, when your groups are randomly assigned, and when the variance (or spread) in the two groups is equal. The independent t-test is calculated using the formula:

\[ t = \frac{M_1 - M_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \]

where:
- \( M_1 \) = mean value of group 1
- \( M_2 \) = mean value of group 2
- \( s_1 \) = standard deviation of group 1
- \( s_2 \) = standard deviation of group 2
- \( n_1 \) = number of participants in group 1
- \( n_2 \) = number of participants in group 2

Parametric test – Pearson product moment correlation coefficient (r)
A correlation is the value of the relationship between two or more variables, which can be positive or negative. Whether it is positive or negative depends on the direction of the line when the results are plotted on a graph. The graphs in Figure 4.12 show examples of perfect positive and perfect negative correlations, but it is rare to record such correlations during data analysis.

Example of an independent t-test
Using the Cooper 12-minute run data in Table 4.8, and the independent t-test formula, see if there is a significant difference between the two groups. Calculate the degrees of freedom (df) using the formula:

\[ df = n_1 + n_2 - 2 \]

and then compare the t value calculated to the table of critical values, Table 4.7.

Where:
- \( s_1 = 238.3 \)
- \( s_2 = 94.6 \)
- \( M_1 = 3183.3 \)
- \( M_2 = 2468.7 \)
- \( s_1^2 = 56786.89 \)
- \( s_2^2 = 8949.16 \)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Group 1 (12-minute run after 70% VO(_2) max training)</th>
<th>Group 2 (12-minute run after 40% VO(_2) max training)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3200 m</td>
<td>2513 m</td>
</tr>
<tr>
<td>2</td>
<td>3600 m</td>
<td>2601 m</td>
</tr>
<tr>
<td>3</td>
<td>2894 m</td>
<td>2444 m</td>
</tr>
<tr>
<td>4</td>
<td>3001 m</td>
<td>2361 m</td>
</tr>
<tr>
<td>5</td>
<td>3187 m</td>
<td>2541 m</td>
</tr>
<tr>
<td>6</td>
<td>3651 m</td>
<td>2486 m</td>
</tr>
<tr>
<td>7</td>
<td>3109 m</td>
<td>2611 m</td>
</tr>
<tr>
<td>8</td>
<td>2997 m</td>
<td>2419 m</td>
</tr>
<tr>
<td>9</td>
<td>3056 m</td>
<td>2400 m</td>
</tr>
<tr>
<td>10</td>
<td>3138 m</td>
<td>2311 m</td>
</tr>
<tr>
<td>Mean</td>
<td>3183.3</td>
<td>2468.7</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>238.3</td>
<td>94.6</td>
</tr>
</tbody>
</table>

Table 4.8: Cooper 12-minute run data.
The Pearson product moment correlation coefficient is a parametric test that is suitable when you have interval or ratio data and you are trying to identify a relationship between two variables. It is a test of association, which means it looks at whether two or more variables are related. The test can be used in two ways. Either you can try to find out a relationship between two variables or you can try to predict one score from another. In a simple correlation which is trying to find out a relationship between two variables, it doesn’t matter which variable is assigned $X$ and which $Y$. If you are trying to predict one score from another, then $X$ is the independent variable and $Y$ is the dependent variable. There are three stages to using the Pearson product moment correlation:

1. Summing each set of scores.
2. Squaring and summing each set of scores.
3. Multiplying each pair of scores and obtaining the cumulative sum of these products.

The formula for this is outlined below:

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2} \sqrt{n\sum Y^2 - (\sum Y)^2}}$$

where:

- $n$ = number of paired scores
- $\sum$ = sum of
- $X$ = scores for one variable
- $Y$ = scores for the other variable
- $\sum X^2$ = sum of raw scores for $X$, squared
- $\sum Y^2$ = sum of raw scores for $Y$, squared
- $\sum X^2$ = sum of all of the $X^2$ scores
- $\sum Y^2$ = sum of all of the $Y^2$ scores

To interpret the significance of your $r$ value, select your level of significance (remember that in sport and exercise science this is normally 0.05) and find your degree of freedom ($df$) for your test. For this test, use the formula $df = n - 2$ and compare your $r$ value to the table of significance (Table 4.7 on page 25) to find whether your results are significant. If your result is equal to or greater than the critical value in the table, your result is significant.

Non-parametric tests

If the data is non-parametric, t-tests cannot be used. In this case, the Wilcoxon matched pairs signed ranks test is used in place of the dependent t-test, and the Mann Whitney U test is used in place of the independent t-test.

### Activity: Pearson product moment correlation coefficient ($r$)

Get together with a group of ten friends. Measure each other’s vertical jump score and then measure each other’s 30 sprint times. Use the Pearson product moment correlation coefficient to find out if there is a relationship between the two variables. Use the table of critical values below (Table 4.9) to assess the strength of the relationship.

<table>
<thead>
<tr>
<th>Level of significance for one-tailed test</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
<th>.005</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>.10</td>
<td>.05</td>
<td>.02</td>
<td>.01</td>
<td>.001</td>
</tr>
<tr>
<td>1</td>
<td>.9877</td>
<td>.9969</td>
<td>.9995</td>
<td>.9999</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>.9000</td>
<td>.9500</td>
<td>.9800</td>
<td>.9900</td>
<td>.9990</td>
</tr>
<tr>
<td>3</td>
<td>.8054</td>
<td>.8783</td>
<td>.9343</td>
<td>.9587</td>
<td>.9912</td>
</tr>
<tr>
<td>4</td>
<td>.7293</td>
<td>.8114</td>
<td>.8822</td>
<td>.9172</td>
<td>.9741</td>
</tr>
<tr>
<td>5</td>
<td>.6694</td>
<td>.7545</td>
<td>.8329</td>
<td>.8745</td>
<td>.9507</td>
</tr>
<tr>
<td>6</td>
<td>.6215</td>
<td>.7067</td>
<td>.7887</td>
<td>.8343</td>
<td>.9249</td>
</tr>
<tr>
<td>7</td>
<td>.5822</td>
<td>.6664</td>
<td>.7498</td>
<td>.7977</td>
<td>.8982</td>
</tr>
<tr>
<td>8</td>
<td>.5494</td>
<td>.6319</td>
<td>.7155</td>
<td>.7646</td>
<td>.8721</td>
</tr>
<tr>
<td>9</td>
<td>.5214</td>
<td>.6021</td>
<td>.6851</td>
<td>.7348</td>
<td>.8471</td>
</tr>
<tr>
<td>10</td>
<td>.4973</td>
<td>.5760</td>
<td>.6581</td>
<td>.7079</td>
<td>.8233</td>
</tr>
<tr>
<td>11</td>
<td>.4762</td>
<td>.5529</td>
<td>.6339</td>
<td>.6835</td>
<td>.8010</td>
</tr>
<tr>
<td>12</td>
<td>.4575</td>
<td>.5324</td>
<td>.6120</td>
<td>.6614</td>
<td>.7800</td>
</tr>
<tr>
<td>13</td>
<td>.4409</td>
<td>.5139</td>
<td>.5923</td>
<td>.6411</td>
<td>.7603</td>
</tr>
<tr>
<td>14</td>
<td>.4259</td>
<td>.4973</td>
<td>.5742</td>
<td>.6226</td>
<td>.7420</td>
</tr>
<tr>
<td>15</td>
<td>.4124</td>
<td>.4821</td>
<td>.5577</td>
<td>.6055</td>
<td>.7246</td>
</tr>
<tr>
<td>16</td>
<td>.4000</td>
<td>.4683</td>
<td>.5425</td>
<td>.5897</td>
<td>.7084</td>
</tr>
<tr>
<td>17</td>
<td>.3887</td>
<td>.4555</td>
<td>.5285</td>
<td>.5751</td>
<td>.6932</td>
</tr>
<tr>
<td>18</td>
<td>.3783</td>
<td>.4438</td>
<td>.5155</td>
<td>.5614</td>
<td>.6787</td>
</tr>
<tr>
<td>19</td>
<td>.3687</td>
<td>.4329</td>
<td>.5034</td>
<td>.5487</td>
<td>.6652</td>
</tr>
<tr>
<td>20</td>
<td>.3598</td>
<td>.4227</td>
<td>.4921</td>
<td>.5368</td>
<td>.6524</td>
</tr>
<tr>
<td>25</td>
<td>.3233</td>
<td>.3809</td>
<td>.4451</td>
<td>.4869</td>
<td>.5974</td>
</tr>
<tr>
<td>30</td>
<td>.2960</td>
<td>.3494</td>
<td>.4093</td>
<td>.4487</td>
<td>.5541</td>
</tr>
<tr>
<td>35</td>
<td>.2746</td>
<td>.3246</td>
<td>.3810</td>
<td>.4182</td>
<td>.5189</td>
</tr>
<tr>
<td>40</td>
<td>.2573</td>
<td>.3044</td>
<td>.3578</td>
<td>.3932</td>
<td>.4896</td>
</tr>
<tr>
<td>45</td>
<td>.2428</td>
<td>.2875</td>
<td>.3384</td>
<td>.3721</td>
<td>.4648</td>
</tr>
<tr>
<td>50</td>
<td>.2306</td>
<td>.2732</td>
<td>.3218</td>
<td>.3541</td>
<td>.4433</td>
</tr>
<tr>
<td>50</td>
<td>.2108</td>
<td>.2500</td>
<td>.2948</td>
<td>.3248</td>
<td>.4078</td>
</tr>
<tr>
<td>70</td>
<td>.1954</td>
<td>.2319</td>
<td>.2737</td>
<td>.3017</td>
<td>.3799</td>
</tr>
<tr>
<td>80</td>
<td>.1829</td>
<td>.2172</td>
<td>.2565</td>
<td>.2830</td>
<td>.3568</td>
</tr>
<tr>
<td>90</td>
<td>.1726</td>
<td>.2050</td>
<td>.2422</td>
<td>.2673</td>
<td>.3375</td>
</tr>
<tr>
<td>100</td>
<td>.1638</td>
<td>.1946</td>
<td>.2301</td>
<td>.2540</td>
<td>.3211</td>
</tr>
</tbody>
</table>

Table 4.9: Critical values of correlation coefficient.
Non-parametric test – Wilcoxon matched pairs signed ranks test

The Wilcoxon matched pairs signed ranks test is used when you are trying to find out if there is a significant difference between two scores that are taken from the same participant (or from matched participants). It is used when the data is ordinal (ranked). To do the test, follow the instructions below.

1. Disregard any results for participants who scored the same in both conditions, then count up the number of paired scores left. This is your n score.
2. Calculate the difference between the two scores of each participant, assigning plus or minus signs (d).
3. Rank the differences, giving the smallest a rank of 1 (ignoring plus or minus signs, i.e. +2 is of the same value as –2). When two scores are tied, each is given the mean of the two ranks and the next rank is missed out (for example, if two participants are in level sixth place, they are both given the rank of 6.5 and the next place is given a rank of 8).
4. Add up the ranks of all the minus scores.
5. Add up the ranks of all the plus scores.
6. Take the smaller of the two figures calculated in points 4 and 5 to gain your w value.
7. Look up your value for w in a significance table (you can find one here: www.social-science.co.uk. If it is equal to or less than the figure in the 0.05 column, the result is significant at that level.

Non-parametric test – Chi square

The Chi square test assesses the significance of the discrepancy between results that were actually achieved and the results that were expected. The formula for calculating the Chi square is given below:

\[ x^2 = \sum [(O - E)^2 / E] \]

where:
- \( \Sigma \) = sum of
- \( O \) = observed frequency
- \( E \) = expected frequency

The degree of freedom (df) is calculated using the formula \( df = c - 1 \), where \( c \) is the number of cells.

Activity: Wilcoxon matched pairs signed ranks

Imagine you are working with a sport psychologist and you want to find out if imagery training has an influence on sprint times in youth 100 metre sprinters. You could do this by using the Wilcoxon matched pairs signed ranks test. Look at the data in Table 4.10 and, using the steps above, find out whether the imagery training programme has been successful.

<table>
<thead>
<tr>
<th>Subject pair</th>
<th>Condition A (run times pre-imagery training)</th>
<th>Condition B (run times post-imagery training)</th>
<th>( d ) (A minus B)</th>
<th>Rank of ( d ) differences</th>
<th>Rank of plus differences</th>
<th>Rank of minus differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.09</td>
<td>11.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11.23</td>
<td>11.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11.55</td>
<td>11.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11.46</td>
<td>11.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>11.22</td>
<td>11.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>11.13</td>
<td>11.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>11.01</td>
<td>10.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10.93</td>
<td>10.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10.99</td>
<td>10.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11.39</td>
<td>11.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.10**: 100m sprint times pre- and post-imagery training.
Imagine you are working with a superstitious squash coach. One of his main players plays on four courts on a regular basis and the coach believes that one of the courts is an unlucky court for the player. He believes the player loses more on that court than on any of the others. The coach aims to prove his point by comparing the number of losses on each of the four courts. The player has lost a total of 40 matches over the course of the season. From this, it could be expected that the player would lose an equal number of matches on each court (i.e. 10).

Using the Chi square test, find out whether there is a significant difference between observed and expected losses. In this example, \( df = 4 - 1 \), as there are four courts.

Use the Table 4.12 to work out your answer.

<table>
<thead>
<tr>
<th>Court number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed losses (O)</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>Expected losses (E)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>((O - E)^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((O - E)^2 \div E)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sum((O - E)^2 \div E))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>df</th>
<th>.10</th>
<th>.05</th>
<th>.02</th>
<th>.01</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.71</td>
<td>3.84</td>
<td>5.41</td>
<td>6.64</td>
<td>10.83</td>
</tr>
<tr>
<td>2</td>
<td>4.60</td>
<td>5.99</td>
<td>7.82</td>
<td>9.21</td>
<td>13.82</td>
</tr>
<tr>
<td>3</td>
<td>6.25</td>
<td>7.82</td>
<td>9.84</td>
<td>11.34</td>
<td>16.27</td>
</tr>
<tr>
<td>4</td>
<td>7.78</td>
<td>9.49</td>
<td>11.67</td>
<td>13.28</td>
<td>18.46</td>
</tr>
<tr>
<td>5</td>
<td>9.24</td>
<td>11.07</td>
<td>13.39</td>
<td>15.09</td>
<td>20.52</td>
</tr>
<tr>
<td>6</td>
<td>10.64</td>
<td>12.59</td>
<td>15.03</td>
<td>16.81</td>
<td>22.46</td>
</tr>
<tr>
<td>7</td>
<td>12.02</td>
<td>14.07</td>
<td>16.62</td>
<td>18.48</td>
<td>24.32</td>
</tr>
<tr>
<td>8</td>
<td>13.36</td>
<td>15.51</td>
<td>18.17</td>
<td>20.09</td>
<td>26.12</td>
</tr>
<tr>
<td>9</td>
<td>14.68</td>
<td>16.92</td>
<td>19.68</td>
<td>21.67</td>
<td>27.88</td>
</tr>
<tr>
<td>10</td>
<td>15.99</td>
<td>18.31</td>
<td>21.16</td>
<td>23.21</td>
<td>29.59</td>
</tr>
<tr>
<td>11</td>
<td>17.28</td>
<td>19.68</td>
<td>22.62</td>
<td>24.72</td>
<td>31.26</td>
</tr>
<tr>
<td>12</td>
<td>18.55</td>
<td>21.03</td>
<td>24.05</td>
<td>26.22</td>
<td>32.91</td>
</tr>
<tr>
<td>13</td>
<td>19.81</td>
<td>22.36</td>
<td>25.47</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>15</td>
<td>22.31</td>
<td>25.00</td>
<td>28.26</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>16</td>
<td>23.54</td>
<td>26.30</td>
<td>29.63</td>
<td>32.00</td>
<td>39.29</td>
</tr>
<tr>
<td>17</td>
<td>24.77</td>
<td>27.59</td>
<td>31.00</td>
<td>33.41</td>
<td>40.75</td>
</tr>
<tr>
<td>18</td>
<td>25.99</td>
<td>28.87</td>
<td>32.35</td>
<td>34.80</td>
<td>42.31</td>
</tr>
<tr>
<td>19</td>
<td>27.20</td>
<td>30.14</td>
<td>33.69</td>
<td>36.19</td>
<td>43.82</td>
</tr>
<tr>
<td>20</td>
<td>28.41</td>
<td>31.41</td>
<td>35.02</td>
<td>37.57</td>
<td>45.32</td>
</tr>
<tr>
<td>21</td>
<td>29.62</td>
<td>32.67</td>
<td>36.34</td>
<td>38.93</td>
<td>46.80</td>
</tr>
<tr>
<td>22</td>
<td>30.81</td>
<td>33.92</td>
<td>37.66</td>
<td>40.29</td>
<td>48.27</td>
</tr>
<tr>
<td>23</td>
<td>32.01</td>
<td>35.17</td>
<td>38.97</td>
<td>41.64</td>
<td>49.73</td>
</tr>
<tr>
<td>24</td>
<td>33.20</td>
<td>36.42</td>
<td>40.27</td>
<td>42.98</td>
<td>51.18</td>
</tr>
<tr>
<td>25</td>
<td>34.38</td>
<td>37.65</td>
<td>41.57</td>
<td>44.31</td>
<td>52.62</td>
</tr>
<tr>
<td>26</td>
<td>35.56</td>
<td>38.88</td>
<td>42.86</td>
<td>45.64</td>
<td>54.05</td>
</tr>
<tr>
<td>27</td>
<td>36.74</td>
<td>40.11</td>
<td>44.14</td>
<td>46.96</td>
<td>55.48</td>
</tr>
<tr>
<td>28</td>
<td>37.92</td>
<td>41.34</td>
<td>45.42</td>
<td>48.28</td>
<td>56.89</td>
</tr>
<tr>
<td>29</td>
<td>39.09</td>
<td>42.56</td>
<td>46.69</td>
<td>49.59</td>
<td>58.30</td>
</tr>
<tr>
<td>30</td>
<td>40.26</td>
<td>43.77</td>
<td>47.96</td>
<td>50.89</td>
<td>59.70</td>
</tr>
</tbody>
</table>

Table 4.12: Critical values of Chi square.
Non-parametric test – Spearman rank-order correlation

The Spearman rank-order correlation test is similar to the Pearson product moment correlation coefficient in its purpose. However, it is a non-parametric equivalent and is used when your data is ordinal (ranked). This test should be used when you want to find a relationship between two sets of ordinal data (for example, goals scored and final league position in football, serving accuracy and final ladder position in badminton, golf driving distance and final leader board position).

The first step is to rank your data (goals scored/serving accuracy/golf driving distance) from highest to lowest, with 1 being the highest. After this, determine the difference between your data and the place in the tournament. This must be squared and then summed.

The formula used for the test is shown below:

\[
r_s = \frac{6 \sum D^2}{n(n^2 - 1)}
\]

where:
- \(n\) = number of ranked pairs
- \(D\) = difference between each pair
- \(\sum D^2\) = the sum of the squared differences between rank

To interpret the significance of your \(r_s\) value, select the level of significance (0.05) and calculate the degree of freedom \((df)\) for your test. For the Spearman rank-order correlation test, this is calculated using the formula \(n - 2\). Compare your value to the table of significance (Table 4.13) to find whether your results are significant.

ICT-based techniques

It is common to use ICT-based techniques to conduct your statistical analysis. However, at this early stage in your research career, it is recommended that you conduct your statistical analysis by hand so you become familiar with data handling. When you are comfortable with manual data handling, the two main ICT programmes you could use for your data analysis are the Statistical Package for Social Sciences (SPSS) and Microsoft Excel. ICT-based techniques are used as they are a powerful way of interpreting statistics and give you more precise results (they can calculate results to a higher level of significance than can sometimes be done by hand).

More statistical tests

Using books and the Internet, research the following examples of statistical tests:
- Mann Whitney U
- ANOVA
- MANOVA

Use the descriptions and worked examples above to guide your reading. You should try to find the following information:
- a description of the test including the formula
- an explanation of the formula
- when you would use each test.

Remember

When two or more scores are tied, each is given the mean of the ranks and then the next rank is missed out. For example, if two participants are in level fourth place, they are both given a rank of 4.5 and the next place is given a rank of 6.

Remember

It is important to select the correct test when conducting statistical analysis. Your choice of statistical test will depend on whether the data you have collected is interval, ratio, ordinal or nominal; the number of groups that you have and/or the number of variables that you have.
Activity: Spearman rank-order correlation

You are interested in the influence of home advantage in sport – a lot of people talk about it but you want to find out if it actually exists. For your favourite sport or a famous sport, find a league table from the end of the season and find out if there is a relationship between final league finishing position and the number of games won at home. Use the formula and the table below to calculate your answer; and use the accompanying table of critical values to interpret the significance of your result.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Rank of league position</th>
<th>Rank of games won at home</th>
<th>$d$</th>
<th>$d^2$</th>
</tr>
</thead>
</table>

![Table of Spearman rank-order correlation](image)

Table 4.13: Critical values of rank order correlation coefficient
You are working as a performance analyst for a football team. You have been asked to find out if there is a relationship between goals scored and the final league finishing position because the manager cannot decide whether to spend his transfer budget on a new striker or a new goalkeeper as your team desperately needs both but can only afford one. To do this, rank teams in relation to their final league position, look at the goals scored and then rank them in relation to the number of goals scored – before investigating the nature of the relationship between the two. You are in the planning stages of the project at the moment and need to discuss the approach you will take for this research with your research team. In your discussion, you need to include the following:

1. The best research design to use and a justification of your choice. **M4**
2. The best data collection technique (or techniques) to use and a justification of your choice. **M4**
3. A description of two contrasting data analysis techniques. **P5**
4. The best way of analysing your data and a justification of your choice. **M4**

**Grading tips**

- To attain M4 you should choose the best data analysis technique for the problem and say why it is the most suitable.
- To attain P6 you should describe one parametric test and one non-parametric test.
- Try to relate the research methods that you select to specific aspects of the project that you want to complete as this will allow you to make a stronger justification.
- If you think it is more appropriate to use a combination of different research methods, make sure that you justify each method.

**Functional skills**

By discussing your ideas for the most appropriate research methods, you could provide evidence of your English skills in speaking and listening.

**PLTS**

By presenting a persuasive case for your choices of research methods, you could develop your skills as an effective participator.
Amy is a performance analyst for a professional football club. One of her key job roles is to analyse football matches to see how well the team and players are performing from a statistics perspective.

‘Professional football is such big business now that football teams are looking for as much detail as they can get about the performance of the team as a whole, and of the individual players, so that they can start to see who is performing well, and who isn’t. This has lots of influence over team selections as the statistics that I produce are often good indicators of whether or not a player is tired and needs a rest, or even if they look like they’re not putting in enough effort. I can also give players detailed breakdowns of what they have done during the game.

For example, how many passes they have completed, how far they have run, how many shots they’ve had on target (you know – like the ones you see on the television when a game is on). I can keep a record of this for them over the course of a season and we can use this to help develop the players.

‘As well as being able to analyse players and our own team, I analyse lots of games of opposition teams if we have a big game coming up (for example, I might look at their star striker and be able to give the staff a breakdown of where they prefer to shoot from and how they score most of their goals) which the manager and the coaching staff find really useful – and sometimes they might even get me to compare the performance statistics of a couple of players that they’re interested in buying.

‘One of the key problems I face is that sometimes the playing and coaching staff don’t understand some of the statistics that I present them with, so I do need to spend some time with them to explain what is going on.’
Just checking

1. What is research?
2. What is qualitative research?
3. What is quantitative research?
4. What are the key issues associated with research methods?
5. What are the different types and classifications of data?
6. What are the common designs used in research?
7. What is the BASES Code of Conduct and why is it important?
8. What are the different stages of qualitative data analysis?
9. What are the different ways of organising quantitative data?
10. What are the different ways of displaying quantitative data?
11. What are the different measures of central tendency and variability?
12. What are the different types of inferential statistics used in quantitative research?

Assignment tips

- To justify something, remember that you need to say why it is appropriate for its intended purpose.
- It is a good idea to try to bring your work to life by using lots of examples as this will demonstrate a greater level of understanding.
- There are some useful websites that can help you with this unit such as www.socialresearchmethods.net, www.sportsci.org and http://openlearn.open.ac.uk/. However these are quite high level so you will need to take your time when reading the information you find.
- Try to read as many journal articles as you can when you are going through this unit. It will benefit your understanding if you can see the different research methods that are used in a range of projects.
- Developing research is all about generating ideas so discuss your ideas with your classmates and try to work with each other to extend each other’s thinking around this topic.
- At this level, when conducting research, you should try to limit yourself to looking at a maximum of two groups or a maximum of two variables so that you are getting a good understanding and a challenging experience but you are not pushing yourself too far.