Census@school

Data Handling

Grades 7, 8 and 9

Statistics South Africa

your leading partner in quality statistics

Foreword

Statistics South Africa (Stats SA) is committed towards building statistical capacity and promoting statistical literacy within the organization, other government departments and within schools and the public as a whole.

The Census@School (C@S) 2009 project was undertaken nationally in all provinces. Data were collected from a sample of 2 500 schools selected from the Department of Basic Education's database of approximately 26 000 registered schools (EMIS database). The main objective of the project is to raise awareness of the national census, how it gathers data, and its benefits to society. In addition, data collected should be used to provide contextual material for teachers and learners to use for teaching and learning of data handling, and promoting statistical literacy relating to a variety of subjects.

The first series of Mathematics Study Guides on Data Handling and Probability for the Senior Phase (Grades 7–9) using the 2009 C@S data has been developed. This milestone has been achieved through the collaboration with and support from the national and provincial Departments of Basic Education with regard to the C@S projects.

This material has been written for Statistics South Africa by
Meg Dickson
Mmatladi Khembo
Farhana Motala
Ingrid Sapire
Elsie Venter
Kerryn Vollmer

EDITOR
Jackie Scheiber

RADMASTE CENTRE

University of the Witwatersrand

CONTENTS

Chapter 1

Collecting data

In this chapter you will:

- *Define data*
- *Learn about the Investigative Cycle*
- *Learn about methods used for collecting data*
- *Learn about the difference between samples and populations*
- *Select and justify appropriate methods for collecting data*
- *Learn about the design and use of questionnaires*

In statistics you often work with tables of data. Statistical results are also often shown in graphs rather than just in words and numbers. These graphs make it easier to interpret and analyse information. In this chapter you will learn about collecting data.

What is data?

NOTE

Data is a collection of facts such as values or measurements.

Data can be numbers, words, measurements, observations or even just a description of things.

Statistics is the name given to the *study of data*. It involves:

1. Collecting data
2. Sorting data
3. Displaying data in diagrams or charts
4. Analysing the results
5. Coming to conclusions.

- ✓ If, for example, someone tells you that 55% of the learners at their school are boys while only 5% of the teaching staff is male, they have given you some statistical information.
- ✓ What these percentages (which are statistics) tell you is that more than half of the children at the school are boys while only a very small percentage (5%) of the teachers are men.

The investigative cycle

When we do a statistical investigation, we use the statistical enquiry cycle, often called the **PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle**.

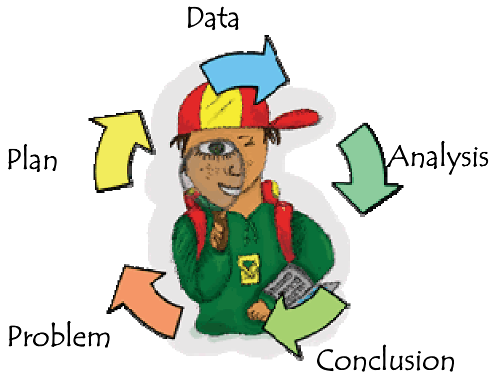
PPDAC reminds us that we have to consider:

- The **P**roblem – what we are going to research
- A **P**lan – how we are going to collect the data
- The **D**ata – how the data is managed and organised
- The **A**nalysis – exploring and analysing the data
- The **C**onclusion – the conclusions that we come to at the end of the cycle.

The South African Census@School activities have been developed using the investigative cycle. Statisticians use this cycle and so can you.

✓ The cycle starts with a problem, and goes on to planning and so on, but it does not have to end at the conclusion.

✓ The conclusion could just be the beginning of a new problem that you find to investigate. That's why it is a cycle.



We are going to go through each of the stages in the PPDAC cycle so that you will get an idea of the stages in the cycle and how they fit together.

Stage 1: Problem

The problem section is where we make decisions about

→ what data to collect
→ who to collect it from
→ why it is important.

Some questions that can help you to think about the problem and to develop a statistical question of your own are:

- How do I go about answering this question?
- What do I need to know?
- How will I find the information that I need?
- What will I do with the information that I collect?
- Who will find this information useful?
- Is this information relevant to the problem?

- ✓ Formulating and defining a statistical question is important as it tells you what to investigate and how to investigate it. *This is where you start*.
- ✓ Most investigations begin with a wondering about something.
  For example: 'I wonder how many different pets there are in the pet shop down the road?'
- ✓ From this general question a statistical question needs to be developed so that a meaningful investigation can be carried out.

Stage 2: Plan

The planning section is about how you will gather the data.

Questions that can help you to think about planning your data gathering project are:

- How will I gather this data?
- What data will I gather?
- How am I going to record this information?

For example: at a certain pet shop there are 205 goldfish, 6 puppies, 15 kittens, 37 budgies, 17 hamsters and 4 cockatiels. The different types of pets and the number of each type is **data**.

- → You need to make predictions and then test them. You might think to yourself "There are mostly kittens and puppies at the pet shop".
- → After that you can reflect on the difference between your prediction and the result.
- → You need to plan things such as the sample size and method of data collection. You think about whether you will count all of the pets (the whole pet shop population), or if the shop is big, maybe only the pets in every second cage (a sample of the pets). But to make this decision you need to plan.

Stage 3: Data

The data section is how the data is managed and organised.

Questions you need to answer here are:

- How am I going to record my data?
- How will I present my data?

Raw Data is data which has been collected but not yet sorted out in any way.

For example: In order to organise the raw data about the animals in the pet shop, you could make a table with the different types of pets in a list down one side and then make a tick for every pet you count.

You may record their data in any format as long as it is clear and easy to work with.

- ✓ A table is usually the best format. Tables are the most common organisational tool that statisticians use.
- ✓ You could count the number of ticks to see how many of each pet type there are and you can record this information in a table.
- ✓ Then you'll think about how to present the data – probably with a graph show that you can give the "picture" of what the information you have found looks like.

Stage 4: Analysis

The analysis section is about exploring the data and reasoning with it.

Some questions that can help you think about how to analyse your data are:

- What do I notice?
- Why do I think it looks like this?
- Are there relationships between some of the variables I have investigated?

We can:

- ✓ Read the data – take information directly off a graph
- ✓ Read ***between*** the data – interpret the data
- ✓ Read ***beyond*** the data – extend, predict or infer
- ✓ Read ***behind*** the data – connect the data to the context.

For example:

→ Look at the table of information and graphs that you may have drawn of the pet shop statistics and see what it tells you.

→ You might realise "Kittens and puppies may be bigger and easier to see from the pet shop window, but there are far more fish for sale in this pet shop than kittens and puppies put together".

→ You might wonder about which pets are most popular (reading beyond), which pets are the easiest/hardest to look after (reading behind) or simply compare the data you see in the graph (reading the data).

Stage 5: Conclusion

The conclusion section is about answering the questions in the problem section and providing reasons based on the analysis of these questions.

Some questions that can help you to think about your conclusions are:

- What do I notice?
- Why do I think it looks like this?
- Are there relationships between some of the variables I have investigated?

✓ Your conclusions should relate back to your original question.

✓ You should mention any features you noticed or wondered about and investigated.

✓ You should give reasons based on what you find out in your investigation.

For example: Based on the pet shop data, you could conclude that fish are the most popular pet and easiest for the pet shop to keep since they have the most fish.

The five stages (**Problem; Plan; Data; Analysis; Conclusion)** take you through the full "investigative research cycle".

The example on the next page illustrates how the five stages of the research cycle can be used:

Example 1.1

The Census@School researchers wanted to investigate whether South African learners had access to libraries, computers and the internet at school. Explain how you can use the investigative cycle to do this.

Solution

1) **PROBLEM:**
   The general question that C@S wanted to investigate was something like – 'I wonder how accessible libraries, computers and the internet are to the learners in South Africa?'
2) **PLAN:**
   C@S had to plan how to collect the data. The project collected data for many questions, including questions about the accessibility of libraries, computers and the internet. They had to plan things such as:
   a) *How to gather this data* – C@S decided to take questionnaires into schools. Two questionnaires were developed, one for Grades 3 – 7 and one for Grades 8 –12.
   b) *What data to gather* – the questionnaire included questions about how easy/difficult it is for learners to get to libraries.
   c) *How to record this information* – the learners would write their answers on the questionnaires which could be collected.
3) **DATA:**
   C@S had to think about the way in which they would organise the data.
   This would involve:
   a) *How to record the data* – C@S entered all of the data onto a table.
   b) *How to present the data* – a bar graph representing the accessibility of libraries, computers and the internet was presented in the C@S report.
4) **ANALYSE** data:
   What the bar graph showed was that more learners have greater access to libraries than to computers and the internet.
   Fewer learners have access to the internet than to libraries or computers.
5) **CONCLUSION:**
   a) The analysis showed that, of the three, libraries are the most accessible and the internet is the least accessible. Computers are somewhere in the middle.
   b) Because learners do not have to pay money to take books out of a library, more learners can access a library. Computers cost a lot and are not always available. Access to the internet costs even more money and so it is even less common.
   c) There is a relationship between ownership of a computer and access to the internet, although internet cafes do provide internet access to people who do not own computers.

Exercise 1.1

Make up your own example of a statistical investigation, like the pet shop example used on the previous pages to illustrate the stages in the investigative cycle.

If you can't think of an example of your own try one of the following:

- What is the range of ages of learners in your class or school?
- What is the range of learner heights in your class or school?

In the process, think of the following:

1) **What is your PROBLEM?** – State your research question**.**
2) **What is your PLAN for data collection?**
   a) How are you going to gather this data?
   b) What data will you gather?
   c) How are you going to record this information?
3) **What DATA will you record and how?**
   a) How are you going to record your data?
   b) How will you present you data?
4) **How do you think you will ANALYSE your data?**
   When you analyse the data, think about:
   a) What you notice?
   b) Why do you think it looks like this?
   c) Are there relationships between some of the variables you have investigated?
5) **What kinds of questions do you hope to answer in your CONCLUSION?**
   When you reach your conclusions, think about:
   a) What you notice?
   b) Reasons for why you think it looks like this.
   c) An explanation of the relationships between some of the variables that you investigated.

Methods used for collecting data

When you collect data, you have to think carefully how you are going to do it.

How you collect data depends on
- ✓ the data you have to find
- ✓ the time available for the research
- ✓ the money available for the research, and so on.

You could use:
- ✓ questionnaires
- ✓ observation
- ✓ experiments
- ✓ telephonic research
- ✓ group or individual interviews which you record on tape.

To be able to make conclusions from data, we need to know how the data were collected; that is, we need to know the ***method(s) of data collection***.

There are **four main methods** of data collection:

1) **A census**.

   A census is a study that gets data from *every* member of a population. In most research that you do yourself, a census is not practical, because of the high cost involved and the time it takes.

2) **A sample survey**.

   A sample survey is a study that gets data from a smaller group within the population, and is used to estimate population characteristics.

3) **An experiment**.

   An experiment is a controlled study in which the researcher tries to understand cause-and-effect relationships.

   Experiments are usually "controlled" in the sense that the researcher controls (1) how subjects are assigned to groups and (2) which treatments each group receives.

   In the analysis phase of an experiment, the researcher compares the different effects of the experimental treatment.

4) **An observational study**.

   Like an experiment, an observational study tries to understand the cause-and-effect of relationships.

   However, in observational studies, the researcher does not have any control over who goes into which group and so the findings may not be as clear to the observer.

Example 1.2

You have been asked to investigate what the favourite sport is of the learners in your school. Think about how you are going to do it, and then answer the following questions about the methods you are going to use:

1) What **data** do you want to collect?
2) How do you **plan** to collect you data? Explain how you would go about it.
3) If you collected data initially as **raw data,** how would you record this data?
4) What form of **data collection could** you use?
   a) Could you have used a **census**? Explain why or why not.
   b) Could you have used a **sample survey**? Explain why or why not.
   c) Could you have used an **experiment**? Explain why or why not.
   d) Could you have used an **observational study**? Explain why or why not.

Solution

1) I want to collect information on favourite sports.
2) I will draw up a questionnaire with questions about different sports so that I can find out about learners' favourite sports. I need to have a list of sports so that learners can say which they like best from the list.
3) If I let the learners each tick on a separate questionnaire I would have raw data. I could make a table in which I could tally (count up) the number of learners who like each sport.
4) Form of **data collection**:
   a) NO, I could NOT have used a **census** because one person cannot carry out a census. With a census, the whole population gets involved.
   b) YES, I could have used a **sample survey** because I could have decided to ask a sample (a smaller representative group) of learners from each of the grades, instead of asking every learner in the whole school. This would have been much easier and more efficient to do.
   c) NO, I could NOT have used an **experiment** because I was not trying to understand cause-and-effect relationships.
   d) NO, I could NOT have used an **observational study** because, again, I was not trying to understand cause and effect relationships.

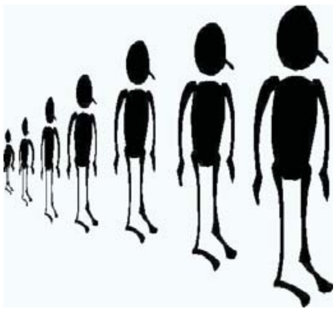
Exercise 1.2

You have been asked to investigate what the favourite tuck shop food is of the learners in your class. Think about how you are going to do it, and then answer the following questions about the methods you are going to use:

1) What **data** do you want to collect?
2) How do you **plan** to collect you data? Explain how you would go about it.
3) If you collected data initially as **raw data,** how would you record this data?
4) What form of **data collection could** you use?
   a) Could you have used a **census**? Explain why or why not.
   b) Could you have used a **sample survey**? Explain why or why not.
   c) Could you have used an **experiment**? Explain why or why not.
   d) Could you have used an **observational study**? Explain why or why not.

The difference between samples and populations

Before you go any further in planning statistical research you need to learn about populations and samples so that you'll be able to plan effectively for your data collection.



A Population

A population is the whole group of people (or things) that you want to speak about or investigate.

Examples:

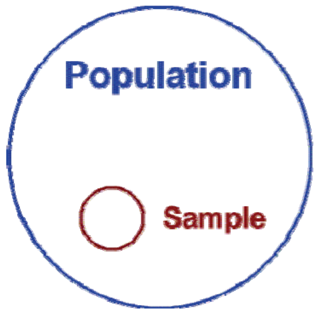
a) In the example about what the favourite foods from the tuck shop are,
   - → If you wanted to find out about the favourite foods of the learners in your class then *your class would be the population*
   - → If you wanted to find out about the favourite foods of the learners in the whole school then *all of the learners in the school would be the population*.

b) In the example about the pet shop, all of the pets in the shop would be the population.

✓ Normally a population is quite large, but the size of your population depends on your statistical research.

✓ The larger your population (for example you might want give information relating to "male South Africans") the more impossible it becomes to ask each member of that population the questions you want to ask. *You then need to develop some way to help you with selection of individuals from your population, and to give a generalised result based on information obtained from your sample.*

A Sample



- When you want to collect information from a certain population, which can sometimes be large, you could take a smaller group, which could represent the whole population.
- Such a group, used to represent the population, is called a **sample**.

Examples:

a) If your favourite food survey was only to be done in your class you would not have to find a sample.
   - → If the survey needed to be extended to the *whole school*, you will not have time to ask everyone in the school the questions that you want to ask.
   - → What you could do is ask *a selection of learners* whose answers you will use to make conclusions about the whole school.

→ You would try to ask learners from each grade in the school, with a good spread across the school, to make your sample representative of the whole school population.

b) In the example of the pet shop, counting pets in every second cage would be a way of sampling.

Example 1.3

Explain the difference between a population and sample by using the investigation to find out about the favourite sport of the learners

a) In your class
b) In your school.

Solution

a) If I wanted to find out about the favourite sports that the learners *in my class* like to play, *the class is the population* for my research. I can ask every learner in my class and then talk about the sports *they* like. I don't have to take a sample of the learners in the class, because the population is small.

b) If I want to find out about the favourite sports that the learners *in my school* like to play, the *population for my research is every learner in the whole school*. This is harder to do! That could be quite a big number. I would rather ask a sample of the learners in the school what their favourite sport is. Then I could use this information to decide about the favourite sport of the learners in the school.

This means that I would ask a smaller group of learners selected from the whole school population. But to be sure that what I say about the school is realistic I would have to be careful to choose a sample that represents the full spread of learners in my school. This sample could consist of:

- Some (a percentage) from each grade
- Some (a percentage)from each class in each grade

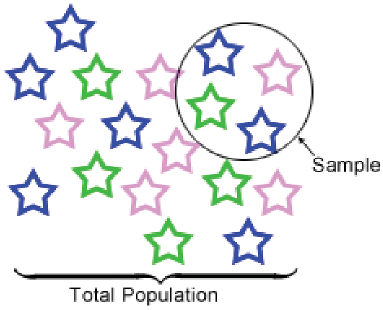
Exercise 1.3

- Thembi is part of a group of 7 learners who are friends.
  It is break time and Thembi is looking forward to playing soccer with them.
  He makes a quick survey of his friends to find out who wants to play soccer with him.
  - → James, Hlengani and Thoko are keen to join in.
  - → Dan and Lebo say that they want to sit and play with their latest computer game.
  - → Habib wants to stand around and talk to some other friends.
- Based on his quick survey, 4 out of the 7 friends (57% of them) wants to play soccer. Thembi wonders whether 57% of South African learners play soccer at break time.
- Habib thinks that Thembi's quick survey is flawed (that it is not accurate) if he wants information on all South African learners.

1) What is a sample?
2) What is a population?
3) What is the statement of Thembi's research problem?
4) Describe Thembi's quick survey. Is the sample he chose representative of the whole population?

Random samples

You need to be able to distinguish between samples and populations.
One way of selecting a sample is by using **random sampling.**



- ✓ If each member of the population has an equal chance of being chosen for a sample, then you have a random sample.
- ✓ Taking a random sample makes sure that the sample represents the population.
- ✓ One way of selecting a random sample is to place all the names in a hat or box; to mix the names up; and to then take the names out of the hat or box without looking. The mixing up of the names is important or else the names which are placed last in the hat will have a better chance of selection than those which were placed first into the hat. This will lead to ***bias*** in your data.

For example
If you want to find out what the favourite tuck shop food is of all the learners in the school, be sure not just to ask your friends (who might all have similar tastes) or people only from one class.

→ You need to think of a way of getting the information from as wide a selection of learners as possible. That way, what you hear from your sample (your selection carefully chosen to be representative) will reflect what is going on in the whole school.

For example
Two ways you could select a random sample of learners from your school are:
1) You could cut up all of the school class lists and then draw 30 names out of the hat. This might not give you a random sample if the hat is not well enough shaken.
2) You could take each class list and select every 10^{th} name on the lists. This could be the easiest way to get a representative random sample.

Exercise 1.4

Think again of Thembi's quick soccer survey in Exercise 1.3.

1) What is the population represented by the sample that Thembi collected?
2) Did Thembi use **random selection** to find his data?
3) What aspects of Thembi's survey does Habib think are flawed? Do you agree with Habib? Explain your answer.
4) How can Thembi improve his survey if he wants information on all South African learners? Explain.
5) Write down some other statistical research problems that you think of.
   Try to think of problems which are different to one another, such as:
   a) Write about one problem where you could use the whole population in your research.
   b) Write about one problem where you could not use the whole population in your research.
   c) Write about one problem that requires random selection to obtain the sample.
   d) Write about one problem that does not require random selection to obtain the sample.

Select and justify appropriate methods for collecting data

- ✓ To select the right method for your research you need to know about the **pros (good things) and cons (bad things)** about the different methods of data collection. Each method of data collection has advantages and disadvantages, depending on the research you want to do.
- ✓ One main factor that affects data collection is the available **resources**. Resources are things like money, time and people.
- ✓ When the population is large, a sample survey is cheaper to do than a census. This is because a sample is smaller and cheaper and easier to work with.
  - A well-designed sample survey can provide very good estimates of population characteristics, but more quickly and cheaply than a census.
  - When the population is small, you can design research that you can manage. This will give you experience in small-scale statistical research.

Exercise 1.5

Jane wants to find out what is the favourite sport of South African children. She spends a week asking every learner she sees at her school what their favourite sport is. She tallies up the information and decides that Athletics is the favourite sport.

My school			
Favourite sport	**Boys**	**Girls**	**Total**
Athletics	34	25	59
Soccer	39	5	44
Netball	2	30	32
Softball	15	12	27
Cricket	16	4	20
No Favourite sport	10	11	21
Other	8	6	14

She decides to check up on her finding. She goes to the South African Census@School website and finds the following table:

Eastern Cape			
Favourite sport	**Male**	**Female**	**Total**
Soccer	6 931	775	7 706
Netball	135	5 626	5 761
Athletics	1 957	2 058	4 015
No Favourite sport	722	1 409	2 131
Rugby	1 276	44	1 319
Unspecified	595	476	1 071
Volleyball	293	435	728
Cricket	450	59	509
Dance Sport	80	387	467
Softball	185	281	465

1) How big is Jane's sample?
2) Did Jane take a random sample?
3) Did Jane collect her data in an appropriate way? Explain.
4) What are the top five favourite sports according to Jane's survey?
5) What are the top five favourite sports according to the Census@School table?
6) How does Jane's finding that athletics is the favourite sport in her school compare with the Census@School findings?
7) Is Jane's table representative of the whole of South Africa? Explain.
8) Is the Census@School table representative of the whole of South Africa? Explain.

The design and use of questionnaires

Statistical research is about asking and answering questions. There are different kinds of questions involved. To do statistical research you need to develop your skills of:

- **Posing questions** (the investigative questions) – thinking about and asking questions that can lead to statistical research.
- **Selecting appropriate sources for the collection of data** (knowing where to find your data) – planning research so that you can find the right data to help you to answer the questions you have posed.
- **Designing and using simple questionnaires to answer questions** (making the tools that you use to carry out the research)

Posing or asking questions

You need to know how to ask or pose questions relating to social, economic, and environmental issues in your own environment.

- ✓ To do this you need to develop your own awareness of these things by reading newspapers, watching TV and listening to the radio.
- ✓ Talking to your family, friends and teachers will also help!
- ✓ Then you need to start thinking about the kinds of questions to ask.

You could ask about:

- Particular things
- Comparative things
- Relationships.

Example 1.4

Give an example of each of the different kinds of questions you could ask when doing research at your school.

<u>Solution</u>

1) **Particular things** – What was the highest score in the Grade 7 maths final exam at my school?
2) **Comparative things** – Are there more 14 year olds or 15 year olds in Grade 8 at my school?
3) **Relationships** – Is there a relationship between the amount of money spent at the tuck shop and the age of the learner at my school?

Exercise 1.6

Give another example of each of the different kinds of questions you could ask when doing research at your school.

1) A particular question
2) A comparative question
3) A relationship question

Selecting appropriate sources for the collection of data

You also need to know how to select appropriate sources for the collection of data.

- ✓ You could ask people such as your friends and family.
- ✓ You could look for data in a selection of newspapers, books, magazines.

When you decide to collect data, you need to decide whether you are working with a sample or a population.

Example 1.5

Where would you find your data to answer the question you posed in Example 1.4?

Solution

1) **Particular question** – What was the highest score in the Grade 7 maths final exam at my school?
   I could find out about this by speaking to the teachers and getting the mark lists for all of the Grade 7 classes at my school.
2) **Comparative question** – Are there more 14 year olds or 15 year olds in Grade 8 at my school?
   I could find out about this by asking the Grade 8 learners at my school about their age.
3) **Relationship question** – Is there a relationship between the amount of money spent at the tuck shop and the age of the learner at my school?
   I could ask a sample of learners who shop at the tuck shop about their age and the amount that they spend. Then I would be able to see if there is a relationship. I would have to be sure to ask a big enough representative sample.

Exercise 1.7

Where would you find your data to answer the questions you posed in Exercise 1.6?
Write down the question and then explain where you would find your data.

Designing and using simple questionnaires to answer questions

✓ A questionnaire consists of a list of questions. These questions are used to collect information and answers from people.

✓ Always try out your questionnaire before you use it to check that it gets you the data that you want. You may be surprised at how some questions are answered!

✓ It is very common in questionnaires to use:
- Questions with yes/no type responses.
- Questions with multiple choice responses.

Both of these types of questions are popular because they are easy to analyse.

✓ Here are some things to think about when you write your questions:

- Don't use abbreviations (shortened forms of words) as some people may not understand them.
- Keep the language as simple as possible.
- Don't ask questions which have two parts in one (for example *Do you like hotdogs and coke?* If they answer "yes" they could mean they like both or either of the two).
- Be careful of asking open questions such as "what is your favourite food" because you may get too many different answers.
- Closed questions (questions where respondents have to choose from given answers) can limit the number answers you get, but this could also limit your research. So you need to think carefully about the list of options you give in a closed question.

✓ A good questionnaire ends with a comments section that allows the respondent to write down any other issues not covered by the questionnaire. This allows them to express any thoughts, questions or concerns they might have.

✓ Lastly, there should be a message at the end thanking the respondents for their time and patience in completing the questionnaire.

Example 1.6

1) Give an example of each of the following two types of questions that could be asked in a questionnaire.
   a) a yes/no response type question.
   b) a multiple choice response type question.
2) Why are these types of questions easy to analyse?

Solution

1) a) Do you eat red meat? Circle yes/no
   b) What is your favourite meat? (Circle the letter that represents your choice)
      i) Beef
      ii) Lamb
      iii) Chicken
      iv) I don't eat meat
2) These types of questions easy to analyse because the responses are limited to the options given on the questionnaire and so you don't have to worry or think about what to do with all sorts of different, uncommon answers that some people might give.

Exercise 1.8

Pick one of your research questions that you posed in Exercise 1.6. Design a 10 question questionnaire to collect data for your research.

- ✓ Your questionnaire should have **six** yes/no questions and **four** multiple choice questions.
- ✓ If you want to design questionnaires for ALL of your questions, go right ahead. The more you do, the better you'll get at doing it.
- ✓ If you had some difficulty making up questions, choose one of the following questions:
  1) What types of transport do the learners in my school use to go to school?
  2) How far do the learners in my school live from the school?

Many of the questions that have been raised in this chapter were researched by the C@S team. Results for South Africa in 2009 are available in the Census at School Results 2009 report which can be found at the Stats SA website ***www.statssa.gov.za***

You will find out lots more about working with data in the chapters that follow.

Chapter 2

Organising and Summarising Data

In this chapter you will:

- *Learn to recognise continuous and discrete data*
- *Organise and record data using tallies*
- *Draw an ungrouped and grouped frequency table*
- *Organise data using a stem and leaf diagram*

The word "data" is the plural of "datum". So data are pieces of information that are given or that one collects. Data can be words, numbers or a mixture of both.

Types of data

NOTE

- **Data** can be numerical or non-numerical.

Data can consist of two types of data: numerical data or non-numerical data. The data you collect in a survey or questionnaire may be varied – it may be about the colour of learners' eyes, their mode of transport to school, an opinion (e.g. which chocolate do you prefer). It may also be numerical (e.g. how many learners come to school by bus or taxi in the morning).

In this book we will be working with numerical data that is either **discrete** or **continuous** as well as non-numerical or **qualitative** data.

- **Qualitative data** consist of descriptions using words.
  Examples are: the colour of hair, the colour of eyes, shoe sizes.
- **Discrete and continuous data** both consist of numerical values.
  - **Discrete data** is information that is collected by counting exact amounts.
    Examples are: the number of children in a family; the number of children with birthdays in October or the number of houses with electricity.
  - **Continuous data** is collected by measurement and the values form part of a continuous scale like a number line.

Examples of continuous data are: measurements of physical quantities such as mass/weight, height, length and time (e.g. the mass of learners in a Grade 8 class measured in kilograms or the length of the right feet of the Grade 7 learners measured in centimetres or the time it takes learners to travel to school each morning measured in hours and minutes).

Organising data

When you first look at numerical data, all you may see is a jumble of information. You need to sort or summarise the data and record it in a way that puts order into it so that it makes more sense.

- ✓ Data in this form are called **raw data**. Raw data haven't been organised in any way.
- ✓ One of the most common ways of sorting data is by making a list. Data is easy to sort into lists that are either numerical or alphabetical.

Exercise 2.1

Given below are the heights (in centimetres) of 90 Grade 8 boys in an Eastern Cape school as recorded in the 2009 Census@School:

165	148	158	150	160	165	150	156	155	164	162	160	158	148	158
140	146	160	148	152	139	165	148	160	156	158	170	155	160	148
155	158	179	170	158	161	155	160	163	178	138	172	170	156	160
160	171	140	160	170	175	148	170	177	155	167	154	160	170	155
136	179	150	167	148	160	164	167	157	165	163	140	162	178	160
170	163	162	165	175	165	152	147	180	148	170	165	167	165	165

1) Write a list of the boys' heights in ascending order.
2) Which height is the most common?
3) How many learners are shorter than 150 cm?
4) How many learners are taller than 170 cm?

Tally tables

- ✓ You can also organise and summarise data using a tally table.
  A tally is a mark which shows how often something happens.
- ✓ For each score, a vertical stroke is entered in the appropriate row, with a diagonal stroke being used to complete each group of five strokes.

Example 2.1

Organise the list of heights (in centimetres) of 90 Grade 8 boys in an Eastern Cape school given in Exercise 2.1 into a tally table.

Solution

Height (in cm)	Tallies	Frequency
136	/	1
137		0
138	/	1
139	/	1
140	///	3
141		1
142		0
143		0
144		0
145		0
146	/	1
147	/	1
148	~~////~~ ///	8
149		0
150	///	3
151		0
152	//	2
153		0
154	/	1
155	~~////~~ /	6
156	///	3
157	/	1
158	~~////~~ /	6
159		0
160	~~////~~ ~~////~~ //	12
161	/	1
162	///	3
163	///	3
164	//	2
165	~~////~~ ////	9
166		0
167	////	4
168		0
169		0
170	~~////~~ ///	8
171	/	1
172	/	1
173		0
174		0
175	//	2
176		0
177	/	1
178	//	2
179	//	2
180	/	1

Frequency Tables

✓ A **frequency table** is usually given *without* the tallies. The word "frequency" may not appear in the table. Instead, the frequency column might be headed "number of learners" or something like that. Always make sure that you identify which numbers are the "values" and which ones are the "frequencies".

Height (in cm)	Frequency
136	1
137	0
138	1
139	1
140	3
141	0
142	0
143	0
144	0
145	0
146	1
147	1
148	8
149	0
150	3
151	0
152	2
153	0
154	1
155	6
156	3
157	1
158	6
159	0
160	12
161	1
162	3
163	3
164	2
165	9
166	0
167	4
168	0
169	0
170	8
171	1
172	1
173	0
174	0
175	2
176	0
177	1
178	2
179	2
180	1
TOTAL	**90**

Exercise 2.2

Conduct a simple survey of the learners in your class to ask about the month of their birthday.

1) Record the information in a frequency table like the one below:

Month	Tally	Frequency
January		
February		
March		
April		
May		
June		
July		
August		
September		
October		
November		
December		

2) In which month do the most birthdays occur?

3) In which month do the least birthdays occur?

Stem and Leaf Diagrams

Tally tables can become very long and difficult to work with if the range of possible values is very large. In the 1960s John Tukey, an American mathematician and statistician, devised a way of organising data called a **stem and leaf** diagram.

- ✓ Stem-and-leaf diagrams can be used both with discrete data and with continuous data (rounded off to the nearest whole number).
- ✓ Stem-and-leaf diagrams retain the original data information, but present it in a compact and more easily understandable way.
- ✓ When entering a number like 56 onto a stem and leaf diagram, the **tens digit** (5) forms the **stem** and the **units digit** (6) forms the **leaf**.

Example 2.2

The members of your class got the following marks in a mathematics test:

1) Organise the data using a stem and leaf diagram.
2) Write down 3 conclusions you can reach about the marks.

Marks scored in a mathematics test

32	56	45	78	77	59	65	54
54	39	45	44	52	47	50	52
40	69	72	36	57	55	47	33
39	66	61	48	45	53	57	56
55	71	63	62	65	58	55	51

Solution

1) Step 1: Decide on the values to use for the stem.

stem	leaf
3	
4	
5	
6	
7	

Look how to write 56 and 78 in the stem and leaf plot

The first number is 56: the **stem** is 5 and the **leaf** is 6

The second number is 78: the **stem** is 7 and the **leaf** is 8

stem	leaf
3	
4	
5	6
6	
7	8

Step 2 : Write down the leaves directly from the data without worrying about the order:

stem	leaf
3	2, 9, 6, 3, 9
4	5, 5, 4, 7, 0, 7, 8, 5
5	6, 9, 4, 4, 2, 0, 2, 7, 5, 3, 7, 6, 5, 8, 5, 1
6	5, 9, 6, 1, 3, 2, 5
7	8, 7, 2, 1

Step 3 : Rewrite the leaves in ***ascending order***. This makes the table easier to read.

stem	leaf
3	2, 3,6, 9, 9
4	0, 4, 5, 5, 5, 7, 7, 8
5	0, 1, 2, 2, 3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8, 9
6	1, 2, 3, 5, 5, 6, 9,
7	1, 2, 7, 8
KEY: 3/2 = 32	

2) We can read the following information from this stem and leaf diagram:
   a) The lowest mark is 32
   b) The highest mark is 78
   c) More learners got marks in the 50s than in any of the other number ranges
   d) More learners got marks in the 40s than in the 60s
   e) The most common marks were 45 and 55 as three learners got 45 and three got 55.
   f) Four learners achieved marks of more than 69.

NOTE:
- ✓ The **leaf** is the digit in the place furthest to the right in the number.
- ✓ The **stem** is the digit or digits that remain when the leaf is dropped.
- ✓ If the list of numbers included numbers like 120 ; 134 ; 127 then 12 and 13 would be the stems and 0, 4 and 7 would be the leaves.
- ✓ If the list of numbers included a single digit number like 2, 3 or 9, then the stem would be 0 and the leaves would be 2, 3 and 9.

Exercise 2.3

Given below are the heights (in centimetres) of 90 Grade 8 boys in an Eastern Cape school as recorded in the 2009 Census@School:

165	148	158	150	160	165	150	156	155	164	162	160	158	148	158
140	146	160	148	152	139	165	148	160	156	158	170	155	160	148
155	158	179	170	158	161	155	160	163	178	138	172	170	156	160
160	171	140	160	170	175	148	170	177	155	167	154	160	170	155
136	179	150	167	148	160	164	167	157	165	163	140	162	178	160
170	163	162	165	175	165	152	147	180	148	170	165	167	165	165

1) Summarise the data using a stem and leaf diagram.
2) Write down three conclusions you can reach about the heights.

Back-to-back stem-and-leaf diagram

✓ When two sets of data are to be compared, the leaves can extend in opposite directions from the same stem.

Example 2.3

The heights of 10 boys and 10 girls in Grade 7 were randomly selected.

a) Draw a back-to-back stem-and-leaf diagram to illustrate the data

b) Write down at least 2 conclusions about the heights.

Heights of a random selection of 20 learners in Grade 7 in cm	
Boys	171; 156; 154; 160; 145; 147; 149; 160; 150; 147
Girls	162; 155; 155; 155; 164; 170; 149; 156; 161; 164

Solution

a) STEP 1 : Decide on the values of the stems and draw up the table.

GIRLS' HEIGHT in centimetres		BOYS HEIGHTS in centimetres
	14	
	15	
	16	
	17	

STEP 2 : Write down the leaves directly from the data.

GIRLS' HEIGHT in centimetres		BOYS HEIGHTS in centimetres
9	14	5 7 9 7
6 5 5 5	15	6 4 0
4 1 4 2	16	0 0
0	17	1

STEP 3 : Rewrite the leaves in ascending order.

GIRLS' HEIGHT in centimetres		BOYS HEIGHTS in centimetres
9	14	5 7 7 9
6 5 5 5	15	0 4 6
4 4 2 1	16	0 0
0	17	1
KEY: 9/14 = 149 cm and 14/5 = 145 cm		

b) The heights of the girls range from 149 cm to 170 cm.
The heights of the boys range from 145 cm to 171 cm
Five of the girls have heights more than 159 cm whereas only 3 of the boys have heights more than 159 cm.
Four of the boys have heights of 149 cm or less than 149 cm, whereas only 1 of the girls has a height that is 149 cm or less.

Grouped Frequency Tables

✓ To obtain a clearer picture of a distribution, you can ***group*** data within **class intervals.**

Example 2.4

Organise the heights (in centimetres) of 90 Grade 8 boys in an Eastern Cape school as recorded in the 2009 Census@School in a ***grouped*** frequency table:

Height (in cm)	Frequency	Height (in cm)	Frequency	Height (in cm)	Frequency
136	1	151	0	166	0
137	0	152	2	167	4
138	1	153	0	168	0
139	1	154	1	169	0
140	3	155	6	170	8
141	0	156	3	171	1
142	0	157	1	172	1
143	0	158	6	173	0
144	0	159	0	174	0
145	0	160	12	175	2
146	1	161	1	176	0
147	1	162	3	177	1
148	8	163	3	178	2
149	0	164	2	179	2
150	3	165	9	180	1

Solution

→ The heights range from 139 cm to 180 cm.

→ We can organise these heights into the following class intervals:

$130 < h \leq 140$
$140 < h \leq 150$
$150 < h \leq 160$
$160 < h \leq 170$
$170 < h \leq 180$

→ **$130 < h \leq 140$** means that the heights are *more than 130 cm, but less than 140 cm or equal to 140 cm.*

→ **$140 < h \leq 150$** means that the heights are *more than 140 cm, but less than 150 cm or equal to 150 cm.*

→ The following grouped frequency table shows the heights of the learners:

Height in centimetres	Frequency
$130 < h \leq 140$	6
$140 < h \leq 150$	13
$150 < h \leq 160$	31
$160 < h \leq 170$	30
$170 < h \leq 180$	10

NOTE:

- ✓ The groups do not overlap at all.
  Notice that the groups were written $130 < h \leq 140$ and then $140 < h \leq 150$.
  In other words, you have to be careful that a particular height (e.g. 140 cm) does not appear in two different intervals.
- ✓ The data in this example – the heights of the Eastern Cape boys – is **continuous data** rounded off to the nearest centimetre.

Exercise 2.4

1) The maths marks of a class of Grade 7 learners are given below.
   a) Copy the following table and use it to organise the marks.

Mathematics Marks
32 ; 56 ; 45 ; 78 ; 77 ; 59 ; 65 ; 54 ; 54 ; 39 ; 45 ; 44 ; 52 ; 47 ; 50 ; 52 ;
51 ; 40 ; 69 ; 72 ; 36 ; 57 ; 55 ; 47 ; 33 ; 39 ; 66 ; 61 ; 48 ; 45 ; 53 ; 57

Mathematics marks	Tally	Frequency
$30 < m \leq 40$		
$40 < m \leq 50$		
$50 < m \leq 60$		
$60 < m \leq 70$		
$70 < m \leq 80$		

   b) How many learners got 50 or less for the test?

2) Lerato wanted to do a survey about the height in metres of learners in her Grade 9 class.

Heights of the Grade 12 learners in metres					
1,82	1,64	1,71	1,86	1,64	1,67
1,73	1,76	1,84	1,52	1,63	1,65
1,80	1,67	1,71	1,64	1,58	1,81
1,67	1,74	1,69	1,56	1,68	1,74
1,79	1,83	1,69	1,58	1,57	1,73

   a) Use intervals $1{,}50 < x \leq 1{,}55$; $1{,}55 < x \leq 1{,}60$; etc to draw a grouped frequency table for Lerato's data.
   b) How many learners are taller than 1,75m?

3) In a class in a school in Mpumalanga, Census@School recorded the time that 30 learners spent watching TV in a week. Their answers (to the nearest hour) are given below:

12	20	13	15	22	3
6	24	20	15	9	12
5	6	8	30	7	12
14	25	2	6	12	20
18	3	18	8	9	20

a) Draw a stem and leaf diagram to organise this data.
b) Write down three conclusions about the data.

4) The metro bus company in Durban did a survey to find out how many learners used a particular bus to come to school in town. They counted the number of learners on the bus each time it arrived in town. The numbers are given below:

11	25	60	58	55	16	23	2	44	26
49	8	14	24	7	16	47	5	30	34
9	33	10	21	1	56	32	19	6	1
21	42	9	35	25	55	37	52	15	7
31	25	6							

a) Draw a grouped frequency table to organise the data. You will need to think of appropriate group intervals.
b) Why do you think the bus company might want to know the numbers of learners on the bus?

5) A survey was conducted to find the colour of eyes of South African learners and the following results were recorded and organised in a table:

Gender	Brown	Green	Blue	Other	Unspecified	Total
Male	60 531	1 615	2 235	5 095	704	**70 180**
Female	66 993	2 015	2 449	4 686	538	**76 681**
Unspecified	789	20	33	55	247	**1 144**

a) What fraction of all the learners has blue eyes?
b) What is the total number of learners with brown eyes?
c) Why do you think that there are such a high number of learners with brown eyes?

Chapter
3

Mode, Mean, Median

In this chapter you will:

- *Learn about measures of central tendency: Mode, Mean, and Median*
- *Learn about measures of dispersion: Range and Extreme values*

When you have a data set, it is possible to summarise the data with one single number (also called a 'statistic'). These single numbers are called either Measures of Central Tendency or Measures of Dispersion or Spread.

Measures of central tendency

Measures of central tendency are numbers that are typical of a given set of data. The three measures of central tendency that we calculate are **the mode**, **the mean** and **the median**.

Mode

- ✓ The mode is the number in your data that **occurs most often**. You can also say the mode is the value that has the highest frequency.
- ✓ Sometimes two scores (or numbers) occur equally often and then the data set has more than one mode. If there are two modes we say that the data set is **bimodal**. If there are more than two modes, we say that the data set is **multi-modal**.
- ✓ Other times there might not be any number that occurs more often than any other number. This means that a data set may have **no mode**.

Example 3.1

For each of these sets of scores, find the mode.

a) 1, 7, 9, 4, 3, 5, 9, 3, 2, 9
b) 3, 7, 9, 4, 3, 5, 9, 3, 2, 9
c) 1, 2, 3, 4, 5, 6, 7, 8, 9

Solution

a) Order the scores from lowest to highest: 1, 2, 3, 3, 4, 5, 7, 9, 9, 9
It is now easy to see which score is the mode:
The score that occurs ***most***, is 9, the mode of the data set.

b) Order the scores from lowest to highest: 2, 3, 3, 3, 4, 5, 7, 9, 9, 9
Both 3 and 9 occur three times in the data set, therefore the data has ***two modes***, 3 and 9.
You can also say this data are ***bi-modal***.

c) The data is already ordered: 1, 2, 3, 4, 5, 6, 7, 8, 9
All the numbers occur only once in the data.
This means that this data has ***no mode***.

Exercise 3.1

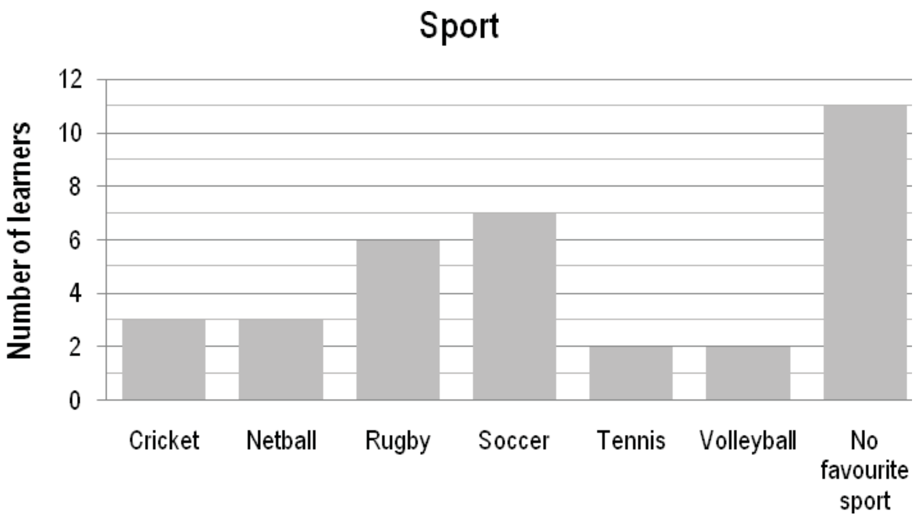
A random sample of ten learners was selected from the 2009 Census@School data base.

1) The following data set contains the ages of ten learners in years:
9, 15, 9, 15, 17, 15, 11, 18, 15, 19
   a) Order the ages from the smallest to the biggest.
   b) How many learners are 9 years old?
   c) What is the highest age?
   d) How many learners are older than 15?
   e) How many learners are 15 years old?
   f) Which age occurs most often?
   g) What is the mode?

2) The following data set contains the heights of the ten learners in centimetres.
138, 161, 121, 170, 170, 165, 142, 160, 140, 182
   a) Rank the heights from the lowest to the highest.
   b) Which is the smallest height?
   c) How many learners are taller than 160 centimetres?
   d) How many learners are 170 centimetres tall?
   e) What height is the mode?

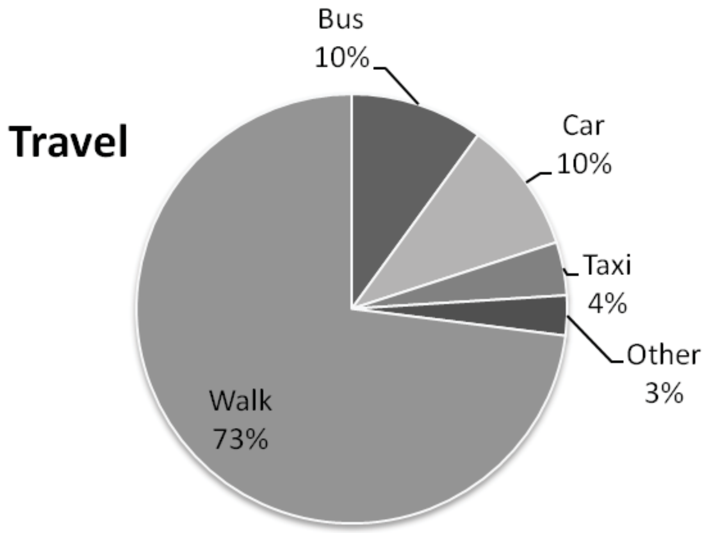
3) Learners were asked how many people lived in their home. The data set below shows their answers: 10, 14, 6, 8, 3, 2, 6, 7, 7, 4
   a) Rank the answers from smallest to largest.
   b) How many learners are in the data set?
   c) What is the fewest number of people in a home?
   d) What is the most number of people in a home?
   e) How many homes have 6 people?
   f) How many homes have 7 people?
   g) What is the mode of the people in the home?
   h) If you only look at the first 7 learners in the data set, what will the mode be?
   i) If you only look at the first 5 learners in the data set, what will the mode be?

4) Grade 9 learners had to say which sport is their favourite sport at school. The frequency of the learners who selected a particular sport is represented in the bar graph below. Study the graph and answer the questions.



   a) How many learners played cricket?
   b) How many learners played volleyball?
   c) Which category was selected by most learners?
   d) Which sport or sports were chosen by the fewest learners?
   e) What is the mode category indicated by the learners?
   f) How many learners took part in the survey?
   g) If the category, “No favourite sport” is ignored, what sport will be the mode?
   h) How many learners took part in the modal sport in question g)?

5) A sample of 174 learners answered the question: "How do you travel to school?" The percentages of learners using different ways of travelling to school are shown in the graph below.



a) What method of travelling to school is the mode?

b) How many of the learners travelled to school by bus?

c) How many of the learners travelled to school by taxi?

Mean

- ✓ The mean is also known as the arithmetic mean or average.
- ✓ To determine the mean you add all the scores together and then divide the sum by the number (n) of scores.
- ✓ We can use the formula: $\textbf{Mean} = \dfrac{\textbf{sum of all the scores}}{\textbf{number of scores}}$ to calculate the mean.
- ✓ The "sum of all the numbers" is calculated by adding all the numbers together.

Example 3.2

Find the mean of the following set of scores:
1, 7, 9, 4, 3, 5, 9, 3, 2, 8

Solution

The sum of all the scores = $1+7+9+4+3+5+9+3+2+8=50$

There are 10 scores in the data set, therefore to calculate the mean, you divide the sum (50) by the number of scores or numbers (10).

$$\textbf{Mean} = \frac{\textbf{sum of all the scores}}{\textbf{number of scores}} = \frac{50}{10} = 5$$

Exercise 3.2

1) The following data set consists of the ages (in years) of 5 learners:

   9, 15, 11, 18, 19

   a) What is the sum of the ages of the learners?
   b) What is mean of the ages?

2) The data set below contains the ages of a few learners in years.

   9, 15, 9, 15, 17, 15, 11, 18, 15, 19

   a) The ages of how many learners are in this data set?
   b) What is the sum of the ages of the learners?
   c) What is mean of the ages?
   d) What is the mean of the first 5 ages?

3) The following data set contains the heights of learners in centimetres.

   138, 161, 121, 170, 170, 165, 142, 160, 140

   a) What is the mean height of the learners in this data set? Give your answer correct to 2 decimal places.
   b) What is the mean height of the first 6 learners in the data set?

Median

- ✓ The median of a data set is the number right in the middle of an ordered data set.
- ✓ If there is an odd number of scores, the median is the middle number.
- ✓ If there is an even number of scores, then two numbers will be the middle numbers. In that case, you add the two numbers together and divide by 2.

Example 3.3

What is the median of each of the following sets of numbers?

a) 1, 7, 9, 4, 3, 5, 8, 3, 2
b) 2, 5, 7, 9, 10, 4, 12, 1, 15, 3

Solution

a) First order (or rank) the numbers in the data set from the smallest to the greatest number: 1, 2, 3, 3, 4, 5, 7, 8, 9
There are 9 items in the data set.
The middle item is the number 4.
So the median = 4

b) First order (or rank) the scores from the smallest to the greatest number.
The data set looks as follows: 1, 2, 3, 4, 5, 7, 9, 10, 12, 15
There are 10 scores in this data set.
There is not only one score in the middle, but there are two, namely 5 and 7.
In this case, you add the two numbers in the middle and divide the sum by two (2) to calculate the median:

Median = $\frac{5+7}{2} = \frac{12}{2} = 6$

NOTE: The median of this data set is 6, although the 6 did not actually appear in the data set.

Exercise 3.3

1) The following data set contains the ages, in years, of a few learners.
   9, 15, 11, 18, 19
   a) How many ages of learners are in this data set?
   b) Is the number of ages an odd or an even number?
   c) Rank the ages from the smallest to the biggest.
   d) What is the median of the ages?

2) The data set below contains the ages of a few learners.
   9, 15, 9, 15, 17, 15, 11, 18, 15, 19
   a) Describe all the steps that you would take to calculate the median of the ages.
   b) What is median of the ages?

3) The following data set contains the heights of Grade 9 learners in metres.
   1,81; 1,53; 1,6; 1,5; 1,28; 1,65; 1,75; 1,58; 1,13; 1,68; 1,77
   a) Rank the heights of the learners from the smallest to tallest.
   b) How many heights of learners are in this data set?
   c) What is the median height in metres of the learners?

4) Grade 8 learners were asked how long it took them to travel to school. The following data set contains these times (in hours).
   0,50; 0,58; 0,67; 0,75; 0,50; 0,83; 1,00; 1,33; 0,58
   a) Rank the travelling times from the shortest to the longest.
   b) How many travelling times are in this data set?
   c) What is the travelling time of the
      (i) 3^{rd} learner?
      (ii) 7^{th} learner?
      (iii) 8^{th} learner?
   d) Which learner(s) travelled 30 minutes to get to school?
   e) What is the position of the learner who travels 0,75 hours to get to school?
   f) What is the position of the modal travelling time?
   g) What is the modal travelling time (in hours)?

5) a) A data set has 121 ordered data values. In which position will the median be?

b) A data set has 502 ordered data values. Explain how you will determine the median's position and how you will calculate the median.

6) Thembi makes the following statement. "The median of a data set with 40 ordered data values is in the 20^{th} position."
Is this a valid statement? Explain your answer.

- ✓ Do you know the song, 'Row, row, row your boat'?
  Use the tune to try to remember the difference between mode, mean and median with the following song:

  Mode, mode, mode is most,
  Average is the mean,
  Median, median, median, median, always in between!

- ✓ If you do not know the song or the melody, write a song yourself to assist you in remembering the difference between the three measures of central tendency.

- ✓ If you have internet available, see also:
  http://www.youtube.com/watch?v=oNdVynH6hcY&NR=1

Measures of dispersion or spread

- ✓ Sometimes we need to know how spread out data is.
- ✓ For this we need to know two other statistics, the ***range*** and ***the extremes***. Both of these are called measures of dispersion (or measures of spread).

Range

The difference between the highest value and the lowest value in a data set is called the range.

Range = highest value – lowest value

Example 3.4

Find the range of the following ranked data set:
1, 2, 3, 4, 5, 7, 9, 10, 12, 15

<u>Solution</u>

The lowest value is 1 and the highest value is 15.
So the Range = highest value – lowest value
= 15 – 1
= 14

Exercise 3.4

1) The following data set contains the ages of a few learners: 9, 15, 11, 18, 19
   a) Write the ages in order.
   b) What is the highest age?
   c) What is the lowest age?
   d) What is the difference between the highest age and the lowest age?
   e) What is the range of this data set?

2) The following data set contains the ages of a few learners.
   9, 15, 9, 15, 17, 15, 11, 18, 15
   a) What is the highest age?
   b) What is the lowest age?
   c) What is the range of this data set?

3) The data set below contains the heights of learners in centimetres.
   138, 161, 121, 170, 170, 165, 142, 160, 140, 182
   a) What is the height of the tallest learner in this data set?
   b) What is the height the shortest learner in this data set?
   c) What is the range of the heights in the data set?

4) Learners were asked how many minutes they travelled to school every day. They answered as follows: 10, 45, 11, 5, 20, 10, 5, 24
   a) What is the shortest travelling time?
   b) What is the longest travelling time?
   c) What is the range of travelling times?

Extreme values

✓ Extreme values are unexpected large or small values in a data set.

Example 3.5

The lengths (in centimetres) of the right feet of a group of 8-year old learners are: 13, 18, 18, 19, 20, 20, 20, 21, 22, 31
What do you notice about this data set?

<u>Solution</u>

If you look at the foot lengths carefully, you can see that the lengths of the middle eight of the foot lengths range from 18 cm to 22 cm.

13, 18, 18, 19, 20, 20, 20, 21, 22, 31

→ The first learner has a very small right foot of only 13 cm.

→ The largest right foot is 31 cm long.

→ The very small value and the very big value are unexpected. They are called ***extreme values***.

Exercise 3.5

1) A group of learners was asked to say how many minutes it took them to travel to school every day.

   They answered as follows: 45, 5, 20, 50, 13, 15, 18, 15, 90, 5, 15, 25, 15, 4

   a) How many values are in the data set?

   b) Rank the travelling times.

   c) Identify the extreme value(s) and say why you think they are extreme.

2) In the Census@School survey, learners had to say how many children in their home were still at school. The results are given in the following table:

Household	Boys	Girls
1	1	1
2	1	1
3	3	1
4	1	2
5	1	1
6	2	0
7	1	2
8	2	1
9	0	0
10	0	1
11	1	2
12	12	1
13	1	3
14	3	3
15	1	4

   a) Which household had the highest number of school-going boys?

   b) Which household had the highest number of school-going girls?

   c) (i) Which household has the fewest school-going children?
      (ii) How many children are school-going in this household?
      (iii) How did you determine your answer?
      (iv) Is this answer possible?

   d) (i) Which household has the most school-going children?
      (ii) How many children are school-going in this household?
      (iv) Is this answer possible?

   e) Are there any extreme values among the boys? Why do you say so?

   f) Are there any extreme values among the girls? Why do you say so?

3) The heights (in centimetres) of Grade 9 learners are given in the table below:

169	181	145	159	160
171	165	109	168	170
173	176	140	178	155
150	170	162	146	151

a) Rank the heights from shortest to tallest.

b) How many heights are in the data set?

c) What is the modal height?

d) What is the median of the heights?

e) What is the mean of the heights of the learners?

f) What is the range of the heights?

g) Identify an extreme value in the data.

h) Delete the extreme value from the data and recalculate the mean of the data to 1 decimal place. What do you notice about the original mean and the new mean?

i) Did the mode and the median change when you deleted the extreme value?

4) Explain the difference between a data set with no mode and a data set with a mode of 0? Give your own example.

Chapter 4

Representing Data

In this chapter you will:

- *Draw graphs using pictograms*
- *Learn to draw 3 different types of bar graphs:*
- *Learn to draw Pie Charts, Histograms, Broken Line Graphs and Scatter plots*
- *Learn about scales on a graph*
- *Learn to choose an appropriate graph to represent given a set of data.*

Graphs are often used to display data. Many people find graphs simpler to understand than a table. They are also more attractive and interesting to look at. A graph can show data clearly without lots of words or figures. This helps you to see any patterns and to compare things easily.

PICTOGRAPHS

NOTE

A pictograph uses simple pictures or symbols to show data

A pictograph gives you a quick impression of the given information.

Pictographs are often used in newspapers, magazines, books and on television because comparing data in a pictograph is easy; just compare how many pictures each item has.

Reading information off a pictograph

EXAMPLE 4.1

The 2009 Census@School was run at Anele's school.
The results were analysed and it was found that the favourite sport of the boys in her school was soccer, rugby, cricket and athletics.

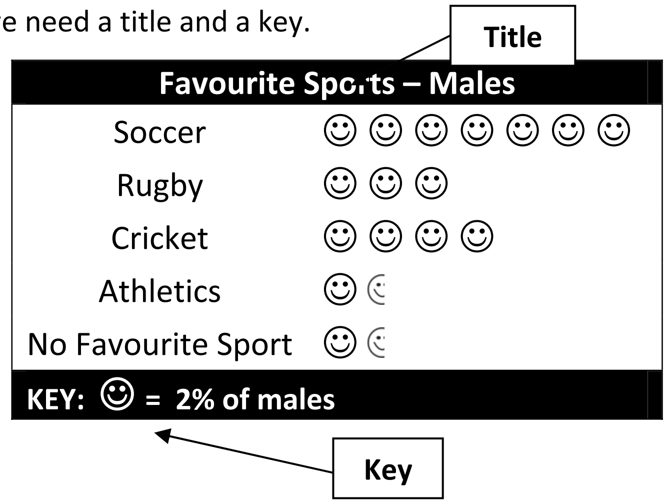
Anele drew up a table to show the percentage of the boys that liked each sport.

Draw a pictograph to represent this data

Favourite Sports – Males	Percentage
Soccer	14
Rugby	6
Cricket	7
Athletics	3
No favourite Sport	3

SOLUTION:

In order to draw a pictograph, we need a title and a key.



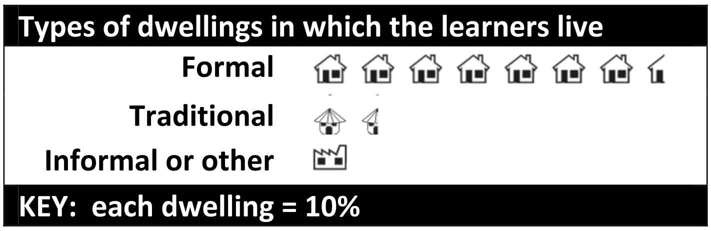
- The title tells you what the pictograph is about.
- The key shows you what each little picture "stands for" or "represents".
- If each ☺ = 2% of the males, then
  ☺ (half) = half of 2%
  = 1% of males
- So to represent 7% who like cricket, we use three full pictures and 1 half picture like this: Cricket ☺ ☺ ☺ ☺(half)

Drawing a pictograph

- Choose a simple picture or symbol that is easy to draw. Try to make it 'look like' the data in some way.
- Always give a key.
  - ✓ You must say what each picture stands for.
  - ✓ Each picture may represent more than 1 piece of data. Choose an easy number to work with.
  - ✓ Part of a picture represents part (or a fraction) of that number
- Draw your pictograph on squared paper. This will help you to keep the pictures in line.
  - ✓ Try to draw each picture the same size and space them out evenly.
  - ✓ Estimate fractions of a picture where necessary
- You could also use different pictures for different kinds of data.

EXAMPLE 4.2

The Census@School questionnaire also asks learners about the type of dwelling or house in which they live. Anele drew a pictograph to show the number of learners in her school living in each type of dwelling.



a) What percentage of the learners lives in Formal Dwellings?
b) What percentage lives in Traditional Dwellings?

<u>SOLUTION</u>

a) The pictograph uses 7 full houses plus half a house to represent the percentage of the learners living in Formal Dwellings.
This means that 10% + 10% + 10% + 10% + 10% + 10% + 10% + 5%
= 75% of learners live in Formal Dwellings.

b) It also uses one full house plus half a house to represent the percentage of the learners living in Traditional Dwellings.
This means that 10% + 5% = 15% of the learners live in Traditional Dwellings.

EXERCISE 4.1

1) Vonani is doing a Travel and Tourism project. He does a survey of 100 people to find out where they like to go on holiday. The table shows the result of his survey.

Place	**Number of people**
The sea	56
The mountains	12
The game reserve	24
Other	8

Draw a pictograph to show the results.

2) The following results were obtained in David's school after the Census@School was run at his school. Draw a pictograph to show the results.

Favourite Subject Boys in Grades 3 to 7	**Percentage**
Maths	15
Literacy	6
Languages	7
Technology	3
Numeracy	3
Arts & Culture	5
Life Orientation	3

Pictographs have two main disadvantages

1) They can take a long time to draw well.
   - ✓ So, unless you are spending time on a special project, use simple symbols that are quick and easy to draw.

2) They are not always a very accurate way of showing data because:
   - ✓ the data is often simplified before drawing;
   - ✓ fractions of pictures are estimated; etc.

Bar Graphs

NOTE

A bar graph uses bars to display data.

The length of the bar stands for the size of the data it represents. This makes the data easy to compare. Just compare the lengths of the bars. The bars can be drawn horizontally or vertically, and gaps are left between the bars.

Example 4.3

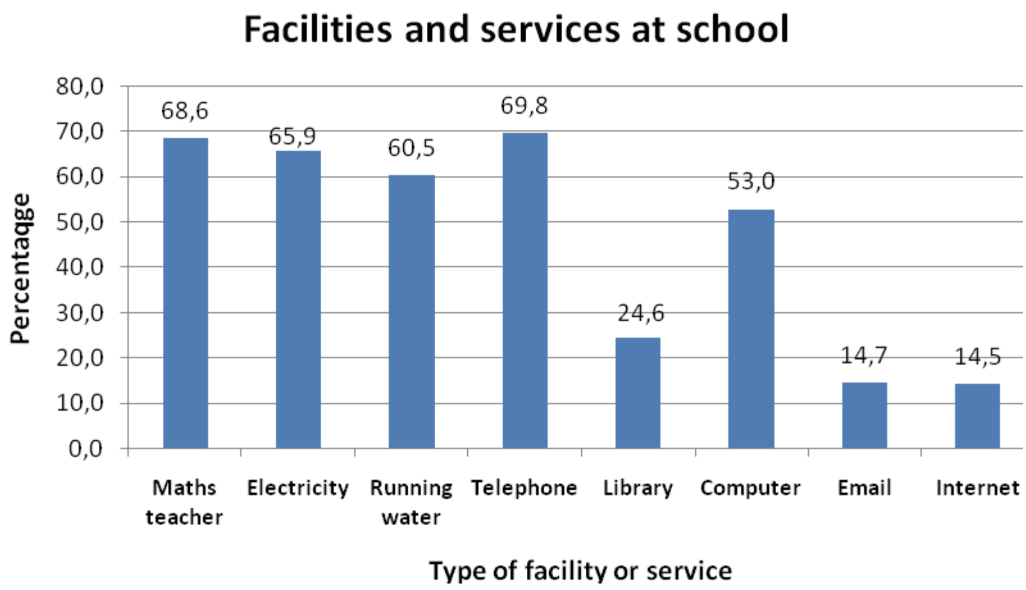
Draw a bar graph to illustrate the facilities and services at the schools that took part in the 2009 Census@School

Facilities and services at school	Percentage
Maths teacher	68,6
Electricity	65,9
Running water	60,5
Telephone	69,8
Library	24,6
Computer	53,0
Email	14,7
Internet	14,5

SOLUTION:

- First, you need a **title for the graph.**
  For this graph the title is "***Facilities and services at school***".
- Next you need **two axes and a label for each axis.**
  In this example we want a vertical bar graph, so the bars will go up the page. Since the percentage values will indicate the height of the graph, this will be on the vertical axis.
  - Vertical axis – Percentage
  - Horizontal axis –Type of facility or service.
- Determine the **scale on the vertical axis**.
  The highest percentage value is 68,6 %, so you can choose the vertical axis to go up to 80. The interval of the scale will be 10 units.
- Finally, **draw in each bar** corresponding to the value of the category.
  So, for the "maths teacher" category we draw a graph that is tall enough to represent 68,6%.
  Keep in mind that often we have to approximate the height of the bar since some scales do not make drawing exact heights possible.

The final graph is on the next page.


Facilities and services at school
Percentage
80,0
70,0
60,0
50,0
40,0
30,0
20,0
10,0
0,0
68,6
65,9
60,5
69,8
24,6
53,0
14,7
14,5
Maths teacher
Electricity
Running water
Telephone
Library
Computer
Email
Internet
Type of facility or service

Select the starting point on the vertical axis carefully

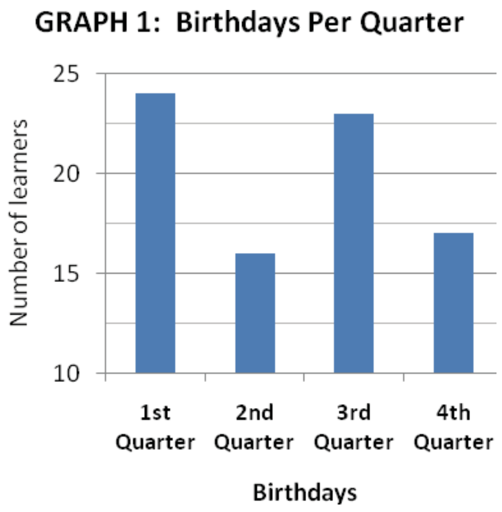
When drawing a bar graph, the scale on the vertical axis has to be selected carefully as it can affect how you interpret the graph.

→ GRAPH 1 and GRAPH 2 contain the same information. They show the birthdays per quarter of a random sample of 80 learners in Grade 8 in Kgomotso's school.

→ Look at how the scale of the vertical axis makes a difference in the way you interpret the information displayed on the graphs.

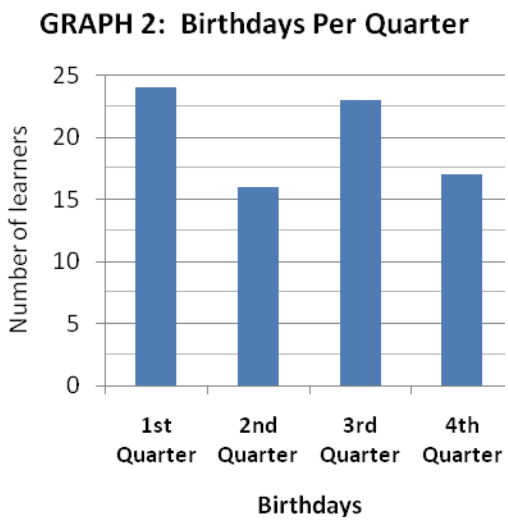
Can you see that in Graph 1 ...

- The vertical scale begins at 10 and ends at 30.
  - ✓ This is because the range of values being graphed is from 16 to 24.
  - ✓ Since we want our smallest bar to have some height, we start the scale at 10.
- It looks like there 2-times as many learners have a birthday in the 1st term than have a birthday in the 4th term.
  - ✓ This is not true because 24 learners have a birthday in the 1st quarter and 17 learners have a birthday in the 4th term.



Can you see that in Graph 2 ...

- The frequency scale begins at zero.
- It is clear that the bars for 1st quarter and the 3rd quarter have nearly the same height.
- The difference between the four quarters does not seem as big as it does in GRAPH 1.

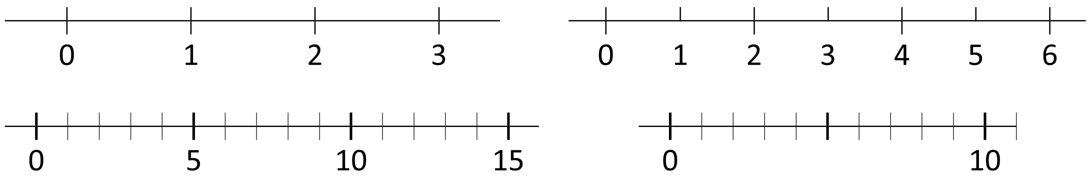


NOTE: Direct comparison can only be made when the scale begins at zero.

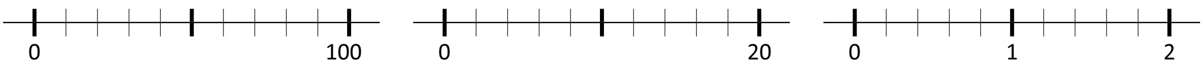
Select the intervals on the vertical axis carefully

✓ When drawing the graph, the intervals (the distance between the marks) on the vertical axis have also have to be selected asthey can affect how you interpret a graph.

- On some bar graphs the scale is very easy. Each 'small space' stands for 1 unit.



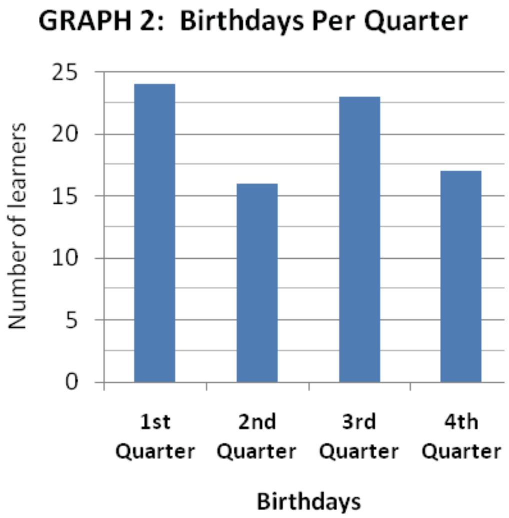
- Some scales are not so easy. Each small space may stand for more than 1, or for a fraction

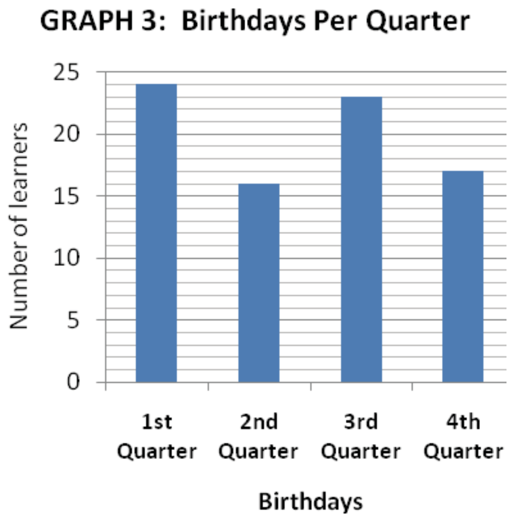


10 spaces → 100	10 spaces → 20	5 spaces → 1
1 space → 100 ÷ 10 = 10	1 space → 20 ÷ 10 = 2	1 space → 1 ÷ 5 = 0,2

→ Changing the interval of the scale also affects the way a bar graph looks.

→ Look again at GRAPH 2 and at GRAPH 3 showing different intervals on the vertical axis ...





2 spaces → 5 units	5 spaces → 5 units
1 space → 5 ÷ 2 = 2,5	1 space → 1 unit
So, each interval is 2,5 units	So, each interval is 1 unit

Exercise 4.2

1) 200 people were asked which their favourite TV channel was. The information is given in the following table. Draw a bar graph to illustrate this information.

SABC1	**78**
SABC 2	44
SABC 3	75
MNET	33
e TV	23

2) 180 learners who took part in the 2009 Census@School were randomly selected from the sample of schools. The following table was drawn showing the provinces in which they were born. Draw a bar graph to illustrate this information

PROVINCE	NUMBER OF LEARNERS
Gauteng	39
Western Cape	25
Mpumalanga	25
Kwa-Zulu Natal	22
Northern Province	19
North West	14
Northern Cape	13
Free State	12
Eastern Cape	9
Outside of South Africa	2

Compound Bar Graphs

NOTE

A compound bar graph can be used to compare different sets of data.

When data is compared, either a compound bar graph or sectional bar graph can be used.

→ If you look at separate bar graphs, it is not very easy to compare them.

→ We can combine two bar graphs and put the bars next to each other. This makes it much easier to compare the two data sets.

EXAMPLE 4.4

The 2009 Census@School asked learners what their favourite sport was. The following table gives the percentage girls and percentage boys who reported that the following sports were their favourites.

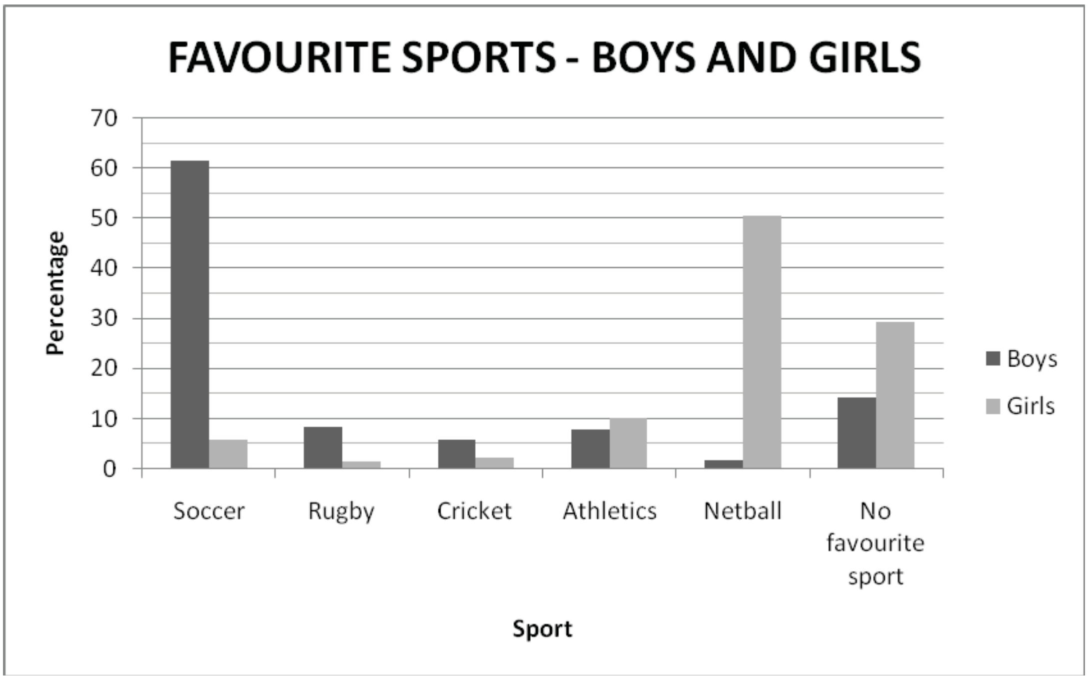
Draw a compound bar graph to represent this information.

FAVOURITE SPORTS	Percentage of the boys	Percentage of the girls
Soccer	61,6	6,0
Rugby	8,4	1,5
Cricket	6,0	2,3
Athletics	7,8	10,2
Netball	1,8	50,5
No favourite Sport	14,4	29,5

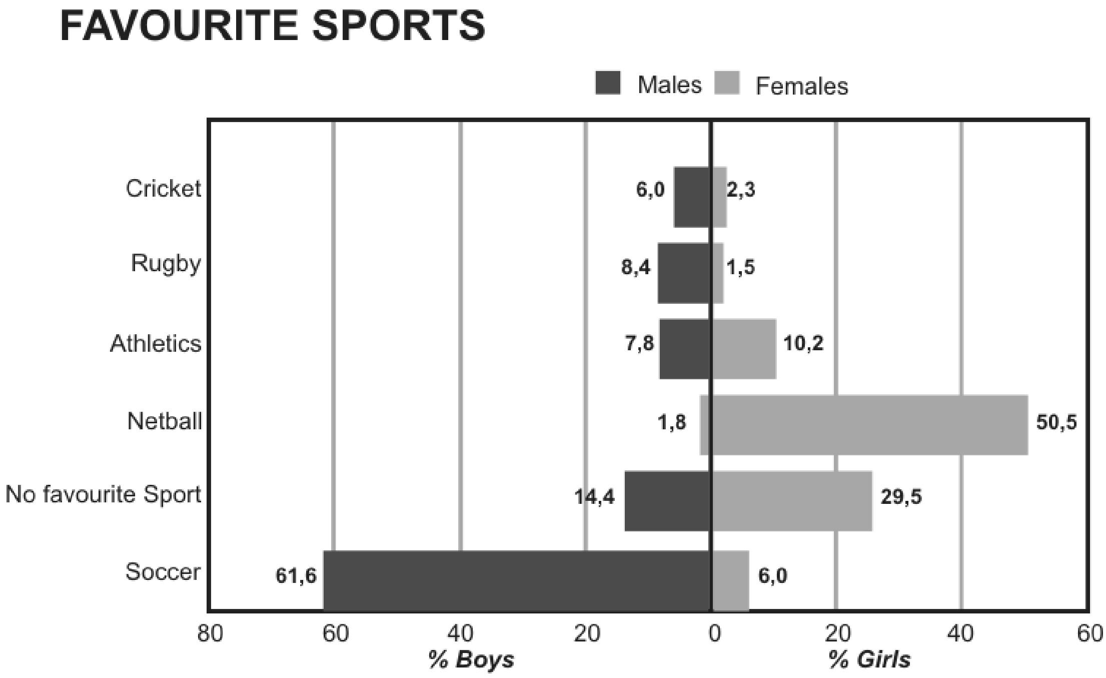
Solution:

- Draw two axes at right angles to one another and chose a scale.
- For each sport draw a bar for the boys and a bar for the girls.
  → So for soccer, first draw a bar with height 61,6 units for the boys.
  → Then draw a second bar next to it without leaving a gap with a height of 6 units for the girls. Make this bar a different colour from the boys bar.
- Similarly, draw the sets of bars for each of the other sport categories.

The final graph is on the next page ...



Here is another type of compound bar graph representing the data in Example 4.4:



- ✓ Notice that the bars are not drawn alongside each other, but instead are drawn on both sides of the vertical axis. This is useful because it helps to make a quick visual comparison across each data category (sport).

- ✓ You can see that 61,6% of boys like soccer, in comparison with just 6% of the girls. In contrast the netball seems to be the most popular sport for girls; 50,5% of the girls chose this as their favourite, while only 0,8 % of the boys chose netball.

Exercise 4.3

Study the table showing the differences in the access to goods and services between the learners in the 2001 Census@School and the 2009 Census@School.

Types of goods & services at home	**Census@School 2001**	**Census @School 2009**
Telephone	47,6	30
Running water	55,3	55,3
Television	83,5	85,8
Radio	94,2	90,2

Draw a compound bar graph to show the percentage of learners' access to goods and services in the 2001 Census@School and the 2009 Census@School.

Sectional or Stacked Bar Graphs

NOTE

A sectional bar graph consists of one bar on top of another.

We use sectional bar graphs when we have two, or more, different sets of information on the same topic. They are particularly useful when we are interested in the total of the two of more bars.

EXAMPLE 4.5

A random sample of learners was taken from the results of the 2009 Census@School, and the following table was drawn.

Draw a sectional bar graph to represent the information.

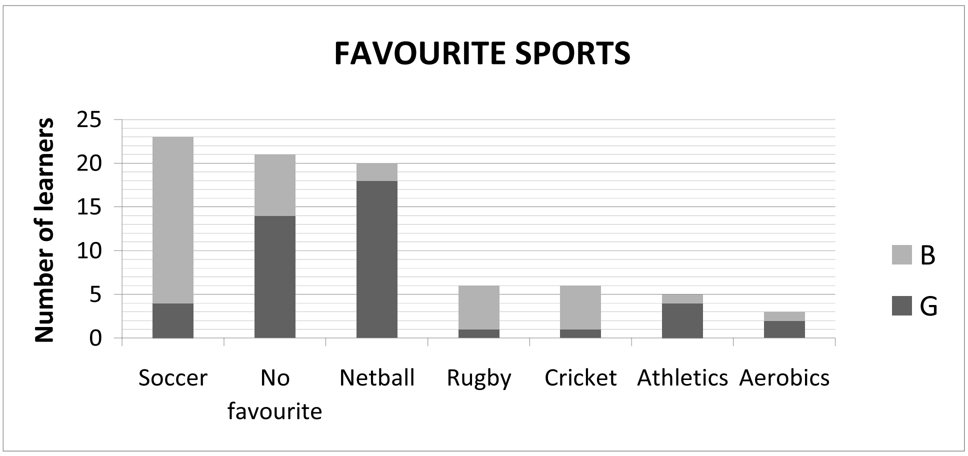
Favourite Sport	Soccer / Football	No Favourite	Netball	Rugby	Cricket	Athletics	Aerobics
Girls	4	14	18	1	1	4	2
Boys	19	7	2	5	5	1	1

Solution:

i. Use the data to draw the bars for the girls just as you would draw a single bar graph.
   → So, for soccer you would draw a bar with a length of 4 units.

ii. Draw all the rest of the bars to represent the girls.

iii. Now, for each category, draw the bar for the boys directly above the bars for the girls, touching the bar for the girls. Shade this in a different colour from the girls' bar.
   → This means that for soccer, you would draw a bar of length 19 units above the girls' bar you drew earlier. So the total height for bar for the soccer category will be 4 + 19 = 23 units.



Exercise 4.4

In the 2009 Census@School, 600 learners from Grades 3 to 7 were asked which subject was their favourite.

The results were as follows:

	females	*males*
Life Orientation	54	38
Languages	154	114
Technology	18	34
Natural Science	26	26
Maths	64	72

1) Draw a suitable sectional bar graph to illustrate the data.

2) Write down two conclusions about this data.

Pie Charts

NOTE

A pie chart is a circular diagram used to display data.

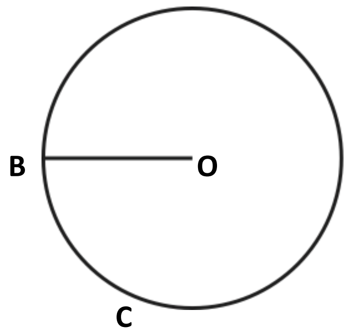
The whole circle stands for the whole amount of data being dealt with.

Each slice stands for part of the data. Its size represents the size of that part of the data.

- ✓ A pie chart is particularly suitable if you want to illustrate how the "whole" of some data is divided up into different parts, and what portion of the whole each part represents.
- ✓ We can write this portion as a fraction, as a decimal fraction, or as a percentage of the whole.

Drawing a Pie Chart

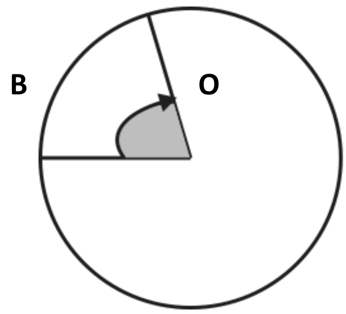
1) Draw circle centre O with a pair of compasses.
(Do not make your circle too small)



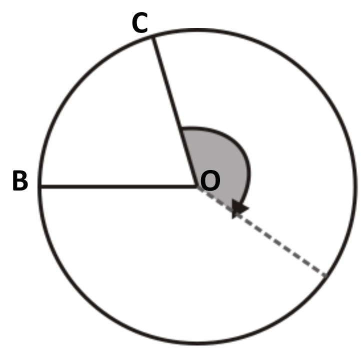
2) Draw radius OB as a starting line for measuring angles.
→ **Remember** that a radius is a straight line from the centre of the circle to outside of the circle (or circumference of the circle).

3) Use a protractor to draw the angle of each sector at the centre of the circle.

→ Measure the first angle from radius BO, by placing the centre of the protractor at the centre of the circle.



4) Measure the second angle from radius OC, then measure the next angle from the new line, and so on.
→ Measure the size of each angle very carefully. Any mistakes will mean that the sectors will not fit into the circle.
→ **Hint**: Draw the smallest angle first, then the next biggest, and so on.



5) Label each slice carefully. If it is difficult to fit the full name of each group on each slice, label each with a letter and use a key to say what each stands for.

6) Give the pie graph a suitable title.

How to work out the angles

1) Look at the given data, and decide what the "whole" or the total amount to be shown on the pie graph (e.g. 24 hours) is going to be.
2) Find out how many different parts or sectors there are.
3) Calculate the angle at the centre of the circle for each of the sectors of the pie graph. Do this by dividing 360º by the "whole".
   For example:
   If we want to show how many hours in a day is spent on different activities, our calculation is $360° \div 24 = 15°$
4) Multiply this answer by the number of items
   For example: the angle for 3 hours is $3 \times 15° = 45°$
5) Always check that all the angles add up to 360°, but remember that when angles are rounded off to the nearest degree, the total may not be exactly 360°.

EXAMPLE 4.6

The eye colours of a random sample of 120 learners who took part in the 2009 Census@School were as follows. Draw a pie chart to display this data.

Eye colour	Frequency
Blue	2
Brown	104
Green	4
Other	10
TOTAL	120

Solution:

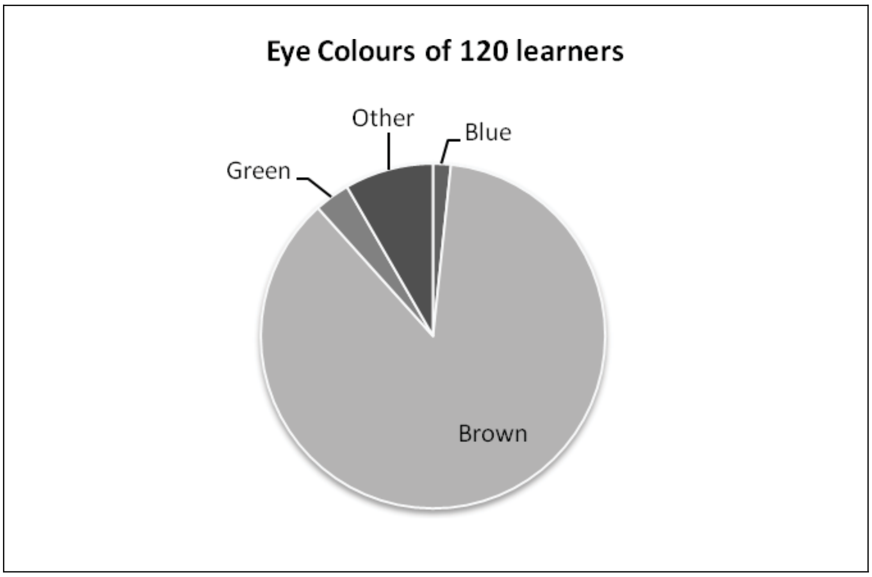
Calculate the angles of the pie chart as follows

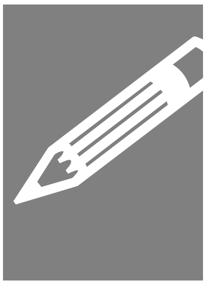
1) The "whole" is 120 students
2) There are 4 different parts; that is, 4 eye colours.
3) Divide 360° by the "whole" : $360° \div 120 = 3°$
4) Multiply this answer by the number of items for each part:

Eye colour	Frequency	Angle
Blue	2	$2 \times 3° = 6°$
Brown	104	$104 \times 3° = 312°$
Green	4	$4 \times 3° = 12°$
Other	10	$10 \times 3 = 30°$
TOTAL	**120**	**360°**

Notice that all the angles add up to 360°

The completed graph is shown on the next page ...





EXERCISE 4.5

1) Thirty learners were asked how they travel to school, and the results were recorded in the given table.

a) Copy the table
b) Work out the angle at the centre of the circle for each learner by calculating 360° ÷ 30
c) Fill in the rest of the table
d) Draw a pie chart to show this information

Method of travel	Number of learners	Angle
Walk	14	14 x 12° = …………….
Bus	7	
Car	6	
Bike	3	
Helicopter	0	
TOTAL:	**30**	

2. 90 people were asked which month they were born in. Here are the results:

Month	No of people	Angle
January	7	
February	4	
March	9	
April	8	
May	7	
June	6	

Month	No of people	Angle
July	6	
August	8	
September	11	
October	7	
November	10	
December	7	

a) Copy the table

b) Work out the angle for one person (i.e. 360° ÷ 90 = ……………)

c) Work out the angle for each month and write it on the table.

d) Draw a pie chart to show this information.

3. A survey was done of 120 learners from Grade 8, to find out what their favourite subject was at school.

It was found that 30 prefer History, 40 prefer Geography and 50 prefer Maths.

a) Illustrate this information by drawing:

i) A table to illustrate this data.

ii) A bar chart

iii) A pie chart

b) Which one of the two graphs do you think represents this information better? Give a reason for your answer.

Histograms

NOTE

A histogram is a graph of grouped data and has no gaps between the bars.

Bars are drawn corresponding in height to the frequency of each group.

The intervals or groups are shown along the horizontal axis
The frequency (or how often something happens) is shown along the vertical axis.

EXAMPLE 4.7

The heights of the heights of 150 learners in Grade 7 at Makhosi's school are recorded in the following table.

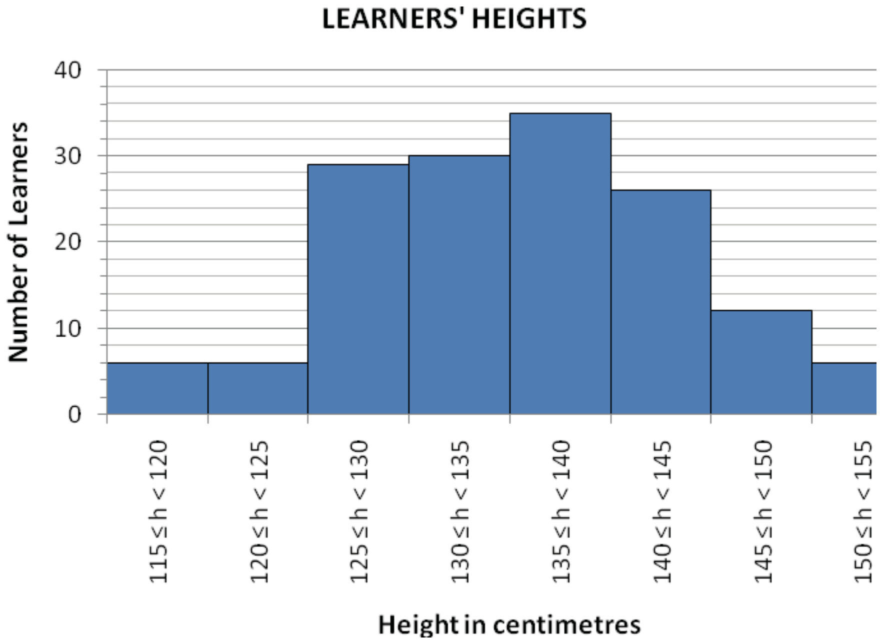
Draw a histogram to display this data.

Height (cm)	Frequency / number of learners
$115 \le h < 120$	6
$120 \le h < 125$	6
$125 \le h < 130$	29
$130 \le h < 135$	30
$135 \le h < 140$	35
$140 \le h < 145$	26
$145 \le h < 150$	12
$150 \le h < 150$	6

Solution:

- Draw the x-axis. Label this axis "Heights of learners".
  - → In this example, the x-axis would have 8 intervals since the table has 8 intervals.
  - → Use the squares on the graph paper to make each interval the same size.
- Draw the y-axis. This will show the frequency.
  - → Label it "Number of learners".
- Draw bars on the histogram with heights corresponding to the frequency of that interval.
  - → So the first interval; $115 \le h < 120$ will have a height of 6 units.
  - → Draw in the rest of the bars.
  - → The histogram bars are drawn next to and touch each other.
- Add a title to complete the histogram.

The required histogram is drawn on the next page…



EXERCISE 4.6

1) In the 2009 Census@School, the learners from Grades 3 to 7 at Ruth's school were asked how long in minutes, it takes for them to travel to school.

   The table shows the results from a sample of 150 learners.

   Draw a histogram to illustrate this data.

Time in minutes	frequency
$0 \le m < 10$	45
10 ≤ m < 20	48
$20 \le m < 30$	35
$30 \le m < 40$	14
$40 \le m < 50$	6
$50 \le m < 60$	1
$60 \le m < 70$	0
$70 \le m < 80$	1

2) The same group of 150 learners were asked to fill in the distance they travelled to school. The table shows the results.

   Draw a histogram to illustrate the data.

Distance in km	frequency
$0 \le d < 2$	72
$2 \le d < 3$	48
$3 \le d < 4$	18
$4 \le d < 5$	12

Scatter Plots

NOTE

A scatter plot is used to display two sets of data in order to find a relationship between them.

Each axis represents a different variable.

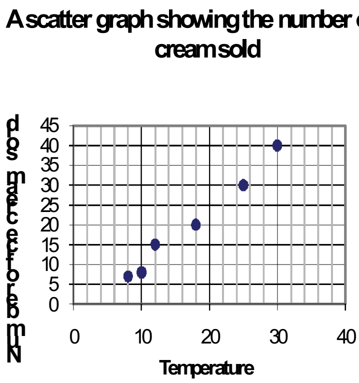
We plot values of one quantity against corresponding values of another quantity.

A scatter plot shows trends

For example ...

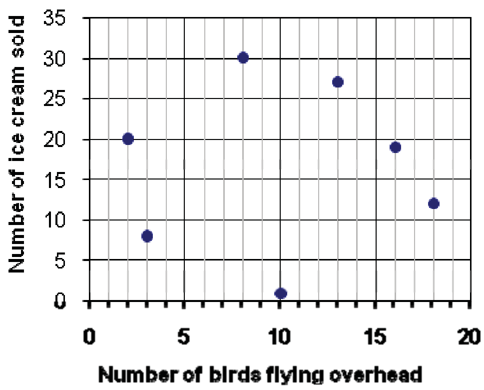
Joe has a job selling ice cream. When will he sell the most ice cream?

→ Joe says he will sell the most ice cream on a hot day and less ice cream on a cooler day.
→ He thinks he will hardly sell any ice cream on a very cold day.
→ Joe calls this common sense. In maths it is called a ***correlation.***



Notice: The points on the graph more or less form a straight line. We say there is a **positive linear correlation** between the number of ice creams sold and the temperature.

The following scatter graph shows no correlation between the points:



EXAMPLE 4.8

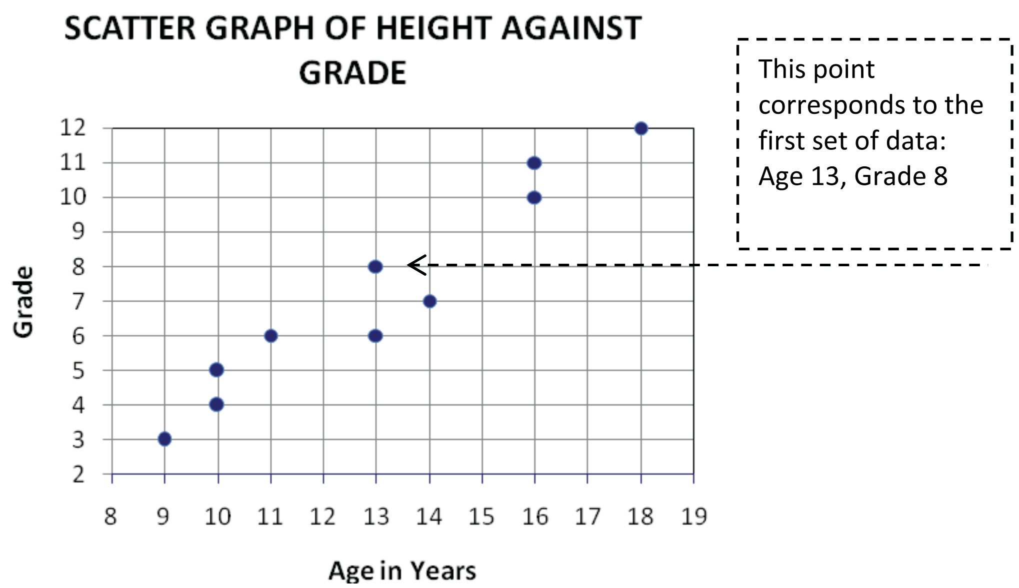
Use these results from a group of randomly selected learners who participated in the 2009 Census@School to draw a scatter plot to see if there is a correlation between age in years and grade number.

Age in years	Grade In school
13	8
16	10
16	11
11	6
14	7
13	6
18	12
10	4
9	3
10	5

Solution:

Draw the graph as follows:

- Draw two axes. Plot the age in years along the horizontal axis and the Grade along the vertical axis.
- Choose an appropriate scale for each axis.
  Here you can make both intervals 1 unit.
- Plot each set of points.
  So the first point is Age = 13 and Grade = 8.
  Similarly plot each set of point given. Here is the completed graph:

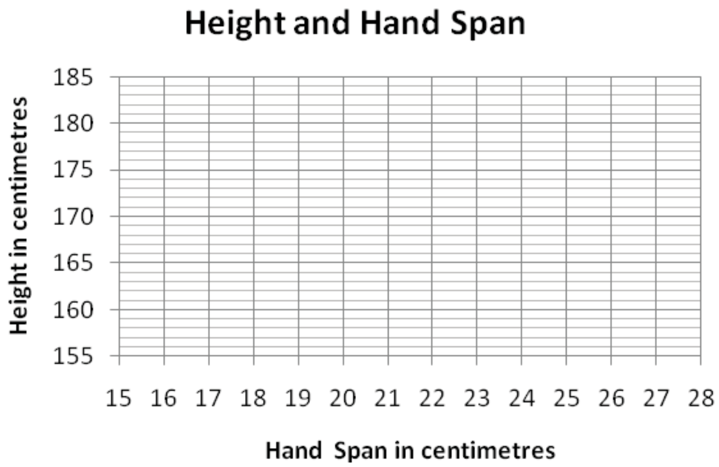


CONCLUSION: There is a correlation between the age of a learner and the grade they are in. As the age increases, the grade number increases.

Exercise 4.7

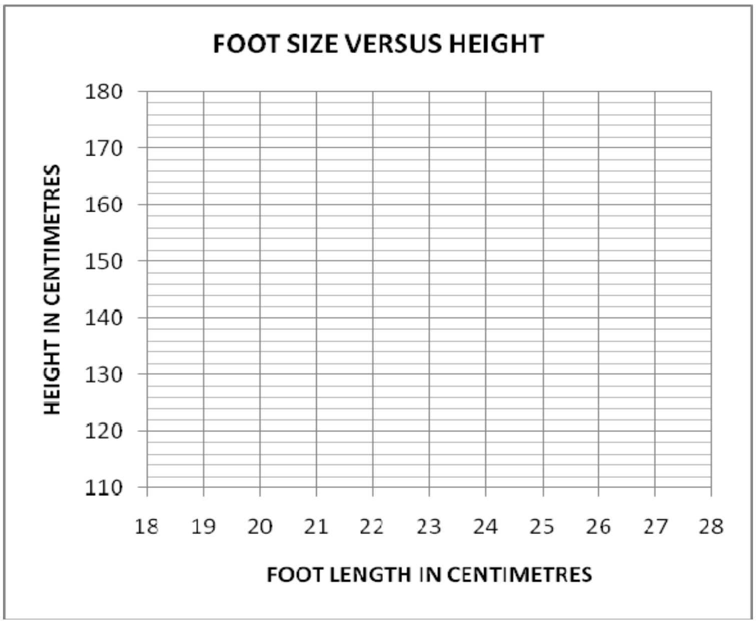
1) 12 learners measure their heights and hand spans and tabulated their results. Copy the given set of axes and draw a scatter plot to see if there is a relationship between height and hand span.

Hand span in cm	26	25	20	19	25	18	22	27	23	24	23
Height in cm	177	175	165	164	172	160	180	164	175	171	169



2) A group of 14 learners measure their heights and foot sizes and tabulated their results. Copy the given set of axes and draw a scatter plot to see whether foot size increases with height.

Foot (cm)	Height (cm)
27	174
24	158
24	163
26	175
23	140
23	126
22	153
25	160
24	160
20	131
23	156
25	165
24	158
20	120



Chapter 5

Interpreting Data

In this chapter you will:

- *Look at bar graphs, pie charts and line graphs and interpret the information.*
- *Consider how information on bar graphs, pie charts and line graphs can be misleading*
- *Know more about errors when collecting and interpreting data*

So far in this workbook you have looked at different ways to collect, organise and present information or data. Remember that the whole point of collecting data is to help you understand more about the world you live in. In particular, by using the Census@School data you find out about different schools and the learners who attend them.

Interpreting bar graphs/bar charts

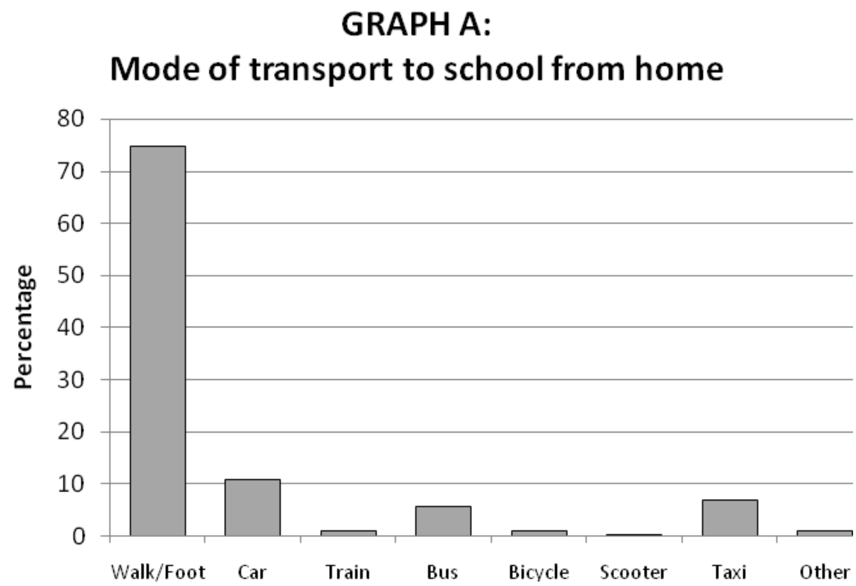
Bar graphs (which are also sometimes called bar charts) are very easy to read.

- ✓ The lengths of the bars stand for the size of the data.
- ✓ Gaps are usually left between each bar and the next.
- ✓ The bars can be horizontal or vertical.
- ✓ Remember that bar graphs are usually used to show **discrete data** (*data that can be counted)* as opposed to **continuous data** (*data that can be measured*).
- ✓ For every bar graph check the following points:
  1. **The title** – what is the bar graph about?
  2. **The axes** – check the labels of the axes. One axis gives the labels of the bars; the other tells you how many items in each bar.
  3. **The scale** – the scale on the number axis tells you how many of each thing has been counted.

Example 5.1

Study the bar graph below taken from the 2009 South African Census@School results. It shows the modes of transport to school from home of the learners who participated in the Census@School.

1) What is the graph about?
2) What information is shown on the axes?
3) Is the scale clear and easy to read?
4) What conclusions can the graph help you make?



Solution

1) In order to find out what the graph is about, we look at the title. The title tells us that the graph shows how learners get to school each day.

2) The axes are clearly labelled.
   The **horizontal axis** shows the different forms of transport: walk, car, train, bus, bicycle, scooter, taxi and other.
   The **vertical axis** shows percentage.

3) The scale is clear and easy to read.
   You cannot read exact percentages for the number of learners walking to school, but it is easy to estimate that nearly 80% (or over 70%) walk to school.

4) You can immediately see that **most** of the learners in the census **walk to school** from home. The other learners **mostly use car, bus or taxis to get to school.**

Misleading bar graphs

✓ Bar graphs are often used to present misleading results.

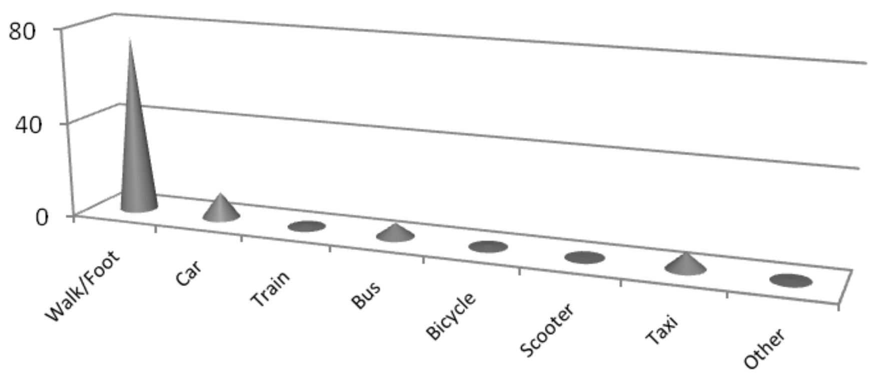
✓ An important thing to look at is the scale on the axes:

- Does the scale start at zero
- Is the scale distorted – too squashed up or too spread out. This is a trick that is often used to make something look better than it is.

Example 5.2

The following graph (Graph B) illustrates the same data as the bar graph on the previous page (Graph A).
Compare the two graphs. Why is it more difficult to read information off Graph B than Graph A?

GRAPH B



Solution

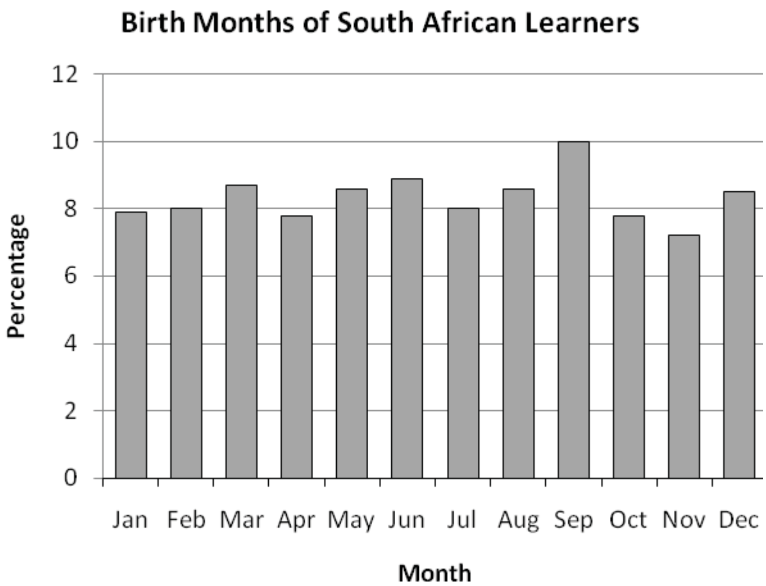
1) There is no heading to the graph so it is difficult to tell what the graph is about.
2) The vertical axis has no label. It might be showing actual numbers or percentages. It is not clear.
3) The cones instead of bars do not give a clear 'picture' of what the information is saying. Three dimensional (3D) shapes like these are usually more difficult to read than two dimensional (2D) shapes like bars.

Exercise 5.1

1) The following table shows the birth months of approximately 45 000 South African learners who took part in the 2009 Census@School. The bar graph that follows has been drawn to illustrate some of the information.

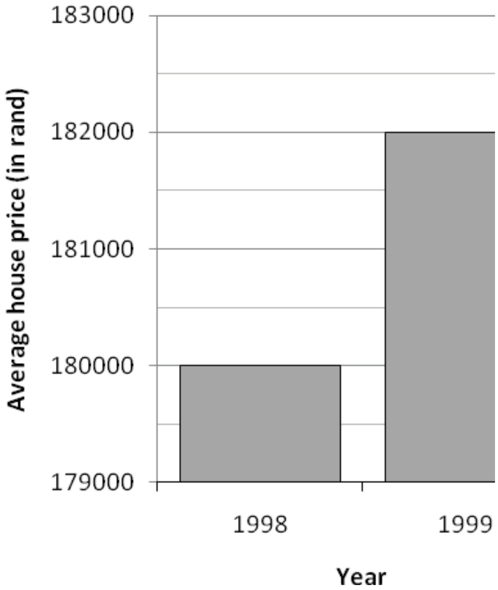
Month	Frequency
January	3 568
February	3 592
March	3 922
April	3 490
May	3 888
June	3 981
July	3 597
August	3 876
September	4 476
October	3 502
November	3 253
December	3 809
TOTAL	**44 954**

Month	Percentage
January	7,9
February	8,0
March	8,7
April	7,8
May	8,6
June	8,9
July	8,0
August	8,6
September	10,0
October	7,8
November	7,2
December	8,5
TOTAL	**100.0**

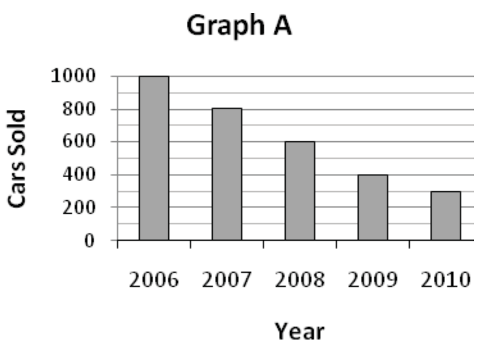


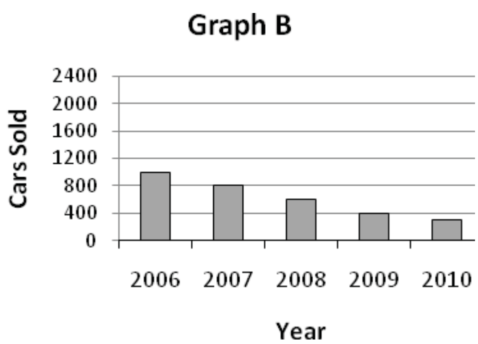
a) In which month were the least learners born?
b) In which month were the most learners born?
c) Why was percentage rather than frequency used for the y-axis?
d) Is it easier to read the most and least values off the graph or the table?

2) What is wrong or missing from this bar graph?



3) Graph A and Graph B show the number of cars sold each year by a particular car dealer.





Explain why the graphs look different

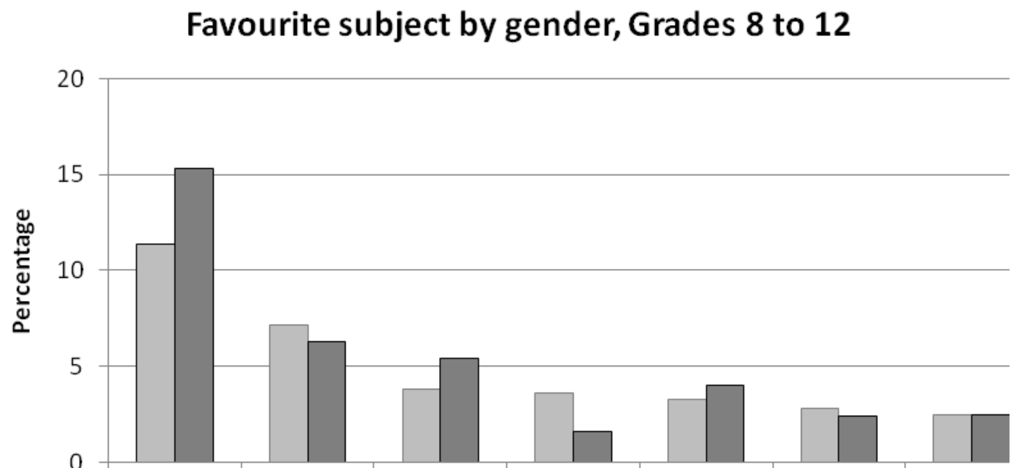
Comparative bar graphs

✓ Comparative bar graphs use adjacent bars to compare two or more subcategories.

Example 5.3

In the graph shown (taken from the South African Census@School 2009 results) the favourite subjects of male and female learners in Grades 8 to 12 were compared.

1) What was the favourite subject of both the boys and the girls?
2) What was the next favourite subject of both the boys and the girls?
3) What percentage of the boy learners and the girl learners say that Technology is their favourite subject?



	Languages	Maths	Life Orientation	Technology	Arts & Culture	Maths Literacy	Natural Science
Boys	11,4	7,2	3,8	3,6	3,3	2,8	2,5
Girls	15,3	6,3	5,4	1,6	4,0	2,4	2,5

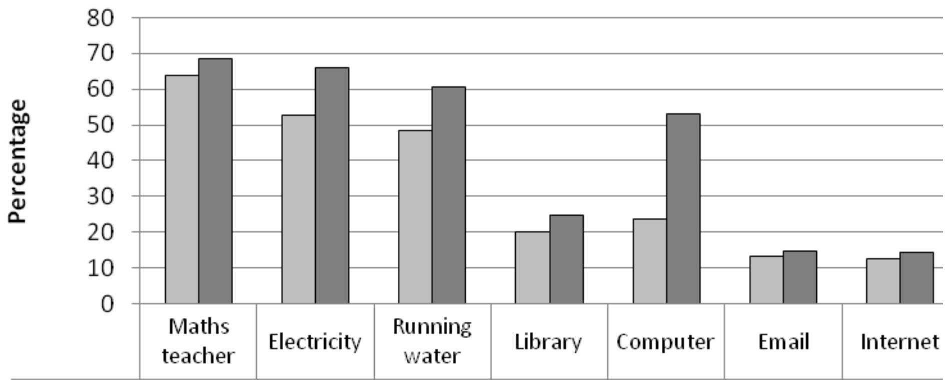
Solution

1) Languages were the favourite subject of both boys and girls
2) The next favourite for both the boys and the girls is Mathematics
3) 3,6% of boys chose Technology as their favourite subject, while only 1,5% of the girls chose it.

Exercise 5.2

1) Look at the graph shown below.
   a) What is the graph about?
   b) What do the different coloured columns represent?
   c) Write 2 sentences describing the information in the graph

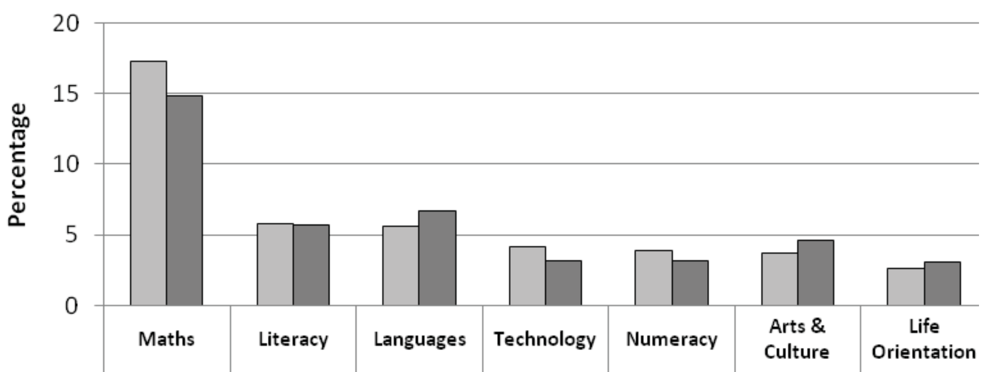
FACILITIES AND SERVICES AT SCHOOL



	Maths teacher	Electricity	Running water	Library	Computer	Email	Internet
C@S2001	64,0	52,7	48,4	20,1	23,6	13,4	12,7
C@S2009	68,6	65,9	60,5	24,6	53,0	14,7	14,5

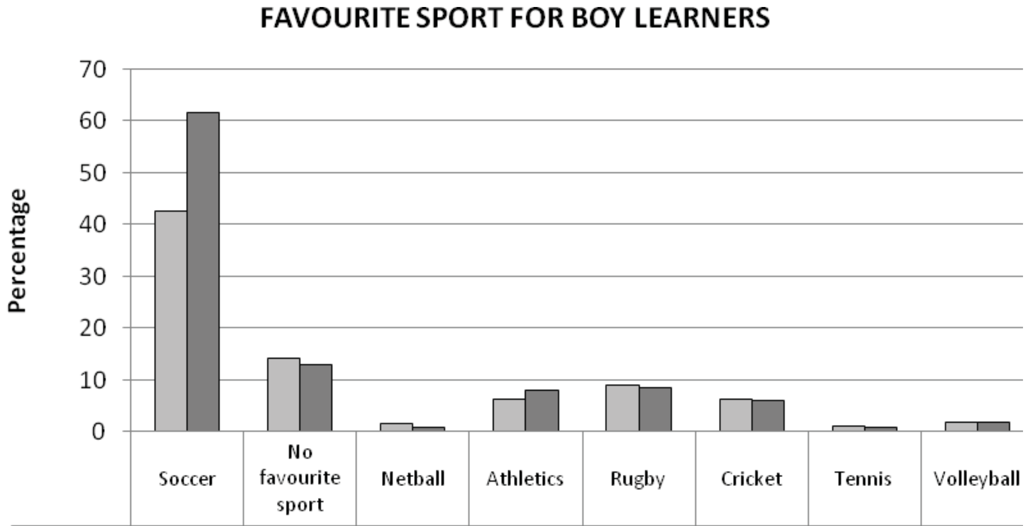
2) The bar graph below shows the favourite subject of the learners in Grades 3 to 7. Compare this graph to the graph showing the favourite subjects of the learners in Grades 8 to 12 on the previous page (page 74).
   a) What subject is the favourite for Grades 3 to 7 and which subject is the favourite for Grades 8 to 12?
   b) List the favourite subjects of the Grade 3 to 7 *girls* from most favourite to least favourite
   c) What was the least favourite subject for *all* Grade 3 to 7 learners?

Favourite subject by gender, Grades 3 to 7

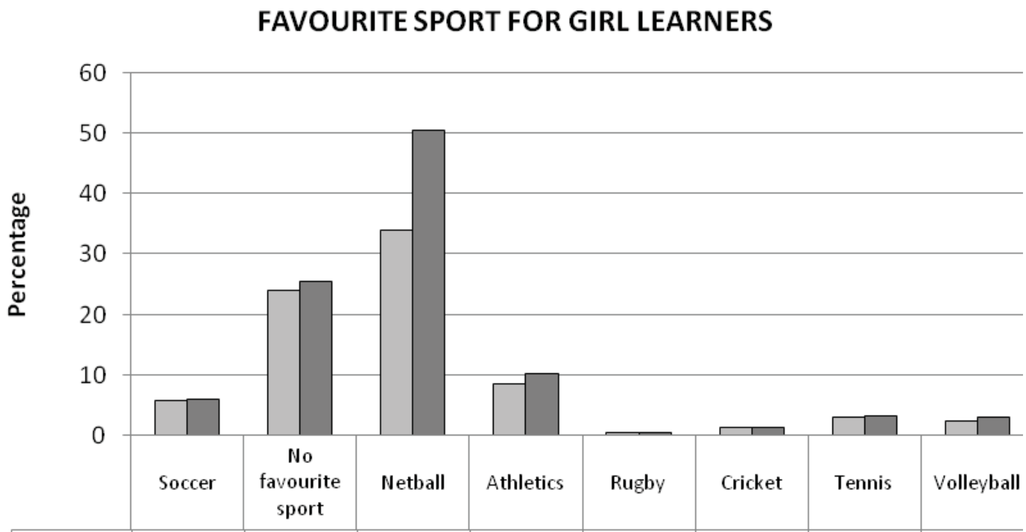


	Maths	Literacy	Languages	Technology	Numeracy	Arts & Culture	Life Orientation
Boys	17,3	5,8	5,6	4,2	3,9	3,7	2,6
Girls	14,8	5,7	6,7	3,2	3,2	4,6	3,1

3) The two graphs below show information about the favourite sport for boy and girl learners in the 2001 and 2009 Census@School.
   a) Write at least 3 sentences describing the differences in the favourite sport for boy and girl learners.
   b) Have the favourite sports changed at all from 2001 and 2009? How can you tell?

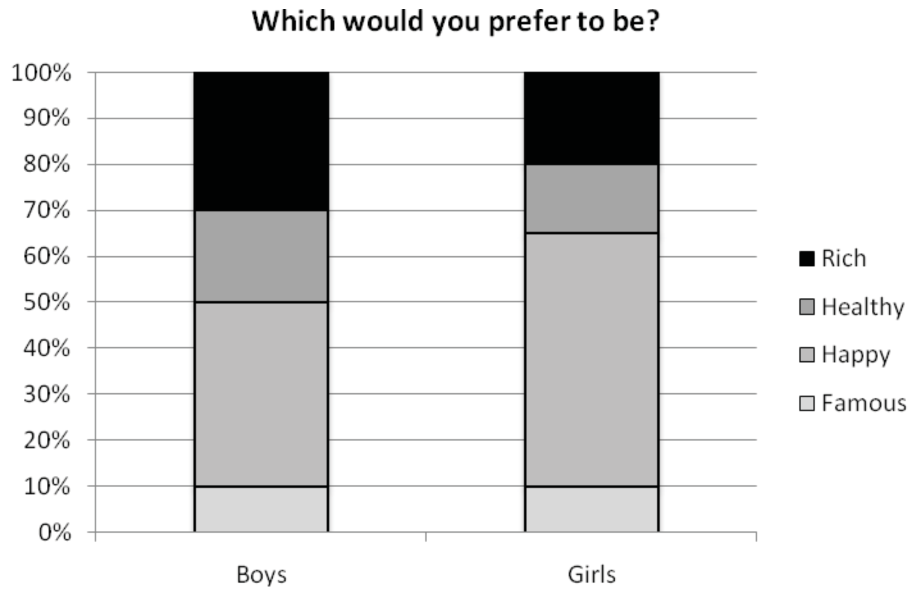


	Soccer	No favourite sport	Netball	Athletics	Rugby	Cricket	Tennis	Volleyball
C@S 2001	42,6	14,0	1,6	6,1	8,8	6,2	1,0	1,7
C@S 2009	61,6	12,9	0,8	7,8	8,4	6,0	0,8	1,8



	Soccer	No favourite sport	Netball	Athletics	Rugby	Cricket	Tennis	Volleyball
C@S 2001	5,8	23,9	34,0	8,4	0,5	1,2	3,0	2,4
C@S 2009	6,0	25,5	50,5	10,2	0,5	1,3	3,1	2,9

4) The graph below is taken from the Grade 8 to 12 Census@School done in Ireland. Write at least three sentences describing what the graph shows about learners' preferences to be famous, happy, healthy or rich.



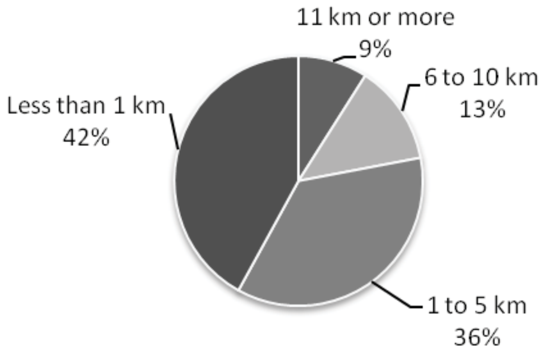
Interpreting Pie Charts

- ✓ The interpretation of pie charts is based on the fact that the largest slice of the pie chart corresponds to the largest item of data and the smallest slice of the pie chart corresponds to the smallest item. It is therefore easy to make comparisons between the relative sizes of data items.
- ✓ When interpreting a pie chart, the following parts of the pie chart are important:
  1. **The title** – what is the pie chart about?
  2. **The sectors or slices** – check the labels of the sectors and any other information given about the sectors.
  3. **The size of the sectors or slices –** the size tells you the 'how many'. Sometimes the size of the slice is easy to estimate by eye, e.g. 90°at the centre of a circle is a quarter of a circle.
- ✓ One of the main advantages of pie charts is that you can see part of the data as a fraction of the whole data.
- ✓ Pie charts that are badly drawn can also display misleading information. As with bar graphs if pie charts are missing headings or labels they can be difficult to interpret.

Example 5.4

The following pie chart is taken from the 2009 South African Census@School results. It shows the distances that learners travelled from home to school each day.

1) What is the title of the bar chart?
2) What do the sectors (or slices) tell us?
3) Compare the sizes of the sectors and say what they tell us.



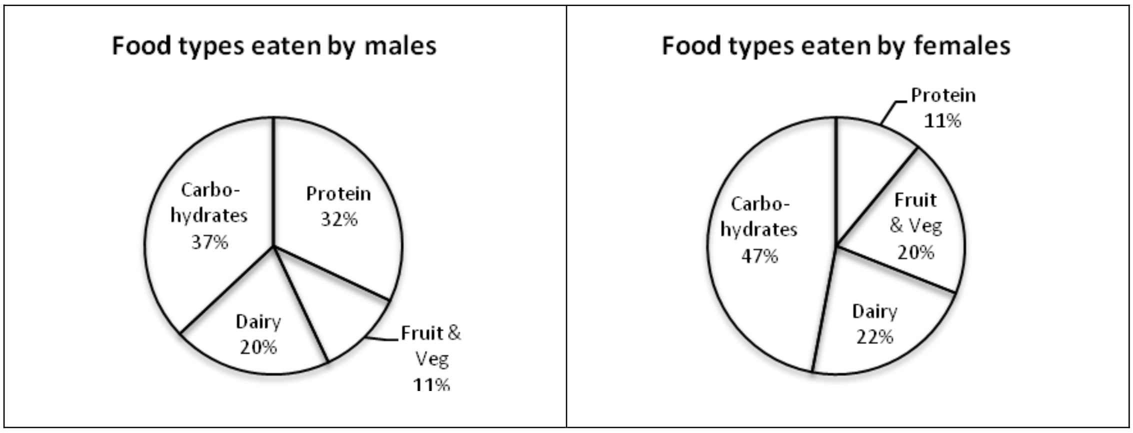
Solution

1) The title is "Travel distance to school from home" and tells us that the pie chart shows the distance travelled by learners to get to school each day.

2) The sectors/slices show different categories of distances: Less than 1 km; 1 to 5 km; 6 to 10 km and 11 km or more. They also show the percentage of learners in each category.
3) There are only 4 sectors it is easy to compare the sizes.
   → The 2 largest slices are almost the same size, so without knowing the percentages it would be difficult to say with certainty which is the bigger.
   → However it is easy to see that the smallest slice is those learners who travel 11 km or more from home to school.
   → **Most learners** (42%) live less than 1 km from school.
   → 42% + 36% = 78% of the learners live 5 km or less from school.

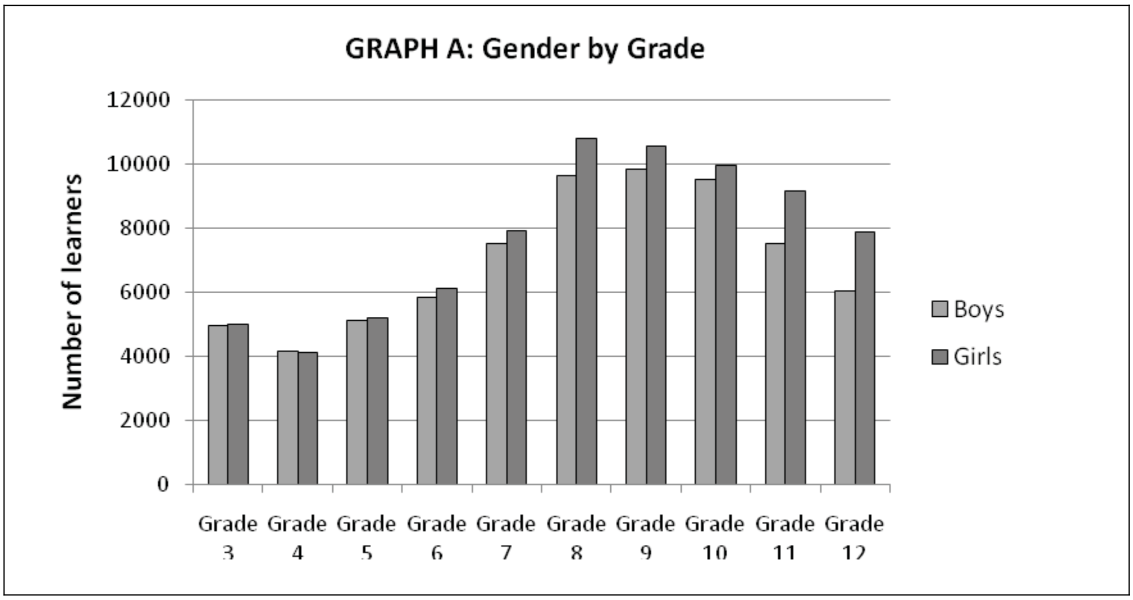
Exercise 5.3

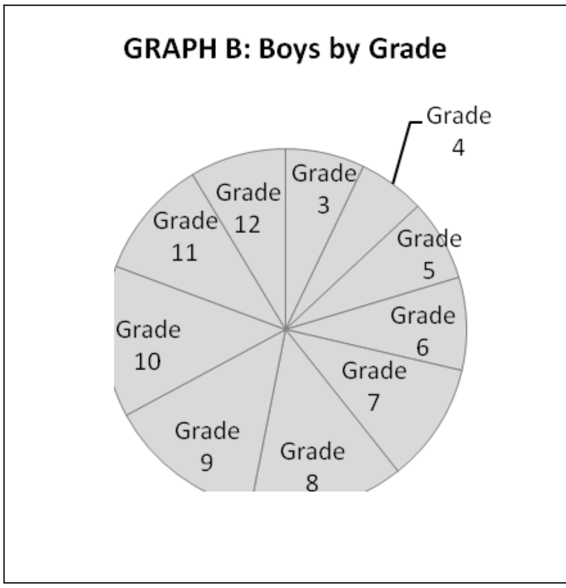
1) Look at the two pie graphs below and then answer the following questions about them:
   a) What do the graphs represent?
   b) Which food type is eaten most by males?
   c) Which food type is eaten least by males?
   d) Which food type is eaten most and eaten least by females?
   e) What is the difference between the dairy consumption of males and the dairy consumption of females?
   f) Write down any other observations you can make about the data represented in the pie charts.

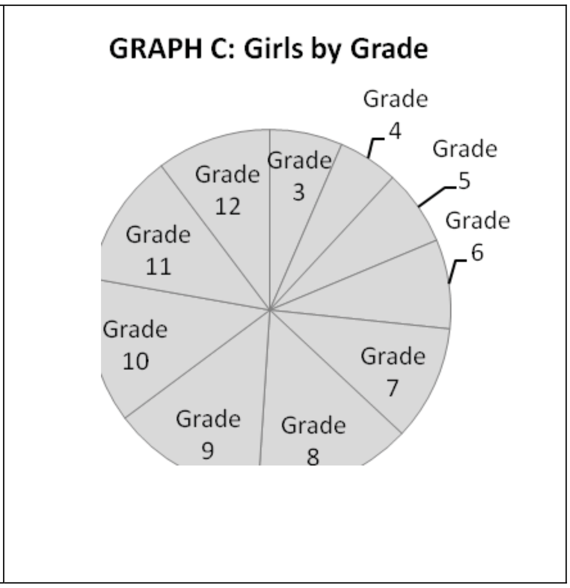


2) Look at the table and graphs shown below that give the gender by grade of learners who participated in Census@School in 2009 and then answer the questions on the next page.

GRADE	GENDER			TOTAL
	Boys	Girls	Unspecified	
Grade 3	4 969	4 983	148	10 100
Grade 4	4 180	4 121	85	8 386
Grade 5	5 126	5 213	84	10 423
Grade 6	5 825	6 128	75	12 028
Grade 7	7 510	7 913	105	15 528
Grade 8	9 654	10 782	118	20 554
Grade 9	9 847	10 547	183	20 577
Grade 10	9 521	9 974	127	19 622
Grade 11	7 521	9 158	96	16 775
Grade 12	6 027	7 862	123	14 012
Total	70 180	76 681	1 144	148 005

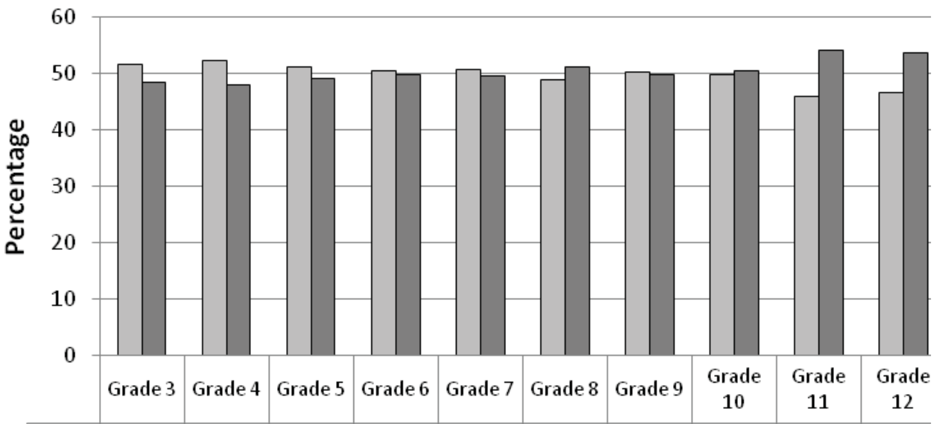






a) Is the table, the bar chart or the pie charts easier to read? Explain your answer.
b) In which grade were there most learners?
c) Did more boys or more girls participate in the survey? How do you know?
d) Can you tell how many learners participated in the census by looking at the bar chart or the pie graph?
e) Look carefully at the graph below. In what way is it different to the bar graph on the previous page? Does it show the same information or different information?

GRAPH D: Gender by Grade



	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
Boys	51,6	52,1	51,0	50,3	50,6	48,9	50,2	49,7	45,9	46,5
Girls	48,4	47,9	49,0	49,7	49,4	51,1	49,8	50,3	54,1	53,5

f) By looking at any of the four graphs for this question, what conclusions can you make about the gender by grade of the learners who participated in the census?

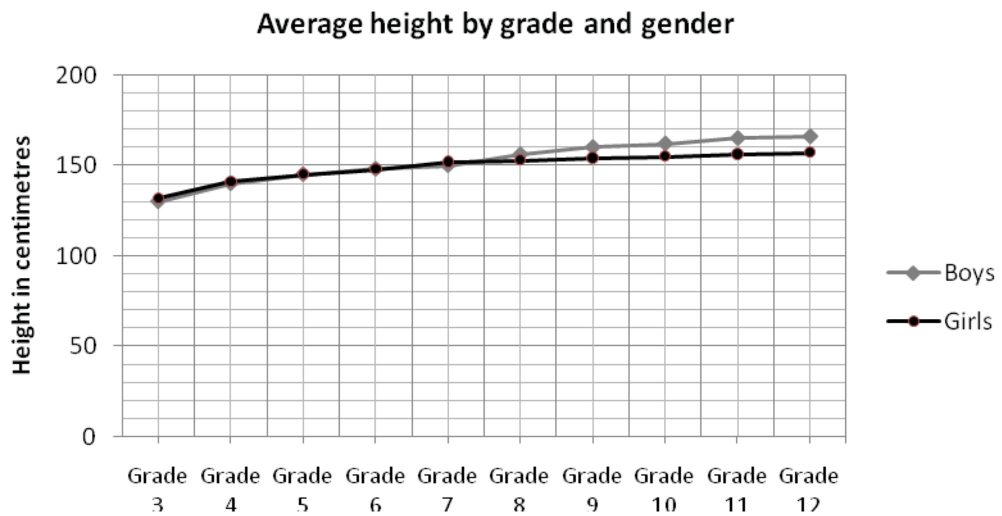
Line Graphs

✓ Line graphs are often used to show how a quantity varies.

✓ Some important things to help you interpret line graphs are:
1. **The title** – what is the line graph about?
2. **The axes** – check the labels of the axes.
3. **The scales on the axes** – do the scales start at zero? What else do the scales tell you?

Example 5.5

The following line graph is taken from the 2009 South African Census@School results. It shows the average height of learners by age and gender. Notice that the heights of the boys and the girls are compared using two lines on one graph.



1) What is the title of the line graph?
2) What information can we read off the vertical and the horizontal axis?
3) How many units does each division on the vertical axis represent?
4) What conclusions can you make about the information shown on the graph?

Solution

1) The graph shows the average height by grade and gender
2) The vertical axis shows the age of learners and the horizontal axis shows the grade.
3) Each division on the vertical axis represents 10 cm. The scales on the axes are easy to read. From the information show you cannot tell actual heights but you give a general comment about the ages.
4) You can easily see that the average height of boy and girl learners is **similar Grades 3 to 7**. However, **boys grow faster from Grade 8 to 12** and the average height of boys in these grades is greater.

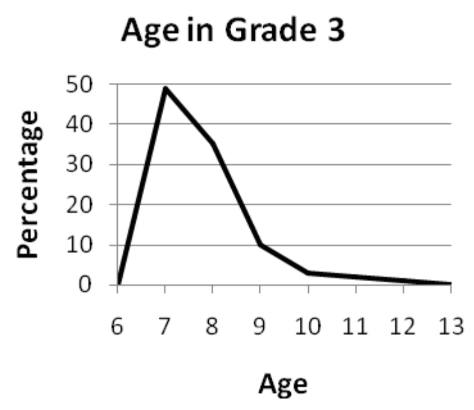
Misleading line graphs

✓ Some line graphs are purposely drawn to convey misleading information.

Example 5.6

The following two line graphs show the same information, but look different.

What is the difference between the two graphs?



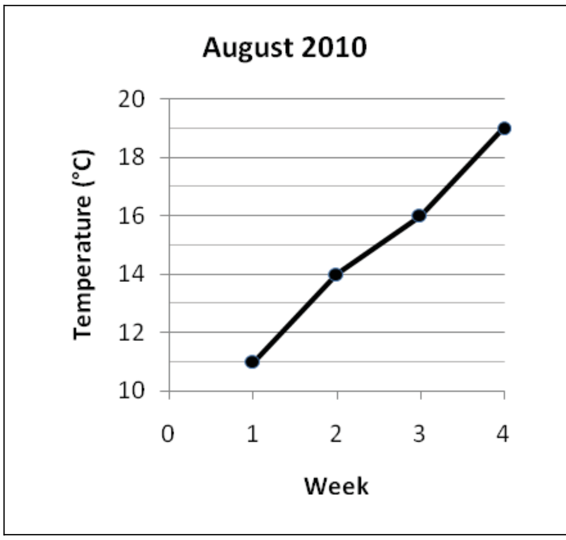
Solution

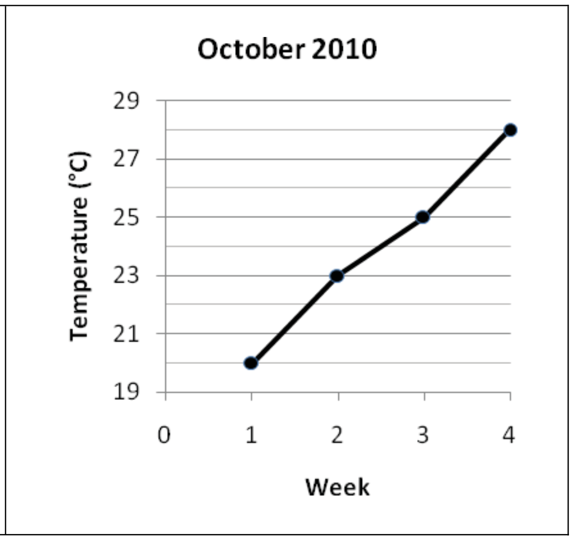
→ The first graph clearly shows that most learners in Grade 3 are 7 years old, but there are learners as old as 12 in Grade 3.

→ In the second graph the information is much more squashed up and the graph suggests that there are no learners older than 10 in Grade 3.

Exercise 5.4

1) What do you think the person who drew these two graphs wants you to believe? How would you correct the graphs so that they don't mislead?





2) a) Why might this graph be misleading?
   b) What might people believe from this graph?

Histograms

A ***bar graph*** is a graph of *ungrouped* data and a ***histogram*** is a graph of *grouped* data.

- ✓ For a histogram, the bars must touch each other, with no spaces between the bars.
- ✓ Bars must begin and end at the boundaries of the intervals.

  For example, if one interval starts just after 1 and ends at 5, the second interval must start just after 5 and end at 10.

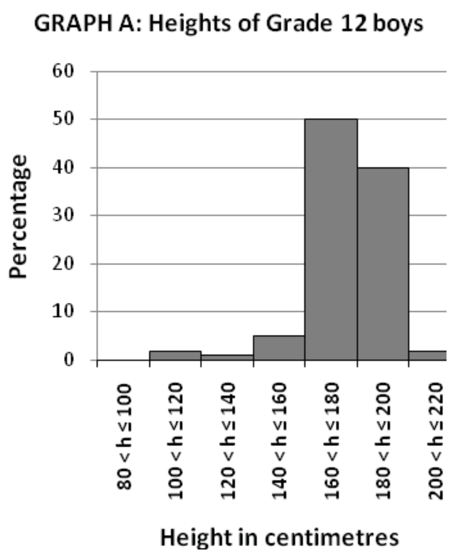
  We write this as $1 < x \leq 5$ and $5 < x \leq 10$

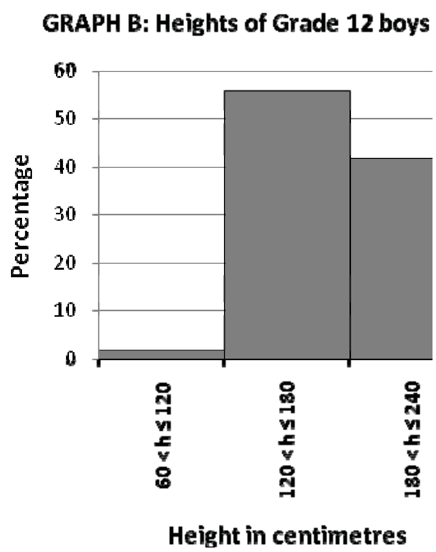
Exercise 5.5

1) The heights of one hundred Grade 12 boys were measured and recorded in the following table.

Interval	Percentage
$80 < h \leq 100$	0
$100 < h \leq 120$	2
$120 < h \leq 140$	1
$140 < h \leq 160$	5
$160 < h \leq 180$	50
$180 < h \leq 200$	40
$200 < h \leq 220$	2
TOTAL	100

Two graphs were drawn to illustrate the data.





a) How wide (in centimetres) is the interval on the horizontal axis in Graph A and how wide is the interval on the horizontal axis in Graph B?

b) Which of the two graphs best represents the spread of heights of the Grade 12 boys? Give reasons for your answer.

2) Two histograms were drawn to show the ages of the boys and the girls who participated in 2009 in an online Census@School questionnaire.

   a) Use the two graphs to answer the following questions:

      (i) What was the age of the oldest boys who participated?

      (ii) What was the age of the youngest boys who participated?

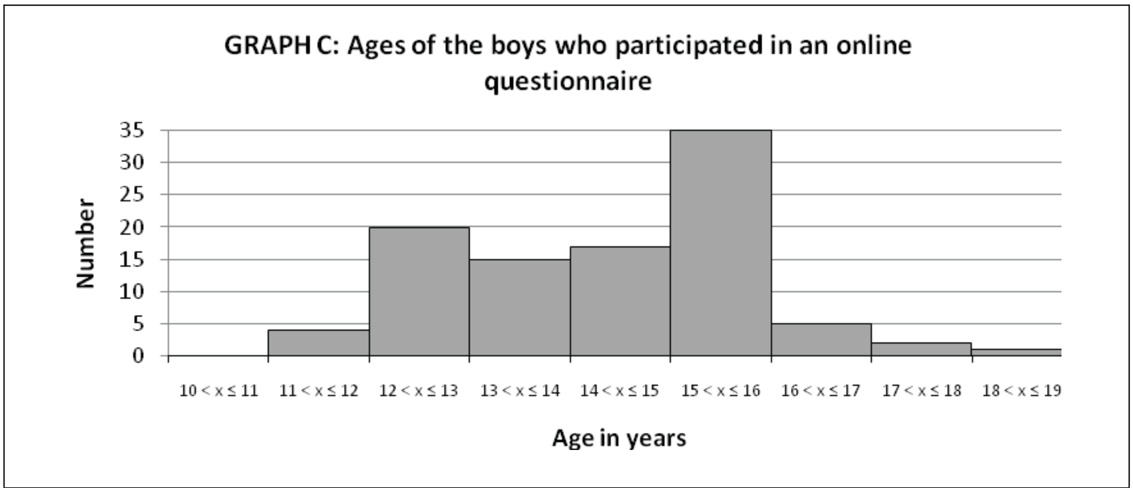
      (iii) What was the age of the oldest girls who participated?

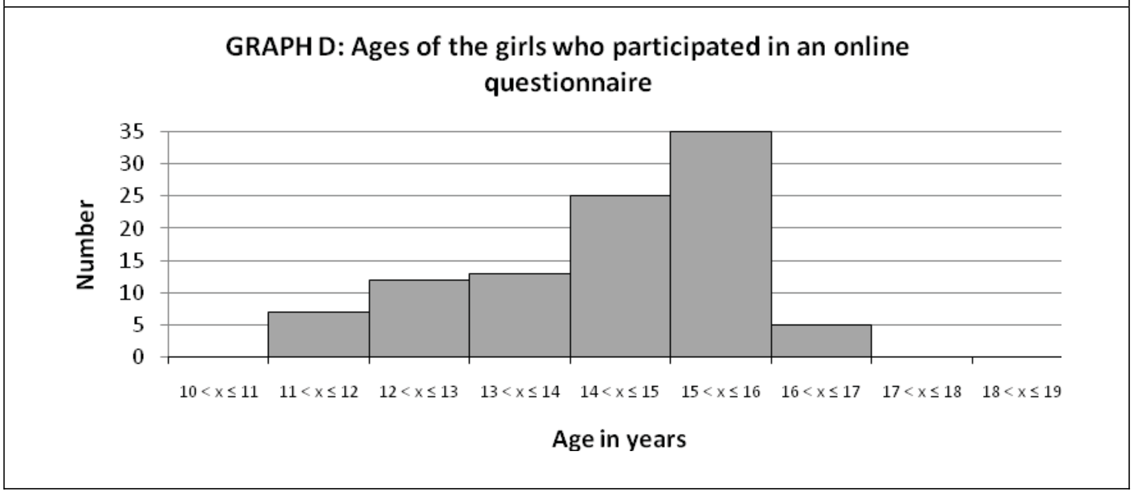
      (iv) What was the age of the youngest girls who participated?

      (v) What age was the mode for the boys?

      (vi) What age was the mode for the girls?

   b) Write down at least two other observations you can make about the information in the two graphs.





Other ways in which errors can be introduced …

Besides graphs being drawn to mislead, errors can be introduced in the following ways:

1) There can be **sampling errors**
2) There can be **dirty data** mixed in with the clean data
3) **Misleading conclusions** can be made.

Sampling

When you take a sample, you need to make sure that you don't introduce bias without thinking, e.g. you might only ask a certain group of people - only white people or only Indian people. Or you might ask only those people who volunteer to answer your questions.

Misleading Conclusions

Suppose someone tests Battery A and one Battery B. The batteries are each put in a torch which is switched on. The time taken for the torch to stop working is measured.

The conclusion is then made that Battery A lasts for up to twice as long as Battery B.

However, this conclusion is misleading. If one Battery A lasts longer than one Battery B, we cannot use this fact to make a conclusion about all "Battery A's".

Dirty data

- ✓ In a large set of data there are almost certainly going to be errors.
- ✓ Errors may arise in many different ways. They could be caused by human errors (such as someone misreading or miscounting data), to errors due to poor sampling.

The following list shows some of the ways that errors can creep into lists of data:

- **Mistyping**:
  A correct value might have been obtained when measuring or counting, but the incorrect number was written down. This can happen when digits are interchanged, e.g. writing 1 765 instead of 1 756, or when digits are written twice, e.g. 772 instead of 72.
- **Mistaken answer:**
  In an interview the respondent may misunderstand a question and so give an incorrect answer, e.g. giving a yearly income instead of a monthly income.
- **Mistaken measurement:**
  Many people like to write down 'nice' numbers, so instead of writing 31 they might write 30.

Exercise 5.6

→ In this exercise we will look at a data set and find out where errors might have crept in.
→ It has been adapted from a worksheet found at http://www.censusatschool.ie/en/resources/strand1statistics/118-cleaning-up-your-data.
→ First read the information below carefully.

Hanging out the Dirty Data

When you have got some data, the first thing you need to do is to check it out and get rid of any obviously wrong or false data. This is called "Cleaning the Data".

In the **Dirty Spreadsheet** on the next page, several playful pupils have been deliberately tampering with the data.

See if you can spot which rows and cells have been interfered with and by which of the following characters:

- **Pointy Pete**: He moves Decimal Points around and adds in unnecessary ones.
- **Obvious Olive**: She puts in **very** obvious errors.
- **Silly Samantha**: She thinks it is funny to answer the question 'Name?' as "Donald Duck".
- **Devious Dave**: He thinks up clever ways to change the data.

1) Study the **Dirty Spreadsheet** on the next page. Find examples of where each of the following characters tampered with (changed) the data.
   a) Pointy Pete
   b) Obvious Olive
   c) Silly Samantha
   d) Devious Dave.

2) The tampered data needs to be corrected. Explain how you will correct the work of:
   a) Pointy Pete
   b) Obvious Olive
   c) Silly Samantha
   d) Devious Dave.

3) Investigate the Spreadsheet further and list any other dirty data on the spreadsheet.

Dirty Spreadsheet

Row 1	Boy or Girl	Date of Birth	Grade	Height in cm	Foot Length in cm	Favourite Subject	Distance to School in km
Row 2	Boy	12/04/91	5	143	26	Art	1-2km
Row 3	Girl	31/02/92	4	132	22	Science	less than 2 km
Row 4	Girl	14/01/91	5,00	14,2	2,3	PE/Sport	2,5423 km
Row 5	Boy	07/09/89	6	136	25	Art	1-2km
Row 6	Boy	13/12/91	4	128	24	PE/Sport	1-2km
Row 7	Boy	14/03/01	5	140	67	PE/Sport	less than 1 km
Row 8	Girl	06/05/89	7	142	24	Art	3-5km
Row 9	Girl	15/08/90	6	138	21	Art	85km
Row 10	Boy	20/02/90	6	192	23	PE/Sport	1-2km
Row 11	Girl	19/05/90	6	140	20	Maths	1-2km
Row 12	Neither	29/06/92	7	48	21	Going Home	3 000km
Row 13	Boy	09/10/91	4	128	21	English	less than 1 km
Row 14	Girl	18/12/90	5	135	21	Geography	less than 1 km
Row 15	Girl	18/07/91	0,5	13,7	20	Art	3-5km
Row 16	Boy	03/06/34	4	129	21	Art	less than 1 km
Row 17	Girl	13/02/89	7	148	23	Art	1-2km
Row 18	Girl	15/09/88	7	150	22,5	PE/Sport	1-2km
Row 19	Girl	07/08/89	7	140	24	Art	less than 1 km
Row 20	Boy	08/06/89	7	142	24	Computing	less than 1 km
Row 21	Boy	31/11/87	11	1 520	22	Computing	5-10km
Row 22	Both	16/07/88	8	142	26	Japanese	2-3 km
Row 23	Girl	28/04/88	8	145	26,5	PE/Sport	1 mile
Row 24	Boy	25/03/92	4,1	132,1	2,4,5	Maths	less than 1 km
Row 25	Boy	26/02/92	4	130	21	PE/Sport	less than 1 km
Row 26	Girl	08/07/99	6	142	22	Art	2-3 km
Row 27	Boy	23/05/90	6	151	25,5	Maths	2-3 km
Row 28	Boy	01/03/87	9	162	25	PE/Sport	less than 1 km
Row 29	Girl	07/08/91	6	150	23	History	2 roads
Row 30	Girl	03/03/92	4	135	21	English	less than 1 km

Chapter 6

Relative Frequency and Probability

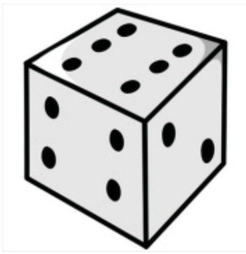
In this chapter you will:

- *List all the possible outcomes of simple experiments.*
- *Determine the relative frequency of actual outcomes for a series of trials.*
- *Determine the probability of outcomes of simple situations using the definition for probability.*
- *Predict the frequency of possible outcomes based on their probability.*
- *Compare relative frequency and probability and explain possible differences.*
- *Determine probabilities for compound events using two-way tables and tree diagrams*

P**robability is the part of mathematics that studies *chance* or *likelihood.*** For example, when you see on the weather report that there is a 60% chance of rain, they are talking about a probability.

✓ Some things will *never happen*. We say that they are **impossible**.

For example: if you throw an ordinary die – it is **impossible** that it will land on 8.



✓ Other things will *definitely happen* and we say that they are **certain** to happen.

For example: it is **certain** for an ordinary die to land on 1, 2, 3, 4, 5, or 6.

✓ Some things are *not certain* and *not impossible*.

For example:

- It is **likely** to rain when the sky is very cloudy.
- It is **unlikely** to snow in Johannesburg in December.
- There is a **50-50 chance** of a coin landing on heads when it is tossed.

- ✓ Rather than using words to describe the chance of an event happening, you can give probability as a number **between 0 and 1**.
- ✓ This number can be written as a *fraction*, *percentag*e or *decimal*.
  - If it is **impossible** for an event to happen, the probability is **0**.
  - If an event is **certain** to happen, the probability is **1**.
  - All other probabilities are greater than 0, but less than 1.

Listing outcomes of an experiment

- ✓ If we do a probability experiment, for example throw a die, toss a coin, spin a spinner, we can list all the possible **outcomes**.
- ✓ An **event** is a particular outcome or group of outcomes.
- ✓ **Favourable outcomes** are the outcomes which give the event you are interested in.
- ✓ If the die is fair, then we can say that the outcomes are **equally likely**.
  In other words all numbers have the same chance of being thrown. One number doesn't have a greater chance than the others of being thrown.
- ✓ If a die is NOT fair, we say that it is **biased**.

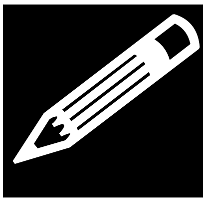
Example 6.1

Suppose you throw a fair die.

1) List all the **possible outcomes**
2) List the **favourable outcomes** for the event
   a) Getting a 4
   b) Getting an even number

Solution

1) The **possible outcomes** are 1, 2, 3, 4, 5 and 6
2) a) The **favourable outcome** of "getting a 4" is 4
   b) The **favourable outcomes** of "getting an even number" are 2, 4 and 6



Exercise 6.1

1) An **eight-sided die** (like the one alongside) is thrown.
   a) List all the possible outcomes.
   b) List all the favourable outcomes for the following events:
      Event A: Getting a 2
      Event B: Getting an odd number
      Event C: Getting a number bigger than 4.

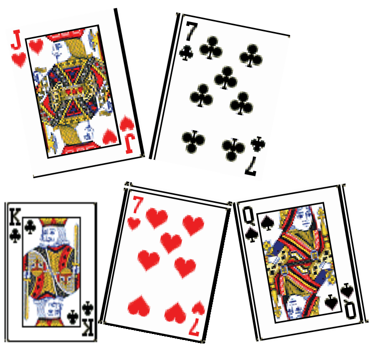
2) A coin is randomly taken from this purse.
   Inside the purse are a R5 coin, a R2 coin, two R1 coins, and a 20c coin.
   a) List all the possible outcomes.
   b) List all the favourable outcomes for the following events:
      Event D: Getting a R1 coin
      Event E: Getting a 'bronze' coin
      Event F: Getting a coin worth more than R4.

3) Randomly select a card from the following five cards.

M	A	T	H	S

   a) List all the possible outcomes.
   b) List all the favourable outcomes for the following events:
      Event G: Getting a T
      Event H: Getting a vowel
      Event J: Getting a consonant

4) Randomly select a card from this "hand" of cards.
   a) List all the possible outcomes.
   b) List all the favourable outcomes for the following events:
      Event K: Getting a 7
      Event L: Getting a heart
      Event M: Getting a picture card
      Event N: Getting a diamond.



5) The table below is from the South African Census@School 2009 Report.
It shows what percentage of learners use particular modes of transport to school from home.

	Walk/Foot	Car	Train	Bus	Bicycle	Scooter	Taxi	Other
South Africa (%)	74,8	10,9	0,6	5,7	0,5	0,1	6,8	0,5

a) The outcomes are the modes of transport from home to school.
List all the possible outcomes.

b) Suppose a South African learner is chosen at random. List all the favourable outcomes for each of the following events:
Event P: Taking a taxi
Event Q: Using public transport
Event R: Using a two-wheeled mode of transport.

Finding the relative frequency

- ✓ We find the **relative frequency** of an experiment by performing the experiment or by collecting information from past records.

- ✓ **Relative frequency** can be calculated using the following formulae:

$$\text{Relative frequency} = \frac{\text{Number of times an outcome occurs in an experiment}}{\text{Total number of trials in the experiment}}$$

OR

$$\text{Relative frequency} = \frac{\text{Number of times an outcome occurs in a survey}}{\text{Total number of observations in the survey}}$$

Example 6.2

In an experiment a drawing pin is dropped
100 times.
It lands with its **point up** 37 times.
What is the *relative frequency* of the drawing pin landing point up?

Solution

$$\text{Relative frequency} = \frac{\text{Number of times an outcome occurs in an experiment}}{\text{Total number of trials in the experiment}}$$

$$= \frac{\text{Number of times the drawing pin lands point up}}{\text{Number of times the drawing pin is dropped}}$$

$$= \frac{37}{100}$$

$= 0{,}37$ or 37%

Example 6.3

50 motor cars are observed passing the school gate.
14 of these motor cars are red.
What is the *relative frequency* of a red motor car passing the school gate?

Solution

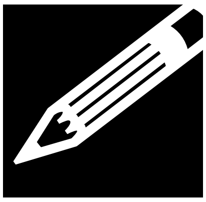
$$\text{Relative frequency} = \frac{\text{Number of times an outcome occurs in a survey}}{\text{Total number of observations in the survey}}$$

$$= \frac{\text{Number of red motor cars}}{\text{Total number of motor cars}}$$

$$= \frac{14}{50}$$

$$= \frac{28}{100}$$

$= 0{,}28$ or 28%



Exercise 6.2

1) In an experiment a gardener planted 40 daffodil bulbs.
36 of these daffodil bulbs produced flowers.
Use these results to find the relative frequency that a daffodil bulb will produce a flower.

Use the formula

$$\text{Relative frequency} = \frac{\text{Number of bulbs producing flowers}}{\text{Total number of bulbs planted}}$$

2) 

Thandi keeps a record of her chess games with Helen.

Out of the first 30 games, Helen wins 21 games.

a) Use the results to work out
(i) the relative frequency that Helen wins

Use the formula:

$$\text{Relative frequency} = \frac{\text{Number of games won by Helen}}{\text{Number of games played}}$$

(ii) the relative frequency that Thandi wins

b) Use the two relative frequencies to predict whether Thandi will win her next game of chess with Helen.

3) The results of games of chess played by four children at a chess club are shown in this table.

Player	Games won	Games drawn	Games lost
Alfred	4	2	6
Busi	8	1	7
Fikile	3	0	1
Pieter	9	2	9

a) Calculate the number of games that each player plays.

b) Calculate the relative frequency of a win for each player.
Use the formula:

$$\text{Relative frequency of a win} = \frac{\text{Number of games won}}{\text{Number of games played}}$$

c) If Alfred plays Pieter, who do you think is more likely to win? Explain your answer.

4) A counter is taken at random from a bag of coloured counters.

Its colour is recorded and the counter is then put back in the bag.

This is repeated 300 times.

The number of yellow counters taken from the bag after every 100 trials is shown in the table.

Number of trials	Number of yellow counters
100	54
200	101
300	135

a) Calculate the relative frequency of getting yellow after

(i) 100 trials

(ii) 200 trials

(iii) 300 trials

Use the formula:

$$\text{Relative frequency of getting a yellow counter} = \frac{\text{Number of yellow counters}}{\text{Number of trials}}$$

b) Suppose there are 10 coloured counters in the bag. How many of these counters do you think are yellow? Explain how you decided on your answer.

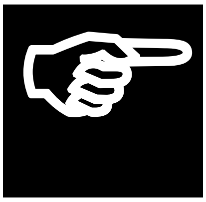
Calculating probability

- ✓ In the previous section we looked at how you could estimate probabilities by doing experiments or making observations.

- ✓ When we have a situations where each outcome is **equally likely** to occur, we can calculate the **probability** of a particular outcome occurring.

Probability can be calculated using the following formula:

$$\text{Probability} = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}}$$



Example 6.4

An ordinary die is rolled. What is the probability of getting:

a) a 6?
b) an odd number?
c) a 2 or 3?

Solution

The possible outcomes on an ordinary die are: 1; 2; 3; 4; 5; and 6.
The total number of possible outcomes is 6.

a) The favourable outcome is 6.
The number of favourable outcomes is 1.

$$P(6) = \frac{1}{6} \approx 0{,}17 \approx 16{,}7\%$$

b) The favourable outcomes are 1; 3 and 5.
The number of favourable outcomes is 3.

$$P(\text{even}) = \frac{3}{6}$$

$$= \frac{1}{2} \text{ or } 0{,}5 \text{ or } 50\%$$

c) The favourable outcomes are 2 and 3.
The number of favourable outcomes is 2.

$$P(2 \text{ or } 3) = \frac{2}{6}$$

$$= \frac{1}{3} \approx 0{,}33 \approx 33{,}3\%$$

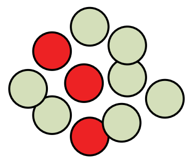
Exercise 6.3

1) A bag contains a red counter, a blue counter and a green counter.
A counter is taken from the bag at random.

a) What are the possible outcomes?

b) What is the probability of taking:
(i) A red counter?
(ii) A red or green counter?
(iii) A counter that is not blue?
(iv) A yellow counter?

2) A bag contains 3 red sweets and 7 green sweets.
A sweet is taken from the bag at random.

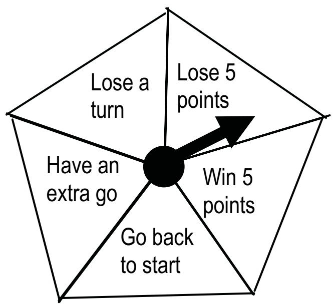


a) What are the possible outcomes?

b) Are you more likely to get a red or a green sweet?

c) What is the probability of taking:
(i) A red sweet?
(ii) A green sweet?

3)



This fair spinner is used in a game.
The player spins the arrow.

a) What are the possible outcomes?

b) What is the probability that the player:
(i) Loses a turn?
(ii) Wins 5 points or loses 5 points?

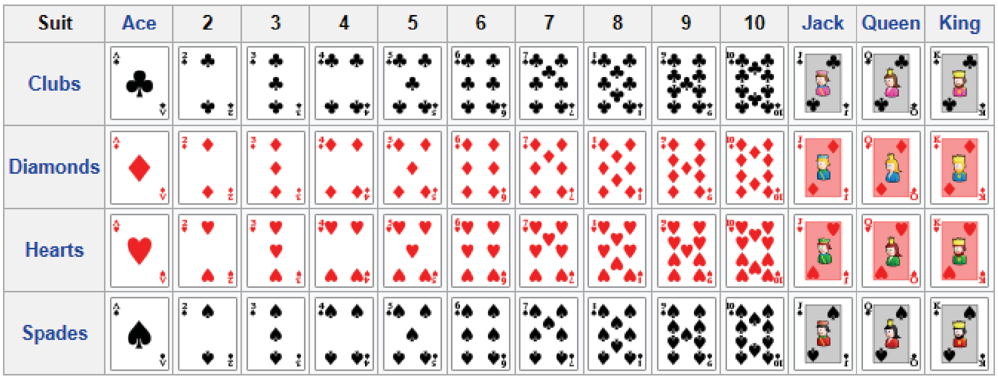
4) The letters of the word PROBABILITY are written on separate cards of the same size.
The cards are shuffled and dealt, face down, onto a table.
A card is selected at random.

a) What are the possible outcomes?

b) What is the probability that the card shows:
(i) The letter T?
(ii) The letter B?

5) A card is randomly taken from a full pack of 52 playing cards with no jokers.

Example set of 52 poker playing cards



- Diamonds and Hearts are red cards
- Clubs and Spades are black cards
- The Ace is the same as the number 1
- The picture cards are Jack, Queen and King.

What is the probability that the card that is taken is:

a) A red card?

b) A heart?

c) The Ace of hearts?

6) A bag contains 4 red counters, 3 white counters and 3 blue counters.
A counter is taken from the bag at random.

a) How many possible outcomes are there?

b) What is the probability that the counter is:
(i) Red?
(ii) White or blue?
(iii) Red, white or blue?
(iv) Green?

7) There are twelve numbered discs in a box.

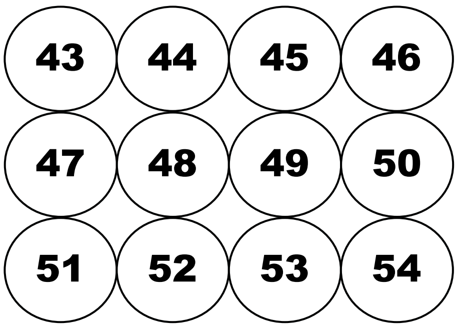
Matlaba closes his eyes and takes a disc from the box at random.

What is the probability that Matlaba takes a disc:

a) With at least one 4 on it?

b) That has not got a 4 on it?

c) That has 3 or 4 on it?



8) Pirates are playing Mamelodi Sundowns.

Victor says that Pirates can either win, draw or lose, so the probability of them winning must be $\frac{1}{3}$.

Explain why Victor is wrong.

9) An 8-sided die has the numbers 1 to 8 painted on its faces. When thrown, it has an equal chance of landing on any one of its faces.

a) How many possible outcomes are there?

b) What is the probability that it lands on:
   (i) 6?
   (ii) A number less than 6?
   (iii) An odd number?
   (iv) A number more than 6?
   (v) A prime number?

10) 

240 tickets are sold in a school raffle.

What is the probability of winning first prize if I buy:

a) One ticket?

b) Seven tickets?

11) In a game of Bingo, the numbers 1 to 90 are written on individual discs and placed in a big box.

The numbers are taken out one by one, and the numbers are called out.

a) How many possible outcomes are there?

b) What is the probability that the first number chosen at random is:
   (i) a 69?
   (ii) a number that ends in 0?
   (iii) a number that is a multiple of 9?

12) Below is a data table showing the birth months of South African learners. It is taken from the 2009 Census@School report.

Month	Frequency
January	3 568
February	3 592
March	3 922
April	3 490
May	3 888
June	3 981
July	3 597
August	3 876
September	4 476
October	3 502
November	3 253
December	3 809
TOTAL	**44 954**

a) If a learner is chosen at random, which month are they most likely to be born in?

b) If a learner is selected at random, what is the relative frequency that this learner was born in March?

Example 6.5

Makgoshi does the following experiment with a bag containing 2 red counters and 8 blue counters.

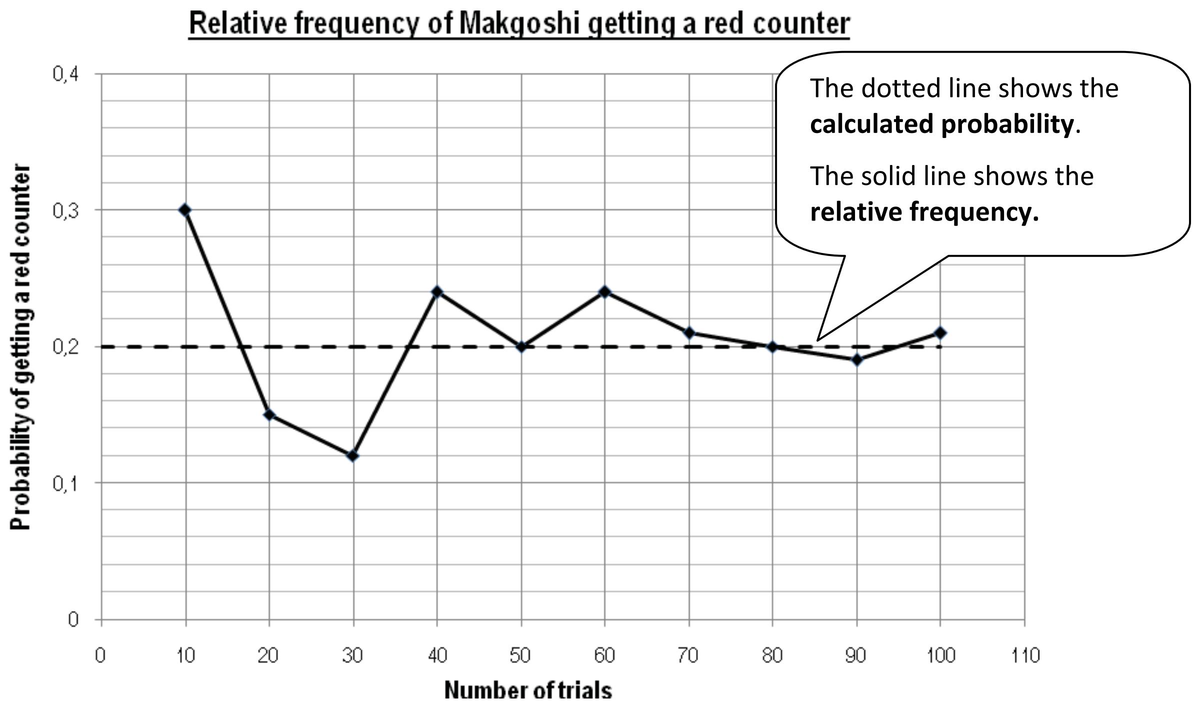
EXPERIMENT

- **Take a counter from the bag at random.**
- **Record the colour then put the counter back in the bag.**
- **Repeat this for 100 trials.**

Makgoshi calculated the relative frequency of getting a red counter after every 10 trials and recorded the results in a table.

Number of throws	10	20	30	40	50	60	70	80	90	100
Relative Frequency	0,3	0,15	0,12	0,24	0,2	0,24	0,21	0,2	0,19	0,21

She then plotted her results on the following graph.



a) What is the **probability** that Makgoshi takes a red counter?

b) What do you notice about the relative frequency as the number of trials increases?

Solution

a) P(red counter) = $\frac{\text{number of red counters}}{\text{total number of counters}} = \frac{2}{10} = \frac{1}{5} = 0{,}2$

b) As the number of trials increases, the relative frequency varies, but eventually gets close to 0,2. In other words, it gets closer to the calculated probability.

NOTE
When a question asks you to *estimate the probability*, it is actually asking you to calculate the relative frequency.

Example 6.5 shows you that we can use relative frequency to estimate the probability of an event happening.

The larger the number of trials or observations, the closer the relative frequency is to the probability.

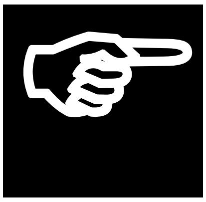
Exercise 6.4

1) Repeat Makgoshi's experiment in the following way:
   a) Cut out 10 circles the same size from a piece of paper.
   b) Mark two of them with R (for red) and eight of them B (for blue).
   c) Take a counter at random. Record the colour and then put the counter back in the bag.
   d) Repeat this for 200 trials.
2) Draw a graph to show the relative frequency after every 10 trials.
3) Are your results similar to Makgoshi's results?

Predicting frequencies based on probability

- ✓ We can use the probability of an outcome occurring to *predict or estimate the frequency of that outcome occurring* in an experiment.

- ✓ We can use the formula to estimate the frequency of an outcome occurring:
  Predicted frequency of an outcome
  = Probability of that outcome occurring × Number of repeats of the experiment



Example 6.6

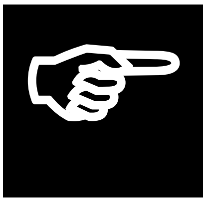
A fair coin is tossed 10 times.

How many heads could you expect to get?

Solution

The probability of getting heads = $\frac{1}{2}$

Predicted frequency of heads
= probability of getting a head × number of times the coin is tossed
= $\frac{1}{2} \times 10$
= 5



Example 7

A bag contains 20 counters of different colours.
A counter is randomly taken out, its colour noted and replaced.

The probability of getting a blue counter is 0,4.

How many blue counters are in the bag?

Solution

Number of blue counters
= probability of getting a blue counter × number of counters
= 0,4 × 20
= 8

Exercise 6.5

1) There are 50 cars in a school car park. Five of the cars are black.

   a) What is the probability that the first car to leave the car park will be black?

   b) If the probability that a red car is the first to leave is 0,2, how many red cars are there in the car park?

2) 500 tickets are sold for a prize draw.

   a) Sam buys one ticket. What is the probability that he wins the first prize?

   b) The probability of Cynthia winning first prize is $\frac{1}{20}$. How many tickets did she buy?

3) Mr Sithole buys 300 calculators from a supplier which he intends to sell at his school.

   He is warned that the probability of the calculators being faulty is 0,02.

   How many of the calculators could he expect to be faulty?

Probabilities of Compound events occurring

✓ So far, in this chapter, we have been looking at probabilities of *single events* occurring.

✓ In some problems, we will have to find the probability of *compound events* occurring. Events are called compound events when two (or more) activities take place.

Examples of compound events:

- We spin a spinner and then select a card.
- We throw two dice together
- We toss a coin twice.

✓ In situations like these, we can find all the possible outcomes using

(i) A list
(ii) A two-way table
(iii) A tree diagram.

Example 6.8

A fair coin is tossed twice.

a) Identify all the possible outcomes

b) Find the probability of getting two heads.

Solution

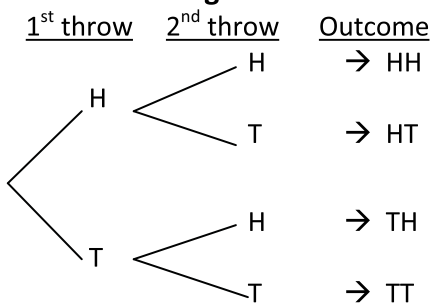
(i) **List** the outcomes systematically.

1st throw	2nd throw
Head (H)	Head (H)
Head (H)	Tail (T)
Tail (T)	Head (H)
Tail (T)	Tail (T)

(ii) Use a **two-way table**.

2nd throw \ 1st throw	H	T
H	HH	TH
T	HT	TT

(iii) Use a **tree diagram**



a) All three of these methods show that there are four possible outcomes, namely: HH, HT, TH and TT.

b) There is only one favourable outcome in the event of getting two heads, so $P(H;H) = \frac{1}{4}$

Example 6.9

Two dice are thrown together.

1) What is the probability of getting a 'double 6'?
2) What is the probability of getting any 'double'?

Solution

→ We first need to find how many possible outcomes there are.

→ We can list them using a two-way table.

		First die					
		1	**2**	**3**	**4**	**5**	**6**
Second die	**1**	1;1	2;1	3;1	4;1	5;1	6;1
	2	1;2	2;2	3;2	4;2	5;2	6;2
	3	1;3	2;3	3;3	4;3	5;3	6;3
	4	1;4	2;4	3;4	4;4	5;4	6;4
	5	1;5	2;5	3;5	4;5	5;5	6;5
	6	1;6	2;6	3;6	4;6	5;6	6;6

There are 36 possible outcomes.

1. There is only one favourable outcome for the event "getting double 6", namely (6;6)

   So P(6;6) = $\frac{1}{36}$

2. There are six favourable outcomes for the event "getting any double", namely (1;1), (2;2), (3;3), (4;4), (5;5) and (6;6)

   So P(any double) = $\frac{6}{36} = \frac{1}{6}$

Exercise 6.6

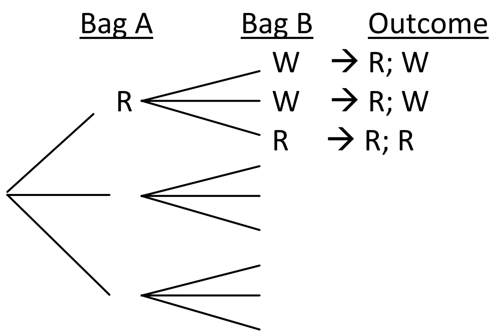
1) A fair coin is tossed and an ordinary die is rolled.
   a) Copy and complete the two-way table to list all the possible outcomes.

		Die					
		1	2	3	4	5	6
Coin	**H**						
	T						

   b) Use the table to calculate the probability of getting:
      (i) A head and 5?
      (ii) A tail and 6?
      (iii) A tail and an even number?
      (iv) A tail and an odd number?
      (v) A head and a number greater than 3?
      (vi) An odd number?

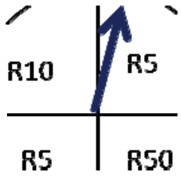
2) Two ordinary dice are rolled and the numbers obtained are added.
   a) Draw a two-way table to show all the possible outcomes.
   b) Use your two-way table to work out:
      (i) The probability of obtaining a total of 10.
      (ii) The probability of obtaining a total greater than 10.
      (iii) The probability of obtaining a total less than 10.
   c) Explain why the probabilities you worked out in (b) should add up to 1.

3) Bag A contains 2 red balls and 1 white ball.
   Bag B contains 2 white balls and 1 red ball.
   A ball is drawn at random from each bag.
   a) Copy and complete the tree diagram to show all the possible pairs of colours.
   b) Calculate the probability that the two balls are the same colour.



4) Tobias has to travel to school in two stages,
   Stage 1: He can get a lift with his neighbour, or he can take a bus, or he can take a train.
   Stage 2: He can either take a taxi, or he could walk.
   a) Use a tree diagram to help you list all the ways that Tobias could travel to school.
   b) Tobias selects one of the ways at random. What is the probability that he takes a bus and then a taxi?

5) In a game, the fair spinner shown alongside is *spun twice*.
   The two amounts that the arrow lands on **are added** to calculate the winnings.
   a) Draw up a two-way table and use it to show the amounts that could be won.
   b) Calculate the probability of winning:
      (i) R20
      (ii) R100
      (iii) R60
      (iv) R15



6) A spinner has an equal probability of landing on red, green, blue, yellow or white.
   The spinner is spun twice.
   a) Use a two-way table or a tree diagram to list all the possible outcomes.
   b) What is the probability that, on both spins,:
      (i) The spinner lands on white?
      (ii) The spinner lands on white at least once?
      (iii) The spinner lands on the same colour?

7) Each day a blue car (B), a yellow car (Y) and a red car (R) are parked one behind the other in a narrow driveway. The order in which they park in the driveway is not fixed and varies from day to day.


Blue Car Yellow Car Red Car

   a) List all the possible orders in which the three cars could be parked.
   b) What is the probability that, on any particular day, the red car is the first in the driveway?

Chapter
7

Answers

CHAPTER 1 – COLLECTING DATA

Note
This chapter is aimed at getting you to think about doing research using data and the many things you need to think about when you do such research.

Exercise 1.1 (page 7)
The solutions here are illustrative and are given for the example question, "What is the spread of ages of learners in your school".

1) The problem is "What is the spread of ages of learners in your school?"
2) Plan:
   a) I will draw up a question sheet where I will be able to record the different ages of the learners at my school.
   b) I will collect information about the ages of the learners at my school.
   c) I will have responses on questionnaires which I will tally.
3) Data recording
   a) I will record my data in tables which I can draw up by hand. If I have access to a computer I would use an Excel spread sheet to record my data.
   b) I would present my data in a table (giving the total numbers of learners with different ages) and then I would also draw a bar graph to represent my information.
4) Analysis
   a) I would look for things like the most common age, the highest age and the lowest age in the sample.
   b) I would think about reasons for why the data looks as it does.
   c) If I noticed any relationships between the variables I would think about them a bit more. Here I might notice that certain ages are more common in certain grades.
5) Conclusion
   a) I would write about what I notice about the most common age, the highest age and the lowest age in the sample.
   b) I would write about the reasons for why I think it looks like this.
   c) I would explain the relationships between some of the variables that I investigated.

Exercise 1.2 (page 10)
The problem you want to investigate is: What is the favourite tuck shop food for learners in your class?

1) Data about favourite tuck shop food for learners in my class
2) I would draw up a questionnaire which I could use to find out about favourite tuck shop food for learners in my class. I would make some lists of food that I know are sold at the tuck shop so that they can tick off their favourites rather than have to write them down in full.
3) I would tally the ticks made by the learners in my class. I would then have raw data. I could make a table in which I could give the totals of all the tallies.

4) What form of data collection could you use?
   a) I could a **census** because the population here is my class and I can manage to ask all of them about their favourite food.
   b) No I could not have used a **sample survey** because if I took a sample from my class, it would be too small. Data needs to be collected from a big enough sample if you use a sample.
   c) No I could NOT have used an **experiment** because I was not trying to understand cause and effect relationships.
   d) No I could NOT have used an **observational study** because, again, I was not trying to understand cause-and-effect relationships.

Exercise 1.3 (page 13)

1) A sample is a smaller group, taken from a population which is used to represent the population.
2) A population is the whole group of people (or things) that you want to speak about.
3) The statement of Thembi's research problem is: What percentage of learners play soccer at break time at school?
4) Thembi asked 7 of his friends whether or not they liked to play soccer at break. This sample is not representative of the whole South African population.

Exercise 1.4 (page 15)

1) Thembi's school.
2) No.
3) (Thembi wonders whether 57% of South African learners play soccer at break time.) Habib thinks that Thembi's sample is not representative. I agree with him. You can't use 7 learners to represent a whole school population. You can't use them to talk about your whole country.
4) Thembi can improve his survey by going to the SA Census at Schools website and finding the information about learners that they have. If they have answered a similar question to the one he has thought about, this could be very useful because he could never carry out a census himself.
5) Answers would vary.
   a) A problem that you investigate in your class or church youth group.
   b) A problem that you investigate in your school or neighbourhood.
   c) A problem when you want to generalise the result for the whole population.
   d) A problem when you will not be making generalisations about what you find. You will keep the results to speak about to the sample that you dealt with only.

Exercise 1.5 (page 16)

1) 217 learners
2) No. She just asked every learner that she saw over a certain period of time.
3) No. Her sample would not have been representative of the school. It was too small and it probably did not have a representative spread of the learners at the school.
4) Athletics, soccer, netball, softball and cricket.
5) Soccer, netball, athletics, no favourite sport and rugby.
6) Athletics in number one in Jane's survey and number 3 in the C@S survey. They are both high on the top 10 list, but not the same.
7) No. her sample is too small and not representative.
8) Yes. It was randomly selected and it is very large (24 172 learners)

Exercise 1.6 (page 18)

Answers will vary

1) A particular question – What are the favourite hobbies of the boys in my Grade 8 class?
2) A comparative question – What is the difference between the favourite hobbies of boys and girls in my Grade 8 class?
3) A relationship question – Do learners with a lot of extra-mural interests perform better at school?

Exercise 1.7 (Page 19)

Answers will vary

b) What are the favourite hobbies of the boys in my Grade 8 class? – I will ask all of the boys in my Grade 8 class.

c) What is the difference between the favourite hobbies of boys and girls in my Grade 8 class? – I will ask all of the boys and girls in my Grade 8 class.

d) Do learners with a lot of extra-mural interests perform better at school? – I will have to ask a representative sample of the learners at my school about their extra mural interests, and find out about how they achieve at school from teachers' records of marks. Then I will be able to see if there is a relationship between academic performance and number of extramural activities.

Exercise 1.8 (page 21)

Answers will vary. The model answers given here apply to the question: "What types of transport do the learners in my school use to go to school?"

Questionnaire – types of transport used to travel to school

Circle the correct answer to each of the questions below

1) Do you walk to school?
   a) Yes b) No
2) Do you drive in a car to school?
   a) Yes b) No
3) Do you drive in a taxi to school?
   a) Yes b) No
4) Do you drive in a bus to school?
   a) Yes b) No
5) Do you ride a bicycle to school?
   a) Yes b) No
6) Do you take a train to school?
   a) Yes b) No
7) If you go to school in a car, who drives the car?
   a) Your mom/dad b) Your guardian c) One of your siblings d) A relative/friend e) I do not drive to school
8) Which would be your favourite way to travel to school
   a) Car b) Taxi or bus c) Bicycle or by foot d) Train e) No favourite
9) How do you choose your favourite method of transport?
   a) It's the quickest
   b) It's the safest
   c) I want to keep fit
   d) It's the easiest to get to from my house
   e) I don't have a favourite
10) Using your form of transport how long do you take to get to school?
   a) 0 – 15 minutes
   b) 16 – 30 minutes
   c) 31 – 45 minutes
   d) 46 – 60 minutes
   e) More than 60 minutes

CHAPTER 2 – ORGANISING AND SUMMARISING DATA

Exercise 2.1 (page 22)

1) 136; 138; 139;
   140; 140; 140; 146; 147; 148; 148; 148; 148; 148; 148; 148; 148
   150; 150; 150; 152; 152; 154; 155; 155; 155; 155; 155; 155; 156; 156; 156; 157; 158; 158; 158; 158; 158; 158
   160; 160; 160; 160; 160; 160; 160; 160; 160; 160; 160; 160; 161; 162; 162; 162; 163; 163; 163; 164; 164; 165; 165; 165; 165; 165; 165; 165; 165; 165; 167; 167; 167; 167
   170; 170; 170; 170; 170; 170; 170; 170; 171; 172; 175; 175; 177; 178; 178; 179; 179
   180
2) The most common height is 160 cm
3) 16 learners are shorter than 150 cm
4) 10 learners are taller than 170 cm

Exercise 2.2 (page 25)

Answers to 1), 2) and 3) are going to vary from class to class.

Exercise 2.3 (page 27)

1)

STEM	LEAVES
13	6; 8; 9
14	0; 0; 0; 6; 7; 8; 8; 8; 8; 8; 8; 8; 8
15	0; 0; 0; 2; 2; 4; 5; 5; 5; 5; 5; 5; 6; 6; 6; 7; 8; 8; 8; 8; 8; 8
16	0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 2; 2; 2; 3; 3; 3; 4; 4; 5; 5; 5; 5; 5; 5; 5; 5; 5; 7; 7; 7; 7
17	0; 0; 0; 0; 0; 0; 0; 0; 1; 2; 5; 5; 7; 8; 8; 9; 9
18	0
KEY: 13/6 = 136 cm	

2) Some conclusions that can be reached are:
   - Fifteen of the boys are shorter than 150 cm
   - Eighteen of the boys are 170 cm or taller
   - The shortest boy is 136 cm and the tallest boy is 180 cm
   - Thirty-four of the boys have heights that range from 160 cm to 167 cm
   - More than half of the boys (56 of the boys) have heights that range from 150 cm to 167 cm.

Exercise 2.4 (page 30)

1) a)

Mathematics Marks	Frequency
$30 < m \leq 40$	6
$40 < m \leq 50$	8
$50 < m \leq 60$	11
$60 < m \leq 70$	4
$70 < m \leq 80$	3
TOTAL	32

   b) 6 + 8 = 14 learners got 50 or less for the test.

2) a)

Heights in metres	Frequency
1,50 < x ≤ 1,55	1
1,55 < x ≤ 1,60	4
1,60 < x ≤ 1,65	5
1,65 < x ≤ 1,70	6
1,70 < x ≤ 1,75	6
1,75 < x ≤ 1,80	3
1,80 < x ≤ 1,85	4
1,85 < x ≤ 1,90	1
TOTAL	30

b) 3 + 4 + 1 = 8 learners are taller than 1,75 m

3) a)

Stem	Leaves
0	2; 3; 3; 5; 6; 6; 6; 7; 8; 8; 9; 9
1	2; 2; 2; 2; 3; 4; 5; 5; 8; 8
2	0; 0; 0; 0; 2; 4; 5
3	0
KEY: 1/2 = 12 hours	

b) Some possible conclusions are

- The learners in the class watched between 2 and 30 hours TV in a week
- More than half of the learners (12 + 10 = 22 learners) watch less than 20 hours of TV in a week
- Four learners watch 20 hours of TV in a week
- Four learners watch 12 hours of TV in a week

4) a)

Number of learners on the bus	Number of buses
0 < l ≤ 10	12
10 < l ≤ 20	6
20 < l ≤ 30	9
30 < l ≤ 40	6
40 < l ≤ 50	4
50 < l ≤ 60	6
TOTAL	

b) The bus company might want to know the numbers of learners on the bus to know whether one bus was enough or if maybe more buses were necessary

5) a) Fraction having blue eyes = $\frac{2\,235 + 2\,449 + 33}{70\,180 + 76\,681 + 1\,144} = \frac{4\,717}{148\,005}$

b) Number with brown eyes = 60 531 + 66 993 + 789 = 128 313 learners

c) The majority of South African learners are black with brown eyes.

CHAPTER 3 – MODE, MEDIAN, MEDIAN

Exercise 3.1 (page 33)

1) a) 9, 9, 11, 15, 15, 15, 15, 17, 18, 19
   b) Two (2) learners are 9 years old.
   c) 19 years
   d) Three (3) learners are older than 15.
   e) Four (4) learners are 15 years old.
   f) The age of 15 years occurs most often.
   g) The mode is the age that occurs most often, therefore 15 years is the mode age.

2) a) 121, 138, 140, 142, 160, 161, 165, 170, 170, 182
   b) 121 cm is the smallest height.
   c) Five (5) learners are taller than 160 cm.
   d) Two (2) learners are 170 cm tall.
   e) 170 cm is the mode of the heights.

3) a) 2; 3; 4; 5; 6; 7; 7; 8; 10; 14
   b) There are ten (10) learners in the data set.
   c) The fewest number of people in a home is two (2).
   d) The most number of people in a home is fourteen (14).
   e) Two (2) homes have 6 people.
   f) Two (2) homes have 7 people.
   g) There are two (2) modes in this data set: 6 and 7 people.
   h) The first seven (7) values are: 10, 14, 6, 8, 3, 2, 6.
   Then you rank the values: 2, 3, 6, 6, 8, 10, 14.
   It is easy to see that six (6) occurs most often. The mode is 6 people.
   i) The first five (5) values are: 10, 14, 6, 8, 3
   Then you rank the values: 3, 6, 8, 10, 14.
   Each value only occurs once, so there is NO mode in this data set.

4) a) Three (3) learners played cricket
   b) Two (2) learners played volleyball
   c) The category "No favourite sport" was selected by the most learners
   d) Tennis and volleyball were chosen by the fewest learners.
   e) "No favourite sport" is the mode. In other words, the most learners indicated this category.
   f) 3 + 3 + 6 + 6 + 2 + 2 + 11 = 34. Thirty four learners took part in the survey.
   g) If the category, "No favourite sport" is ignored, the modal sport is Soccer.
   h) Seven (7) learners chose soccer as their favourite sport.

5) a) Walking is the mode.
   b) 10% × 174 learners = 17,4 learners.
   You cannot have a 17,4 learners, therefore you round 17,4 down to 17 learners travelling to school by bus.
   c) 4% × 174 learners = 6,96 learners.
   You cannot have 6,96 learners, therefore you round 6,96 up to 7 learners travelling to school by taxi.

Exercise 3.2 (page 37)

1) a) The sum of the ages of the learners ins 72
   b) The mean of the ages is: $\frac{72}{5} = 14,4$ years.
2) a) There are ten (10) ages of learners in this data set
   b) The sum of the ages is 143 years

c) The mean of the ages = $\frac{143}{10} = 14,3$ years

d) The mean of the first 5 ages = $\frac{9+15+9+15+17}{5} = \frac{65}{5} = 13$ years

3) a) The mean height of all the learners

$= \frac{138+161+121+170+170+165+142+160+140}{9} = \frac{1\,367}{9} \approx 151,89$ cm.

b) The mean height of the first 6 learners

= $\frac{138+161+121+170+170+165}{6} = \frac{925}{6} \approx 154,17$ cm

Exercise 3.3 (page 39)

1) a) There are 5 learners' ages in this data set
   b) Odd number
   c) 9, 11, 15, 18, 19
   d) 15 years

2) a) Rank the ages from youngest to oldest: 9, 9, 11, 15, 15, 15, 15, 17, 18, 19
   Count and see that there are 10 ages, an even number.
   Because there is not only one age right in the middle of the data set, you add the two middle ages, and divide by 2 to calculate the median.
   b) Median = $\frac{15+15}{2} = \frac{30}{2}$ = 15 years

3) a) 1,13 1,28 1,5 1,53 1,58 1,6 1,65 1,68 1,75 1,77 1,81
   b) 11 heights
   c) Median = 1,65 m

4) a) 0,50; 0,50; 0,58; 0,58; 0,67; 0,75; 0,83; 1,00; 1,33
   b) 9
   c) (i) 0,58 h
   (ii) 0,83 h
   (iii) 1,00 h
   d) Learners in positions 1 and 2 as 0,50 h = 30 minutes
   e) 6^{th} position
   f) The median is in the 5^{th} position
   g) 0,67 h

5) a) Median is in the 61^{st} position
   b) You have to add the number in the 251^{st} position and the number in the 252^{nd} position together and divide the answer by 2.

6) If the total number of data values is an even number, the median will be the sum of the values in the 20^{th} and 21^{st} positions, divided by 2. Therefore, this statement is not valid.

Exercise 3.4 (page 42)

1) a) 9, 11, 15, 18, 19
   b) 19 years
   c) 9 years
   d) 19 – 9 = 10
   e) Range = 10 years

2) First arrange the data values from small to large. 9, 9, 11, 15, 15, 15, 15, 17, 18
   a) 18 years
   b) 9 years
   c) Range = 18 – 9 = 9 years

3) Arrange the data values from small to large: 121, 138, 140, 142, 160, 161, 165, 170, 170, 182
   a) 182 cm
   b) 121 cm
   c) Range = 182 – 121 = 61 cm

4) First order the number of minutes from short to long: 5, 5, 10, 10, 11, 20, 24, 45
   a) 5 minutes
   b) 45 minutes
   c) Range = 45 – 5 = 40 minutes

Exercise 3.5 (page 44)

1) a) 14 values
   b) 4, 5, 5, 13, 15, 15, 15, 15, 18, 20, 25, 45, 50, 90
   c) Most learners take from 4 minutes to 50 minutes to travel to school.
   90 minutes (or 1 ½ hours) is an unexpectedly long time to travel to school. 90 minutes is therefore an extreme value.

2) a) Household 12 had 12 school-going boys
   b) Household 15 had 4 school-going girls
   c) (i) Household 9
   (ii) 0 children
   (iii) Add the number of school-going boys and girls together.
   (iv) This answer is not possible, because the learner who answered the question is still school-going, therefore the answer should be at least 1.
   d) (i) Household 12
   (ii) 13 children
   (iii) Yes, 13 children in a household are possible.
   e) 12 boys are extreme, because all the other households only have from 0 to 3 school-going boys.
   f) There are no extreme (unexpected) values among the girls, because the number of girls only varies from 0 to 4.

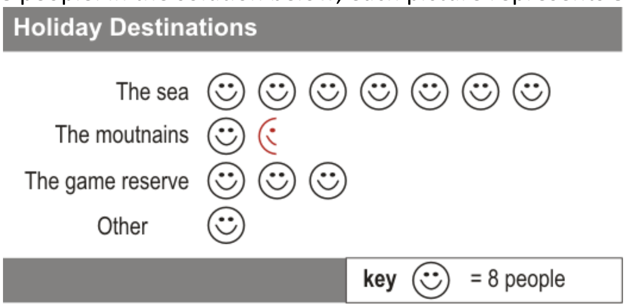
3) a) 109; 140; 145; 146; 150; 151; 155; 159; 160; 162; 165; 168; 169; 170; 170; 171; 173; 176; 178; 181
   b) 20 heights
   c) Modal height = 170 cm
   d) There are 20 values in the data set. This is an even number of values. The middle two numbers of this data set are 162 cm and 165 cm.
   The median = $\frac{162+165}{2} = \frac{327}{2} = 163,5 \text{ cm}$.
   e) Remember, you do not have to rank the data to calculate the mean.
   Mean = $\frac{3\ 198}{20} = 159,9 \text{ cm}$
   f) Range = 181 – 109 = 72 cm
   g) 109 cm is an extreme value. It is a lot less than the rest of the heights.
   h) New mean = $\frac{3\ 198-109}{19} = \frac{3\ 089}{19}$ = 162,6 cm.
   The new mean is larger than the old mean, because the extreme value which was an unexpected low value, was deleted.
   i) The mode is still 170 cm, but the median is now the 10^{th} value and is now 162 cm.

4) In a data set with no mode, all values occur equally often, or there is not one value that occurs more often than the rest. In a data set where the number 0 occurs most often, 0 will be the mode of the data set. For example: in the data set: 0; 1; 2; 3; 4; 5; 6; 7 there is no mode, but in the data set 0; 2; 4; 0; 6; 8; 0, the mode = 0.

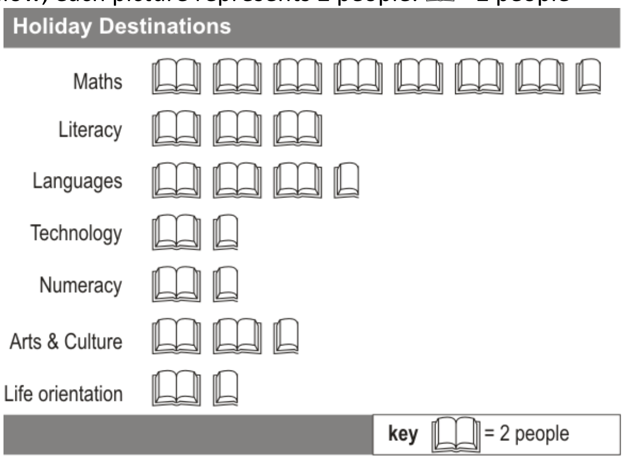
CHAPTER 4: REPRESENTING DATA

Exercise 4.1 (page 46)

1) Note: Answers will vary, depending on what key you choose. You could choose each picture to represent 2, 4 or 8 people. In the solution below, each picture represents 8 people.

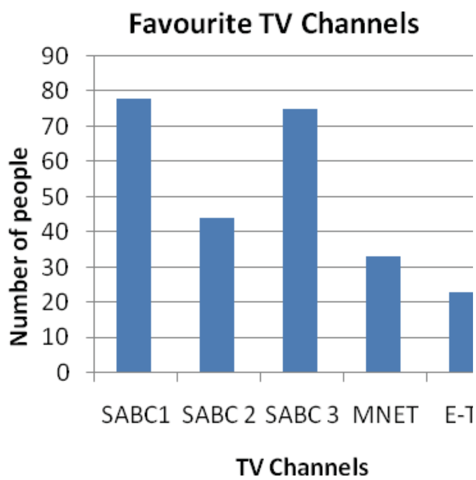


2) Note: Answers will vary, depending on what key you choose
In the solution below, each picture represents 2 people: 🕮= 2 people

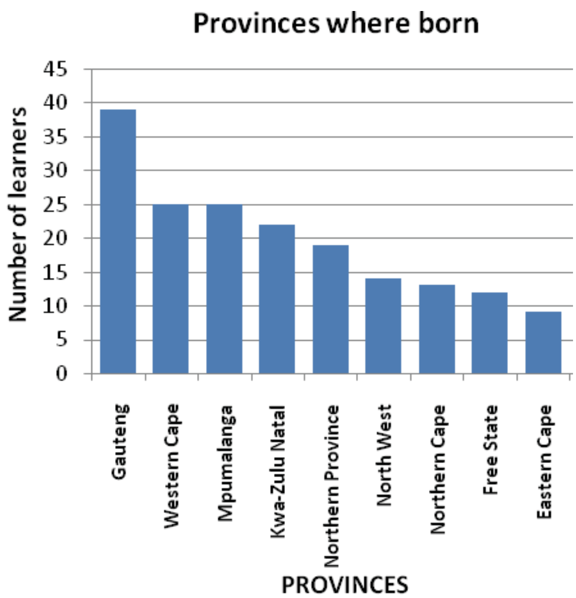


Exercise 4.2 (page 54)

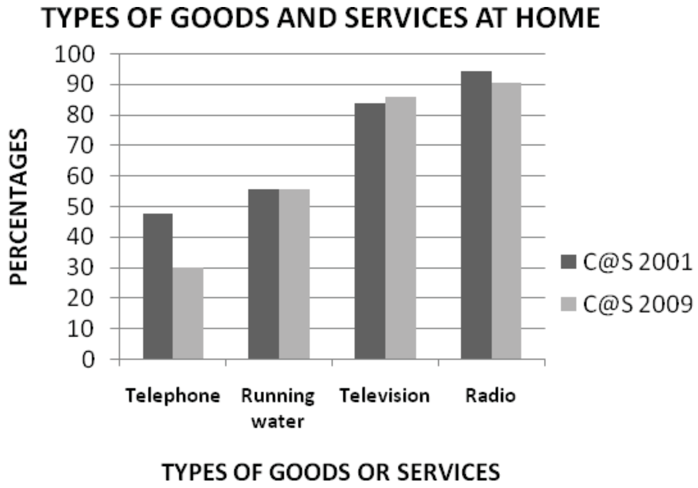
1)



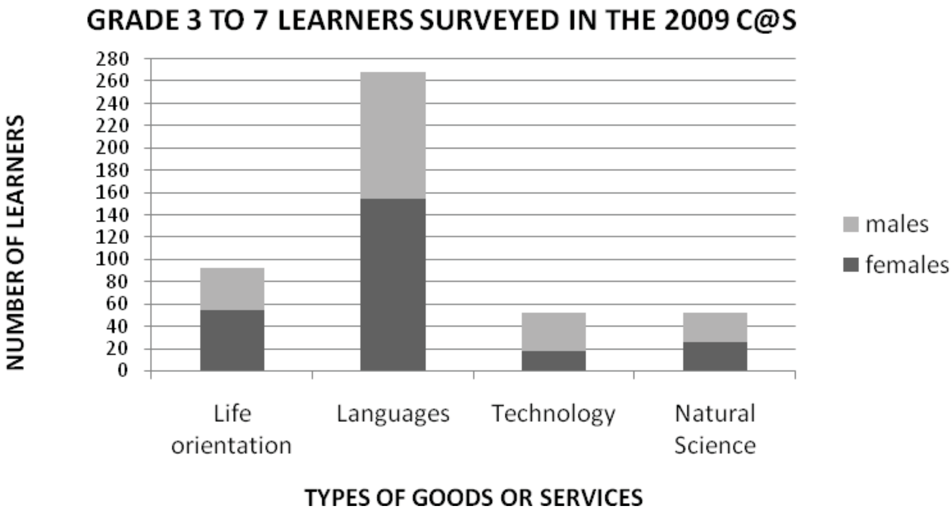
2)



Exercise 4.3 (page 57)
1)



Exercise 4.4 (page 59)
1)

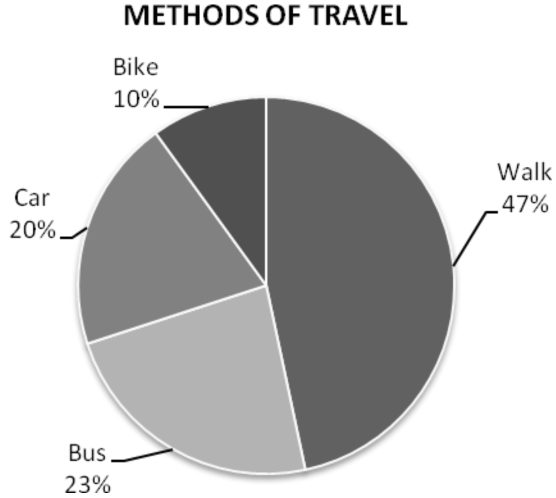


2) Answers will vary, but may include:
- The languages are favourite subjects for both boys and girls.
- Natural Science and Technology overall are the least favourites across both genders
- Natural Science is rated equally by girls and boys in the survey.

Exercise 4.5 (page 62)
1)

Method of travel	Number of learners	Angle
Walk	14	14 × 12° = 168°
Bus	7	7 × 12° = 84°
Car	6	6 × 12° = 72°
Bike	3	3 × 12° = 36°
Helicopter	0	0 × 12° = 0°

1) continued

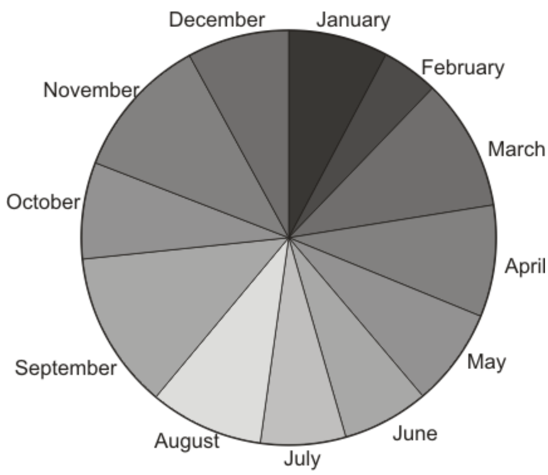


2)

Month	No of people	Angle
January	7	28°
February	4	16°
March	9	36°
April	8	32°
May	7	28°
June	6	24°

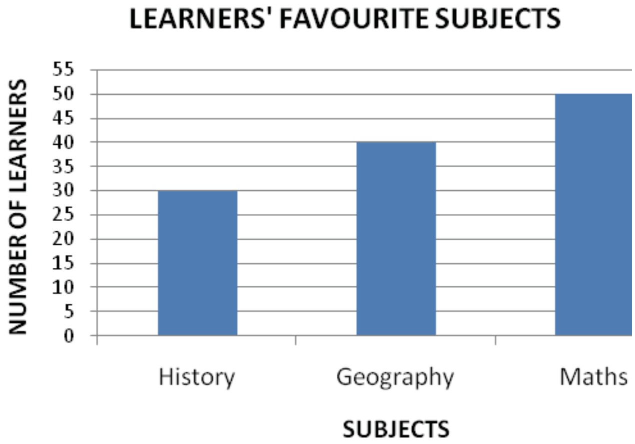
Month	No of people	Angle
July	6	24°
August	8	32°
September	11	44°
October	7	28°
November	10	40°
December	7	28°

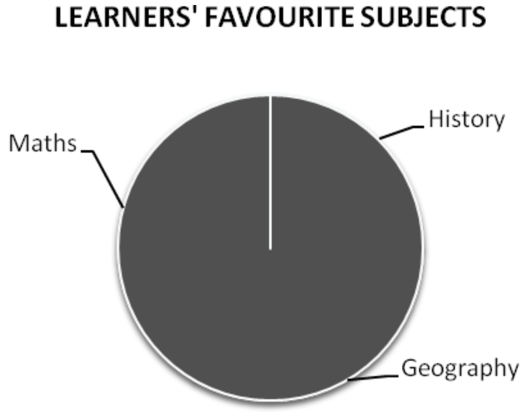
NO OF BIRTHDAYS PER MONTH



3) a)

FAVOURITE SUBJECT	NUMBER OF LEARNERS
History	30
Geography	40
Maths	50
TOTAL	**120**

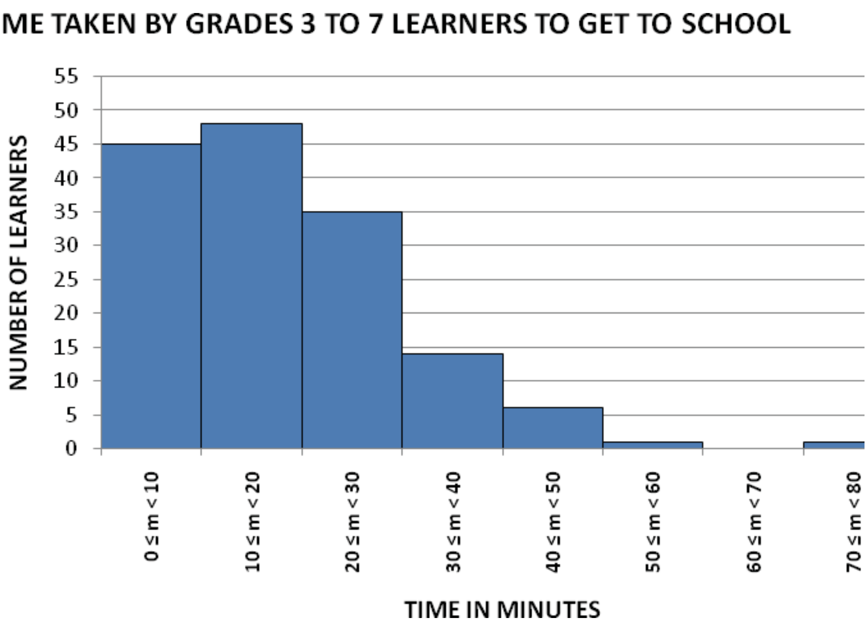




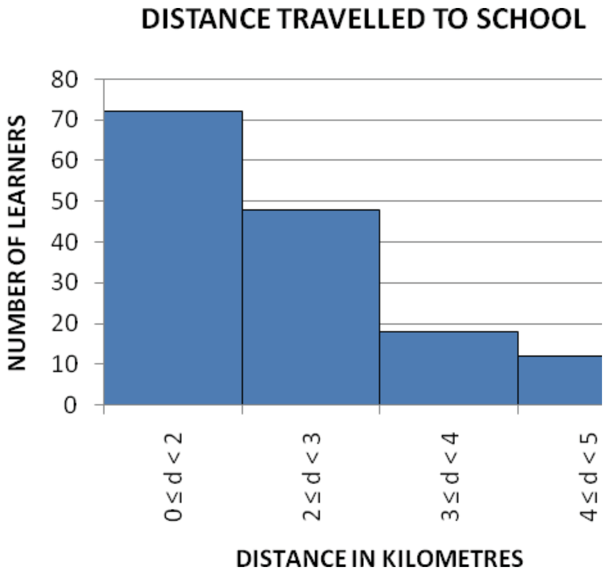
b) Answers will vary. The pie chart show both the difference between the groups and how each group compares to the whole, so this seems to be a good representation of the data.

EXERCISE 4.6 (page 65)

1)

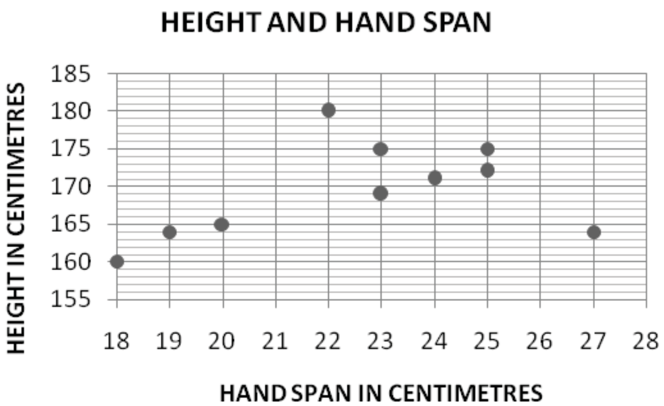


2)

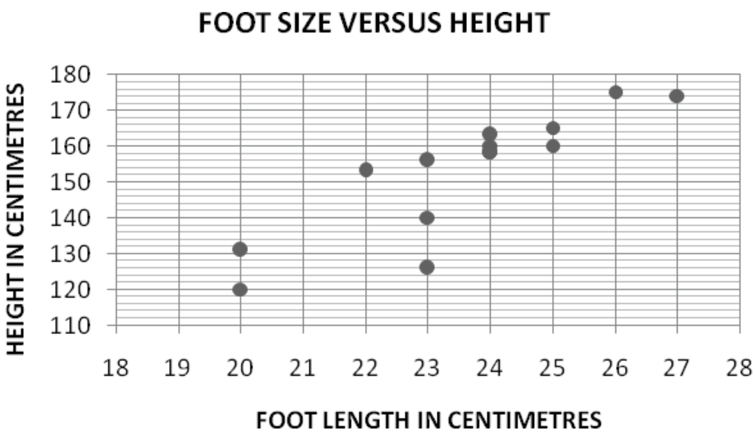


Exercise 4.7 (page 68)

1) Except for the people who have measurements (22; 153) and (27; 174), the taller people generally have wider hand spans.



2) Except for the people who have measurements (20; 120), and (23; 126), the taller people generally have longer foot lengths.



CHAPTER 5 – INTERPRETING DATA

Exercise 5.1 (page 72)

1) a) The least learners were born in November
   b) The most learners were born in September
   c) If numbers were used, the scale would take up too much space
   d) The graph may be easier to read, but the table gives the best representation. This is because we can only read approximate percentages off the graph.
2) The heading is missing so we don't know what the graph is about
3) They look different because the values on the vertical axes are different.
   Graph A: values go from 0 to 1 000 in steps of 200; Graph B: values go from 0 to 2 400 in steps of 400

Exercise 5.2 (page 75)

1) a) The graph compares the facilities and services at schools from the 2001 Census@School survey and the 2009 Census@School survey.
   b) The first column represents results from the 2001 Census@School survey and the second column represents the results from the 2009 Census@School survey.
   c) Sentences you could write include:
      - The percentage of facilities and services at schools has improved in all cases from 2001 to 2009
      - The highest improvement has been in the percentages of computers at school
      - Only a small percentage of the schools surveyed have access to email and/or internet
      - Approximately 70% of the schools surveyed in 2009 had access to a maths teacher.

2) a) Grades 3 to 7: Maths;
      Grades 8 to 12: Languages.
   b) Maths; Languages; Literacy; Arts & Culture; Technology; Numeracy; Life Orientation
   c) Life Orientation

3) a) Some of the sentences you could write include:
      - Netball is the girls' favourite sport for females, whereas the boys' favourite sport is soccer
      - No girls prefer rugby as their favourite sport, whereas approximately 18% of the boys prefer rugby as their favourite sport
      - Tennis is more popular amongst girls than amongst boys
      - There are many more girls who do not have a favourite sport as compared to the boys
   b) Yes they have. We know this because the lengths of most of the bars touching each other are different.

4) Some possible sentences you could write include:
   - The largest percentage of boys and girls would prefer to be happy
   - The smallest percentage of boys and girls would prefer to be famous
   - The percentage of learners who would prefer to be famous is the same for boys as for girls
   - There is a higher percentage of boys that would prefer to be rich than girls
   - There is a higher percentage of girls that would prefer to be happy than the percentage of boys.

Exercise 5.3 (page 79)

1) a) The graphs show the percentage of food types eaten by males and females
   b) Carbohydrates
   c) Fruit and vegetables
   d) Most: carbohydrates; least: Protein
   e) Females eat 2% more dairy than males

2) a) The bar graph gives a better visual representation of the data, but the table gives more detail
   b) The most learners were in Grade 9
   c) There were more girls. We can see this from the table.
   d) You will find out approximately how many learners participated in the census by adding up all the bars. However, the scale makes it impossible to find the exact number represented by each bar.
      You could not use the pie graph to say how many learners participated in the survey. The pie graph only shows percentages of the whole census.
   e) The bar graph shows the **percentage** of boy and girl learners in each grade as opposed to the **number** of learners in each grade shown in the first bar graph.
   f) In Grades 3 and 4 there are more boys than girls.
      In the rest of the grades there are more girls than boys.

Exercise 5.4 (page 84)

1) They want you to believe that the temperatures in August 2010 were about the same as the temperatures in October 2010. In reality, the temperatures in August were about 9° lower than those in October.

2) a) The scale on the vertical axis doesn't start at 0
   a) One possible answer: It looks like there were no visitors to the park on Mondays and Fridays

Exercise 5.5 (page 86)

1) a) Graph A: 20 cm; Graph B: 60 cm
   b) Graph A shows clearly that the heights of most of the learners lie between 160 cm and 200 cm. This is not shown in Graph B

2) a) (i) The oldest boys who participated in the survey were 18 or 19 years old
   (ii) The youngest boys were 11 or 12 years old
   (iii) The oldest girls were 16 or 17 years old
   (iv) The youngest girls were 11 or 12 years old
   (v) The mode for the boys is 15 to 16 years
   (vi) The mode for the girls is 15 to 16 years
   b) Some of the sentences you could write are:
   - The boys are older than the girls
   - More girls who were 14 or 15 responded than boys of the same age.

Exercise 5.6 (page 89)
The shaded cells are the one where there are errors.

Dirty Spreadsheet

Row 1	Boy or Girl	Date of Birth	Grade	Height in cm	Foot Length in cm	Favourite Subject	Distance to School in km
Row 2	Boy	12/04/91	5	143	26	Art	1-2km
Row 3	Girl	31/02/92	4	132	22	Science	less than 2 km
Row 4	Girl	14/01/91	5,00	14,2	2,3	PE/Sport	2,5423 km
Row 5	Boy	07/09/89	6	136	25	Art	1-2km
Row 6	Boy	13/12/91	4	128	24	PE/Sport	1-2km
Row 7	Boy	14/03/01	5	140	67	PE/Sport	less than 1 km
Row 8	Girl	06/05/89	7	142	24	Art	3-5km
Row 9	Girl	15/08/90	6	138	21	Art	85km
Row 10	Boy	20/02/90	6	192	23	PE/Sport	1-2km
Row 11	Girl	19/05/90	6	140	20	Maths	1-2km
Row 12	Neither	29/06/92	7	48	21	Going Home	3 000km
Row 13	Boy	09/10/91	4	128	21	English	less than 1 km
Row 14	Girl	18/12/90	5	135	21	Geography	less than 1 km
Row 15	Girl	18/07/91	0,5	13,7	20	Art	3-5km
Row 16	Boy	03/06/34	4	129	21	Art	less than 1 km
Row 17	Girl	13/02/89	7	148	23	Art	1-2km
Row 18	Girl	15/09/88	7	150	22,5	PE/Sport	1-2km
Row 19	Girl	07/08/89	7	140	24	Art	less than 1 km
Row 20	Boy	08/06/89	7	142	24	Computing	less than 1 km
Row 21	Boy	31/11/87	11	1 520	22	Computing	5-10km
Row 22	Both	16/07/88	8	142	26	Japanese	2-3 km
Row 23	Girl	28/04/88	8	145	26,5	PE/Sport	1 mile
Row 24	Boy	25/03/92	4,1	132,1	2,4,5	Maths	less than 1 km
Row 25	Boy	26/02/92	4	130	21	PE/Sport	less than 1 km
Row 26	Girl	08/07/99	6	142	22	Art	2-3 km
Row 27	Boy	23/05/90	6	151	25,5	Maths	2-3 km
Row 28	Boy	01/03/87	9	162	25	PE/Sport	less than 1 km
Row 29	Girl	07/08/91	6	150	23	History	2 roads
Row 30	Girl	03/03/92	4	135	21	English	less than 1 km

CHAPTER 6 – RELATIVE FREQUENCY AND PROBABILITY

Exercise 6.1 (page 93)

1) a) Possible outcomes: 1; 2; 3; 4; 5; 6; 7; 8
   c) Event A: 2; Event B: 1; 3; 5; 7; Event C: 5; 6; 7; 8

2) a) Possible outcomes: R5; R2; R1; R1; 20c
   b) Event D: R1; R1; Event E: 20c; Event F: R5

3) a) Possible outcomes: M; A; T; H; S
   b) The vowels are a, e, i, o and u. All other letters of the alphabet are called consonants,
   Event G: T; Event H: A; Event J: M; T; H; S

4) a) Possible outcomes: J♥; 7♣; K♣; 7♥; Q♠
   b) Event K: 7♣; 7♥; Event L: J♥; 7♥; Event M: J♥; K♣; Q♠; Event N: None

5) a) Possible outcomes: Walk/Foot; Car; Train; Bus; Bicycle; Scooter; Taxi; Other
   b) Event P: Taxi; Event Q: Train; Bus; Taxi; Event R: Bicycle; Scooter

Exercise 6.2 (page 96)

You only have to give one answer, unless your teacher asks specifically for your answer to be a decimal fraction or a percentage. If you leave your answer as a common fraction – make sure it is in its simplest form. All three answers are given here.

1) Relative frequency of flowers $= \dfrac{\text{Number of times a bulb flowers}}{\text{Total number of bulbs planted}} = \dfrac{36}{40} = \dfrac{9}{10}$ = 0,9 = 90%

2) We could calculate the relative frequency after 10 games, or after 30 games. It is better to calculate the relative frequency after 30 games, as the more times an experiment is repeated, the more accurate our prediction will be.

   a) (i) Relative frequency of Helen's wins

   $= \dfrac{\text{Number of games won by Helen}}{\text{Number of games played}} = \dfrac{21}{30} = \dfrac{7}{10}$ = 0,7 = 70%

   (ii) Relative frequency of Thandi's wins

   $= \dfrac{\text{Number of games won by Thandi}}{\text{Number of games played}} = \dfrac{9}{30} = \dfrac{3}{10}$ = 0,3 = 30%

   b) It is unlikely that Thandi will win her next match against Helen as 0,3 is much lower than 0,7. We cannot however, say for certain that Thandi will not win.

3) a) Alfred plays 4 + 2 + 6 = 12games
   Busi plays 8 + 1 + 7 = 16 games
   Fikile plays 3 + 0 + 1 = 4 games
   Pieter plays 9 + 2 + 9 = 20 games.

   b) Relative frequency that Alfred wins

   $= \dfrac{\text{Number of times Alfred has won}}{\text{Number of games Alfred has played}}$

   $= \dfrac{4}{12} = \dfrac{1}{3}$ ≈ 0,333 (correct to three decimal places) ≈ 33% (correct to nearest %)

   Relative frequency that Busi wins

   $= \dfrac{\text{Number of times Busi has won}}{\text{Number of games Busi has played}} = \dfrac{8}{16} = \dfrac{1}{2}$ = 0,5 = 50%

   Relative frequency that Fikile wins

$= \dfrac{\text{Number of times Fikile has won}}{\text{Number of games Fikile has played}} = \dfrac{3}{4}$ = 0,75 = 75%

Relative frequency that Pieter wins

$= \dfrac{\text{Number of times Pieter has won}}{\text{Number of games Pieter has played}} = \dfrac{9}{20}$ = 0,45 = 45%

c) Pieter is more likely to win than Alfred because the relative frequency of Pieter winning (0,45) is higher than the relative frequency of Alfred winning (0,333)

4) a) (i) Relative frequency of getting a yellow counter after 100 trials

$= \dfrac{\text{Number of yellow counters}}{\text{Number of trials}} = \dfrac{54}{100}$ = 0,54 = 54%

(ii) Relative frequency of getting a yellow counter after 200 trials

$= \dfrac{\text{Number of yellow counters}}{\text{Number of trials}} = \dfrac{101}{200}$ = 0,505 = 50,5%

(iii) Relative frequency of getting a yellow counter after 300 trials

$= \dfrac{\text{Number of yellow counters}}{\text{Number of trials}} = \dfrac{135}{300} = \dfrac{9}{20}$ = 0,45 = 45%

b) The most accurate of these relative frequencies is 0,45 because we used the most trials to calculate this relative frequency.
The number of yellow counters in a bag is 0,45 × 10 = 4,5
As there are no fractions of counters, we can predict that there are 4 or 5 yellow counters in the bag.

Exercise 6.3 (page 99)

1) a) Possible outcomes: Red; Blue; Green OR R; B; G

b) (i) The number of possible outcomes is: 3; The favourable outcome is: R;
There is 1 favourable outcomes

$$P(R) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}}$$

$= \dfrac{1}{3}$ ≈ 0,333 (correct to three decimal places) ≈ 33% (correct to nearest %)

(ii) The number of possible outcomes is: 3; The favourable outcomes are: R; G;
There are 2 favourable outcomes

$$P(RorG) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}}$$

$= \dfrac{2}{3}$ ≈ 0,667 (correct to three decimal places) ≈ 67% (correct to nearest %)

(iii) The number of possible outcomes is: 3; The favourable outcome is: R; G;
There are 2 favourable outcomes

$$P(notB) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}}$$

$= \dfrac{2}{3}$ ≈ 0,667 (correct to three decimal places) ≈ 67% (correct to nearest %)

(iv) The number of possible outcomes is: 3; There is NO favourable outcomes;
There are 0 favourable outcomes

$P(Y) = \dfrac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}} = \dfrac{0}{3}$ = 0 = 0%

2) a) Possible outcomes: R; R; R; G; G; G; G; G; G; G

b) You are more likely to get a green sweet because there are more of them.

c) (i) The number of possible outcomes is: 10; The favourable outcome is: R: R; R;
There are 3 favourable outcomes

$$P(R) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}} = \frac{3}{10}$$ = 0,3 = 30%

(ii) The number of possible outcomes is: 10; The favourable outcome is: G: G; G; G; G; G; G
There are 7 favourable outcomes

$$P(G) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}} = \frac{7}{10}$$ = 0,7 = 70%

3) a) Possible outcomes:
Lose a turn; Lose 5 points; Win 5 points; Go back to start; Have an extra go

b) (i) The number of possible outcomes is: 5; The favourable outcome is: Loses a turn
There is 1 favourable outcome

$$P(\text{Loses a turn}) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}} = \frac{1}{5}$$ = 0,2 = 20%

(ii) The number of possible outcomes is: 5;
The favourable outcomes are: Lose 5 points; Win 5 points
There are 2 favourable outcomes

$$P(\text{Lose or Win 5 pts}) = \frac{\text{Number of favourable outcomes in an event}}{\text{Total number of possible outcomes}} = \frac{2}{5}$$ = 0,4 = 40%

4) a) Possible outcomes: P; R; O; B; A; B; I; L; I; T; Y

b) (i) The number of possible outcomes is: 11; The favourable outcome is: T
There is 1 favourable outcome

P(T) = $\frac{1}{11}$ = 0,091 (correct to three decimal places) = 9% (correct to nearest %)

(ii) The number of possible outcomes is: 11; The favourable outcomes are: B; B
There are 2 favourable outcomes

P(B) = $\frac{2}{11}$ = 0,182 (correct to three decimal places) = 18% (correct to nearest %)

5) a) Number of possible outcomes: 52; Number of favourable outcomes: 26

P(red card) = $\frac{26}{52} = \frac{1}{2}$ = 0,5 = 50%

b) Number of possible outcomes: 52; Number of favourable outcomes: 13

P(heart) = $\frac{13}{52} = \frac{1}{4}$ = 0,25 = 25%

c) Number of possible outcomes: 52; Number of favourable outcomes: 1

P(Ace of hearts) = $\frac{1}{52}$

= 0,019 (correct to three decimal places) = 2% (correct to nearest %)

6) a) There are 10 possible outcomes.

b) (i) P(red) = $\frac{4}{10} = \frac{2}{5}$ = 0,4 = 40%

(ii) P(white or blue) = $\frac{6}{10}$ = 0,6 = 60%

(iii) P(red or white or blue) = $\frac{10}{10}$ = 1 = 100%

(iv) P(green) = $\frac{0}{10}$ = 0 = 0%

7) a) There are 12 possible outcomes; Favourable outcomes: 43; 44; 45; 46; 47; 48; 49; 54
There are 8 favourable outcomes: 8
P(at least one 4)

$= \frac{8}{12} = \frac{2}{3}$ ≈ 0,667 (correct to three decimal places) ≈ 67% (correct to nearest %)

b) There are 12 possible outcomes; Favourable outcomes: 50; 51; 52; 53
There are 4 favourable outcomes

P(no 4) $= \frac{4}{12} = \frac{1}{3}$ ≈ 0,333 (correct to three decimal places) ≈ 33,3% (correct to nearest %)

c) There are 12 possible outcomes; Favourable outcomes: 43; 44; 45; 46; 47; 48; 49; 53; 54
There are 9 favourable outcomes

P(3 or 4) = $\frac{9}{12}$ = 0,75 = 75%

8) The outcomes of winning, drawing or losing are ***not equally likely***.
We can only calculate probabilities using the Probability formula when the outcomes are equally likely. In this case there are other factors which would influence the chances of Pirates winning.

9) a) There are 8 possible outcomes.

b) (i) P(6) = $\frac{1}{8}$ = 0,125 = 12,5%

(ii) P(number less than 6) = $\frac{5}{8}$ = 0,625 = 62,5%

(iii) P(odd number) = $\frac{4}{8} = \frac{1}{2}$ = 0,5 = 50%

(iv) P(number more than 6) = $\frac{2}{8} = \frac{1}{4}$ = 0,25 = 25%

(v) P(prime number) = $\frac{4}{8} = \frac{1}{2}$ = 0,5 = 50%

10) a) Number of possible outcomes: 240; There is 1 favourable outcome

P(winning) $= \frac{1}{240}$ = 0,00417 (correct to 5 decimal places) = 0,4% (correct to 0,1%)

b) Number of possible outcomes: 52; There are 7 favourable outcomes

P(winning) $= \frac{7}{240}$ = 0,02917 (correct to 5 decimal places) = 2,92% (correct to 0,01 %)

11) a) There are 90 possible outcomes.

b) (i) Favourable outcome: 69; there is 1 favourable outcome

P(69) $= \frac{1}{90}$ = 0,011 (correct to three decimal places) = 1% (correct to nearest %)

(ii) Favourable outcome: 10; 20; 30; 40; 50; 60; 70; 80; 90;
There are 9 favourable outcomes

P(ends in 0) = $\frac{9}{90} = \frac{1}{10}$ = 0,1 = 10%

(iii) Favourable outcome: 9; 18; 27; 36; 45; 54; 63; 72; 81; there are 9 favourable outcomes

P(multiple of 9) = $\frac{9}{90} = \frac{1}{10}$ = 0, 1 = 10%

12) a) They are most likely to be born in September

b) Number of possible outcomes: 44 954; there are 3 922 favourable outcomes

P(March) = $\frac{3\ 922}{44\ 954} = \frac{1\ 961}{22\ 477}$

= 0,0872 (correct to 4 decimal places) = 8,7% (correct to 0,1 %)

Exercise 6.4 (page 104)
You may find that your results are similar to Makgoshi's results.
However, because this is an experiment, your results may be very different to Makgoshi's results.

Exercise 6.5 (page 106)

1) a) P(black) = $\frac{5}{50} = \frac{1}{10}$ = 0,1 = 10%

b) Predicted number of red cars = P(red) × Number of cars = 0,2 × 50 = 10

2) a) P(win) = $\frac{1}{500}$ = 0,002

b) Predicted number of tickets = P(win) × Total number of tickets sold = $\frac{1}{20}$ × 500 = 25

3) Predicted number of faulty calculators = P(faulty calculator) × Total number of tickets sold
= 0,02 × 300 = 6

Exercise 6.6 (page 109)

1) a)

		Die					
		1	2	3	4	5	6
Coin	**H**	H; 1	H; 2	H; 3	H; 4	H; 5	H; 6
	T	T; 1	T; 2	T; 3	T; 4	T; 5	T; 6

i) P(H; 5) = $\frac{1}{12}$ = 0,083 (correct to three decimal places) = 8% (correct to nearest %)

ii) P(T; 6) = $\frac{1}{12}$ = 0,083 (correct to three decimal places) = 8% (correct to nearest %)

iii) P(T; even) = $\frac{3}{12} = \frac{1}{4}$ = 0,25 = 25%

iv) P(T; odd) = $\frac{3}{12} = \frac{1}{4}$ = 0,25 = 25%

v) P(H: number greater than 3) = $\frac{3}{12} = \frac{1}{4}$ = 0,25 = 25%

vi) P(odd) = $\frac{6}{12} = \frac{1}{2}$ = 0,5 = 50%

2) a)

+		2nd die					
		1	**2**	**3**	**4**	**5**	**6**
1st die	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

b) (i) P(total of 10)

= $\frac{3}{36} = \frac{1}{12}$ = 0,083 (correct to 3 decimal places) = 8% (correct to nearest %)

(ii) P(greater than 10)

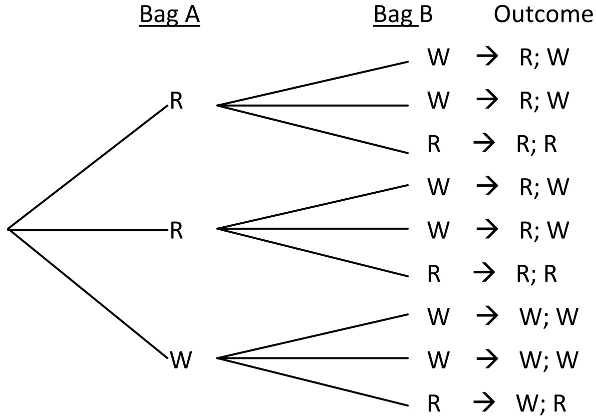
= $\frac{30}{36} = \frac{5}{6}$ = 0,833 (correct to three decimal places) = 83% (correct to nearest %)

(iii) P(less than 10)

= $\frac{3}{36} = \frac{1}{12}$ = 0,083 (correct to three decimal places = 8% (correct to nearest %)

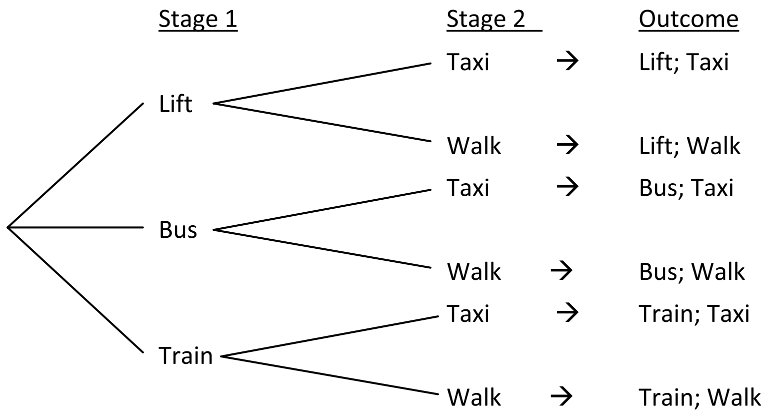
c) The probability is equal to 1 because less than 10 and greater than 10 includes all the possible outcomes.

3) a)



b) P(RR or WW) $= \frac{4}{9}$ = 0,444 (correct to four decimal places) = 44% (correct to nearest %)

4) a)



b) P(bus; taxi) $= \frac{1}{6}$ = 0,167 (correct to four decimal places) = 17% (correct to nearest %)

5) a)

		First Spin			
		R10	**R5**	**R5**	**R50**
Second Spin	**R10**	R20	R15	R15	R60
	R5	R15	R10	R10	R55
	R5	R15	R10	R10	R55
	R50	R60	R55	R55	R100

b) (i) P(R20) = $= \frac{1}{16}$ = 0,0625 = 6% (correct to nearest %)

(ii) P(R100) $= \frac{1}{16}$ = 0,0625 = 6%

(iii) P(R60) = $\frac{2}{16} = \frac{1}{8}$ = 0,125 = 12,5%

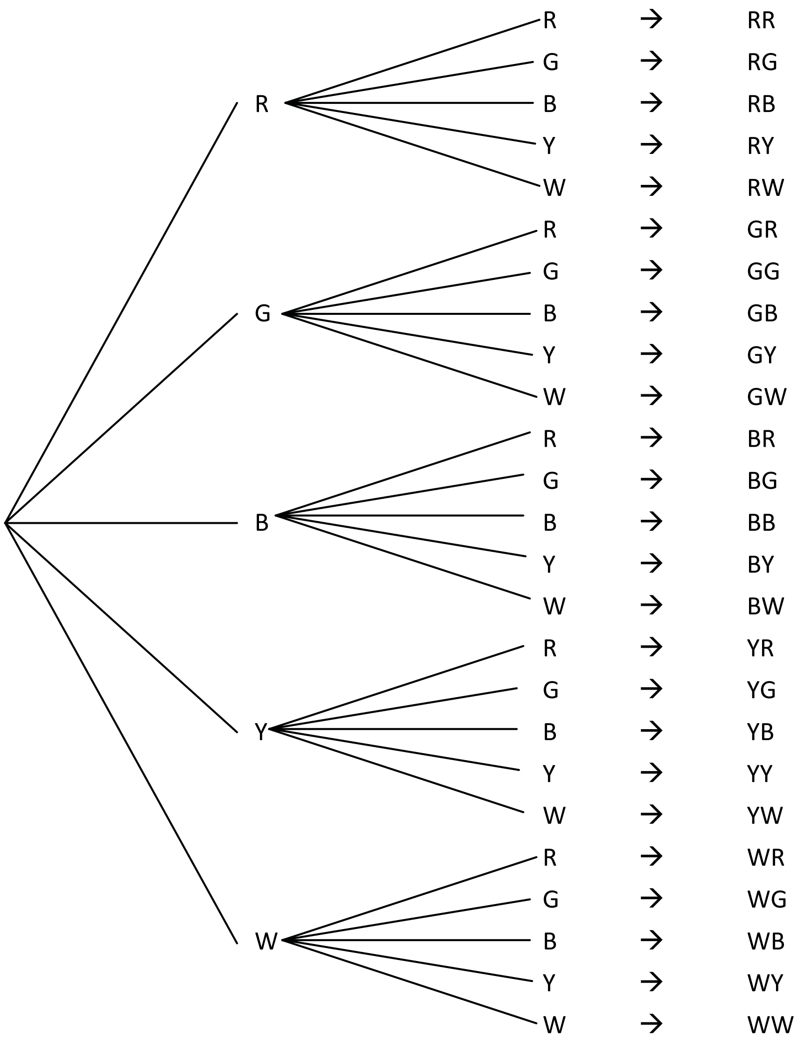
(iv) P(R15) $= \frac{4}{16} = \frac{1}{4}$ = 0,25 = 25%

6) a) TWO-WAY TABLE:

		1st Spin				
		R	**G**	**B**	**Y**	**W**
2nd Spin	**R**	R; R	G; R	B; R	Y; R	W; R
	G	R; G	G; G	B; G	Y; G	W; G
	B	R; B	G; B	B; B	Y; B	W; B
	Y	R; B	G; Y	B; Y	Y; Y	W; Y
	W	R; W	G; W	B; W	Y; W	W; W

TREE DIAGRAM:

1st Spin 2nd Spin Outcomes



b) (i) P(W; W) $= \frac{1}{25}$ = 0,04 = 4%

(ii) The favourable outcomes here are: RW; GW; BW; YW; WR; WG; WB; WY; WW

P(at least one W) $= \frac{9}{25}$ = 0,36 = 36%

(iii) Here the favourable outcomes are: RR; GG; BB; YY; WW.

P(both the same colour) = $\frac{5}{25} = \frac{1}{5}$ = 0,2 = 20%

7) a) R; B; Y R; Y; B B; R; Y B; Y; R Y; R; B Y; B; R

b) P(red is first) = $\frac{2}{6} = \frac{1}{3}$ ≈ 0,333 (correct to three decimal places) ≈ 33% (correct to nearest %)