Census@school

Data Handling & Probability

Grades 10, 11 and 12

Statistics South Africa

your leading partner in quality statistics

Census@School – Data Handling and Probability (Grades 10, 11 and 12)

Statistics South Africa, 2013

Pali Lehohla, Statistician-General

Census@School – Data Handling and Probability (Grades 10, 11 and 12) / Statistics South Africa

Published by Statistics South Africa, Private Bag X44, Pretoria 0001

© Statistics South Africa, 2013
Users may apply or process this data, provided Statistics South Africa (Stats SA) is acknowledged as the original source of the data; that it is specified that the application and/or analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.

Stats SA Library Cataloguing-in-Publication (CIP) Data

Census@School – Data Handling and Probability (Grades 10, 11 and 12) / Statistics South Africa. Pretoria: Statistics South Africa, 2013

238pp

ISBN: **978-0-621-41837-8**

A complete set of Stats SA publications is available at Stats SA Library and the following libraries:

National Library of South Africa, Pretoria Division
National Library of South Africa, Cape Town Division
Library of Parliament, Cape Town
Bloemfontein Public Library
Natal Society Library, Pietermaritzburg
Johannesburg Public Library
Eastern Cape Library Services, King William's Town
Central Regional Library, Polokwane
Central Reference Library, Nelspruit
Central Reference Collection, Kimberley
Central Reference Library, Mmabatho

This publication is available on the Stats SA website: www.statssa.gov.za

Copies are obtainable from: Printing and Distribution, Statistics South Africa

Tel: (012) 310 8093
(012) 310 8619
Email: inadp@statssa.gov.za
millies@statssa.gov.za

Foreword

H G Wells, sometimes called the father of modern science fiction, opined that at the turn of the 21st Century numeracy will be as important to humanity as is the ability to read and write. That Statistics South Africa (Stats SA) is committed towards building society wide statistical capacity therefore comes as no surprise in that it fulfils the demand that this prophecy of H G Wells inspired. By looking outside-in StatsSA took a conscious decision to create a programme that focuses on schools and the public to engender the love for statistics and its use. Statistical evidence is a critical decision support that creates possibilities for rational behaviour, decision making and development of society.

A practical way in which StatsSA undertakes and commits to execute this strategy is by embarking on a range of catalytic projects under ISIbalo, a legacy programme created by South Africa at the 57th Session of the International Statistical Institute (ISI), hosted in Durban, South Africa in August of 2009. The 2009 Census@School (C@S) project, first undertaken in 2001 in support of popularising Census 2001, was a repeat project 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 predecessor project was so successful that one of the learners added that before Census@Schools she did not know how tall she was, thus fulfilling albeit anecdotally awareness of how data are gathered and its benefits to individuals and society. In addition and importantly, data collected provides contextual material for teachers and learners for teaching and learning of data handling, and promoting statistical literacy relating to a variety of subjects.

The first series of Mathematics Study Guide on Data Handling and Probability for the Senior Phase (Grades 7–9) using the 2009 C@S data was developed in 2011. Stats SA has achieved yet another milestone by developing a Further Education and Training (FET) (Grades 10–12) Mathematics Study Guide on Data Handling and Probability, as a contribution towards teaching and learning support material. Examples and exercises are based on the 2009 C@S and Census 2011 data. It is anticipated that teachers and learners would interact with data that was collected from their real-life situations and thus make learning interesting and enjoyable. A fast growing and innovative way of getting into schools is the Soccer4Stats series which bring the love for Mathematics and Statistics through gaming and this will be added as part of the essential material for building capacity. Johnny Masegela, Black Sunday of Orlando Pirates came with up with this innovation that has inspired not only South African schools but the world.

This milestone has been achieved through the collaboration and support from the national and provincial Departments of Basic Education with regard to the Census@School project.

Pali Lehohla
Statistician-General

CONTENTS

1

Grade 10 Data Handling

In this chapter you will:

- Revise the language of data handling
- Determine measures of central tendency of lists of data, of data in frequency tables and data in grouped frequency tables
- Determine quartiles and the five number summary
- Determine percentiles
- Determine measures of dispersion (range and inter-quartile range)
- Illustrate the five number summary with a box and whisker diagram

WHAT YOU LEARNED ABOUT DATA HANDLING IN GRADE 9

In Grade 9 you covered the following data handling concepts:

- ***Collecting data :*** *including distinguishing between samples and populations*
- ***Organising and summarising data :*** *using tallies, tables and stem-and-leaf displays; determining measures of central tendency (mean, median, mode); determining measures of dispersion (range, extremes, outliers)*
- ***Representing data :*** *drawing and interpreting bar graphs, double bar graphs, histograms, pie charts, broken-line graphs, scatter plots.*
- ***Interpreting data :*** *critically reading and interpreting two sets of data represented in a variety of graphs.*
- ***Analysing data :*** *critically analysing data by answering questions related to data collection methods, summaries of data, sources of error and bias in the data*
- ***Reporting data*** *by drawing conclusions about the data; making predictions based on the data; making comparisons between two sets of data; identifying sources of error and bias in the data; choosing appropriate summary statistics (mean, median, mode, range) for the data and discussing the role of extremes and outliers in the data*

THE LANGUAGE OF DATA HANDLING

✓ The word ***data*** is the plural of the word ***datum*** which means "a piece of information". So data are pieces of information.

a) Organising Data

✓ In order to make sense of the data, we need to ***organise*** the data.

✓ Different sets of data can be sorted in different ways:

- You can write the data items in either ***alphabetical*** or ***numerical*** order.

  For example:
  - The words elephant; lion; frog and crocodile can be ordered in ***alphabetical order*** as follows: crocodile; elephant; frog; lion
  - The numbers 32,1; 32,001; 32,0001 and 32,01 can be ordered in ***ascending numerical order*** as follows: 32,0001; 32,001; 32,01 and 32,1

- You can sort data items using a ***tally table***.

  A ***tally*** is a way of collecting information by making an appropriate mark for each item.
  A line is drawn for each item counted : ||||
  Every ***fifth*** line is drawn across the other four : ~~||||~~. This makes it easy to add up the number of items checked.

  For example:
  The following ***tally table*** shows the favourite fruit of sixteen Grade 10 learners.

<table>
<tr><th>Favourite Fruit</th><th>Number of learners</th></tr>
<tr><td>Apple</td><td>~~||||~~ ||</td></tr>
<tr><td>Banana</td><td>~~||||~~ ||||</td></tr>
</table>

✓ When collecting data, the number of times a particular item occurs is called its ***frequency***.

For example:
The following ***frequency tables*** show the same information about the favourite fruit of the sixteen Grade 10 learners.

Favourite Fruit	Frequency
Apple	7
Banana	9
TOTAL	16

Favourite Fruit	Apple	Banana	**TOTAL**
Frequency	7	9	16

✓ We can use a ***stem-and-leaf diagram*** to organise data.

With a stem-and-leaf diagram, we organise the data by using place value:

- The digits in the ***largest place*** are referred to as the ***stem.***
- The digits in the ***smallest place*** are referred to as the ***leaf*** (or leaves).
- The leaves are displayed to the right of the stem.

This means that for the number 45, the digit 4 is the stem and the 5 is the leaf.

EXAMPLE 1

Organise the following set of 25 data items using a stem-and-leaf-diagram:

6; 9; 12; 12; 14; 15; 16; 18; 18; 18; 19; 20; 20;
21; 21; 21; 22; 23; 28; 28; 29; 32; 33; 33; 37

<u>SOLUTION</u>:

Stem	Leaves
0	6; 9
1	2; 2; 4; 5; 6; 8; 8; 8; 9
2	0; 0; 1; 1; 1; 2; 3; 8; 8; 9
3	2; 3; 3; 7
KEY: 1/4 = 14	

b) Populations and Samples

✓ We can carry out a ***survey*** to find out information. We find out the information by asking questions.

✓ The word ***population*** is used in statistics *for the set of data being investigated.*

So, if we want to find out information about all the learners in a school, we could ask every single learner in the school. This group is called the ***population.***

✓ A ***sample*** is a subset of a population.
This means that a sample is much smaller than a population.

A ***subset*** is a set that is part of a larger set.

So, to find out information about all the learners in a school, we could ask selected learners in each grade instead of every learner in the school. These selected learners would be a ***sample*** and, if the sample is selected correctly, the results could be used to reach conclusions about the whole school.

MEASURES OF CENTRAL TENDENCY OF LISTS OF DATA

- ✓ An ***average*** or ***measure of central tendency*** is a single number which is used to represent a collection of numerical data. The commonly used averages are the ***mean, median*** and ***mode***.

- ✓ When we calculate the mean, median and mode we are finding the value of a ***typical item*** in a data set.

a) Mean

- ✓ The ***mean*** is the sum of all the values divided by the total number of values.

- ✓ The mean is the ***equal shares*** average. To find the mean you find the total of all the data items and share the total out equally.

- ✓ The mean is usually written as $\bar{x}$ (often read as "x bar").

  $\bar{x}$ is worked out using the following formula:

$$\bar{x} = \frac{sum\ of\ all\ the\ values}{total\ number\ of\ values}$$

$$= \frac{\sum x}{n}$$

- The symbol Σ (sigma) tells us to add all the values in the data set.
- The symbol *n* is the total number of items in the data set.

EXAMPLE 2

Fourteen of the learners in a Grade 10 class were asked to work out how many kilometres they lived from school. The following list of data shows the distances in km:

4; 7; 1; 9; 4; 8; 11; 10; 19; 2; 5; 7; 19; 3

a) Calculate the mean distance these fourteen learners live from school.
b) What does the mean tell us about the distances travelled?
c) Use your scientific calculator to determine the mean distance travelled.

SOLUTION:

a) Σx = 4 + 7 + 1 + 9 + 4 + 8 + 11 + 10 + 19 + 2 + 5 + 7 + 19 + 3 = 109 km

There are 14 terms in the data set, so n = 14

Mean = $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = \frac{109}{14}$

$\bar{x}$ = 7,7857...

$\bar{x} \approx$ 7,79

The mean distance that the 14 learners live from their school is 7,79 km.

EXAMPLE 2 (continued)

An ***outlier*** is a value far from most others in a set of data

b) The mean tells us that if all the distances are added together and shared out equally, each learner would travel 7,79 km.

Two of these learners live 19 km away from the school. These two ***outliers*** (of 19 km) affect the value of the mean and make it larger than it should be if only the distances that the other twelve learners live from the school were considered.

c) The key sequences for the Casio *fx-82ZA PLUS* and the Sharp EL-W535HT are as follows:

CASIO:
[MODE] [2 : STAT] [1: 1 – VAR]
4 [=] 7 [=] 1 [=] 9 [=] 4 [=] 8 [=] 11 [=] 10 [=] 19 [=] 2[=] 5 [=] 7 [=] 19 [=] 3 [=] [AC]
[SHIFT] [1 : STAT] [4 : VAR] [2 : $\bar{x}$] [=]

SHARP
[MODE] [1 : STAT] [0 : SD] [2ndF] [CA]
4 [DATA] 7 [DATA] 1 [DATA] 9 [DATA] 4 [DATA] 8 [DATA] 11 [DATA]
10 [DATA] 19 [DATA] 2 [DATA] 5 [DATA] 7 [DATA] 19 [DATA] 3 [DATA]
[ALPHA] [=] [$\bar{x}$]

Both calculators give the value $\bar{x}$ = 7,7857... ≈ 7,79 km

EXAMPLE 3
A representative sample of Secondary Schools in South Africa took part in the 2009 Census@School. The mean number of schools per province for 8 of the 9 provinces (Free State is excluded) was 78.

a) What is the total number of schools in the 8 provinces that took part in 2009 Census@School?
b) The number of schools in the Free State (54) is now added to the total in a).
   i) What is the total number of schools in the sample now?
   ii) What is the mean number of schools per province now?
c) What does this mean tell us about the number of schools in the sample?

SOLUTION:

a) Mean = $\frac{\text{total number of schools}}{\text{number of provinces}}$

$78 = \frac{\text{total number of schools}}{8}$

Total number of schools = $8 \times 78 = 624$

b) Number of provinces with the inclusion of the Free State = 8 + 1 = 9
   i) New total number of schools = 624 + 54 = 678
   ii) New mean = $\frac{\text{total number of schools}}{\text{number of provinces}} = \frac{678}{9}$ = 75,333... ≈ 75

c) The mean tells us that if the 678 schools were shared out equally amongst the 9 provinces, each province would get approximately 75 schools.

b) Median

- ✓ The median is the middle value when all values are placed in ascending or descending order.

- ✓ There are as many values above the median as below it.
  - If there is an *odd number* of data items, the median is one of the data items.
  - If there is an *even number* of data items, the median is found by adding the two middle data items and dividing it by two.

EXAMPLE 4
Find the median of the following two sets of data:
a) 4; 7; 1; 9; 4; 9; 11; 10; 19; 2; 5; 8; 19
b) 4; 6; 1; 9; 4; 8; 11; 10; 19; 2; 5; 7; 19; 3

SOLUTION:

a) First arrange the data in ascending order: 1; 2; 4; 4; 5; 7; 8; 9; 9; 10; 11; 19; 19
There are 13 data items, and 13 is an odd number.
The middle item is the 7^{th} one: 1; 2; 4; 4; 5; 7; **8**; 9; 9; 10; 11; 19; 19
The median = 8
Note that there are six data items to the left of 8 and six data items to the right of 8.

b) First arrange the data in ascending order: 1; 2; 3; 4; 4; 5; 6; 7; 8; 9; 10; 11;19; 19
There are 14 data items, and 14 is an even number.
The 7^{th} and 8^{th} terms are the two middle data items:
1; 2; 3; 4; 4; 5; **6; 7**; 8; 9; 10; 11;19; 19
The median is ***halfway*** between 6 and 7, so the ***median*** $= \frac{6+7}{2} = \frac{13}{2} = 6{,}5$
Note that 50% of the data items are less than 6,5 and 50% of the data items are more than 6,5.

c) Mode

- ✓ The mode is the data item that occurs most frequently.
  - If there are two modes, then the data set is said to be ***bimodal***.
  - If there are more than two modes, then the data set is said to be ***multimodal***.
  - All the data items in a set may be different. In this case it has ***no mode***.
- ✓ The associated adjective is ***modal*** so we are sometimes asked to find the ***modal value.***

EXAMPLE 5
Find the mode of the following sets of data:
a) 3; 8; 9; 12; 17; 11; 9; 1; 10; 18
b) 1; 2; 3; 4; 4; 5; 7; 7; 8; 9; 10; 11; 19; 19
c) 1; 7; 8; 10; 51; 18; 2; 19; 11; 45

SOLUTION:

a) First arrange the data in ascending order: 1; 3; 8; 9; 9; 10; 11; 12; 17; 18
Look for the value that occurs most frequently: 1; 3; 8; **9; 9**; 10; 11; 12; 17; 18
Mode = 9

b) The data is already arranged in ascending order:
1; 2; 3; 4; 4; 5; 7; 7; 8; 9; 10; 11; 19; 19
Look for the value that occurs most often:
1; 2; 3; **4; 4**; 5; **7; 7**; 8; 9; 10; 11; **19; 19**
There are three modes, so the data set is multimodal.
Modes = 4; 7 and 19

c) First arrange the data in ascending order: 1; 2; 7; 8; 10; 11; 18; 19; 45; 51
Look for the value that occurs most frequently: 1; 2; 7; 8; 10; 11; 18; 19; 45; 51
None of the values are repeated.
So there is no mode

d) Advantages and disadvantages of the mean, median and mode

	Advantages	Disadvantages
Mean	• Easy to work out with a calculator • Uses all the data • What most people think of as the average	• Can only be used for numbers and measurements • *Not always one of the values* • A few very large or small numbers can affect its size
Median	• Easy to find when the values are in order • Is one of the values if you have an odd number of values	• Can only be used for numbers and measurements • A lot of values can take a long time to put in order • *May not be one of the values if you have an even number of values*
Mode	• Can be found for any kind of data • Simple to find because you count, not calculate • Always one of the items in the data • Quick and easy to find from a frequency table, bar graph or pie chart.	• Not very useful for small amounts of data • May be more than one item • Does not exist if there is an equal number of each item

- ***In practice, much more use is made of the mean than of either of the other two measures of central tendency.***

EXERCISE 1.1

Round all decimal answers to 2 decimal places.

1) For each of the following sets of data find:
   i) The mean
   ii) The median
   iii) The mode
   a) 2; 5; 8; 4; 3; 4; 7; 6; 2; 4; 4
   b) R2,50; R3,00; R6,50; R1,25; R6,50; R2,50; R6,50
   c) 12 cm; 15 cm; 7 cm; 6 cm; 11 cm; 7 cm; 13 cm; 12 cm
   d) 120 kg; 112 kg; 118 kg; 111 kg; 113 kg; 114 kg; 119 kg; 125 kg; 109 kg; 130 kg

2) The mean height of a group of 10 learners is 166,8 cm.
   The tallest person in the group is 169,9 cm.
   Calculate the average height of the remaining 9 members in the group.

3) There are 4 children in a family. The two oldest children are twins. The mean of the 4 children's ages is 14,25 years, the median is 15,5 years and the mode is 16 years. Use this information to work out the ages of the 4 children.

4) Mr Molefe was reading the section in the 2009 Census@School that shows *Favourite Subject by Gender, Grade 8 to 12*. He was astonished at how few students in Secondary Schools in the sample chose Mathematics as their favourite subject.

   In an attempt to find out if the learners in his school shared the same feelings about Mathematics, he asked a representative sample of 100 boys and 100 girls in each grade what their favourite subject was. The following table shows the results of Mr Molefe's survey:

Percentage of the learners in the sample whose favourite subject is Mathematics		
Grades	**Boys**	**Girls**
8	12,2%	10,3%
9	9,8%	11,2%
10	11,1%	7,8%
11	8,3%	6,9%
12		12,5%

   a) What is the mean percentage of girls in the table?
   b) Mr Molefe is still waiting for the results for the Grade 12 boys. He would like the mean for the boys in the whole sample to be 10%. What must the minimum percentage be from the Grade 12 boys in order for Mr Molefe to get the results that he wants?

5) A representative sample of 1 000 learners in each grade from Grades 3 to 12 was surveyed to determine the number of boys in each grade. The mean percentage of the boys in each of the grades, from Grade 3 to Grade 11 (9 grades), is 50%. The mean percentage of the boys in each of the grades, from Grades 3 to 12 (10 grades), is 49,68%. What percentage of the learners in the Grade 12 sample are boys?

MEASURES OF CENTRAL TENDENCY OF DATA IN A FREQUENCY TABLE

✓ We can find the mode, median and mean of data in a ***frequency table.***

a) Mode

✓ The mode is the value in the table that occurs most often.

✓ It is the value with the ***greatest frequency***.

✓ ***Remember:*** The mode is the ***value***, not the frequency.

EXAMPLE 6
Zanele did a survey of 10 of her friends.
She asked them how many siblings they had.
The frequency table shows the results of her survey:

Your ***sibling*** is your brother or sister.

Number of siblings	0	1	2	3
Frequency	2	3	4	1

Find the mode of the number of siblings.

SOLUTION:
The greatest frequency in the table is 4.
This means that four of her friends had 2 siblings.
So the mode = 2 siblings.

b) Median

- ✓ One way to find the median is to list all the values in the frequency table in order of size.

- ✓ Another way is to work directly with the table.

EXAMPLE 7
Look again at the frequency table showing the results of Zanele's survey.

Number of siblings	0	1	2	3
Frequency	2	3	4	1

Find the median of the data in the frequency table.

SOLUTION:

METHOD 1: List all the values in order of size:

0; 0; 1; 1; **1; 2**; 2; 2; 2; 3

The median = $\frac{1+2}{2} = \frac{3}{2}$ ***= 1,5 siblings***

METHOD 2: Find the median directly from the table.

- The values in the table are already in order of size.
- Count along the frequencies to find where the middle value or values lie.
  For Zanele's data, adding the frequencies gives 2 + 3 + 4 + 1 = 10
  There are 10 values in the data. Half of 10 is 5.
  So the median is half-way between the 5^{th} and 6^{th} values.
  Count along the frequencies to find these values.

Number of siblings	0	1	2	3
Frequency	2	3	4	1
	2 values to here	2+3=5 values to here	6^{th} value must be in here	

So the 5^{th} value must be 1 and the 6^{th} value must be 2.

The median = $\frac{1+2}{2} = \frac{3}{2}$ ***= 1,5 siblings***

NOTE:

- *When there is an even number of data items, there is a possibility that the median is **not** one of the data items, and that it is a decimal value. As a result we can often end up with an answer like **1,5 siblings**.*
- *We **don't** need round off an answer like this because our interpretation of the situation is that "50% of Zanele's friends have less than 1,5 siblings (in other words 0 or 1), and 50% of Zanele's friends have more than 1,5 siblings (in other words 2 or 3)".*

c) Mean

- ✓ One way to find the mean is to list all the values in the frequency table, add up these values, and then divide the answer by the number of values

- ✓ Another way is to work directly with the table using the formula $\bar{x} = \frac{\sum f.x}{n}$ where *f* is the frequency, *x* is the data item, and *n* is the number of data items in the set of data.

EXAMPLE 8
Look again at the frequency table showing the results of Zanele's survey.

Number of siblings	0	1	2	3
Frequency	2	3	4	1

Find the mean of the data in the frequency table.

<u>**SOLUTION**</u>**:**

<u>METHOD 1</u>: List all the values in the table: 0; 0; 1; 1; 1; 2; 2; 2; 2; 3

The mean = $\frac{\textbf{\textit{total number of siblings}}}{\textbf{\textit{frequency}}} = \frac{\mathbf{0+0+1+1+1+2+2+2+2+3}}{\mathbf{10}} = \frac{\mathbf{14}}{\mathbf{10}}$ ***= 1,4 siblings***

<u>METHOD 2</u>: Find the mean directly from the table.

- To find the *total number of siblings* you must take the frequency of each value into account. We multiply each *number of siblings* by *its frequency*, and then add the numbers.

- Change the table to a vertical one, and add in another column:

VALUE **Number of siblings** $\boldsymbol{x}$	**FREQUENCY** **Number of Zanele's friends** $\boldsymbol{f}$	**FREQUENCY × VALUE** $\boldsymbol{f \times x}$
0	2	2 × 0 = 0
1	3	3 × 1 = 3
2	4	4 × 2 = 8
3	1	1 × 3 = 3
	n = 10	$\boldsymbol{\sum f.x = 14}$

The mean = $\boldsymbol{\bar{x}} = \frac{\boldsymbol{\sum f.x}}{\boldsymbol{n}} = \frac{14}{10}$ = *1,4 siblings*

NOTE:
- *The mean is* ***not necessarily*** *one of the data items. This means that it is possible to end up with an answer of* ***1,4 siblings****.*
- *Again, we do not round off an answer like this because our interpretation of the situation is that "if the total number of siblings were shared out equally, each friend would get 1,4 siblings."*

EXAMPLE 9

A certain school provides buses to transport the learners to and from a nearby village. A record is kept of the number of learners on each bus for 26 school days.

27	25	27	29	31	24	25	27	28	29	24	26	30
28	31	25	25	27	28	28	28	26	28	31	24	30

a) Organise the data in a frequency table
b) Use the table to calculate the total number of learners that were transported to school by bus.
c) Calculate the mean number of learners per trip, correct to one decimal place.
d) Explain what the mean represents.
e) Find the mode.
f) Find the median and explain what the median represents.

SOLUTION:

a)

Number of learners per trip (x)	**Frequency (f)**	$f \times x$
24	3	$3 \times 24 = 72$
25	4	$4 \times 25 = 100$
26	2	$2 \times 26 = 52$
27	4	$4 \times 27 = 108$
28	6	$6 \times 28 = 168$
29	2	$2 \times 29 = 58$
30	2	$2 \times 30 = 60$
31	3	$3 \times 31 = 93$
	$n = 26$	$\sum f.x = 711$

b) Total number of learners that were transported to school by bus = 711

c) Mean number of learners on each bus $= \frac{\text{total number of learners on all the bus trips}}{\text{total number of bus trips}}$

$$= \frac{\sum f.x}{n}$$
$$= \frac{711}{26}$$
$$= 27{,}3461\ldots$$
$$\approx 27{,}3$$

d) The mean tells us that if each bus had exactly the same number of people each time, there would be approximately 27 learners on each bus.

e) The largest value in the frequency column is 6 and it goes with 28 learners.
This means that the mode = 28 learners.

<u>EXAMPLE 9</u> (continued)

f) There are 26 bus trips.
Half of 26 = 13, so the median lies between the 13th and the 14th values on the table.

Number of learners per trip	24	25	26	27	28	29	30	31
Frequency	3	4	2	4	6	2	2	3

	3 values to here	3 + 4 = 7 values to here	7 + 2 = 9 values to here	9 + 4 = 13 values to here	13 + 6 = 19 values to here

So the 13th value is 27, and the 14th value is 28.
Median = $\frac{27+28}{2} = \frac{55}{2}$ = 27,5 learners.

The median tells us that for half of the bus trips there were less than 27,5 learners (which means 27 and less) on the bus and for half of the bus trips there were more than 27,5 learners (which means 28 or more) on the bus.

d) Using a Scientific Calculator to Find the Mean of Data

- ✓ A scientific calculator makes it quicker and easier to find the mean of data in a frequency table.
- ✓ The key sequences for the CASIO *fx-82ZA* PLUS and the SHARP EL-W535HT that can be used to find the mean in Example 9 are as follows:

CASIO	SHARP
• First add in a frequency column: [SHIFT] [SETUP] [▼] [3:STAT] [1:ON] • Then enter the data [SETUP] [2:STAT] [1:1-VAR] 24 [=] 25 [=] 26 [=] 27 [=] 28 [=] 29 [=] 30 [=] 31 [=] [▼] [▶] 3 [=] 4 [=] 2 [=] 4 [=] 6 [=] 2 [=] 2 [=] 3 [=] [AC] [SHIFT] [STAT] [1] [4:VAR] [2 :$\bar{x}$]	[MODE] [1 : STAT] [0 : SD] [2ndF] [MODE] [CA] 24 [(x ; y)] 3 [DATA] 25 [(x ; y)] 4 [DATA] 26 [(x ; y)] 2 [DATA] 27 [(x ; y)] 4 [DATA] 28 [(x ; y)] 6 [DATA] 29 [(x ; y)] 2 [DATA] 30 [(x ; y)] 2 [DATA] 31 [(x ; y)] 3 [DATA] [ALPHA] [4] [$\bar{x}$]

EXERCISE 1.2

1) In the 2009 Census@School survey, learners were asked to measure the length of their right foot. A group of Grade 10 learners measured the lengths of each other's feet and recorded the lengths obtained correct to the nearest 0,5 cm. The results are summarised in the table below:

Length of foot (in cm)	22,5	23	23,5	24	24,5	25	25,5	26
Number of learners	2	4	6	8	10	5	3	1

   a) Use the table to determine the following:
      i) The value of the mode
      ii) The value of the median
      iii) The mean foot length.

   b) What does each average tell you about the foot lengths for the group of Grade 10 learners?

2) Grade 11B did an Investigation Task to determine the effect of the Consumer Price Index (CPI) on inflation. The marks they obtained for their Investigations are as follows:

Investigation mark (%)	46	68	72	75	78	82	85	90	91
Frequency (number of learners)	1	3	5	6	4	2	2	1	1

   a) Determine the marks that are the mean, the median and the mode.
   b) Explain what each measure of central tendency tells you about the results of the investigation.

3) The table below is adapted from Census 2011 and shows the ***unemployment rate*** of 25 municipalities in the Western Cape. (The unemployment rate is the percentage of the total labour force that is **unemployed** but actively seeking employment and willing to work.) The rate has been rounded off to the nearest whole number.

Unemployment rate (%)	7	11	13	14	15	17	18	19	21	23	24	25	26	30
Frequency (number of municipalities)	1	3	1	4	2	1	2	1	1	3	1	2	1	1

   a) Determine the mean, median and mode of the unemployment rate for the 25 municipalities.
   b) Explain what each measure of central tendency tells you about the results in the table.

MEASURES OF CENTRAL TENDENCY IN A GROUPED FREQUENCY TABLE

a) Discrete Data, Continuous Data and Categorical Data

✓ There are three common types of data: ***discrete data***, ***continuous data*** and ***categorical data***.

- ***Discrete data*** consists of numerical values that are found ***by counting***. They are often whole numbers.
  Examples of discrete data are:
  - Number of children in a family
  - Number of rooms in a house
  - Marks scored in a maths test

- ***Continuous data*** consists of numerical values that are found ***by measuring***. They are given to a certain *degree of accuracy*. They cannot be given *exactly*. They are often decimals.
  Examples of continuous data are:
  - Foot lengths
  - Time taken to travel to school
  - Mass of books in a school bag
  - Amount of water drunk in a day.

- ***Categorical data*** consists of descriptions using names.
  Examples of categorical data are:
  - Heads or Tails
  - Boy or Girl
  - House, shack, flat or informal dwelling.

b) Grouped Frequency Tables

✓ We can group discrete data and continuous data in ***grouped frequency tables***. In a grouped frequency table the values are grouped in ***class intervals***. The ***frequencies*** show the number of values in each class interval.

✓ We can find the mode, median and mean of data in a ***grouped frequency table.***

✓ If you do not have the raw data, you do not know the actual values in each class interval. So you cannot find the actual mode, median and mean of the data. However, you can make ***reasonable estimates*** of these answers.

c) Mode

- ✓ The ***modal interval*** is the interval in the table that occurs most often. It is the group of values with the ***greatest frequency***.
- ✓ Note: ***Mode*** refers to a single value that occurs most often; ***modal interval*** refers to the group of values that occurs most often.
- ✓ ***Remember:*** The mode is the ***value***, not the frequency.

EXAMPLE 10
The choir teacher kept a record of the number of learners who attended the 40 choir practices during the year.

This frequency table gives a summary of the attendance.

Number of learners at choir practice (x)	**Frequency (f)**
$0 < x \leq 10$	1
$10 < x \leq 20$	2
$20 < x \leq 30$	11
$30 < x \leq 40$	9
$40 < x \leq 50$	14
$50 < x \leq 60$	3
	$\boldsymbol{n = 40}$

Find the modal interval.

<u>SOLUTION</u>:
The interval with the greatest frequency is $40 < x \leq 50$.

This means that there were more times when there were from 40 up to and including 50 learners at the choir practices than any other interval.

So, the ***modal interval*** is $40 \leq x \leq 50$.

d) Median

- ✓ You cannot find the median directly from a grouped frequency table. You can find the class interval that contains the median, and then find the approximate value of the median by finding the midpoint of the interval.

EXAMPLE 11
Look again at the choir teacher's summary of attendance. This time find the median number of learners at choir practice.

Number of learners at choir practice (x)	**Frequency (f)**
$0 < x \leq 10$	1
$10 < x \leq 20$	2
$20 < x \leq 30$	11
$30 < x \leq 40$	9
$40 < x \leq 50$	14
$50 < x \leq 60$	3
	$\boldsymbol{n = 40}$

<u>SOLUTION</u>:
There were 40 choir practices.
Half of 40 is 20, so the median is half-way between the 20th and 21st values.

Count down the frequencies to find the position of the median

Number of learners at choir practice (x)	**Frequency (f)**	
$0 < x \leq 10$	1	... 1 value to here
$10 < x \leq 20$	2	... 1 + 2 = 3 values to here
$20 < x \leq 30$	11	... 3 + 11 = 14 values to here
$30 < x \leq 40$	9	... 14 + 9 = 23 values to here
$40 < x \leq 50$	14	
$50 < x \leq 60$	3	
TOTAL	40	

The 15th to 23rd values are in the class interval $30 < x \leq 40$
So the 20th and 21st values are in this class.
So the ***median class is $30 < x \leq 40$***

To find ***an estimate of the median***, find the midpoint of the interval.
The midpoint of the interval $30 < x \leq 40$ is $\frac{30+40}{2} = \frac{70}{2} = 35$
So the ***median ≈ 35 learners***

The median tells us that 50% of the choir practices were attended by less than 35 learners and 50% of the choir practices were attended by more than 35 learners.

e) Mean

✓ To find *an estimate of the mean:*

i) Find the midpoint of each interval to represent each class (usually written as *X*). Then you assume that each item in the interval has that value.

ii) Multiply the midpoint by the frequency in order to work out an estimate of the total of the values in the class.

iii) Add these values together to get an estimate of the total of all the values.

iv) Substitute the values directly into the formula $\bar{X} = \frac{\sum f.X}{n}$ where $\bar{X}$ is the approximate value of the mean, *X* is the value of the midpoint of each interval, and *f* is the frequency of that interval.

EXAMPLE 12
Look again at the choir teacher's summary of attendance.

Number of learners at choir practice (x)	Frequency (f)
$0 < x \leq 10$	1
$10 < x \leq 20$	2
$20 < x \leq 30$	11
$30 < x \leq 40$	9
$40 < x \leq 50$	14
$50 < x \leq 60$	3
	$n = 40$

Find the approximate value of the mean number of learners who attended choir practice.

SOLUTION:
First add in another column and work out the ***midpoint of each interval***. Then, add another column and calculate ***frequency × mid-point value***.

Number of learners at choir practice (x)	Frequency f	Midpoint of the interval X	$f \times X$
$0 < x \leq 10$	1	$\frac{0+10}{2} = 5$	$1 \times 5 = 5$
$10 < x \leq 20$	2	$\frac{10+20}{2} = 15$	$2 \times 15 = 30$
$20 < x \leq 30$	11	$\frac{20+30}{2} = 25$	$11 \times 25 = 275$
$30 < x \leq 40$	9	$\frac{30+40}{2} = 35$	$9 \times 35 = 315$
$40 < x \leq 50$	14	$\frac{40+50}{2} = 45$	$14 \times 45 = 630$
$50 < x \leq 60$	3	$\frac{50+60}{2} = 55$	$3 \times 55 = 165$
	$n = 40$		$\sum f.X = 1\,420$

<u>EXAMPLE 12</u> (continued)

Mean = $\bar{X} = \frac{\sum f.X}{n} \approx \frac{1\ 420 \text{ learners}}{40}$ = 35,5 learners

The mean tells us that if the total number of learners attending the 40 choir practices were shared out equally, then approximately 35,5 learners would have attended each choir practice.

> ***NOTE:***
> - *Once again, we don't round this amount off.*
> - *We use the value of the mean to **interpret** the situation, so don't have to end up with a value that is a whole number.*

f) Using a Scientific Calculator to Find the Mean of Data

- ✓ A scientific calculator makes it quicker and easier to find the mean of grouped data.
- ✓ The key sequences for finding the mean in Example 12 using the CASIO *fx-82ZA* PLUS and the SHARP EL-W535HT are as follows:

CASIO	SHARP
[SETUP] [2:STAT] [1:1-VAR] ***First enter the midpoints*** 5 [=] 15 [=] 25 [=] 35 [=] 45 [=] 55 [=] [▼] [▶] ***Then enter the frequencies*** 1 [=] 2 [=] 11 [=] 9 [=] 14 [=] 3 [=] [AC] [SHIFT] [STAT] [1] [4:VAR] [2 : $\bar{x}$]	[MODE] [1 : STAT] [0 : SD] [2ndF] [MODE] [CA] ***Enter the midpoints and frequencies together*** 5 [(x ; y)] 1 [DATA] 15 [(x ; y)] 2 [DATA] 25 [(x ; y)] 11 [DATA] 35 [(x ; y)] 9 [DATA] 45 [(x ; y)] 14 [DATA] 55 [(x ; y)] 3 [DATA] [ALPHA] [4] [$\bar{x}$]

EXAMPLE 13

In a particular primary school in Pietermaritzburg, it was found that ninety of their Foundation Phase learners (Grades 1, 2 and 3) were accompanied to school by someone. The ages of the person accompanying the child were recorded, as shown in the table below.

Age (in years) (x)	**Frequency (f)**
$0 < x \leq 10$	12
$10 < x \leq 20$	30
$20 < x \leq 30$	18
$30 < x \leq 40$	12
$40 < x \leq 50$	9
$50 < x \leq 60$	6
$60 < x \leq 70$	3

Use the information given in the table to

a) Determine the modal interval.
b) Estimate the mean age of the person accompanying a learner from the Foundation Phase.
c) Estimate the median age of the person accompanying a learner from the Foundation Phase.

<u>SOLUTION</u>:

a) The modal interval is $10 < x \leq 20$.
This means that more people in this age group accompanied the learners to school than any other age group.

b) To find the mean we have to take the midpoint of each class interval and then calculate *frequency* $\times$ *midpoint* for each class interval.

Age (in years) x	**Midpoint X**	**Frequency f**	***f.X***
$0 < x \leq 10$	$\frac{0+10}{2} = 5$	12	$12 \times 5 = 60$
$10 < x \leq 20$	$\frac{10+20}{2} = 15$	30	$30 \times 15 = 450$
$20 < x \leq 30$	$\frac{20+30}{2} = 25$	18	$18 \times 25 = 450$
$30 < x \leq 40$	$\frac{30+40}{2} = 35$	12	$12 \times 35 = 420$
$40 < x \leq 50$	$\frac{40+50}{2} = 45$	9	$9 \times 45 = 405$
$50 < x \leq 60$	$\frac{50+60}{2} = 55$	6	$6 \times 55 = 330$
$60 < x \leq 70$	$\frac{60+70}{2} = 65$	3	$3 \times 65 = 195$
		$n = 90$	$\sum f.X = 2\,310$

Mean $= \bar{X} = \frac{\sum f.X}{n} \approx \frac{2\,310 \text{ years}}{90} = 25{,}7$ years old

The mean tells us that if all the ages were added together, and then shared out equally amongst the 90 people, then each one would be 25,7 years old.

EXAMPLE 13 (continued)

c) 90 people accompanied the learners to school.
$90 \div 2 = 45$. So the median lies half-way between the 45th and 46th person.

Age (years)	Frequency	
$0 < x \leq 10$	12	… 12 people to here
$10 < x \leq 20$	30	… 12 + 30 = 42 people to here
$20 < x \leq 30$	18	… 42 + 18 = 60 people to here.
$30 < x \leq 40$	12	
$40 < x \leq 50$	9	
$50 < x \leq 60$	6	
$60 < x \leq 70$	3	

Both the 45th and 46th people lie in the interval $20 < x \leq 30$.
So the median interval is 20 years $< x \leq$ 30 years, and the approximate value of the median is 25 years.

The value of the median tells us that 50% of the people accompanying the Grade 1, 2 and 3 learners to school are less than or equal to 25 years old, and 50% are more than or equal to 25 years old.

EXERCISE 1.3

1) The table below represents the ages of the 90 people accompanying Foundation Phase learners to another primary school in Pietermaritzburg:

Age (in years) (x)	**Frequency (*f*)**
$0 \leq x < 10$	4
$10 \leq x < 20$	12
$20 \leq x < 30$	25
$30 \leq x < 40$	14
$40 \leq x < 50$	10
$50 \leq x < 60$	20
$60 \leq x < 70$	5
	$n = 90$

   a) Use the information given in the table to
      i) Determine the modal interval.
      ii) Estimate the mean age of the people accompanying a learner from the Foundation Phase.
      iii) Estimate the median age of the people accompanying a learner from the Foundation Phase.
   b) What do the modal class, the mean and the median tell you about the ages of the people who accompany the children to school?

2) The table below has been adapted from Census 2011. It lists the property values of 262 properties in a part of the Ikwezi municipality.

Property Value in Rand (x)	**Frequency (number of households) (*f*)**
$0 \leq x < 50\,000$	172
$50\,000 \leq x < 100\,000$	51
$100\,000 \leq x < 150\,000$	18
$150\,000 \leq x < 200\,000$	12
$200\,000 \leq x < 250\,000$	9
	$n = 262$

   a) Use the information given in the table to
      i) Determine the modal interval.
      ii) Estimate the mean property value of the properties.
      iii) Estimate the median property value of the properties.
   b) What does the modal class, the mean and the median tell you about these property values?

QUARTILES AND THE FIVE NUMBER SUMMARY

a) Quartiles

- ✓ ***Quartiles*** are the three values Q_1, Q_2 and Q_3 that divide a data set into ***four*** approximately equal parts. Each part consists of approximately 25% of the elements of the data set.

- ✓ $\boldsymbol{Q_1}$ is the ***lower quartile***; $\boldsymbol{Q_2}$ is the middle quartile or ***median*** and $\boldsymbol{Q_3}$ is the ***upper quartile***.
  - ✓ The ***median*** divides an ordered data set into *two halves*.
  - ✓ The ***quartiles*** divide an ordered data set into *four quarters*.
  - ✓ The ***median*** is also the 2nd quartile ($\boldsymbol{M}$ or $\boldsymbol{Q_2}$).

25% of the data ↑ *25% of the data* ↑ *25% of the data* ↑ *25% of the data*

Lower quartile Q_1	**Median M or Q_2**	**Upper quartile Q_3**

- ✓ From the above diagram, one can see that:
  - Approximately one quarter or 25% of the data is less than Q_1.
  - Approximately three quarters or 75% of the data is more than Q_1.
  - Approximately one half or 50% of the data is less than Q_2 and one half or 50% is more than Q_2.
  - Approximately three quarters or 75% of the data is less than Q_3.
  - About one quarter or 25% of the data is more than Q_3.
  - Approximately one half or 50% of the data lies between Q_1 and Q_3.

HOW TO FIND THE QUARTILES:

i) Put the data items in order and find the median.

ii) Find the midpoint of the data items to the left of the median. This is the ***lower quartile (Q_1).***

iii) Find the midpoint of the data items to the right of the median. This is the ***upper quartile (Q_3)***.

EXAMPLE 14

For each of the following sets of data

a) 23 65 33 101 23 21 102 18 26 9

b) 65 33 101 23 21 102 18 26 9

i) Find the median (M)

ii) Find the lower quartile (Q_1) and the upper quartile (Q_3)

SOLUTION:

a) First arrange the data in ascending or decending order.

9 18 21 23 23 26 33 65 101 102

i) There are ten data items (an even number of data items). To find the median, we have to find the mean of the middle two numbers.

$10 \div 2 = 5$, so the median lies between the 5th and the 6th terms.

9 18 21 23 **23 26** 33 65 101 102

So the median = M = $\frac{23+26}{2} = \frac{49}{2} = 24{,}5$

ii) To find the lower quartile (Q_1), take the data before the median and find the median of that.

There is an odd number of data items below the median. Take the middle one.

9 18 **21** 23 23 | 26 33 65 101 102

So, $Q_1 = 21$.

To find the upper quartile (Q_3), take the data after the median and find the median of that:

There is an odd number of data items above the median; take the middle one.

9 18 **21** 23 23 | 26 33 **65** 101 102

So, $Q_3 = 65$.

So, the three quartiles are 21; 24,5 and 65.

b) First arrange the data in ascending or decending order.

9 18 21 23 26 33 65 101 102

i) There are nine data items (an odd number of data items).

$9 \div 2 = 4{,}5$ so the median is the 5th term.

9 18 21 23 **26** 33 65 101 102

So the median = 26

ii) To find the lower quartile (Q_1), take the data before the median (26) and find the median of that.

There is an even number of data items below the median.

$4 \div 2 = 2$, so the lower quartile, Q_1, lies between the 2nd and 3rd terms.

9 18 | 21 23 **26** 33 65 101 102

So the $Q_1 = \frac{18+21}{2} = \frac{39}{2} = 19{,}5$

<u>EXAMPLE 14</u> (continued)

To find the upper quartile (Q_3), take the data after the median and find the median of that.
There is an even number of data items above the median.
$4 \div 2 = 2$, and $5 + 2 = 7$, so Q_3 lies between the 7^{th} and the 8^{th} terms.

9 18 | 21 23 **<u>26</u>** 33 65 | 101 102

So the $Q_3 = \frac{65+101}{2} = \frac{166}{2} = 83$

So, the three quartiles are 19,5; 26 and 83.

b) The Five Number Summary

- ✓ The ***five number summary*** consists of 5 items
  1) The minimum value in the data set;
  2) $Q_{1,}$ the lower quartile;
  3) M, the median;
  4) $Q_{3,}$ the upper quartile;
  5) The maximum value in the data set.

- ✓ Use the following method to find a five number summary:

STEPS	EXAMPLE
1) Put your numbers in the data set in order.	Find the five-number summary of the following data set: 1, 3, 5, 6, 12, 15, 23, 28, 31
2) Find the minimum and maximum values.	Minimum value = 1 Maximum value = 31
3) Find the median	There are 9 data items. $9 \div 2 = 4,5$ so the median is the 5^{th} data item 1 3 5 6 **<u>12</u>** 15 23 28 31 So the median = 12
4) Find Q_1 and Q_3.	There are 4 items below the median. $4 \div 2 = 2$, so Q_1 lies between the 2^{nd} and 3^{rd} terms. 1 3 | 5 6 **<u>12</u>** 15 23 28 31 So $Q_1 = \frac{3+5}{2} = 4$ There are 4 items above the median. $5 + 2 = 7$, so Q_3 lies between the 7^{th} and 8^{th} terms. 1 3 | 5 6 **<u>12</u>** 15 23 | 28 31 So $Q_2 = \frac{23+28}{2} = 25,5$
5) Write down the five number summary.	**Minimum = 1** **Q_1 = 4** **Median = 12** **Q_3 = 25,5** **Maximum = 31**

EXERCISE 1.4

1) Find the five number summary for each of the following data sets:
   a) 1 6 6 9 15 17 23 24 33 33 38 38 38 45 46 51
   b) 9 14 19 21 24 29 29 32 33 35 36 40 46 49
   c) 45 15 43 19 26 25 15 36 27 32 41 25 48
   d) 4 46 6 44 10 17 34 35 31 22 10 16

2) Johan is asked to find the five number summary for the following set of numbers:
   2 23 24 12 11 23 34 12 34 12 33 19 48 25 37 38 59

   Johan's answer is as follows:

   2 11 12 12 12 19 23 23 24 25 33 34 34 37 38 48 59

   Min = 2
   Q_1 = 12
   Median = 24
   Q_3 = 34 or 37
   Max = 59

   a) Give the correct solution to the question.
   b) Identify the mistake that Johan has made. Describe the misconception and explain to John why he is incorrect.

PERCENTILES

- ✓ When you write a test and get a mark of 75%, it tells you how many questions you got right. ***But***, it doesn't tell you how well you did compared to the other people who wrote the test. ***Percentiles*** are values from 0 to 99 that tell you the ***percentage of the marks that are less than a particular mark.***

  If the ***percentile of your test mark is 75***, it tells that
  - ***75% of the marks are LESS than yours***.
  - ***100% – 75% = 25% of the marks are MORE than yours***

- ✓ Percentiles can be used to compare values in any set of ***ordered*** data. You can calculate percentiles for income, mass, etc. Percentiles are often used in education and health-related fields to indicate how one person compares with others in a group.

> ***NOTE:***
> *There is a difference between a mark of* ***60%*** *(****a percentage*** *telling you that you got 60 out of 100) and a mark at the* ***60th percentile*** *(which tells us that approximately 60% of the marks are less than yours).*

- ✓ The following are some special percentiles:
  - The ***median*** is at the 50th percentile
  - The ***lower quartile*** is at the 25th percentile
  - The ***upper quartile*** is at the 75th percentile.

- ✓ Scores that are in the ***95th percentile*** and above are unusually high while those in the ***5th percentile*** and below are unusually low.

  Example:

  a) The ***89th percentile*** is a number that 89% of the data items are below. We then also know that 100% – 89% = 11% of the data items are above that number.

  b) If the ***20th percentile*** is 12, then 20% of the data items are less than 12, and 80% of the data items are more than 12.

  c) If the ***90th percentile*** is 17, then 90% of the data items are less than 17, and 10% of the data items are more than 17.

- ✓ We usually find percentiles of a ***large*** number of items.

EXAMPLE 15

For the following set of 19 data items

72 71 65 60 62 58 67 57 70 73

50 61 51 55 64 68 69 59 63

a) At what percentile is 70?

b) Find the data item that is at the 20th percentile.

c) What is the median score (the score at the 50th percentile)?

<u>SOLUTION</u>:

A stem and leaf diagram can be used to get the data into ascending order:

Stem	**Leaf**
5	0 1 5 7 8 9
6	0 1 2 3 4 5 7 8 9
7	0 1 2 3
KEY: 6/2 = 62	

a) ***To calculate at which percentile a data item lies:***

Percentile $= \frac{number\ of\ data\ items\ that\ are\ less\ than\ or\ equal\ to\ 70}{total\ number\ of\ data\ items} \times 100\%$

$= \frac{16}{19} \times 100\%$

$= 84{,}210\ldots\%$

$\approx 84\%$

So 70 is at the 84th percentile.

a) ***To calculate the data item which is at a given percentile:***

i) *Find out which term corresponds to the 20th percentile:*

Data item corresponding to the 20th percentile

$= \frac{percentile}{100} \times number\ of\ data\ items$

$= \frac{20}{100} \times 19$ data items

$= 3{,}8$

≈ 4

ii) *Count along until you get to the 4th data item* – It is 57

So 57 is the 20th percentile

b) ***To calculate the data item which is at a given percentile:***

The data item corresponding to the 50th percentile

$= \frac{percentile}{100} \times number\ of\ data\ items$

$= \frac{50}{100} \times 19$

$= 9{,}5$

≈ 10

Count along to find the 10th data item – It is 63.

So 63 is the 50th percentile or the median.

EXERCISE 1.5

1) Given the following set of data:

24	34	35	37	39	40	41	45	46	48
50	52	54	55	56	56	59	59	60	61
63	66	67	68	69	70	73	75	77	77
78	79	79	80	84	84	86	86	86	89
94	95	96	97	98	98	100	101	102	103

a) At which percentile is:
   i) 102?
   ii) 34?
   iii) 96?
   iv) 70?

b) Find the item that corresponds to
   i) The 30^{th} percentile
   ii) The 50^{th} percentile
   iii) The 10^{th} percentile
   iv) The 65^{th} percentile

2) For the following stem and leaf diagram

STEM	LEAF
10	0 0 2 3 3 3 4 5 7 7 7 7 8 8 9
11	1 1 2 3 5 5 5 6 6 6 8 9
12	0 0 0 1 3 4 6 7 7 9 9 9 9 9
13	1 2 2 2 5 6 7 7 7 8 8 8 9
14	0 1 1 1 1 2 3 5 5 5 5 6 8 8 8 9
KEY: 10/4 = 104	

a) What is the score that corresponds to the 11^{th} percentile?
b) Find the data item that corresponds to the 44^{th} percentile
c) At which percentile is a score of 135?

MEASURES OF DISPERSION

- ✓ A ***measure of central tendency*** such as the mean, median and mode gives you a single measurement to stand for a set of data. A ***measure of dispersion*** or ***measure of spread*** tells you *how spread out the data is.*

- ✓ The data can either:
  - Be grouped closely together around the measure of central tendency, or
  - Be spread widely apart around the measure of central tendency.

a) Range

- ✓ The range is the simplest measure of spread. It is the difference between the largest and smallest items of data.

Range = Largest Value – Smallest Value

- ✓ There are some limitations to using range:
  - It does not take into account anything about the distribution of any other piece of data except the smallest and largest value.
  - When data is given in a grouped frequency table, the range cannot be used.

b) Interquartile Range

- ✓ The interquartile range (or ***IQR***) is the difference between the upper quartile and the lower quartile.

Interquartile Range = $Q_3 - Q_1$

- ✓ The interquartile range is a better measure of dispersion than the range. It is not affected by any extreme values (very small or very large values). It is based on the ***middle half*** of the data. It is the *range between the upper and lower quartiles.*

- ✓ The ***semi-interquartile range*** is sometimes used. It is half of the interquartile range.

Semi-interquartile range = $\frac{Q_3 - Q_1}{2}$

EXAMPLE 16

a) For the following set of data: 22 17 28 19 23 18 25 29 19 29
Find
   i) the range
   ii) the interquartile range
   iii) the semi-interquartile range

b) Approximately what percentage of the data items lie within the interquartile range?

<u>SOLUTION</u>:

a)

i) Arrange the data in order: 17 18 19 19 22 23 25 28 29 29

Range = Largest Value – Smallest Value= 29 – 17 = 15

ii) First find the median:
There are 10 terms.
$10 \div 2 = 5$ which means that the median lies between the 5th and 6th terms
Median = $\frac{22+23}{2} = \frac{45}{2} = 22{,}5$

The find the two quartiles

17 18 **<u>19</u>** 19 22 | 23 25 **<u>28</u>** 29 29
There are five terms to the left of the median and five terms to the right of the median.
So $Q_1 = 19$ and $Q_3 = 29$

Interquartile range = $Q_3 - Q_1 = 28 - 19 = 9$

iii) ***Semi-interquartile range = $\frac{Q_3 - Q_1}{2} = \frac{28-19}{2} = \frac{9}{2} = 4{,}5$***

b) Approximately 50% of the data items lie within the interquartile range.

EXERCISE 1.6

1) Find the median, lower quartile, upper quartile and the interquartile range for each of the following sets of data:
   a) 9 11 12 12 13 13 14 16 21 22 24
   b) 4 5 5 6 7 9 10 11 12 12 13 14 14 14
   c) 12 7 1 3 2 12 2 9 14 5 6 5 4 8 11 14

2) A group of 21 learners attending extra mathematics classes were required to write a test which was out of 50.
   Their results were:

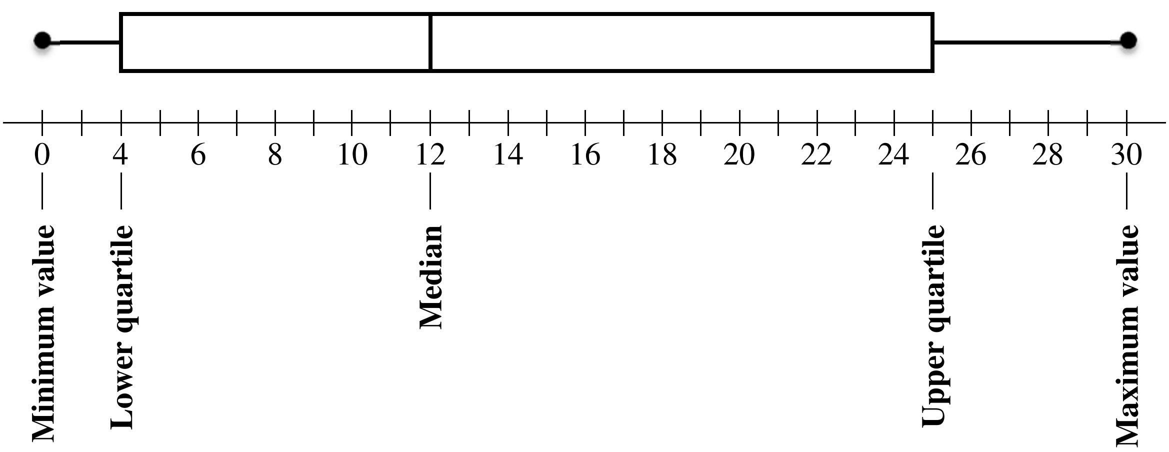
17	8	19	9	12	28	11	16	20	14	29
23	37	23	26	4	35	26	18	45	7	

   a) Find the range.
   b) Find the lower and upper quartiles.
   c) Calculate the interquartile range.
   d) What do the lower and upper quartiles indicate about the results of the test?

BOX AND WHISKER DIAGRAMS

✓ A graphical representation of the five number is known as a ***box and whisker diagram***, also sometimes called a ***box plot.***

- Vertical lines mark the two quartiles and the median. These are joined to make a ***box*** containing the middle half of the data. The box illustrates the ***interquartile range***.
- From the quartiles, horizontal lines are drawn to the minimum and maximum values. These lines are the ***whiskers***.



HOW TO DRAW A BOX AND WHISKER DIAGRAM	
STEP 1:	Make sure that the data is arranged in ascending order
STEP 2:	Find the five number summary
STEP 3:	Draw a number line long enough to fit the minimum and maximum values. Make sure that the units are plotted correctly on the number line.
STEP 3:	Draw vertical lines at Q_1, M and Q_3 and then draw two horizontal lines to make the box.
STEP 4:	From the middle of the box, first draw a horizontal line to the minimum value and then draw a horizontal line to the maximum value.

EXAMPLE 17

Draw a box-and-whisker diagram of the following set of data:

506 503 507 504 510 511 526 513
517 508 515 513 508 509 516

SOLUTION:

STEP 1: Arrange the data in ascending order

503 504 506 507 508 508 509 510 511 513 513 515 516 517 526

STEP 2: Find the five number summary.

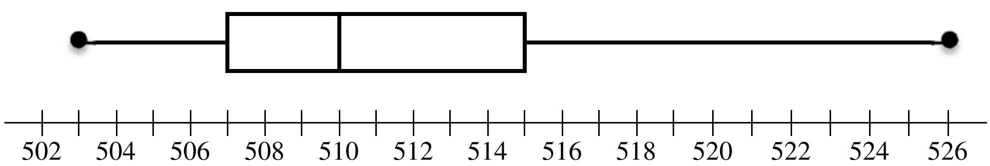
503 504 506 **507** 508 508 509 **510** 511 513 513 **515** 516 517 526

Minimum value = 503
Q_1 = 507
M = 510
Q_3 = 515
Maximum value = 526

STEP 3: Draw a number line long enough to go from 503 to 526.

STEP 4: Draw vertical lines at Q_1, M and Q_3 and form the box

STEP 5: Join the box to the minimum and maximum values to form the whiskers.



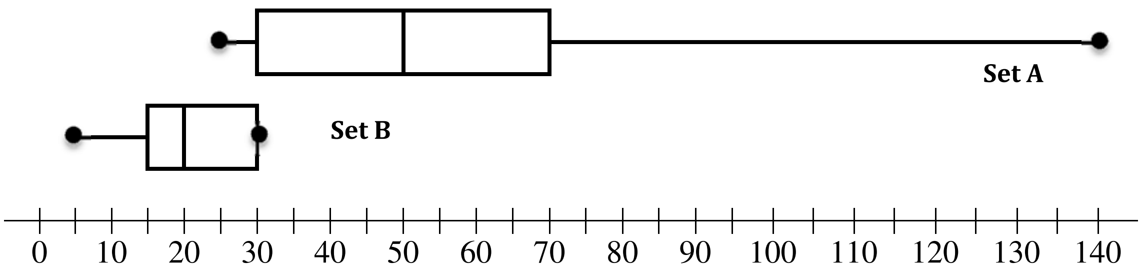
EXAMPLE 18

Two box and whisker diagrams are drawn on the same number line

a) List the five number summary of
   i) Set A
   ii) Set B

b) Find the range and interquartile range of
   i) Set A
   ii) Set B

c) Explain why the dot is on the edge of the box in Set B



<u>SOLUTION</u>:

a)

i) <u>SET A:</u>
Minimum value = 25
$Q_1 = 30$
M = 50
$Q_3 = 70$
Maximum value = 140

ii) <u>SET B</u>
Minimum value = 5
$Q_1 = 15$
M = 20
$Q_3 = 30$
Maximum value = 30

b)

iii) Range of Set A = maximum value – minimum value = 140 – 25 = 115

Interquartile range of Set A = $Q_3 - Q_1$ = 70 – 30 = 40

iv) Range of Set B = maximum value – minimum value = 30 – 5 = 25

Interquartile range of Set B = $Q_3 - Q_1$ = 30 – 15 = 15

c) The dot is on the edge of the box in Set B because the upper quartile and the maximum value in Set B are identical (they are both 30).

EXERCISE 1.7

1) The percentages achieved by a learner for a series of mathematics tests that he wrote throughout his Grade 9 year are as follows:

35	45	50	28	39	49	55	35	56	49
43	37	28	53	55	38	47	51	30	58

   a) Calculate the five number summary
   b) Draw a box-and-whisker diagram to illustrate the five number summary.
   c) Find
      i) the range of the set of data
      ii) the interquartile range.

2) As part of the 2009 Census@School, the 26 Grade 10A learners measured their heights.

   The girls' heights (in centimetres) were:

150	150	153	155	156	158	160	161	164	164	166	170	170

   The boys' heights in centimetres were:

140	142	151	157	158	159	160	162	165	180	180	180	180

   a) Find the five number summary and the interquartile range for the girls and for the boys
   b) On the same number line draw two box and whisker diagrams to illustrate the girls' heights and the boys' heights.
   c) Use the five number summaries, the interquartile ranges and the box and whisker diagrams to write down two conclusions you can make about the heights of the girls and the boys.

REFERENCES

Pike M. et al. (2011). *Classroom Mathematics Grade 10*. Heinemann.
Graham D. and Graham C. (1996). *Mainstream Mathematics for GCSE*. Macmillan
Speed B. (1997). *Higher Mathematics for GCSE*. Collins Educational.
Statistics South Africa (2010) *Census At School Results (2009).*

2

Grade 11 Data Handling

In this chapter you:
- Draw histograms
- Draw frequency polygons
- Draw ogives (cumulative frequency curves)
- Calculate variance and standard deviation of ungrouped data;
- Determine whether data is symmetric or skewed
- Identify the values of the outliers.

WHAT YOU LEARNED ABOUT DATA HANDLING IN GRADE 10

In Grade 10 you covered the following data handling concepts:
- Measures of central tendency of lists of data, of data in frequency tables and of data in grouped frequency tables.
- The range, percentiles, quartiles, interquartile and semi-interquartile range
- The five number summary and box-and-whisker diagram
- Using statistical summaries (measures of central tendency and dispersion) to analyse and make meaningful comments on the context associated with the given data.

STATISTICAL GRAPHS

✓ Organised data can often be presented in graphical form.

- Statistical graphs are used to describe data or to analyse it.
- The purpose of graphs in statistics is to communicate the data to the viewers in pictorial form. It is easier for most people to understand data when it is presented as a graph than when it is presented numerically in tables.

✓ In earlier grades you dealt with the following graphs
- Bar graphs and double bar graphs
- Histograms
- Pie charts
- Broken-line graphs.

- ✓ In Grade 11 you study three statistical graphs often used in research: the ***histogram***, the ***frequency polygon*** and the ***cumulative frequency graph or ogive***.

HISTOGRAMS

- ✓ A ***histogram*** gives us a visual interpretation of data. It looks very similar to a ***bar graph***, but there are definite differences between them.

HISTOGRAM	BAR GRAPH
• It is a representation of grouped data	• It is a representation of ungrouped data that does not have to be numerical
• There is no gap between the bars	• There is generally a gap between the bars
For example*, you draw a HISTOGRAM to show the number of people whose heights (h) lie in the following intervals (measured in cm): $150 \leq h < 160$; $160 \leq h < 170$; etc*	***For example*** *you draw a BAR GRAPH to show the number of learners in a class who wear glasses and the number who do not wear glasses.*

EXAMPLE 1

The following table lists the marks (given as percentage) obtained by the Grade 11 learners of Musi High School in their mathematics test:

24	70	50	22	63	45	48	52	56	38
65	68	65	17	32	60	62	53	63	45
49	44	56	12	55	83	54	22	67	54
34	77	46	50	58	80	81	39	84	75
55	76	73	80	66	71	62	40	23	76

a) Organise the data using a grouped frequency table.
b) Draw a histogram to illustrate the data.
c) Calculate the modal interval. What does this measure of central tendency tell you about the learners' marks?
d) Estimate the median. What does this measure of central tendency tell you about the learners' marks?

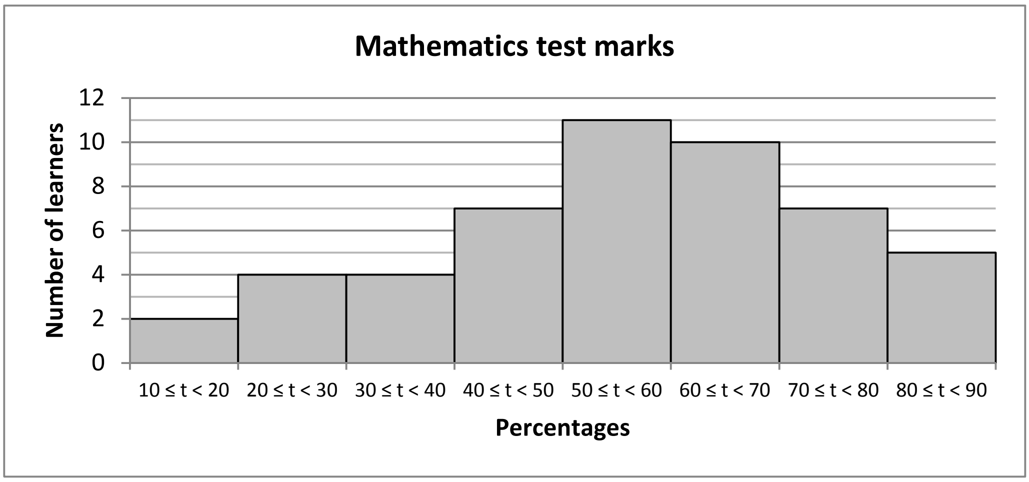
<u>**SOLUTION:**</u>

a) The lowest mark was 12% and the highest mark was 84%
It is often easiest to use multiples of 10 as the class intervals, so start the first interval at 10% and end the last interval at 90%

<u>EXAMPLE 1</u> (continued)

Percentages (t)	**Frequency (Number of learners)**
$10 \leq t < 20$	2
$20 \leq t < 30$	4
$30 \leq t < 40$	4
$40 \leq t < 50$	7
$50 \leq t < 60$	11
$60 \leq t < 70$	10
$70 \leq t < 80$	7
$80 \leq t < 90$	5
TOTAL	50

b) Draw the histogram as follows:
STEP 1: Draw and label the horizontal and vertical axes.
STEP 2: Represent the frequency on the vertical axis and the classes on the horizontal axis.
STEP 3: Using the frequencies (or number of learners) as the heights, draw vertical bars for each class.



c) The modal interval is the interval with the largest frequency or largest number of learners. So the modal interval is $50 \leq t < 60$.
This tells us that more learners got marks in the interval $50 \leq t < 60$ than in any of the other intervals.

d) There are 50 data items (marks/percentages).
The median lies between the 25^{th} and the 26^{th} marks.
Add up the frequencies until you reach 25 (or more than 25):

$$2 + 4 + 4 + 7 + 11 = 28$$

The 28^{th} mark lies in the interval $50 \leq t < 60$
So the median lies in the interval $50 \leq t < 60$
The median ≈ 55% (the midpoint of the interval)
This tells us that 50% of the learners got marks that were less ***than 55%*** and 50% of the learners got marks that were ***more than 55%***

NOTE:

A histogram should have the following:

- A title which describes the information that is contained in the histogram.
- A horizontal axis with a label which shows the scale of values into which the data fit (grouped data intervals)
- A vertical axis with a label which shows the number of times the data within the interval occurred (frequency)
- Adjacent bars (i.e. there are no gaps between the bars).

EXERCISE 2.1

1) The frequency table below represent the distribution of the amount of time (in hours) that 80 high school learners spent in one week watching their favourite sport.

Time in hours	**Frequency**
$10 < t \leq 15$	8
$15 < t \leq 20$	28
$20 < t \leq 25$	27
$25 < t \leq 30$	12
$30 < t \leq 35$	4
$35 < t \leq 40$	1

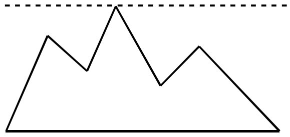
   a) Draw a histogram to represent the data
   b) Calculate
      i) the modal interval
      ii) an estimate of the median
   c) What do these two measures of central tendency tell you about the amount of time the learners devote to watching their favourite sport?

2) In the 2009 Census@School, learners were asked which their favourite subjects at school were. Fifty Grade 11 learners from a certain school in Limpopo chose Science as their favourite subject. The following are their Science marks (as percentages):

31	62	51	44	61	63	59	47	59	67
50	54	61	41	48	74	53	53	53	36
60	42	50	48	42	27	43	42	43	54
49	47	51	28	54	48	83	65	54	35
61	56	57	32	38	32	40	63	56	59

   a) Organise the data in a grouped frequency table.
   b) Draw a histogram to represents the data.
   c) Calculate the modal interval and an estimate of the median and say what these two measures of central tendency tell you about the learners' mark.

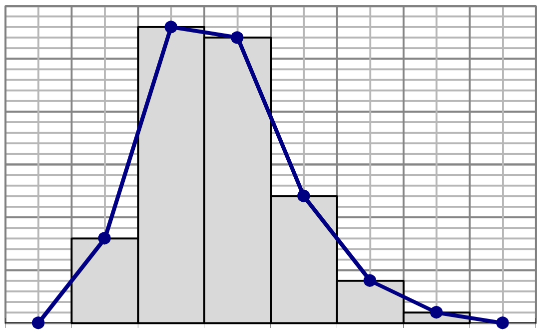
FREQUENCY POLYGONS


A **polygon** is a closed geometric shape made up of line segments.

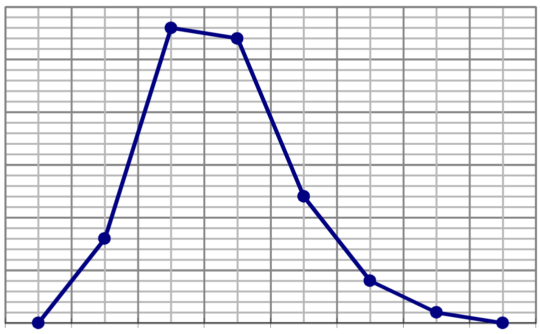
- ✓ A ***frequency polygon*** can be used instead of a histogram for illustrating *grouped data.*

  NOTE: It is called a frequency ***polygon*** because of its shape.

- ✓ One way of drawing a frequency polygon is to
  - a) Draw a histogram
  - b) Join the *midpoints* of the top of the columns of the histogram
  - c) Extend the line to the midpoint of the *class interval below the lowest value* and to the midpoint of the *class interval above the highest value* so that the line touches the horizontal axis on both sides.



- ✓ Another way of drawing a frequency polygon is to
  - a) Calculate the midpoint of each interval and then to plot the ordered pair (midpoint of the interval; frequency)
  - b) Plot the midpoint of the interval below the lowest interval and the interval above the highest interval and plot the points (midpoint of the interval; 0)
  - c) Join these points with straight lines.



EXAMPLE 1

Eighty of the learners at Alexandra High School were surveyed to find out how many minutes each week they spent collecting waste material for recycling. The grouped frequency table shows the results of the survey.

a) Use the frequency table to draw a histogram and to then draw a frequency polygon on the histogram.

b)

i) Find the midpoint of the intervals

ii) Use the table to draw a frequency polygon on a separate set of axes.

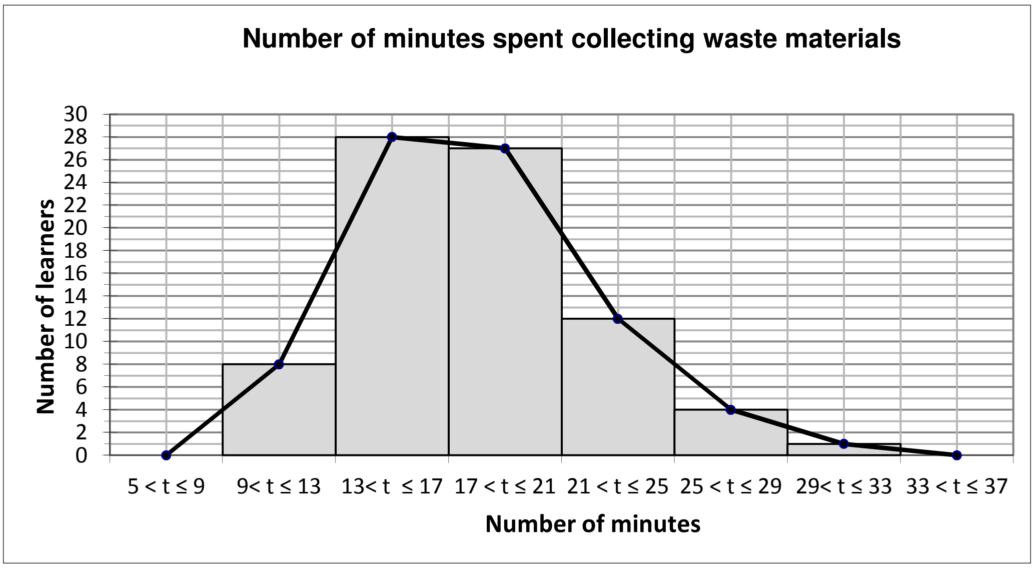
Number of minutes (*t*)	**Number of learners** (*f*)
$9 < t \leq 13$	8
$13 < t \leq 17$	28
$17 < t \leq 21$	27
$21 < t \leq 25$	12
$25 < t \leq 29$	4
$29 < t \leq 33$	1

SOLUTION

a) **Step 1:** Add in two classes with a frequency of zero:

Number of minutes (*t*)	**Number of learners** (*f*)
$5 < t \leq 9$	0
$9 < t \leq 13$	8
$13< t \leq 17$	28
$17 < t \leq 21$	27
$21 < t \leq 25$	12
$25 < t \leq 29$	4
$29 < t \leq 33$	1
$33 < t \leq 37$	0

Step 2: Draw the histogram and then join the midpoints of the top of the columns to form the frequency polygon.



<u>EXAMPLE 2</u> (continued)

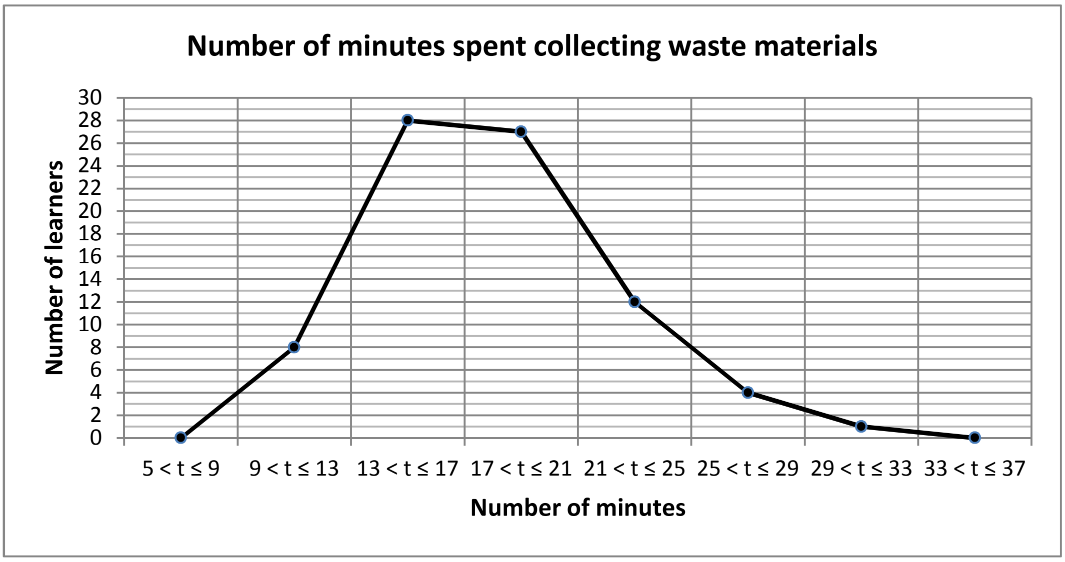
b)

i) Calculate the midpoint of each interval using the formula:

Midpoint = $\frac{lower\ limit\ of\ interval + upper\ limit\ of\ interval}{2}$

Number of minutes (t)	**Mid points**	**Frequency (f)**	**Ordered pairs**
$5 < t \leq 9$	$\frac{5+9}{2} = \frac{14}{2} = 7$	0	(7; 0)
$9 < t \leq 13$	$\frac{9+13}{2} = \frac{22}{2} = 11$	8	(11; 8)
$13 < t \leq 17$	$\frac{13+17}{2} = \frac{30}{2} = 15$	28	(15; 28)
$17 < t \leq 21$	$\frac{17+21}{2} = \frac{38}{2} = 19$	27	(19; 27)
$21 < t \leq 25$	$\frac{21+25}{2} = \frac{46}{2} = 23$	12	(23; 12)
$25 < t \leq 29$	$\frac{25+29}{2} = \frac{54}{2} = 27$	4	(27; 4)
$29 < t \leq 33$	$\frac{29+33}{2} = \frac{62}{2} = 31$	1	(31; 1)
$33 < t \leq 37$	$\frac{33+37}{2} = \frac{70}{2} = 35$	0	(35; 0)

ii) Plot the ordered pairs (midpoint; frequency) and join them with straight lines. Make sure that the graph touches the horizontal axis on both sides.



NOTE:

The ***main advantage*** of using a frequency polygon instead of a histogram is that you can easily draw *two or more* frequency polygons on the same set of axes and make comparisons between the sets of data.

EXAMPLE 3
The Grade 10 and Grade 11 learners were surveyed to find out the approximate number of hours every week they spend doing their Mathematics and Science homework. The results are summarised in the following grouped frequency table:

Number of hours spent on Mathematics and Science homework each week (*t*)	**Number of Grade 10 learners**	**Number of Grade 11 learners**
$5 \le t < 10$	3	12
$10 \le t < 15$	4	22
$15 \le t < 20$	7	10
$20 \le t < 25$	19	6
$25 \le t < 30$	16	8
$30 \le t < 35$	1	2
TOTAL	50	60

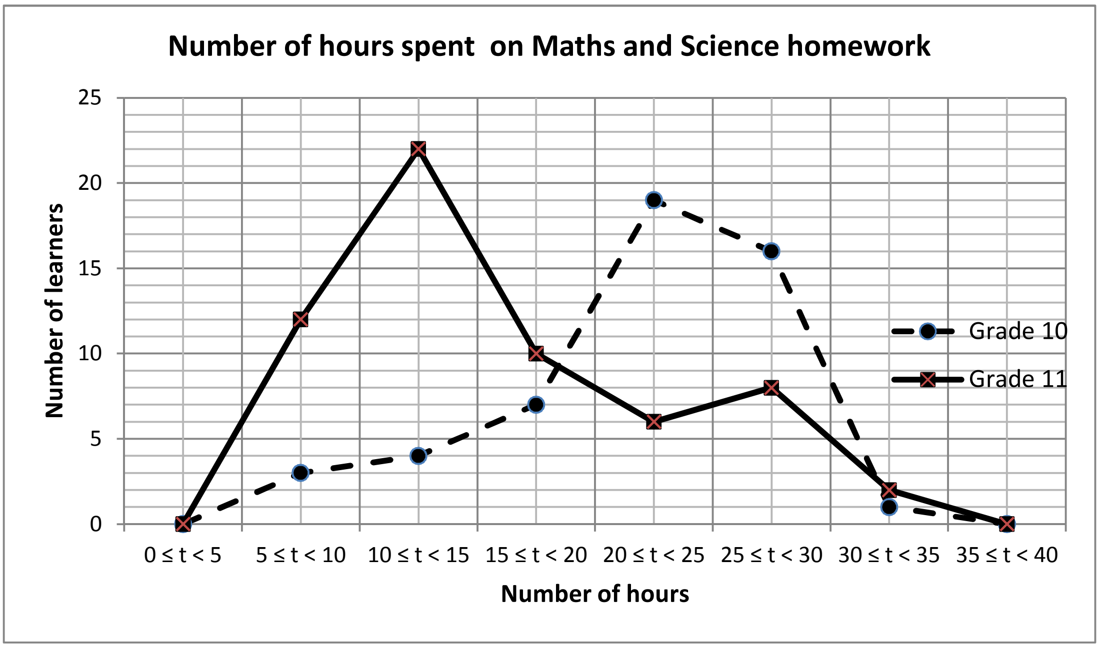
a) Draw two frequency polygons on the same set of axes to illustrate this data.
b) Use the table and the graphs to answer the following:
   i) What is the modal interval for Grade 10 and also for Grade 11?
   ii) Approximately how many more Grade 11 learners than Grade 10 learners spent between 15 and 20 hours doing their homework each week?
   iii) Which grade spent more time doing their homework?

SOLUTION:
a)

Number of hours spent on Mathematics and Science homework each week (*t*)	**Mid-point of the interval**	**Number of Grade 10 learners**	**Number of Grade 11 learners**
$0 \le t < 5$	2,5	0	0
$5 \le t < 10$	7,5	3	12
$10 \le t < 15$	12,5	4	22
$15 \le t < 20$	17,5	7	10
$20 \le t < 25$	22.5	19	6
$25 \le t < 30$	27,5	16	8
$30 \le t < 35$	32,5	1	2
$35 \le t < 40$	37,5	0	0

- *Note that it is not essential to have the same number of data items in the two sets of data.*

<u>EXAMPLE 3</u> (continued)



b)

i) The modal interval for Grade 10 is $20 < t \leq 25$
The modal interval for Grade 11 is $10 < t \leq 15$

ii) Difference in the number of learners who spent between 15 and 20 hours doing homework each week = Number in Grade 11 – Number in Grade 10
= 10 – 7
= 3

So 3 more Grade 11 learners than Grade 10 learners spent between 15 and 20 hours doing homework each week

iii) According to the table:

36 out of 50 Grade 10 learners (72% of them) spent 20 hours or more doing homework each week.

16 out of 60 Grade 11 learners (27% of them) spent 20 hours or more doing homework each week.

So the Grade 10s spent more time on homework than the Grade 11s.

EXERCISE 2.2

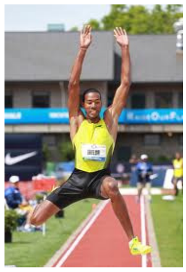
Athletics is a collection of sporting events that involve competitive running, jumping and throwing.

1) The learners at Mjolo High School enjoy taking part in athletics.
Some of the learners took part in the long jump.
The distances they jumped (in metres) are:

5,46	5,97	6,72	6,26	5,13	6,36
6,11	6,38	6,55	5,84	6,20	6,34
5,80	5,43	5,93	6,64	5,67	6,00
6,05	6,88	5,50	5,51	6,10	5,49

a) Copy and complete the following grouped frequency table:

Distance in metres (m)	Frequency	Midpoints
$5{,}00 < m \leq 5{,}50$		
$5{,}50 < m \leq 6{,}00$		
$6{,}00 < m \leq 6{,}50$		
$6{,}50 < m \leq 7{,}00$		



b) Draw a frequency polygon to illustrate the data.
c) Write down the modal interval.

2) Some of the learners took part in the javelin competition. The best distances (in metres) thrown by each competitor in 2011 and 2012 are shown.

Distance thrown in metres (m)	Number of competitors 2011	Number of competitors 2012
$10 < m \leq 20$	0	1
$20 < m \leq 30$	3	4
$30 < m \leq 40$	14	19
$40 < m \leq 50$	21	13
$50 < m \leq 60$	7	11
$60 < m \leq 70$	0	2
TOTAL	**45**	**50**

a) On the same set of axes, draw frequency polygons to illustrate the 2011 and 2012 results.
b) By referring to the table and the frequency polygons, comment on the performance of the competitors in 2011 and 2012.

OGIVES / CUMULATIVE FREQUENCY CURVES

FREQUENCY

Frequency tells us *how many of each item there are in a data set*.

For example
As part of the Census@School, 170 learners were surveyed to find out the type of dwelling that they lived in.

The following table shows the result of the survey:

Type of house that you live in	**Frequency (number of learners)**
Traditional dwelling	7
House on separate yard	76
Tent	1
Informal dwelling in an informal settlement	86
	TOTAL = 170

CUMULATIVE FREQUENCY

Cumulative frequency shows the number of results that are *less than* (<) or *less than or equal to* (≤) a stated value in a set of data.

To find the ***cumulative frequency***,
- Add up the frequencies as you go down the frequency table.
- Write each *running total* or *cumulative frequency* in your table.

For example
Using the above information, we can find the cumulative frequency.

Type of house that you live in	**Frequency (number of learners)**	**Cumulative Frequency**
Traditional dwelling	7	**7**
House on separate yard	76	7 + 76 = **83**
Tent	1	83 + 1 = **84**
Informal dwelling in an informal settlement	86	84 + 86 = **170**
	TOTAL = 170	

Can you see that the last cumulative frequency is equal to the total frequency? (This is a useful check of your addition.)

You can find ***cumulative frequencies*** of discrete data and continuous data.

- ✓ An ***ogive*** or ***cumulative frequency curve*** is a graph that shows the information in a cumulative frequency table. The graph is useful for estimating the median and inter-quartile range of the grouped data.

- ✓ You can draw an ***ogive*** of ungrouped discrete data, grouped discrete data or grouped continuous data. It can be drawn from a grouped frequency table or an ungrouped frequency table.

EXAMPLE 4

The following frequency table shows the time (in minutes) taken by learners to travel to school.

Time taken to travel to school	Frequency	Cumulative Frequency	Ordered Pairs
$0 < t \leq 10$	4		
$10 < t \leq 20$	12		
$20 < t \leq 30$	28		
$30 < t \leq 40$	32		
$40 < t \leq 50$	29		
$50 < t \leq 60$	15		

a) Complete the table.
b) Draw an ogive to illustrate the information.

SOLUTION:

a) Steps to follow when completing the table:

- Add in an interval with a frequency of 0 before the first interval.
- Find the cumulative frequency by adding the frequencies.
- List the ordered pairs where *the first coordinate* = *upper limit of the interval* and *the second coordinate* = *cumulative frequency*.

Note: A cumulative frequency of 105 means that 105 learners or less spent 50 minutes or less to walk to school.

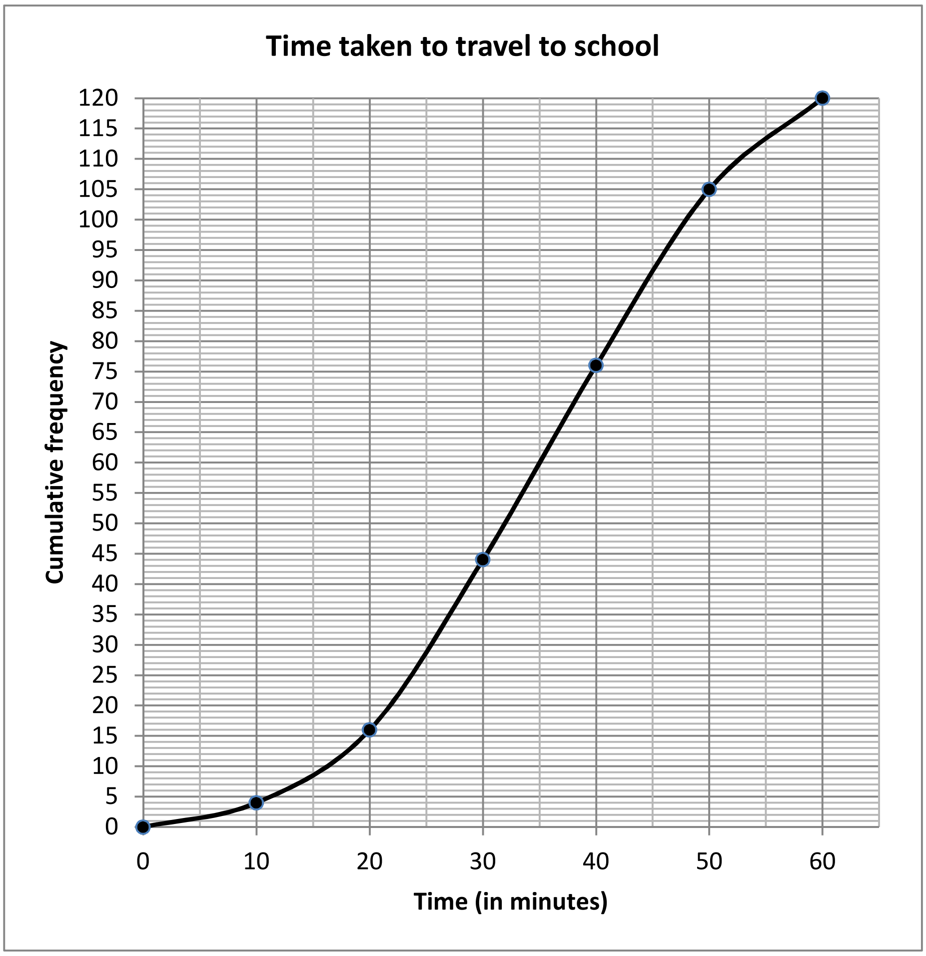
Time taken to travel to school	Frequency	Cumulative Frequency	Ordered Pairs
$-10 < t \leq \mathbf{0}$	0	**0**	**(0;0)**
$0 < t \leq \mathbf{10}$	4	**4**	**(10;4)**
$10 < t \leq \mathbf{20}$	12	**4 + 12 = 16**	**(20;16)**
$20 < t \leq \mathbf{30}$	28	**16 + 28 = 44**	**(30;44)**
$30 < t \leq \mathbf{40}$	32	**44 + 32 = 76**	**(40;76)**
$40 < t \leq \mathbf{50}$	29	**76 + 29 = 105**	**(50;105)**
$50 < t \leq \mathbf{60}$	15	**105 + 15 = 120**	**(60;120)**

b) Draw the ogive as follows:

i) Draw the axes and label the variable on the x-axis and the cumulative frequency on the y-axis.
ii) Plot the ordered pairs.
iii) Join the points to form a smooth curve.

EXAMPLE 4 (continued)

The ogive:



✓ Always remember when drawing cumulative frequency curve from a table of grouped data, the ***cumulative frequencies*** are plotted at the ***upper limit*** of the interval**.**

EXAMPLE 5

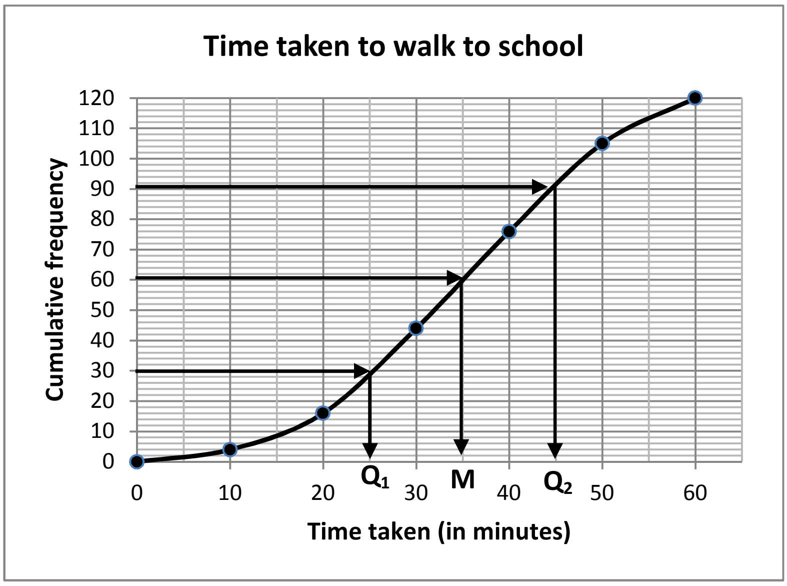
Use the ogive drawn in Example 4 to

a) Determine the approximate values of
   i) the median
   ii) the lower quartile
   iii) the upper quartile of the set of data.

b) What does each of these values tell you about the time taken by the learners?

SOLUTION:

a) This is the ogive drawn in Example 4:



i) To find the approximate value of the ***median (M)***, find the midpoint of the values plotted on the ***cumulative frequency axis.***
- The maximum value is 120, so the median lies between the 60th and 61st term.
- Draw a horizontal line from just above 60 until it touches the ogive.
- From that point draw a vertical line down to the *horizontal* axis.

So the ***median $\approx$ 35 minutes***.

ii) To find the approximate value of the ***lower quartile (Q_1)***, find the midpoint of the lower half of the values plotted on the ***cumulative frequency axis.***
- There are 60 terms in the lower half of the data, so the lower quartile lies between the 30th and the 31st term.
- Draw a horizontal line from just above 30 until it touches the ogive.
- From that point draw a vertical line down to the *horizontal* axis.

So the ***lower quartile $\approx$ 25 minutes***.

<u>EXAMPLE 5</u> (continued)

iii) To find the approximate value of the ***upper quartile (Q_3)***, find the midpoint of the upper half of the values plotted on the ***cumulative frequency axis.***
- There are 60 terms in the upper half of the data, so the upper quartile lies between 60 + 30 = 90th and the 91st term.
- Draw a horizontal line from just above 90 until it touches the ogive.
- From that point draw a vertical line down to the *horizontal* axis.

So the ***upper quartile ≈ 45 minutes***.

b)

i) The ***median*** tells us that 50% of the learners took 35 minutes or less or to walk to school.

ii) The ***lower quartile*** tells us that 25% of the learners took 25 minutes or less to walk to school.

iii) The ***upper quartile*** tells us that 75% of the learners took 45 minutes or less to walk to school.

EXERCISE 2.3

1) In the 2009 Census@School learners were asked what their arm span was, correct to the nearest centimetre. The results of two hundred of the Grade 10, 11 and 12 learners who took part were recorded as follows:

Arm span in cm	Frequency	Cumulative Frequency
$130 < h \leq 135$	16	
$135 < h \leq 140$	26	
$140 < h \leq 145$	42	
$145 < h \leq 150$	54	
$150 < h \leq 155$	26	
$155 < h \leq 160$	22	
$160 < h \leq 165$	14	

To find your arm span: Open arms wide, measure the distance across your back from the tip of your right hand middle finger to the tip of your left hand middle finger.

a) Copy and complete the table.

b) Draw an ogive to illustrate the data.

c) Use your ogive to determine approximately how many learners have arm spans that are less than or equal to 152 cm.

d) Use your graph to determine approximately how many learners have arm spans of between 138 cm and 158 cm.

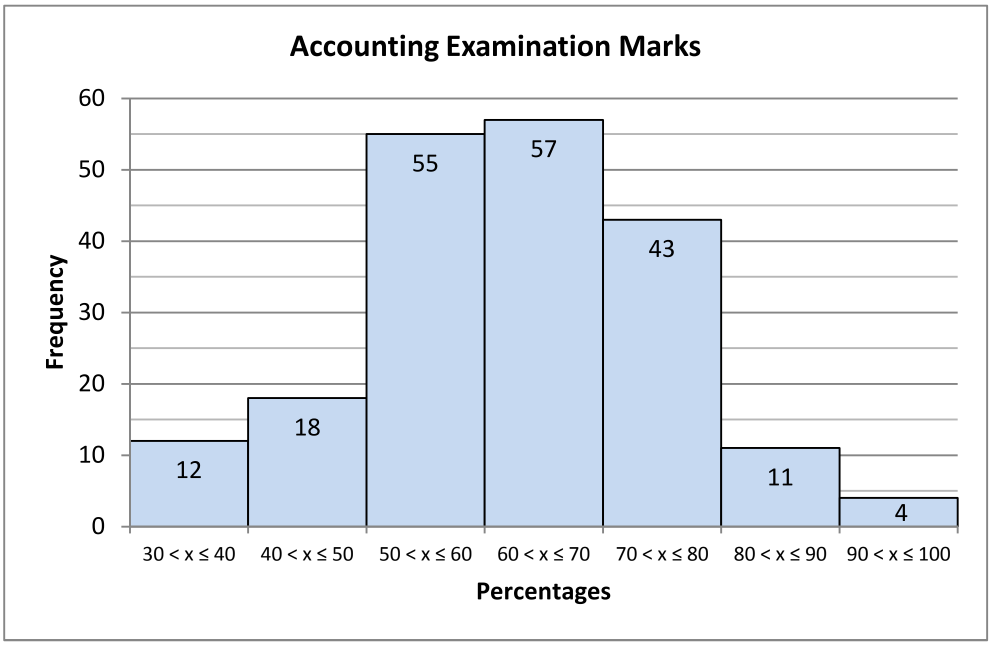
EXERCISE 2.3 (continued)

2) Fifty learners who travel by car to school were asked to record the number of kilometres travelled to and from school in one week. The following table shows the results:

Number of kilometres	Number of learners	Cumulative frequency
$10 < x \leq 20$	2	2
$20 < x \leq 30$		9
$30 < x \leq 40$		13
$40 < x \leq 50$		26
$50 < x \leq 60$		42
$60 < x \leq 70$		50
	TOTAL = 50	

a) Copy the table and then fill in the second column of the table.
b) Draw an ogive to illustrate the data.
c) Use your graph to estimate the median number of kilometres travelled per week.

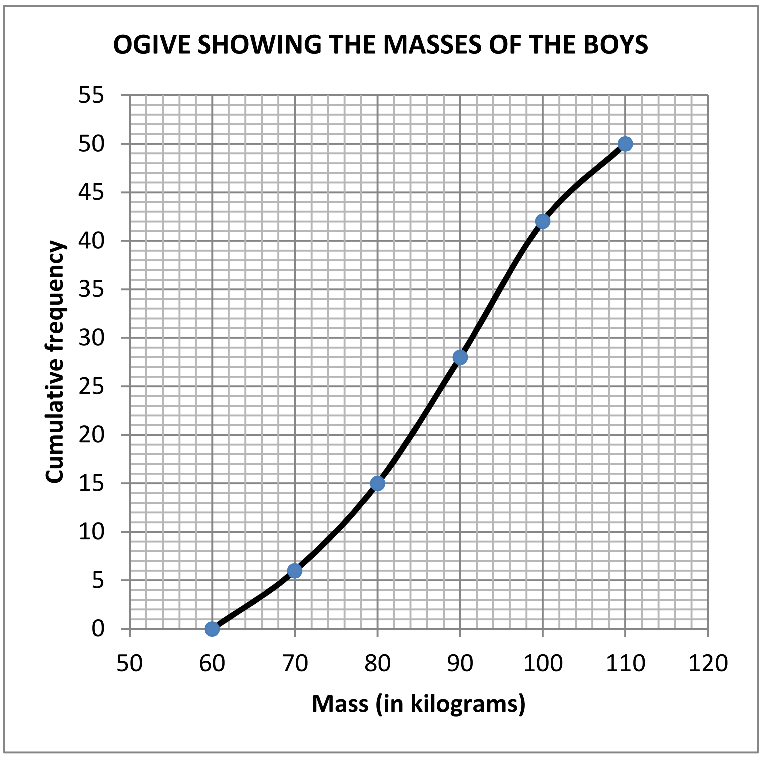
3) The histogram below shows the distribution of the Accounting examination marks for 200 learners.



a) Draw a grouped frequency table to record the data shown on the histogram.
b) Draw an ogive to illustrate the data in the frequency table.
c) Use the ogive to estimate how many learners scored 72% or more for the examination.

EXERCISE 2.3 (continued)

4) The masses of a random sample of 50 boys in Grade 11 were recorded. This cumulative frequency graph (ogive) represents the recorded masses.



a) How many of the boys had a mass between 90 and 100 kilograms?
b) Estimate the median mass of the boys.
c) Estimate how many of boys had mass less than 80 kilograms.

CHOOSING WHICH DISPLAY TO USE

The following table will help you when you have to select the appropriate diagram or graph for your data by identifying the diagrams most commonly associated with different types of data.

DATA TYPE	DESCRIPTION	EXAMPLE	TYPE OF DISPLAY
Qualitative data	Data that can be arranged into categories that are not numerical such as physical traits, gender, and colours.	*To show frequencies* e.g. 10 girls in this class have blonde hair, 18 black hair, 12 brown hair, etc.	Bar graph
		To show proportions e.g. area by province in South Africa: Western Cape 10,6%; Gauteng 1,5%; Free State 10,6%; KwaZulu Natal 7,7%; Limpopo 10,3%; Mpumalanga 6,3%; North West 8,6%; Northern Cape 30,5%.	Pie chart
Discrete Data	Data that has a finite number of different responses such as the number of people in a household.	Few different values 2, 4, 67, 34, 69	Tally table for counting, bar graph for display
		Many different values 4, 24, 25, 26, 45, 37, 38, 48, 53, 120, 75, 67, 100, 89, 47, 58, 87, 55, 45,	Stem-and-leaf diagram or a bar graph or a histogram or a box and whisker diagram or a frequency polygon
Grouped data	Data which have been arranged in groups or classes rather than showing all the original figures.	*Equal intervals* $3 \leq x < 9$ $9 \leq x < 15$ $15 \leq x < 21$	Histogram or frequency polygon
Cumulative data	Data that is increasing by successive additions of the same numbers	*Continuous variable* Time, height, etc.	Ogive
Two samples	Qualitative data	For example comparing the number of learners in the class with brown eyes, blue eyes, and green eyes.	Compound bar graphs or side-by-side pie charts.
	Discrete data	For example comparing the English exam marks for boys and girls in a Grade 11 class.	For ungrouped data use back-to-back stem-and-leaf diagrams or compound bar graphs. For grouped data use frequency polygons
	Continuous data	For example comparing the heights of the girls and the boys in a Grade 11 class.	Ogives or frequency polygons
Two variables	Used to decide whether there is a relationship between the two variables	*Data in pairs* Shoe size and the age of a person.	Scatter plot

VARIANCE AND STANDARD DEVIATION

✓ The ***interquartile range (IQR)*** measures the spread of the *middle half* of the data and is closely linked to the ***median***.

Interquartile range = upper quartile – lower quartile
Or
IQR = $Q_3 - Q_1$

✓ We can define ***two more measures of dispersion***, taking into account all of the data, which are linked to the ***mean***. They are the ***variance*** and the ***standard deviation***.

✓ The ***variance*** is the *mean of the sums of the squares of the deviations from the mean.*
We find the variance by:

i) Finding the mean: $\bar{x} = \frac{\sum x}{n}$
ii) Finding the deviation from the mean of each item of the data set:
Deviation = data item – mean = $x - \bar{x}$
iii) Squaring each deviation : ***(deviation)^2 = $(x - \bar{x})^2$***
iv) Finding the sum of the squares of the deviations:
$\sum(\boldsymbol{deviation})^2 = \sum(\boldsymbol{x} - \bar{\boldsymbol{x}})^2$
v) Finding the mean of the squares of the deviations by dividing by the number of terms in the data set:
$$\textit{Variance} = \frac{\sum(\boldsymbol{deviation})^2}{\boldsymbol{number\ of\ data\ items}} = \frac{\sum(\boldsymbol{x}-\bar{\boldsymbol{x}})^2}{\boldsymbol{n}}$$

✓ The ***standard deviation*** is the square root of the variance:
$$\textit{Standard Deviation} = \sqrt{\boldsymbol{variance}} = \sqrt{\frac{\sum(\boldsymbol{deviation})^2}{\boldsymbol{number\ of\ data\ items}}} = \sqrt{\frac{\sum(\boldsymbol{x}-\bar{\boldsymbol{x}})^2}{\boldsymbol{n}}}$$

✓ When data elements are ***tightly clustered together***, the standard deviation and variance are ***small***; when they are ***spread apart***, the standard deviation and the variance are ***relatively large***.

- A data set with ***more data near the mean*** will have ***less spread*** and a ***smaller standard deviation***
- A data set with ***lots of data far from the mean*** which will have a ***greater spread*** and a ***larger standard deviation***.

EXAMPLE 6

a) Calculate the variance and the standard deviation of the following two data sets:

Set A	182	182	184	184	185	185	186
Set B	152	166	176	184	194	200	216

b) Use the two standard deviations to compare the distribution of data in the two sets.

SOLUTION:

a) **Step 1:** Find the mean of each set

Mean of Set A = $\frac{182+182+184+184+185+185+186}{7} = \frac{1\,288}{7} = 184$

Mean of Set B = $\frac{152+166+176+184+194+200+216}{7} = \frac{1\,288}{7} = 184$

Step 2: Find the deviation from the mean of each item in the data set

Step 3: Square each deviation

Step 4: Find the variance

Step 5: Find the standard deviation

GROUP A			**GROUP B**		
Data item	**Deviation from the mean**	**(Deviation)^2**	**Data item**	**Deviation from the mean**	**(Deviation)^2**
182	182 – 184 = –2	$(-2)^2 = 4$	152	152 – 184 = – 32	$(-32)^2 = 1\,024$
182	182 – 184 = –2	$(-2)^2 = 4$	166	166 – 184 = – 18	$(-18)^2 = 324$
184	184 – 184 = 0	$0^2 = 0$	176	176 – 184 = – 8	$(-8)^2 = 64$
184	184 – 184 = 0	$0^2 = 0$	184	184 – 184 = 0	$0^2 = 0$
185	185 – 184 = 1	$1^2 = 1$	194	194 – 184 = 10	$10^2 = 100$
185	185 – 184 = 1	$1^2 = 1$	200	200 – 184 = 16	$16^2 = 256$
186	186 – 184 = 2	$2^2 = 4$	216	216 – 184 = 32	$32^2 = 1\,024$
		$\sum(\text{deviations})^2 = 14$			$\sum(\text{deviations})^2 = 2\,792$
$\boldsymbol{Variance = \frac{\sum(deviation)^2}{number\ of\ data\ items}} = \frac{14}{7} = \mathbf{2}$			$\boldsymbol{Variance = \frac{\sum(deviation)^2}{number\ of\ data\ items}} = \frac{2\,792}{7} = \mathbf{398,857\ ...}$		
$\boldsymbol{Standard\ Deviation = \sqrt{variance}} = \sqrt{2} \approx 1{,}414$			$\boldsymbol{Standard\ Deviation = \sqrt{variance}} = \sqrt{\frac{2\,792}{7}} \approx \mathbf{19{,}971}$		

b) The larger standard deviation in Group B indicates that the data items are generally much further from the mean than the data items in Group 1.
This means that the data items in Group B are more spread out than the data items in Group A.

EXAMPLE 7
Use a scientific calculator to calculate the standard deviation of 9, 7, 11, 10, 13 and 7.

SOLUTION:

CASIO *fx-82ZA PLUS* calculator	**SHARP EL-W535HT calculator**
Press the following keys [MODE] [2: STAT] [1: 1 – VAR] 9 (=) 7 (=) 11 (=) 10 (=) 13 (=) 7 (=) [AC] [SHIFT : 1] [STAT] [4: VAR] [3: σx] [=]	Press the following keys: [MODE] [1: STAT] [0:SD] 9 [CHANGE] 7 [CHANGE] 11 [CHANGE] 10 [CHANGE] 13 [CHANGE] 7 [CHANGE] [ALPHA] [6: σx] [=]

So the Standard Deviation ≈ 2,141

EXERCISE 2.4

Where necessary, give decimal answers correct to 1 decimal place

1) The arm spans (in cm) of the eleven players in each of two different soccer teams A and B are recorded.
   a) The arm spans for **TEAM A** are:
      203, 214, 187, 188, 196, 199, 205, 203, 199, 194 and 206
      i) Calculate the mean of the arm spans using the formula: $\bar{x} = \frac{\sum x}{n}$.
      ii) Copy and complete the table given.
      iii) Calculate the standard deviation of the arm spans using the formula: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$.
   b) For **TEAM B**, the variance is 875 cm^2. Calculate the standard deviation of the arm spans of **TEAM B.**
   c) Make a comment about the dispersion of the arm spans of the players in both teams.

$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$	$\boldsymbol{(x - \bar{x})^2}$
203		
214		
187		
188		
196		
199		
205		
203		
199		
194		
206		
$n =$		$\sum(x - \bar{x})^2 =$

EXERCISE 2.4 (continued)

2) The time (in minutes) taken by a group of athletes from Lesiba High School to run a 3 km cross country race is: 18 21 16 24 28 20 22 29 19 23
   Use your calculator to determine
   a) The mean time taken to complete the race.
   b) The standard deviation of the time taken to complete the race.

3) The following tables show the masses (in kilograms) of the A and B rugby teams at Sir John Adamson High School:

   TEAM A

51	82	71	64	81	81	76	77	62	68
70	74	81	61	68	69	67	71	68	74
80	62	70	68	62					

   TEAM B

83	79	67	79	87	62	60	83	76	79
94	110	73	97	70	68	103	85	74	55
47	63	62	87	74					

   a) Use your calculator to determine the mean and the standard deviations of each data set.
   b) Is the standard deviation a good measure for determining which team plays better? Give reasons for your answer.

4) As part of the Census@School, learners had to record the length (in centimetres) of their right foot without a shoe. The girls (G) and boys (B) in Grade 11C measured their foot lengths and recorded the results in the following table.

G:	29	22	28	23	23	29	29	25	27	23	27	21	24	21	20	25	22	29
B:	28	30	26	29	25	28	26	25	28	22	30	25	21	27	25	23		

   a) Use your calculator to determine the mean and standard deviation of the foot lengths of
      i) The girls
      ii) The boys.
   b) Use the mean and the standard deviation of the foot lengths to comment on the differences in foot sizes of the two groups.

SYMMETRIC AND SKEWED DATA

✓ A ***measure of shape*** describes the *distribution of the data* within a data set.

✓ A distribution of data values can be ***symmetric*** or ***skewed***.

- In a ***symmetric distribution***, the two sides of the distribution are a mirror image of each other

- In a ***skewed distribution***, the two sides of the distribution are NOT mirror images of each other.

✓ Both ***frequency polygons*** and ***box-and-whisker diagrams*** can be used to illustrate symmetric and skewed data.

KEY FEATURES OF A SYMMETRIC DISTRIBUTION

- ***The shape is symmetrical***
- ***The mode, median and mode have the same value***.
- ***Most of the data are clustered around the centre.***

  In fact, about 68% of the data lie within 1 standard deviation of the mean
  About 95% of the data lie within 2 standard deviations of the mean
  About 99,7% of the data lie within 3 standard deviations of the mean.

KEY FEATURES OF SKEWED DATA
Skewness is the tendency for the values to be more frequently around the high or low ends of the *x*-axis.

- With a ***positively skewed distribution***, the ***tail on the right side is longer*** than the left side

  Most of the values tend to cluster toward the left side of the x-axis (i.e. the smaller values) with increasingly fewer values on the right side of the x-axis (i.e. the larger values).

- With a ***negatively skewed distribution***, the ***tail on the left side is longer*** than the right side.

  Most of the values tend to cluster toward the right side of the x-axis (i.e. the larger values) with increasingly fewer values on the left side of the x-axis (i.e. the smaller values).

EXAMPLE 8

The Grade 10 learners of Leihlo Secondary School, Helen Frans Secondary School and Pitseng Secondary School attended a meeting at a hall in Senwabarwana about the problems they have encountered with the bus company which transports them to school.

The following table shows the time the learners spent in the meeting:

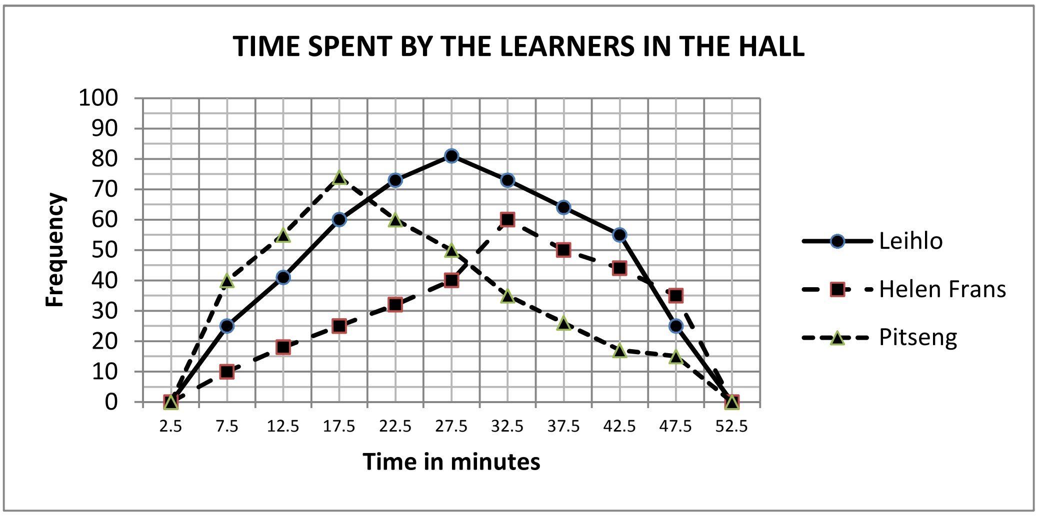
Time spent in the hall (in minutes)	Midpoint of the intervals (in minutes)	Frequency		
		Leihlo Secondary School	Helen Frans Secondary School	Pitseng Secondary School
$0 < t \leq 5$	2,5	0	0	0
$5 < t \leq 10$	7,5	25	10	40
$10 < t \leq 15$	12,5	41	18	55
$15 < t \leq 20$	17,5	60	25	74
$20 < t \leq 25$	22,5	73	32	60
$25 < t \leq 30$	27,5	81	40	50
$30 < t \leq 35$	32,5	73	60	35
$35 < t \leq 40$	37,5	64	50	26
$40 < t \leq 45$	42,5	55	44	17
$45 < t \leq 50$	47,5	25	35	15
$50 < t \leq 55$	52,5	0	0	0

a) Draw frequency polygons to represent the time spent by learners of each school in the hall.

b) Describe the shapes of the polygons.

SOLUTION:

a)



b) The data from Leihlo Secondary is symmetric
The data from Helen Frans is not symmetric. It is more spread out on the left and clustered more closely together on the right. We say that it is ***skewed left.***
The data from Pitseng is also not symmetric. It is more spread out on the right and clustered more closely together on the left. We say that it is ***skewed right***.

✓ Note that if the mean and the median of a data set are known, then

- If ***mean – median ≈ 0,*** then the distribution is ***symmetric***
- If ***mean – median > 0,*** then the distribution is ***positively skewed***
- If ***mean – median < 0,*** then the distribution is ***negatively skewed***

✓ Note that for a box-and-whisker diagram

- If the distribution is symmetric, the median is in the middle of the box and the whiskers are equal in length
- When data is more spread out on the left side and clustered on the right, the distribution is said to be negatively skewed or skewed to the left.
- When the data is more spread out on the right side clustered on the left, the distribution is said to be positively skewed or skewed to the right.

EXAMPLE 9

Use the data given in Example 8 for the following:

a) Calculate the mean and the five-number-summary for the time spent by learners in each school.

b) Draw box-and-whisker diagrams to represent the data.

c) State whether each data set is symmetric, positively skewed or negatively skewed.

<u>**SOLUTION**</u>**:**

<u>**Leihlo Secondary School**</u>

a) $\bar{x} \approx 27{,}5$ minutes

Minimum value ≈ 7,5 minutes

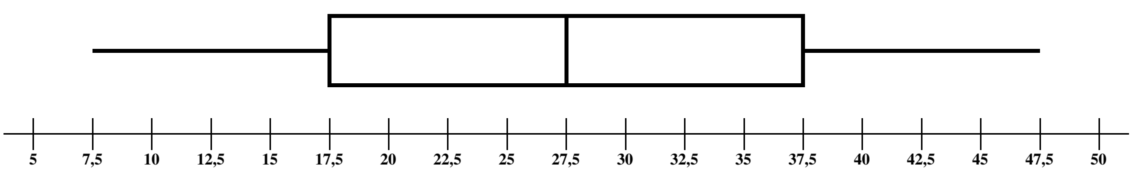
Lower quartile = $Q_1 \approx 17{,}5$ minutes

Median ≈ 27,5 minutes

Upper quartile = $Q_3 \approx 37{,}5$ minutes

Maximum value ≈ 47,5 minutes

b)



c) Mean – median = 27,5 minutes – 27,5 minutes = 0

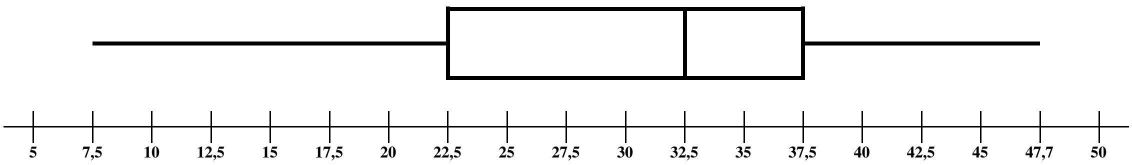
This means that the distribution is symmetric.

<u>EXAMPLE 9</u> (continued)

<u>Helen Frans Secondary School</u>

a) $\bar{x} \approx 31{,}6$ minutes
Minimum value ≈ 7,5 minutes
Lower quartile = $Q_1 \approx 22{,}5$ minutes
Median ≈ 32,5 minutes
Upper quartile = $Q_3 \approx 37{,}5$ minutes
Maximum value ≈ 47,5 minutes

b)

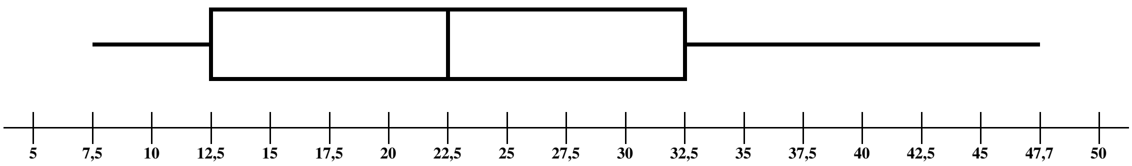


c) Mean – median = 31,6 minutes – 32,7 minutes = – 1,1
This means that the distribution is negatively skewed (or skewed left).

<u>Pitseng Secondary School</u>

a) $\bar{x} \approx 22{,}99$ minutes
Minimum value ≈ 7,5 minutes
Lower quartile = $Q_1 \approx 12{,}5$ minutes
Median ≈ 22,5 minutes
Upper quartile = $Q_3 \approx 32{,}5$ minutes
Maximum value ≈ 47,5 minutes

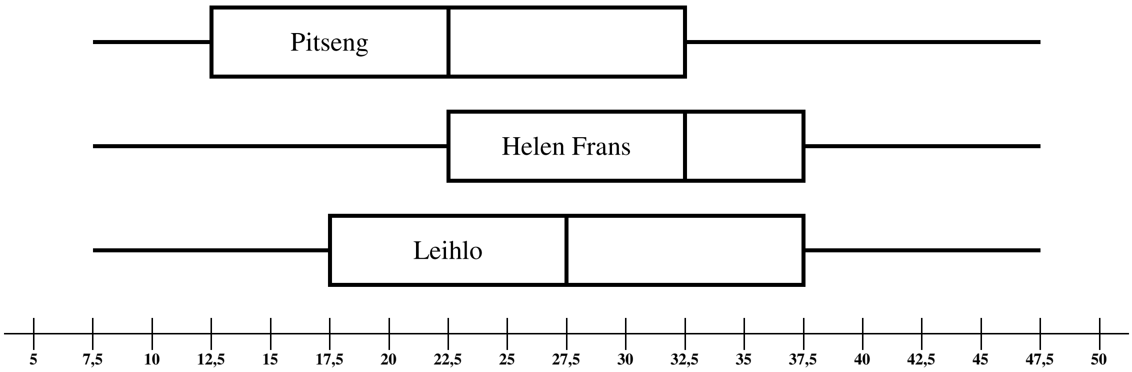
b)

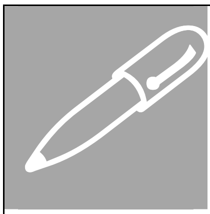


c) Mean – median = 22,99 minutes – 22,5 minutes = 0,49
This means that the distribution is positively skewed (or skewed right).

EXAMPLE 9 (continued)

When we draw all three box and whisker diagrams on the same page, we can immediately see that the Leihlo data is symmetric, the Helen Frans data is negatively skewed, and the Pitseng data is positively skewed.





EXERCISE 2.5

1) The box and whiskers diagrams of two sets A and B are shown below.

   a) Write down what is common to both sets of data.
   b) Which data set is symmetrical? State the reasons.
   c) Is the other data set skewed left or right? State the reasons.

2) For the 2009 Census@School, 47 Grade 11 learners recorded how long (in minutes) it took them to travel to school. The following data was obtained:

Time (in minutes)	Frequency
$5 < t \leq 10$	1
$10 < t \leq 15$	5
$15 < t \leq 20$	9
$20 < t \leq 25$	13
$25 < t \leq 30$	11
$30 < t \leq 35$	8

   a) Use the given information to determine the five number summary.
   b) Draw a box and whisker diagram to illustrate the five number summary.
   c) Comment on the spread of the time taken to complete the task.

<u>EXERCISE 2.5</u> (continued)

3) Three high schools in Limpopo have a total number of 132 Grade 12 learners. These learners completed the question in the 2009 Census@School where they were asked to record the distance (in kilometres) they travel each day from home to school. The results of the survey are shown in the grouped frequency below.

Distance in kilometres (x)	***Frequency***	***Midpoint of intervals***
$0 < x \leq 5$	12	
$5 < x \leq 10$	29	
$10 < x \leq 15$	13	
$15 < x \leq 20$	63	
$20 < x \leq 25$	12	
$25 < x \leq 30$	3	

a) Copy and complete the table.
b) Draw a frequency polygon to illustrate the data.
c) Determine the median of the data in the table.
d) Use your calculator to determine the mean of the data.
e) Calculate ***mean – median***
f) By referring to the shape of the polygon and the relationship between the mean and the median, state whether the distribution of the data is symmetric, positively skewed or negatively skewed.

OUTLIERS

- ✓ An ***outlier*** is a data entry that is ***far removed from the other entries*** in the data set e.g. a data entry that is much smaller or much larger than the rest of the data values.

- ✓ An outlier has an influence on the ***mean*** and the ***range*** of the data set, but has no influence on the median or lower or upper quartiles.

- ✓ An outlier can affect the ***skewness*** of the data.

- ✓ Any data item that is **less than Q_1 – 1,5 × IQR** <u>OR</u> **more than Q_3 +1,5 × IQR** is an outlier.

EXAMPLE 1

Investigate the following data set:

1, 8, 12, 14, 14, 15, 17, 17, 19, 26, 32

a) Calculate (where necessary correct to 1 decimal place)

i) The mean

ii) The median

iii) The interquartile range

b) Are any of the entries in the data set outliers?

<u>SOLUTION</u>:

a)

i) Mean = $\bar{x} = \frac{1+8+12+14+14+15+17+17+19+26+32}{11}$

$= \frac{175}{11}$

$= 15{,}9090...$

$\boldsymbol{\bar{x} \approx 15{,}9}$

ii) There are 11 terms so the median is the 6^{th} term.

Median = 15

iii) There are 5 terms less than the median so Q_1 is the 3^{rd} term. So $Q_1 = 12$.

To find Q_3 we add 3 terms to the position of the median and get the 9^{th} term.

So $Q_3 = 19$

IQR = 19 – 12 = 7

b) Lower outlier < **Q_1 – 1,5 × IQR**

$< 12 - 1{,}5 \times 7$

$< 1{,}5$

So 1 is an outlier

Upper outlier > **Q_3 +1,5 × IQR**

$> 19 + 1{,}5 \times 7$

$> 29{,}5$

And 32 is also an outlier

EXERCISE 2.6

1) Determine the interquartile range and then find outliers (if there are any) for the following set of data:
   10,2 ; 14,1 ; 14,4 ; 14,4 ; 14,5 ; 14,5 ; 14,6 ;
   14,7 ; 14,7 ; 14,9 ; 15,1 ; 15,9 ; 16,4 ; 18,9

2) A class of 20 learners has to submit Mathematics assessment tasks over the course of the year. While some learners were conscientious others were not.

   The following table shows the number of assessment tasks each learner handed in:

9	5	11	8	12	2	6	9	15	10
12	6	9	3	9	13	14	16	4	7

   a) Determine the IQR
   b) Determine the outliers (if any).

3) The following are the ages of boys in one of the Grade 8 class of Dendron Secondary School:
   12 12 13 14 14 13 12 15 15 14 12 19 14 12 9
   a) Determine the five number summary.
   b) Determine the outliers, if any.

REFERENCES

Bowie L. et al. (2007). *Focus on Mathematical Literacy Grade 12*. Maskew Miller Longman.
Freund J. E. (1999). *Statistics A First Course*. Prentice Hall, New Jersey.
Larson R. and Farber B. (2006). *Elementary Statistics Picturing the World*. Third Edition. Pearson, Prentice Hall.
Upton G and Cook I (2001) *Introducing Statistics 2nd edition*. Oxford
Statistics South Africa (2010) *Census At School Results (2009).*
The Answer Series *Grade 12 Mathematics Paper 3, notes, questions and answers.*

3

Grade 12 Data Handling

In this chapter you:
- Look at the difference between univariate and bivariate data.
- Draw scatter plots.
- Describe the correlation between any two sets of data in a scatter plot.
- Determine whether the correlation is linear, quadratic or exponential.
- Calculate the correlation coefficient.
- Draw an intuitive line of best fit.
- Find the equation of regression lines.
- Use regression lines to make predictions.

WHAT YOU LEARNED ABOUT DATA HANDLING IN GRADE 11

In Grade 11 you covered the following data handling concepts:
- Histograms
- Frequency polygons.
- Ogives (also called cumulative frequency curves)
- Symmetric and skewed data
- The identification of outliers.

UNIVARIATE AND BIVARIATE DATA

✓ In previous chapters you analysed ***univariate data***.

- ***Univariate*** means "one variable" (one type of data).
- Univariate data involves a ***single variable***.
- You can graph univariate data using a pictograph, bar graph, pie chart, histogram, frequency polygon, line graph, broken line graph or ogive.

Examples of univariate data are:
- o The time taken by learners to get to school.
- o Mathematics test marks.

✓ ***Bivariate*** means "two variables".

- Bivariate data involves ***two variables***, and you investigate these two variables in order to find out if there are connections between sets of data.

  For example, if you investigate height and mass, you probably will find that *taller learners have a greater mass than shorter learners*.

- You can graph bivariate data on a **scatter plot** (also called a **scatter diagram** or **scatter graph**).

  Examples of bivariate data are:
  - shoe size and the length of a person's forearm,
  - time spent watching television and mathematics test scores,
  - hand span and foot length,
  - the speed of a car and the petrol consumption.
  - the age of a learner and the grade she is in.

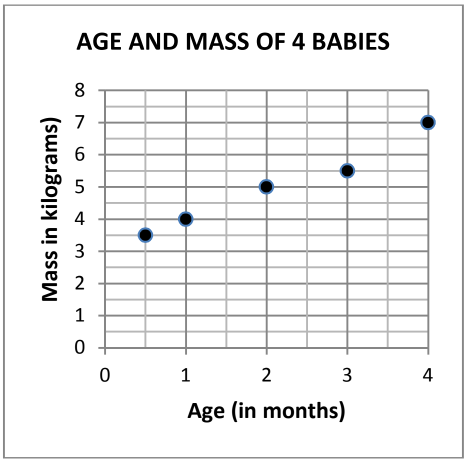
✓ The following table shows the differences between ***univariate data*** and ***bivariate data***.

UNIVARIATE DATA	BIVARIATE DATA
Data that consist of a **single variable** ***For example:*** mathematics test scores.	Data that consist of **two variables** ***For example:*** time spent studying mathematics and the corresponding mathematics test marks.
Data is analysed by: • Determining measures of central tendency – mean, mode, median • Determining measures of dispersion – range, interquartile range, variance and standard deviation. • Drawing bar graphs, histograms, frequency polygons, ogives, pie charts, broken line graphs and box-and-whisker diagrams.	Data is analysed by: • identifying the independent and dependent variables • determining if a relationship or correlations exist between the variables • determining the strength of the relationship or correlation
Answers **questions** like: How many of the learners in the Physical Sciences class are females?	Answers **questions** like: Describe the relationship/correlation between the time spent studying and test scores.

SCATTER PLOTS AND TYPES OF CORRELATION

✓ A **scatter plot** is a graph that helps you to see whether there is a **correlation** (relationship) between any set of two numeric data.

- It has two axes, one for each variable.
- Each ordered pair of values is plotted as a point on the graph.
  The *x*-coordinate is the **independent variable**.
  The *y*-coordinate is called the **dependent variable.**



- An **independent variable** is the variable that determines the value of another variable (the dependent variable). This variable can often be manipulated.

- A **dependent variable** is the variable whose values depend on the corresponding values of the domain or independent variable.

✓ If there is a relationship between the two variables then you can say the variables are **correlated**.

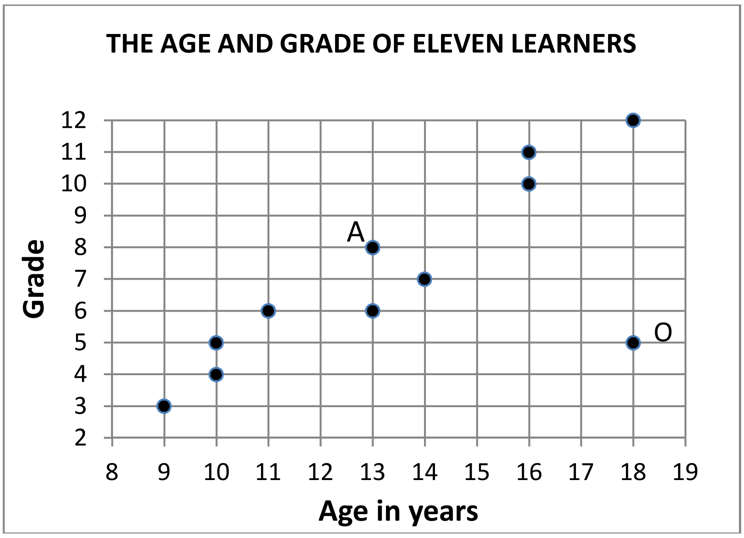
a) Drawing a Scatter Plot

✓ As part of the 2009 Census@School, learners were asked to give their date of birth and to say what grade they were in at school. The ages and grades of a random sample of 13 girls who took part in the Census@School are given the in the following table:

Age (in years)	13	16	16	11	14	13	18	10	9	10	18
Grade	8	10	11	6	7	6	12	4	3	5	2

✓ To see if there is a correlation between age and grade, we can draw a scatter plot. It is often assumed that older learners are in higher grades, hence the age of the learners is the independent variable and the grade of the learners is the dependent variable.

DRAWING A SCATTER PLOT	
STEP 1:	Draw two axes. Plot the independent variable (age in years) along the horizontal axis and the dependent variable (grade) along the vertical axis.
STEP 2:	Choose an appropriate scale for each axis. Here you can make both intervals 1 unit.
STEP 3:	Plot each set of points. For example, the first learner listed in the table is 13 years old and is in Grade 8, so the point (13; 8) is plotted at A on the graph.



- In this scatter plot the points are spread in a rough 'corridor' going up from left to right. This shows a ***positive relationship*** between the two variables. This means that generally an older child is in a higher grade.

- The point O (18; 5) stands out and is well away from the other points and clearly is not part of the main trend of the points indicated above. Such a point is called an ***outlier***. An outlier can have an extreme x-value, an extreme y-value, or both.

- The learner at O is 18 years old and is in Grade 5. This outlier may have occurred because a learner had not progressed due to sickness or maybe because the data was collected incorrectly.

b) Types of Correlation

✓ In analysing the scatter plot, you look for a pattern in the way the points lie. Certain patterns tell you that correlations (relationships) exist between the two variables.

✓ When describing the relationship between two variables displayed on a scatter plot, we should comment on:
a) The form – whether it is linear or non-linear (either a quadratic or exponential curve).
b) The direction – whether it is positive or negative
c) The strength – whether it is strong, moderate or weak.

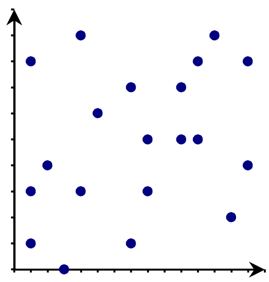
✓ This table shows different types of correlation.

ZERO CORRELATION

The points are scattered randomly over the graph indicating no pattern between the two sets of data. This tells you that there is **no correlation** between the two variables.

Examples showing this type of correlation:

- Marks on chemistry exam and marks on art exam.
- Type of fruits that people prefer and their shoe sizes.

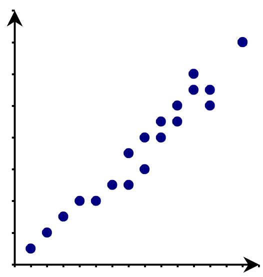


STRONG POSITIVE CORRELATION

The points show a 'band' that slopes upwards from bottom left to top right. As one variable increases, the other variable also increases. Such a pattern shows a **strong positive correlation**.

Examples showing this type of correlation

- Number of pages in a newspaper and the mass of the newspaper.
- Time spent studying and mathematics marks.

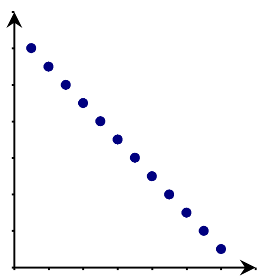


STRONG NEGATIVE CORRELATION

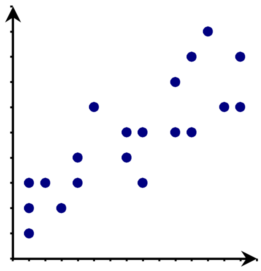
The points show a 'band' that slopes downwards from top left to bottom right. As one variable increases, the other variable decreases. Such a pattern shows a **strong negative correlation**.

Examples:

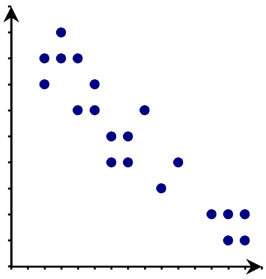
- Time spent watching TV and test marks.



MODERATLY POSITIVE CORRELATION
The points are obviously clustered from bottom left to top right, but are not clustered together as closely as with the strong positive correlation.



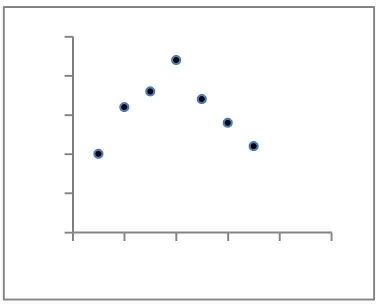
MODERATE NEGATIVE CORRELATION
The points are obviously clustered from top left to bottom right, but are not clustered together as closely as with the strong negative correlation.



- ✓ In many real-life situations, scatter plots follow patterns that are approximately **linear**. However, it might sometimes look as though there is no correlation between the variables. The points might look as though they are randomly scattered over the plane. However, on closer inspection you may be able to recognise a **quadratic** or an **exponential** shape to the pattern of points or any other pattern.

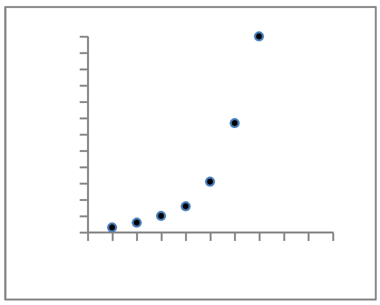
- ✓ Consider the examples given below.

QUADRATIC RELATIONSHIP



The points move upwards from left to right until they reach a peak point. From the peak point, they follow a downwards movement.

EXPONENTIAL RELATIONSHIP



The points in this scatter plot follow a curve from left to right, and show an exponential correlation.

EXAMPLE 1

Besides giving their date of birth, the learners who took part in the 2009 Census@School also had to say how tall they were without their shoes on. This measurement had to be given correct to the nearest centimetre. The following tables give data on the age and corresponding height of 12 South African girls taken from 2009 Census@school data base.

Age (years)	8	19	11	12	17	13
Height (cm)	128	180	150	157	166	160

Age (years)	15	16	9	18	10	14
Height (cm)	112	173	140	178	190	163

a) Is age or height the dependent variable? Give a reason for your answer.

b) Draw a scatter plot and use it to describe the relationship between the ages and heights of the girls.

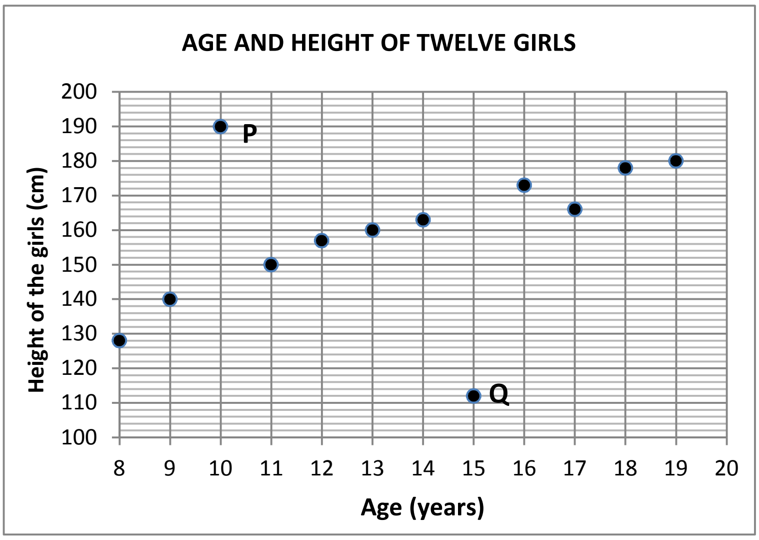
c) Do you think you would see this same correlation amongst a group of women aged from 25 to 35 years of age? Explain why or why not.

d) Which point shows a girl that is an outlier? Label the point(s) representing the outliers with a P and a Q.

<u>SOLUTION</u>:

a) Height is the dependent variable – usually the height of a person depends on how old the person is.

b)



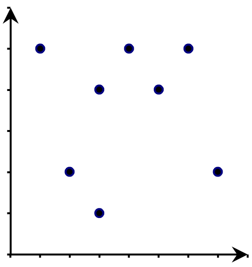
The points fall in an upward-sloping diagonal band. This indicates that as a child gets older her height increases. Older girls are taller hence there is a **strong positive linear correlation** between age and height

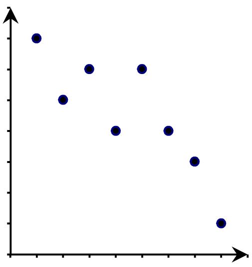
c) No, we won't see the same correlation amongst a group of women aged from 25 to 35 years because girls stop growing by about 18 or 19. So, as age increases, we are going to find that the height remains constant.

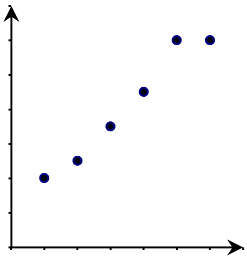
d) The points P(10;190) and Q(15;112) represent girls who are outliers. **P** represents a very tall 10-year old girl while **Q** represents a very short 15-year old girl.

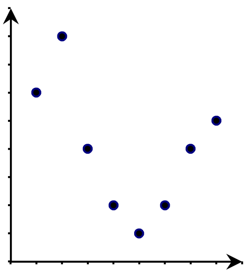
EXERCISE 3.1

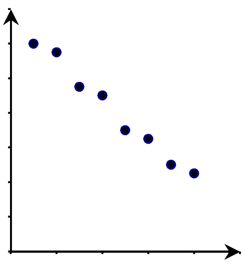
1) Describe the relationship that exists between the variables in the scatter plots below. Say whether there is:
   - Zero correlation
   - A strong positive correlation
   - A strong negative correlation
   - A moderate positive correlation
   - A moderate negative correlation
   - A non-linear correlation (either a quadratic or exponential relationship).

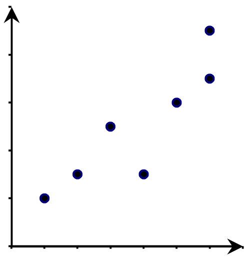
a)


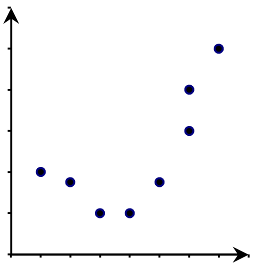
b)


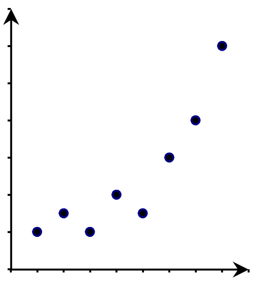
c)


d)


e)


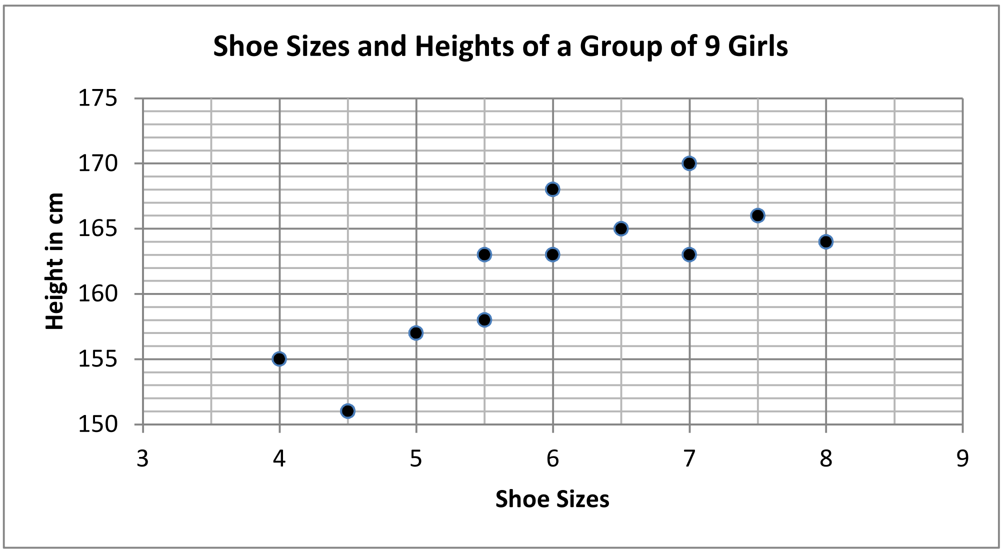
f)


g)


h)


EXERCISE 3.1 (continued)

2) The scatter graph below shows the shoe sizes and heights of a group of 9 girls.



a) What shoe size does the tallest girl wear?
b) How tall is the girl with the largest shoe size? Give your answer in metres.
c) Does the shortest girl wear the smallest shoes?
d) What do you notice about the shoe sizes of the taller girls compared to the shoe sizes of the shorter girls?
e) Describe the correlation shown by this scatter graph.

3) Given below are heights and foot lengths (both rounded off to the nearest centimetre) of eleven learners from Eastern Cape schools as recorded in the 2009 Census@School:

Foot length (cm)	27	24	24	26	23	22	19	24	20	23	25
Height (cm)	174	158	163	175	126	153	170	160	131	156	165

a) Draw a scatter plot to show the relationship between foot length and height.
b) Are any of these points outliers? Explain why they are outliers.
c) What does the graph tell you about the correlation between foot length and height?

CORRELATION COEFFICIENT

✓ Where a linear association exists between two variables, we say that the two variables ***correlate***. A commonly used statistical measure of association is called the ***correlation coefficient***.

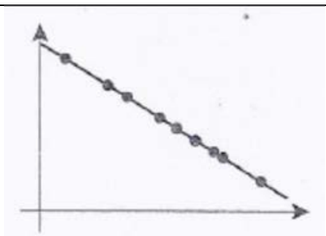
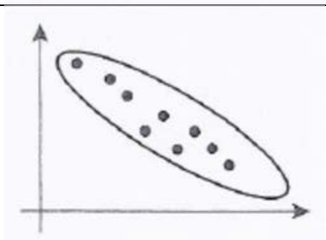
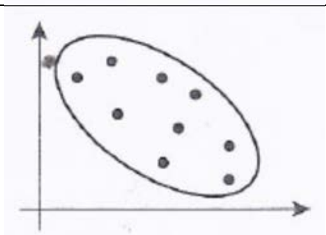
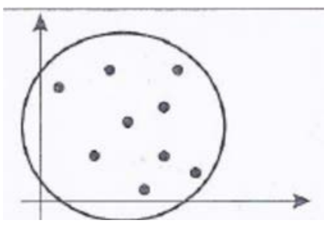
✓ The correlation coefficient is ***a measure of the strength and direction of the linear relationship between two variables***.

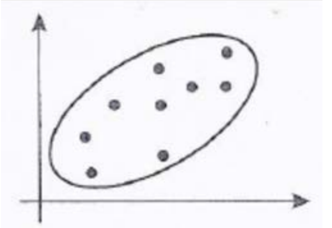
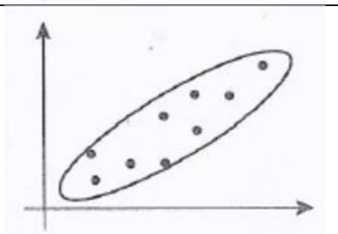
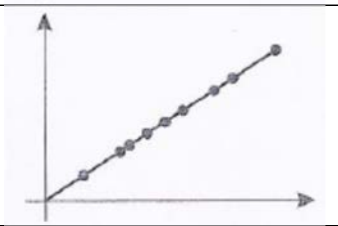
✓ The symbol 'r' is used to represent the sample correlation coefficient.
The formula for r is $\boldsymbol{r} = \frac{n\sum xy-(\sum x)(\sum y)}{\sqrt{n\sum x^2-(\sum x)^2}\,\sqrt{n\sum y^2-(\sum y)^2}}$ where n is the number of pairs of data.

✓ The range of the correlation coefficient is –1 to 1.

- If x and y have strong positive correlation, r is close to 1.
- If x and y have strong negative linear correlation, r *is* close to –1.
- If there is no linear correlation or there is a weak linear correlation, r is close to 0.

Consider the examples given below:

	$r = 1$	Perfect negative linear correlation
	$-1 < r \leq -0{,}7$	Strong negative linear correlation
	$-0{,}7 < r < -0{,}3$	Weak negative linear correlation
	$-0{,}3 < r \leq 0{,}3$	No significant linear correlation

	$0,3 < r \leq 0,7$	Weak positive linear correlation
	$0,7 < r < 1$	Strong positive linear correlation
	$r = 1$	Perfect positive linear correlation

✓ The correlation coefficient r has no units.

✓ The correlation coefficient only measures the strength of a ***linear association***.

a) Finding the Correlation Coefficient Using the Formula

✓ The correlation coefficient can be found as follows:

In Words	Symbols
Find the sum of the x-values	$\sum \boldsymbol{x}$
Find the sum of the y-values	$\sum \boldsymbol{y}$
Multiply each x-value by its corresponding y-value and then find the sum	$\sum \boldsymbol{xy}$
Square each x-value and find the sum of the squares	$\sum \boldsymbol{x^2}$
Square each y-value and find the sum of the squares	$\sum \boldsymbol{y^2}$
Use these five sums to calculate the correlation coefficient	$\boldsymbol{r = \dfrac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}}$

EXAMPLE 2

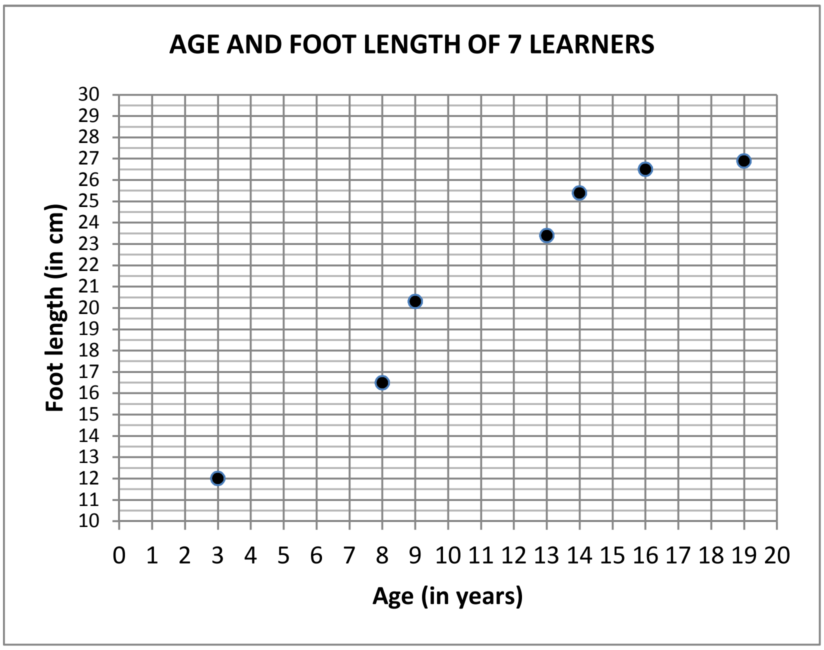
As part of the 2009 Census@School learners were asked the date of their birth and the length of their right foot without shoes on, correct to the nearest centimetre. The following table shows the data collected from 7 learners randomly selected from the data base.

Age (years)	3	8	9	13	14	16	19
Foot length (cm)	12	16,5	20,3	23,4	25,4	26,5	26,9

a) Draw a scatter plot to illustrate the data
b) Use the formula to determine the value of *r*.
c) Use the value of *r* to describe the type of linear correlation that exists between the age and the foot length.

SOLUTION:

a)



b)

Age (x)	Foot length (y)	$x\,y$	x^2	y^2
3	12	$3 \times 12 = 36$	$(3)^2 = 9$	$(12)^2 = 144$
8	16,5	$8 \times 16,5 = 132$	$(8)^2 = 64$	$(16,5)^2 = 272,25$
9	20,3	$9 \times 20,3 = 182,7$	$(9)^2 = 81$	$(20,3)^2 = 412,09$
13	23,4	$13 \times 23,4 = 304,2$	$(13)^2 = 169$	$(23,4)^2 = 547,56$
14	25,4	$14 \times 25,4 = 355,6$	$(14)^2 = 196$	$(25,4)^2 = 645,16$
16	26,5	$16 \times 26,5 = 424$	$(16)^2 = 256$	$(26,5)^2 = 702,25$
19	26,9	$19 \times 26,9 = 511,1$	$(19)^2 = 361$	$(26,9)^2 = 723,61$
$\sum x = 82$	$\sum y = 151$	$\sum xy = 1\,945,6$	$\sum x^2 = 1\,136$	$\sum y^2 = 3\,446,92$

<u>EXAMPLE 2</u> (continued)

With these sums and $n = 7$, the correlation coefficient is:

$$\boldsymbol{r} = \frac{\boldsymbol{n\sum xy-(\sum x)(\sum y)}}{\sqrt{\boldsymbol{n\sum x^2-(\sum x)^2}}\sqrt{\boldsymbol{n\sum y^2-(\sum y)^2}}}$$

$$= \frac{7\times 1\,945{,}6-(82)(151)}{\sqrt{(7\times 1\,136{,}0)-(82)^2}\times\sqrt{7\times 3\,446{,}92-(151)^2}}$$

$$= \frac{1\,237{,}2}{\sqrt{1\,228{,}0}\sqrt{1\,327{,}44}}$$

$= 0{,}96902\ldots$

$\approx 0{,}97$

c) $r \approx 0{,}97$ tells us that a strong positive linear correlation exists between age and foot length.

EXAMPLE 3

In the 2011 Household Survey, a representative sample of people was asked how many rooms they have in their homes. The table below shows the data taken from the Gauteng Province.

Number of rooms	1	2	3	4	5	6	7	8	9	10
Percentage of the households	20,6	13,9	10,9	17,9	13,1	9,8	6,1	3,8	2,6	1,3

Use your calculator to find the correlation coefficient ***r***

<u>SOLUTION</u>:

CASIO *fx*-82 *ZA*	SHARP EL-W535HT
• First get the calculator in STAT mode: [MODE] [2: STAT] [2: A + B*X*] • Enter the x-values: 1 [=] 2 [=] 3 [=] 4 [=] 5 [=] 6 [=] 7 [=] 8 [=] 9 [=] 10 [=] • Enter the *y*-values [▼] [►] 20,6 [=] 13,8 [=] 10,9 [=] 17,9 [=] 13,1 [=] 9,8 [=] 6,1 [=] 3,8 [=] 2,6 [=] 1,3 [=] [AC] • Find the value for ***r*** [SHIFT] [1:STAT] [5:Reg] [3: ***r***] [=] ***r* = – 0,917 952 1043**	• First get the calculator into STAT mode: <MODE> [1:STAT] [1:LINE] • Enter *x*-values and the *y*-values together: 1 [(*x*;*y*)] 20,6 [CHANGE] 2 [(*x*;*y*)] 13,9 [CHANGE] 3 [(*x*;*y*)] 10,9 [CHANGE] 4 [(*x*;*y*)] 17,9 [CHANGE] 5 [(*x*;*y*)] 12 [CHANGE] 6 [(*x*;*y*)] 9,7 [CHANGE] 7 [(*x*;*y*)] 6,1 [CHANGE] 8 [(*x*;*y*)] 3,8 [CHANGE] 9 [(*x*;*y*)] 2,3 [CHANGE] 10 [(*x*;*y*)] 10 [CHANGE] • Find the value for ***r***, the y-intercept [ALPHA] [÷] [=] ***r* = – 0,917 952 1043**

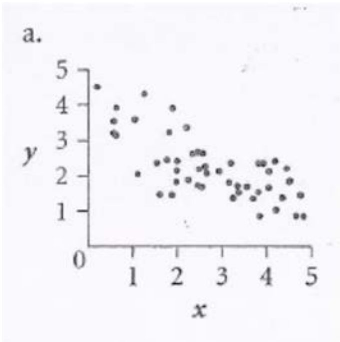
So $r = -\,0{,}917\ 952\ 1043 \approx -\,0{,}9$

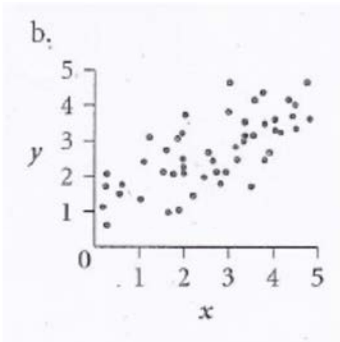
EXERCISE 3.2

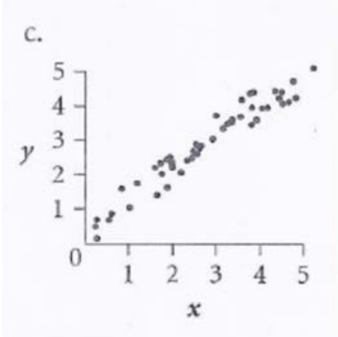
For decimal answers, round off answers to TWO decimal places.

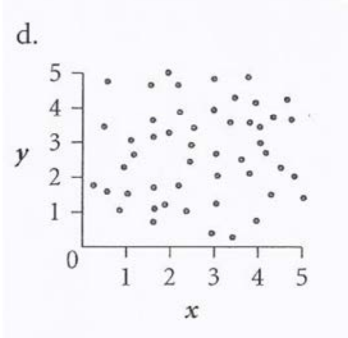
1) The scatter plots of paired data sets are shown below. Match the r values given with the scatter plots below.

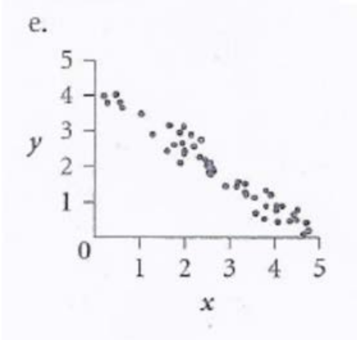
$r = -0{,}95$ $r = -0{,}5$ $r = 0$ $r = 0{,}5$ $r = 0{,}95$











2) Given below are the foot lengths (correct to the nearest cm) and heights (without their shoes on and correct to the nearest cm) of 10 learners in KZN recorded during the 2009 Census@School:

Foot length (cm)	22	19	24	20	23	27	24	24	26	25
Height (cm)	153	170	160	131	156	174	158	163	175	165

a) Draw a scatter plot to show the relationship between the foot length and the height of the learners.

b) Use the formula $r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$ to calculate the correlation coefficient. Explain what the value of r tells us.

EXERCISE 3.2 (continued)

3) The results from a group of randomly selected girls who participated in the 2009 Census@School. Each learner was asked to give her age and the grade she was in. The table below shows the results:

Age (years)	17	16	16	14	11	13	18	10	9	18	13
Grade	11	10	11	7	6	6	12	4	3	3	9

   a) Draw a scatter plot to illustrate the data
   b) Use your calculator to determine the correlation coefficient
   c) Use the value of r to describe the correlation that exists between the two variables.

4) The following table provides data on the age and corresponding height without shoes on of 12 learners selected from the 2009 Census@School results.

Age (in years)	8	19	11	12	17	13	15	16	9	18	10	14
Height (in cm)	128	180	150	157	166	160	165	173	140	178	130	150

   a) Draw a scatter plot to illustrate the data.
   b) Use your calculator to determine the correlation coefficient.
   c) Use r to describe the correlation that may exist between age and height.

5) The table shows the percentage of the households in Mpumalanga that had a land line telephone in the 2001, 2007, 2010 and 2011 Community Survey.

Year of survey (x)	2001	2007	2010	2011
Percentage of the households (y)	15	10	20	5

   a) Draw a scatter plot illustrate the given data.
   b) Calculate the correlation coefficient
   c) Describe the correlation that exists between the variables.
   d) Judging from the trend shown in the scatter plot, what would be the percentage of households with land line telephones by 2015? Explain your answer.

LINE OF BEST FIT

- ✓ If there is a strong linear correlation between the two aspects of the data, it may be possible to draw a line that best models the data. The line is called the **line of best fit** or **the regression line** or **the least squares regression line**.
- ✓ This line can be used to **predict/estimate** the value of one variable, given the value of the other variable.
- ✓ The slope of the line shows the **trend** of the points

a) Drawing an intuitive line of best fit

- ✓ A line of best fit can be drawn **intuitively or** 'by eye'. This line of best fit is *an approximation or estimate* of the line.
  - To draw a line of best fit, try to draw the line so that there is the same number of points above it as below it.
  - Outliers should clearly be ignored when fitting a line to the points.

EXAMPLE 4

In the 2009 Census@School survey of 15 to 19 year olds, learners were asked for their height without their shoes on and the length of their arm spans, both to the nearest centimetre. The table gives the data from five boys in the data base:

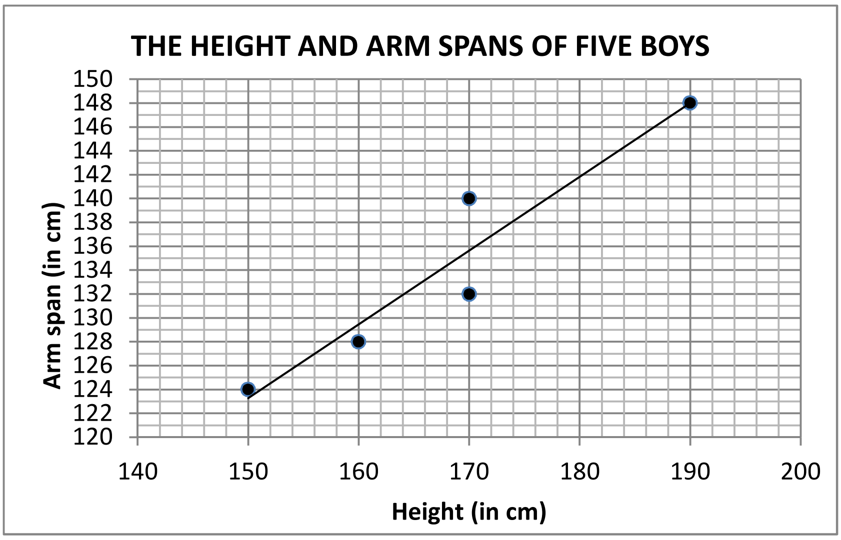
Height (cm)	150	170	190	160	170
Arm span (cm)	124	132	148	128	140

a) Use your calculator to determine the value of *r*.
b) Draw a scatter plot and use it to describe the correlation that exists between the height of boys and their arm spans.
c) Draw a line of best fit on the scatter plot. What relationship can you establish from the line of best fit?

<u>SOLUTION</u>:

a) $r = 0{,}8986... \approx 0{,}90$

b)



Both *r* and the graph indicate a strong positive linear correlation between height and arm span. Generally, as height increases, the arm span increases.

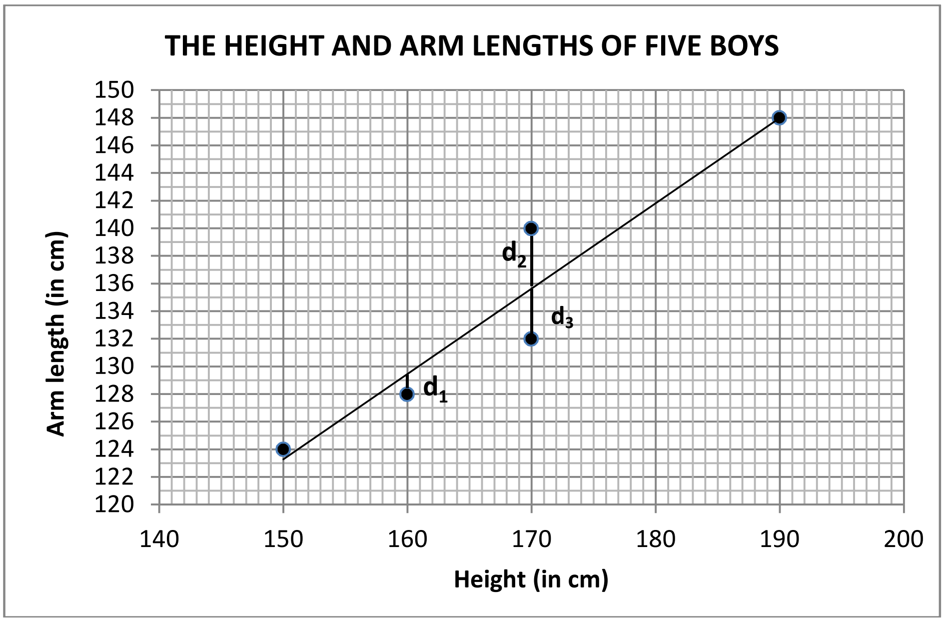
c) Note: one line of best fit has been drawn. Other ones are possible, but all should show a strong positive correlation.

b) Finding the Equation of the Regression Line

- ✓ Because your line of best-fit may not be the same as someone else's, it is helpful to have a systematic method that always gives the same result. One procedure commonly used is the **'method of least squares'**.

- ✓ The equation of the linear regression line is $\hat{y} = a + bx$ where $\hat{y}$ is ***'y-hat***, ***b*** is the gradient of the line and ***a*** is the cut on the *y*-axis (or the *y*-intercept).

- ✓ Consider the scatter plot and the line of best fit shown below



- Unless $r = 1$ or $r = -1$ (perfect positive or negative correlation), there will be no difference between the plotted points and the line of best fit. These differences are called ***residuals***.

- The residuals = d = (observed *y*-value) – (predicted *y*-value).

  The residuals can be positive, negative or zero.
  When the point is above the line, *d* is positive.
  When the point is below the line, *d* is negative
  When the point is on the line, $d = 0$.

- ✓ The aim of regression is to obtain an expression for the relationship that keeps the residuals as small as possible.

  The regression line, also called the line of best fit, is *the line for which the sum of the squares of the residuals is a minimum (i.e. as close to 0 as possible).*

✓ For the ***least squares regression line***, if $\bar{x}$ is the mean of the *x* values and $\bar{y}$ is the mean for the *y* values.

- The slope of the line is given by: $b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$
- The *y*-intercept is given by $a = \bar{y} - b\bar{x}$.

EXAMPLE 5:

The table below shows the masses and the heights of seven learners.

Mass (kg)	49	65	82	60	65	94	88
Height (cm)	156	176	183	153	163	192	180

a) Use your calculator to determine *r*, the correlation coefficient.
b) Use the method of least squares to determine the equation of the regression line.
c) Draw a scatter plot to illustrate the data.
d) Draw the regression line on the scatter plot.
e) Use your graph to determine Joan's height if her mass is 50 kg.

<u>**SOLUTION**</u>**:**

a) $r = 0{,}9048... \approx 0{,}90$

b) First calculate $\bar{x}$ and $\bar{y}$, then find $\sum(x-\bar{x})(y-\bar{y})$ and $\sum(x-\bar{x})^2$.
It helps to draw up a table and to fill everything onto it:

mass	**height**				
x	***y***	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
49	156	–22,86	–15,86	362,56	522,58
65	176	–6,86	4,14	–28,4	47,06
82	183	10,14	11,14	112,96	102,82
60	153	–11,86	–18,86	223,68	140,66
65	163	–6,86	–8,86	60,78	47,06
94	192	22,14	20,14	445,9	490,18
88	180	16,14	8,14	131,38	260,5
$\bar{x} = \frac{503}{7}$ = 71,86	$\bar{y} = \frac{1\,203}{7}$ = 171,86	$\sum(x-\bar{x})$ = 0	$\sum(y-\bar{y})$ = 0	$\sum(x-\bar{x})(y-\bar{y})$ = 1 308,86	$\sum(x-\bar{x})^2$ = 1 610,86

Use the formula for *b* to get the slope (or gradient) of the regression line:

$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{1\,308{,}86}{1\,610{,}86} = 0{,}812\,541\,1271... \approx 0{,}81$$

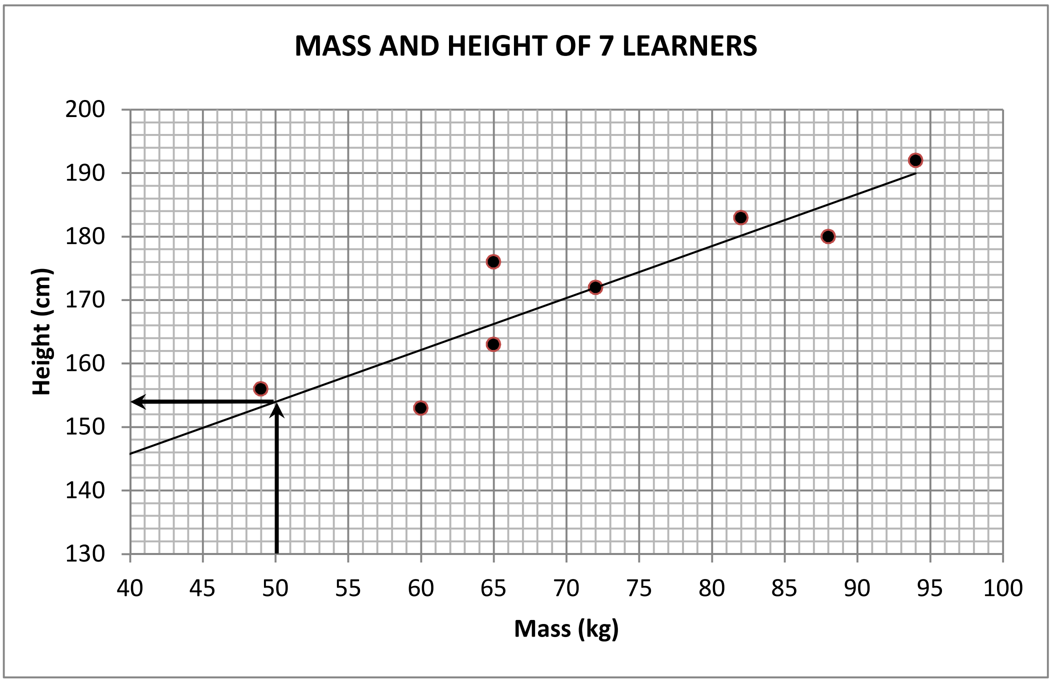
Substitute ($\bar{x}$; $\bar{y}$) and *b* into $a = \bar{y} - bx$

a = 171,86 – 71,86 (0,812 541 1271)
= 113,4707946
≈ 113,47

The equation of the line of regression is therefore: $\hat{y}$ = 113,47 + 0,81 *x*

<u>EXAMPLE 5</u> (continued)

c)



d) In order to draw the regression line, substitute any two x-values that lie between the minimum and maximum x-values into the equation of the regression line, plot the two points and then join them up.
When $x = 45$, $\hat{y} = 113{,}47 + 0{,}81(45) = 149{,}92 \approx 150$
When $x = 94$, $\hat{y} = 113{,}47 + 0{,}81(94) = 189{,}61 \approx 190$
So, to draw the regression line, plot the points (45; 150) and (94; 190) and join them up.

e) The graph shows that at 50 kg, the height is 154 cm.

The same value would be obtained using the equation of the regression line:
When $x = 50$ $\hat{y} = 113{,}47 + 0{,}81(50) = 153{,}97 \approx 154$

c) Using a Least Squares Regression Line to Make Predictions

- ✓ When a value for one of the variables that was not originally in the data is found, you are making a prediction.

- ✓ The required value can be read off from the scatter plot or by using the equation of the regression line. Predictions made from the equation of the line can be made through the process of ***interpolation*** and ***extrapolation***.
  - **Interpolation** is a method of predicting/estimating new data value(s) **within** the known range of data values.
  - **Extrapolation** on the other hand is a method of estimating new data value(s) **beyond** a discrete set of known data values.

- ✓ ***Note that*** data values that are the result of extrapolation from statistical data are often less valid than those that are the result of interpolation. This is because the values are often estimated outside the tabulated or observed range of data.

EXAMPLE 6

a) Use the equation of the regression line $\hat{y} = 113{,}47 + 0{,}81x$ from Example 5 to estimate the height in each of the cases:
   i) 81 kg   ii) 52 kg   iii) 20 kg

b) Give reasons for considering each of the predictions in (a) to be good or not good.

<u>**SOLUTION**</u>**:**

a) $\hat{y} = 113{,}47 + 0{,}81\,x$
   i) When $x = 81$ kg, $\hat{y} = 113{,}47 + 0{,}81\,(81) = 179{,}08 \approx 180$ cm
   ii) When $x = 52$ kg, $\hat{y} = 113{,}47 + 0{,}81\,(52) = 155{,}59 \approx 156$ cm
   iii) When $x = 20$ kg, $\hat{y} = 113{,}47 + 0{,}81\,(20) = 129{,}67 \approx 130$ cm

b) The predictions in (i) and (ii) are good because they are within the data range, hence the resulting height is within the range. The prediction (iii) is not as good because it is outside the domain of the original data of 49 to 88 kg.

EXAMPLE 7

Use your calculator to find the equation of the regression line for the following set of data:

Mass (kg)	49	65	82	60	65	94	88
Height (cm)	156	176	183	153	163	192	180

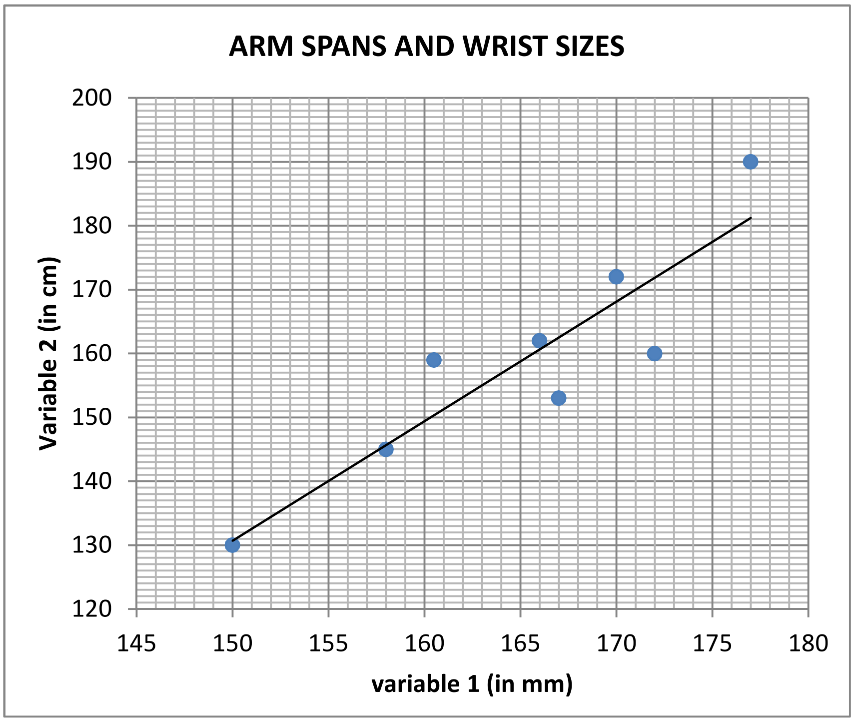
<u>SOLUTION</u>:

CASIO *fx*-82 *ZA*	SHARP EL-W535HT
• First get the calculator in STAT mode: [MODE] [2: STAT] [2: A + B*X*] • Enter the x-values: 49 [=] 65 [=] 82 [=] 60 [=] 65 [=] 94 [=] 88 [=] • Enter the *y*-values [▼] [▶] 156 [=] 176 [=] 183 [=] 153 [=] 163 [=] 192 [=] 180 [=] [AC] • Find the value for ***a***, the y-intercept [SHIFT] [1:STAT] [5:Reg] [1: A] [=] ***a* = 113, 4716211 ≈ 113,47.** • Get the value for ***b***, the gradient [SHIFT][1:STAT] [5:Reg] [2: B] [=] ***b* = 0,812522171 ≈ 0,81**	• First get the calculator in STAT mode: <MODE>[1:STAT] [1:LINE] • Enter *x*-values and the *y*-values together: 49 [(*x;y*)] 156 [CHANGE] 65 [(*x;y*)] 176 [CHANGE] 82 [(*x;y*)] 183 [CHANGE] 60 [(*x;y*)] 153 [CHANGE] 65 [(*x;y*)] 163 [CHANGE] 94 [(*x;y*)] 192 [CHANGE] 88 [(*x;y*)] 180 [CHANGE] • Find the value for ***a***, the y-intercept [ALPHA] [( ] [=] ***a* = 113, 4716211 ≈ 113,47.** • Find the value for ***b***, the gradient [ALPHA] [)] [=] ***b* = 0,812522171 ≈ 0,81**

So the equation of the line of best-fit is: $\hat{y} = 113{,}47 + 0{,}81\ x$ (the same equation as before).

EXERCISE 3.3

1) The scatter plot below shows the correlation between the arm span (correct to the nearest cm) and a wrist size (correct to the nearest mm) of a group of boys.



a) Identify the two variables.
b) List the coordinates of the points on the scatter plot.
c) Use your calculator to determine the correlation coefficient and describe the correlation that exists between the two units.
d) Use the formulae $b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$ and $a = \bar{y} - b\bar{x}$ to determine the equation of the linear regression line.

2) In the 2011 Household Survey, people were asked how many rooms they have in their homes. The table below shows the data taken from the North West Province:

Number of rooms per household	1	2	3	4	5	6	7	8	9
Percentage of the households	15	14	13,6	19,4	12,6	11,4	6,7	3,4	1,8

a) Draw a scatter plot to display the data
b) Calculate the value of r, the correlation coefficient and use it to describe the correlation that exists between the two variables
c) Use the calculator to determine the equation of the regression line. Give your answer to TWO decimal places.
d) Draw the regression line onto your scatter plot.

EXERCISE 3.3 (continued)

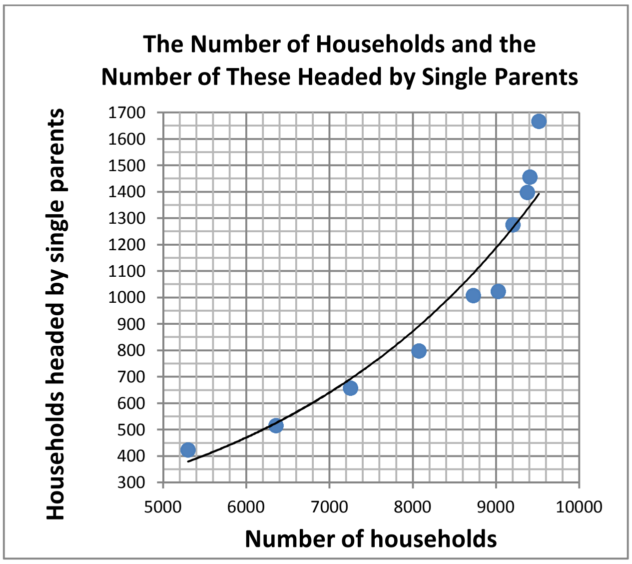
3) The table below shows the age and foot length of the right foot for a group of 10 boys as recorded in the 2009 Census@School.

Age (in years)	12	14	13	11	16	19	15	18	11	19
Foot length (in cm)	23,4	25,4	24,8	22,9	26,5	28	24	27	22,9	30

a) Draw a scatter plot for the given data.
b) Calculate the value of r, the correlation coefficient and use it to describe the correlation that exists between the two variables
c) Calculate the equation of the regression line. Give your answer to TWO decimal places.
d) Use your regression line to determine:
   i) The foot length of a boy 9 years old.
   ii) The foot length of a boy who is 17 years and 6 months old.
e) State whether you think each of the predictions in (c) is good or not. Give reasons for your answer.

4) The table and scatter plot below shows information taken from the 2011 Census. They show the number of households surveyed in ten areas in KwaZulu-Natal and the number of these households that are headed by single parents.

Total number of households	Number of these households headed by single parents
5 302	423
6 359	516
7 256	657
8 077	798
8 729	1 008
9 028	1 023
9 205	1 275
9 379	1 398
9 412	1 456
9 517	1 667



As you can see, an exponential regression function best fits the data.

a) Use your calculator to determine the equation of the exponential regression function $y = \text{A.B}^x$ as follows:

CASIO *fx-82ZA PLUS*	SHARP EL-W535HT
[MODE] [2 : STAT] [6 : A.B^X] *Now enter the data in the normal way and find A and B as follows:* [AC] [SHIFT] [STAT] [5 : Reg] [1 : A]	[2ndF] [CA} <MODE> [1 : STAT] [3 : E_EXP] *Now enter the data in the normal way and find A and B as follows:* [ALPHA] [( ] (a) [=]

	[=] [SHIFT] [STAT] [5 : Reg] [2 : B] [=]	[ALPHA] [ )] (b) [=]

<u>EXERCISE 3.3</u> (continued)

b) Use your calculator to predict the number of households (correct to the nearest unit) headed by single parents if there are 15 000 households.

CASIO *fx-82ZA PLUS*	SHARP EL-W535HT
The data remains in the calculator Enter the value: 15 000 Then press: [SHIFT] [STAT] [5 : REG] [5 : $\hat{y}$] [=]	*The data remains in the calculator.* Enter the value: 15 000 Then press: [2ndF] [ )] (*y'*)

c) Would you have confidence in this prediction? Justify your answer.

REFERENCES

Bluman, A. (2004) *Elementary Statistics*. MacGraw Hill
Buckwell, G., (1997) *Mastering Mathematics*. Macmillan Master Series
Carter P. Et al (2007): *FOCUS in Mathematics*, Maskew Miller Longman (Pty) Ltd
Haighton, J. et al(2003) *Use of Maths Statistics*. Nelson Thornes Ltd.
Statistics South Africa (2010). *Census At School Results 2009*.
Watkins E., et al (2004) *Statistics in Action*. Von Hoffmann Press

4

Grade 10 Probability

In this chapter you:
- Revise the language of probability
- Calculate theoretical probability of events happening
- Calculate the relative frequency of events happening
- Draw and interpret Venn diagrams
- Use Venn diagrams to determine the probability of events happening
- Define mutually exclusive events
- Use the addition rule for probability and the complementary rule to determine probabilities.

WHAT YOU LEARNED ABOUT PROBABILITY IN GRADE 9

In Grade 9 you covered the following probability concepts for situations with equally probable outcomes:
- *Determining probabilities for compound events using two-way tables and tree diagrams*
- *Determining the probabilities for outcomes of events and predicting their relative frequency in simple experiments*
- *Comparing relative frequency with probability and explaining possible differences*

THE LANGUAGE OF PROBABILITY

✓ In this Study Guide we will use the term ***dice*** for both one dice or many dice.

One dice

Two dice

a) What is Probability?

- ✓ Probability is a branch of mathematics that deals with calculating how likely it is that a given event occurs or happens. Probability is expressed as a number between 1 and 0. The words ***chance*** or ***likelihood*** are often used in place of the word probability.
  - Tossing a coin is an ***activity*** or ***experiment***.
    If both Heads (H) and Tails (T) have an equal chance of landing face up, it is called a ***fair coin***

  - Throwing a dice is an ***activity*** or ***experiment***.
    If each number on the dice has an equal chance of landing face up, it is called a ***fair dice***.

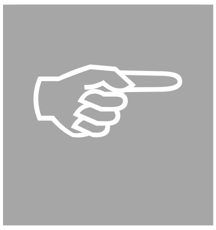
- ✓ When we talk about the probability of something happening, we call the something an **event**
  - Getting tails when tossing a coin is an **event**.

b) Listing Outcomes

- ✓ For any activity or probability experiment you can usually list all the ***outcomes***.

  The set of ***all*** possible outcomes of a probability experiment is a ***sample space***.

  An ***event*** consists of one or more outcomes and is a subset of the sample space. Outcomes of the event you are interested in are called the ***favourable outcomes*** for that event.



EXAMPLE 1
List the ***sample space, event*** and ***favourable outcomes*** of the following probability experiments:
a) Throw a dice and get a 6
b) Throw a dice and get an even number
c) Toss a coin and get a head (H)

<u>SOLUTION:</u>

a) The **activity** is *throw a dice*
The **sample space** is 1; 2; 3; 4; 5 and 6
The **event** you are interested in is ***get a 6***
The **favourable outcome** is 6.

b) The **activity** is *throw a dice*
The **sample space** is 1; 2; 3; 4; 5 and 6
The **event** you are interested in is ***get an even number***
The **favourable outcomes** are 2; 4 and 6.

<u>EXAMPLE 1</u> (continued)

c) The **activity** is *toss a coin*
The **sample space** is heads (H) and tails (T)
The **event** you are interested in is ***get a head*** (H).
The **favourable outcome** is H.

c) Probability Scales

✓ Some events ***always happen.*** We say that they are **certain** to happen and give them a probability of 1.

It is **certain** that the day after Monday is Tuesday
The probability that the day after Monday is Tuesday is 1.

Sunday
Monday
Tuesday

✓ Some events ***never happen.*** We say that they are **impossible** and give them a probability of 0.

If you throw an ordinary dice, it is **impossible** to get a 7.
The probability of getting a 7 when you throw an ordinary dice is 0.

✓ Some events are ***not certain,*** but are ***not impossible*** either. They may or may not happen. These probabilities lie between 0 and 1.

If you toss a fair coin it may land on heads or it may not.
The chances are **equally likely.**
We say that there is a **50-50 chance** that it will land on heads.

✓ We can write probabilities **in words** or as **common fractions**, **decimal fractions** or **percentages**.

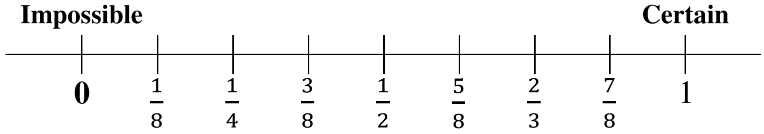
The following number line shows **words**:

0		**½**		**1**
Impossible	**Unlikely**	**Equally likely**	**Likely**	**Certain**

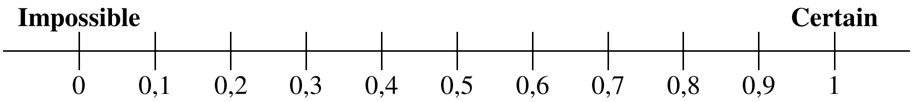
To **compare probabilities**, we compare the sizes of the fractions, decimal fractions or percentages.

- The *less likely* an event is to happen, the *smaller* the fraction, decimal fraction or percentage.
- The *more likely* an event is to happen, the *larger* the fraction, decimal fraction or percentage.

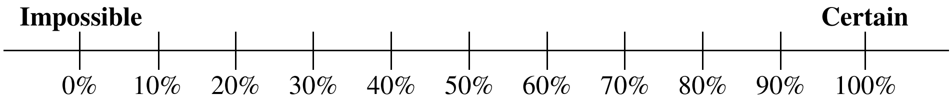
The following number line shows **common fractions**:



The following number line shows **decimal fractions**:



The following number line shows **percentages**:



Remember that 100% = 100 ÷ 100 = 1

CALCULATING PROBABILITY

- ✓ The method you use to calculate probabilities depends on the type of probability you are dealing with. We can find ***theoretical probability*** (also called ***actual probability***) and ***relative frequency*** (also called ***experimental probability***).

- ✓ The probability that Event E will occur is written ***P(E)*** and is read ***the probability of Event E occurring***. The same terminology is used for both theoretical probability and relative frequency.

a) Theoretical Probability (or Actual Probability)

- ✓ Theoretical probability is used when each outcome in a sample space is equally likely to occur.

- ✓ The theoretical probability for an Event E is given by:

Probability of Event E happening

$$= \frac{\textbf{\textit{number of outcomes for Event E}}}{\textbf{\textit{total number of possible outcomes in the sample space}}}$$

EXAMPLE 2
Calculate the probability of getting a head (H) when a fair coin is tossed.
Write the answer as a fraction in simplest form, as a decimal and as a percentage.

<u>SOLUTION</u>:
Because this is a fair coin, each outcome is equally likely to occur, so we can find the theoretical probability.

The **event** is *getting a head* (H).
The **possible outcomes** or **sample space (S)** are heads and tails (*H and T*).
The **total number of possible outcomes in the sample space = *n(S)*** = 2.

The **favourable outcome** is *heads (H)*.
The **number of favourable outcomes = *n(H)*** = 1.

We use the formula:

$$\boldsymbol{Probability\ of\ an\ event\ happening} = \frac{\boldsymbol{number\ of\ favourable\ outcomes\ for\ that\ event}}{\boldsymbol{total\ number\ of\ possible\ outcomes\ in\ the\ sample\ space}}$$

$$= \frac{\boldsymbol{number\ of\ heads}}{\boldsymbol{total\ number\ of\ possible\ outcomes\ in\ the\ sample\ space}}$$

$$P(H) = \frac{n(H)}{n(S)}$$

$$= \frac{1}{2}$$

$$= 0{,}5$$

$$= 50\%$$

NOTE:

- If P(E) stands for the probability of event E occurring then $\mathbf{0 \leq P(E) \leq 1.}$
- In other words, the probability of event E occurring is a rational number from 0 up to and including 1.

EXAMPLE 3
A regular six-sided fair dice is thrown once.
a) List the sample set.
b) How many elements are there in the sample set?
c) List all the favourable outcomes for getting a score of 3 or more.
d) How many favourable outcomes are there?
e) What is the probability of getting a score of 3 or more? Give your answer as a fraction in simplest form, as a decimal and as a percentage both correct to 2 decimal places.

- *You can make your own 6 - sided dice using the net on the last page of this chapter.*

<u>SOLUTION</u>:
There are 6 numbers on a dice and each number has an equal chance of landing face up. Because this is a fair dice, each outcome is equally likely to occur, so we can find the theoretical probability (also just called ***probability***).

a) The ***sample set***, S, is 1, 2, 3, 4, 5 and 6 or {1; 2; 3; 4; 5; 6}

b) n(S) = 6

c) The *favourable outcomes* for this event are the numbers that are 3 or more.
So the *favourable outcomes* are 3; 4; 5; 6 or {3; 4; 5; 6}

d) n(3 or more) = 4

e) $Probability\ of\ an\ event\ happening = \frac{number\ of\ favourable\ outcomes\ for\ that\ event}{total\ number\ of\ possible\ outcomes\ in\ the\ sample\ set}$

$$\begin{aligned} P(3\ or\ more) &= \frac{n(3\ or\ more)}{n(S)} \\ &= \frac{4}{6} \\ &= \frac{2}{3} \\ &= 0{,}66666\ldots \\ &\approx 0{,}67 \\ &\approx 66{,}67\% \end{aligned}$$

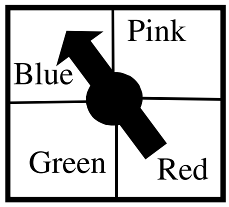
EXERCISE 4.1

Give each of the answers in this exercise

- i) as a common fraction in simplest form,
- ii) as a decimal fraction (correct to 2 decimal places)
- iii) as a percentage (correct to 1 decimal place).

1) A fair dice is rolled once.
   a) List the elements of the sample space.
   b) What is the probability that you will get
      i) A six?
      ii) An odd number?
      iii) A seven?
      iv) More than 2?
      v) Less than 10?

2) The spinner alongside is spun.
   a) List the elements of the sample space
   b) What is the probability of the spinner (arrow) landing on
      i) Green?
      ii) Yellow?



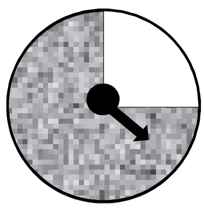
3) Each letter of the word **MATHEMATICS** is written on a separate piece of paper of the same shape and size and put in a box. Nomsa closes her eyes and takes one piece of paper out of the bag at random.
   a) List the elements of the sample space
   b) What is the probability that she takes a piece of paper with:
      i) An M on it?
      ii) A vowel on it?

> ***At random*** means you choose without method or without thinking about your choice.

4) Six counters in a bag are numbered 3 4 7 9 10 11.
   One counter is drawn at random from the bag.
   a) What does the sample space consist of?
   b) Calculate the probability that the number drawn is
      i) An odd number
      ii) A prime number
      iii) A square number

5) A learner is chosen at random from a group of 18 boys and 12 girls.
   a) Determine $n(S)$ where S is the sample space.
   b) What is the probability that this learner is
      i) A boy?
      ii) A girl?

6) The spinner alongside is spun.
   a) Determine $n(S)$ where S is the sample space.
   b) Calculate the probability of the spinner landing on the shaded area.



b) Relative Frequency (or Experimental Probability)

✓ Sometimes we *calculate* probability and sometimes we *estimate* probability.

- Probability that is *calculated* is called ***theoretical probability*** or just ***probability***.
- Probability that is *estimated* is calculated after performing a very large number of trials of an experiment or conducting a survey involving a very large number of items, and is called ***relative frequency***.

Examples of experiments that can be used to calculate ***relative frequency:***

i) Tossing a coin 500 times and counting the number of times it lands on heads.

ii) Throwing a dice 200 times and counting the number of times it lands on an even number.

iii) Repeatedly taking a counter out of a bag containing ten counters numbered from 34 to 43. Recording the number on the counter, replacing the counter into the bag, and seeing how many times you get a multiple of 3 after taking out and returning a counter 1 000 times.

✓ When an experiment is repeated over and over, the relative frequency of an event approaches the theoretical or actual probability of the event.

- If you want to *estimate the probability* of an event by using an *experiment*, you need to perform *a very large number of trials* as a pattern often does not become clear until you observe a ***large*** number of trials
- If you want to *estimate the probability* of an event by using *the results of a* survey, the survey should involve a large number of items as a pattern often does not become clear until you observe a ***large*** number of items

✓ Relative frequency can be used even if each outcome of an event is not equally likely to occur.

✓ We find relative frequency using the following formula:

$$\textbf{\textit{Relative frequency of an event happening}} = \frac{\textit{number of times the event happens}}{\textit{total number of trials in the experiment}}$$

✓ Relative frequency is a fraction of the occurrences.

Like probability, $0 \leq$ ***relative frequency*** ≤ 1

✓ The results of an experiment or of a survey are often shown in a table as in the following example.

EXAMPLE 4

In the 2009 Census@School, learners were asked how they usually travel to school in the morning. The following table shows the responses from 443 590 learners who live less than 1 km from the school.

How they get to school	Number of learners
Walk/foot	399 220
Car	28 555
Train	782
Bus	3 052
Bicycle	1 415
Scooter	296
Taxi	10 270
TOTAL	**443 590**

a) Which are the 3 most popular ways of getting to school?
b) Determine $n(S)$ where S is the sample set.
c) Estimate the probability (as a percentage correct to 2 decimal places) that one of these learners selected at random
   i) walks to school
   ii) comes to school by helicopter
   iii) comes to school by car and taxi
   iv) comes to school by car or taxi

SOLUTION:

a) The 3 most popular ways of getting to school are walk/foot, car and taxi.

b) 443 590 learners responded so $n(S) = 443\ 590$

c)

i) P(walks to school) = $\frac{n(walk\ to\ school)}{n(S)} = \frac{399\ 220}{443\ 590} \approx 90{,}00\%$

ii) P(comes to school by helicopter)

$$= \frac{n(come\ to\ school\ by\ helicopter)}{n(S)} = \frac{0}{443\ 590} = 0{,}00\%$$

iii) Nobody comes to school by car AND by taxi.
P(comes to school by car AND by taxi)

$$= \frac{n(come\ to\ school\ by\ car\ AND\ by\ taxi)}{n(S)} = \frac{0}{443\ 590} = 0{,}00\%$$

iv) 28 555 learners come by car and 10 270 learners come by taxi.
n(come to school by car OR by taxi) = 28 555 + 10 270 = 38 825.
P(comes to school by car or by taxi)

$$= \frac{n(come\ to\ school\ by\ car\ or\ by\ taxi)}{n(S)} = \frac{38\ 825}{443\ 590} \approx 8{,}75\%$$

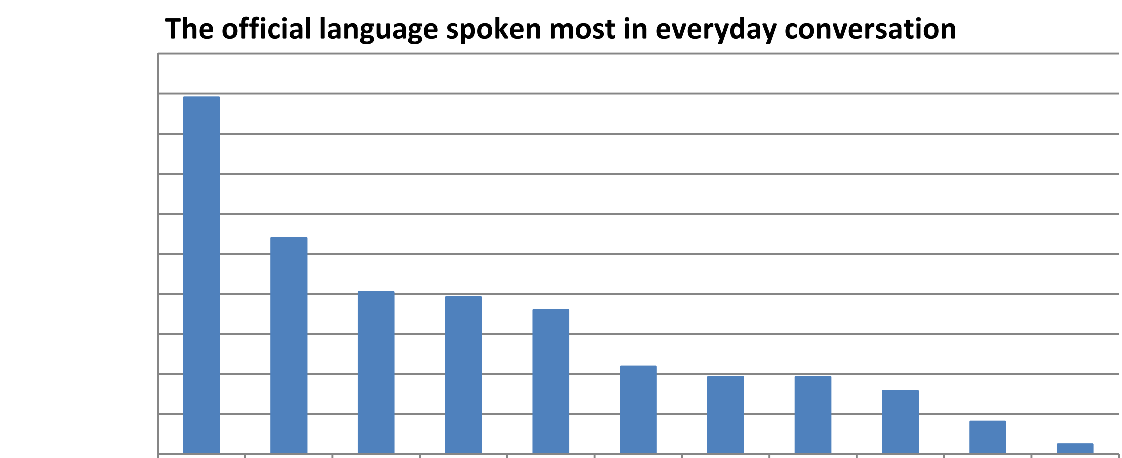
EXERCISE 4.2

1) This information comes from the table in the previous example giving the responses from 443 590 learners who live less than 1 km from the school.

How they get to school	Number of learners
Bus	3 052
Bicycle	1 415

a) Determine $n(S)$ if S is the sample set
b) Estimate the probability (as a percentage correct to 2 decimal places) that one of these learners selected at random from the sample
   i) comes to school by bus
   ii) comes to school by bicycle
   iii) comes to school by bus or bicycle
c) You should find the probabilities in b) low. Why do you think this is so?

2) The bar graph below is taken from 2009 Census@School. A sample of all the learners in South Africa was asked which of the official languages they spoke most in everyday conversation. (*The language used in everyday conversation is the language you use most of the time when talking and listening to others*). The bar graph shows their answers.



	English	isiZulu	Afrikaans	Setswana	Sesotho	isiXhosa	Sepedi	siSwati	Xitsonga	isiNdebele	Tshivenda
No: of learners	445492	270062	202570	196259	180020	109596	96538	96538	79137	40961	12310

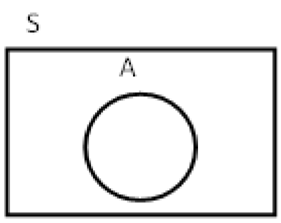
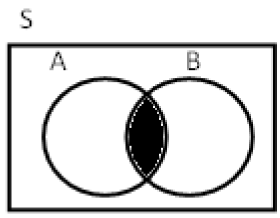
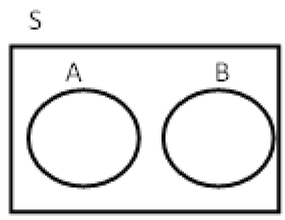
a) How many learners were surveyed?
b) Estimate the probability (as a percentage correct to 1 decimal place) that a learner selected at random from the sample
   i) Speaks mainly English in everyday conversation
   ii) Speaks mainly isiZulu OR Afrikaans in everyday conversation
c) In Census 2011 it was found that in South Africa, with a population of 51 770 560, 9,6% speak English and 36,2% speak isiZulu or Afrikaans in everyday conversation. Which results, 2009 Census@School or Census 2011, give better estimates? Give reasons for your answer.

VENN DIAGRAMS

- ✓ ***Venn diagrams*** were introduced in 1880 by John Venn as a way of picturing relationships between different groups of items.
- ✓ Venn diagrams use overlapping circles or closed curves within an enclosing rectangle to represent the items that are common to the groups of items.
- ✓ You can use Venn diagrams to help you work out the probability of an event occurring.

a) Drawing Venn diagrams

- ✓ We generally use a rectangle to represent a ***sample space (S)***. However, any closed shape could be used.
- ✓ The circles represent ***events*** within the sample space. However any closed shape could be used.

<table>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>This Venn diagram shows Event A in sample space S.</td>
<td>This Venn diagram shows Events A and B which have common values.

The part where they overlap is called the intersection of A and B.

The section that is shaded belongs to Event A and to Event B.</td>
<td>This Venn diagrams shows Events A and B which have no common values.

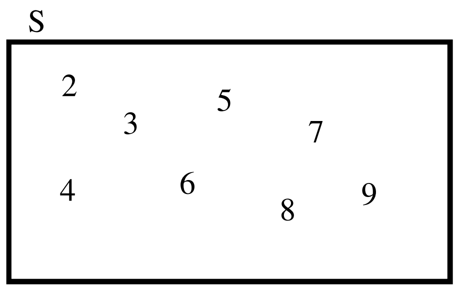
A and B are called disjoint sets.</td>
</tr>
</table>

EXAMPLE 5
Sample space S consists of the whole numbers from 2 to 9 inclusive.
In other words, S = {2; 3; 4; 5; 6; 7; 8; 9}
Event P consists of multiples of 3 in S. So P = {3; 6; 9}
Event Q consists of factors of 6 in S. So Q = {2; 3; 6}
Event R consists of multiples of 4 in S. So R = {4; 8}
Draw a Venn diagram to show each of the following:

a) The sample space S
b) Event P in the sample space S
c) Events P and Q in the sample space S
d) Events P and R in the sample space S.

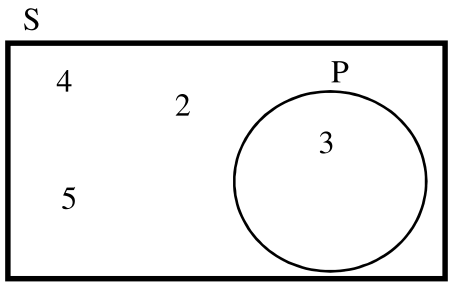
<u>SOLUTION</u>:

a)



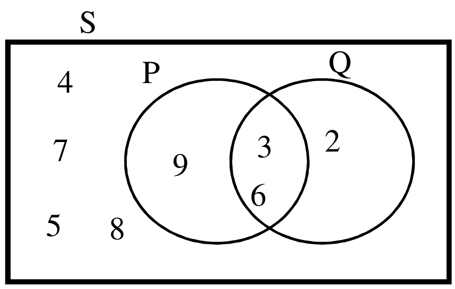
- *The sample space is labelled S.*
- *Values may be written in any order.*

b)



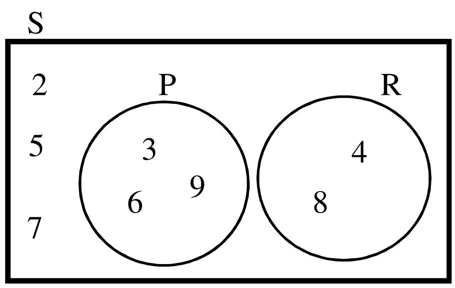
- *Draw a circle inside S.*
- *Label the circle P and write the outcomes of event P inside it.*
- *The values outside of P are those values in the sample space that are not multiples of 3.*

c)



- *P = {3; 6; 9} and Q = {2; 3; 6}. 3 and 6 belong to both sets. So draw two intersecting circles, P and Q.*
- *The outcomes common to events P and Q are written in the intersection.*
- *The values outside of P and Q are those values in the sample space that are not multiples of 3 and are not factors of 6.*

d)



- *P = {3; 6; 9} and R = {4; 8}. These two events have no common values so draw 2 separate circles P and R.*
- *Write 3, 6 and 9 in P and 4 and 8 in R.*
- *The values outside of P and R are those values in the sample space that are not multiples of 3 and are not multiples of 4.*

- Notice that each Venn diagram shows ***all*** the values in the sample space.

EXERCISE 4.3

1) A sample space S consists of whole numbers from 20 to 29 inclusive.
   Event A consists of the multiples of 4 in S.
   Event B consists of the factors of 420 in S.
   Event C consists of the multiples of 5 in S.
   Event D consists of the multiples of 3 in S.
   a) List the elements in
      i) S   ii) A   iii) B   iv) C   v) D
   b) Draw Venn diagrams to show
      i) Sample space S and event A
      ii) Sample space S, event A and event B
      iii) Sample space S, event C and event D

2) A fair eight-sided dice is rolled. [*You can make your own 8-sided dice using the net given on the last page of this chapter*].
   Event P is scoring a prime number.
   Event E is scoring a multiple of 2
   Event F is scoring more than 3
   Event G is scoring an even number
   Event H is scoring an odd number.
   a) List S, the possible outcomes of throwing a fair 8-sided dice.
   b) List the elements in
      i) P   ii) E   iii) F
      iv) G   v) H
      vi) The intersection of E and F
      vii) The intersection of G and H
   c) Draw Venn diagrams to illustrate each of the following in sample space S:
      i) Event P
      ii) Events E and F
      iii) Events G and H.

3) A sample space S consists of the letters of the alphabet.
   Event P consists of the letters of the word PROBABILITY.
   Event M consists of the letters of the word MATHEMATICS
   a) Which letters are common to the words PROBABILITY and MATHEMATICS?
   b) Draw a Venn diagram to show the information.
   Hint: If a letter in Event P or Event M occurs more than once you must write it down each time it occurs.

b) Interpreting Venn Diagrams

- ✓ ***A and B*** is the overlap or ***intersection*** of two sets or events. We sometimes use the symbol ∩ to show the intersection between two sets or events.

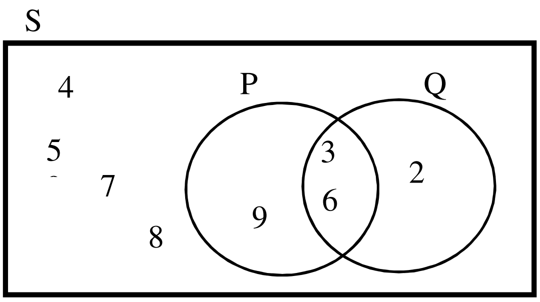
- ✓ ***A or B*** is the combination or ***union*** of the events or sets. We sometimes use the symbol ∪ to show the union between two sets or events.

- ✓ ***Neither A nor B*** could also be called ***not in A or B*** or ***not(A or B)***.

- ✓ ***A only*** excludes the intersection and ***B only*** excludes the intersection.

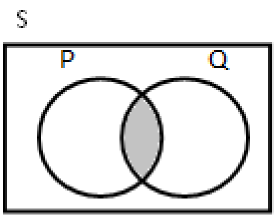
- ✓ ***A*** includes the intersection and ***B*** includes the intersection.

EXAMPLE 6
Use the Venn diagram to list the values which lie in
a) P and Q *(where the two sets overlap or intersect)*
b) P or Q *(the combination of the two sets, their union)*
c) P only *(in P but not in Q)*
d) Q only *(in Q but not in P)*
e) Neither P nor Q *(not in P or Q)*

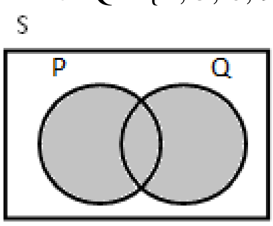


<u>SOLUTION</u>:
The values in P are {3; 6; 9} and the values in Q are {2; 3 6}

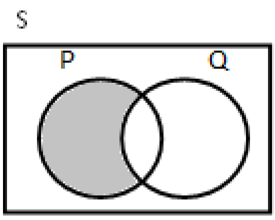
a) Values in P and Q = {3; 6}



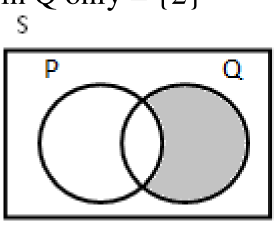
b) Values in P or Q = {2; 3; 6; 9}



c) Values in P only = {9}

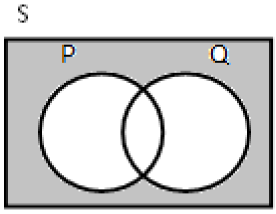


d) Values in Q only = {2}



EXAMPLE 6 (continued)

e) Values in neither P nor Q = {4; 5; 7; 8}

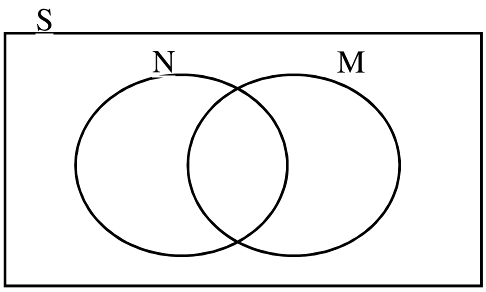


EXERCISE 4.4

Draw six Venn diagrams like the one given.

On each one shade one of the following:

1) N
2) N *and* M
3) N *or* M
4) N but not M
5) M but not N
6) Neither M nor N



c) Venn Diagrams Showing the Number of Outcomes in Events

✓ Sometimes there are too many outcomes to list in the Venn diagram. When this happens you write the ***number*** of outcomes in the Venn diagram.

EXAMPLE 7

150 people were asked which type of movie they like to watch.

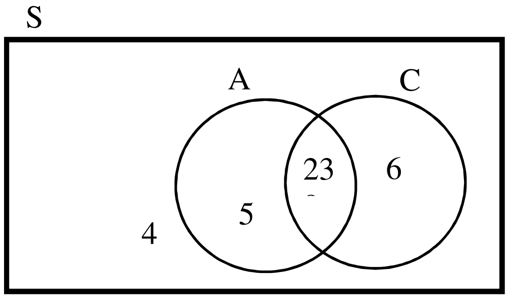
- 80 said they liked Action (A)
- 55 said they liked Comedy (C)
- 23 said they liked both.

Draw a Venn diagram to illustrate this information.

<u>SOLUTION</u>:

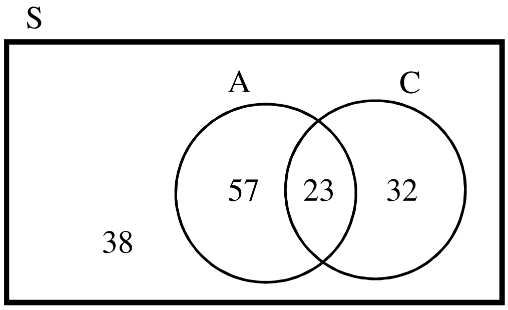
<u>Step1:</u>

- Because there are some people who like Action (A) *and* Comedy (C), we draw 2 intersecting circles in the sample space.
- Fill in 23 where the circles overlap (*in the intersection of A and B*) because 23 people like **both Action and Comedy**.



<u>Step 2:</u>

- The number of people who like ***A only*** $= 80 - 23 = 57$.
- The number of people who like ***C only*** $= 55 - 23 = 32$.
- The number of people who like ***neither*** $= 150 - (57 + 23 + 32) = 38$.
- Fill these values in on the Venn diagram.



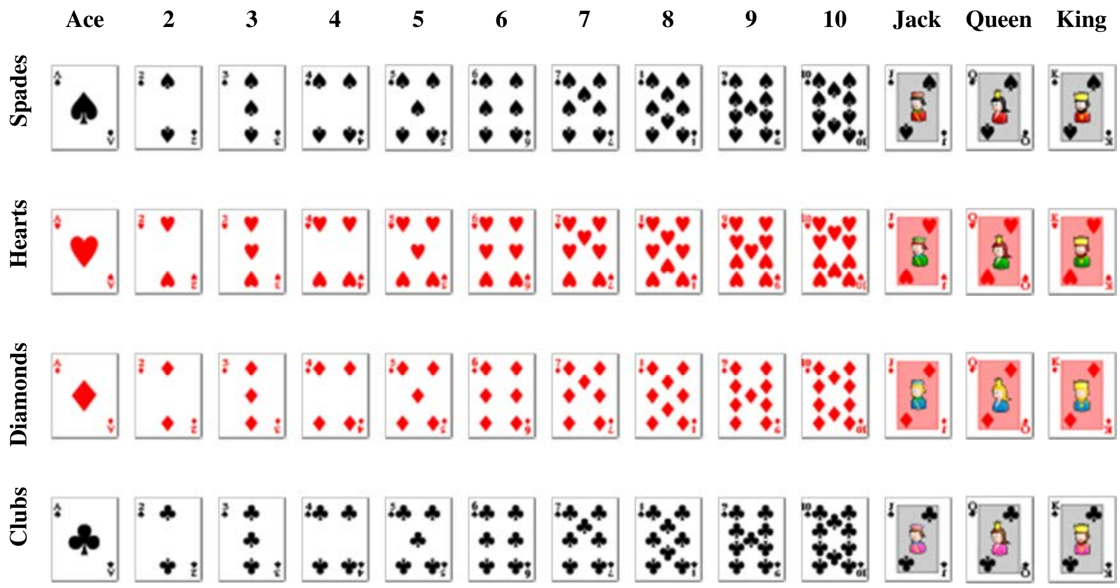
<u>Step 3:</u>

Check:

- o Are there 80 people altogether in A?
- o Are there 55 altogether in C?
- o Does 38 + 57 + 23 + 32 give 150?

✓ We often use playing cards in probability experiments

A SET OF 52 PLAYING CARDS



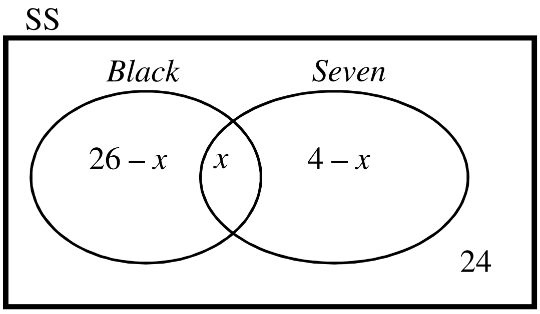
http://en.wikipedia.org/wiki/Playing_card

Notice that

- A set of playing cards contains 52 cards.
- Thirteen of the cards are marked with a black spade (♠)
- Thirteen of the cards are marked with a red heart (♥)
- Thirteen of the cards are marked with a red diamond (♦)
- Thirteen of the cards are marked with a black club (♣)
- Each set of thirteen cards is called a ***suit***
- Each suit consists of cards numbered 1 to 10, a card marked with a J (for Jack), Q (for Queen) and K (for King). The 1 is generally marked A and is called the Ace.

EXERCISE 4.5

1) The Venn diagram illustrates the number of playing cards in a pack of playing cards which are black as well as the number of cards that are sevens. Use the Venn diagram to answer the following:



   a) How many cards are there in a pack of cards?
   b) How many black cards are there in a pack of playing cards?
   c) How many sevens are there in a pack of playing cards?
   d) How many cards are black *or* seven?
   e) Find the value of x; where x is the number of black sevens in a pack of playing cards.
   f) Check your answers by substituting for x and adding.

2) The 2009 Census@School was completed by 124 975 Grade 10 learners. 190 168 of the learners who completed it were 15 years old. 82 426 of the 15 year olds were in Grade 10.
   a) How many Grade 10s were not 15 years old?
   b) How many 15 year olds were not in Grade 10?
   c) Draw a Venn diagram to illustrate the situation.
   d) How many learners were in the sample set?
   e) Suppose one of the learners in the sample set was selected at random, what is the probability (written as a percentage correct to 2 decimal places) that this learner is in Grade 10 AND is 15 years old?

3) 24 learners in a class were invited to Adam and Nisha's birthday parties.
   13 learners decided to go to Adam's party.
   12 learners decided to go to Nisha's party.
   3 learners decided to not go to either of the parties.
   a) How many learners went to the two parties (either Adam's or Nisha's or both)?
   b) How many learners went to both Adam's party ***and*** to Nisha's party?
   c) Draw a Venn diagram to illustrate the situation.
   d) Suppose one of the 24 learners is selected at random, what is the probability (written as a decimal correct to 3 decimal places):
      i) That the learner only goes to Adam's party?
      ii) That the learner goes to both parties?
      iii) That the learner doesn't go to either of the parties?

d) Finding A Relationship Between The Number Of Outcomes In Different Events

Remember:

- To find $n(P)$ count the outcomes in P only *and* in the intersection.
- To find $n(Q)$ count the outcomes in Q only *and* in the intersection.

EXAMPLE 8

a) Use the given Venn diagram to calculate:

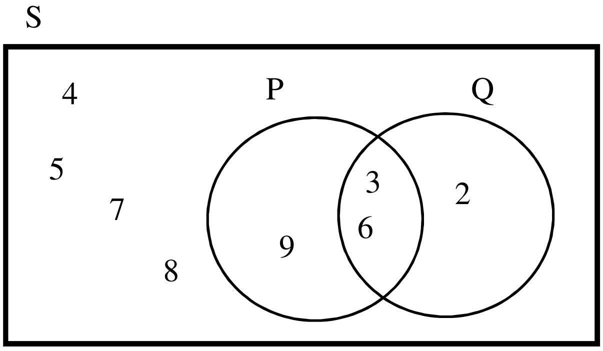
i) $n(P)$

ii) $n(Q)$

iii) $n(P\ and\ Q)$

iv) $n(P\ or\ Q)$

v) $n(P) + n(Q) - n(P\ and\ Q)$



b) Is $n(P) + n(Q) - n(P\ and\ Q) = n(P\ or\ Q)$?

<u>SOLUTION</u>:

a)

i) P consists of the elements 3, 6 and 9.
So $n(P) = 3$

ii) Q consists of the elements 2, 3 and 6
So $n(Q) = 3$

iii) P *and* Q consists of the elements in the intersection of P and Q.
So $n(P\ and\ Q) = 2$

iv) P *or* Q consists of the elements in P only, in P *and* Q and in Q only.
So P *or* Q consists of the elements 2, 3, 6 and 9 and $n(P\ or\ Q) = 4$

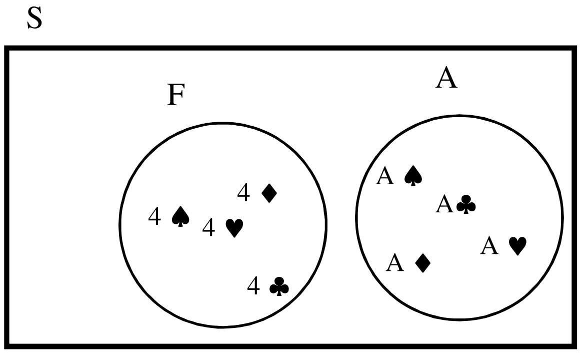
v) $n(P) + n(Q) - n(P\ and\ Q) = 3 + 3 - 2 = 4$

b) $n(P) + n(Q) - n(P\ and\ Q) = 4$
$n(P\ or\ Q) = 4$
So $n(P) + n(Q) - n(P\ and\ Q) = n(P\ or\ Q)$

EXERCISE 4.6

The Venn diagram illustrates the number of playing cards in a pack of playing cards that are fours as well as the number of cards that are aces.

Let F be the set of fours in a pack of playing cards
Let A be the set of aces in a pack of playing cards



1) Use the Venn diagram to find the following:
   a) $n(F)$
   b) $n(A)$
   c) $n(F\ and\ A)$
   d) $n(F\ or\ A)$
   e) $n(F) + n(A) - n(F\ and\ A)$
2) Is $n(F) + n(A) - n(F\ and\ A) = n(F\ or\ A)$?
3) You select a card at random from the pack of cards.
   a) Determine $n(S)$ where S is the sample set.
   b) What is the probability (written as a fraction in simplest form) that this card is:
      i) A four?
      ii) An ace?
      iii) A four *and* an ace?
      iv) A four *or* an ace?
   c) Calculate $P(F) + P(A) - P(F\ and\ A)$
   d) Is $P(F\ or\ A) = P(F) + P(A) - P(F\ and\ A)$?

NOTE:

- The relationship between the number of outcomes in events A and B is:
  $n(A\ or\ B) = n(A) + n(B) - n(A\ and\ B)$
- This relationship is true
  - when the events have *common outcomes* as in EXAMPLE 8
  - when the events have *no outcomes in common* as in EXERCISE 4.6.
- The relationship is also true for probabilities. The relationship
  $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$ is called the ***addition rule for probability.***

e) Venn Diagrams Showing The Probability Of Events Happening

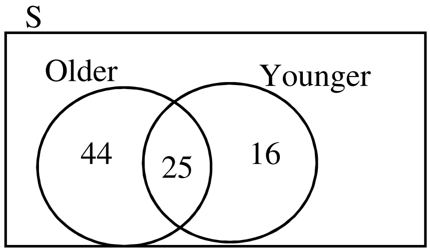
- ✓ You have worked with Venn diagrams
  - i) That list all the outcomes of the events
  - ii) That show the number of outcomes in the events.

- ✓ You can also show probabilities in Venn diagrams. The sum of the probabilities in the Venn diagram must be 1 or 100%.

EXAMPLE 9

A survey of 100 learners shows that 69 learners have an older sibling, 41 learners have a younger sibling and 25 learners have both. *(A sibling is a brother or a sister).*

The Venn diagram shows the results of the survey.



a) How many of the learners have no siblings?

b) Determine the probability (as a decimal correct to 2 decimal places) that a learner chosen at random has:
   - i) an older sibling
   - ii) a younger sibling
   - iii) an older sibling *and* a younger sibling
   - iv) no siblings

c) Redraw the Venn diagram to show the probabilities.

d) Use the Venn diagram to determine $P(an\ older\ sibling\ or\ a\ younger\ sibling)$

e) Is *P(older) + P(younger) – P(older* ***and*** *younger)*
   = P(older ***or*** *younger)?*

<u>SOLUTION</u>:

a) $n(no\ siblings) = 100 - (44 + 25 + 16) = 15$

b)

i) $P(older\ sibling) = \frac{n(older\ sibling)}{n(sample\ space)} = \frac{69}{100} = 0{,}69$

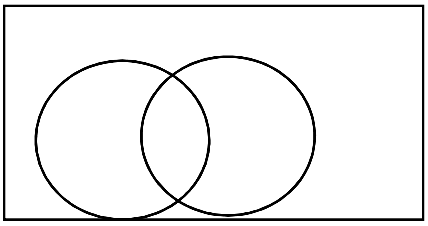
ii) $P(younger\ sibling) = \frac{n(younger\ sibling)}{n(sample\ space)} = \frac{41}{100} = 0{,}41$

iii) $P(older\ and\ younger\ sibling) = \frac{n(older\ and\ younger\ sibling)}{n(sample\ set)} = \frac{25}{100} = 0{,}25$

iv) $P(no\ siblings) = \frac{n(no\ siblings)}{n(sample\ set)} = \frac{15}{100} = 0{,}15$

EXAMPLE 9 (continued)

c)



d) $P(an\ older\ sibling\ or\ a\ younger\ sibling)$ = 0,44 + 0,25 + 0,16 = 0,85

e) *P(older)* + *P(younger)* – *P(older and younger)* = 0,69 + 0,41 – 0,25 = 0,85

So $P(older) + P(younger) - P(older\ and\ younger) =$ *P(older or younger).*

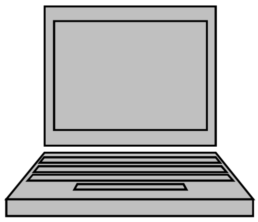
EXERCISE 4.7

1) Of the 420 Grade 10 learners at Farhana's high school, 126 do Geography (G) and 275 do Maths (M). 55 of the learners do both Geography and Maths.
   a) Draw a Venn diagram to illustrate this situation.
   b) What percentage (correct to 1 decimal place) of the learners do both Geography *and* Maths?
   c) How many learners do neither Geography nor Maths?

2) Two events A and B have the following probabilities:
   $P(A) = 0{,}2;\ P(B) = 0{,}4$ and $P(A\ and\ B) = 0{,}08$
   b) Draw a Venn diagram to illustrate the situation
   c) Determine $P(A\ or\ B)$
   d) Is $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$?

3) Results from Census 2011 show that of the 14 450 133 people in South Africa who have computer access, 16% connect to the internet using their cell phones and 9% connect to the internet using use their home computers. 3% use their cell phones to connect their home computers to the internet.

16% use cellphone

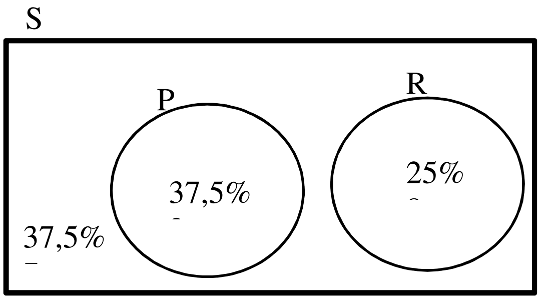


9% use their home computer

   a) Draw a Venn diagram to show this information.
   b) What is the probability that a person chosen at random uses neither a cell phone nor a home computer for internet access?
   c) How many South Africans do not use cell phones or home computers for internet access? Give your answer correct to the nearest ten thousand people.

EXERCISE 4.7 (continued)

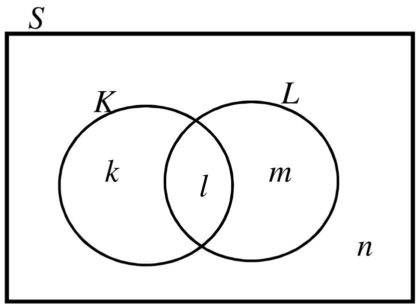
4) The following Venn diagram shows that the probability of Event P occurring is 37,5% and the probability of event R occurring is 25%.



a) Calculate the following probabilities:
   i) $P(P\ and\ R)$
   ii) $P(P\ or\ R)$

b) Show that $P(P) + P(R) - P(Pand\ R) = P(P\ or\ R)$

5) The Venn diagram shows the relationship between events K and L.
Suppose one of the items in the events is selected at random, then the probability of the item being in event K is $P(K) = \frac{n(K)}{n(S)} = \frac{k+l}{k+l+m+n}$



a) Use the Venn diagram to find:
   i) $P(L)$
   ii) $P(K\ and\ L)$
   iii) $P(K\ or\ L)$

b) Show that $P(K\ or\ L) = P(K) + P(L) - P(K\ and\ L)$

MUTUALLY EXCLUSIVE EVENTS

✓ ***Mutually exclusive*** events are events that cannot happen at the same time.

Examples of mutually exclusive events:

- Turning left and turning right are mutually exclusive because you can't do both at the same time
- Taking a 4 and taking a 7 from a pack of cards are mutually exclusive because a card cannot be a 4 and a 7 at the same time.

Examples of events that are NOT mutually exclusive:

- Turning left and scratching your head can happen at the same time, so they are **NOT** mutually exclusive
- Taking a King (K) and taking a heart (♥) from a pack of cards in **NOT** mutually exclusive as the card could be the king of hearts (K♥).

✓ If events are **NOT** mutually exclusive, then $\boldsymbol{P(A \text{ and } B) \neq 0}$ and $\boldsymbol{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$

✓ Mutually exclusive sets are ***disjoint sets***. If events **ARE** mutually exclusive, then $\boldsymbol{P(A \text{ and } B) = 0}$.

✓ If events are mutually exclusive, then $\boldsymbol{P(A \text{ or } B) = P(A) + P(B)}$

EXERCISE 4.8

1) You randomly select a single card from a pack of playing cards.
   a) Can it be red and black at the same time?
   b) Can it be black and a 10 at the same time?
   c) Can it be a King (K) and a Queen (Q) at the same time?
2) Are the following events mutually exclusive or not? Explain your answer.
   a) Randomly drawing a red card and a black Jack (J) from a pack of playing cards.
   b) Randomly drawing a red card and a card marked with a diamond (♦) from a pack of playing cards.
   c) Tossing a coin and getting a heads (H) and getting a tails (T) at the same time.
   d) Rolling a fair dice and getting a 3 and a 4 at the same time.
   e) Eating a sandwich and eating jam.
   f) 'Living in Kwa-Zulu Natal' and 'speaking English at home'.

PROBABILITY RULES

a) The Addition Rule for Probability (often called the "OR LAW")

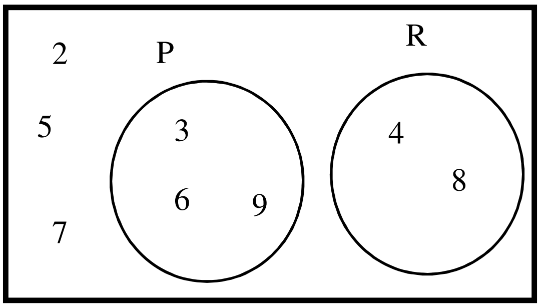
- ✓ If two events are ***NOT mutually exclusive***, then
  P(A or B) = P(A) + P(B) – P(A and B)

- ✓ If two events are ***mutually exclusive***, then
  - ***P(A and B) = 0***
  - ***P(A or B) = P(A) + P(B)***

EXAMPLE 10
In the Venn diagram, Sample Space S contains the numbers from 2 to 9. The elements of Event P are the multiples of 3 in S.
The elements of Event R are the multiples of 4 in S.

a) Are events P and R mutually exclusive?
b) Are sets P and R disjoint?
c) Determine *n(S)* where S is the Sample Space.
d) Use the addition rule P(P or R) = P(P) + P(R) – P(P and R) to determine P(P or R).

S



<u>SOLUTION</u>:

a) Events P and R are *mutually exclusive* – they have no elements that are the same.
b) Sets P and R are *disjoint* – mutually exclusive events are represented as disjoint sets on a Venn diagram.
c) *N(S)* = 8
d) $P(P) = \frac{n(P)}{n(S)} = \frac{3}{8}$

$P(R) = \frac{n(P)}{n(S)} = \frac{2}{8}$

$P(P \text{ and } R) = \frac{n(P \text{ and } R)}{n(S)} = \frac{0}{8} = 0$

So $P(P \text{ or } R) = P(P) + P(R) - P(P \text{ and } R) = \frac{3}{8} + \frac{2}{8} - 0 = \frac{5}{8}$

b) The Complementary Rule:

- ✓ ***Complementary events*** are events that cannot occur at the same time.

- ✓ If the event A occurs, then the complement of A is ***not A***.

  Examples of complementary events are
  - Speaking English in everyday conversation or not speaking English in everyday conversation.
  - Walking to school on Monday morning or not walking to school on Monday morning.
  - Tossing a coin and getting a Head or not getting a Head.
  - Rolling a fair dice and scoring a 3 or not scoring a 3.

- ✓ The event $\boldsymbol{A}$ and its complement ***not A*** are mutually exclusive.

- ✓ We write: $\boldsymbol{P(A) + P(not\ A) = 1}$

  $\boldsymbol{or \qquad P(not\ A) = 1 - P(A)}$

  $\boldsymbol{or \qquad P(A) = 1 - P(not\ A)}$

EXAMPLE 11

a) The probability that a Grade 10 learner, chosen at random, will pass English is 80%. What is the probability that this learner will not pass English?

b) The probability of getting a white ball from a bag of balls is $\frac{1}{4}$. What is the probability of ***not*** getting a white ball?

c) A bag contains red and blue cards. The probability of taking a red card is $\frac{2}{5}$. What is the probability of taking a blue card?

<u>SOLUTION</u>:

a) You either pass English or you do not pass English, so the events are complementary.

$P(not\ pass\ English) = 1 - P(pass\ English) = 100\% - 80\% = 20\%.$

This learner has a 20% chance of not passing English.

b) $P(ball\ is\ not\ white) = 1 - P(ball\ is\ white) = 1 - \frac{1}{4} = \frac{3}{4}$

c) Taking a red card and taking a blue card from a bag of red and blue cards are complementary events.

$$\begin{aligned} P(taking\ a\ blue\ card) &= 1 - P(not\ taking\ a\ blue\ card) \\ &= 1 - P(taking\ a\ red\ card) \\ &= 1 - \frac{2}{5} \\ &= \frac{3}{5} \end{aligned}$$

EXERCISE 4.9

1) Out of a group of 27 girls, 10 play netball (N) and the rest do not play netball.
   a) How many in the group of girls ***do not*** play netball.
   b) Represent this information in a Venn diagram.
   c) Are the events play netball and do not play netball complementary?
   d) Determine *n(S)* where S is the sample set.
   e) One of the 27 girls is chosen at random. Determine the probability as a common fraction that:
      i) The chosen girl plays netball.
      ii) The chosen girl does not play netball.

2) A number is chosen at random from a set of numbers from 1 to 50, including 1 and 50. Event A is choosing a number that is a perfect square.
   a) Determine *n(S)* where S is the sample space.
   b) List the elements of A.
   c) Determine *P(A)*, written as a fraction in simplest form.
   d) Hence calculate the probability that the chosen number is ***not*** a perfect square.

3) In the 2009 Census@School, the Grade 10 to 12 learners were asked what type of home they stay in most of the time. 58,6% of the learners answered that they live in a house (H) and 14,8% live in a traditional dwelling (T).
   a) Are ‘living in a house (H)’ and ‘living in a traditional dwelling (T)’ mutually exclusive?
   b) Calculate the percentage of the Grade 10 to 12 learners who do not live in a house or traditional dwelling.
   c) Draw a Venn diagram to show the percentage of learners who live in a house (H), the percentage of learners who live in a traditional dwelling (T) and the percentage of learners who do not live in a house or in a traditional dwelling.
   d) Suppose one of the Grade 10 to 12 learners is selected at random. Determine
      i) *P(H)*
      ii) *P(T)*
      iii) *P(not H)*
      iv) *P(not T)*
      v) *P(H or T)*

c) Using the probability rules to determine probabilities

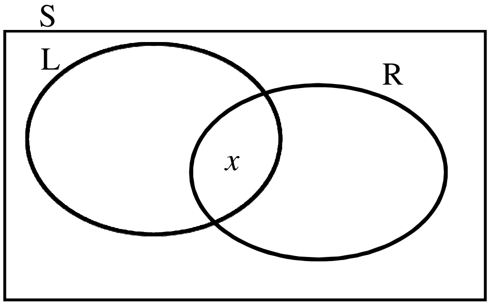
✓ Venn diagrams and rules can be used to find unknown values.

EXAMPLE 12
During the 2009 Census@School, 69 469 male learners were asked about whether they wrote with their right hand, left hand or both.
9 759 of the male learners said that they were left-handed (L).
61 368 of the male learners said that they were right-handed (R).
x learners wrote with both their right and left hand.

a) Draw a Venn diagram to illustrate the information about the male learners who participated in this 2009 Census@ School project and determine the value of x.
b) Determine $n(S)$ where S is the sample set.
c) Calculate the probability (as a percentage correct to 1 decimal place) that a male learner chosen at random from this sample set writes with both hands.

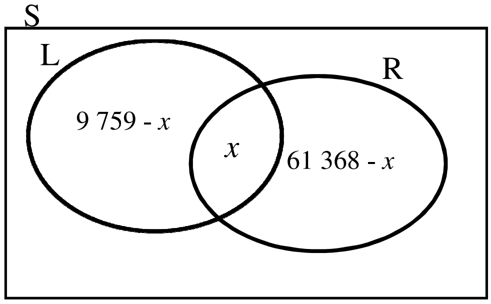
<u>SOLUTION</u>:
The events of writing with your left hand and writing with your right hand are not mutually exclusive because some people are able to write with both hands.

a) Draw 2 intersecting circles in the rectangle.



Fill in x where the circles **intersect** because x males write with both hands.

Use the information '9 759 are left handed' to determine the number of males who write with their left hand **only**
Number who write with their left hand only
$= 9\ 759 - x$
Fill this in on the Venn diagram.



Use the information '61 368 are right handed' to determine the number of males who write with their right hand **only**:

Number who write with their right hand only = $61\ 368 - x$
Fill this in on the Venn diagram.

Form an equation using the fact that: $n\ (L \text{ or } R) = n\ (L) + n\ (R) - n\ (L \text{ and } R)$

$$69\ 469 = 9\ 759 + 61\ 368 - x$$
$$69\ 469 = 71\ 127 - x$$
$$\therefore x = 1\ 658$$

So, 1 658 male learners write with both hands.

b) $n(S) = 69\ 469$

c) $P(writes\ with\ both\ hands)$
$= \frac{n(writes\ with\ both\ hands)}{n(S)} = \frac{1\ 658}{69\ 469} = 2{,}38\ \ldots\% = 2{,}4\%$

EXERCISE 4.10

1) At a fast food outlet 75 people were surveyed to find out what food they had eaten. 23 had eaten burgers (B), 30 had eaten fried chicken (F) and 7 had eaten both.
   a) Are these two events mutually exclusive?
   b) Calculate the number of people interviewed who had eaten neither fried chicken nor burgers.
   c) Draw a Venn diagram showing all the given information.
   d) Determine $n(S)$ where S is the sample set.
   e) Calculate the probability that one of these people selected at random had eaten neither burgers nor fried chicken. Give your answer as a percentage correct to the nearest whole number.

2) A bag contains thirty five counters of the same shape and size; 10 counters are yellow and 25 counters are green. You draw one counter at random.
   a) Are these events mutually exclusive?
   b) Draw a Venn diagram to represent the information.
   c) Determine $n(S)$ where S is the sample set.
   d) Suppose you put your hand in the bag and take out a counter. What is the probability (as a fraction in simplest form)
      i) That you take a yellow counter
      ii) That you take a green counter?
   e) Determine
      i) $P(Y \; and \; G)$
      ii) $P(Y \; or \; G)$

3) Given P(A) = 0,5; P(B) = 0,4 and P(A and B) = 0,3.
   a) Use the formula $P(A \; or \; B) = P(A) + P(B) - P(A \; and \; B)$ to determine P(A or B). Write the answer in a decimal form.
   b) Draw a Venn diagram showing these probabilities.

4) A card is drawn at random from a pack of 52 playing cards.
   What is the probability (giving your answer as a common fraction in simplest form) that this card is
   a) A black card?
   b) A seven?
   c) A black 7?
   d) Not a black 7?

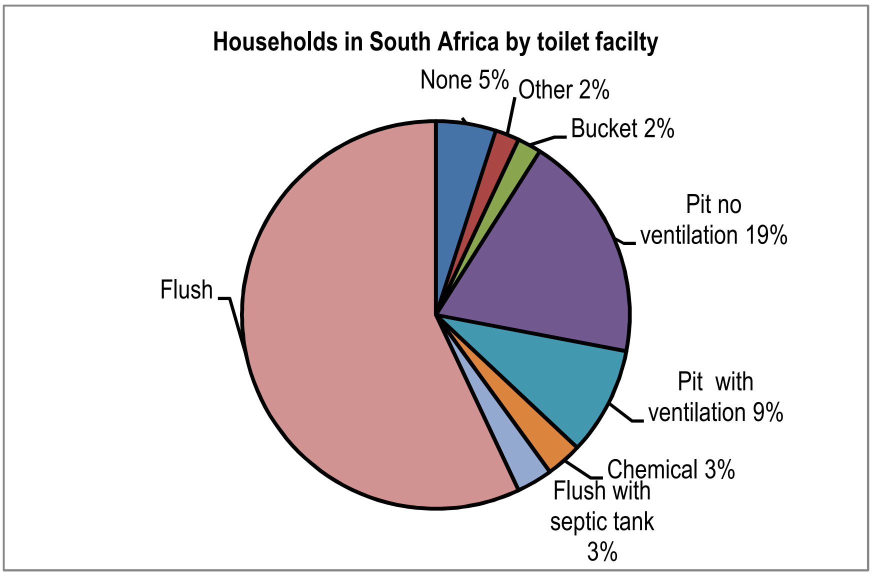
EXERCISE 4.10 (continued)

5) In a the 2009 Census@School survey of 15 to 19 year olds, learners were asked what sport they would like to take part in. Below is data adapted from the database:

Age 15 -19 males and females sport	%
Athletics (A)	x
Volleyball (V)	21%
Athletics and Volleyball	12%
Neither of these sports	65%

a) Draw a Venn diagram to illustrate the data given in the table.
b) Use the Venn diagram to determine the value of x.
c) Calculate the probability that a learner chosen at random:
   i) likes Athletics but not Volleyball
   ii) likes Athletics or Volleyball

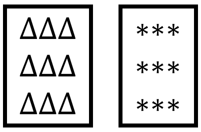
6) Census 2011 gives the following results about the types of toilet facilities in South African homes:



a) What percentage of households has toilets that flush?
b) Assuming that households have only one type of toilet facility, calculate the probability that a household selected at random
   i) has a toilet that flushes or has a chemical toilet
   ii) has a pit toilet or a bucket toilet
c) Use your answers to b) to write down two conclusions about the types of toilets found in South African homes.

<u>EXERCISE 4.10 (continued)</u>

7) A set of cards of the same shape and size have either triangles or stars on them. A card is drawn at random from the set of cards.

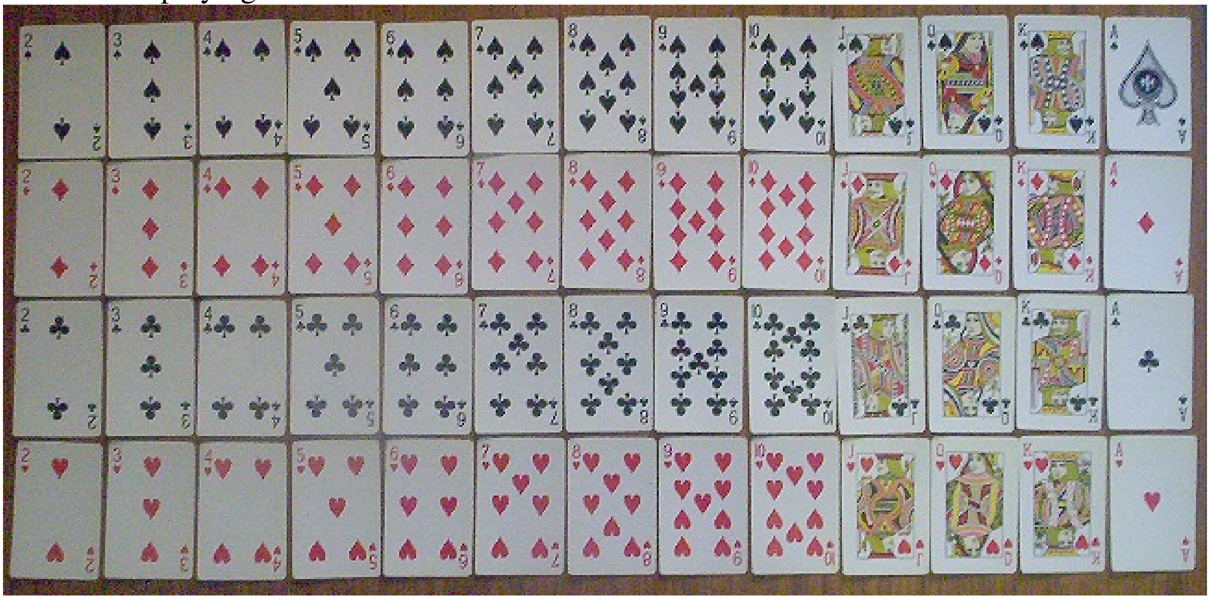


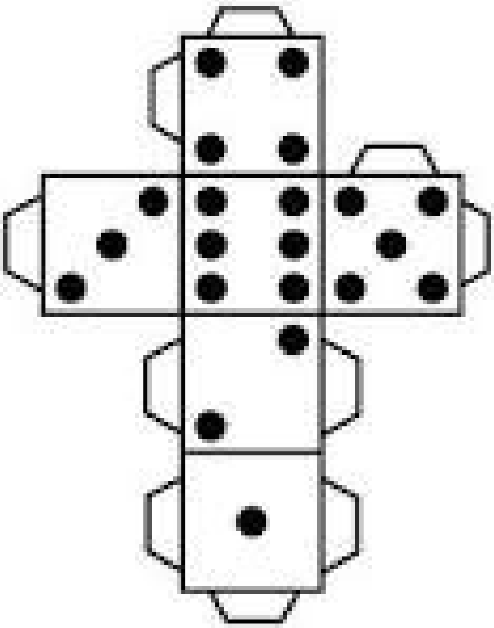
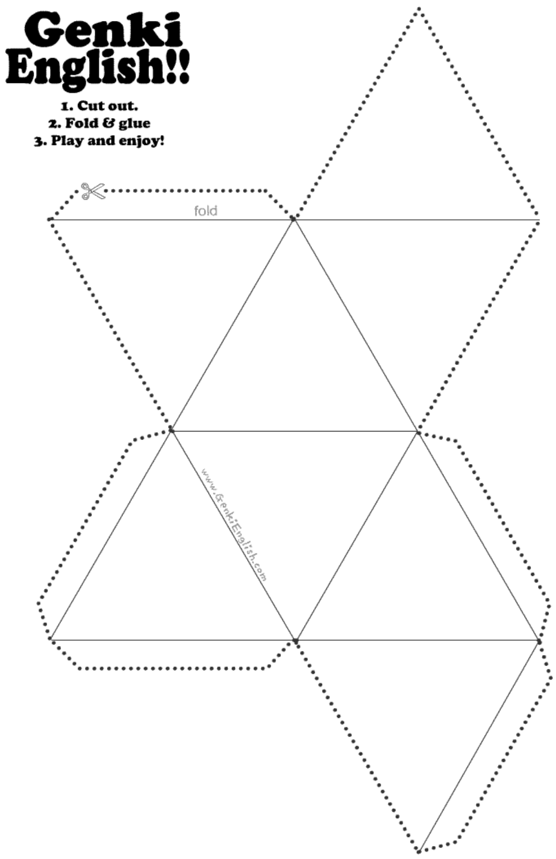
   a) Is it possible to draw a card that has both triangles ***and*** stars on it?
   b) There are 8 triangle cards (Δ). Let the number of star cards ($*$) be x. Draw a Venn diagram to show the number of cards.
   c) Determine the value of x if the probability of drawing a star card = $P(*) = \frac{5}{7}$.

Bibliography

Barton, J. (2007) *Gamma Mathematics Workbook NCEA Level 1*. Pearson Longman
Barton, J. (2005) *Theta Mathematics Workbook NCEA Level 2.* Pearson Longman
Barton, J. (2007) *Sigma Mathematics Workbook NCEA Level 3*. Pearson Longman
Buckwell, G (1997) *Mastering Mathematics*. Macmillan Master Series
Speed, B. (1997) *Higher Mathematics for GCSE.* Collins Educational
Statistics South Africa (2010) *Census@School Results 2009*
2011 Census in South Africa
www.onlinemathlearning.com/probability-venn-diagram.html
http://www.haesemathematics.com.au/samples/ibmyp4_10.pdf

A full set of playing cards



The net of a six sided dice	The net of an eight sided dice
	

5

Grade 11 Probability

In this chapter you:

- Revise the language of probability
- Use Venn diagrams showing two events to determine probabilities
- Identify independent and dependent events
- Use 2-way tables to determine probabilities of events happening
- Investigate the multiplication law for probability
- Use tree diagrams to determine probabilities of events happening
- Use Venn Diagrams showing three events to determine probabilities.

WHAT YOU LEARNED ABOUT PROBABILITY IN GRADE 10

In Grade 10 you covered the following probability concepts:

- *The use of probability models to compare the relative frequency of events with the theoretical probability*
- *The use of Venn diagrams to solve probability problems*
- *Rules for any **two events** A and B in sample space S:*
  - *P(A or B) = P(A) + P(B) – P(A and B)*
  - *A and B are mutually exclusive if P(A and B) = 0*
  - *A and B are complementary if they are mutually exclusive*
    And P (A) + P (B) = 1
    Or P(B) = P(not A) = 1 – P(A)
    Or P(A) = P(not B) = 1 – P(B)

THE LANGUAGE OF PROBABILITY

✓ A ***statistical experiment*** is one in which there are a number of possible outcomes and we have no way of predicting which outcome will actually occur.

✓ A ***sample space*** is the set of all the possible outcomes in an experiment.

✓ An ***event*** is any set of possible outcomes of an experiment.

For example:
If we roll a dice, the ***sample space*** consists of the numbers 1, 2, 3, 4, 5 and 6. Getting an even number is an ***event***, and is a subset of the sample space.

✓ $The\ probability\ of\ any\ event\ A\ occurring$
$$= \frac{number\ of\ outcomes\ in\ A}{total\ number\ of\ outcomes\ in\ the\ sample\ space}$$

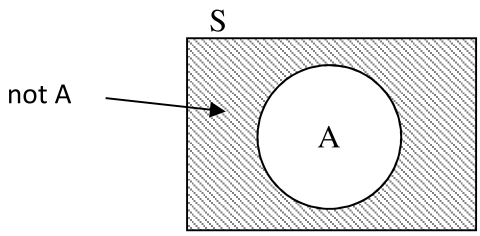
We write $P(A) = \frac{n(A)}{n(S)}$ where $0 \leq P(A) \leq 1$.

- This means the probability cannot be negative or greater than 1
- If the event A is ***impossible*** then its probability is 0.
- If the event A is ***certain***, then its probability is 1.

✓ ***Complementary events:***

The events 'A' and 'not A' cannot happen at the same time. We say that they are ***complementary***.

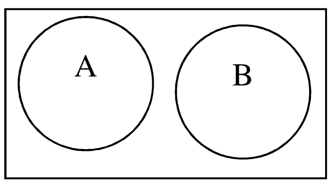
So $P(not\ A) = 1 - P(A)$
or
$P(A) + P(not\ A) = 1$
or
$P(A) = 1 - P(not\ A)$



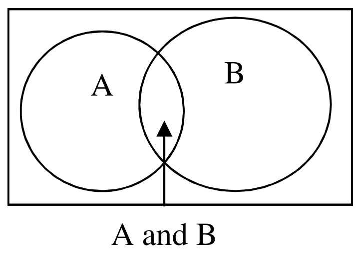
✓ ***Mutually exclusive events***

Two events are ***mutually exclusive*** if they cannot happen at the same time. If one of them happens, it excludes the other.

- If events are mutually exclusive then $P(\text{A and B}) = 0$.
- Mutually exclusive sets are ***disjoint sets.***



A and B are mutually exclusive



A and B are NOT mutually exclusive

✓ The ***addition law for probability:***

- If the events A and B are ***NOT*** mutually exclusive, the probability that A *or* B will occur is given by: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$
- If the events A and B ***ARE*** mutually exclusive, then the rule can be simplified to $P(A\ or\ B) = P(A) + P(B)$

VENN DIAGRAMS SHOWING TWO EVENTS

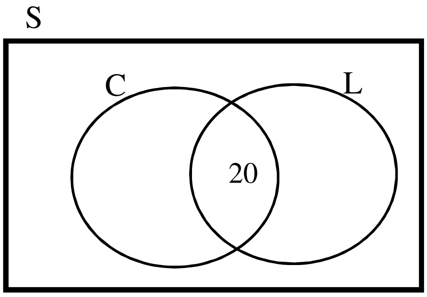
- ✓ A ***Venn diagram*** is a simple representation of a sample space.
  - Usually a rectangle is used to represent a Sample Space (S).
  - Circles are used to represent events within the sample space.
- ✓ We can use Venn diagrams to assist us to work out the probability of an event happening.

EXAMPLE 1
In the 2009 Census@School learners were asked to indicate whether they have access to a telephone or a cellphone at home.
290 Grade 11 learners in one high school responded as follows:
150 have access to a cellphone
80 have access to a telephone at home
20 have access to both cellphone and to a telephone at home
Draw a Venn diagram to represent the situation.

<u>SOLUTION</u>:
Let C be the event that a learner has access to a cellphone and L be the event that a learner has access to a telephone at home.

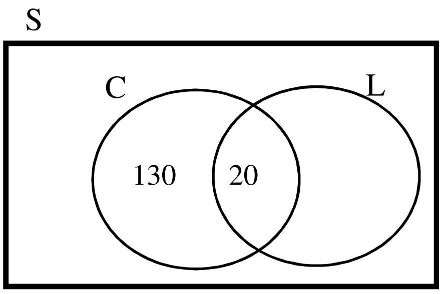
STEP 1:
- Draw a rectangle to represent the Sample Space. Inside the rectangle draw two intersecting circles
- Start at the intersection:
  20 learners have access to both a cellphone and a landline phones at home i.e. n(C and L) = 20.

Write 20 in the intersection.



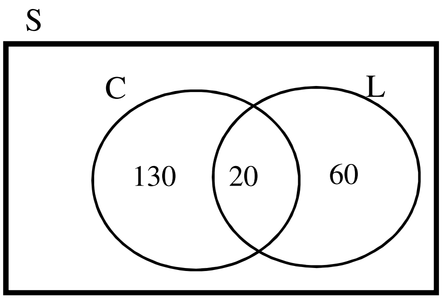
STEP 2
- 150 have access to cellphone i.e. n(C) = 150
- But 20 has already been entered in the intersection, so $150-20=130$ must be entered into the other part of circle C.

Write 130 into the remaining part of C



STEP 3
- 80 learners have access to a telephone at home i.e. n(L) = 80
- But 20 of the 80 have been entered into the intersection, which means that $80-20=60$ must be entered into the remaining part of circle L.

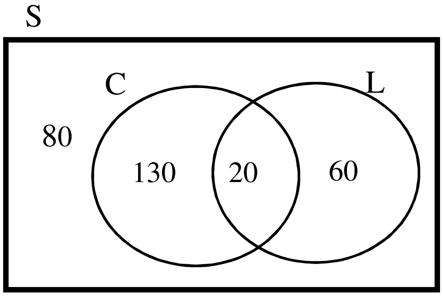
Write 60 into the remaining part of L



<u>EXAMPLE 1</u> (continued)

STEP 4:

- The total number of learners is 290 i.e. n(S)=290
- The total number of learners who neither have access to a Cellphone nor to a telephone at home = 290 – (130 + 20 + 60) = 80



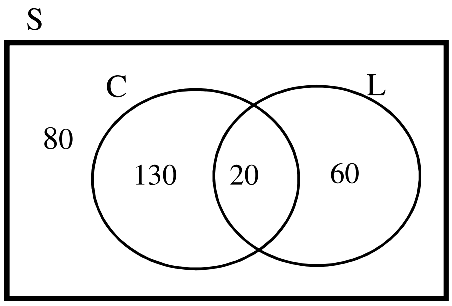
EXAMPLE 2

Use the Venn diagram to answer the following.

a) Determine *n(S)*

b) Determine the following probabilities as a percentage correct to 1 decimal place

i) P(C)
ii) P(L)
iii) P(C and L)
iv) P(C or L)
v) P(not C)
vi) P(not L)
vii) P(not (C or L))
viii) P(not (C or L)
ix) P(at least one type of phone)



<u>SOLUTION</u>:

a) *n(S)* = 80 + 130 + 20 + 60 = 290

b)

i) P(C) = $\frac{n(C)}{n(S)} = \frac{130+20}{290} = \frac{150}{290} \approx 51{,}7\%$

ii) P(L) = $\frac{n(L)}{n(S)} = \frac{20+60}{290} = \frac{80}{290} \approx 27{,}6\%$

iii) P(C and L) = $\frac{n(C\ and\ L)}{n(S)} = \frac{20}{290} \approx 6{,}9\%$

iv) P(C or L) = P(C) + P(L) – P(C and L) ≈ 51,7% + 27,6% – 6,9% = 72,4%

v) P(not C) = 1 – P(C) ≈ 100% – 51,7% = 48,3%

vi) P(not L) = 1 – P(L) ≈ 100% – 27,6% = 72,4%

vii) P(not (C or L)) = 1 – P(C or L) ≈ 100% – 72,4% = 27,6%

viii) P(L only) = $\frac{60}{290} \approx 20{,}7\%$

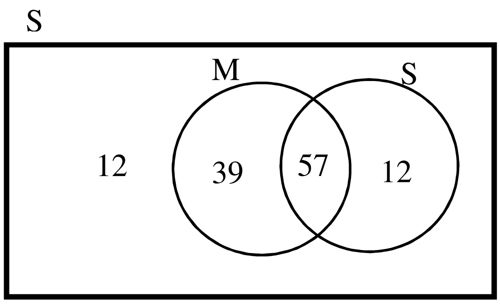
ix) P(at least one type of phone) = P(C or L) = 72, 4%

Remember that
1 = 100% = $\frac{100}{100}$

EXERCISE 5.1

1) A survey was conducted with the Grade 11 learners at a certain school to find out which subjects they are taking this year. It was found that
   - 96 learners were taking Maths (M),
   - 69 were taking Physical Sciences (S),
   - 57 were taking Maths and Physical Sciences
   - 12 were taking neither Maths nor Physical Sciences.

   The results are shown in the Venn diagram.



   a) How many learners were interviewed?
   b) A learner is selected at random from the learners who were interviewed. Determine the following as a percentage:
      i) P(M)
      ii) P(S)
      iii) P(M and S)
      iv) P(M or S)
      v) P(not M)
      vi) P(not S)
      vii) P(not (M or S))

2) During 2009 Census@School, learners were asked to indicate their favourite subjects. 1 320 learners in a certain school responded as follows:
   - 474 learners like Maths (M)
   - 894 learners like English (E)
   - 105 learners like both Maths and English

   a) Draw a Venn diagram to show the results of the survey
   b) How many learners like neither Maths nor English
   c) What is the probability (as a decimal fraction correct to 2 decimal places and as a percentage correct to 1 decimal place) that a randomly picked learner in the school likes
      i) both Maths and English
      ii) Maths only
      iii) English only?

INDEPENDENT AND DEPENDENT EVENTS

✓ ***Independent events*** are events where the outcome of the second event is not affected by the outcomes of the first event.

For example

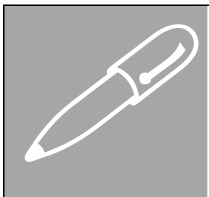
a) Suppose a dice is thrown twice. The second throw of the dice is not affected by the first throw of the dice. Thus, ***the two events are independent***.

b) Suppose you toss a coin and throw a dice. The number that you throw on the dice is not dependent on whether you get heads or tails on the coin. Thus, ***the two events are independent***

c) If you randomly select a card from a pack of playing cards, ***replace it*** then randomly select a second card, ***the events are independent.***

✓ ***Dependent events*** are events that are not independent. The probabilities of dependent events DO affect each other.

For example,
If you select a card from a pack of playing cards and then randomly select another card ***without replacing the first card*** these events are ***dependent events.***



EXERCISE 5.2

Decide whether the events described below are independent or dependent. Give a reason for your answer.

1) Tossing a coin and taking a card from a pack of playing cards.
2) Throwing a dice and spinning a spinner.
3) Taking a card and spinning a spinner.
4) Getting 7 on the spin of a spinner, and then getting 3 when it is spun a second time.
5) Driving over 120 km/h, and causing a car accident.
6) Exercising frequently and having a low resting heart rate.
7) Randomly selecting a ball numbered from 1 to 40 from a box, and then selecting a second numbered ball from the remaining balls in the box.
8) Selecting a ball numbered from 1 and 52 from a box, and then selecting a second numbered ball between 1 and 52 from a second box.

TWO-WAY TABLES

✓ In some probability problems we look at the combined or compound outcomes of several activities.

These activities are often ‘games’ such as

- Spinning a spinner and taking a card
- Throwing two dice
- Tossing a coin three times.

✓ To find out the possible outcomes (the sample set), you can make a list, draw up a table, or use a tree diagram. In order to make sure that we don’t miss any of the combined outcomes, it is best to use a table or a tree diagram.

✓ For two or more activities, a table is often the easiest way to list the outcomes.

We draw up a ***two-way table*** as follows:

- Write one activity down the side of the table
- Write the other activity along the top.
- Write the combined result from the two activities in each cell of the table.

For example:
The following table shows the results obtained when two coins are tossed.

		Coin 2	
		H	**T**
Coin 1	**H**	H ; H	H ; T
	T	T ; H	T ; T

A two-way table

EXAMPLE 3

Bag A contains 2 red balls (R) and 1 white ball (W).
Bag B contains 2 red balls (R) and 1 white ball (W).
A ball is taken at random for each bag.

a) Copy and complete the table to show all possible pairs of colours.

		Bag A		
		R	**R**	**W**
Bag B	**R**			
	R			
	W			

b) Determine n(S) where S is the sample set.

c) A ball is selected at random from each of the two bags. Determine the following as a fraction in simplest form:
   i) P(R; R)
   ii) P(R; W)
   iii) P(W; R)
   iv) P(W; W)

b) Determine the probability that the two balls are the same colour

SOLUTION:

a)

		Bag A		
		R	**R**	**W**
Bag B	**R**	R; R	R; R	R; W
	R	R; R	R; R	R; W
	W	W; R	W; R	W; W

b) n(S) = 9

c)

i) $P(R;R) = \frac{n(R;R)}{n(S)} = \frac{4}{9}$

ii) $P(R;W) = \frac{n(R;W)}{n(S)} = \frac{2}{9}$

iii) $P(W;R) = \frac{n(W;R)}{n(S)} = \frac{2}{9}$

iv) $P(W;W) = \frac{nW;W)}{n(S)} = \frac{1}{9}$

d) Two balls of the same colour were taken out of the two bags five times – four times the two balls were red and one time the two balls were white.

$P(two\ balls\ are\ the\ same\ colour) = \frac{5}{9}$

EXAMPLE 4

A coin is tossed and a dice is thrown

a) Are the events 'toss a coin' and 'throw a dice' independent?
b) Use a two-way table to determine all the possible outcomes
c) Determine n(S) where S is the sample set.
d) Suppose a coin and a dice are thrown together and the results noted. Determine the following probabilities, writing them as fractions in simplest form:
    i) P(H)
    ii) P(3)
    iii) P(H and 3)
e) Determine P(H) × P(3)
f) What do you notice about the answers P(H and 3) and P(H) × P(3)

SOLUTION:

a) The two events are independent as the number on the dice does not depend on whether you get heads or tails on the coin and getting heads or tails on a coin does not affect the number that comes up on a dice.

b)

		DICE					
		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

c) n(S) = 12

d)

i) $P(H) = \frac{n(events\ having\ H\ as\ one\ of\ its\ outcomes)}{n(S)} = \frac{6}{12} = \frac{1}{2}$

ii) $P(3) = \frac{n(events\ having\ 3\ as\ one\ of\ its\ outcomes)}{n(S)} = \frac{2}{12} = \frac{1}{6}$

iii) $P(H\ and\ 3) = \frac{n(H\ and\ 3)}{n(S)} = \frac{1}{12}$

e) $P(H) \times P(3) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

f) The answers to P (H and 3) and P (H) ×P (3) are the same.

EXERCISE 5.3

1) A coin and a dice are used in a game. The counter has two faces, one marked 1 and the other one marked 2. The counter and dice are thrown together. The numbers obtained are multiplied together.
   a) Copy and complete this table to show all the possible outcomes.

×		Number on dice					
		1	2	3	4	5	6
Number on counter	1						
	2						

   b) Determine n(S) where S is the sample space
   c) The counter and the dice are thrown together and the results noted. Determine the following as as a fraction in simplest form:
      i) P(2)
      ii) P(5)
      iii) P(2 or 5)
      iv) P(even number)

2) A supermarket runs a competition. Each customer is given a card with two whole numbers printed on it. Each of these whole numbers are chosen at random from the numbers 1; 2; 3; 4; 5; 6; 7; 8; 9. The customer has to add her two whole numbers together to find her total.

   a) Copy and complete the table to show all the totals of the pairs of numbers.

+		Second number								
		1	2	3	4	5	6	7	8	9
First number	1									
	2									
	3									
	4									
	5									
	6									
	7									
	8									
	9									

   b) How many pairs of numbers give a total of 14?
   c) Determine the probability (as a percentage of the sample space, correct to 1 decimal place) that these numbers would win the prize:
      i) 14
      ii) 15
   d) The supermarket decides to make 2 and 10 the lucky numbers of the day. Which one of these two numbers has a greater chance of being chosen? Why.

<u>EXERCISE 5.4</u> (continued)

3) A coin is tossed and a dice is thrown. All the possible outcomes are recorded in the two-way table.
   a) Are these two events independent?
   b) Determine n(S) where S is the sample set.
   c) Determine the following, writing each one as a fraction in simplest form:
      i) P(T)
      ii) P(6)
      iii) P(T) × P(6)
   d) Is P (T and 6)
      = P (T) × P (6)?

		DICE					
		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

   e) Determine the following, writing them as fractions in simplest form.
      i) P(T and even)
      ii) P(even)
      iii) P(T) × P(even)
   f) Is P (T and even) = P (T) × P (even)?
   g) Determine the following, writing them as fractions in simplest form.
      i) P(H)
      ii) P(H and a multiple of 3)
      iii) P(multiple of 3)
      iv) P(H) ×P(multiple of 3)
   h) Is P (H) × P (multiple of 3) = P (H and multiple of 3)?
4) A spinner has the shape of a regular pentagon. The five sections of the spinner are numbered 2; 2; 3; 5; 5. When the spinner is used, it is equally likely to stop on any one of its five edges.

   a) In a game, Nomsa spins the spinner twice and notes the results. Is spinning the spinner twice two independent events?
   b) Draw a two-way table to show all the possible results for the two spins.
   c) Determine the following as fractions in simplest form
      i) P(the first number is a 2)
      ii) P(the second number is a 5)
      iii) P(the first number is a 2 and the second number is a 5)
      iv) P(the first number is a 2) × P(the second number is a 5)
   d) Is P(the first number is a 2 and the second number is a 5)
      = P(first number is a 2) × P(second number is a five)?
5) Suppose you have two independent events M and N. Complete the following:

   P(M and N) = …………………………………

THE MULTIPLICATION LAW FOR PROBABILITIES

- ✓ The ***multiplication law*** for probability only holds for ***independent events.***

  If A and B are independent events, then P(A and B) = P(A) × P(B).

- ✓ The multiplication law is often called the '***and law***'. It gives the probability that one event **AND** another happens

- ✓ This multiplication law can be used for ***more than two*** independent events.
  If A, B and C are independent events, then
  P(A and B and C) = P(A) × P(B) × P(C)

- ✓ The multiplication law is a quick way to find the probability that one event ***and*** another event happens. It saves you having to list all the possible combined results for the two events and then picking out the favourable results for a combined event.

  However, before you use it, ***make sure that the events to be combined are independent***!

DIFFERENT ACTIVITIES	REPEATED ACTIVITIES
When two events result from ***two completely separate and different activities***, it is clear that they must be ***independent***.	***Repeating an activity in exactly the same way and under the same circumstances*** results in independent events.
Here are some examples of ***different activities:***	Here are some examples of ***repeated activities:***
Tossing a coin and taking a card 	Tossing a coin twice OR tossing two coins 
Throwing a dice and spinning a spinner. 	Throwing a dice several times OR throwing several dice 

Using two-way tables to show probabilities

EXAMPLE 5
Two dice are thrown.
a) Draw a tree diagram to show all the totals when the dice are added.
b) Determine
   i) n(S)
   ii) P(7)
   iii) P(4)
c) Are P(7) and P(4) independent events?
d) Use a two-way table to find
   i) The probability of getting a sum of 7 on the first throw of the two dice **OR** a sum of 4 on the second throw.
   ii) The probability of getting a sum of 7 on the first throw of the two dice **AND** a sum of 4 on the second throw.
e) Which is greater: the probability of getting a 7 ***or*** a 4, OR the probability of getting a 5 ***and*** a 4.

<u>SOLUTION</u>:
a) The two-way table shows the totals when the numbers on the two dice are ***added***

+		**DICE 2**					
		1	**2**	**3**	**4**	**5**	**6**
DICE 1	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) There are 36 outcomes, so *n(S)* = 36
   ii) Six of the outcomes give 7 as the total: P(7) = $\frac{n(7)}{n(S)} = \frac{6}{36} = \frac{1}{6}$
   iii) Three of the outcomes give 4 as the total: P(4)= $\frac{n(4)}{n(S)} = \frac{3}{36} = \frac{1}{12}$

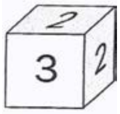
c) The probability of getting a sum of 7 on the first throw of 2 dice and the probability of getting a sum of 4 on the second throw are ***independent events*** because getting a 7 has no effect on getting a 4

d)
   i) P(7 or 4) = P(7) + P(4) = $\frac{1}{6} + \frac{1}{12} = \frac{2+1}{12} = \frac{3}{12} = \frac{1}{4}$
   ii) P(7 and 4) = P(7) × P(4) = $\frac{1}{6} \times \frac{1}{12} = \frac{1}{72}$

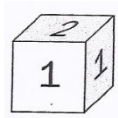
e) So the probability of getting a 7 or a 4 is greater than getting a 7 and a 4.

EXAMPLE 6

A fair BLUE dice has its faces marked with the numbers 2; 2; 2; 2; 3; 3



A fair RED dice has its faces marked with the numbers 1; 1; 2; 2; 2; 3



The two dice are thrown together.

a) Are these two events independent? Explain.

b) Copy and complete this probability table for this situation

		RED DICE		
		P(1) =	P(2) =	P(3) =
BLUE DICE	P(2) =			
	P(3) =			

c) Use your table to find:
   i) The probability of getting a 2 on the blue die and a 1 on the red dice
   ii) The probability of getting a 3 on the blue dice and a 2 on the red dice
   iii) The probability of getting a double 2

<u>SOLUTION</u>:

a) These two events are independent. A number thrown on the blue dice is independent of a number thrown on the red dice.

b) On the BLUE dice, P(2) = $\frac{4}{6} = \frac{2}{3}$ and P(3) = $\frac{2}{6} = \frac{1}{3}$

On the RED dice, P(1) = $\frac{2}{6} = \frac{1}{3}$; P(2) = $\frac{3}{6} = \frac{1}{2}$ and P(3) = $\frac{1}{6}$

As the two events are independent, we can use the multiplication law.

		RED DICE		
		P(1) = $\frac{2}{6}$	P(2) = $\frac{3}{6}$	P(3) = $\frac{1}{6}$
BLUE DIE	P(2) = $\frac{4}{6}$	P(2 and 1) $= \frac{4}{6} \times \frac{2}{6} = \frac{2}{9}$	P(2 and 2) $= \frac{4}{6} \times \frac{3}{6} = \frac{1}{3}$	P(2 and 3) $= \frac{4}{6} \times \frac{1}{6} = \frac{1}{9}$
	P(3) = $\frac{2}{6}$	P(3 and 1) $= \frac{2}{6} \times \frac{2}{6} = \frac{1}{9}$	P(3 and 2) $= \frac{2}{6} \times \frac{3}{6} = \frac{1}{6}$	P(3 and 3) $= \frac{2}{6} \times \frac{1}{6} = \frac{1}{18}$

c)
   i) P(2 and 1) = $\frac{2}{9}$
   ii) P(3 and 2) = $\frac{1}{6}$
   iii) P(double 2) = P(2 and 2) = $\frac{1}{3}$

EXERCISE 5.4

1) Bernice takes two cards without looking at them from a full pack of 52 cards. She replaces the first card before taking the second card.
   a) Calculate the probability that the first card taken is
      i) Red (R)
      ii) Black (B)
   b) Calculate the probability that the second card taken is
      i) A picture card (P)
      ii) Not a picture card (not P)
   c) Copy and complete the probability table
   d) Use your table to find the probability that Bernice takes
      i) First a red card and then a picture card
      ii) First a black card and then not a picture card.

> A picture card is a playing card which has a picture on it (usually the Jack, Queen and King).

		Second card	
		P(P) =	**P(not P) =**
First card	**P(R) =**		
	P(B) =		

2) In a soccer match, a team can win, draw or lose. United and City are due to play a match next week. Kenneth estimates the probabilities of the two teams winning, losing and drawing and writes them in a table.

		UNITED		
		P(w) = $\frac{1}{2}$	**P(d) =** $\frac{1}{6}$	**P(l) =** $\frac{1}{3}$
CITY	**P(W) =** $\frac{7}{10}$	P(W; w) =	P(W; d) =	P(W; l) =
	P(D) = $\frac{1}{5}$			
	P(L) = $\frac{1}{10}$			

   a) Copy and complete the table to show the probability of each combined event.
   b) What is the probability that City wins *and* United wins?
   c) What is the probability that City loses *and* United draws?
   d) Calculate the following for next week's match:
      i) P(both teams win or both teams lose)
      ii) P(City wins)
      iii) P(only one team loses)
      iv) P(only one team draws)

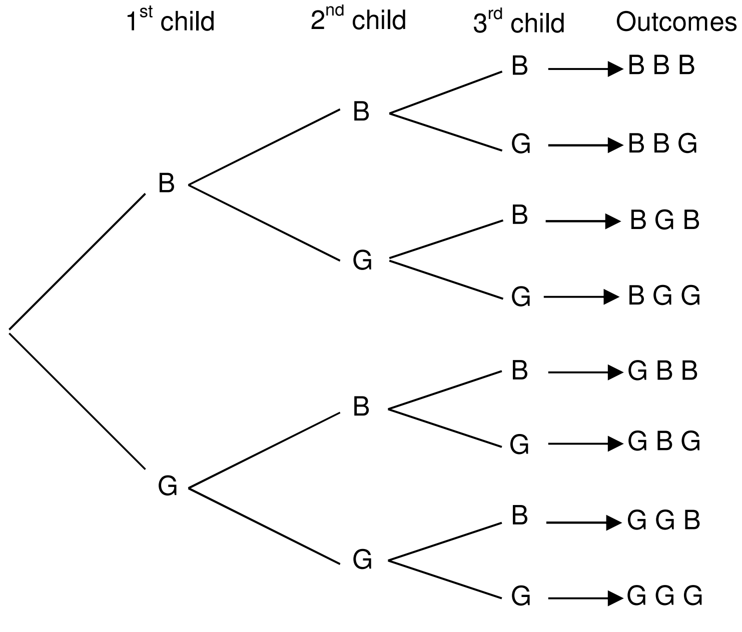
TREE DIAGRAMS

- ✓ When listing the combined outcomes of 2 or more activities, drawing a ***tree diagram*** is often the easiest method to use.

- ✓ We can use a ***tree diagram*** to work out combined outcomes of events that are either independent or dependent.

EXAMPLE 7
Suppose a family has three children. There are many combinations of boys (B) and girls (G) that can make up these three children.
a) Draw a tree diagram to find all the possible combinations of three children in the family.
b) Determine n(S) where S is the sample space.
c) Use the tree diagram to work out:
   i) The probability that all three children are girls
   ii) The probability that all three children are of the same gender
   iii) The probability that at least two of the three children are boys
   iv) The probability that one of the three children is a girl

<u>SOLUTION</u>:
a)



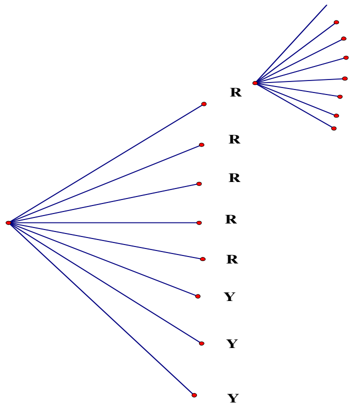
a) n(S) = 8
b)
   i) So, P(GGG) = $\frac{n(GGG)}{n(S)} = \frac{1}{8}$
   ii) P(same gender) = P(BBB or GGG) = $\frac{2}{8} = \frac{1}{4}$
   iii) P(at least two children are boys) = P(BBB or BBG or BGB or GBB) = $\frac{4}{8} = \frac{1}{2}$
   iv) P(one child is a girl) = P(BBG or BGB or GBB) = $\frac{3}{8}$

a) Events without replacement

- ✓ In some situations we work with probabilities where ***object are replaced*** before being selected again.

- ✓ When objects are replaced we can often end up with tree diagrams that can become very complicated (like the one in the following example)

Example:
Suppose a bag contains 5 red counters and 3 yellow counters of same shape and size.
A counter is taken from the bag, replaced and then another counter is taken.



- ✓ To simplify the drawing of a tree diagram like this, we can show probabilities are along the branches of the tree diagram, as shown in the next example.

EXAMPLE 8

A bag contains five red counters (R) and three yellow counters (Y) of the same shape and size.
A counter is picked at random and then **replaced** in the bag. Another counter is picked and replaced in the bag.

a) Are these events of picking a red counter (R) and picking a yellow counter (Y) independent? Explain.
b) Are these events of picking a red counter (R) and picking a yellow counter (Y) mutually exclusive?
c) Draw a tree diagram and use it to calculate the probability of each event occurring
d) Determine
   i) The probability that a red counter was picked first followed by a yellow counter
   ii) The probability that a yellow counter was picked first followed by another yellow counter
   iii) The probability that a yellow counter was picked first followed by a red counter
   iv) The probability that a red counter was picked and then a yellow OR a yellow counter and then a red

EXAMPLE 8 (continued)

SOLUTION:

a) Picking a red counter and picking a yellow counter are independent as the second event is not affected by the outcome of the first event. There are the same numbers of counters to choose from each time because the counters are replaced after being taken out.

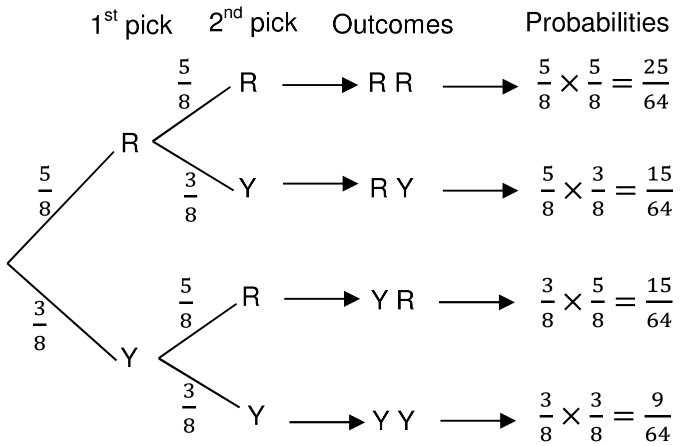
b) The events picking a red counter and picking a yellow counter are mutually exclusive – only one counter is picked at a time so the events cannot happen at the same time.

c) You can have one branch of the tree diagram for each counter, ***or*** we can write the probability of getting each colour counter on each branch.

→ There are 8 counters for each draw, five red and three yellow, so P(R) = $\frac{5}{8}$ and P(Y) = $\frac{3}{8}$. Write these probabilities along the branches.

→ To find out the probability of choosing a red and then choosing another red counter, work out P(R and R). To do this we follow the route along the branches and get P(R and R) = $\frac{5}{8} \times \frac{5}{8} = \frac{25}{64}$

→ We find the rest of the probabilities in the same way.



→ The four combined results cover all the possible outcomes, so their probabilities must add up to 1: $\frac{25}{64} + \frac{15}{64} + \frac{15}{64} + \frac{9}{64} = \frac{64}{64} = 1$. ***This is a useful check to make sure that you have worked out the probabilities correctly.***

d)

i) $\mathrm{P(R\ and\ Y)} = \frac{5}{8} \times \frac{3}{8} = \frac{15}{64}$

ii) $\mathrm{P(Y\ and\ Y)} = \frac{3}{8} \times \frac{3}{8} = \frac{9}{64}$

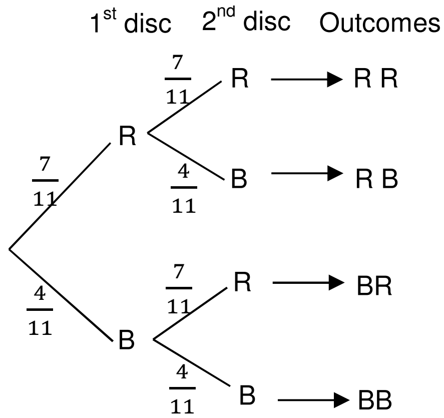
iii) $\mathrm{P(Y\ and\ R)} = \frac{3}{8} \times \frac{5}{8} = \frac{15}{64}$

iv) $\mathrm{P\big((R\ and\ Y)}\ \mathit{or}\ \mathrm{(Y\ and\ R)\big)} = \mathrm{P(R\ and\ Y)} + \mathrm{P(Y\ and\ R)}$

$$= \frac{15}{64} + \frac{15}{64} = \frac{30}{64}$$

EXERCISE 5.5

1) A coin is flipped and a dice is rolled.
   a) Are the two events independent? Explain.
   b) Draw a tree diagram to illustrate the situation where the first set of outcomes is H and T, and the second set of outcomes is 1, 2, 3, 4, 5 and 6.
   c) Manuel flips a coin and rolls a dice. Determine the probability as a common fractions in simplest form that Manuel gets:
      i) a H and a 4
      ii) a T and a 6
      iii) a H and an odd number
2) A jar contains 7 red discs and 4 blue discs.
   Two discs are selected, and then replaced. This means that the first disc is returned to the jar before the second disc is selected.
   a) Are these two events independent? Why or why not?
   b) A tree diagram is drawn to show the probabilities.



   Calculate the probability as a common fraction in simplest form that
      i) The two discs are red
      ii) The first disc is red and the second disc is blue
      iii) Only one of the discs is red
      iv) At least one of the discs is red.
3) A bag contains 3 red balls, 2 blue balls and 5 white balls.
   A ball is selected, its colour noted, and then it is replaced.
   A second ball is selected, its colour noted and then it is replaced.
   a) Are the two events independent? Explain.
   b) Draw a tree diagram to illustrate the situation. Fill in the probabilities along the branches of the tree diagram
   c) Find the probability as a common fraction in simplest form of:
      i) Selecting 2 blue balls
      ii) Selecting a blue ball and then a white ball
      iii) Selecting a red ball and then a blue ball
      iv) Selecting two balls of the same colour
      v) Selecting at least one blue ball

b) Events without replacement

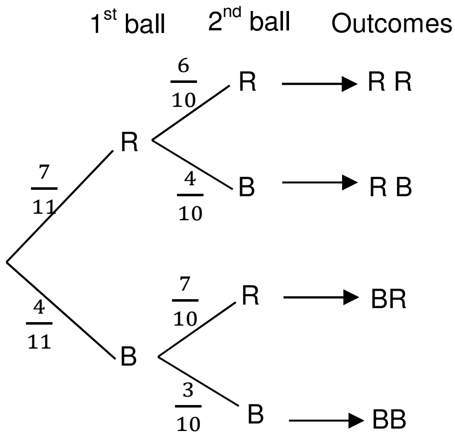
✓ In some situations, we need to be able to work out probabilities when ***there is no replacement.***

EXAMPLE 9

Consider a jar containing seven red balls and four blue balls.
Two balls are selected and are NOT replaced.
This means that the first ball is NOT returned to the jar before the second ball is selected.

a) Draw a tree diagram to illustrate the situation. Fill in the probabilities along the branches.
b) Use the tree diagram to find the probability that
   i) Both of the balls are red.
   ii) The first ball is red and the second ball is blue
   iii) At least one of the balls is red

SOLUTION:



a) There are seven red balls and four blue balls in the jar.
   - The probability of choosing a red ball the first time = $\frac{7}{11}$ and the probability of choosing a blue ball the first time = $\frac{4}{11}$
   - The balls are NOT replaced, so ***if a red ball is chosen the first time***, there are ten balls left in the jar with six of them being red and four of them blue.
     So the probability of choosing a red ball the second time = $\frac{6}{10}$ and the probability of choosing a blue ball the second time = $\frac{4}{10}$.
   - ***If a blue ball is chosen for the first time***, there are ten balls left in the jar with seven of them being red and three of them being blue.
     So the probability of choosing a red ball the second time = $\frac{7}{10}$ and the probability of choosing a blue ball the second time = $\frac{3}{10}$.

b)

i) P(R and R) = $\frac{7}{11} \times \frac{6}{10} = \frac{42}{110} = \frac{21}{55}$

ii) P(Rand B) = $\frac{7}{11} \times \frac{4}{10} = \frac{28}{110} = \frac{14}{55}$

iii) P(at least one R) = P((R and R) or (R and B) or (B and R))
= P(R and R) + P(R and B) + P(B and R)
$= \left(\frac{21}{55}\right) + \left(\frac{14}{55}\right) + \left(\frac{4}{11} \times \frac{7}{10}\right) = \frac{21}{55} + \frac{14}{55} + \frac{14}{55} = \frac{49}{55}$

<u>OR</u> P(at least one R) = 1 – P(B and B) = $1 - \left(\frac{4}{11} \times \frac{3}{10}\right) = 1 - \frac{6}{55} = \frac{49}{55}$

EXERCISE 5.6

1) Nomvula owns a collection of 30 CDs, of which 5 are gospel music (G). She selects one CD at random and then selects another CD ***without replacing the first CD***.
   a) How many CDs are NOT gospel music (NG)?
   b) Draw a tree diagram to illustrate DRAW 1: select a CD without replacing it and DRAW 2: select a second CD. Add in the probabilities along the branches.
   c) Nomvula selects two CDs at random. Determine the probability as a percentage correct to 1 decimal place:
      i) That both CDs are gospel music.
      ii) That the first CD is gospel music and the second one is not gospel music.

2) A bag contains one blue cube (B), two green cubes (G) and three red cubes (R). One cube is taken at random from the bag. Its colour is recorded and ***it is not replaced***. A second cube is then taken at random.
   a) Draw a tree diagram that represents all the possible outcomes in this situation. Fill in the probabilities along the branches.
   b) Use the tree diagram to determine probabilities as a common fraction in simplest form that
      i) two blue cubes are taken
      ii) two red cubes are taken
      iii) at least one green cube is taken
      iv) at most 1 red cube is taken.

3) Patrice enters both the 100 m sprint (S) and the long jump (J) at a school sports competition. The probability that he wins the 100 m sprint (SW) is 0,4. The probability that he wins the long jump (JW) is 0,7. The two events are independent.
   a) What is the probability (written as a decimal to 2 decimal places) that Patrice:
      i) Does not win the 100 m sprint (SL)
      ii) Does not win the long jump (JL)?
   b) Draw a probability tree diagram to show the situation.
   c) Use your tree diagram to determine the probability that, at the school event:
      i) Patrice wins both competitions
      ii) Patrice wins the 100 m sprint but does not win the long jump
      iii) Patrice does not win either competition
      iv) Patrice wins only one of the competitions.

VENN DIAGRAMS SHOWING THREE EVENTS

- ✓ We can solve probability problems involving ***three events*** using Venn diagrams.

EXAMPLE 10

During a survey at a certain school, 350 learners were asked about their favourite sport. It was found that

180 like Tennis (T)
180 like Rugby (R)
240 like Cricket (C)
78 like all three of these sports
60 like Tennis and Cricket but not Rugby
54 like Rugby and Cricket but not Tennis
30 like Tennis and Rugby but not Cricket

a) Calculate how many learners:
   i) Like Tennis only
   ii) Like Cricket only
   iii) Do not like any of the three sports.
b) Represent the given information in a Venn diagram
c) Calculate the probability (as a percentage correct to 2 decimal places) that one of these learners, chosen at random
   i) Likes rugby only
   ii) Likes none of these sports

SOLUTION:

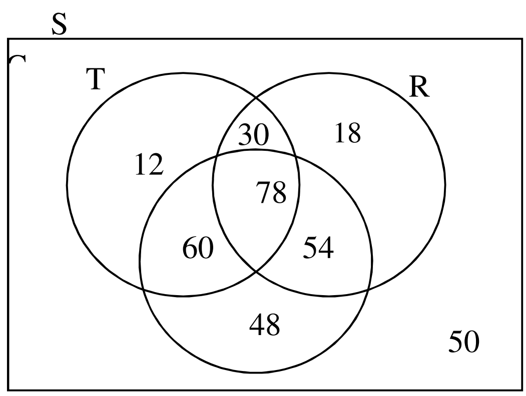
a)

i) 180 like Tennis, 78 like all 3 sports, 60 like playing Tennis and Cricket but not Rugby, and 30 like Tennis and Rugby but not Cricket.
So number of learners who like Tennis only = 180 – (78 + 60 + 30)
= 12

ii) 240 like Cricket, 78 like Cricket, Rugby and Tennis, 60 like Tennis and Cricket, 54 like Rugby and Cricket but not Tennis.
So number of learners who like Cricket only = 240 – (78 + 60 + 30) = 48

iii) The number of learners who do not like any of the three sports
= 350 – (12 + 30 + 18 + 78 + 54 + 48 + 60) = 50

b)



c)

i) P(Rugby only) = $\frac{n(Rugby\ only)}{n(sample\ set)} = \frac{18}{350}$ = 0,051 ... ≈ 0,05

ii) P(likes none of the 3 sports) = $\frac{n(likes\ none\ of\ the\ 3\ sports)}{n(sample\ set)} = \frac{50}{350}$ = 0,142 ... ≈ 0,14

EXAMPLE 11

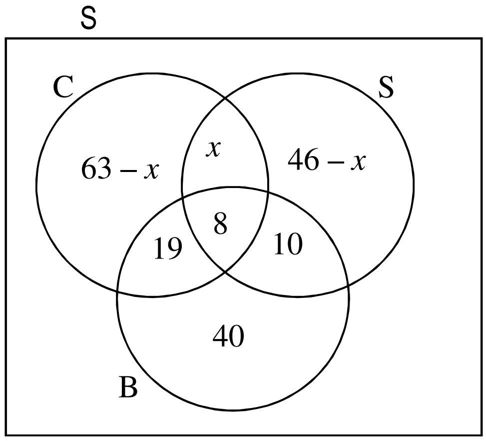
During the 2009 Census@School, 174 learners were asked about their favourite sport. It was found that:

90 prefer chess (C)
64 prefer swimming (S)
77 prefer boxing (B)
8 prefer all three sports: chess, swimming and boxing
18 prefer swimming and boxing
27 prefer chess and boxing
26 do not like any of the three sports
x prefer chess and swimming but don't like boxing

a) Draw a Venn diagram to illustrate the results of this survey
b) Calculate the value of x, the number of learners who prefer chess and swimming but don't like boxing.
c) What is the probability that a randomly selected learner prefers two types of sporting codes? Write your answer as a percentage correct to 1 decimal place.

<u>SOLUTION</u>:

a) STEP 1:
Fill 8 into the intersection of all 3 events.
STEP 2:
18 prefer swimming and boxing, so 10 prefer swimming and boxing but don't like chess
STEP 3:
27 prefer chess and boxing, so 19 prefer chess and boxing but don't like swimming
STEP 4:
x prefer chess and swimming but do not like boxing



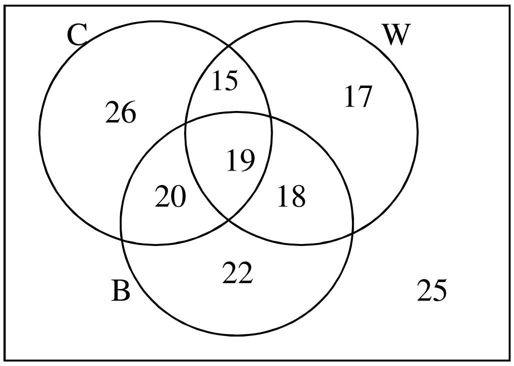
STEP 5:
77 prefer boxing, so the number who prefer boxing only = 77 – (19 + 8 + 10) = 40
64 prefer swimming, so the number who prefer swimming only
= 64 – (x + 8 + 10) = 46 – x
90 prefer chess, so the number who prefer chess only = 90 – (x + 8 + 19) = 63 – x

b) Add up all the values and set them equal to the total number of learners surveyed.
$(63 - x) + x + (46 - x) + 19 + 8 + 10 + 40 = 174$
$186 - x = 174$
$x = 12$
So 12 learners play chess and swim but do not box.

c) Number of learners who prefer two types of sporting codes = 19 + 8 + 12 + 10 = 49
P(prefer two types) = $\frac{n(prefer\ two\ types)}{n(sample\ set)} \frac{49}{174} \approx 0{,}281\,6\ldots \approx 28{,}2\%$

EXERCISE 5.7

1) The Venn diagram shows the numbers of learners who came to school using different modes of transport. Some learners walk (W), others ride a bicycle (B) and others come by car (C).
   a) Are these three events independent?
   b) Determine n(S) if S is the sample set
   c) Use the information on the Venn diagram to find the probability (as a common fraction in simplest form) that one of these learners, picked at random
      i) Walks
      ii) Comes by car
      iii) Uses a bicycle
      iv) Walks only
      v) Does not walk
      vi) Walks or uses a bicycle



2) A survey of 80 people at a local food outlet indicated their meal preferences.
   - 44 like Pap and Wors (P)
   - 33 like Burgers (B)
   - 39 like Fried Chicken (FC)
   - x people like Pap and Wors and Burgers but not Fried Chicken
   - 23 like Pap and Wors as well as Fried Chicken
   - 19 like Burgers and Fried Chicken
   - 9 like Pap and Wors, Burger and Fried Chicken
   - 69 like at least one of these meals

   a) How many people did not like any of these meals?
   b) Draw the Venn diagram to represent the meal preferences of these learners.
   c) Determine the value of x.
   d) What is the probability, written as a percentage, that one of these people, chosen at random,
      i) Like Burgers?
      ii) Likes Fried Chicken only?
      iii) Like Burgers and Fried Chicken?
      iv) Likes Pap and Wors or likes Fried Chicken?

REFERENCES

Banks T and Alcorn D (2004) *Mathematics for AQA GCSE*, Causeway Press
Charles, E. (1986) *Advanced Mathematical Concepts*. Merrill Publishing Co.
Graham D and C (1996) *Mainstream Maths for GCSE*, Macmillan
The Answer Series Grade 12 Mathematics Paper 3
Speed, B. (1997) *Higher Mathematics for GCSE*, Collins Educational
Statistics South Africa (2010); *Census@School Results 2009*
2011 Census in South Africa
www.onlinemathlearning.com/probability-venn-diagram.html
http://www.haesemathematics.com.au/samples/ibmyp4_10.pdf

6

Grade 12 Probability

In this chapter you:
- Use Venn diagrams to determine the probability of events happening
- Identify independent and dependent events
- Identify mutually exclusive events
- Identify complementary events
- Use tree diagrams to determine the probabilities of independent and dependent events happening
- Use two-way contingency tables to calculate probabilities and test for independence
- Apply counting principles to solve probability problems

WHAT YOU LEARNED ABOUT PROBABILITY IN GRADE 11

In Grade 11 you covered the following probability concepts:
- *The addition rule for mutually exclusive events*
- *The complementary rule: P(not A) = 1 – P(A)*
- *The probability identity: P(A or B) = P(A) + P(B) – P(A and B)*
- *Dependent and independent events*
- *The product rule for independent events: P(A and B) = P(A) $\times$ P(B)*
- *The use of Venn diagrams to solve probability problems*
- *The use of tree diagrams to determine the probability of consecutive or simultaneous events which are not necessarily independent*

VENN DIAGRAMS

✓ Venn diagrams are a graphical way of depicting sets and are very useful when you are solving probability questions.

EXAMPLE 1

Cards numbered from 1 to 15 are put into a box. A card is randomly selected from the box. Consider the following events:

- Event E: an even number is drawn
- Event F: a factor of 14 is drawn.
- Event M: a multiple of 7 is drawn.

a) Draw a Venn Diagram to represent events E, F and M

b) Calculate the following, giving your answer as a percentage correct to the nearest whole number:

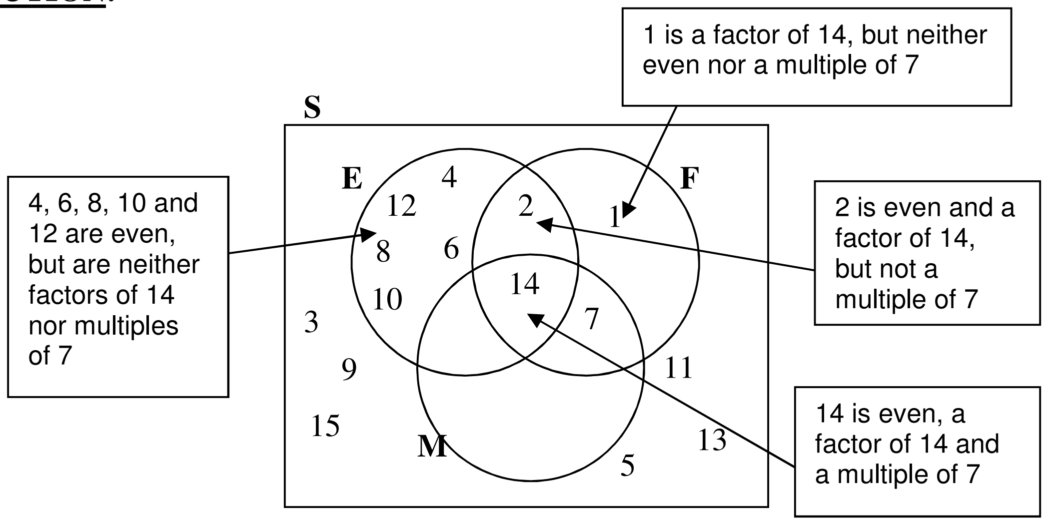
  i) P(F).

  ii) Calculate P(M and F)

  iii) Calculate P(M or F or E)

<u>SOLUTION</u>:

a)



b) There are 15 possible outcomes, so n(S) = 15

i) F = {1 ; 2 ; 7 ; 14}, so n(F) = 4

$P(\mathrm{F}) = \frac{n(\mathrm{F})}{n(\mathrm{S})} = \frac{4}{15} \approx 0{,}2667 \approx 27\%$

ii) The elements in M and F = {7 ; 14}, so n(M and F) = 2

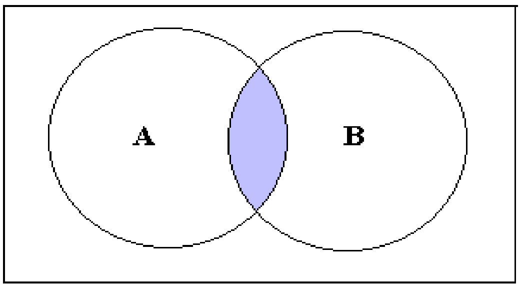
$\mathrm{P(M\ and\ F)} = \frac{n(\mathrm{M\ and\ F})}{n(\mathrm{S})} = \frac{2}{15} \approx 0{,}1333 \approx 13\%$

iii) M or F or E = {1; 2; 4; 6; 7; 8; 10; 12; 14}, so n(M or F or E) = 9

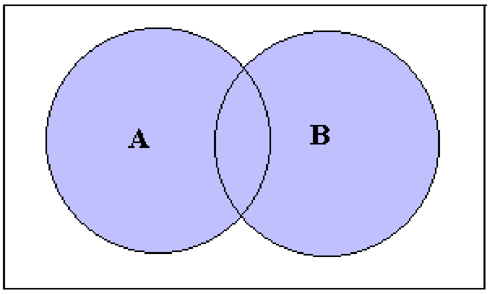
$\mathrm{P(M\ or\ F\ or\ E)} = \frac{n(\mathrm{M\ or\ F\ or\ E})}{n(\mathrm{S})} = \frac{9}{15} = \frac{3}{5} = 0{,}6 = 60\%$

✓ A Venn diagram can help you find P(A *and* B) and P(A *or* B)

The shaded area below shows A and B.

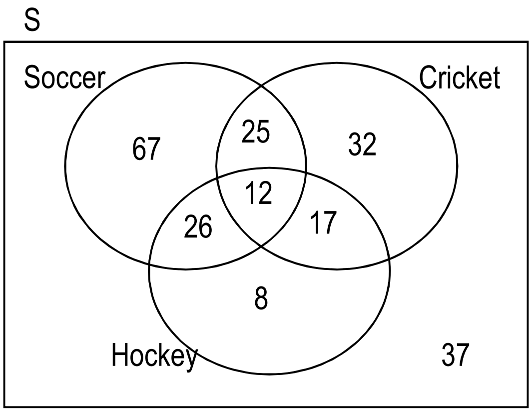


The shaded area below shows A or B.



EXAMPLE 2

The Venn diagram shows the number of learners in a school who play soccer, cricket and hockey.

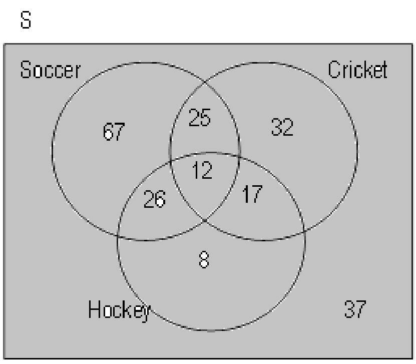


a) Determine n(S), where S is the sample set.
b) How many learners play soccer, cricket *and* hockey?
c) How many learners play cricket *and* hockey?
d) How many learners play cricket *or* hockey?
e) What is the probability that one of these learners selected at random plays cricket *or* hockey?

<u>SOLUTION</u>:

a) This shaded section shows you all the learners in the school:

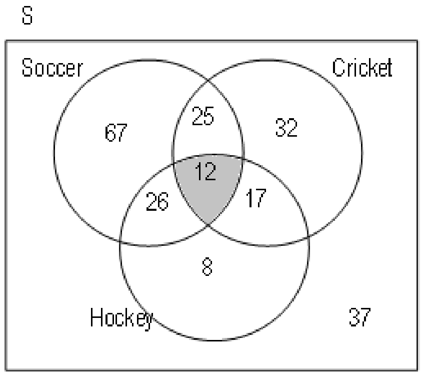
n(S) = 67 + 25 + 32 + 26 + 12 + 17 + 8 + 37
= 224



<u>EXAMPLE 2</u> (continued)

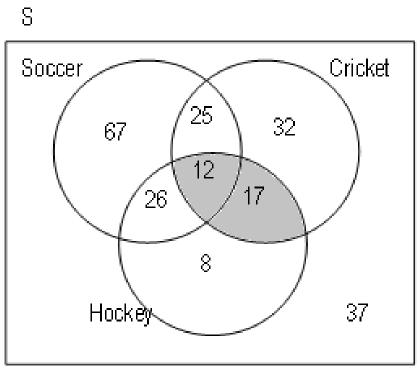
b) The shaded section shows the learners who play soccer, cricket *and* hockey.

n(soccer, cricket *and* hockey) = 12



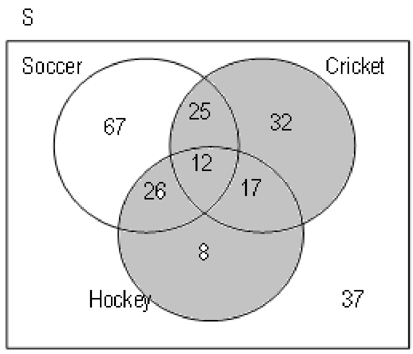
c) The shaded section shows you all the learners who play cricket *and* hockey:

n(cricket *and* hockey) = 12 + 17 = 29



d) The shaded section shows you all the learners who play cricket *or* hockey:

n(cricket *or* hockey)
= 25 + 32 + 26 + 12 + 17 + 8
= 120

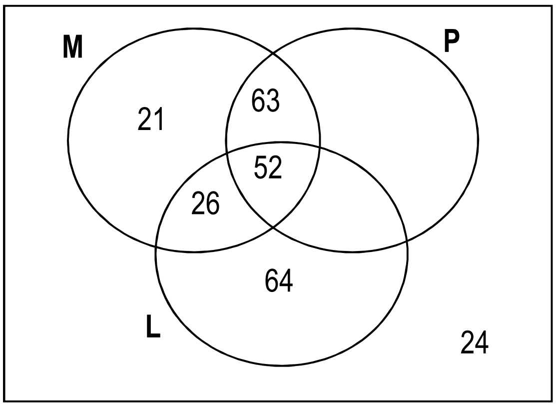


e) P(cricket or hockey) = $\frac{\text{n(cricket or hockey)}}{\text{n(S)}} = \frac{120}{224} = \frac{15}{28} \approx 0{,}5357 \approx 54\%$

This means that the chances of a learner selected at random who plays either cricket or hockey is just over half.

EXERCISE 6.1

1) There are 250 Grade 12 learners at Greenwood Secondary School. The Venn diagram below shows the number of learners in the school who take Mathematics (M), Physical Sciences (P) and Life Sciences (L).



   a) How many learners take Mathematics?
   b) How many learners take Mathematics, Physical Sciences and Life Sciences?
   c) What is the probability, as a fraction in simplest form, that one of these Grade 12 learners, selected at random takes
      i) Physical Sciences *or* Life Sciences?
      ii) Physical Sciences *and* Life Sciences?

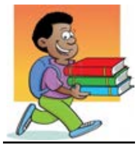
2) 120 shoppers are surveyed at a till to determine whether they buy full-cream (FC), low fat (LF) or skimmed milk (S). They determined that:
   - 75 shoppers bought full cream milk
   - 41 shoppers bought low fat milk
   - 7 shoppers bought skimmed milk
   - 16 shoppers bought full cream and low fat milk
   - 3 shoppers bought low-fat and skimmed milk
   - 1 shoppers bought full-cream and skimmed milk
   - 1 shopper bought full-cream, low-fat and skimmed milk.

   a) Draw a Venn diagram to illustrate the given data.
   b) How many shoppers did not buy any milk?
   c) What is the probability, written as a fraction in simplest form, that one of these shoppers, selected at random, bought low-fat or skimmed milk?

INDEPENDENT AND DEPENDENT EVENTS

- ✓ Independent events are events where the outcomes of the second event is not affected by the outcomes of the first event.

  For example:
  Three Grade 12 learners (Sipho, Mpho and Jacob) are randomly selected before school and after school to carry books for a teacher.

  
  www.123rf.com

  The event of Sipho being selected to carry books before school will not affect the probability of him being selected to carry books after school.

  The events of carrying books before school and carrying books after school are ***independent events***.

- ✓ Dependent events are events that are not independent. The probability of independents events DO affect each other.

  For example:
  The probability of taking an umbrella to work is dependent on the probability that it will rain.

  
  www.flickr.com

- ✓ When events are independent you can work out the probability that they **both** occur by multiplying the probabilities.

  i.e. For events A and B, $\boldsymbol{P(A \text{ and } B) = P(A) \times P(B)}$

EXAMPLE 3
Mmatladi and Eva run a 1 km race. The probability that Mmatladi finishes the race within 5 minutes is 0,67 and the probability that Eva finishes the race within 5 minutes is 0,82.

a) Are the events of Mmatladi finishing the race within 5 minutes and Eva finishing the race within 5 minutes independent?
b) Who is a better runner, Mmatladi or Eva?
c) What is the probability that both Mmatladi and Eva finish the race within 5 minutes?

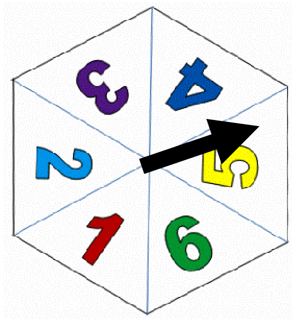
<u>SOLUTION</u>:

a) The two events are independent. Mmatladi's finishing time will not affect Eva's time.
b) Eva is a better runner. The probability that she finishes the race within 5 minutes is greater than Mmatladi's probability.
c) P(both finish within 5 min)
   = P(Mmatladi finishes within 5 minutes *and* Eva finishes within 5 minutes)
   = P(Mmatladi finishes within 5 min) × P(Eva finishes within 5 min)
   = 0,67 × 0,82
   ≈ 0,5494
   ≈ 55%

EXERCISE 6.2

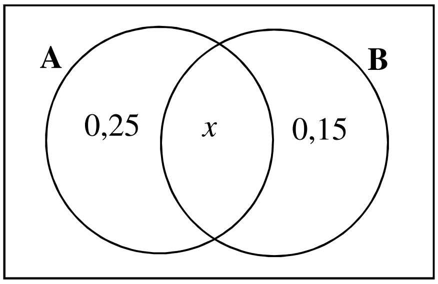
1) Are the events described below dependent or independent? Give a reason for your answer.
   a) Randomly drawing a red card from a standard deck, replacing it and then randomly drawing a heart.
   b) Randomly picking a red counter from a cup containing 3 red counters, 2 blue counters and 1 green counter, not returning it and then randomly selecting a blue counter from the cup.
   c) Running in a thunderstorm and being struck by lightning.
   d) Swimming in the ocean and being bitten by a shark
   e) Watching television and wearing a red T-shirt

2) The fair spinner shown alongside is spun twice.
   a) Is the result of the second spin dependent or independent of the first spin?
   b) Calculate the probability that the spinner will land on **4** twice.



www.norma-pages.blogspot.com

3)



Events A and B are independent.
The Venn diagram shows that

- P(A) = $x + 0{,}25$
- P(B) = $x + 0{,}15$.
- P(A and B) = x.

Form an equation and work out the possible value(s) of x.

MUTUALLY EXCLUSIVE EVENTS

✓ Events that are ***mutually exclusive*** cannot happen at the same time.

For example:

- A learner cannot live in Mpumalanga and in Northern Cape at the same time.
- A learner cannot be in Grade 10 and in Grade 12 at the same time.

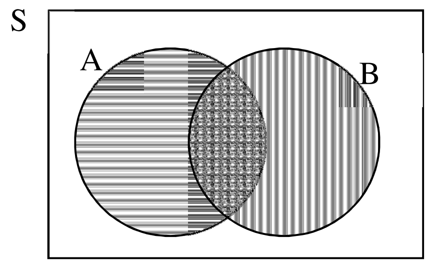
✓ Events that are ***not mutually exclusive*** could occur at the same time.

For example:
A learner could live in Mpumalanga and be in Grade 10.

✓ When events are ***not mutually exclusive*** we can find the probability of one or the other occurring by using the identity:

P(A or B) = P(A) + P(B) – P(A and B)

As can be seen by looking at a Venn diagram, we need to subtract P(A and B) from P(A) + P(B), otherwise it will be included twice.



✓ When events are ***mutually exclusive*** we can find the probability of one event *or* the other occurring by adding the probabilities of each happening.

i.e. For events A and B, P(A or B) = P(A) + P(B), because P(A and B) = 0.

The Addition Rule for mutually exclusive events is really just a special case of the Addition Rule for not mutually exclusive events.
Since A and B cannot occur together, P(A and B) = 0.

<u>Remember:</u>

- ***Independent events*** are events where one does not affect the other. For example, the probability of winning a race is not affected by the probability of wearing a red T-shirt.

- ***Mutually exclusive events*** are events that do not share any elements. The probability of wearing a red T-shirt and the probability of wearing a blue T-shirt are mutually exclusive because a whole T-shirt cannot be red and blue at the same time.

EXAMPLE 4

The data in this table gives the number of Primary Schools and Secondary Schools in Education Districts in Gauteng.

District	Primary School	Secondary School	Total
EKURHULENI EAST	2 356	1 990	4 346
EKURUHLENI WEST	3 240	4 709	7 949
GAUTENG NORTH	112	289	401
GAUTENG WEST	813	1 735	2 548
JOHANNESBURG EAST	1 266	3	1 269
JOHANNESBURG NORTH	861	2 131	2 992
JOHANNESBURG SOUTH	2 607	2 171	4 778
JOHANNESBURG WEST	1 783	1 691	3 474
SEDIBENG EAST	581	395	976
SEDIBENG WEST	1 400	1 377	2 777
TSHWANE NORTH	2 555	2 416	4 971
TSHWANE SOUTH	2 312	4 029	6 341
GAUTENG TOTAL	20 441	27 034	47 475

a) Are the events randomly selecting a school in Ekurhuleni East and randomly selecting a school in Johannesburg North mutually exclusive? Explain your answer.

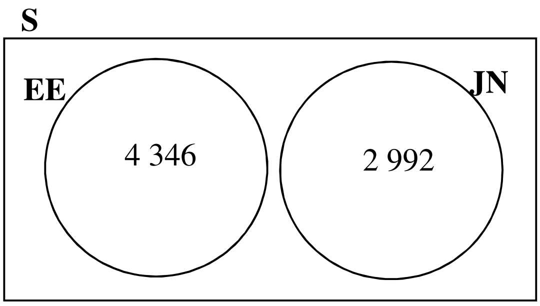
b) What is the probability of randomly selecting a school that is in Ekurhuleni East or in Johannesburg North?

c) Are the events randomly selecting a Secondary School and randomly selecting a school from Tshwane North mutually exclusive? Explain your answer.

d) What is the probability of randomly selecting a school in Gauteng that is a Secondary School or a school in Tshwane North?

<u>SOLUTION</u>:

a) The events randomly selecting a school in Ekurhuleni East and randomly selecting a school in Johannesburg North are mutually exclusive. A school cannot be located in two districts.
This is shown in the Venn diagram:



b) P(EE or JN) = P(EE) + P(JN)

$= \frac{n(EE)}{n(S)} + \frac{n(JN)}{n(S)}$

$= \frac{4\ 346}{47\ 475} + \frac{2\ 992}{47\ 475}$

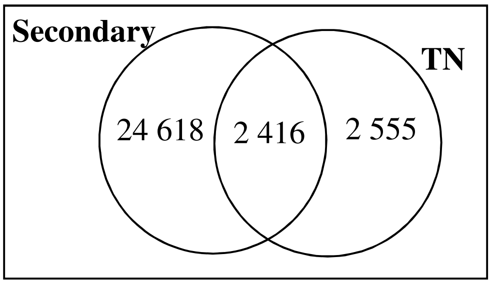
$= \frac{2\ 446}{15\ 825}$

$\approx 0{,}1546$

$\approx 15\%$

<u>EXAMPLE 4</u> (continued)

c) The two events 'randomly selecting a Secondary School' and 'randomly selecting a school in Tshwane North' are not mutually exclusive. There are schools that are both Secondary Schools and in Tshwane North.
This is shown in the Venn diagram.



d) P(Sec or TN) = P(Sec) + P(TN) – P(Sec and TN)

$$= \frac{\text{n(Sec)}}{\text{n(S)}} + \frac{\text{n(TN)}}{\text{n(S)}}$$

$$= \frac{27\,034}{47\,475} + \frac{4\,971}{47\,475} - \frac{2\,416}{47\,475}$$

$$= \frac{29\,589}{47\,475}$$

$$= \frac{9\,863}{15\,825}$$

$$\approx 0{,}6233$$

$$\approx 62\%$$

EXERCISE 6.3

1) Are the events described below mutually exclusive or not mutually exclusive? Give a reason for your answer
   a) A red counter is taken from a cup containing 3 red counters, 2 blue counters and 5 yellow counters, and then a blue counter is taken.
   b) A red card is drawn from a standard deck of cards and then a King is drawn.

2) You select a card at random from a standard deck of cards. What is the probability that the card is a 3 or a 7? Give your answer as a common fraction in simplest form.

3) You select a card at random from a standard deck of cards. What is the probability that the card is 3 or red? Give your answer as a common fraction in simplest form.

4) Sibongile throws a biased dice. *A biased dice is an unfair dice. Each number is NOT equally* likely. On this dice the probabilities of throwing each number are shown in the table below:

Number on dice	**1**	**2**	**3**	**4**	**5**	**6**
Probability	0,08	0,04	0,125	0,167	0,33	0,258

   a) On which number is the dice most likely to land?
   b) Are the events of the dice landing on 5 and the dice landing on 6 mutually exclusive? Explain why or why not.
   c) What is the probability, as a decimal correct to 2 decimal places, that Sibongile throws:
      i) a 5 or 6?
      ii) an even number?

5) 250 people were asked whether they watched rugby and/or cricket on television. 180 people watched rugby, 99 watched cricket and 87 watched both rugby and cricket. What is the probability, correct to the nearest full percentage, that one of these people, chosen at random, watched rugby or cricket?

COMPLEMENTARY EVENTS

- ✓ ***Complementary events*** are events that cannot happen at the same time.

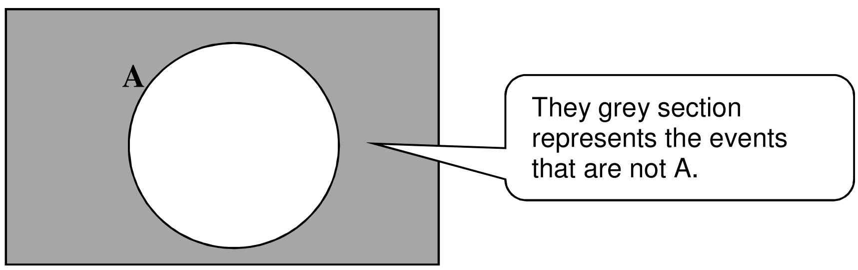
  Examples of complementary events are:
  - *Rolling a 5 on a dice and not rolling a 5 on a dice*. It is certain that either the dice will land on 5 or not on 5.
  - *Randomly selecting a learner who lives in a house and randomly selecting a learner who does not live in a house*. It is certain that a learner will either live in a house or not in a house.
  - *Randomly selecting a learner who is a girl and randomly selecting a learner who is not a girl (i.e. a boy).* It is certain that the learner selected will be either a boy or a girl.

  Examples of events that are NOT complementary:
  - *Rolling a 5 on a dice and rolling a 3 on a dice*. It is not certain that the dice will only land on 5 or 3 as it could also land on 1, 2, 4 or 6. The events of rolling a 5 and rolling a 3 are mutually exclusive, but not complementary.
  - *Randomly selecting a learner who lives in a house and randomly selecting a learner who lives in a traditional dwelling*. It is not certain that a learner lives only in a house or a traditional dwelling. They may also live in a flat, townhouse, a room at the back of a house, etc.

- ✓ When events are complementary, the event either happens (A) or does not happen (not A). It is *certain* that they will happen, or not happen so P(A) + P(not A) = 1.

  Another way to say this is P(not A) = 1 – P(A) or P(A) = 1 – P(not A).

- ✓ Complementary events can also be shown on a Venn Diagram:



EXAMPLE 5

Out of 3 330 976 learners surveyed in 2009 Census@Schools, 199 610 lived in an informal dwelling. What is the probability that a learner selected at random did not live in an informal dwelling?

SOLUTION:

P(informal dwelling) = $\frac{199\ 610}{3\ 330\ 976}$

P(not informal dwelling) = 1 – P(informal dwelling)

$$= 1 - \frac{199\ 610}{3\ 330\ 976}$$

$$= \frac{3\ 131\ 366}{3\ 330\ 976}$$

$$\approx 0{,}9401$$

$$\approx 94\%$$

It is very likely that a learner selected at random will not live in an informal dwelling.

EXAMPLE 6

a) What is the probability of throwing ***no sixes*** when rolling a dice four times?
b) What is the probability of throwing ***at least one six*** in four rolls of a regular fair six-sided dice?

SOLUTION:

Each roll of the dice is independent of the roll before.

a) In one roll, P(6) = $\frac{1}{6}$

So, P(no 6) = $1 - P(6) = 1 - \frac{1}{6} = \frac{5}{6}$

In four rolls, P(no 6 and no 6 and no 6 and no 6)

=P(no 6) × P(no 6) × P(no 6) × P(no 6)

$$= \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6}$$

$$= \frac{625}{1\ 296}$$

$$\approx 48\%$$

"At least one" means you could get 1, 2, 3 or 4 sixes when you roll the dice four times.

b) When you throw a dice, either you will get a 6 or you won't get a 6.

When you throw a dice four times, either you get no 6 or you get ***at least one*** 6.

i.e. The events 'get no 6' and 'get at least one 6' are complementary.

P(at least one 6 in 4 rolls of the dice)

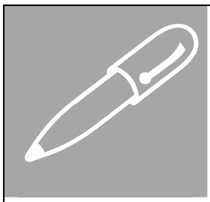
= 1 – P(no 6 in 4 rolls of the dice)

$$= 1 - \frac{625}{1\ 296}$$

$$= \frac{671}{1\ 296}$$

$$\approx 52\%$$

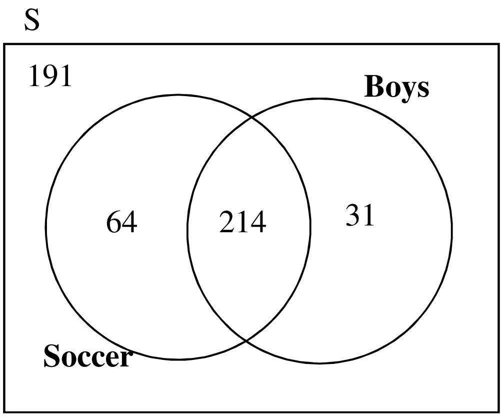
There is a greater chance that you will get at least one 6 than of getting no 6 when you throw a dice four times.



EXERCISE 6.4

Work correct to 2 decimal places where necessary.

1) If P(A) = 0,35 and P(B) = 0,22, find
   a) P(not A)
   b) P(A or B) if A and B are mutually exclusive events.
   c) P(A or B) if A and B are independent events.
2) The Venn diagram below illustrates the number of learners in a school who play soccer and those who are boys.



   Use the Venn diagram to determine:
   a) The number of learners in the school.
   b) The number of girls in the school.
   c) The number of learners who do not play soccer.
   d) The probability that one of these learners, chosen at random:
      i) is a boy and plays soccer.
      ii) is a girl or plays soccer.

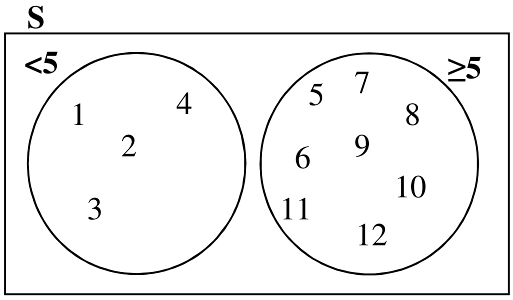
TREE DIAGRAMS

- ✓ Tree diagrams are very useful for solving probability problems where events are independent and where events are dependent.

EXAMPLE 7
In a game, two numbers from 1 to 12, have to be randomly selected from the numbers in a box. A number is drawn, put back and then the second number is drawn.
a) Find the probability that both numbers are less than 5.
b) Find the probability that at least one of the numbers is less than 5.

SOLUTION:
There are 12 possible outcomes each time a number is drawn, so n(S) = 12.
There are 4 favourable outcomes that are less than five: {1; 2; 3 and 4}.
There are 8 possible outcomes that are five or more than 5: {5; 6; 7; 8; 9; 10; 11; 12}.

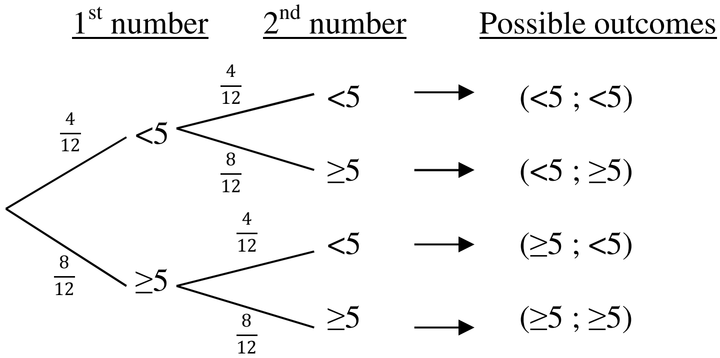
- The possible outcomes can be represented in a Venn diagram as follows:



Key:	
Numbers less than 5	**<5**
Number 5 and numbers more than 5	**≥5**

Notice that the sets are disjoint. This means that they are mutually exclusive. A number cannot be less than 5 and 5 or more than 5 at the same time.

- You can also construct a tree diagram of the outcomes of the first number selected and the second number selected. The second number selected is independent of the first number selected because the first number is returned to the box. This means the denominators on each branch of the tree diagram will be 12 each time.



a) P(<5 and <5) = $\frac{4}{12} \times \frac{4}{12} = \frac{1}{9} \approx 0{,}1111$

EXAMPLE 7 (continued)

b) *Method 1*:

P(at least one number less than 5)

= P((<5 and <5) OR (<5 and ≥5) OR (≥5 and <5))

= P(<5 and <5) + P(<5 and ≥5) + P (≥5 and <5)

$= \left(\frac{4}{12} \times \frac{4}{12}\right) + \left(\frac{4}{12} \times \frac{8}{12}\right) + \left(\frac{8}{12} \times \frac{4}{12}\right)$

$= \frac{16}{144} + \frac{32}{144} + \frac{32}{144}$

$= \frac{80}{144}$

$= \frac{5}{9}$

≈ 0,55561

Method 2:

The events 'getting at least one number less than 5' and 'getting no numbers less than 5' (i.e. the event of getting both numbers more than or equal to 5) are complementary.

P(≥5 and ≥5) = $\frac{8}{12} \times \frac{8}{12} = \frac{4}{9}$

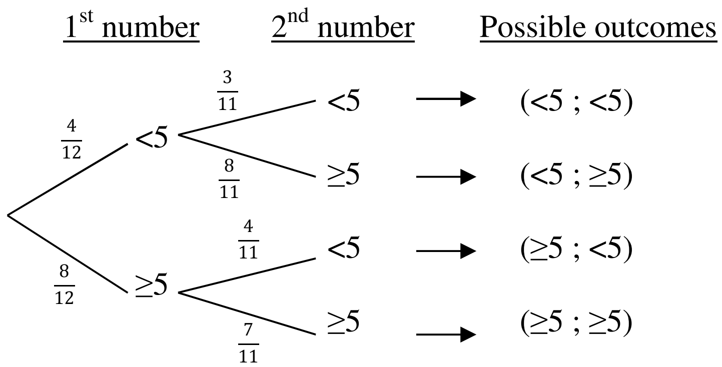
P(at least one number less than 5) = 1 – P(≥5 and ≥5) = $1 - \frac{4}{9} = \frac{5}{9} \approx 0{,}5556$

EXAMPLE 8

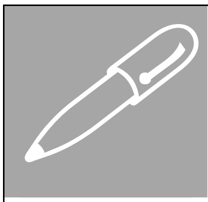
In a game, two numbers from 1 to 12, are to be randomly selected.
A number is drawn. It is *not* put back and then the second number is drawn.
Find the probability that both numbers are less than 5.

SOLUTION:

There are 12 possible outcomes the first time a number is drawn and 11 possible outcomes the second time a number is drawn because the first number is not replaced. This means that the second number selected depends on the first number selected.

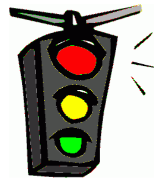


P(<5 ; <5) = $\frac{4}{12} \times \frac{3}{11} = \frac{1}{11} \approx 0{,}0909$



EXERCISE 6.5

- Give each of the answers in this exercise as a percentage correct to the nearest whole number.



jerrin-it.blogspot.com

1) Vusi rides his bike to school. He has to go through two intersections with robots (traffic lights). He has to stop at the first robot 55% of the time. He has to stop at the second robot 20% of the time. The robots are independent of each other.
   a) What percentage of the times does he
      i) Not stop at the first robot?
      ii) Not stop at the second robot?
   b) Draw a tree diagram to illustrate the situation and fill in the values along the branches.
   c) What is the probability that, on any particular day, he will have to stop
      i) At both of the robots?
      ii) At one of the robots?

2) The probability that a tennis player has no injuries (NI) is 0,7.
   The probability that she will win a game (W) if she has no injuries is 0,9.
   When the tennis player has injuries (I) the probability of her winning becomes 0,45.
   a) Draw a tree diagram to illustrate the situation
   b) Calculate the probability of her winning her next tennis game.

3) The South African Weather Bureau predicts that the days in Johannesburg in summer could be:
   ✓ HOT or
   ✓ WET or
   ✓ HOT and WET.
   The chance that any day will be hot (H) is 70%.
   The chance that any day will be wet (W) is 43%.
   The chance that a day will be hot and wet is 27%.
   Without the use of a tree diagram calculate the probability of any day in summer being HOT or WET.

4) Raesetja and Makoena have entered a Maths quiz. They are asked two questions.
   The probability that they get the first answer correct (C) is 0,8.
   If the first answer is correct, the probability of them getting the next answer correct is 0,7.
   However, if they get the first answer wrong (W), the probability of them getting the next answer correct is only 0,4.
   a) Draw a tree diagram to illustrate the situation.
   b) Determine the probability that they get the second answer correct.

5) A drawer contains twenty envelopes.
   Eight of the envelopes each contain five blue and three red sheets of paper.
   The other twelve envelopes each contain six blue and two red sheets of paper.
   One envelope is chosen at random and a sheet of paper is chosen at random from it.
   a) Draw a tree diagram first showing the possible outcomes when selecting the envelopes and then the possible outcomes when selecting the sheets of paper.
   b) What is the probability that this sheet of paper is red?

CONTINGENCY TABLES

✓ A contingency table represents the frequencies of events involving two or more variables. The categories of the variables are listed along the top and down the side.

EXAMPLE 9
The following data was collected by Census@Schools 2009. The data gives the ages of learners in each grade in the Eastern Cape.

	Age (in years)									
	13	14	15	16	17	18	19	20	21	Total
Grade 8	478	600	316	78	0	0	0	0	0	1 472
Grade 9	155	696	1 019	197	0	0	0	0	0	2 067
Grade 10	0	459	1 764	565	0	0	0	0	0	2 788
Grade 11	0	0	1 518	706	0	0	0	0	0	2 224
Grade 12	0	0	0	26	300	335	79	88	18	846
TOTAL	633	1 755	4 617	1 572	300	335	79	88	18	9 397

a) How many learners were surveyed?
b) How many learners were in Grade 10?
c) How many learners were 15 years old?
d) How many learners were in Grade 10 and 15 years old?
e) What is the probability that one of these learners selected at random
   i) would have been in Grade 10
   ii) would have been 15 years old
   iii) would have been in Grade 10 and 15 years old?

<u>SOLUTION</u>:

	Age (in years)										
	13	14	15	16	17	18	19	20	21	Total	
Grade 8	478	600	316	78	0	0	0	0	0	1 472	
Grade 9	155	696	1 019	197	0	0	0	0	0	2 067	
Grade 10	0	459	1 764	565	0	0	0	0	0	2 788	B
Grade 11	0	0	1 518	706	0	0	0	0	0	2 224	
Grade 12	0	0	0	26	300	335	79	88	18	846	
Total	633	1 755	4 617	1 572	300	335	79	88	18	9 397	A

C D

a) 9 397 learners from Eastern Cape were surveyed. (*Read at A*)
b) 2 788 learners were in Grade 10. (*Read at B*)
c) 4 617 learners were 15 years old. (*Read at C*)
d) 1 764 learners were in Grade 10 and 15 years old. (*Read at D*)
e)
   i) $P(\text{Gr } 10) = \frac{n(\text{Gr } 10)}{n(S)} = \frac{2\,788}{9\,397} \approx 0{,}296\ 7$
   ii) $P(15 \text{ years}) = \frac{n(15 \text{ yr olds})}{n(S)} = \frac{4\,617}{9\,397} \approx 0{,}4913$
   iii) $P(\text{Grd } 10 \text{ and } 15 \text{ yrs}) = \frac{n(\text{Gr } 10 \text{ and } 15 \text{ yrs})}{n(S)} = \frac{1\,764}{9\,397} \approx 0{,}1877$

- ✓ Remember, when events are independent we can work out the probability that they both occur by multiplying the probabilities.
  i.e. For events A and B, P(A and B) = P(A) × P(B)

- ✓ We can also use this rule to determine whether or not events are independent.
  i.e. If P(A) × P(B) = P(A and B), the events A and B are independent.

EXAMPLE 10
Use the contingency table in Example 9 to determine whether the events of a learner being in Grade 10 and a learner being 15 years old are independent. Justify your answer.

<u>SOLUTION</u>:

- *To test whether events are independent you need to check whether P(A) × P(B) = P(A and B)*

P(Gr 10) × P(15 yrs) $= \frac{2\,788}{9\,397} \times \frac{4\,617}{9\,397} \approx 0{,}1458$

P(Gr 10 and 15 yrs) ≈ 0,1877 ← From question g) in the previous example

∴ P(Gr 10) × P(15 yrs) ≠ P(Gr 10 and 15 yrs)

So, the events of being in Grade 10 and being 15 years old are ***not*** independent (or we can say they are dependent)

- *The probability of being in Grade 10 depends on your age and you are more likely to be in Grade 10 when you are 15 years old than when you are 6 years old.*

EXERCISE 6.6

Give each of the answers in the exercise (where necessary) as a percentage correct to 2 decimal places.

1) The table below from Census@Schools 2009 shows the number of male and female learners and the colour of both their eyes. Study the table and answer the questions below:

	Brown	Green	Blue	Other	**Total**
Male	60 531	1 615	2 235	5 095	**69 476**
Female	66 993	2 015	2 449	4 686	**76 143**
Total	**127 524**	**3 630**	**4 684**	**9 781**	**145 619**

a) Are the events of selecting a learner with brown eyes and selecting a learner with blue eyes mutually exclusive? Explain your answer.
b) Are the events of selecting a learner with brown eyes and selecting a learner with blue eyes complementary? Explain your answer.
c) Are the events of selecting a male learner and selecting a female learner complementary? Explain your answer.
d) What is the probability that one of these learners, selected at random:
   i) Has blue eyes?
   ii) Is female?
e) Are the events of having blue eyes and being female independent? Justify your answer with calculations.

2) The table below shows data collected by 2009 Census@School about the favourite sports amongst girls and boys in the Free State.

	Boys	Girls	**Total**
Soccer	3 068	425	**3 493**
No Favourite sport	801	1 922	**A**
Netball	29	2 259	**2 288**
Athletics	**B**	394	**688**
Rugby	375	9	**384**
Volleyball	154	227	**381**
Cricket	274	57	**331**
Dance Sport	86	**C**	**326**
Other	63	249	**312**
Total	**5 144**	**D**	**10 926**

a) Calculate the value of A, B, C and D.
b) One of these learners is randomly selected. What is the probability that this learner:
   i) Prefers netball?
   ii) Is a girl?
   iii) Prefers netball and is a girl?
c) Without doing any calculations, say whether you think the events of preferring netball and being a girl are independent. Explain why.
d) Show with calculations whether the events of preferring netball and being a girl are independent or dependent.

COUNTING PRINCIPLES

✓ It is possible to list all possible outcomes using a tree diagram. When you have many possible outcomes, a tree diagram can become very messy and it becomes difficult to count the possible outcomes.

Counting principles help you to count the possible outcomes without drawing a tree diagram.

a) Arrangements with Repeats

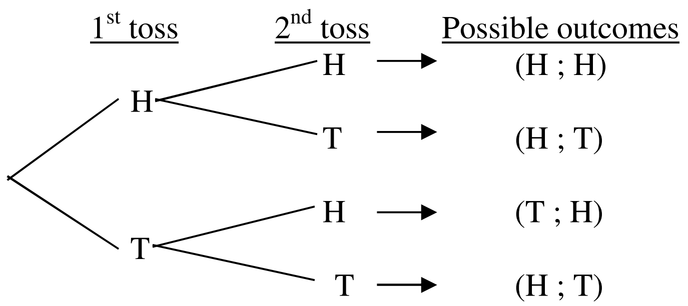
✓ It will help you understand the counting principles by thinking back to the tree diagrams that you used in previous grades.

EXAMPLE 11

a) A coin is tossed twice. How many outcomes are there?
b) A coin is tossed three times. How many outcomes are there?
c) A coin is tossed four times. How many outcomes are there?
d) A coin is tossed twenty times. How many outcomes are there?

<u>SOLUTION</u>:

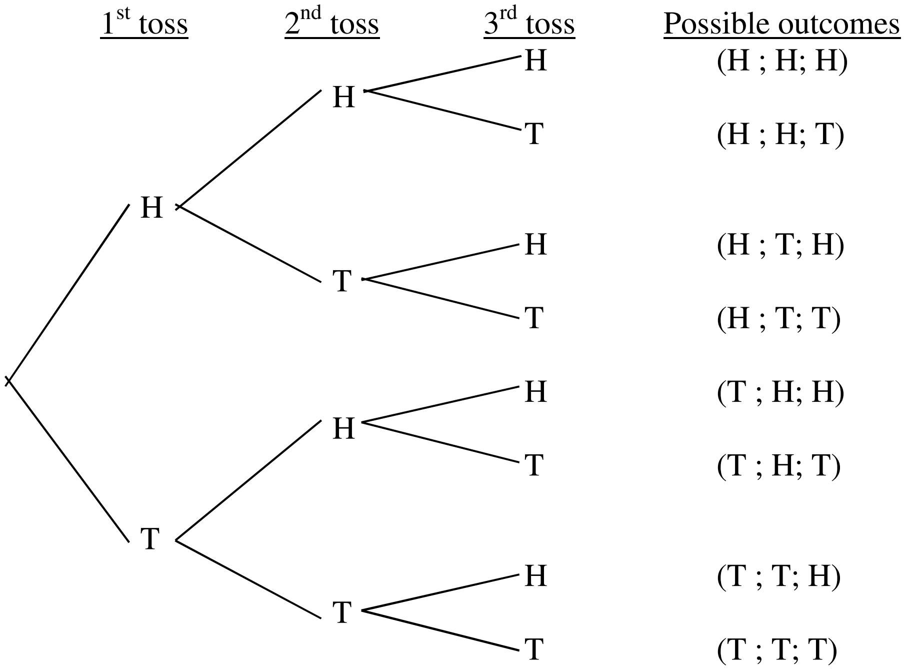
a) You can use a tree diagram to help you find the number of outcomes when a coin is tossed twice:



There are two possible outcomes for the first toss.
There are two possible outcomes for the second toss.
There are 2 × 2 = 4 outcomes in total.

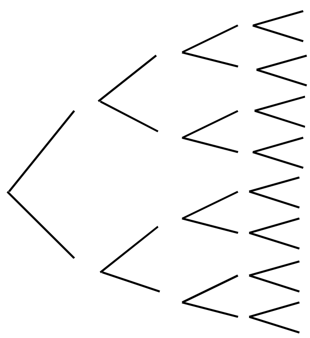
EXAMPLE 11 (continued)

b) You can use a tree diagram to help you find the number of outcomes when a coin is tossed three times:



There are two possible outcomes for the first toss.
There are two possible outcomes for the second toss.
There are two possible outcomes for the third toss.
There are $2 \times 2 \times 2 = 8$ outcomes in total.

c) It becomes more difficult to draw a tree diagram to show all the possible outcomes when a coin is tossed 4 times. Below is a simple sketch of what one might look like:



When a coin is tossed four times:
There are two possible outcomes for the first toss.
There are two possible outcomes for the second toss.
There are two possible outcomes for the third toss.
There are two possible outcomes for the fourth toss.
There are $2 \times 2 \times 2 \times 2 = 16$ outcomes in total.

d) It is not easy to draw a tree diagram for 20 tosses, but you should be seeing a pattern.
If a coin is tossed 20 times, there will be $2 \times 2 \times 2$…. (to 20 terms) possible outcomes, i.e. number of possible outcomes = 2^{20} = 1 048 576.

- ✓ The ***fundamental counting principle*** is a quick method for calculating numbers of outcomes using multiplication.

- ✓ The fundamental counting principle states:

  Suppose there are n_1 ways to make a choice, and for each of these there are n_2 ways to make a second choice, and for each of these there are n_3 ways to make a third choice, and so on.
  The product $n_1 \times n_2 \times n_3 \times ... \times n_k$ is the number of possible outcomes.

- ✓ In simple language the fundamental counting principle says:
  "If you have several stages of an event, each with a different number of outcomes, then you can find the TOTAL number of outcomes by multiplying the number of outcomes of each stage."

  For example:
  - o We can use the fundamental counting principle to find the number of outcomes when a coin is tossed four times.
    - There are 2 ways to get the first outcome (H or T).
    - There are 2 ways to get the second outcome (H or T).
    - There are 2 ways to get the third outcome (H or T).
    - There are 2 ways to get the fourth outcome (H or T).

    The number of possible outcomes when a coin is tossed four times
    $= 2 \times 2 \times 2 \times 2 = 16$

  - o We can use the fundamental counting principle to find the number of outcomes when a dice is tossed three times.
    - There are 6 ways to get the first outcome (1; 2; 3; 4; 5 or 6)
    - There are 6 ways to get the second outcome (1; 2; 3; 4; 5 or 6)
    - There are 6 ways to get the third outcome (1; 2; 3; 4; 5 or 6)

    The number of possible outcomes when a dice is tossed three times
    $= 6 \times 6 \times 6 = 216$

b) Arrangements without Repeats

✓ Sometimes we have examples where an event can only be used once.

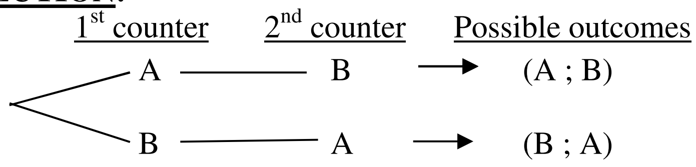
EXAMPLE 12

a) Two counters marked A and B are randomly drawn from a box. When a counter is taken, it is not returned. How many ways can these letters be drawn, i.e. how many possible outcomes are there?

b) Three counters marked A, B and C are randomly drawn from a box. When a counter is taken, it is not returned. How many possible outcomes are there?

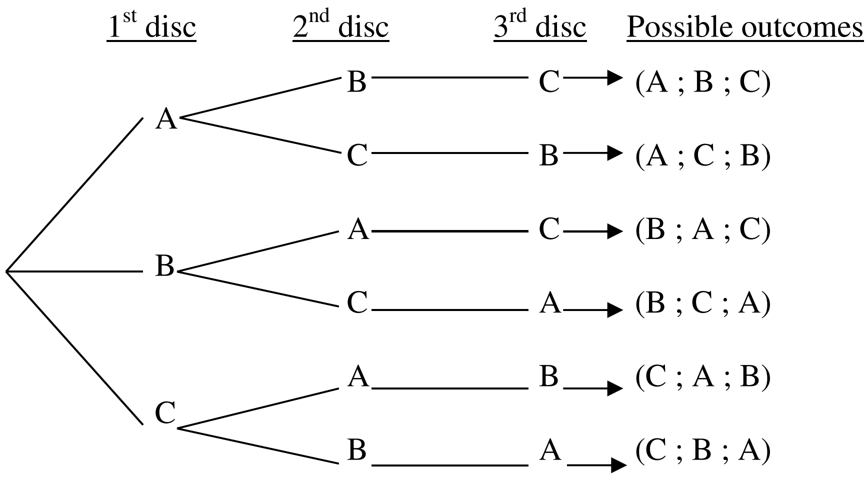
SOLUTION:

a)



If A is the first letter, then B must be the second letter.
If B is the first letter, then A must be the second letter.
There are two possible outcomes for the first letter.
There is only one possible outcome for the second letter.
So the number of possible outcomes = 2 × 1 = 2.

b)

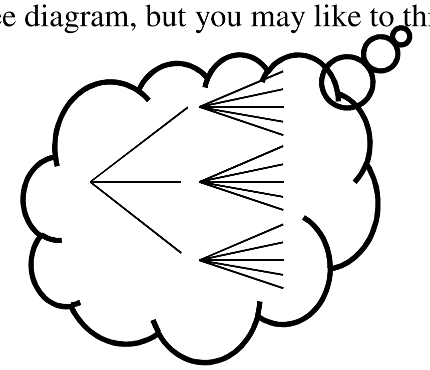


There are three possible outcomes for the first letter.
There are two possible outcomes for the second letter.
There is only one possible outcome for the third letter.
So the number of possible outcomes = 3 × 2 × 1 = 6.

- ✓ Multiplying the number of possible outcomes of events each time can also be used in other contexts.

EXAMPLE 13
You are going out with your friends. You are going to watch a movie and then you are going to a restaurant.
There are three movies that you would like to watch.
There are five restaurants where you would like to eat.
You can only choose one movie and one restaurant.
How many different combinations of movie and restaurant are there?

<u>SOLUTION</u>:
You do not have to draw a tree diagram, but you may like to think about a tree diagram.



Number of combinations = number of movies × number of restaurants
$= 3 \times 5$
$= 15$

c) Factorial Notation

- ✓ The arrangement of numbers $4 \times 3 \times 2 \times 1$ can be written as 4!
  You say ***'4 factorial'***.

  $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$

  $n! = n \times (n - 1) \times (n - 2) \times (n - 3) \times (n - 4) \times ..... \times 4 \times 3 \times 2 \times 1$

- ✓ Factorial notation is used for finding the total number of outcomes without repeats.

- ✓ Most scientific calculators have a factorial key.

  - On the Casio fx-82ZA PLUS, the factorial key ($x!$) is next to the (x^{-1}).
    To calculate 6!, enter the number (e.g. 6) then press [SHIFT] [x^{-1}] ($x!$) [=]

  - On the Sharp EL-W535HT, the factorial key ($n!$) is next to the 1.
    To calculate 6!, enter the number (e.g. 6) then press: [2nd F] [4] ($n!$) [=]

EXAMPLE 14

a) A three digit code is made up of numbers 3, 5 and 7. The digits may be repeated. How many different codes are possible?

b) A three digit code is made up of numbers 3, 5 and 7. Each digit is used only once. How many different codes are possible?

<u>SOLUTION</u>:

a) There are three places to fill ______ ______ ______

The first place may be filled by 3, 5 or 7, so there are 3 possible numbers for the first place.
The second space may be filled by 3, 5 or 7, so there are 3 possible numbers for the second place.
And the third space may be filled by 3, 5 or 7, so there are 3 possible numbers for the third place.

So, because the numbers can be repeated, there are $3 \times 3 \times 3 = 27$ possible codes.

These 27 different codes are:

333	335	337	353	355	357	373	375	377
533	535	537	553	555	557	573	575	577
733	735	737	753	755	756	773	775	777

b) There are three places to fill ______ ______ ______

The first space may be filled by 3, 5 or 7, so there are 3 possible numbers for the first place.
The second space can only be filled in 2 ways and the third space can only be filled in 1 way.

Because the numbers ***cannot*** be repeated, there are $3! = 3 \times 2 \times 1 = 6$ codes possible.

These 6 different codes are: 357; 375; 537; 573; 735; and 753.

- ✓ WHEN REPEATS ***ARE ALLOWED***, *you find the total number of outcomes by multiplying the number of possible outcomes in each stage of an event.*
- ✓ WHEN REPEATS ***ARE NOT ALLOWED***, *you find the total number of outcomes by multiplying the number of possible outcomes* ***that are left*** *in each stage of an event.*

✓ Sometimes you don't want to arrange all the items, but only some of them.

EXAMPLE 15

a) How many three-digit numbers can be formed from the digits 2; 3; 5; 7 and 8 if each digit can be used only once?

b) How many four-digit codes can be formed if the first character must be a letter of the alphabet and the following three characters must be digits?

<u>SOLUTION</u>:

a) There are three places to fill ______ ______ ______

The first place may be filled by 2; 3, 5, 7 or 8, so the first place can be filled 5 ways.
Once that place has been filled, there are four numbers left over.
This means that the second space can only be filled in 4 ways.
Once that place has been filled, there are three numbers left over.
This means that the third space can be filled in 3 ways.

So, the number of three-digit numbers that can be formed = $5 \times 4 \times 3 = 60$.

You could find all sixty 3-digit numbers using a systematic list.
Start at 2, then 3, etc.

235	237	238
253	257	258
273	275	278

etc.

b) There are four places to fill ______ ______ ______ ______

Since it does not say the digits may not be repeated, we can presume that they can be repeated.

The first space may be filled in 26 ways because there are 26 letters in the English alphabet.
The second, third and fourth space can be filled in 10 ways because there are 10 digits to choose from).

So, the number of four-digit numbers that can be formed
$= 26 \times 10 \times 10 \times 10 = 26\ 000$.

You could find all 26 000 3-digit numbers using a systematic list.

A000	A001	A002	A003	A004	A005	A006	A007	A008	A009
A010	A011	A012	A013	A014	A015	A016	A017	A018	A019
A020	A021	A022	A023	A024	A025	A026	A027	A028	A029

etc.

EXERCISE 6.7

1) Write each of these as a product of digits (e.g. $3! = 3 \times 2 \times 1$)
   a) 4!
   b) 7!
   c) 10!

2) Use your calculator to determine each of the following (e.g. 5! = 120)
   a) 6!
   b) 8!
   c) 12!
   d) 25!

3) The following menu is offered at a restaurant:

Starter	**Main**	**Dessert**
Chicken livers Tomato soup Salad	Peri-peri chicken Lamb chops Beef rump Fish and chips	Fruit salad and ice-cream Chocolate pudding

   If one starter, one main and one dessert is selected from the menu. How many combinations of starter, main and dessert could be chosen?

4) The new Gauteng number plate has two letters (from the alphabet A – Z, but not using vowels), two digits from 0 – 9 and then another two letters (excluding vowels).
   a) How many possible combinations can be created of this number plate?
   b) The previous Gauteng number plate had three letters (from the alphabet A – Z, but not using vowels) and then three digits from 0 – 9. Are there more combinations of the old or new number plates?

5) Given the numbers: 0; 1; 2; 3; 4 and 5
   a) How many 4 digit numbers can be formed if the first digit may not be 0 and the numbers may not be repeated?
   b) How many of these numbers will be divisible by 5?

6) The soccer coach needs the goal posts to be moved on the field. He randomly chooses five boys out of a group of twenty to help him. How many different groups of 5 boys can be selected?

7) How many three-character codes can be formed if the first character must be a letter and the second two characters must be digits?

d) Special Conditions

- ✓ Sometimes when we are counting the number of arrangements, we are given special conditions, for example similar groups must be arranged together, or two or more elements must be put together in the arrangement.

EXAMPLE 16

A photograph needs to be taken of the Representative Council of Learners (RCL) at a school.
There are three girls and two boys in the RCL and all of them need to sit in one row for the photograph.

moooi.com

a) Suppose there is no restriction on the order in which the RCL sits. In how many ways can the RCL be arranged in a row?
b) Suppose the President and the Vice-President of the RCL must be seated next to each other, in how many different ways can the RCL be arranged in a row?
c) Suppose all the girls must sit next to each other, and all the boys must sit next to each other. In how many different ways can the RCL be arranged in a row?

SOLUTION:

a) There are five learners in total and 5 places to fill.

____ ____ ____ ____ ____

Once a learner has been seated in the row, they cannot be seated a second time. Therefore, there is no repetition.

Number of different arrangements = $5 \times 4 \times 3 \times 2 \times 1 = 5! = 120$ different ways.

b) If the President and Vice-President must be seated next to each other, we consider them as one unit, but with two ways of arranging them: AB or BA.

Once they have been seated, there are 3 places left to fill

________ ____ ____ ____

This means that there are four units/people to arrange: $4 \times 3 \times 2 \times 1 = 4! = 24$
But because the President and Vice-President can be arranged in two ways, we have to double this number of arrangements.

Number of different arrangements = $(4 \times 3 \times 2 \times 1) \times 2 = 4! \times 2! = 48$ ways.

c) The three girls can be arranged in $3 \times 2 \times 1 = 3! = 6$ ways.
The two boys can be arranged in $2 \times 1 = 2! = 2$ ways.
But we could seat the girls on the left and the boys on the right, or we could seat the boys on the left and the girls on the right – so there are 2 further ways of arranging them.

So, the number of different arrangements $= (3 \times 2 \times 1) \times (2 \times 1) \times 2$
$= 3! \times 2! \times 2!$
$= 24$ ways.

EXERCISE 6.8

1) Balls are numbered 1 to 12 and placed in a bag.
   A ball is drawn at random and placed on a table in a row.
   a) How many different arrangements of the twelve balls are possible?
   b) If the balls numbered 8 and 11 are placed next to each other, in how many ways can the balls be rearranged?
2) Six different red mugs and five different blue mugs have to be arranged on a shelf. How many arrangements can be made if all the red mugs are put together and all the blue mugs are put together?
3) Jarred, Rikus, Gareth, Mathope and Fishana are in a hockey team.
   The five of them are sitting on a bench waiting for their turn to play.
   Jarred and Rikus are sitting next to each other.
   In how many different ways can these boys be arranged on the bench?

e) Identical Items in a list

✓ Consider how many arrangements of the letters there are in the word LEEK.

- Here is a list of some of the possible arrangements:

LEEK	LEKE	LEEK	LEKE	
LKEE	LKEE	ELEK	EELK	etc

- Because the letter E is repeated, we *cannot* say that there are 4! different arrangements. In fact, because 2 letters are repeated, there is half the number of different arrangements than there would be if all four letters were different.

✓ We say that

- If there are n different items that are all different, then there are $n \times (n - 1) \times (n - 2) \ldots n$ terms or $n!$ arrangements.
- If there are n different items, but ***one item is repeated twice***, then there are $n \times (n - 1) \times (n - 2) \ldots n$ terms **divided by** 2 or 2! arrangements.
- If there are n different items, but ***one item is repeated three times***, then there are $n \times (n - 1) \times (n - 2) \ldots n$ terms **divided by** $3 \times 2 \times 1$ or 3! arrangements.

✓ Sometimes more than one letter is repeated.

- For example, in the word DODO, D is repeated twice and O is repeated twice
- Here you can say that if there are n different items, but two items are repeated twice, then there are $n \times (n - 1) \times (n - 2) \ldots n$ terms **divided by** $2! \times 2!$ arrangements.

EXAMPLE 17

a) In how many ways can you arrange the letters in the word MEDIAN?
b) In how many ways can you arrange the letters in the word DATA?
c) In how many ways can you arrange letters in the word PERCENTILE?
d) In how many ways can you arrange the letters in the term CENSUS@SCHOOL?
e) The letters of the word STATSSA are re-arranged. How many arrangements will start and end with the letter 'T'?

<u>SOLUTION</u>:

a) There are 6 letters in the word MEDIAN and none of these are repeated.
Number of arrangements = 6! = 720.

b) There are 4 letters in the word DATA and A is repeated twice.
Number of arrangements $= \frac{4!}{2!} = 12$

c) There are 10 letters in the word PERCENTILE and E is repeated three times.
Number of arrangements $= \frac{10!}{3!} = 604\ 800$

d) There are 12 letters and an '@' in the term Census@School. S is repeated 3 times, C is repeated twice and O is repeated twice.
Number of arrangements = $\frac{13!}{3! \times 2! \times 2!}$ = 259 459 200
Note: You must enter the × sign between each factorial on your calculator.

e) If the arrangement of the word STATSSA starts and ends in T, then there are 5 letters left to arrange. S is repeated three times and A is repeated twice.
Number of arrangements $= \frac{5!}{3! \times 2!} = 10$

EXERCISE 6.9

1) In how many ways can you arrange the letters in the word CUMULATIVE?

2) In how many ways can you arrange the letters in the word PROBABILITY?

3) How many arrangements of the letters in the words STANDARD DEVIATION start and end in N?

f) Using Counting Principles to Find Probability

- ✓ You can use these counting principles to find the number of possible outcomes, and you can also use them to find the number of favourable outcomes.

- ✓ When you know the number of possible outcomes and the number of favourable outcomes, you can work out the probability of the favourable event using the formula

$$\text{Probability of a favourable event} = = \frac{\text{Number of favourable outcomes}}{\text{Number of possible outcomes}}$$

EXAMPLE 18

Suppose a four-digit number is formed by randomly selecting four digits without repetition from 1; 2; 3; 4; 5; 6; 7 and 8.

What is the probability that the number formed lies between 4 000 and 5 000?

<u>SOLUTION</u>:

Number of possible outcomes = 8 × 7 × 6 × 5 = 1 680

A number that lies between 4 000 and 5 000 must start with a 4, so there are seven digits left to arrange and only three places to fill.

Number of favourable outcomes = 7 × 6 × 5 = 210

$$\text{Probability that the number lies between 4 000 and 5 000} = \frac{\text{Number of favourable outcomes}}{\text{Number of possible outcomes}}$$

$$= \frac{210}{1\ 680}$$

$$= \frac{1}{8}$$

EXERCISE 6.10

Give each of the answers correct to the nearest percentage.

1) What is the probability that a random arrangement of the letters in the name ‘PHILLIPINE’, start and end in ‘L’?
2) If a four digit number is created from the digits 1, 2, 3, 4, 5, 7 and 9, what is the probability that a randomly chosen number:
   a) Is even?
   b) Is greater than 4 000?
   c) Has digits arranged in ascending order?
3) What is the probability that a random arrangement of the letters BAFANA starts and ends with an A?

REFERENCES

Department of Basic Education. (2011). *Curriculum and Assessment Policy Statement, Grade 10 – 12, Mathematics.*
Statistics South Africa. (2010). *Census at Schools Results 2009*
Murdock, J; Kaminschke, E and Kaminschke E: (2004) *Discovering Advanced Algebra*, Key Curriculum Press
Berry, J; Fentem, R; Francis, B and Graham, T ((2000) *AS Mathematics*, Collins

7

Worked Solutions to Exercises

CHAPTER 1 – GRADE 10 DATA HANDLING

Exercise 1.1

1)

a)

i) $\bar{x} = \frac{49}{11} = 4{,}4545.... \approx 4{,}45$

ii) 2; 2; 3; 4; 4; 4; 4; 5; 6; 7; 8
Median = 4

iii) Mode = 4

b)

i) $\bar{x} = \frac{R28{,}75}{7} = R4{,}1071.... \approx R4{,}11$

ii) R1,25; R2,50; R2,50; R3,00; R6,50; R6,50; R6,50
Median = R3,00

iii) Mode = R6,50

c)

i) $\bar{x} = \frac{83}{8} = 10{,}375 \text{ cm} \approx 10{,}38 \text{ cm}$

ii) 6 cm; 7 cm; 7 cm; 11 cm; 12 cm; 12 cm; 13 cm; 15 cm
Median $= \frac{11 \text{ cm}+12 \text{ cm}}{2} = 11{,}5$ cm

iii) Mode = 7 cm and 12 cm

d)

i) $\bar{x} = \frac{1\,171}{10} = 117{,}1$ kg

ii) 109 kg; 111 kg; 112 kg; 113 kg; 114 kg; 118 kg; 119 kg; 120 kg; 125 kg; 130 kg
Median $= \frac{114+118}{2} = 116$ kg

iii) No mode

2) Total height of 10 learners = 10 × 166,8 cm = 1 668 cm
Total height minus the tallest learner = 1 668 cm – 169,9 cm = 1 498,1 cm
Average height of the remaining 9 learners $= \frac{1\,498{,}1}{9} = 166{,}4555... \approx 166{,}46$ cm

3) Let the ages of the four children be *a; b; c* and *d*.
The mode is 16 years. The oldest children are twins, so *c* = *d* = 16 years.
The median $= \frac{b+16}{2} = 15{,}5$ years
$b + 16 = 31$
So $b = 15$ years
The mean $= \frac{a+15+16+16}{4} = 14{,}25$
$a + 47 = 57$
So $a = 10$ years
The ages of the children are 10 years, 15 years, 16 years and 16 years.

4)

a) Mean of the girls $= \frac{10,3\% + 11,2\% + 7,8\% + 6,9\% + 12,5\%}{5} = \frac{48,7\%}{5} = 9,74\%$

b) Mean of the boys $= \frac{12,2\% + 9,8\% + 11,1\% + 8,3 + x}{5} \geq 10$

$$41,4 + x \geq 50$$
$$x \geq 50 - 41,4$$
$$x \geq 8,6\%$$

The minimum percentage for the Grade 12 boys must be 8,6%.

5) Total for 9 grades = 9 × 50% = 450%
Total for 10 grades = 10 × 49,68% = 496,8%
Percentage of boy learners in Grade 12 = 496,8% – 450% = 46,8%

Exercise 1.2

1)

a)

i) Mode = 24,5 cm

ii) There are 39 learners. $\frac{39}{2} = 19\frac{1}{2}$. So 20th term is the median.
Count along the frequencies: 2 + 4 + 6 + 8 = 20
So the median = 24 cm

iii) Mean = $\frac{942,5}{39}$ = 24,1666...≈ 24,2 cm

b)

i) The mode is the most common length of foot that occurs in the group.

ii) The median is the middle item. 50% of the foot lengths are less than 24,2 cm and 50% of the foot lengths are more than 24,2 cm.

iii) The mean is a single measurement that is calculated using all the data items. The mean is found by adding all the lengths together (totalling 942,5 cm) and sharing them out equally so that each of the 39 learners would get 24,2 cm.

2)

a)

i) Mean = $\frac{1\,887\%}{25}$ = 75,48%

ii) There are 25 learners. $\frac{25}{2} = 12\frac{1}{2}$, so the 13th term is the median
Count the frequencies: 1 + 3 + 5 + 6 = 15, so the median is 75%

iii) Mode = 75%

b)

i) The mean is found by adding all the marks together and then sharing them out equally so that each learner gets 75,48%.

ii) 50% of the learners got less than or equal to the median of 75% and 50% of the learners got more than or equal to 75%

iii) The mode tells us that the mark that occurred most often was 75%

3)

a)

i) Mean = $\frac{431\%}{24}$ = 17,9583...% ≈ 18,0%

ii) There are 24 municipalities. $\frac{24}{2}$ = 12, so the median lies between the 12th and 13th terms.
12th term = 17% and the 13th term = 18%, so the median = $\frac{17\% + 18\%}{2}$ = 17,5%

iii) Mode = 14%

b)

i) The mean unemployment rate of 18,0% is found by totalling all the unemployment rates and then sharing them out equally amongst all the municipalities. This means that each municipality would have an unemployment rate of 18,0%.

ii) 50% of the municipalities have unemployment rates that are less than 17,5% and 50% of the municipalities have unemployment rates that are more than 17,5%.

iii) The mode of 14% is the unemployment rate that occurs most often.

Exercise 1.3

1)

a)

i) The modal interval is 20 years ≤ x < 30 years

ii)

Age (in years) (x)	Midpoint of the interval (X)	Frequency (f)	$f.X$
$0 \leq x < 10$	5	4	$4 \times 5 = 20$
$10 \leq x < 20$	15	12	$12 \times 15 = 180$
$20 \leq x < 30$	25	25	$25 \times 25 = 625$
$30 \leq x < 40$	35	14	$14 \times 35 = 490$
$40 \leq x < 50$	45	10	$10 \times 45 = 450$
$50 \leq x < 60$	55	20	$20 \times 55 = 1\ 100$
$60 \leq x < 70$	65	5	$5 \times 65 = 325$
		n = 90	$\sum f.X = 3\ 190$

Mean $= \bar{X} = \frac{\sum f.X}{n} = \frac{3\ 190 \text{ years}}{90} \approx$ 35,4444... years ≈ 35,4 years

iii) There are 90 people. $\frac{90}{2} = 45$, so the median lies between the 45th and 46th terms.
The 45th and 46th terms lie in the interval 30 years ≤ x < 40 years.
So both the 45th and the 46th terms are approximately equal to 35 years,
and the median ≈ 35 years.

b)

i) The modal interval of 20 years ≤ x < 30 years is the most common age category.

ii) The mean tells us that if all the ages are added together and shared out equally, each person would be approximately 35,4 years old.

iii) 50% of the people in this group are younger than or equal to 35 years. 50% of the people are older than or equal to 35 years.

2)

a)

i) The modal interval is 0 ≤ x < R50 000

ii)

Property value in Rand x	Midpoint of the interval X	Frequency f	$f.X$
$0 \leq x < 50\ 000$	25 000	172	$172 \times 25\ 000 = 4\ 300\ 000$
$50\ 000 \leq x < 100\ 000$	75 000	51	$51 \times 75\ 000 = 3\ 825\ 000$
$100\ 000 \leq x < 150\ 000$	125 000	18	$18 \times 125\ 000 = 2\ 250\ 000$
$150\ 000 \leq x < 200\ 000$	175 000	12	$12 \times 175\ 000 = 2\ 100\ 000$
$200\ 000 \leq x < 250\ 000$	225 000	9	$9 \times 225\ 000 = 2\ 025\ 000$
		n = 262	$\sum f.X = 14\ 500\ 000$

Mean property values $= \frac{\sum f.X}{n} = \frac{\text{R}14\ 500\ 000}{262} \approx$ R55 343,51

iii) There are 262 properties. $\frac{262}{2} = 131$, so the median lies between the 131st and 132nd terms.
The 131st and 132nd terms lie in the interval R0 ≤ x < R50 000.
So both the 131st and the 132nd terms are approximately equal to R25 000,
and the median ≈ R25 000.

b)

i) The modal interval tells us that more of the property prices lie in the interval R0 ≤ x < R50 000 than in any other of the intervals.

ii) The mean tells us that if all the property values were added together and shared out equally, that each property would have a value of approximately R55 343,51.

iii) Approximately 50% of the property values are less than R25 000 and approximately 50% of the property values are more than R25 000.

Exercise 1.4

1)

a) 1 6 6 9 | 15 17 23 24 | 33 33 38 38 | 38 45 46 51
Minimum value = 1
$Q_1 = \frac{9+15}{2} = \frac{24}{2} = 12$
$M = \frac{24+33}{2} = \frac{57}{2} = 28{,}5$
$Q_1 = \frac{38+38}{2} = 38$
Maximum value = 51

b) 9 14 19 **21** 24 29 29 | 32 33 35 **36** 40 46 49
Minimum value = 9
$Q_1 = 21$
$M = \frac{29+32}{2} = \frac{61}{2} = 30{,}5$
$Q_3 = 36$
Maximum value = 49

c) 15 15 19 | 25 25 26 **27** 32 36 41 | 43 45 48
Minimum value = 15
$Q_1 = \frac{19+25}{2} = \frac{44}{2} = 22$
$M = 27$
$Q_3 = \frac{41+43}{2} = \frac{84}{2} = 42$
Maximum value = 48

d) 4 6 10 | 10 16 17 | 22 31 34 | 35 44 46
Minimum value = 4
$Q_1 = \frac{10+10}{2} = 10$
$M = \frac{17+22}{2} = \frac{39}{2} = 19{,}5$
$Q_1 = \frac{34+35}{2} = \frac{69}{2} = 34{,}5$
Maximum value = 46

2)

a)
Minimum value = 2
$Q_1 = \frac{12+12}{2} = 12$
$M = 24$
$Q_1 = \frac{34+37}{2} = \frac{71}{2} = 35{,}5$
Maximum value = 59

b) Johan gave two answers for the upper quartile. He does not understand that in this case you have to find the mean of the two middle numbers.

Exercise 1.5

1)

a)

i) Percentile = $\frac{\text{number of data items that are less than or equal to 102}}{\text{total number of data items}} \times 100\% = \frac{49}{50} \times 100\% = 98^{\text{th}}$

ii) Percentile = $\frac{\text{number of data items that are less than or equal to 34}}{\text{total number of data items}} \times 100\% = \frac{2}{50} \times 100\% = 4^{\text{th}}$

iii) Percentile = $\frac{\text{number of data items that are less than or equal to 96}}{\text{total number of data items}} \times 100\% = \frac{43}{50} \times 100\% = 86^{\text{th}}$

iv) Percentile = $\frac{\text{number of data items that are less than or equal to 70}}{\text{total number of data items}} \times 100\% = \frac{26}{50} \times 100\% = 52^{\text{nd}}$

b)

i) Data item corresponding to the 30^{th} percentile = $\frac{\text{percentile}}{100} \times 100\% = \frac{30}{100} \times 50 = 15$
The 15^{th} data item is 56.

ii) Data item corresponding to the 50^{th} percentile = $\frac{\text{percentile}}{100} \times 100\% = \frac{50}{100} \times 50 = 25$

The 25^{th} data item is 69.

iii) Data item corresponding to the 10^{th} percentile $= \frac{\text{percentile}}{100} \times 100\% = \frac{10}{100} \times 50 = 5$.
The 5^{th} data item is 39.

iv) Data item corresponding to the 65^{th} percentile $= \frac{\text{percentile}}{100} \times 100\% = \frac{65}{100} \times 50 = 32{,}5 \approx 33$.
The 33^{rd} data item is 79.

2)

a) Data item corresponding to the 11^{th} percentile $= \frac{\text{percentile}}{100} \times 100\% = \frac{11}{100} \times 70 = 7{,}7 \approx 8$.
The 8^{th} data item is 105.

b) Data item corresponding to the 44^{th} percentile $= \frac{\text{percentile}}{100} \times 100\% = \frac{44}{100} \times 70 = 30{,}8 \approx 31$.
The 31^{st} data item is 121.

c) Percentile $= \frac{\text{number of data items that are less than or equal to 135}}{\text{total number of data items}} \times 100\% = \frac{46}{70} \times 100\% = 65{,}7 \approx 66^{th}$

Exercise 1.6

1)

a) 9 11 **12** 12 13 **13** 14 16 **21** 22 24
Median = 13
$Q_1 = 12$ and $Q_3 = 21$
Interquartile range = 21 – 12 = 9

b) 4 5 5 **6** 7 9 10 | 11 12 12 **13** 14 14 14
Median $= \frac{10+11}{2} = \frac{21}{2} = 10{,}5$
$Q_1 = 6$ and $Q_3 = 13$
Interquartile range = 13 – 6 = 7

c) 1 2 2 3 | 4 5 5 6 | 7 8 9 11 | 12 12 14 14
Median $= \frac{6+7}{2} = \frac{13}{2} = 6{,}5$
$Q_1 = \frac{3+4}{2} = \frac{7}{2} = 3{,}5$
$Q_3 = \frac{11+12}{2} = \frac{23}{2} = 12{,}5$
Interquartile range = 12,5 – 3,5 = 9

2) 4 7 8 9 11 | 12 14 16 17 18 **19** 20 23 23 26 26 | 28 29 37 35 45

a) Range = 45 – 4 = 41

b) $Q_1 = \frac{11+12}{2} = \frac{23}{2} = 11{,}5$ and $Q_3 = \frac{26+28}{2} = \frac{54}{2} = 27$

c) $IQR = Q_3 - Q_1 = 27 - 11{,}5 = 15{,}5$

d) 25% of the data items are less than 11,5 and 75% of the data items are more than 11,5.
75% of the data items are less than 27 and 25% of the data items are more than 27.
50% of the data items lie between 11,5 and 27.

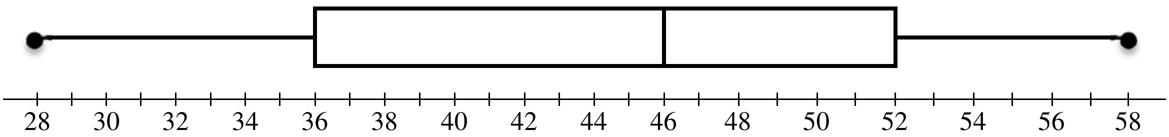
Exercise 1.7

- *Always make sure that the units on the number line are equally spaced.*

1) 28 28 30 35 35 | 37 38 39 43 45 | 47 49 49 50 51 | 53 55 55 56 58

a) Minimum value = 28
$Q_1 = \frac{35+37}{2} = \frac{72}{2} = 36$
$M = \frac{45+47}{2} = \frac{92}{2} = 46$
$Q_3 = \frac{51+53}{2} = \frac{104}{2} = 52$
Maximum value = 58

b)



c)

i) Range = 58 – 28 = 30

ii) IQR = 52 – 36 = 16

2)

a)

<u>Girls' Heights</u>: 150 150 153 | 155 156 158 **<u>160</u>** 161 164 164 | 166 170 170

Minimum value =- 150 cm

$Q_1 = \frac{153+155}{2} = \frac{308}{2} = 154$ cm

M = 160 cm

$Q_3 = \frac{164+166}{2} = \frac{330}{2} = 165$ cm

Maximum value = 170 cm

IQR = 165 cm – 154 cm = 11 cm

<u>Boys' Heights</u>: 140 142 151 | 157 158 159 **<u>160</u>** 162 165 180 | 180 180 180

Minimum value = 140 cm

$Q_1 = \frac{151+157}{2} = \frac{308}{2} = 154$ cm

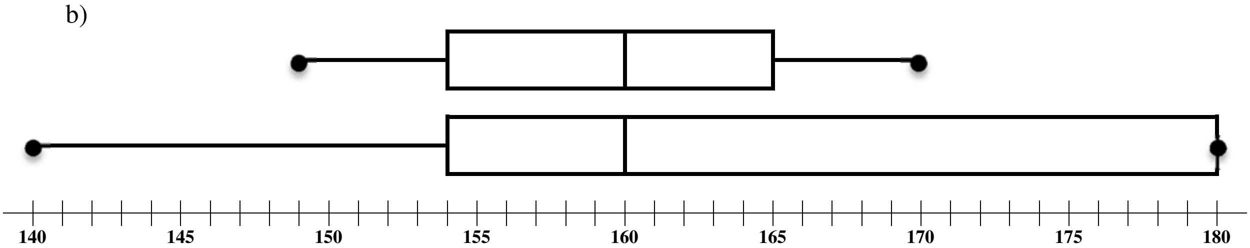
M = 160 cm

$Q_3 = \frac{180+180}{2} = 180$ cm

Maximum value = 180 cm

IQR = 180 cm – 154 cm = 26 cm

b)



c) *Some examples of conclusions:*

The girls' median height and the boys' median height it the same

The girls' heights are not as widely spread apart as the boys' heights.

The boys' interquartile range (26 cm) is far greater than the girls' interquartile range (11 cm) indicating that the boys' heights are more spread out than the girls' heights.

The heights of 25% of the girls and 25% of the boys lie between 154 cm and 160 cm.

25% of the boys' heights are 180 cm.

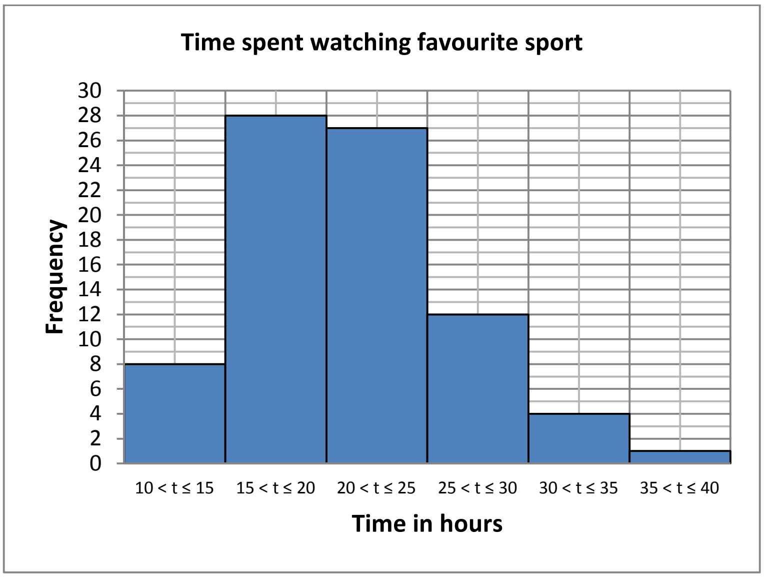
The boys' minimum height is much less than the girls' minimum height, and the boys' maximum height is much more than the girls' maximum height.

CHAPTER 2 – GRADE 11 DATA HANDLING

<u>Exercise 2.1</u>

1)

a)



b)

i) Modal interval is $15 < t \leq 20$

ii) There are 80 learners so the median lies between the 40th and 41st data items.
So the median is in the interval $20 < t \leq 25$
Therefore the median $\approx \frac{20+25}{2} = \frac{45}{2} =$ 22,5 hours

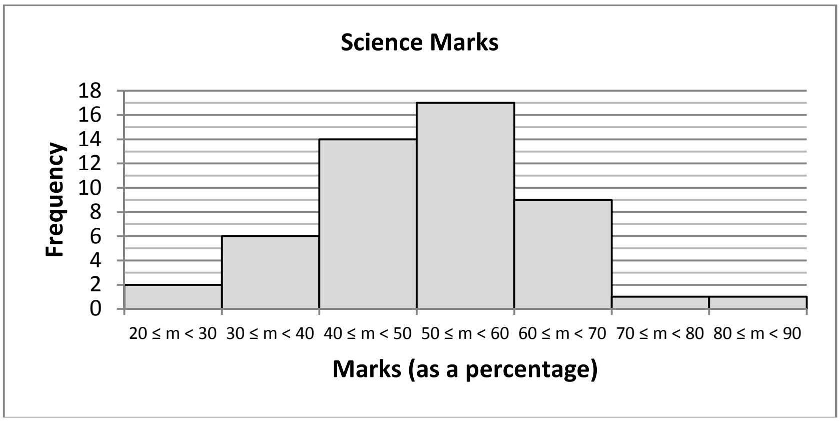
c) The modal interval tells us that more learners spent between 15 and 20 hours watching their favourite sport than any other interval.
The median tell us that approximately 50% of the learners spend less than 22,5 hours watching their favourite sport and approximately 50% of the learners spend more than 22,5 hours watching their favourite sport.

2)

a)

Marks	Frequency
$20 < m \leq 30$	2
$30 < m \leq 40$	6
$40 < m \leq 50$	14
$50 < m \leq 60$	17
$60 < m \leq 70$	9
$70 < m \leq 80$	1
$80 < m \leq 90$	1
TOTAL	50

b)



c) Modal Interval: $50\% \leq m < 60\%$
This tells us that there were more learners who got marks in that interval than in any other interval.

There are 50 learners so the median lies between the 25th and the 26th learners.
Add up the frequencies until you reach 25 and 25: $2 + 6 + 14 + 17 = 39$.
So the median is in the interval $50\% \leq m < 60\%$
The median $= \frac{50+60}{2} = \frac{110}{2} = 55\%$

This tells us that approximately ½ of the learners got less than 55% and approximately ½ of the learners got more than 55%.

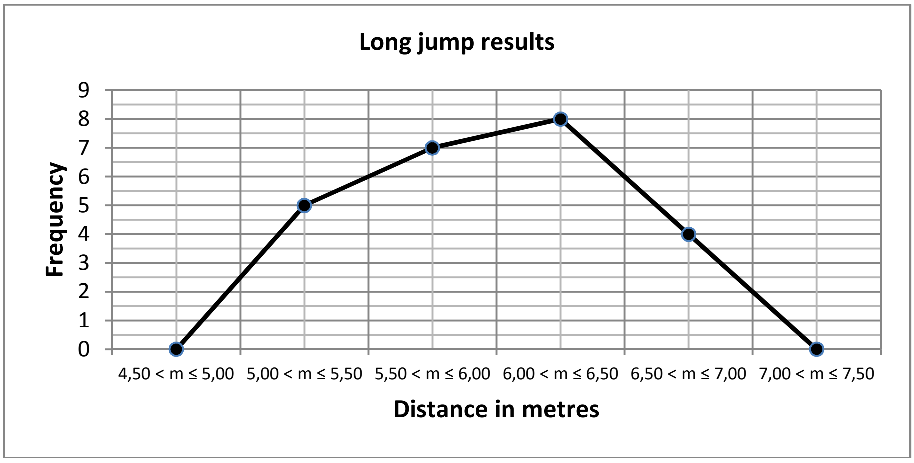
Exercise 2.2

1)

a)

Distance in metres (in m)	Frequency	Midpoints (in m)
$4{,}50 < m \leq 5{,}00$	0	4,75
$5{,}00 < m \leq 5{,}50$	5	5,25
$5{,}50 < m \leq 6{,}00$	7	5,75
$6{,}00 < m \leq 6{,}50$	8	6,25
$6{,}50 < m \leq 7{,}00$	4	6,75
$7{,}00 < m \leq 7{,}50$	0	7.25

b)

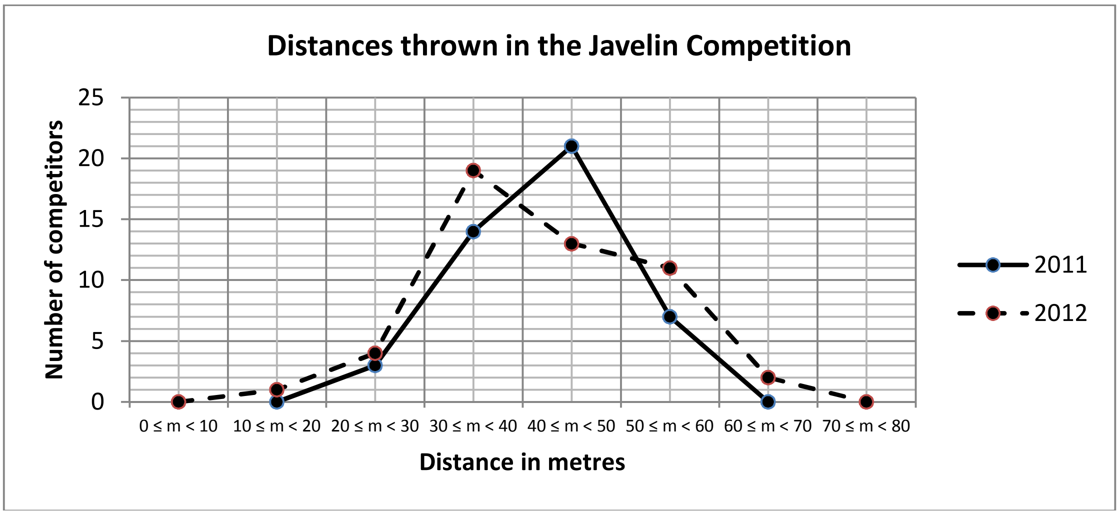


c) The modal interval is $6{,}00 \text{ m} < m \leq 6{,}50 \text{ m}$

2)

a)

Distance thrown in metres (m)	Midpoints	Number of competitors 2011	Number of competitors 2012
$0 \le m < 10$	5	0	0
$10 \le m < 20$	15	0	1
$20 \le m < 30$	25	3	4
$30 \le m < 40$	35	14	19
$40 \le m < 50$	45	21	13
$50 \le m < 60$	55	7	11
$60 \le m < 70$	65	0	2
$70 \le m < 80$	75	0	0



b) In 2011, $\frac{21+7}{45} = \frac{28}{45} = 62\%$ of the competitors threw the javelin for more than 50 metres.

In 2012, $\frac{13+11+2}{50} = \frac{26}{50} = 52\%$ of the competitors threw the javelin more than 50 metres.
So more competitors in 2011 threw more than 50 m than in 2012.

In 2011, $\frac{7}{45} = 16\%$ of the competitors threw the javelin for more than 70 metres.

In 2012, $\frac{13}{50} = 26\%$ of the competitors threw the javelin more than 70 metres.
So more competitors in 2012 threw more than 70 m than in 2012.

So, although there were more people threw long distances in 2011 than in 2012, there were more people in 2012 than in 2011 who threw very far.

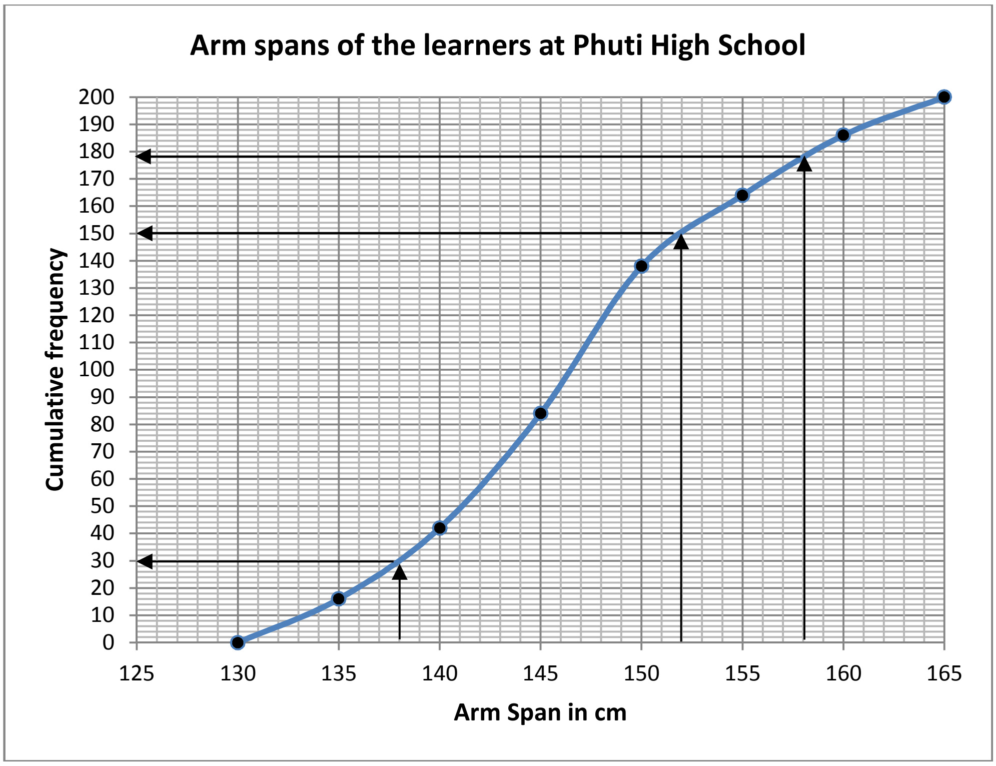
Exercise 2.3

1)

a) *Don't forget to add in another interval with a frequency of 0.*

Arm span in cm	Frequency	Cumulative Frequency	Ordered Pairs
$125 < h \le 130$	0	0	(130; 0)
$130 < h \le 135$	16	16	(135; 16)
$135 < h \le 140$	26	42	(140; 42)
$140 < h \le 145$	42	84	(145; 84)
$145 < h \le 150$	54	138	(150; 138)
$150 < h \le 155$	26	164	(155; 164)
$155 < h \le 160$	22	186	(160; 186)
$160 < h \le 165$	14	200	(165; 200)

b)



c) From the ogive, the number of learners with an arm span of 152 cm or less = 150

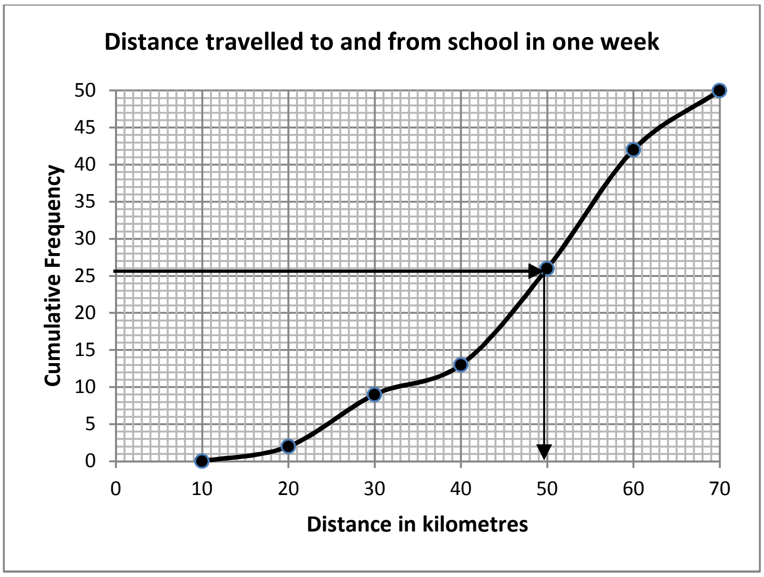
d) Approximately 178 learners are 158 cm or less.
Approximately 30 learners are 138 cm or less.
Number of learners between 138 cm and 158 cm ≈ 178 – 30 = 148

2)

a) *Don't forget to add in another interval with a frequency of 0.*

Number of kilometres	Number of learners	Cumulative frequency	Ordered pairs
$0 < x \leq 10$	0	0	(10; 0)
$10 < x \leq 20$	2	2	(20; 2)
$20 < x \leq 30$	7	9	(30; 9)
$30 < x \leq 40$	4	13	(40; 13)
$40 < x \leq 50$	13	26	(50; 26)
$50 < x \leq 60$	16	42	(60; 42)
$60 < x \leq 70$	8	50	(70; 50)
	TOTAL = 50		

b)



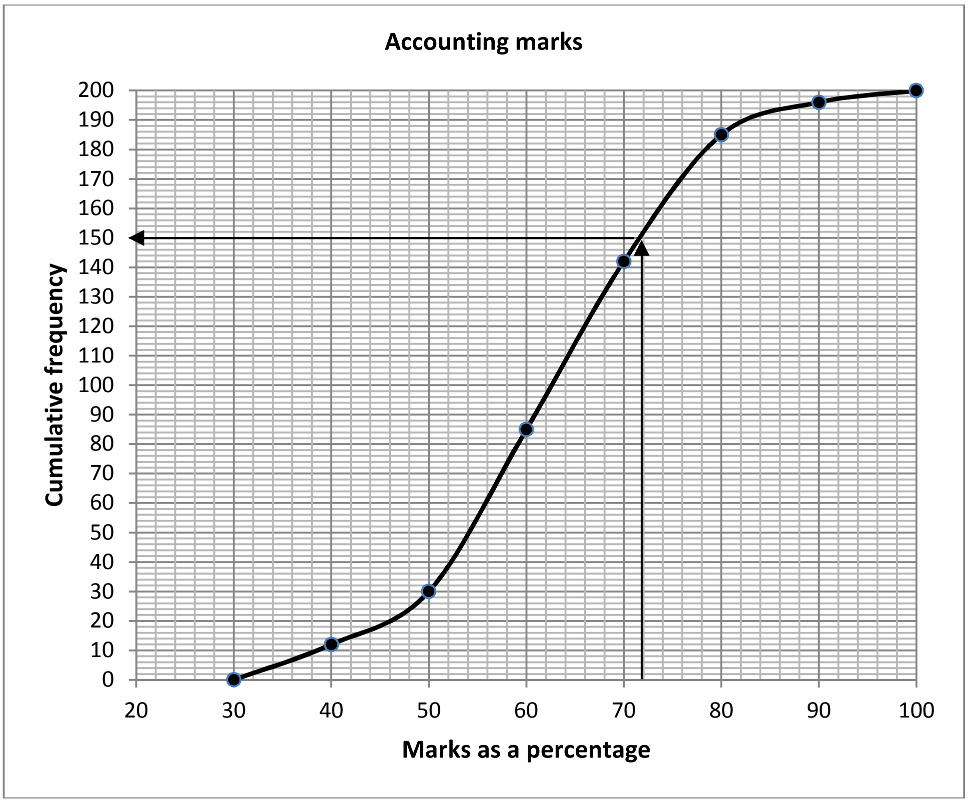
c) 50 learners travelled by car. The median lies between the 25[th] learner and the 26[th] learner. From the graph we can see that the median ≈ 50 km per week.

3)

a) *Don't forget to add in another interval with a frequency of 0.*

	Frequency	**Cumulative Frequency**	**Ordered pairs**
$20 < x \leq 30$	0	0	(30; 0)
$30 < x \leq 40$	12	12	(40; 12)
$40 < x \leq 50$	18	30	(50; 30)
$50 < x \leq 60$	55	85	(60; 85)
$60 < x \leq 70$	57	142	(70; 142)
$70 < x \leq 80$	43	185	(80; 185)
$80 < x \leq 90$	11	196	(90; 196)
$90 < x \leq 100$	4	200	(100; 200)

b)

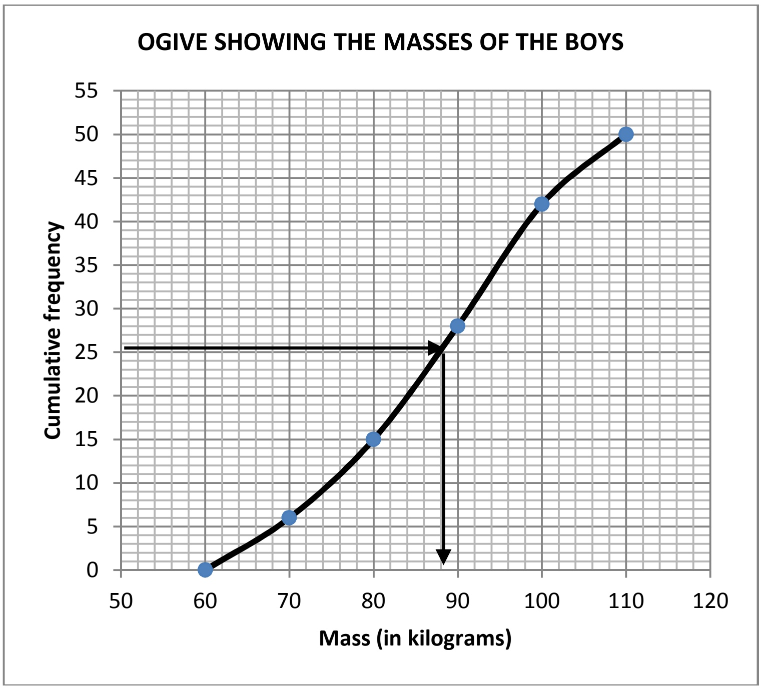


c) Number of learners who scored 72% or less for the examination ≈ 152.
Number of learners who scored 72% or more for the examination ≈ 200 – 152 = 48

4)

a) From the graph:
Approximately 28 boys had a mass of 90 kg or less
Approximately 42 boys had a mass of 100 kg or less
Number of boys with a mass between 90 and 100 kg = 42 – 28 = 14

b)



The graph shows the mass of 50 boys. The median lies between the 25th and 26th boys.

The median ≈ 88 kg

c) From the graph: Number of boys with a mass less than 80 kg ≈ 15

Exercise 2.4

1)

a)

i) $\bar{x} = \frac{\sum x}{n} = \frac{203+214+187+188+196+199+205+203+199+194+206}{11}$ = 199,4545… ≈ 199,5 cm

ii)

x	$x - \bar{x}$	$(x - \bar{x})^2$
203	203 – 199,5 = 3,5	$(3,5)^2$ = 12,25
214	214 – 199,5 = 14,5	$(14,5)^2$ = 201,25
187	187 – 199,5 = – 12,5	$(– 12,5)^2$ = 156,25
188	188 – 199,5 = – 11,5	$(– 11,5)^2$ = 132,25
196	196 – 199,5 = – 3,5	$(– 3,5)^2$ = 12,25
199	199 – 199,5 = – 0,5	$(– 0,5)^2$ = 0,25
205	205 – 199,5 = 5,5	$(5,5)^2$ = 30,25
203	203 – 199,5 = 3,5	$(3,5)^2$ = 12,25
199	199 – 199,5 = – 0,5	$(– 0,5)^2$ = 0,25
194	194 – 199,5 = – 5,5	$(– 5,5)^2$ = 30,25
206	206 – 199,5 = 6,5	$(6,5)^2$ = 42,25
n = 11		$\sum(x - \bar{x})^2$ = 629,75

iii) Standard deviation = $\sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{629,75}{11}} = \sqrt{57,25}$ = 7,5663… ≈ 7,6 cm.

b) For Team B, $\sigma = \sqrt{\text{variance}} = \sqrt{875}$ = 29,5803… ≈ 29,6 cm.

c) The standard deviation of Team B is greater than that of Team A. This shows that the lengths of the arm spans of team B are more variable and spread out than those from Team A.

2)

a) $\bar{x}$ = 22 minutes

b) Standard deviation = 3,9496… ≈ 4,0 minutes

3)

a) Team A: $\bar{x}$ = 70,32 kg and σ = 7,5350… ≈ 7,5 kg
Team B: $\bar{x}$ = 76,68 kg and σ = 14,6743… ≈ 14,7 kg

b) Both the mean and the standard deviation for team B is greater than that of A.
This does not mean that team B plays better than team A.
The mass of a rugby players do not guarantee better play, hence the mean and standard deviation are not good measures for determining which team plays better.
What these measures tell us that, on a whole, the players in Team B are heavier than the players in Team A, and there is a greater spread of masses in Team B than in Team A.

4)

a)

i) For the girls:
$\bar{x}$ = 24,8333… ≈ 24,8 cm
σ =3,0595… ≈ 3,1 cm

For the boys:
$\bar{x}$ = 26,125 ≈ 26,1 cm
σ =2,5950… ≈ 2,6 cm

b) On the whole, the foot lengths of the girls are less than the foot lengths of the boys.
However, the foot lengths of the boys are more closely grouped together than of the girls. The foot lengths of the girls are more spread out.

Exercise 2.5

1)

a) For both box and whisker diagrams, the minimum values are 2 and the lower quartiles are 4. 25% of the values in both sets lies between 2 and 4.

b) Data set A is symmetrical. The median is in the middle of the box and the whiskers are equal in length.

c) Data set B is skewed to the right (more stretched out to the right). The data is clustered more closely on the left and more spread out on the right.

2)

a) The minimum value lies in the interval $5 < t \leq 10$
Minimum value ≈ 7,5 minutes

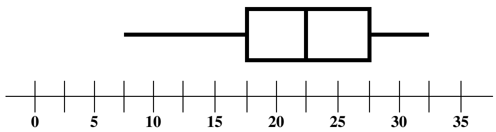
The maximum value lies in the interval $30 < t \leq 35$
Maximum value ≈ 32,5 minutes

47 learners answered the question.
$47 \div 2 = 23\frac{1}{2}$, so the median is the 24^{th} learner.
Add the frequencies to find the interval that the median lies in: $1 + 5 + 9 + 13 = 28$
So the median lies in the interval $20 < t \leq 25$.
Median = 22,5 minutes.

There are 23 learners below the median.
$23 \div 2 = 11\frac{1}{2}$, so Q_1 is the 12^{th} term.
Add the frequencies to find the interval that Q_1 lies in: $1 + 5 + 9 = 15$
So Q_1 lies in the interval $15 < t \leq 20$.
Q_1 ≈ 17,5 minutes

There are 23 learners above the median.
$23 \div 2 = 11\frac{1}{2}$, so Q_3 is 12 learners above the median. The $(24 + 12)^{th}$ term = 36^{th} term
Add the frequencies to find the interval that Q_3 lies in: $1 + 5 + 9 + 13 + 11 = 39$
So Q_3 lies in the interval $25 < t \leq 30$.
Q_3 ≈ 27,5 minutes

b)



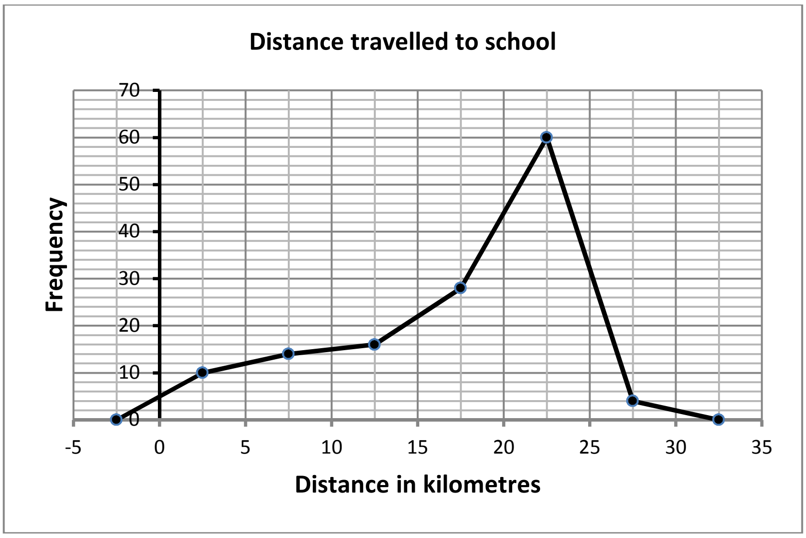
c) The median is in the middle of the box but the left hand whisker is slightly longer than the right hand whiskers. So the data is skewed slightly left. The data is clustered more closely on the right and is slightly more spread out on the left.

3)

a)

Distance in kilometres (x)	**Frequency (f)**	**Midpoint of intervals (X)**
$-5 < x \leq 0$	0	– 2,5
$0 < x \leq 5$	12	2,5
$5 < x \leq 10$	29	7,5
$10 < x \leq 15$	13	12,5
$15 < x \leq 20$	63	17,5
$20 < x \leq 25$	12	22,5
$25 < x \leq 30$	3	27,5
$30 < x \leq 35$	0	32,5

b)



c) 132 learners took part in the survey
$132 \div 2 = 66$, so the median is half-way between the 66th and 67th term.
Add the frequencies to find the interval that the median lies in: $12 + 29 + 13 + 63 = 117$
So the median lies in the interval $15 < x \leq 20$
Median ≈ 17,5

d) ***Mean ≈ 14,1287... ≈ 14,1***

e) ***Mean – median = 14,1 – 17,5 = – 3,4 < 0***

f) The frequency polygon shows data that is negatively skewed. The data is clustered more closely on the right and is slightly more spread out on the left. The difference between the mean and the median is negative which also indicates that the data is negatively skewed.

Exercise 2.6

1) There are 14 data items.
The median lies between the 7th and the 8th terms.
Median $= \frac{14{,}6+14{,}7}{2} = \frac{29{,}3}{2} = 14{,}65$
There are 7 data items before the median. Q_1 is the 3rd term. $Q_1 = 14{,}4$
There are 7 data items after the median. Q_3 is the 3rd term after the median. $Q_3 = 15{,}1$
IQR = 15,1 – 14,4 = 0,7
Lower outlier $< Q_1 - 1{,}5 \times \text{IQR}$
$< 14{,}4 - 1{,}5 \times 0{,}7$
$< 13{,}35$
Upper outlier $> Q_3 + 1{,}5 \times \text{IQR}$
$> 15{,}1 + 1{,}5 \times 0{,}7$
$> 16{,}15$
Therefore 10,2; 16,4 and 18,9 are outliers.

2)

a) First arrange the data in order: 2 3 4 5 6 | 6 7 8 9 9 | 9 9 10 11 12 | 12 13 14 15 16
There are 20 data items. The median lies half way between the 10th and 11th data items.
So the median $= \frac{9+9}{2} = 9$ assignments
There are 10 data items less than the median, so Q_1 lies between the 5th and 6th data items.
So $Q_1 = \frac{6+6}{2} = 6$ assignments
There are 10 data items more than the median, so Q_3 lies between the 15th and 16th data items.
So $Q_3 = \frac{12+12}{2} = 12$ assignments
IQR = 12 – 6 = 6 assignments.

b) Lower outlier $< Q_1 - 1{,}5 \times IQR$
$< 6 - 1{,}5 \times 6$
< -3
Upper outlier $> Q_3 + 1{,}5 \times IQR$
$> 12 + 1{,}5 \times 6$
> 21
Therefore there are ***no outliers***.

3)
a) First arrange the data in order: 9 12 12 **12** 12 12 13 **13** 14 14 14 **14** 15 15 19
There are 15 data items. The median is the 8^{th} data item. So the median = 13 years old
There are 7 data items less than the median, so Q_1 is the 4^{th} data item. So Q_1 = 12 years old
There are 7 data items more than the median, so Q_3 is the 12^{th} data item. So Q_3 = 14 years old
IQR = 14 – 12 = 2 years
So, the five number summary is 9; 12; 13; 14; 19.

b) Lower outlier $< Q_1 - 1{,}5 \times IQR$
$< 12 - 1{,}5 \times 2$
< 9 years old
Upper outlier $> Q_3 + 1{,}5 \times IQR$
$> 14 + 1{,}5 \times 2$
> 17
So 19 is an outlier.

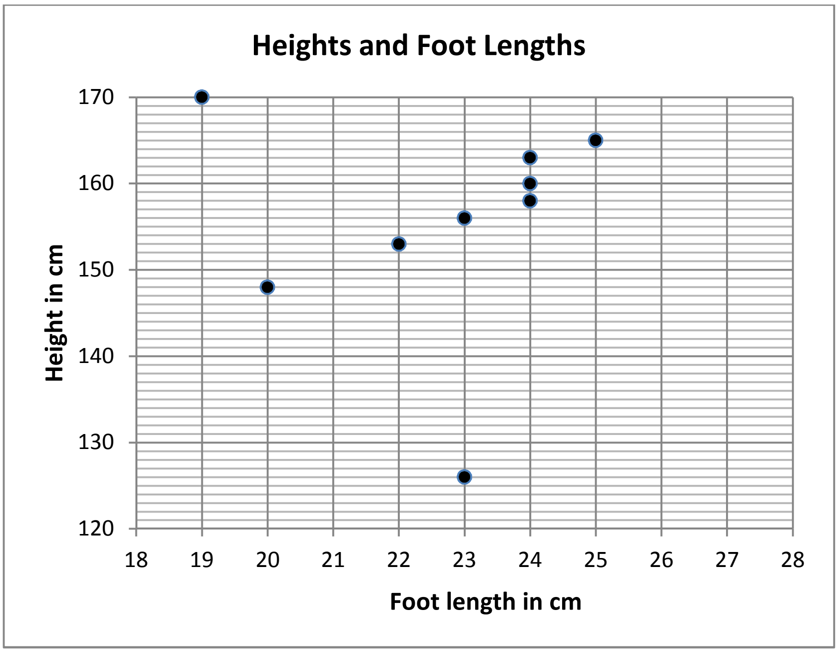
CHAPTER 3 – GRADE 12 DATA HANDLING

Exercise 3.1
1)
a) Zero correlation
b) A moderate negative correlation
c) A strong positive correlation
d) A non-linear correlation (a quadratic relationship)
e) A strong negative correlation
f) A moderate positive correlation
g) A non-linear correlation (a quadratic relationship)
h) A non-linear correlation (an exponential relationship)

2)
a) The tallest girl (170 cm) wears a size 7 shoe
b) The girl with the largest shoe size (8) is 164 cm = 1,64 m
c) The shortest girl does not wear the smallest shoes. The shortest girl wears size 4 ½ shoes.
d) The shoe sizes of the shorter girls are generally smaller than those of taller girls
e) There is a moderate positive correlation.

3)

a)



b) Points (19; 170) and (23; 126) are outliers. These points stand out and are clearly not part of the main trend of the other points.

c) Since the points slope upwards from left to right, the graph shows a strong positive correlation between foot length and height of the learners. Because the points cluster along an obvious straight line, it shows that it is a strong positive linear correlation.

<u>Exercise 3.2</u>

1)

a) $r = -0{,}5$

b) $r = 0{,}5$

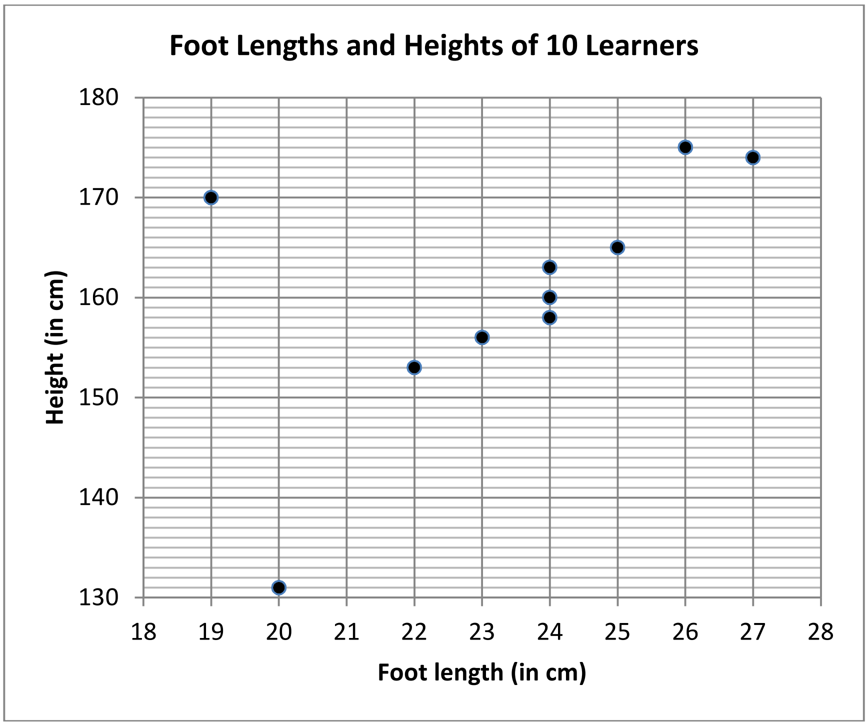
c) $r = 0{,}95$

d) $r = 0$

e) $r = -0{,}95$

2)

a)



b)

Foot length (x)	Height (y)	$x\,y$	x^2	y^2
22	153	3366	484	23409
19	170	3230	361	28900
24	160	3840	576	25600
20	131	2620	400	17161
23	156	3588	529	24336
27	174	4698	729	30276
24	158	3792	576	24964
24	163	3912	576	26569
26	175	4550	676	30625
25	165	4125	625	27225
$\sum x = 234$	$\sum y = 1\,605$	$\sum xy = 37\,721$	$\sum x^2 = 5\,532$	$\sum y^2 = 259\,065$

With these values and $n = 10$, the correlation coefficient is:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{10\times 37\,721 - (234)(1\,605)}{\sqrt{10\times 5\,532 - (234)^2}\sqrt{10\times 259\,065 - (1\,605)^2}}$$

$$r = \frac{1\,640}{\sqrt{564}\sqrt{14\,625}}$$

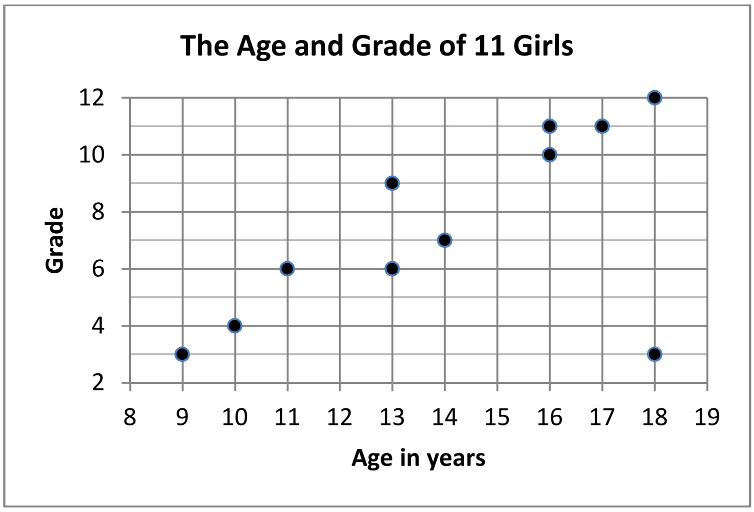
$= 0{,}57102\ldots$

$\approx 0{,}57$

The value of r tells us that there is a weak positive linear correlation between these foot lengths and heights.

3)

a)

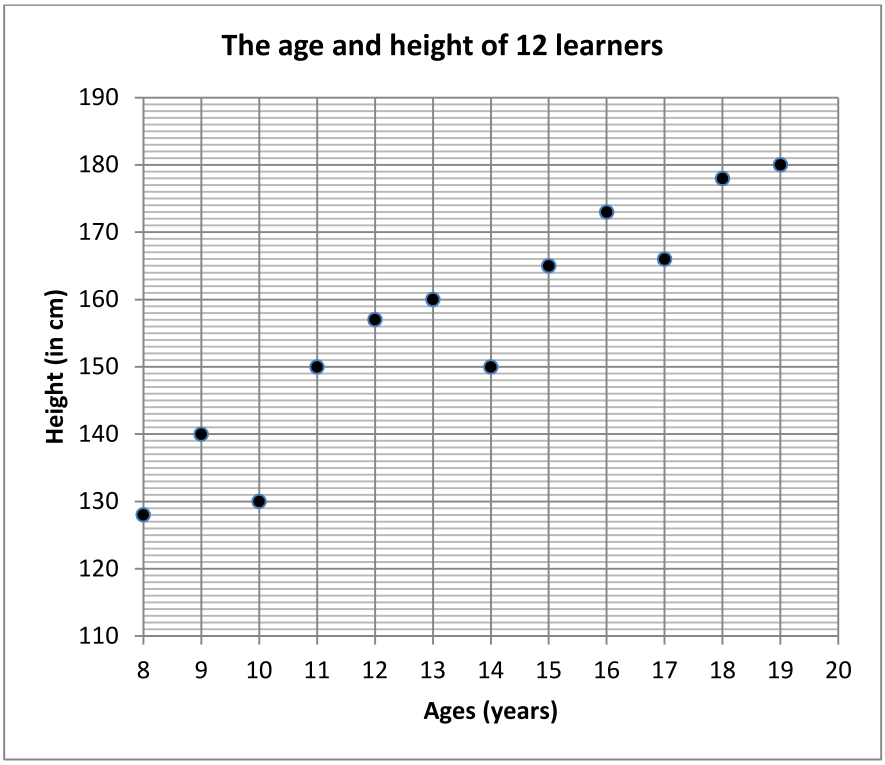


b) $r = 0{,}6011\ldots \approx 0{,}60$

c) There is a weak positive correlation between the age and grade of the eleven girls.

4)

a)

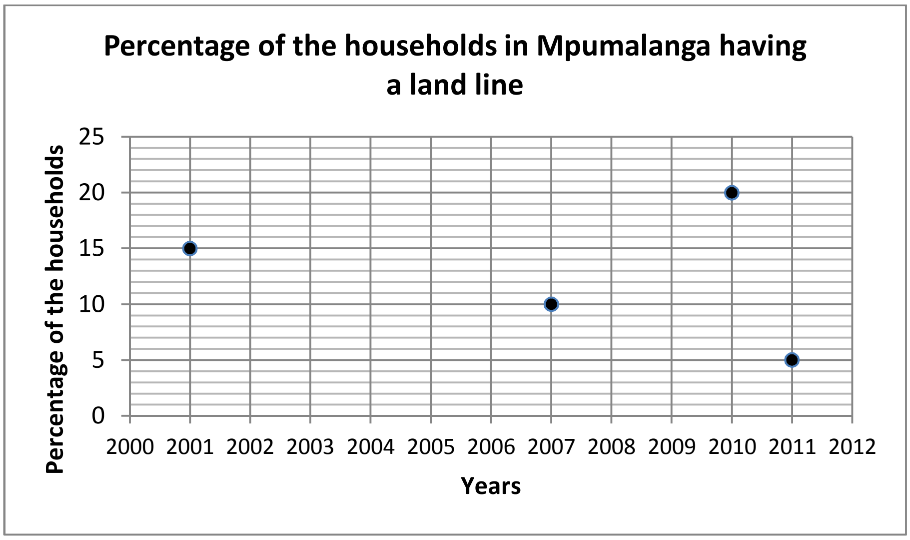


b) $r = 9379\ldots \approx 0{,}94$

c) The value of r indicates that there is a strong positive linear correlation between the age and height of the learners in this sample.

5)

a)



b) $r = -0{,}2581\ldots \approx -0{,}26$

c) There is no significant correlation between the years and the percentage of the households having a land line. Looking at the points from 2001 and 2007, we see that there was a decline in the use of landline telephones. Also between 2010 and 2011 there was a decline. This shows that generally there is a negative correlation between the years and the number of household using landline telephones. This could be as a result that people use cell phones hence the decrease in the use of landline telephones. The point (2010; 20) is likely to be an outlier.

d) If we take the point (2010; 20) as an outlier, we can predict that by 2015 there will be about 2% households that will still be using landline telephones.

Exercise 3.3

1)

a) Variable 1 is the wrist size and Variable 2 is the arm span

b) (150; 130) (158; 145) (161; 159) (166; 162) (167; 153) (170; 172) (172; 160) (177; 190)

c) $r = 0{,}9087\ldots \approx 0{,}91$

There is a strong positive linear correlation

d)

Wrist size in mm	**Arm span in cm**				
x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$
150	130	–15,125	–28,875	436,734 375	228,765 625
158	145	–7,125	–13,875	98,859 375	50,765 625
161	159	–4,125	0,125	–0,515 625	17,015 625
166	162	0,875	3,125	2,734 375	0,765 625
167	153	1,875	–5,875	–11,015 625	3,515 625
170	172	4,875	13,125	63,984 375	23,765 625
172	160	6,875	1,125	7,734 375	47,265 625
177	190	11,875	31,125	369,609 375	141,015 625
$\bar{x}$ = 165,125	$\bar{y}$ = 158,875	0	0	$\sum(x-\bar{x})(y-\bar{y})$ = 968,125	$\sum(x-\bar{x})^2$ = 512,875

Use the formula for b to get the slope: $b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{968{,}125}{512{,}875} = 1{,}8876\ldots \approx 1{,}89$

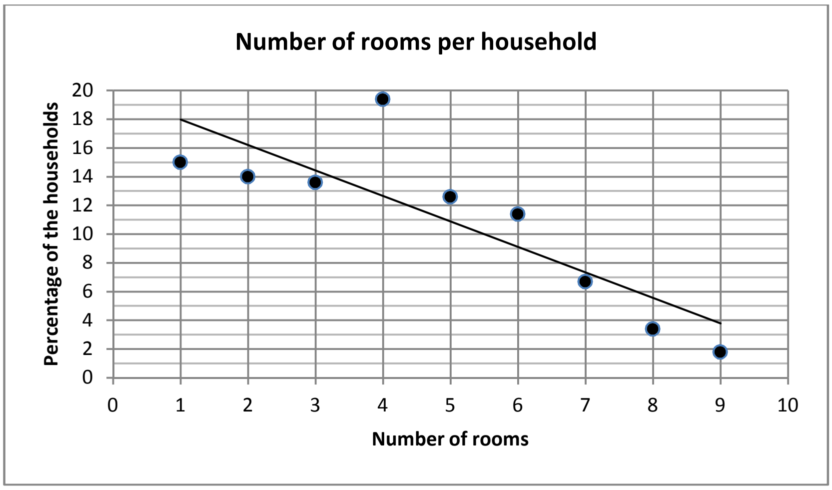
Use the formula for a to get the cut on the y-axis:

$a = \overline{y} - b\,\overline{x} = 158{,}875 - \frac{968{,}125}{512{,}875}(165{,}125) = -152{,}8220\ldots \approx -152{,}82$

The equation of the line of regression is therefore: $\hat{y} = -152{,}82 + 1{,}89\,x$

2)

a)



b) $r = -0{,}8425\ldots \approx -0{,}84$

There is a strong negative correlation

c) $a = 19{,}7444\ldots \approx 19{,}74$

$b = -1{,}7733\ldots \approx -1{,}77$

$\hat{y} = 19{,}75 - 1{,}77\,x$

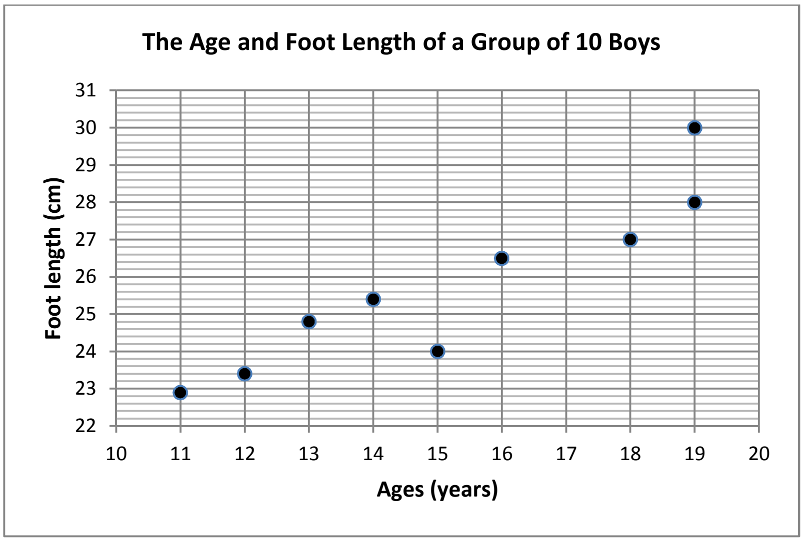
d) When $x = 1$, $y = 19{,}74 - 1{,}77(1) = 17{,}97 \approx 18$

When $x = 9$, $y = 19{,}74 - 1{,}77(9) = 3{,}81 \approx 4$

Plot the points (1; 18) and (9; 4) and draw the line through the two points.

3)

a)



b) $r = 0{,}9222\ldots \approx 0{,}92$

There is a strong positive correlation.

c) $a = 14{,}8976\ldots \approx 14{,}89$

$b = 0{,}7147\ldots \approx 0{,}71$

$\hat{y} = 14{,}89 - 0{,}71\,x$

d)
  i) For a 9 year old boy, the foot length = 14,89 – 0,71(9) = 21,3 cm
  ii) For a boy of 17 years and 6 months or 17,5 years, the foot length = 14,89 – 0,71(17,5)
  = 27,4 cm

e) Prediction (i) may not be accurate because a boy of 9 years is outside the given domain. Prediction (ii) is good because 17 years and 6 months lies in the given range of the domain.

4)
a) A = 73,7816… ≈ 73,78
B = 1,0003…
The equation of the exponential regression function is y = 73,78 $(1,0003)^x$

b) When there are 15 000 households, we can use the exponential regression function to predict that 7 566,3685… ≈ 7 566 of these households are headed by single parents.

c) The total number of households under observation range from 5 000 to 10 000. 15 000 falls out of this range. Hence the number of households headed by single parents is a prediction, not an actual measurement.

CHAPTER 4 – GRADE 10 PROBABILITY

Exercise 4.1

1)
a) S = {1; 2; 3; 4; 5; 6}
b)
  i) $P(6) = \frac{n(6)}{n(S)} = \frac{1}{6} = 0{,}1666\ldots \approx 0{,}17 \approx 16{,}7\%$
  ii) Favourable outcomes = {1; 3; 5}
  $P(odd\ no) = \frac{n(odd\ numbers)}{n(S} = \frac{3}{6} = \frac{1}{2} = 0{,}50 = 50{,}0\%$
  iii) Favourable outcomes = { } (an empty set), so n(7) = 0
  $P(7) = \frac{n(7)}{n(S)} = \frac{0}{6} = 0{,}00 = 0{,}0\%$
  iv) Favourable outcomes = {3; 4; 5; 6}
  $P(more\ than\ 2) = \frac{n(more\ than\ 2)}{n(S)} = \frac{4}{6} = \frac{2}{3} = 0{,}6666\ldots \approx 0{,}67 \approx 66{,}7\%$
  v) Favourable outcomes = {1; 2; 3; 4; 5; 6}
  $P(less\ than\ 10) = \frac{n(less\ than\ 10)}{n(possible\ outcomes)} = \frac{6}{6} = 1{,}00 = 100{,}0\%$

2)
a) S = {blue; pink; red; green}
b)
  i) $P(green) = \frac{n(green)}{n(S)} = \frac{1}{4} = 0{,}25 = 25{,}0\%$
  ii) $P(yellow) = \frac{n(yellow)}{n(S)} = \frac{0}{4} = 0{,}00 = 0{,}0\%$

3)
a) S = {M; A; T; H; E; M; A; T; I; C; S}
b)
  i) Favourable outcomes = {M; M}
  $P(M) = \frac{n(M)}{n(S)} = \frac{2}{11} = 0{,}1818\ldots \approx 0{,}18 \approx 18{,}2\%$
  ii) Favourable outcomes = {A; E; A; I}
  $P(vowels) = \frac{n(vowels)}{n(S)} = \frac{4}{11} = 0{,}3636\ldots \approx 0{,}36 \approx 36{,}4\%$

4)
a) S = {3; 4; 7; 9; 10; 11}
b)
  i) Favourable outcomes = {3; 7; 9; 11}
  $P(odd\ number) = \frac{n(odd\ number)}{n(S)} = \frac{4}{6} = \frac{2}{3} = 0{,}6666\ldots \approx 0{,}67 \approx 66{,}7\%$
  ii) Favourable outcomes = {3; 7;11}
  $P(prime\ number) = \frac{n(prime\ number)}{n(S)} = \frac{3}{6} = \frac{1}{2} = 0{,}50 = 50\%$

iii) Favourable outcomes = {4; 9}
$P(square\ number) = \frac{n(square\ number)}{n(S)} = \frac{2}{6} = \frac{1}{3} = 0{,}3333\ldots \approx 0{,}33 \approx 33{,}3\%$

5)
a) 18 boys and 12 girls add up to 30 children. So $n(S)$ = 30
b)
i) $P(boy) = \frac{n(boys)}{n(S)} = \frac{18}{30} = \frac{3}{5} = 0{,}60 = 60{,}0\%$
ii) $P(girl) = \frac{n(girls)}{n(S)} = \frac{12}{30} = \frac{2}{5} = 0{,}40 = 40{,}0\%$

6)
a) One quarter of the circle is white and three quarters of the circle is shaded.
So the sample space is four quarters of the circle and $n(S) = 4$
b) $P(land\ on\ shaded\ area) = \frac{n(shaded\ areas)}{n(S)} = \frac{3}{4} = 0{,}75 = 75{,}0\%$

Exercise 4.2

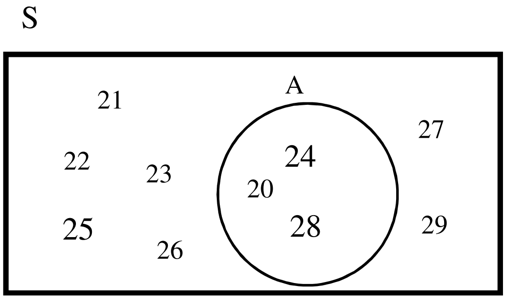
1)
a) The sample set consists of 443 590 learners, so $n(S) = 443\ 590$
b)
i) $P(come\ by\ bus) = \frac{n(come\ by\ bus)}{n(S)} = \frac{3\ 052}{443\ 590} = 0{,}006\ 88\ldots \approx 0{,}69\%$
ii) $P(come\ by\ bicycle) = \frac{n(come\ by\ bicycle)}{n(S)} = \frac{1\ 415}{443\ 590} = 0{,}003\ 18 \approx 0{,}32\%$
iii) n(come by bus or bicycle) = 3 053 + 1 415 = 4 467
$P(come\ by\ bus\ or\ bicycle) = \frac{n(come\ by\ bus\ or\ bicycle)}{n(S)} = \frac{4\ 467}{443\ 590} = 0{,}010\ 07\ldots \approx 1{,}01\%$
c) The probabilities in b) are low because these learners live only 1 km away from school which means they could walk to school. Walking is cheaper than going by bus and bicycles cost money so walking is the better option.

2)
a) Add up all the numbers of learners. 1 729 483 learners were surveyed.
So $n(S)$ = 1 729 483
b)
i) $P(speaks\ mainly\ English) = \frac{n(speaks\ mainly\ English)}{n(S)} = \frac{445\ 492}{1\ 729\ 483} = 0{,}257\ 58\ldots \approx 25{,}8\%$
ii) $n(speaks\ mainly\ isiZulu\ or\ Afrikaans) = 270\ 062 + 202\ 570 = 472\ 632$
$P(speaks\ mainly\ isiZulu\ or\ Afrikaans) = \frac{n(speaks\ mainly\ isiZulu\ or\ Afrikaans)}{n(S)}$
$= \frac{472\ 632}{1\ 729\ 483} = 0{,}273\ 28\ldots \approx 27{,}3\%$
c) Census 2011 gives a better probability estimate. This census involved a larger number of people than the 2009 Census@School. (The more data you have, the better your estimate of the probability).

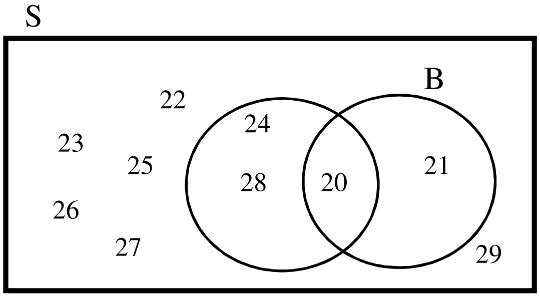
Exercise 4.3

1)
a)
i) S = {20; 21; 22; 23; 24; 25; 26; 27; 28; 29}
ii) A = {20; 24; 28}
iii) B = {20; 21}
iv) C = {20; 25}
v) D = {21; 24; 27}

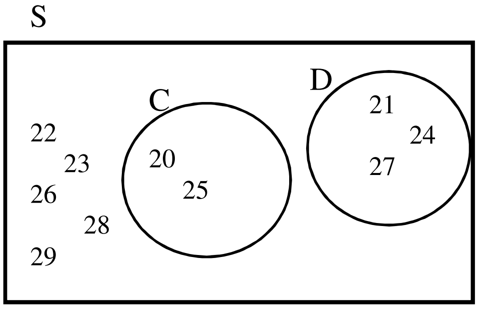
b) i)



ii)



iii)



- *Check that all the diagrams contain all the numbers in the sample space.*

2)

a) S = {1; 2; 3; 4; 5; 6; 7; 8}

b) i) P={2; 3; 5;7}

ii) E = {2; 4; 6; 8}

iii) F = {4; 5; 6; 7; 8}

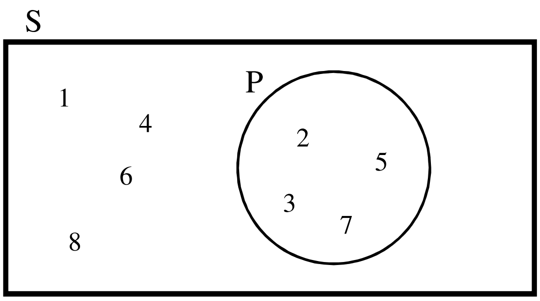
iv) G = {2; 4; 6; 8}

v) H = {1; 3; 5; 7}

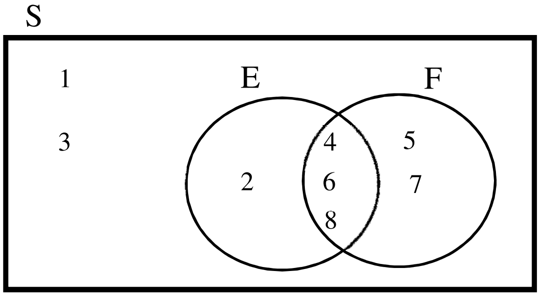
vi) Intersection of E and F = {4; 6; 8}

vii) G and H don't intersect. So there are no elements in the intersection.

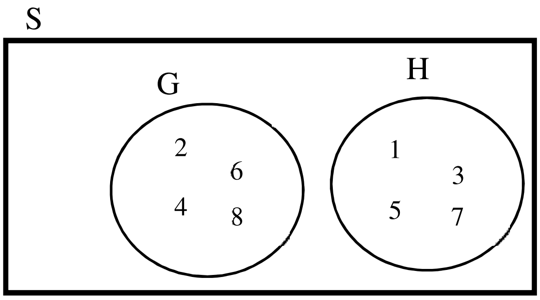
c) i) Event P



ii) Events E and F



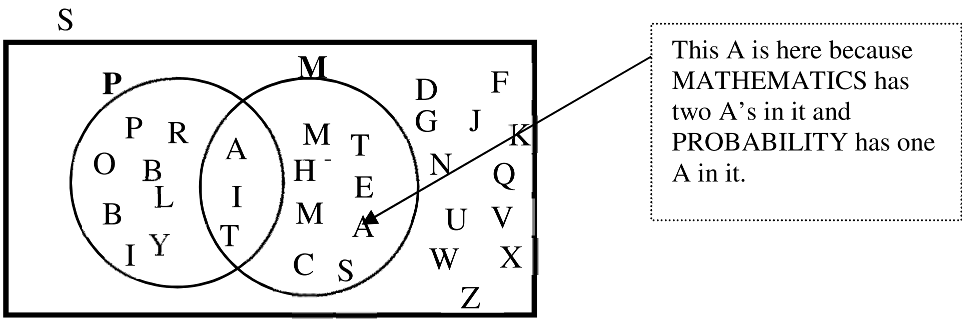
iii) Events G and H



- *Check that all the diagrams contain all the numbers in the sample space.*

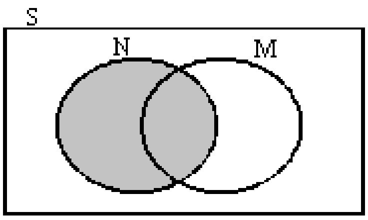
3 a) Letters common to PROBABILITY and MATHEMATICS = {A; I; T}

b)

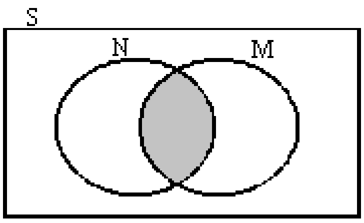


<u>Exercise 4.4</u>

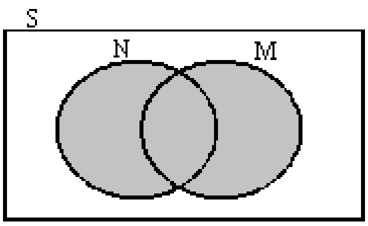
1) N



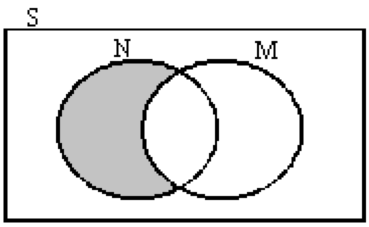
2) N ***and*** M



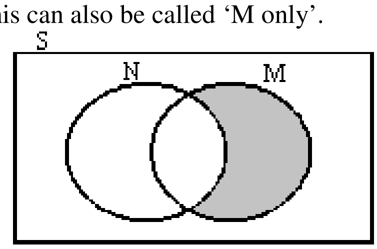
3) N ***or*** M



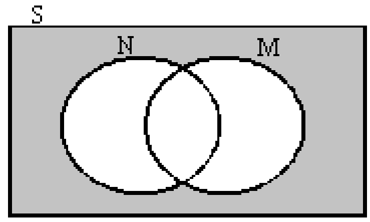
4) N but not M
This can also be called 'N only'.



5) M but not N
This can also be called 'M only'.



4) Neither M nor N



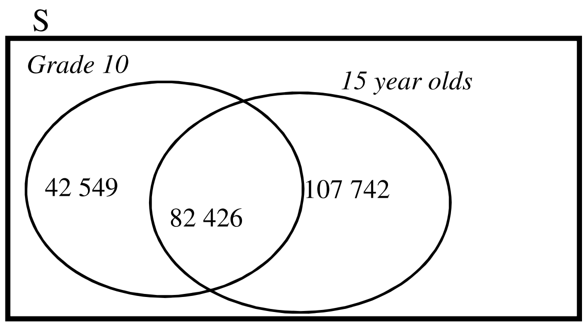
Exercise 4.5

1)

a) 52 cards
b) Number of black cards = $26 - x + x = 26$. (There are 13 clubs and 13 spades)
c) Number of sevens = $x + 4 - x = 4$. (There is one seven of clubs, one seven of spades, one seven of hearts and one seven of diamonds)
d) n(black or seven) = $(26 - x) + x + (4 - x) = 30 - x$
e) $$\begin{aligned} 26 - x + x + 4 - x + 24 &= 52 \\ 54 - x &= 52 \\ 2 &= x \end{aligned}$$
f) Check: 24 + 2 +2 + 24 = 52 (which is correct)

2)

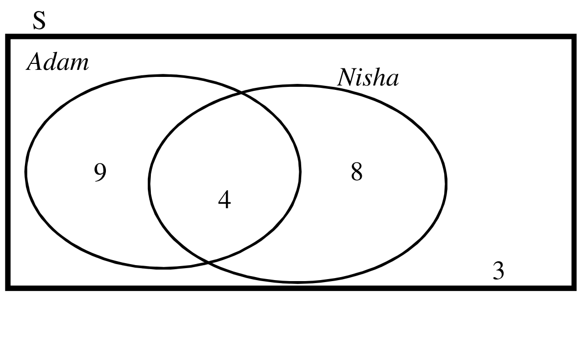
a) n(Grade 10s that were not 15 years old) = 124 975 – 82 426 = 42 549
b) n(15 year olds that were not in Grade 10) = 190 168 – 82 426 = 107 742
c)



d) n(learners in the sample set) = 42 549 + 82 426 +107 742 = 232 717
e) $P(learner\ is\ in\ Grade\ 10\ AND\ is\ 15\ years\ old)$
$$= \frac{n(learners\ in\ Grade\ 10\ AND\ 15\ years\ old}{n(learners\ in\ the\ sample\ set)} = \frac{82\ 426}{232\ 717} \approx 35{,}42\%$$

3)

a) n(learners that went to both parties) = 24 – 3 = 21
b) Of these 21 learners, 13 went to Adam's party and 12 went to Nisha's party and 13 + 12 = 25
n(learners that went to Adam's party *and* to Nisha's party) = 25 – 21 = 4
c)



d)

i) $P(only\ goes\ to\ Adam's\ party) = \frac{n(only\ goes\ to\ Adam's\ party)}{n(sample\ set)} = \frac{9}{24} = 0{,}375$

ii) $P(goes\ to\ both\ parties) = \frac{n(goes\ to\ both\ parties)}{n(sample\ set)} = \frac{4}{24} \approx 0{,}167$

iii) $P(doesn't\ go\ to\ either\ party) = \frac{n(doesn't\ go\ to\ either\ party}{n(sample\ set)} = \frac{3}{24} = 0{,}125$

Exercise 4.6

1)

a) $n(F) = 4$

b) $n(A) = 4$

c) $n(F\ \boldsymbol{and}\ A) = 0$

d) $n(F\ \boldsymbol{or}\ A) = 4 + 4 = 8$

e) $n(F) + n(A) - n(F\ and\ A) = 4 + 4 - 0 = 8$

2) $n(F) + n(A) - n(F\ and\ A) = 8$ ... (from 1 (e))

$n(F\ or\ A) = 8$ ... (from 1 (d))

so $n(F) + n(A) - n(F\ and\ A) = n(F\ or\ A)$

3)

a) The sample set is the pack of playing cards so $n(S) = 52$

b)

i) $P(F) = \frac{n(F)}{n(S)} = \frac{4}{52} = \frac{1}{13}$

ii) $P(A) = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$

iii) $n(F\ \boldsymbol{and}\ A) = 0$

$P(F\ \boldsymbol{and}\ A) = \frac{n(F\ and\ A)}{n(S)} = \frac{0}{52} = 0$

iv) $n(F\ \boldsymbol{or}\ A) = 4 + 4 = 8$

$P(F\ \boldsymbol{or}\ A) = \frac{n(F\ or\ A)}{n(S)} = \frac{8}{52} = \frac{2}{13}$

c) $P(F) + P(A) - P(F\ and\ A) = \frac{1}{13} + \frac{1}{13} - 0 = \frac{2}{13}$

d) $P(F\ or\ A) = \frac{2}{13}$ ... (from 3 b) (iv))

$P(F) + P(A) - P(F\ and\ A) = \frac{2}{13}$ ... (from 3 c))

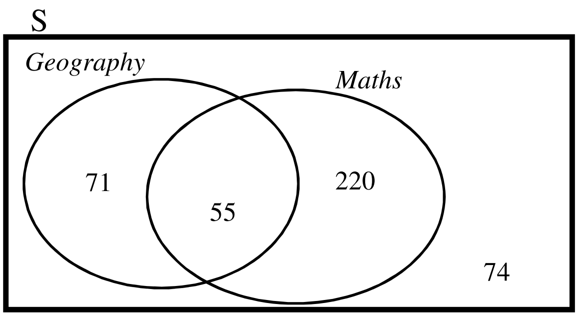
$so\ P(F\ or\ A) = P(F) + P(A) - P(F\ and\ A)$

Exercise 4.7

1)

a) Number doing geography only = 126 – 55 = 71

Number doing maths only = 275 – 55 = 220



b) Number of learners doing both geography ***and*** maths = 55

Learners doing both geography and maths = $\frac{55}{420} \times 100\% \approx 13{,}1\%$

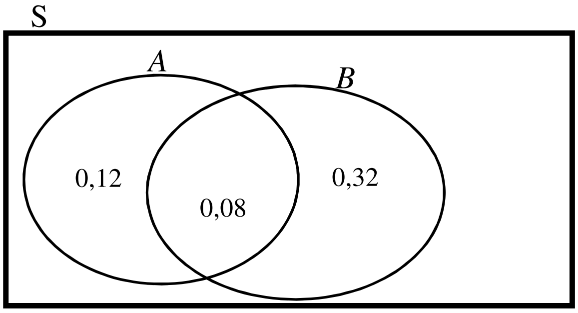
c) Number of learners doing geography ***or*** maths = 71 + 55 + 220 = 346

d) Number of learners doing ***neither*** geography nor maths = 420 – 346 = 74.

2)

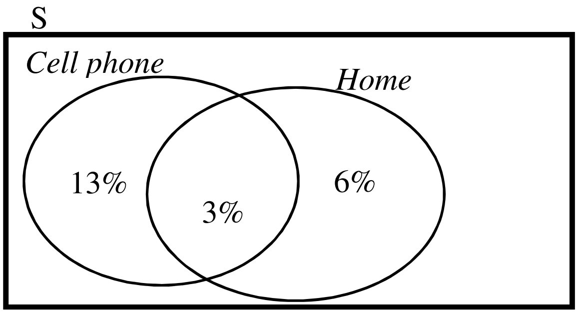
a) $P(A\ only) = 0{,}2 - 0{,}08$

$P(B\ only) = 0{,}4 - 0{,}08 = 0{,}32$



b) $P(A\ or\ B) = 0{,}12 + 0{,}08 + 0{,}32 = 0{,}52$
c) $P(A) + P(B) - P(A\ or\ B) = 0{,}2 + 0{,}4 - 0{,}08 = 0{,}52$
So $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$

3)
a) Percentage who connect to the internet using cell phones only =16% – 3% = 13%
Percentage who connect to the internet using home computers only = 9% –3% = 6%



b) Percentage using neither a cellphone or a home computer for connecting to the internet
= 100% - (13% + 3% + 6%) = 78%
P(Uses neither a cellphone nor a home computer for connecting to the internetl) = 78%
c) The number of South Africans who do not use cell phones or home computers for internet access
$= 78\% \times 14\ 450\ 133$
$= 11\ 271\ 103{,}74$
$\approx 11\ 270\ 000$

4)
a)
i) $P(P\ and\ R) = 0\%$
ii) $P(P\ or\ R)$ = 37,5% + 25% = 62,5%
b) $P(P) + P(R) - P(P\ and\ R)$ = 37,5% + 25% $-$ 0% = 62,5%
So $P(P) + P(R) - P(P\ and\ R) = P(P\ or\ R)$

5)
a)
i) $P(L) = \frac{l+m}{k+l+m+n}$
ii) $P(K\ and\ L) = \frac{l}{k+l+m+n}$
iii) $P(K\ or\ L) = \frac{k+l+m}{k+l+m+n}$
b) $P\ (K) + P\ (L) - P\ (K\ and\ L) = \frac{k+l}{k+l+m+n} + \frac{l+m}{k+l+m+n} - \frac{l}{k+l+m+n} = \frac{k+l+m}{k+l+m+n}$
$so\ P(K\ or\ L) = P\ (K) + P\ (L) - P\ (K\ and\ L)$

Exercise 4.8

1)
a) A playing card cannot be red and black at the same time
b) A playing card can be black and a 10 at the same time. Both the 10 of spades and the 10 of clubs are black and 10 at the same time.
c) A playing card cannot be a King and a Queen at the same time.

2)
a) Mutually exclusive as a playing card cannot be a red card and a black Jack at the same time.

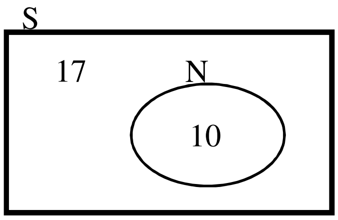
b) NOT mutually exclusive. All cards marked with a diamond are also red so it is possible to draw a card that is both red and marked with a diamond.
c) Mutually exclusive. You cannot toss a coin and get a Head and a Tail at the same time.
d) Mutually exclusive. When you roll a dice you cannot get the events “getting a 3” and “getting a 4” at the same time.
e) NOT mutually exclusive as it is possible that the sandwich that you eat has jam on it.
f) NOT mutually exclusive as there are many people who live in Kwa-Zulu Natal who speak English at home.

Exercise 4.9

1)

a) 27 – 10 = 17 girls do not play netball

b)



c) The events ‘play netball’ and ‘do not play netball’ are complementary. The girls in the group either play netball or they don’t play netball.

d) There are 27 girls in the sample so *n(S)* = 27

e)

i) $P(N) = \frac{n(N)}{n(S)} = \frac{10}{27}$

ii) $P(not\ N) = 1 - P(N) = 1 - \frac{10}{27} = \frac{17}{27}$

2)

a) *n(S)* = 50

b) The elements of A are 1; 4; 9; 16; 25; 36 and 49

c) $P(A) = \frac{n(A)}{n(S)} = \frac{7}{50}$

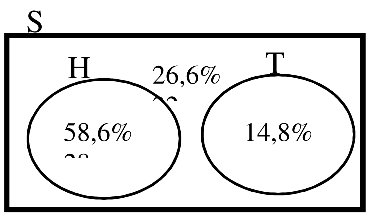
d) $P(not\ A) = 1 - P(A) = 1 - \frac{7}{50} = \frac{43}{50}$

3)

a) H and T are mutually exclusive. Either you live in a house (H) or you live in a traditional dwelling (T), not both.

b) Do not live in a house or a traditional dwelling = 100% – (58,6% + 14,8%) = 26,6%

c)



d)

i) $P(H) = 58{,}6\%$

ii) $P(T) = 14{,}8\%$

iii) $P(not\ H) = 1 - P(H) = 100\% - 58{,}6\% = 41{,}4\%$

iv) $P(not\ T) = 1 - P(T) = 100\% - 14{,}8\% = 85{,}2\%$

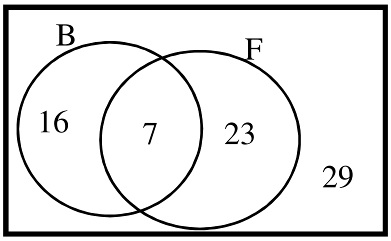
Exercise 4.10

1)

a) The two events are NOT mutually exclusive because some of the people surveyed had eaten both burgers and chicken.

b) Neither eaten burgers nor fried chicken = 75 – (16 + 7 + 23) = 29

c)



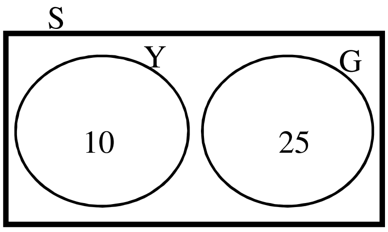
d) $n(S) = 75$

e) $P(eaten\ neither\ burgers\ not\ fried\ chicken) = \frac{29}{75}$ = 38,66...% ≈ 39%

2)

a) The events are mutually exclusive. A counter cannot be yellow and green at the same time.

b)



c) $n(S)$=35

d)

i) $P(Y) = \frac{n(Y)}{n(S)} = \frac{10}{35} = \frac{2}{7}$

ii) $P(G) = \frac{n(G)}{n(S)} = \frac{25}{35} = \frac{5}{7}$ OR $P(G) = P(not\ yellow) = 1 - P(Y) = 1 - \frac{10}{35} = \frac{25}{35} = \frac{5}{7}$

e)

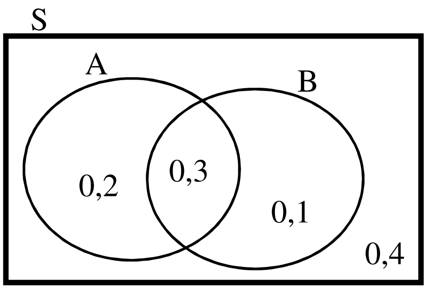
i) $P(A\ and\ B) = \frac{n(A\ and\ B)}{n(S)} = \frac{0}{35} = 0$

ii) $P(Y\ or\ G) = P(Y) + P(G) = \frac{2}{7} + \frac{5}{7} = \frac{7}{7} = 1$

3)

a) $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B) = 0{,}5 + 0{,}4 - 0{,}3 = 0{,}6$

b)



4)

a) $P(black\ card) = \frac{n(black\ cards\ in\ a\ pack)}{n(cards\ in\ a\ pack)} = \frac{26}{52} = \frac{1}{2}$

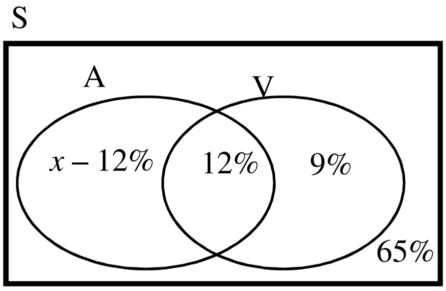
b) $P(seven) = \frac{n(sevens\ in\ a\ pack)}{n(cards\ in\ a\ pack)} = \frac{4}{52} = \frac{1}{13}$

c) $P(black\ seven) = \frac{n(black\ sevens\ in\ a\ pack)}{n(cards\ in\ a\ pack)} = \frac{2}{52} = \frac{1}{26}$

d) $P(not\ a\ black\ seven) = 1 - P(black\ seven) = 1 - \frac{2}{52} = \frac{50}{52} = \frac{25}{26}$

5)

a)



b) $(x - 12\%) + 12\% + 9\% + 65\% = 100\%$
$x + 74\% = 100\%$
$x = 26\%$

c) P(likes Athletics but not Volleyball)
= P(A only) = x – 12% = 26% – 12% = 14%

d) P(likes Athletics or Volleyball) = P(Athletics) + P(Volleyball) – P(Athletics and Volleyball)
= (26% – 12%) + 12% + 9% = 35%
OR P(A or V) = 1 – P(neither of the sports) = 1 – 65% = 35%

6)

a) Percentage of households with toilets that flush
$= 100\% - (3\% + 3\% + 9\% + 19\% + 2\% + 2\% + 5\%) = 57\%$

b)

i) $P(flush\ toilet\ or\ chemical\ toilet)$
$= P(flush\ toilet) + P(chemical\ toilet) - P(flush\ and\ chemical\ toilet)$
$= 57\% + 3\% - 0\% = 60\%$

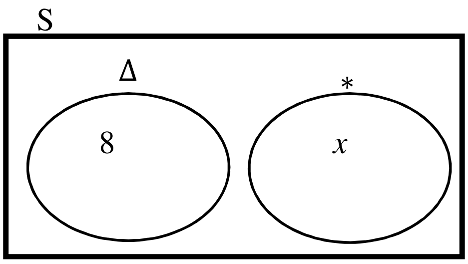
ii) $P(pit\ toilet\ or\ bucket\ toilet)$
$= P(pit\ toilet) + P(bucket\ toilet) - P(pit\ toilet\ and\ bucket\ toilet)$
$= (9\% + 19\%) + 2\% - 0\%$
$= 30\%$

c) More people in South Africa have flush toilets than any other toilet.
Just over half (60%) of South Africans have toilets that flush or chemical toilets.
Almost one third (30%) of South Africans have pit or bucket toilets.
5% of South Africans have no toilets. This is a high number as 5% of 50 000 000 = 2 500 000.

7)

a) It is not possible to draw a card that has both triangles and stars on it. A card has either triangles OR stars on it.

b)



c) $n(S) = 8 + x$
$P(\star) = \dfrac{x}{8 + x} = \dfrac{5}{7}$
So $7x = 5(8 + x)$
$7x = 40 + 5x$
$2x = 40$, and $x = 20$

CHAPTER 5 – GRADE 11 PROBABILITY

<u>Exercise 5.1</u>

1)

a) Number of learners interviewed = 12 + 39 + 57 + 12 = 120

b)

i) $P(M) = \frac{n(M)}{n(S)} = \frac{39+57}{120} = \frac{96}{120} = 80\%$

ii) $P(S) = \frac{n(S)}{n(S)} = \frac{57+12}{120} = \frac{69}{120} = 57{,}5\%$

iii) $P(M \text{ and } S) = \frac{n(M \text{ and } S)}{n(S)} = \frac{57}{120} = 47{,}5\%$

iv) $P(M \text{ or } S) = P(M) + P(S) - P(M \text{ and } S) = 80\% + 57{,}5\% - 47{,}5\% = 90\%$

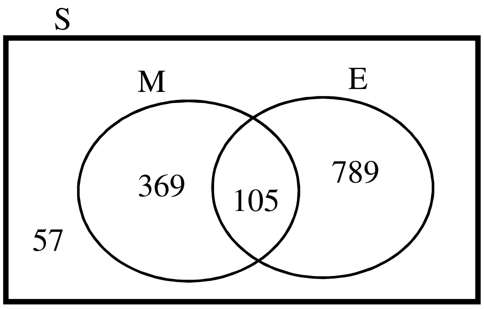
v) $P(not\ M) = 1 - P(M) = 100\% - 80\% = 20\%$

vi) $P(not\ S) = 1 - P(S) = 100\% - 57{,}5\% = 42{,}5\%$

vii) $P\big(not\ (M \text{ or } S)\big) = 1 - P(M \text{ or } S) = 100\% - 90\% = 10\%$

2)

a)



b) $\text{n(neither Maths nor English)} = 1\,320 - (369 + 105 + 789) = 57$

c)

i) P(likes both Maths and English) = $\frac{n(M \text{ and } E)}{n(S)} = \frac{105}{1\,320} = 0{,}079\ 54... \approx 0{,}08 \approx 8{,}0\%$

ii) P(likes M only) = $\frac{n(M\ only)}{n(S)} = \frac{369}{1\,320} = 0{,}279\ 54... \approx 0{,}28 \approx 28{,}0\%$

iii) P(likes E only) = $\frac{n(E\ only)}{n(S)} = \frac{789}{1\,320} = 0{,}597\ 72... \approx 0{,}60 \approx 59{,}8\%$

<u>Exercise 5.2</u>

1) Independent. The outcome of tossing a coin does not affect the outcome of taking a card from a pack of cards.
2) Independent. The outcome of throwing a dice does not affect the outcome of spinning a spinner.
3) Independent. The outcome of taking a card does not affect the outcome of spinning a spinner.
4) Independent. The outcome of selecting 7 does not affect the outcome of selecting 3
5) Dependent. You have a bigger chance of causing an accident by driving over 120 km/h
6) Dependent. Frequent exercise increases the chance of a low resting heart rate
7) Dependent. Selecting a second ball is affected by the first ball as there is one less ball to choose from
8) Independent. The probability of selecting the second ball from another bin is not affected by the first ball that is selected.

<u>Exercise 5.3</u>

1)

a)

×		Number on dice					
		1	**2**	**3**	**4**	**5**	**6**
Number on counter	**1**	1	2	3	4	5	6
	2	2	4	6	8	10	12

b) n(S) = 12

c)

i) P(2) = $\frac{n(2)}{n(S)} = \frac{2}{12} = \frac{1}{6}$

ii) P(5) = $\frac{n(5)}{n(S)} = \frac{1}{12}$

iii) P(2 or 5) = $\frac{n(2\ or\ 5)}{n(S)} = \frac{3}{12} = \frac{1}{4}$

iv) P(even number) = $\frac{n(even\ number)}{n(S)} = \frac{9}{12} = \frac{3}{4}$

2)

a)

+		Second number								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
First number	**1**	2	3	4	5	6	7	8	9	10
	2	3	4	5	6	7	8	9	10	11
	3	4	5	6	7	8	9	10	11	12
	4	5	6	7	8	9	10	11	12	13
	5	6	7	8	9	10	11	12	13	14
	6	7	8	9	10	11	12	13	14	15
	7	8	9	10	11	12	13	14	15	16
	8	9	10	11	12	13	14	15	16	17
	9	10	11	12	13	14	15	16	17	18

b) Five pairs of numbers give a total of 14

c)

i) $P(14) = \frac{n(14)}{n(S)} = \frac{5}{81} = 6{,}2\%$

ii) $P(15) = \frac{n(15)}{n(S)} = \frac{4}{81} = 4{,}9\%$

d) $P(2) = \frac{n(2)}{n(S)} = \frac{1}{81} = 1{,}2\%$ and $P(10) = \frac{n(10)}{n(S)} = \frac{9}{81} = 11{,}1\%$. There is a greater chance of the supermarket choosing 10 than choosing 2 because there are more ways of getting a total of 10 from the nine whole numbers than of getting a total of 2.

<u>Exercise 5.4</u>

1)

a) The events are independent. Tossing a coin does not affect the number obtained when the dice is thrown.

b) n(S) = 10

c)

i) $P(T) = \frac{n(events\ having\ a\ T\ as\ one\ of\ its\ outcomes)}{n(S)} = \frac{6}{12} = \frac{1}{2}$

ii) $P(6) = \frac{n(events\ having\ a\ 6\ as\ one\ of\ its\ outcomes)}{n(S)} = \frac{2}{12} = \frac{1}{6}$

iii) $P(T) \times P(6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

d) $P(T\ and\ 6) = \frac{n(T\ and\ 6)}{n(S)} = \frac{1}{12}$

So $P(T\ and\ 6) = P(T) \times P(6)$

e)

i) $P(T\ and\ even) = \frac{n(T\ and\ even)}{n(S)} = \frac{3}{12} = \frac{1}{4}$

ii) $P(even) = \frac{n(even)}{n(S)} = \frac{6}{12} = \frac{1}{2}$

iii) $P(T) \times P(even) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

f) $Both\ P(T\ and\ even) and\ P(T) \times P(even) are\ equal\ to\ \frac{1}{4}$.

$So\ P(T) \times P(even) = P(T\ and\ even)$

g)

i) $P(H) = \frac{n(H)}{n(S)} = \frac{6}{12} = \frac{1}{2}$

ii) $P(H\ and\ a\ multiple\ of\ 3) = \frac{n(H\ and\ a\ multiple\ of\ 3)}{n(S)} = \frac{2}{12} = \frac{1}{6}$

iii) $P(multiple\ of\ 3) = \frac{n(multiples\ of\ 3)}{n(S)} = \frac{4}{12} = \frac{1}{3}$

iv) $P(H) \times P(multiple\ of\ 3) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

h) $Both\ P(H) \times P(multiple\ of\ 3) and\ P(H\ and\ multiple\ of\ 3) are\ equal\ to\ \frac{1}{6}$.
$So\ P(H) \times P(multiple\ of\ 3) = P(H\ and\ multiple\ of\ 3)$.

2)

a) The two events are independent as the number the spinner lands on the first time does not affect the number the spinner lands on the second time.

b)

		Second spin				
		2	**2**	**3**	**5**	**5**
First spin	**2**	2; 2	2; 2	2; 3	2; 5	2; 5
	2	2; 2	2; 2	2; 3	2; 5	2; 5
	3	3; 2	3; 2	3; 3	3; 5	3; 5
	5	5; 2	5; 2	5; 3	5; 5	5; 5
	5	5; 2	5; 2	5; 3	5; 5	5; 5

c)

i) $P(first\ number\ is\ 2) = \frac{n(first\ number\ is\ 2)}{n(S)} = \frac{10}{25} = \frac{2}{5}$

ii) $P(second\ number\ is\ 5) = \frac{n(second\ number\ is\ 5)}{n(S)} = \frac{10}{25} = \frac{2}{5}$

iii) $P(first\ number\ is\ 2\ and\ the\ second\ number\ is\ 5) = \frac{4}{25}$

iv) $P(first\ number\ is\ 2) \times P(second\ number\ is\ 5) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$

d) Both $P(first\ number\ is\ 2\ and\ the\ second\ number\ is\ 5)$ and $P(first\ number\ is\ 2) \times P(second\ number\ is\ 5)$ are equal to $\frac{4}{25}$.
So $P(first\ number\ is\ 2\ and\ the\ second\ number\ is\ 5)$
$= P(first\ number\ is\ 2) \times P(second\ number\ is\ 5)$

3) If M and N are independent events, the P(M and N) = P(M) × P(N)

<u>Exercise 5.5</u>

1)

a)

i) $P(R) = \frac{n(red\ cards\ in\ the\ pack)}{n(cards\ in\ the\ pack)} = \frac{26}{52} = \frac{1}{2}$

ii) $P(B) = \frac{n(black\ cards\ in\ the\ pack)}{n(cards\ in\ the\ pack)} = \frac{26}{52} = \frac{1}{2}$

b)

i) $P(P) = \frac{n(picture\ cards\ in\ the\ pack)}{n(cards\ in\ the\ pack)} = \frac{12}{52} = \frac{3}{13}$

ii) $P(not\ P) = 1 - P(P) = 1 - \frac{3}{13} = \frac{10}{13}$

c)

		Second card	
		$\boldsymbol{P(P) = \frac{3}{13}}$	$\boldsymbol{P(not\ P) = \frac{10}{13}}$
First card	$\boldsymbol{P(R) = \frac{1}{2}}$	$P(R;P) = \frac{1}{2} \times \frac{3}{13} = \frac{3}{26}$	$P(R;not\ P) = \frac{1}{2} \times \frac{10}{13} = \frac{10}{26}$
	$\boldsymbol{P(B) = \frac{1}{2}}$	$P(B;P) = \frac{1}{2} \times \frac{3}{13} = \frac{3}{26}$	$P(B;not\ P) = \frac{1}{2} \times \frac{10}{13} = \frac{10}{26}$

d) $P(red\ then\ picture) = P(R;P) = \frac{3}{26}$

e) $P(black\ then\ not\ a\ picture) = P(B;not\ P) = \frac{10}{26} = \frac{5}{13}$

2)

a)

		UNITED		
		$P(w)=\frac{1}{2}$	$P(d)=\frac{1}{6}$	$P(l)=\frac{1}{3}$
CITY	$P(W)=\frac{7}{10}$	P(W; w) $=\frac{7}{10}\times\frac{1}{2}=\frac{7}{20}$	P(W; d) $=\frac{7}{10}\times\frac{1}{6}=\frac{7}{60}$	P(W; l) $=\frac{7}{10}\times\frac{1}{3}=\frac{7}{30}$
	$P(D)=\frac{1}{5}$	P(D; w) $=\frac{1}{5}\times\frac{1}{2}=\frac{1}{10}$	P(D; d) $=\frac{1}{5}\times\frac{1}{6}=\frac{1}{30}$	P(D; l) $=\frac{1}{5}\times\frac{1}{3}=\frac{1}{15}$
	$P(L)=\frac{1}{10}$	P(L; w) $=\frac{1}{10}\times\frac{1}{2}=\frac{1}{20}$	P(L; d) $=\frac{1}{10}\times\frac{1}{6}=\frac{1}{60}$	P(L; l) $=\frac{1}{10}\times\frac{1}{3}=\frac{1}{30}$

b) $P(City\ wins\ and\ United\ wins) = P(W; w) = \frac{7}{20}$

c) $P(City\ loses\ and\ United\ draws) = P(L; d) = \frac{1}{60}$

d)

i) $P(both\ teams\ win\ or\ both\ teams\ lose) = P(W; w) + P(L; l) = \frac{7}{20} + \frac{1}{30} = \frac{23}{60}$

ii)
$$\begin{aligned} P(City\ wins) &= P(W; w\ or\ W; d\ or\ W; l) \\ &= P(W; w) + P(W; d) + P(W; l) \\ &= \frac{7}{20} + \frac{7}{60} + \frac{7}{30} \\ &= \frac{7}{10} \end{aligned}$$

iii)
$$\begin{aligned} P(only\ one\ team\ loses) &= P(W; l\ or\ D; l\ or\ L; w\ or\ L; d) \\ &= P(W; l) + P(D; l) + P(L; w) + P(L; d) \\ &= \frac{7}{30} + \frac{1}{15} + \frac{1}{20} + \frac{1}{60} \\ &= \frac{11}{30} \end{aligned}$$

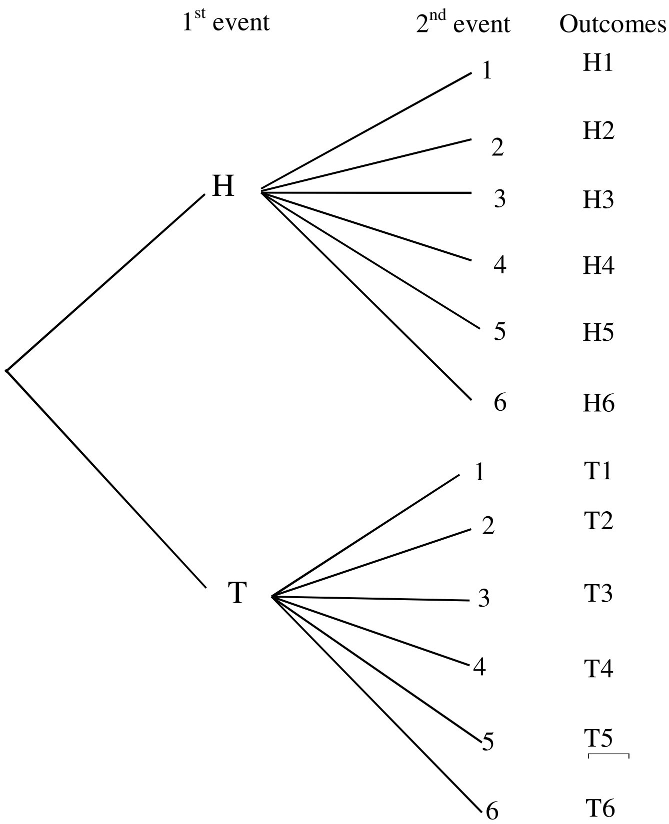
iv)
$$\begin{aligned} P(only\ one\ team\ draws) &= P(W; d\ or\ D; w\ or\ D; l\ or\ L; d) \\ &= P(W; d) + P(D; w) + P(D; l) + P(L; d) \\ &= \frac{7}{60} + \frac{1}{10} + \frac{1}{15} + \frac{1}{60} \\ &= \frac{3}{10} \end{aligned}$$

<u>Exercise 5.6</u>

1)

a) The two events are independent. Flipping a coin and rolling a dice do not affect each other.

b)



c)

i) $P(H;4) = \frac{n(H;4)}{n(S)} = \frac{1}{12}$

ii) $P(T;6) = \frac{n(T;6)}{n(S)} = \frac{1}{12}$

iii) Outcomes that are H and odd are H1, H3, and H5.
$P(H;odd) = \frac{n(H;odd)}{n(S)} = \frac{3}{12} = \frac{1}{4}$

2)

a) The two events are independent. The colour of the first disc taken does not affect the colour of the second disc taken because the first disc is returned before the second disc is taken.

b)

i) $P(RR) = \frac{7}{11} \times \frac{7}{11} = \frac{49}{121}$

ii) $P(RB) = \frac{7}{11} \times \frac{4}{11} = \frac{28}{121}$

iii) $$\begin{aligned} P(only\ one\ disc\ is\ red) &= P(RB\ or\ BR) \\ &= P(RB) + P(BR) \\ &= \left(\frac{7}{11} \times \frac{4}{11}\right) + \left(\frac{4}{11} \times \frac{7}{11}\right) \\ &= \frac{28}{121} + \frac{28}{121} \\ &= \frac{56}{121} \end{aligned}$$

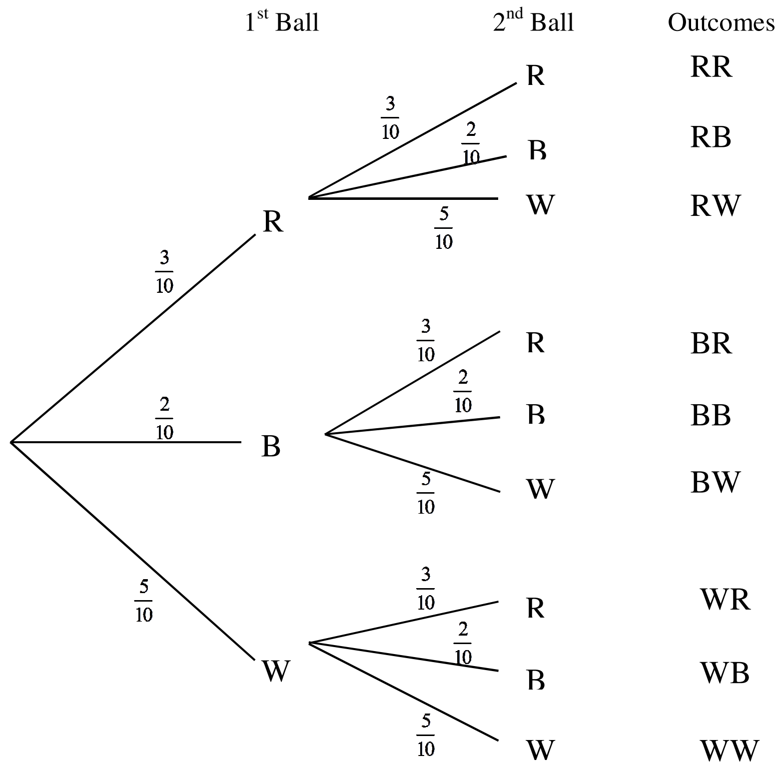
iv) $$\begin{aligned} P(at\ least\ one\ is\ red) &= P(RR\ or\ RB\ or\ BR) \\ &= P(RR) + P(RB) + P(BR) \\ &= \left(\frac{7}{11} \times \frac{7}{11}\right) + \left(\frac{7}{11} \times \frac{4}{11}\right) + \left(\frac{4}{11} \times \frac{7}{11}\right) \\ &= \frac{49}{121} + \frac{28}{121} + \frac{28}{121} \\ &= \frac{105}{121} \end{aligned}$$

OR $P(at\ least\ one\ red) = 1 - P(BB) = 1 - \left(\frac{4}{11} \times \frac{4}{11}\right) = 1 - \frac{16}{121} = \frac{105}{121}$

3)

a) The two events are independent. Because each time the ball selected is replaced, each draw has the same number of red, blue and white balls.

b)



c)

i) $P(BB) = \frac{2}{10} \times \frac{2}{10} = \frac{4}{100} = \frac{1}{25}$

ii) $P(BW) = \frac{2}{10} \times \frac{5}{10} = \frac{10}{100} = \frac{1}{10}$

iii) $P(RB) = \frac{3}{10} \times \frac{2}{10} = \frac{6}{100} = \frac{3}{50}$

iv)
$$\begin{aligned} P(two\ balls\ the\ same\ colour) &= P(RR\ or\ BB\ or\ WW) \\ &= P(RR) + P(BB) + P(WW) \\ &= \left(\frac{3}{10} \times \frac{3}{10}\right) + \left(\frac{2}{10} \times \frac{2}{10}\right) + \left(\frac{5}{10} \times \frac{5}{10}\right) \\ &= \frac{9}{100} + \frac{4}{100} + \frac{25}{100} \\ &= \frac{38}{100} \\ &= \frac{19}{50} \end{aligned}$$

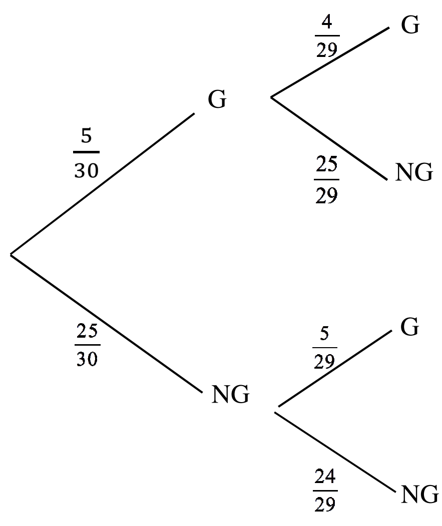
v)
$$\begin{aligned} P(at\ least\ one\ B) &= P(RB\ or\ BR\ or\ BB\ or\ BW\ or\ WB) \\ &= P(RB) + P(BR) + P(BB) + P(BW) + P(WB) \\ &= \left(\frac{3}{10} \times \frac{2}{10}\right) + \left(\frac{2}{10} \times \frac{3}{10}\right) + \left(\frac{2}{10} \times \frac{2}{10}\right) + \left(\frac{2}{10} \times \frac{5}{10}\right) + \left(\frac{5}{10} \times \frac{2}{10}\right) \\ &= \frac{6}{100} + \frac{6}{100} + \frac{4}{100} + \frac{10}{100} + \frac{10}{100} \\ &= \frac{36}{100} \\ &= \frac{9}{25} \end{aligned}$$

<u>Exercise 5.7</u>

1)

a) Number of CDs that are NOT gospel music = 30 – 5 = 25

b) Draw 1 Draw 2 Outcomes Probabilities



Outcomes	Probabilities
G;G	$\frac{5}{30}\times\frac{4}{29}=\frac{20}{870}$
G;NG	$\frac{5}{30}\times\frac{25}{29}=\frac{125}{870}$
NG;G	$\frac{25}{30}\times\frac{5}{29}=\frac{125}{870}$
NG;NG	$\frac{25}{30}\times\frac{24}{29}=\frac{600}{870}$

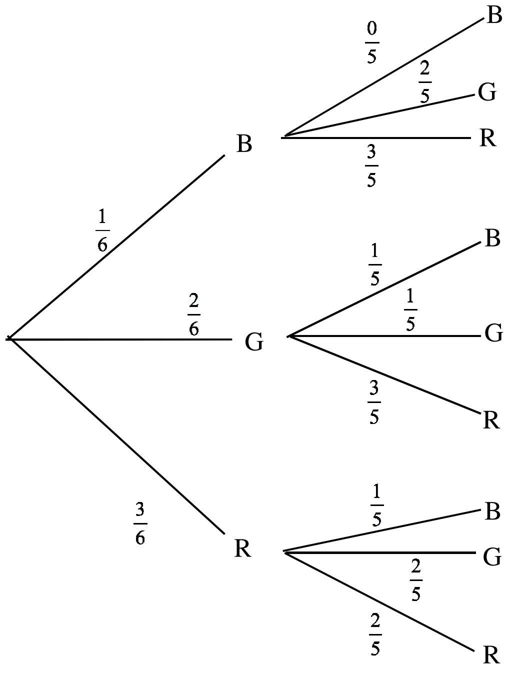
c)

i) $P(G;G)=\frac{20}{870}=\frac{2}{87}$

ii) $P(G;NG)=\frac{125}{870}=\frac{25}{174}$

2)

a) 1st draw 2nd draw Outcomes



BB
BG
BR
GB
GG
GR
RB
RG
RR

b)

i) $P(BB)=\frac{1}{6}\times\frac{0}{5}=0$

ii) $P(RR)=\frac{3}{6}\times\frac{2}{5}=\frac{6}{30}=\frac{1}{5}$

iii) $P(at\ least\ one\ green)=P(BG\ or\ GB\ or\ GG\ or\ GR\ or\ RG)$

$$=P(BG)+P(GB)+P(GG)+P(GR)+P(RG)$$
$$=\left(\frac{1}{6}\times\frac{2}{5}\right)+\left(\frac{2}{6}\times\frac{1}{5}\right)+\left(\frac{2}{6}\times\frac{1}{5}\right)+\left(\frac{2}{6}\times\frac{3}{5}\right)+\left(\frac{3}{6}\times\frac{2}{5}\right)$$
$$=\frac{1}{15}+\frac{1}{15}+\frac{1}{15}+\frac{3}{15}+\frac{3}{15}$$
$$=\frac{9}{15}=\frac{3}{5}$$

iv) $P(at\ most\ 1\ red) = 1 - P(RR) = 1 - \left(\frac{3}{6} \times \frac{2}{5}\right) = 1 - \frac{1}{5} = \frac{4}{5}$

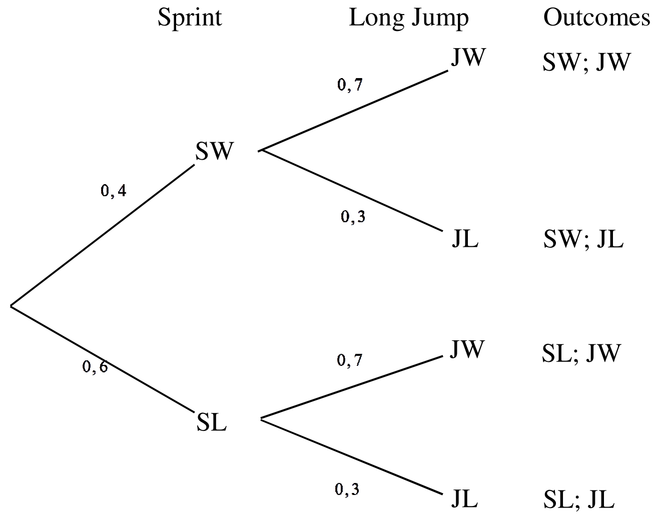
3)

a)

i) $P(SL) = 1 - P(SW) = 1 - 0{,}4 = 0{,}6$

ii) $P(JL) = 1 - P(JW) = 1 - 0{,}7 = 0{,}3$

b)



c)

i) $P(SW; JW) = 0{,}4 \times 0{,}7 = 0{,}28$

ii) $P(SW; JL) = 0{,}4 \times 0{,}3 = 0{,}12$

iii) $P(SL; JL) = 0{,}6 \times 0{,}3 = 0{,}18$

iv) $P(SW; JL\ or\ SL; JW) = P(SW; JL) + P(SL; JW)$
$= (0{,}4 \times 0{,}3) + (0{,}6 \times 0{,}7)$
$= 0{,}12 + 0{,}42$
$= 0{,}54$

Exercise 5.8

1)

a) The events are not independent as some learners use all three modes of transport.

b) n(S) = 26 + 15 + 17 + 20 + 19 + 18 + 22 + 25 = 162

c)

i) P(W) = $\frac{n(W)}{n(S)} = \frac{15+17+19+18}{162} = \frac{69}{162} = \frac{23}{54}$

ii) P(C) = $\frac{n\ (C)}{n(S)} = \frac{26+15+20+19}{162} = \frac{80}{162} = \frac{40}{81}$

iii) P(B) = $\frac{n\ (B)}{n(S)} = \frac{20+19+18+22}{162} = \frac{79}{162}$

iv) P(walk only) = $\frac{n\ (walk\ only)}{n(S)} = \frac{17}{162}$

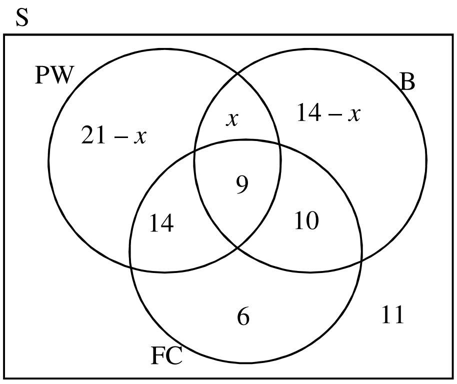
v) P(does not walk) = 1 – P(W) = 1 – $\frac{23}{54} = \frac{31}{54}$

vi) P(W or B) = P(W) + P(B) – P(W and B)
$= \frac{69}{162} + \frac{79}{162} - \frac{(19+18)}{162} = \frac{111}{162} = \frac{37}{54}$

2)

a) Number who don't like any of the meals = number in the survey – number who likes at least one
= 80 – 69
= 11 people

b)



c) $(21 - x) + x + (14 - x) + 14 + 9 + 10 + 6 + 11 = 80$
$85 - x = 80$
$x = 5$

d)

i) $P(Burgers) = \frac{n(Burgers)}{n(S)} = \frac{5+9+9+10}{80} = \frac{33}{80} = 41{,}25\%$

ii) $P(Fried\ Chicken\ only) = \frac{n(chicken\ only)}{n(S)} = \frac{6}{80} = 7{,}5\ \%$

iii) $P(Burgers\ and\ Fried\ Chicken) = \frac{n(Burgers\ and\ Fried\ Chicken)}{n(S)} = \frac{9+10}{80} = \frac{19}{80} = 23{,}75\%$

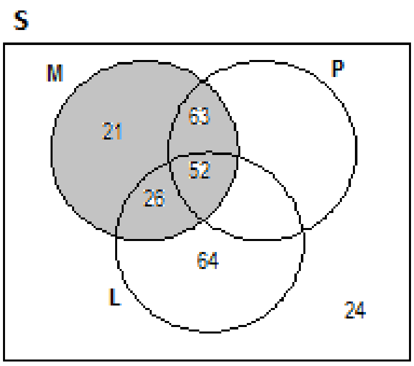
iv) $P(Pap\ and\ Wors\ or\ Fried\ Chicken)$
$= P(Pap\ and\ Wors) + (P(Fried\ Chicken) - P(Pap\ and\ Wors\ and\ Fried\ Chicken)$
$= \frac{44}{80} + \frac{39}{80} - \frac{(14+9)}{80}$
$= \frac{60}{80}$
$= 75\%$

CHAPTER 6 – GRADE 12 PROBABILITY

Exercise 6.1

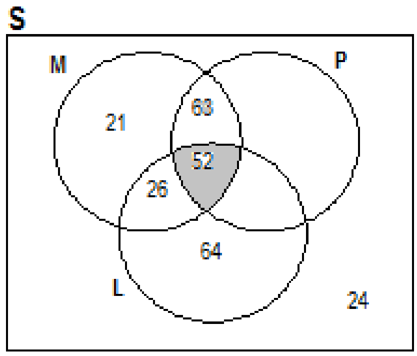
1. a) This section of the Venn diagram shows you the learners who take Mathematics.

n(M) = 21 + 63 + 52 + 26 = 162 learners.



b) This section of the Venn diagram shows you learners who take Mathematics, Physical Sciences and Life Sciences.

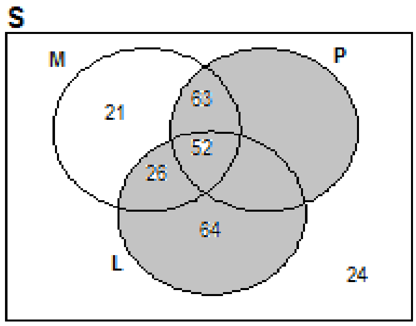
n(M and P and L) = 52 learners



c) i) This section of the Venn diagram shows you the learners who take Physical Sciences or Life Sciences.

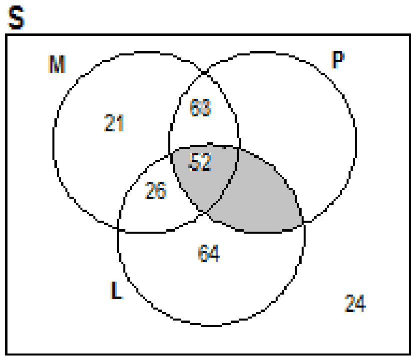
n(P or L) = 63 + 52 + 26 + 64 = 205

$P(P \text{ or } L) = \frac{n(P \text{ or } L)}{n(S)} = \frac{205}{250} = \frac{41}{50}$

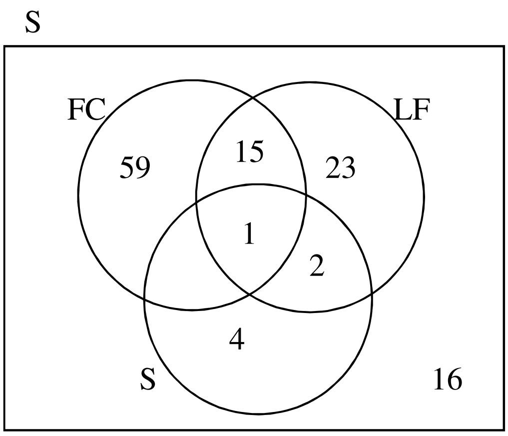


ii) This section of the Venn diagram shows you the learners who take Physical Sciences and Life Sciences.

$P(P \text{ and } L) = \frac{n(P \text{ and } L)}{n(S)} = \frac{52}{250} = \frac{26}{125}$

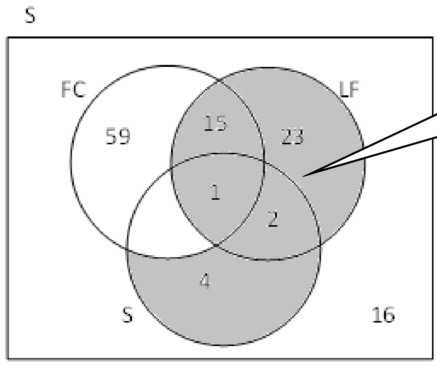


2. a)



b) Number of shoppers who did not buy milk = 120 – (59 + 15 + 23 + 1 + 2 + 4) = 16

c)



P(LF or S)

$= \frac{\text{Number of shoppers who bought LF or S}}{\text{Total number of shoppers}}$

$= \frac{15+23+1+2+4}{120}$

$= \frac{45}{120}$

$= \frac{3}{8}$

Another method could be:

$\text{P(LF or S)} = \text{P(LF)} + \text{P(S)} - \text{P(LF and S)}$

$= \frac{41}{120} + \frac{7}{120} - \frac{3}{120} = \frac{45}{120} = \frac{3}{8}$

Exercise 6.2

1)

a) Independent, because the first card was replaced.

b) Dependent, because the first counter was not returned so the colour of the second counter selected depends on the colour of the first one selected.

c) Dependent, because you are more likely to be struck by lightning by running in a thunderstorm than by staying indoors in a thunderstorm or running when there is no thunderstorm.

d) Dependent, because the chances of being bitten by a shark are higher if you swim in the ocean than if you don't swim in the ocean (although, it is still HIGHLY unlikely!)

e) Independent, because wearing a red T-shirt is not influenced by watching television.

2)

a) Independent because the number that the spinner lands on the second time it is spun is not dependent on the number that the spinner lands on the first time.

b) $\text{P(4 and 4)} = \text{P(4)} \times \text{P(4)} = \frac{1}{6} \times \frac{1}{6} = \frac{1}{3} \approx 33\%$

3) P(A and B) = P(A) × P(B) because we are told that they are independent.

$x = (x + 0{,}25) \times (x + 0{,}15)$

$x = x^2 - 0{,}4x + 0{,}0375$

$0 = x^2 - 0{,}4x + 0{,}0375$

$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$= \frac{-(-0{,}4) \pm \sqrt{(-0{,}4)^2 - 4(1)(0{,}0375)}}{2(1)}$

$x = 0{,}25$ or $x = 0{,}15$ (OR $x = \frac{3}{20}$ or $x = \frac{1}{4}$)

Exercise 6.3

1)

a) They are mutually exclusive, because one counter cannot be both red and blue.

b) They are not mutually exclusive, because a red card could also be a King i.e. the King of Hearts and the King of Diamonds.

2) Selecting a 3 and selecting a 7 are mutually exclusive because a card cannot be a 3 and a 7 at the same time.

P(3 or 7) = P(3) + P(7) = $\frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$

3) Selecting a 3 and selecting a red card are NOT mutually exclusive because a card can be a 3 and a red card at the same time.

P(3 or R) = P(3) + P(R) – P(3 and R)= $\frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13}$

4)

a) This dice is most likely to land on the 5

b) Yes they are mutually exclusive. Dice cannot land on a 5 and 6 in the same throw.

c)

i) P(5 or 6) = P(5) + P(6) = 0,33 + 0,258 = 0,588 ≈ 0,59

ii) P(even) = P(2) + P(4) + P(6) = 0,04 + 0,167 + 0,258 = 0,465 ≈ 0,47

5) P(R or C) = P(R) + P(C) – P(R and C)

$= \frac{n(R)}{n(S)} + \frac{n(C)}{n(S)} - \frac{n(R\ and\ C)}{n(S)}$

$= \frac{180}{250} + \frac{99}{250} - \frac{87}{250}$

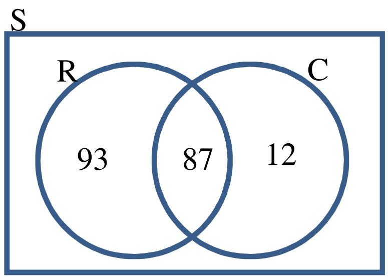
$= \frac{192}{250}$

$= \frac{96}{125}$

$= 0{,}768$

$\approx 77\%$

A Venn diagram may help you see what is happening:



Exercise 6.4

1)

a) P(not A) = 1 – P(A) = 1 – 0,35 = 0,65

b) A and B are mutually exclusive, so P(A or B) = P(A) + P(B) = 0,35 + 0,22 = 0,57

c) A and B are NOT mutually exclusive, so P(A or B) = P(A) + P(B) – P(A and B)
A and B are independent, so P(A and B) = P(A) × P(B) = 0,35 × 0,22 = 0,077
So P(A or B) = P(A) + P(B) – P(A) × P(B) = 0,35 + 0,22 – 0,077 = 0,493

2)

a) Number of learners in the school = n(S) = 191 +64 + 214 + 310 = 500

b) Number of girls = n(girls) = 500 – number of boys = 500 – (214 +31) = 255

c) Number of learners who do not play soccer = 500 – number of learners who play soccer
= 500 – (64 + 214)
= 222

d)

i) n(boy and plays soccer) = 214

$P(\text{boy and plays soccer}) = \frac{n(boy\ and\ plays\ soccer)}{n(S)} = \frac{214}{500} = \frac{107}{250} = 0{,}428 \approx 0{,}43$

ii) P(girl or plays soccer) = P(girl + P(soccer) – P(girl and plays soccer)

$= \frac{n(girl)}{n(S)} + \frac{n(soccer)}{n(S)} - \frac{n(girl\ and\ plays\ soccer)}{n(S)}$

$= \frac{255}{500} + \frac{278}{500} - \frac{64}{500}$

$= \frac{469}{500}$

$= 0{,}938$

$\approx 0{,}94$

OR P(girl and plays soccer) = 1 – P(boy that does not play soccer) $= 1 - \frac{31}{500} = \frac{469}{500} \approx 0{,}94$

Exercise 6.5

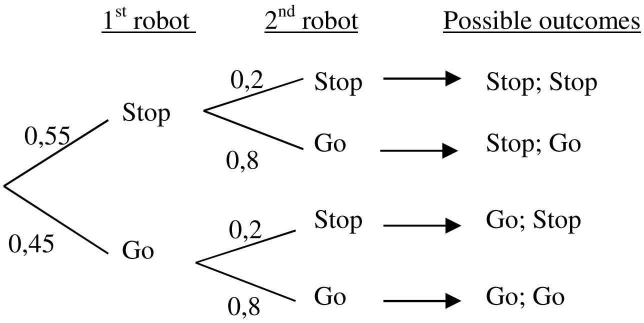
1)

a)

i) He does not stop at the first robot for 100% - 55% = 45% of the time

ii) He does not stop at the second robot for 100% - 20% = 80% of the time.

b) The percentages have been converted to decimals to make them easier to multiply.



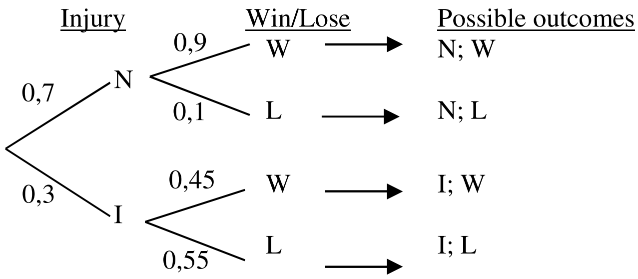
c) These events are mutually exclusive because Vusi cannot be at both intersections at the same time.

i) P(stop at both robots) = P(S and S) = P(S) × P(S) …..you are told the events are independent
$= 0{,}55 \times 0{,}2 = 0{,}11 = 11\%$

ii) P(stop at one of the robots) = P(SG or GS) = P(SG) + P(GS)
$= P(S) \times P(G) + P(G) \times P(S)$
$= 0{,}55 \times 0{,}8 + 0{,}45 \times 0{,}2$
$= 0{,}53$
$= 53\%$

2)

a) N stands for 'no injury' and I stands for 'injury'. W stands for 'win' and L stands for 'lose'>

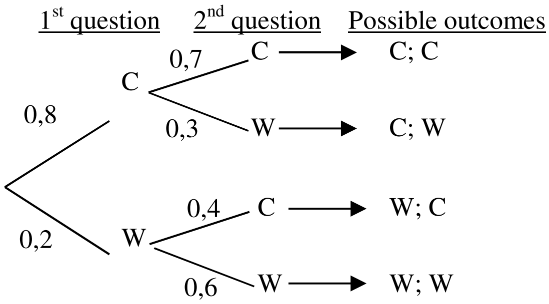


b) P(Win) = P(NW or IW)
$= (P(N) \times P(W)) + (P(I) \times P(W))$
$= (0{,}7 \times 0{,}9) + (0{,}3 \times 0{,}45)$
$= 0{,}765$
$\approx 77\%$

3) P(H or W) = P(H) + P(W) – P(H and W)
$= 0{,}7 + 0{,}43 - 0{,}27$
$= 0{,}86$
$= 86\%$

4)

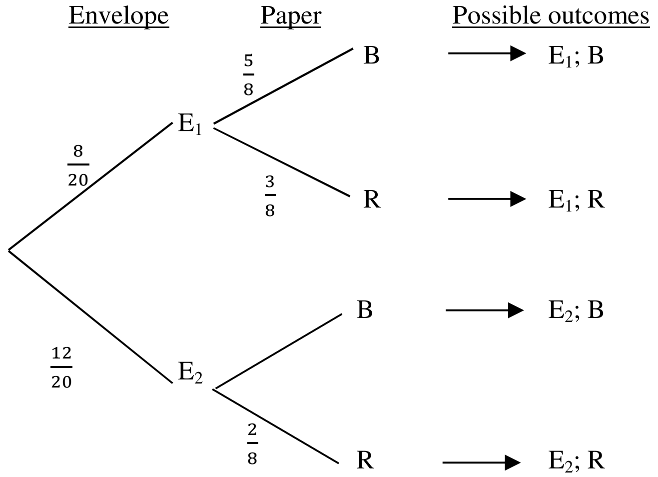
a) C stands for 'correct' and W stands for 'wrong'.



b) P(CC or WC) = P(C and C) + P(W and C)
$= P(C) \times P(C) + P(W) \times P(C)$
$= 0{,}8 \times 0{,}7 + 0{,}2 \times 0{,}4$
$= 0{,}64$
$= 64\%$

5)

a) E_1 stands for the first envelope and E_2 stands for the second envelope. B stands for blue paper and R stands for red paper.



b) $P(E_1R \text{ or } E_2R) = P(E_1R) + P(E_2R)$

$= P(E_1) \times P(R) + P(E_2) \times P(R)$

Exercise 6.6

1)

a) The events 'selecting a learner with brown eyes' and 'selecting a learner with blue eyes' are mutually exclusive because a learner cannot have both eyes brown and blue at the same time.

b) The events 'Selecting a learner with brown eyes' and 'selecting a learner with blue eyes' are not complementary because there are other possibilities besides brown eyes and blue eyes. The learner's eyes could also be green.

c) Selecting a male learner and a female learner are complementary events because it is certain that the learner will be either male or female.

d)

i) $\text{P(blue eyes)} = \frac{4\,684}{145\,619} = 0{,}032166 \approx 3{,}22\%$

ii) $\text{P(female)} = \frac{76\,143}{145\,619} = 0{,}52289 \approx 52{,}29\%$

e) $\text{P(blue eyes and female)} = \frac{2\,449}{145\,619} = 0{,}016\,817\,86031\ldots$

$\text{P(blue eyes)} \times \text{P(female)} = \frac{4\,684}{145\,619} \times \frac{76\,143}{145\,619} = 0{,}016\,819\,410\,94\ldots$

Notice that these numbers are SLIGHTLY different.

It is important when answering questions like this that you don't round off.

So, $\text{P(blue eyes and female)} \neq \text{P(blue eyes)} \times \text{P(female)}$.

Therefore the event "having blue eyes" and "being female" are NOT independent,

2)

a) A = 801 + 1 922 = 2 723
B = 688 – 394 = 294
C = 326 – 86 = 240
D = 10 926 – 5 144 = 5 782

b)

i) $\text{P(prefers netball)} = \frac{2\,288}{10\,926} = 0{,}20940\ldots \approx 20{,}94\%$

ii) $\text{P(girl)} = \frac{5\,782}{10\,926} = 0{,}52919\ldots. \approx 52{,}92\%$

iii) $P(\text{prefers netball and a girl}) = \frac{2\,259}{10\,926} = 0{,}20675\ldots \approx 20{,}68\%$

c) It seems likely that the events of being a girl and preferring netball are dependent, because netball is played by more girls than boys.

d) P(prefer netball and a girl) = 0,20675... (From question d)
$P(\text{prefers netball}) \times P(\text{a girl}) = \frac{2\,288}{10\,926} \times \frac{5\,782}{10\,926} = 0{,}11081\ldots$
$P(\text{prefers netball and a girl}) \neq P(\text{prefers netball}) \times P(\text{a girl})$
Therefore the event "preferring netball" and "being a girl" are NOT independent.

<u>Exercise 6.7</u>

1)
a) $4! = 4 \times 3 \times 2 \times 1$
b) $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$
c) $10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$

2)
a) $6! = 720$
b) $8! = 40\,320$
c) $12! = 479\,001\,600$
d) $25! = 1{,}551\,121\,004 \times 10^{25}$

3) Number of combinations of meals = $3 \times 4 \times 2 = 24$

4)
a) Number of possible combinations for the new number plates
$= (21 \times 21) \times (10 \times 10) \times (21 \times 21) = 19\,448\,100$.
b) Number of possible combinations of old number plates:
$= (21 \times 21 \times 21) \times (10 \times 10 \times 10) = 9\,261\,000$.
There are more combinations of the new number plate.

5)
a) There are 6 digits to choose from.
There are only 5 options for the first digit (it cannot be 0).
After a digit has been used, there are five remaining for the second digit, four for the third digit and three for the fourth digit.
Number of possible combinations = $5 \times 5 \times 4 \times 3 = 300$.
b) In order to be divisible by 5, the last digit must be a 5 or a 0.
So the number will be __ __ __ 5 or __ __ __ 0.
So, there are 4 digits left to start the first number (because the first number cannot be 5 or 0)
There are 5 digits left to start the second number (because 0 is already at the end).
Number of possible outcomes that are divisible by 5 = $(4 \times 4 \times 3) + (5 \times 4 \times 3) = 108$.

6) There are 20 boys that have 5 spaces to fill. The same boy cannot be selected twice.
Number of different groups of boys = $20 \times 19 \times 18 \times 17 \times 16 = 1\,860\,480$.

7) There are three spaces to fill.
26 letters can fill the first space and 10 digits can fill the second and third space.
We are not told otherwise, so the digits can be repeated.
Number of codes = $26 \times 10 \times 10 = 2\,600$.

<u>Exercise 6.8</u>

1)
a) There are twelve spaces to fill with twelve balls. Each ball cannot be in more than one space.
Number of arrangements = $12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$
$= 12! = 479\,001\,600$
b) 8 and 11 are a single entity that can be arranged in two ways.
There are 11 spaces left to fill.
Number of different arrangements = $(11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1) \times (2 \times 1)$
$= 11! \times 2! = 79\,833\,600$

2) There are six spaces to fill with six red mugs. A mug cannot fill more than one space.
There are five spaces to fill with five blue mugs. A mug cannot fill more than one space.
The red mugs can be placed on the shelf first, or the blue mugs. In other words, there are two ways of organising the different coloured mugs.
Number of different arrangements = $(6 \times 5 \times 4 \times 3 \times 2 \times 1) \times (5 \times 4 \times 3 \times 2 \times 1) \times (2 \times 1)$
$= 6! \times 5! \times 2! = 172\ 800$
3) Jarryd and Rikus are one entity that can be arranged in two ways (Jarryd and Rikus or Rikus and Jarryd). There are four spaces left to fill.
Number of different arrangements = $(4 \times 3 \times 2 \times 1) \times (2 \times 1) = 4! \times 2! = 48$.

Exercise 6.9

1) There are 10 letters in the word CUMULATIVE. The letter "U" is repeated twice.
Number of arrangements = $\frac{10!}{2!} = 1\ 814\ 400$
2) There are 11 letters in the word PROBABILITY. The letter "B" is repeated twice and the letter "I" is repeated twice.
Number of arrangements = $\frac{11!}{2! \times 2!} = 9\ 979\ 200$
3) There are 17 letters in the words STANDARD DEVIATION. The first and last letter is an "N", so there are 15 letters left to arrange.
The letter "T" is repeated twice. The letter "I" is repeated twice. The letter "A" is repeated three times and the letter "D" is repeated three times.
Number of arrangements = $\frac{15!}{2! \times 2! \times 3! \times 3!} = 9\ 081\ 072\ 000$

Exercise 6.10

1) There are 10 letters in the word PHILLIPINE.
The letter 'P' is repeated twice, the letter 'I' is repeated three times, the letter 'L' is repeated twice,
Number of possible outcomes = $\frac{10!}{2! \times 3! \times 2!} = 151\ 200$
The arrangements must start and end with an L.
The two Ls are regarded as a single entity, so there are 8 places to fill, with 'P' repeated twice and 'I' repeated three times.
Number of favourable outcomes = $\frac{8!}{2! \times 3!} = 3\ 360$
Probability (start and end in L)= $\frac{3\ 360}{151\ 200} = \frac{1}{45} = 0{,}0222\ldots \approx 2\%$
2) Possible outcomes = $7 \times 6 \times 5 \times 4 = 840$
   a) In order for the number to be even, the last digit must be a 2 or a 4.
   So the number will be __ __ __ ② or __ __ __ ④
   There are 6 digits left to write in the three places.
   Number of favourable outcomes = number of even numbers = $(6 \times 5 \times 4) \times (6 \times 5 \times 4) = 240$.
   Probability (even number) = $\frac{240}{840} = \frac{2}{7} = 0{,}28571\ldots \approx 29\%$
   b) In order for the number to be greater than 4 000, the first digit must be a 4, 5, 7 or 9. So the number will be ④ __ __ __ or ⑤ __ __ __ or ⑦ ___ ___ ___ or ⑨ ___ ___ ___
   Number of favourable outcomes = number of numbers greater than 4 000
   $= (6 \times 5 \times 4) + (6 \times 5 \times 4) + (6 \times 5 \times 4) + (6 \times 5 \times 4) = 480$.
   Probability (number greater than 4 000) = $\frac{480}{840} = \frac{4}{7} = 0{,}57142\ldots \approx 57\%$
   c) There is only one possible way that digits can be arranged in ascending order.
   Number of favourable outcomes = 1
   Probability (ascending order) = $\frac{1}{840} = 0{,}0012$ (correct to four decimal places) $\approx 0{,}12\%$
   In other words, it is almost impossible!
3) There are 6 letters in the word BAFANA with the letter 'A' repeated three times
Number of possible outcomes = $\frac{6!}{3!} = 120$
The arrangement starts and ends with an A which means there are 4 places to fill.
Number of favourable outcomes = $4! = 24$
Probability (start and end with an A) = $\frac{24}{120} = \frac{1}{5} = 0{,}02 = 2\%$