UNIVERSITY OF THE FREE STATE
SOUTH CAMPUS

Advanced Certificate
In Teaching

Senior Phase
Mathematics Subject Content VI
Data Handling

Compiled by: Riana Nel

NQF level: 5
Total credits: 4

EMBC2521 | SOUTH CAMPUS

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS
UV

Adapted May 2019

Published by CDD
South Campus
University of the Free State

Dalene Meintjes, Erica Odendaal, Jodie Teise
Louise Mitchell, Noluvo Mokoena, Ronél Theron, Thabo Ncapodi

May 2019

TABLE OF CONTENTS

1. Introduction

A word of welcome!

Welcome to the next module in our series on the content of Senior Phase Mathematics. In this module we are going to discover valuable strategies and information that you can use to enhance learner performance at your school.

In this module we are going to deepen our knowledge of data handling. We will discuss the collection, organisation and representation of information. We will have a look at the analysis and interpretation of information. In the last unit we will discuss the concept of probability. We will also do calculations and predictions based on calculations of probability.

In the Senior Phase we will ensure that learners can use data handling and can do all the relevant calculations. Learners must understand the concepts and be able to apply their knowledge.

2. Where are we headed?

This is one of six content modules. In this content module we are going to discuss data handling and will include a discussion on probability. We will have a look at the collection of data and the various types of data we work with in the Senior Phase. We will learn how to organise and summarise data in order to represent data.

We will discuss the different types of graphs and when to use a specific graph. We will discuss the analysis and interpretation of data which includes calculations done on data. In the last unit we will have a detailed discussion on the calculations and interpretations of probability.

As a teacher, you must be the master of your subject in terms of the terminology, content and the methodology. I hope that this module will bring you closer to be a master of mathematics in your school!

The outcomes of the module are that teachers will demonstrate deepened knowledge and insight into Data handling and Probability in Senior Phase Mathematics. The focus of this module is on the deepening of the subject content knowledge of teachers and to serve as reference for applications in the core modules.

It includes content on Data Handling and Probability such as:

- collecting, organising and summarising data;
- representing data;
- interpreting, analysing and reporting data; and
- doing calculations on probability.

3. Module-content map

In this module the student will show a deepened knowledge of the concepts and teaching of the following:

- To collect and organise data.
- To summarise data.
- To represent data.
- To interpret, analyse and report data.
- To do calculations on probability.

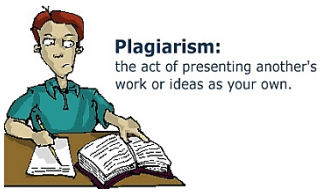
4. Work schedule and learning

Contact session	Duration	Action	Learning hours
1	2 hours	Orientation Unit 1: Collect and organise data Planning for assessment 1	8 hours
2	2 hours	Unit 2: Summarise data Planning for assessment 2	7 hours
3	2 hours	Assessment 1 Unit 3: Represent data	10 hours
4	2 hours	Assessment 2 Unit 4: Interpret, analyse and report data	8 hours
5	2 hours	Unit 5: Probability – what is the chance?	7 hours
		Total	**40**

5. Assessment guidelines and schedule

Unit	Type	Due date	Notes
1 & 2	Assessment 1	Submit: Planning at the end of contact session 1 Submit: Final at the beginning of contact session 3	See page: 68
1 - 5	Assessment 2	Submit: Planning at the end of contact session 2 Submit: Final at the beginning of contact session 4	See page: 72

6. Plagiarism



https://goo.gl/bPCZ5g

What is plagiarism? The word 'plagiarism' is derived from the Latin term *plagiaries*, which literally means **'kidnapper'**, and refers to the **theft** of someone else's work. If you submit your work as if it is your own, but all or part of it is copied from somebody else's work another student, the internet, or other published works, it is called plagiarism.

According to Lemon (n.d. online), "to plagiarize means

- to steal and pass off (the ideas or words of another) as one's own;
- to use (another's production) without crediting the source;
- to commit literary theft;
- to present as new and original an idea or product derived from an existing source."

> Plagiarism is stealing someone else's work.

The UFS makes a distinction between plagiarism and academic writing misconduct. Plagiarism refers to

- acts of plagiarism committed unknowingly/unintentionally; and/or
- improper or no acknowledgement of a limited section of the work.

Academic writing misconduct refers to repeated incidents of collusion, deliberate dishonesty, and more serious forms of plagiarism.

Against this background, students have to refrain from deliberate dishonesty in the form of

- cribbing in tests and examinations;
- collusion and fabrication or falsification of data;
- purchasing assignments, dissertations, and/or theses on the internet and presenting such documents as your own work; and/or
- submitting the same work for more than one course or in consecutive years.

The programme booklet outlines the consequences that students, who are guilty of plagiarism or academic writing misconduct, will have to face.

> When you cheat, you are only cheating yourself.
> Lemon, B.

Why must you do and submit your own work, besides the fact that you can be formally disciplined? The ACT is designed to develop agents of change. The ACT aims to empower teachers who are honest, who take responsibility for their own work and their circumstances, who reflect on situations, and then want to improve. By doing your own work, you learn something, you reflect, you grow, and you become an agent of change for your learners, your community, education and your country!

https://goo.gl/WUTYv2

> Be the change you wish to see in the world.
> Gandhi

Unit 1: Collect and organise data

Outcomes

At the end of this unit, you will be able to guide your learners to

- ✓ develop a vocabulary to define data sources;
- ✓ apply and develop the skills to set different types of questionnaires; and
- ✓ define and use different techniques to sort data.

1.1 Introduction

We are exposed to information every day. For example, a person might consider starting a business and needs information regarding various aspects of the business in order to assess whether it will be a worthwhile venture or not. Other people might want to investigate how South Africans feel about rhino *poaching* or how many people are unemployed in South Africa. We need to find information, which is referred to as the collection of *data*, in order to answer these questions. When we need information, we ask questions. Data handling starts with a question – you want to find information about something.

1.2 What is data?

The result of statistical investigation or collected information is referred to as *data*. The test marks of learners, for example, can be considered as data. We can identify different types of data.

Activity 1: Types of data

Work in pairs

 **You should use about 15 minutes for this activity.**

1. Write in your own words what you understand about data.

2. Study **Appendix A**, p. 78, and explain the meaning of each of the following types of data:

Types of data	Explanation with examples
Discrete data	
Continuous data	

Now that you have more knowledge on the different types of data, we can continue our discussion on the sources of data.

1.3 Sources of data

That brings us to the first question: what are the different sources of data? If you need information, where do you get it from?

Activity 2: Sources

Work alone

You should use about 15 minutes for this activity.

1. Following are different types of information. Indicate where you will find this information.

Information needed	Where to find information
The pocket money of learners	
The number of house burglaries in your community.	
What is the average shoe size of Grade 8 learners?	
How many people read the newspaper?	
The *location* of the Livingstone hospital.	
What is the average temperature increase in South Africa per year?	

2. Can you identify more sources of data?

3. A learner asked you the difference between a *population* and a *sample*. How will you answer the learner? Give an example.

A data *source* should be suitable for the population you will sample. Meaning it should be able to provide the data you are looking for. From time to time there is an official count or survey of the whole population of a country to record different details about individuals – this is called a *census.* Your peers, family and community can also supply you with information. A census is more accurate, but hard to do. A sample survey is not as accurate, but may be good enough, and it is a lot easier to execute. We can also get information from the internet or newspapers.

1.4 Collecting data

Now that you know where to find information, we can move on to the important aspect of how to get the information from the source. Let us say you want to know the favourite chocolate of your learners to give as a gift to learners. You know the source will be the learners at your school. How will you get the information from them?

Activity 3: Collecting data

Work in groups

You should use about 20 minutes for this activity.

1. Brainstorm in the group to find ways to collect the information from the learners.

2. Set a questionnaire to collect the information from the learners.

3. Advise learners on tips to use when setting questionnaires.

4. A learner asked you the following: Must all questionnaires be done face to face with people? Explain your answer to the learner.

5. What is a biased question?

6. What *skills* and *prior knowledge* should learners have to successfully compile questionnaires?

When we set a questionnaire for people to complete, we must remember that we are taking valuable time from a person.

Here are some important tips:

- ensure you have enough copies to complete;
- use short, multiple-choice questions;
- give people the option to comment on questions; and
- to save time, you can complete the survey yourself while you ask people the questions.

Now that we have the information, we must revise our knowledge on how to organise the data in order to get answers to our survey questions.

1.5 Organise data

After collecting data, we must sort the data to make it easier to draw useful conclusions. Organising your data set makes it easier to work with. There are a number of tools you can use to organise your data.

Activity 4: Organising data

Work in groups

You should use about 40 minutes for this activity.

1. Study the cartoon and write down what you understand about organising data.

2. In the text we have mentioned that there are different tools to organise data. One method of organising data makes use of *tally tables* and *frequency tables.*
   a) What is your knowledge of tally tables and frequency tables?

   b) We use the poem, "Learning" by Modikwe Dikobe, as a data source to collect information:

Learning has no limitations
Last year I made mistakes
The previous year I blundered
This year, I shall do better.

In the Garden of Eden
I learned to make fire

Clothe myself
And have children

The world is boundless
Each one of us

Can contribute
Provided
We pluck off
Complexes, fear and mistrust.

Now organise the data by completing the following table:

Vowel	Tally	Frequency
A		
E		
I		
O		
U		

c) The number of cool drink cans Thabo collected per day over a period of time is:

8; 15; 2; 30; 14; 21; 14; 30; 5; 7; 17; 10; 9; 25; 26; 36; 21; 14

Complete the following table to organise the data: Study **Appendix B**, p. 81.

Number of cans	Tally	Frequency
0 – 10		
11 – 20		
21 – 30		
31 – 40		

d) Explain to learners in your class why it is sometimes necessary to group data into intervals.

e) What will be the effect on the frequency table if the data is not grouped into intervals?

f) Compile a rule on when to *group data* in class intervals.

3. Another tool that we use to organise data is called a *stem-and-leaf diagram* or display. Study **Appendix A**, p. 78, for an example.
   a) Explain what a stem-and-leaf diagram is.

   b) Use the information given to organise the data in a stem-and-leaf diagram. The data shows the marks (out of 50) of learners in a mathematics competition:

      24; 38; 18; 45; 44; 18; 10; 23; 20; 44; 9; 50; 39; 42

c) Write down some *misconceptions*/mistakes that learners make when organising data. In each case give possible *prevention strategies*.

4. Suggest a *teaching strategy* that you will use to help learners to gain *conceptual knowledge* of organising data.

When we organise data, we sort it to make it easier to summarise and represent in order to answer survey questions. You can use tally tables and frequency tables to help count and organise the raw data. The stem-and-leaf diagram is also a useful tool to help sort data from small to large. When we have a large number of different data, we can group it together in intervals, to minimise the number of *data*.

1.6 Conclusion

Choose teaching strategies e.g. *cooperative learning* or group work to give learners the chance to actively participate. Plan for possible misconceptions or mistakes and make sure that you have prevention strategies in place. Learners must understand the concepts and not only the procedure when you organise data in frequency tables and stem-and-leaf diagrams. Let learners practically set questionnaires and complete them. Ask questions like "why?" or "when?" to make sure learners can explain why they do something in a specific way. This can be an effective method of immediate assessments where you can rectify misconceptions and problems.

Effective learning starts with effective teaching. You must choose the best teaching strategy to help learners with clear and deep understanding. Learners must be actively involved in the teaching of Mathematics in order to make meaning of the mathematical concepts and procedures.

At the end of this unit I can guide my learners to			**Reason**
✓ develop a vocabulary to define data sources;			
✓ apply and develop the skills to set different types of questionnaires; and			
✓ define and use different techniques to sort data.			

Glossary

Conceptual knowledge	learning that involves understanding and interpreting concepts and the relationships between them
Cooperative learning	when learners learn from each other and help each other
Misconception	a view or opinion that is incorrect based on faulty thinking or understanding
Poaching	the illegal practice of trespassing on another's property to hunt or steal game without the landowner's permission
Prevention strategies	mechanisms or plans to help learners not to make unnecessary and continuous mistakes in Mathematics
Prior knowledge	the skills and knowledge the learner needs to have mastered in order to construct new knowledge and understanding
Source	the "place" to collect raw data for a survey, e.g. learners in a team
Teaching strategy	the method of teaching that is used to help learners understand knowledge, e.g. group work or direct instruction

References

Groenewald, M., Otto, H., Roos, H. & Van der Westhuizen, G. (2012). *Mathematics Today, Grade 8 Learner's Book.* Maskew Miller Longman: Cape Town.

Morrison, K., Slamang, M., Greenstein, L. & Carter, P. (2013). *Ken & Verstaan Wiskunde, Leerdersboek Graad 7.* Cambridge University Press: Cape Town.

Phillips, M.D., Basson, J. & Odendaal, J. (2014). *Mind Action Series. Mathematics Textbook Grade 9*. Allcopy publishers: Sanlamhof.

Yule, D., Harris, L., Pretorius, J., Kriel, D. & Ebrahim, R. (2013). *Oxford Successful Mathematics Grade 8 Learner's Book*. Oxford University Press: Cape Town.

Unit 2: Summarise data

Outcomes

At the end of this unit, you will be able to guide your learners to

- ✓ define, name, calculate and summarise data using measures of central tendency;
- ✓ name, calculate and summarise data using measures of dispersion; and
- ✓ develop and use techniques to *compare* data using calculations.

2.1 Introduction

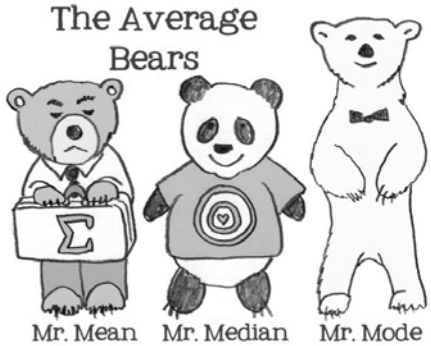
What is the first thing that comes to your mind when I mention the phrase "summarising data"? Many of you will answer "calculations!" Is that what we want in the *effective* understanding of Mathematics?

There are two types of knowledge in Mathematics, namely conceptual knowledge and procedural knowledge. In conceptual knowledge the learner understands concepts and the relationship between concepts. Procedural knowledge is knowledge about the steps and processes we use in Mathematics.

Is summarising data just about calculations and how do you teach this topic to your learners? We will come back to this important question at the end of the unit.

2.2 Measures of central tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, *measures of central tendency* are sometimes called measures of central location. By calculating the measures of central tendency, we can describe what the middle part of a data set "looks" like. The *mean* (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the *median* and the *mode.*



2.2.1 Mean

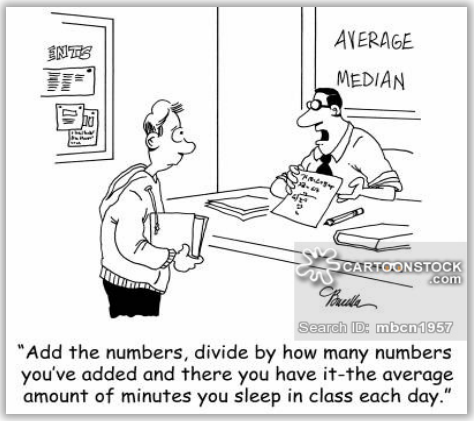
The mean of a data set is the average value. The most common use of mean or average for teachers is to calculate the class average or mean after each assessment. Why do you need to calculate the mean of the assessment? Let us find the answer in the following activity.

Activity 1: Calculate the mean

Work alone

You should use about 35 minutes for this activity.

1. How do we calculate the mean or average of a set of data? Do you agree with the calculation of mean as shown in the cartoon?



2. Consult **Appendix C**, p. 84, and state the conditions under which the mean can be used as a measure of central tendency.

3. The following statement was made by a learner: "The mean is seldom one of the actual values in a set of data." How will you respond to your learner?

4. Name two advantages and two disadvantages for calculating the mean of a data set. Study **Appendix C**, p. 84, for more ideas.

5. Identify common mistakes that learners make when they calculate the mean of a data set. Give possible prevention actions.

Now that we have revised the calculations of the mean, when to use it and when not, we can have a look at the other two measures of central tendency that we can also use.

2.2.2 Median

As you have discovered in the previous unit, the mean is not the only measure of central tendency we can use and it has disadvantages under certain conditions. As an alternative, we can calculate the median to represent the set of data and to tell us more about the middle of the data.

Activity 2: Calculate the median

Work in groups

You should use about 35 minutes for this activity.

1. Write down, in your own words, the meaning of median.

2. Explain how to calculate the median of a set of data. Study **Appendix C**, p. 84, for guidance.

3. Provide possible mistakes learners might make when calculating the median of a set of data.

4. The following data set represents the annual salaries of people working at a company. Calculate the median of the following set of data:

R159 000; R259 000; R137 000; R651 000; R462 010; R351 002; R463 200; R229 000; R239 000; R526 412; R413 100; R345 000; R189 000; R196 000; R210 000; R125 000

5. The following data set represents the body temperature (in °C) of patients in a hospital. The normal body temperature is between 36,5°C and 37,0°C.
   a) Calculate the median of the following data set.

38,9	36,7	40,2	39,5	38,2	36,8	37,0	36,7	38,2	38,6	40,3
40,5	37,6	38,5	37,9	38,0	38,5	40,2	39,4	38,9	36,9	38,4

   b) Calculate the percentage of patients with body temperature in the normal range.

6. A teacher gave learners the following to do:
   Calculate the median of the data set:

19	12	8	25	30	34	54	67	90	12	54	64	19	27	28	21	37

   One of the learners got 90 as an answer. Evaluate the answer and give feedback to the learner.

7. Name two advantages and one disadvantage of using median as a measure of central tendency. Study **Appendix C**, p.84.

8. What effect will grouped data have on the median?

The median and the mean may be used to provide more information on the middle values in a data set. The median is not influenced by outliers and skewed data. These are not the only calculations to compare sets of data.

We can also compare data by referring to the mode, the last measure of central tendency to discuss.

2.3 Mode

The mode of a set of data is the value (or values) that occurs the most, meaning the value with the highest *frequency*. Many learners would prefer to use mode to compare different sets of data as it actually requires no calculation at all.

Activity 3: Calculate the mode

Work in pairs

You should use about 25 minutes for this activity.

1. How will you explain to learners the use of mode in data handling? Use an example and explain your answer to your peer.

2. Explain to learners how to find the mode of a set of data? Study **Appendix C**, p. 84, for tips.

3. Can there be more than one mode for a set of data? Explain. Study **Appendix C**, p. 84.

4. Identify possible mistakes that learners make when they calculate the mode of a data set.

You must ensure learners understand why we calculate the measures of central tendency – we want to compare sets of data with each other. The emphasis is on comparison to determine which set of data is the "best". Learners can choose any of the three calculations (mean, median or mode) to determine which set of data is the "best". Each of these calculations has its own unique shortcomings. But we must remember that the aim is to be able to compare the data from different sets.

Learners must not only have procedural knowledge about the calculations on mean, median and mode. It is your responsibility to also develop their conceptual knowledge of these concepts. Learners must understand when and why a specific measure of central tendency is used. Do you ask your learners "why?" and "when?" when you cover this topic?

2.4 Measures of dispersion

Do we only compare sets of data according to the middle part? What if the middle values of the sets of data are similar? Does this mean that the data sets are identical without any differences? Let us do the following activity to answer these important questions.

Activity 4: Dispersion

Work in groups

You should use about 25 minutes for this activity.

Do the following exercise:

1. Analyse the given sets of data by calculating the mean, median and mode of each:

A:	5	8	9	12	15	21	21	21	24	30	31	35
B:	8	9	10	11	18	21	21	22	24	26	30	32

Mean of A	
Mean of B	
Median of A	
Median of B	
Mode of A	
Mode of B	

2. What do you notice about the central measures of tendency?

3. Are the two sets of data identical? Explain.

4. Study **Appendix C**, p. 84, and give a definition for the range as a *measure of dispersion*.

5. Calculate the range of the two sets of data given in number 1.

6. What will a large range value tell you about the set of data?

7. What will the effect of grouped data be on the range of data?

Measures of dispersion assist us to identify possible *outliers* in a set of data. The outliers tell us which measures of central tendency to use. The measures of dispersion assist us to find the boundaries of a set of data. Not all sets of data have outliers, but all sets of data have a range. To identify an outlier, we need to make a *box-and-whisker plot* of the data; all values outside of 1,5 times the *interquartile range* are considered to be outliers. This is covered in the FET phase of Mathematics.

2.5 Skewed distribution of data

When we draw a histogram of data (we will discuss how to do this in the representation of data in the next unit) we can comment on the distribution of data. Data can either show a normal distribution, or be skewed to the left or to the right. Let us investigate this in the following activity.

Activity 5: Distribution of data


Work in groups

You should use about 10 minutes for this activity.

1. Study **Appendix C**, p. 84, and explain what a normal distribution of data will look like.

2. Comment on the statement: If the mean is smaller than the median, the data is skewed to the right.

Let me take you back to the beginning of this unit. If the central part of data is identical, the measures of central tendency are the same. It does not mean the whole data set is identical. We calculate the measure of dispersion to find the spread of data. This leads to a difference between data sets.

Although the distribution of data is not in the syllabus for learners in the Senior Phase, we must have the knowledge to be prepared for possible questions from our more gifted learners.

2.6 Conclusion

Let me take you back to the question at the start of this unit. How do you teach the summarising of data to your learners? Are you only focused on the calculations, and the procedural knowledge? Or do you emphasise the importance of understanding the concepts behind the calculations? As agents of change, I hope you realise the importance of effective understanding and *constructive learning*.

How do we know that effective learning has taken place? I believe that a lot of you might give answers like informal tests, homework and class work. But are those the only way? It will take some time before you will know if effective learning has taken place. What about asking questions in the class? Remember the "why?" and "when?" questions. By asking questions during your teaching, you can immediately detect misconceptions and problems. You can save yourself a lot of time in the end.

I hope that you are inspired as an agent of change to focus on the effective learning and constructive understanding of mathematical concepts.

To summarise, data enables us to form a picture of what the data set looks like in order to compare it to another data set. We must be careful not to teach this section only as formulas, but must ensure that learners understand the concepts behind the calculations. Learners usually enjoy the calculations, although they can sometimes confuse the different measures of central tendency. In the next unit, we will look at how to represent and interpret data.

At the end of this unit I can guide my learners to			**Reason**
✓ define, name, calculate and summarise data using measures of central tendency;			
✓ name, calculate and summarise data using measures of dispersion; and			
✓ develop and use techniques to compare data using calculations.			

Glossary

Box-and-whisker plot	plotting the minimum value, lower quartile, median, upper quartile and maximum value on a number line
Compare	finding the differences and similarities between sets of data
Constructive learning	when learners build on prior knowledge and make meaning of what they have learnt
Frequency	the number of times a value appears in a given set of data
Interquartile range	the difference between the upper and lower quartiles. Divide data into parts with the median and each part is halved again by the upper and lower quartile
Mean	the sum of the numbers in a set, divided by how many numbers are in the set
Measures of central tendency	summaries to compare sets of data in terms of the middle part of the sets of data – they include calculating the mean, median and mode
Measures of dispersion	the spread of data, how wide or narrow data values are apart
Median	the middle value in a sorted data set

References

Lund, A. (2013a). *Measures of central tendency.* Available at https://bit.ly/1zX41bK. Accessed: 4 April 2017.

Lund, A. (2013b). *Measures of Spread.* Available at https://bit.ly/2Sf53zQ. Accessed: 4 April 2017.

Unit 3: Represent data

Outcomes

At the end of this unit, you will be able to guide your learners to

- ✓ draw, compile, understand and represent various graphical representations of *univariate* and *bivariate data*; and
- ✓ develop the skill to compare different graphical representations, including their advantages and disadvantages.

3.1 Introduction

In the previous unit we discussed how to summarise data. In this unit we will investigate the representation of a set of data. If I mention the word "representation", what is the first thing that comes to your mind? Graphs! I believe this is the answer for most of you. Let us orientate ourselves on the aspects of data handling that we have discussed so far: We have a problem and a survey question, so we can collect data, organise the data and summarise the data. The next step is to make a visual representation of the data.

What do you understand about prior knowledge? Why do you think it is important to determine if learners have the necessary skills and knowledge before you can start a topic? In the following activity we will put the topic of representing data into practice.

Activity 1: Prior knowledge for drawing graphs

Work alone

 **You should use about 10 minutes for this activity.**

1. What prior knowledge and skills must learners have before they draw graphs?

2. How can you assess if learners have the necessary pre-knowledge?

We can differentiate between *univariate data* and bivariate data. We will start with the representation of univariate data.

3.2 Bar graphs

A *bar graph* is a visual display that makes use of vertical and horizontal bars whose length represents discrete data. In unit 1 we discussed the meaning of discrete data – revisit this work before you move on. Ensure that you emphasise that the bars do not touch each other when you teach your learners to draw bar graphs.

Here are guidelines to help you when you construct bar graphs:

- identify the title or heading for the graph;
- identify the labels on the axes;
- determine the scale; and
- draw the bars.

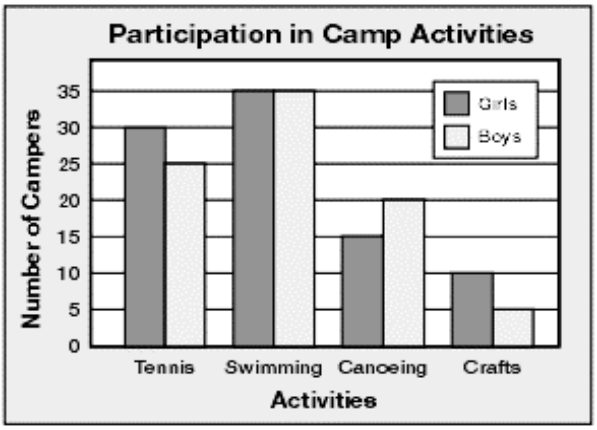
We can have different types of bar graphs, including vertical, horizontal, double and compound bar graphs. Let us investigate.

Activity 2: Graphs

Work in groups

You should use about 30 minutes for this activity.

1. Study the following example of a bar graph and answer the questions that follow:

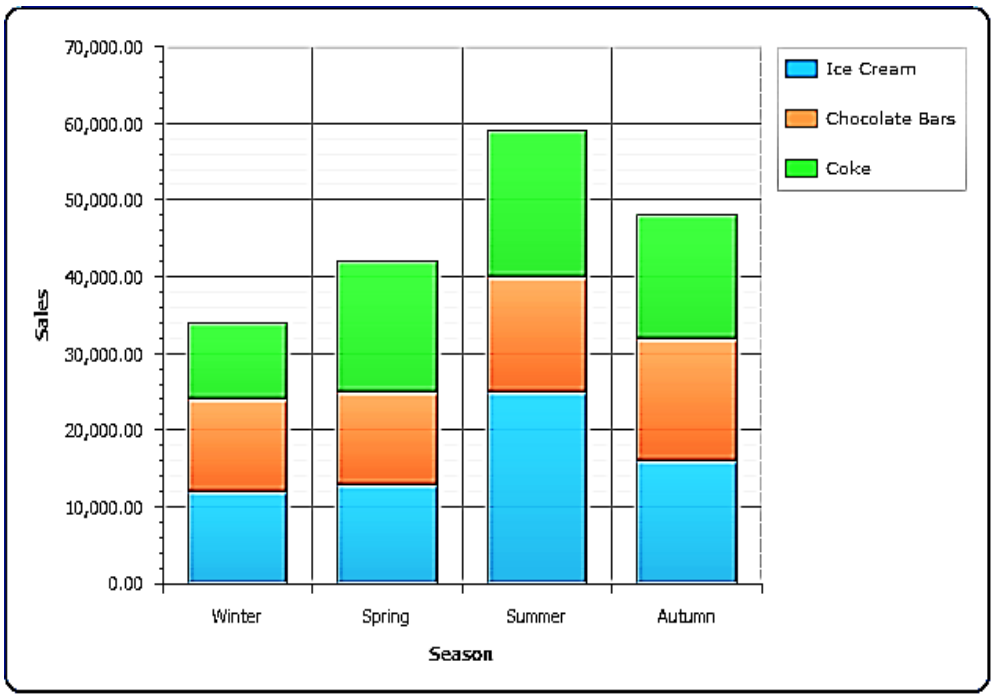


https://bit.ly/2Q3OqDO

a) Identify possible mistakes that learners can make. Explain.

b) What other type of graph(s) can be drawn to represent the same information? Will the effect be similar to this graph?

2. Study the bar graph:



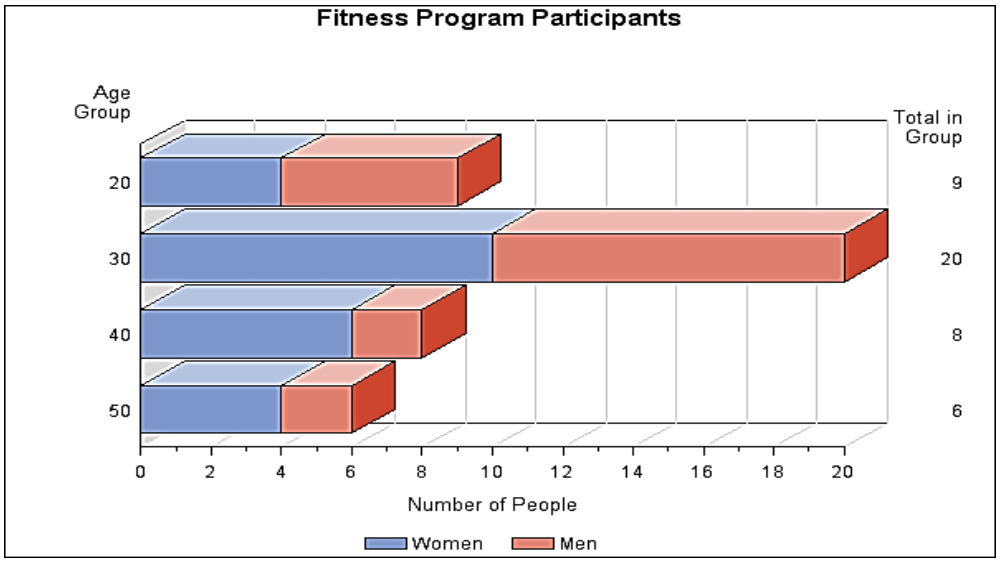
https://bit.ly/2Jxiwhv

a) Identify the type of bar graph.

b) Calculate the sales from Coke during spring.

c) Write down how difficult or easy it was to get the answer of b. Give tips to the learners to help them with similar questions.

3. Another type of bar graph:

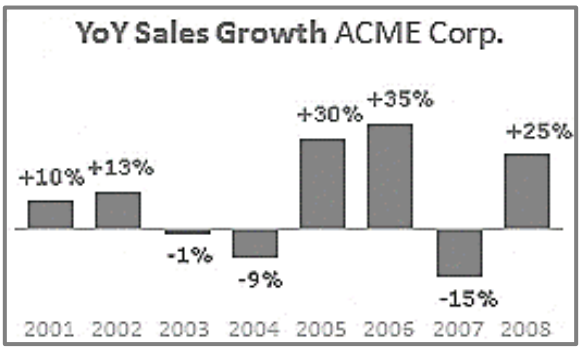


https://bit.ly/2WCrqOF

a) Why do you think it is called a horizontal 3D- bar graph?

b) The following statement is based on the graph: "In the age group 50, there are 3 men". Is the statement correct? Explain.

4. Study the following graph:



Some of the bars of this bar graph are negative. Explain the meaning of negative values.

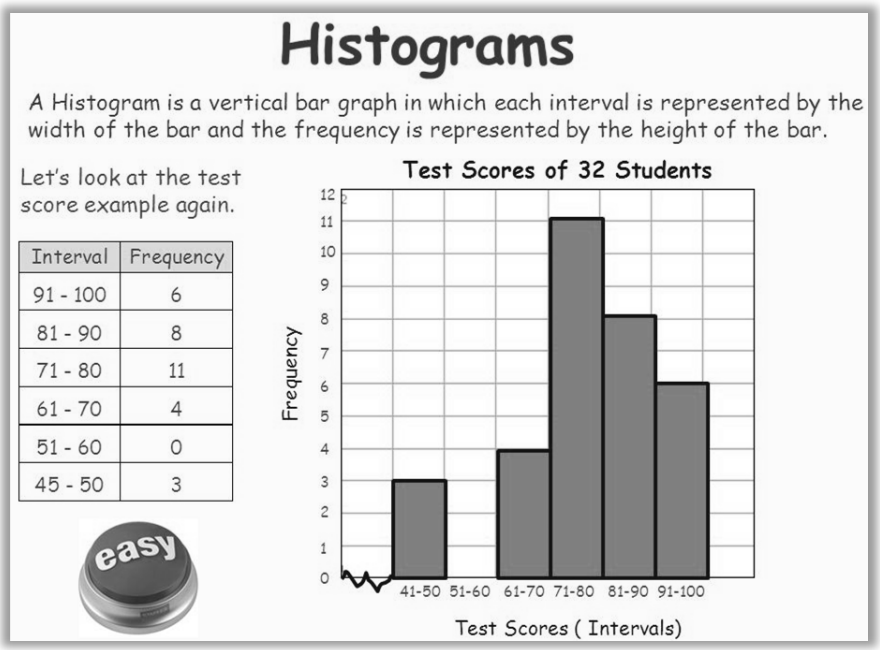
5. Identify common mistake learners make when drawing bar graphs and indicate possible prevention strategies to help these learners in the table below.

Mistakes of learners	Prevention actions

Bar graphs are easy to read and construct. A bar graph shows the trend of the data by looking at the rise or fall in the length of the bars. We can see the effect of different categories on the same set of axes to make conclusions easier for the viewer.

3.3 Histogram

A *histogram* makes use of bars, similar to the bar graph, although in this case the bars are touching each other. Why? The bars will touch because we use a histogram to represent *continuous data* which are grouped in consecutive intervals. When we determine the intervals, we ensure that all values fit into the intervals, explaining why the bars will touch.



https://bit.ly/2E3Jn1j

When you construct a histogram **with given intervals**, you must follow the following guidelines:

- plot the data in the given intervals;
- the interval data will be plotted on the horizontal axis;
- determine the scale to use for the vertical axis;
- use the frequency values to decide on the scale; and
- indicate the graph title and axes labels.

When you construct a histogram **without given intervals** you must follow these guidelines:

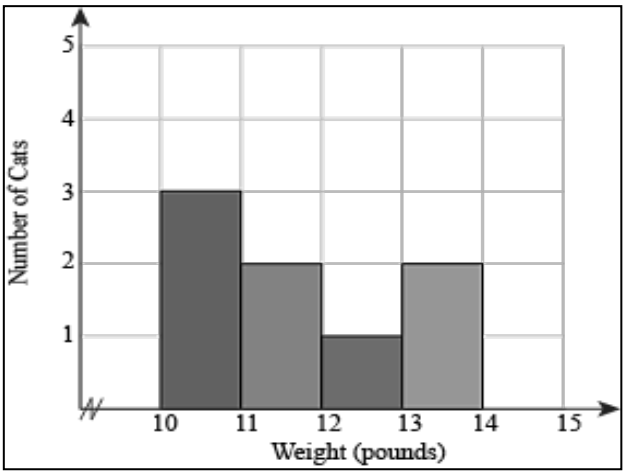
- work out the range of the data set;
- decide how many intervals to use, usually between 5 and 10;
- decide on the width size of the intervals, use whole numbers; and
- draw up a table with the class intervals and plot the graph.

Activity 3: Drawing a histogram

Work in pairs

You should use about 25 minutes for this activity.

1. Study the example of a histogram and answer the questions that follow:

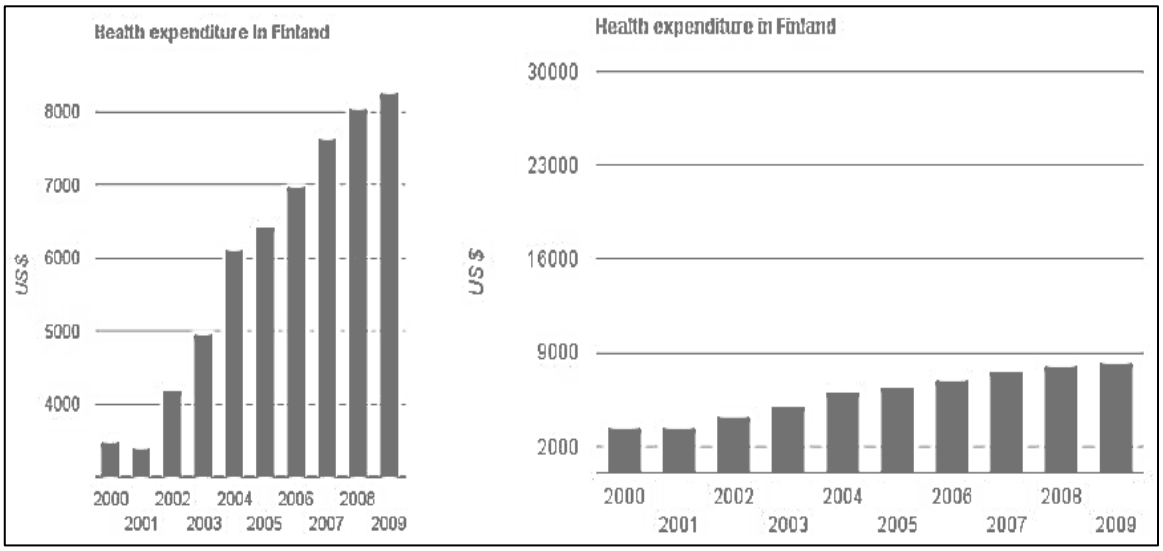


a) Explain the meaning of the symbol shown where the two axes intersect.

b) Do you agree with the following statement: the horizontal axis of a histogram is a number line? Motivate your answer.

c) How will you explain to learners the *general trend* of the graph?

2. Study the following two graphs showing the health expenditure in Finland:



https://schoolofdata.org/handbook/courses/common-misconceptions/

a) A person wants to show that the health expenditure in Finland has not increased much over a period of time. Which graph will the person use and why?

b) Do you agree with the following statement? "The two graphs represent the same information". Explain your reasoning.

c) What do you understand about "misleading graphs"?

3. Compare a bar graph and histogram by completing the following tables:

Differences	
Bar graph	**Histogram**

Similarities

4. Is it possible to have double histograms like double bar graphs, or compound histograms like compound bar graphs?

Histograms are used when we want to represent continuous data, causing bars to touch each other. They focus on the general trend of the data. The histogram is a graph that is easy to read and draw. Different sets of data require different histograms as we cannot use double and compound bars. This is one disadvantages of using a histogram.

3.4 Pie charts

The *pie chart* is a circular chart consisting of sectors representing the information in the data set. Learners often find the drawing of the pie chart difficult as it requires some calculations. The angle of each sector corresponds with the value of the item in percentage or fraction form. The total of all the sectors will always add up to 100%, which is 360° on the pie chart.

Pie charts are convenient when different parts of a whole amount are compared. Representing the different expenses of a budget is one such example.

Activity 4: Prior knowledge for pie charts

Work alone

You should use about 10 minutes for this activity.

1. Reflect on the pre-knowledge learners must have before you can allow them to draw pie charts. Give a reason in each case.

Prior knowledge	Reason for prior knowledge

2. What instruments are needed when learners draw pie charts?

When you construct pie charts, the following guidelines will help:

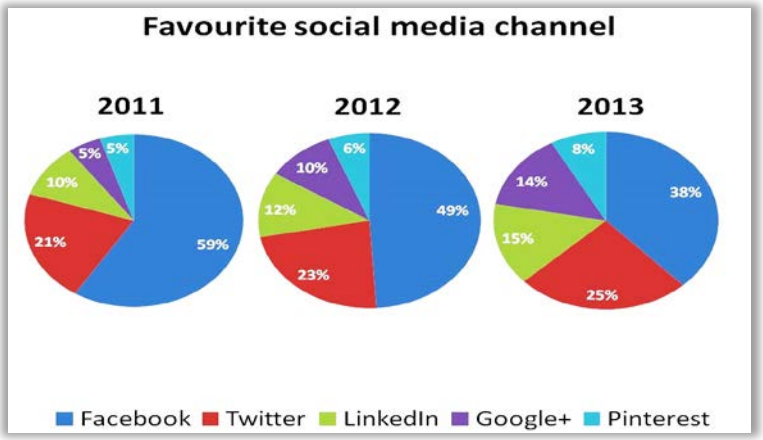
- convert all the data to percentages; then convert all the percentages to angles in degrees;
- draw a circle, using a pair of compasses;
- draw a radius using a *protractor* and measure the size of each sector; and
- label each sector.

Activity 5: Draw a pie chart

Work in groups

You should use about 30 minutes for this activity

1. Study the following pie charts and answer the questions that follow:

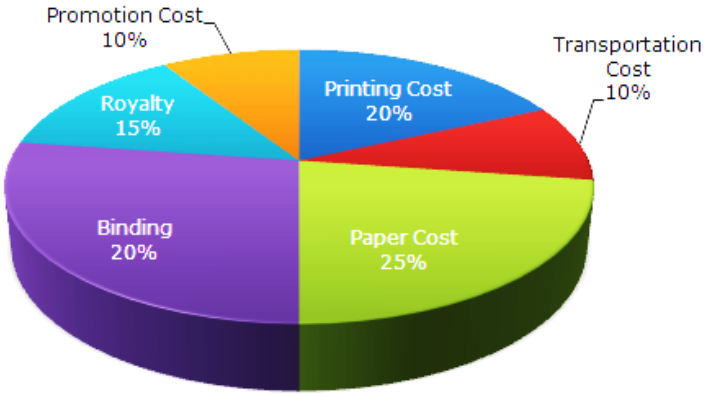


https://bit.ly/2vQTaTR

a) Do you agree with the following statement: Twitter has increased by 4% from 2011 to 2013? Explain your answer.

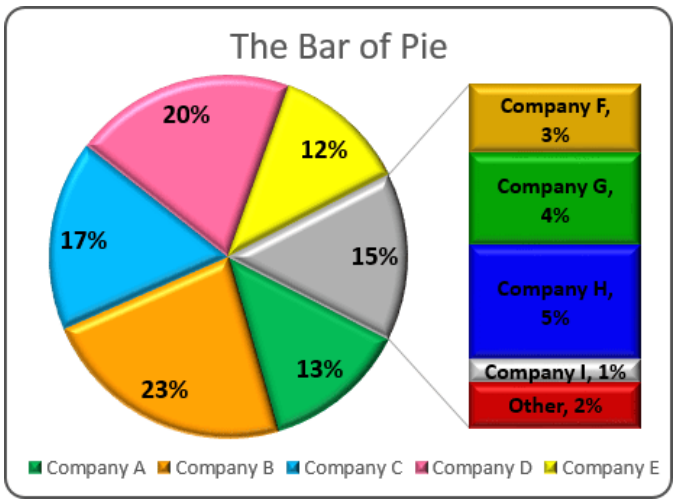
b) Do you agree with the following statement: The number of people using Twitter increased by 4% from 2011 to 2013? Explain.

2. Study the following pie chart:



This is a 3D pie chart. What is the difference between this pie chart and a basic pie chart?

3. Here is a pie chart with a bar graph to illustrate information.



https://bit.ly/2VmlkRa

a) Why do you think the 15% slice of the pie chart is represented by a bar graph?

b) What other graphs could be used to represent the 15% slice of the pie chart? Explain.

4. Reflect on common mistakes learners make when drawing a pie chart. Indicate what you as a teacher can do to prevent these mistakes.

Possible mistakes	Prevention actions

Pie charts are easiest to interpret when each segment of the pie chart is shaded or coloured differently. Ensure that your learners can read pie charts and answer questions about them.

We must accommodate learners with different *learning styles*. Pie charts drawn in different colours are an excellent opportunity to accommodate the *visual learning* style of learners in your class. The visual learner enjoys colourful pictures to understand and summarise information.

3.5 Line and broken-line graphs

In a *line graph* the data points lie on a line. The *broken-line graph* is a graph where plotted points are joined by line segments. It can only be used if the values on the horizontal axis represent points on a continuous function. This type of graph is useful as it clearly shows trends and can easily be extended. The broken-line graph is used for data that changes continuously, like the temperature during a month. Each day the temperature is taken and plotted. The points are then joined with a broken line as we don't have the information of the in-between temperatures.

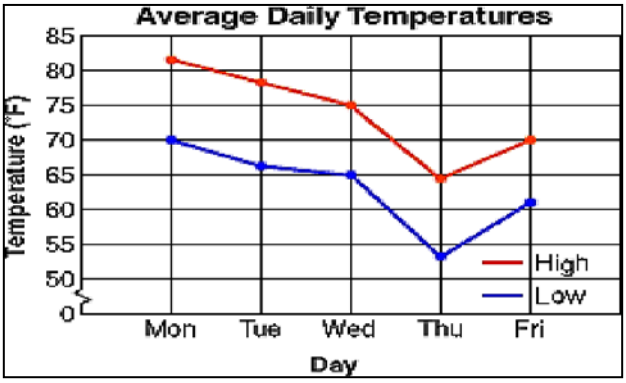
Activity 6: Draw a broken-line graph

Work in pairs

You should use about 25 minutes for this activity.

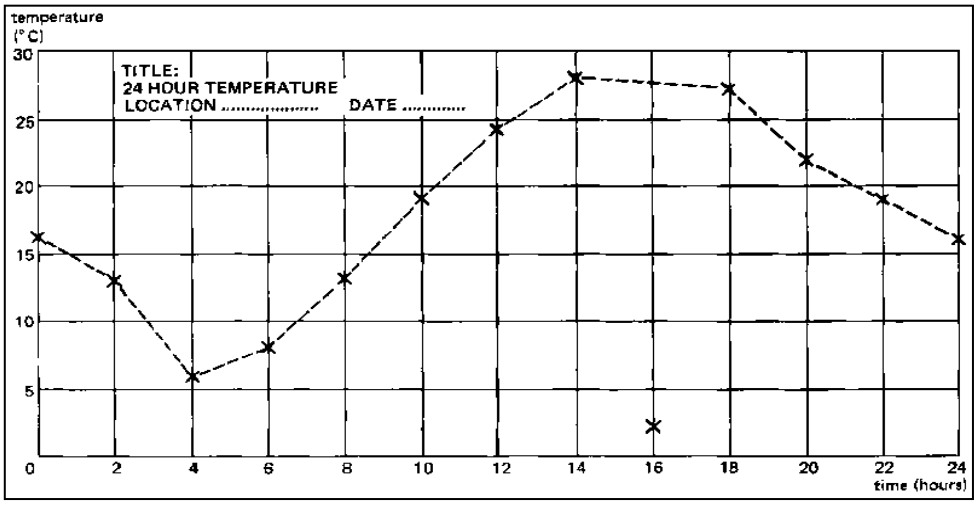
1. Study the following graphs and answer the questions that follow:

Graph A



http://www.biologyjunction.com/images/clip0079.gif

Graph B



http://www.fao.org/docrep/r4082e/r4082e1k.gif

a) Give a reason for the broken line in graph B.

b) Graph A and B represents daily temperatures. What is the reason for the difference in these types of line graphs? Explain.

c) A learner makes the following statement: It was about 10°C at 07:00. How will you answer this learner?

2. Reflect on possible mistakes of learners when drawing line and broken-line graphs, and provide prevention actions from the teacher.

Possible mistakes	Prevention actions

Broken-line graphs are useful to read trends and patterns in data and for predictive purposes. We can usually predict if the temperature will go up or down in the next month. Readings are taken at regular intervals, which are shown on the horizontal axis. The data values are shown on the vertical axis. The points on the broken-line graph will not lie in a straight line.

These are the types of graphical representations for univariate data. In the next section we will have a look at what type of graphical representation to draw for bivariate data.

3.6 Scatter plot

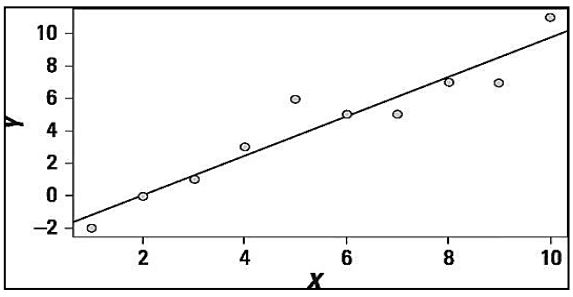
Scatter plots represent the relationship between two sets of data. The *scatter plot* reveals if there is a relationship or correlation between two sets of data. Any of the two sets can be represented on the horizontal axis. You plot the points and look for the correlation between the two sets of data.

Activity 7: Draw a scatter plot

Work in groups

You should use about 20 minutes for this activity.

1. Study the following scatter plot and answer the questions that follow:



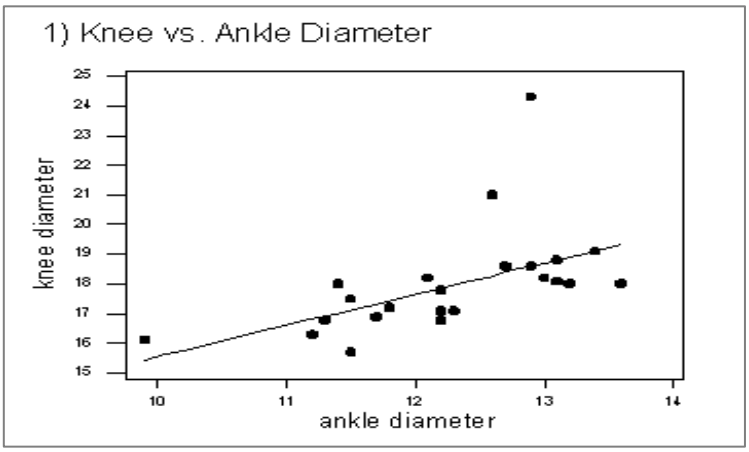
https://bit.ly/2WCu9aR

a) What is represented by the solid line in the graph?

b) Design your own story to fit with the trend of this scatter plot.

c) Give the general trend shown by the scatter plot.

2. Here is another scatter plot:



https://bit.ly/2VPwD8L

a) Explain the general trend of the scatter plot.

b) Explain to your learners what an outlier is and how you will identify it on the graph.

3. Reflect on possible mistakes learners can make, and indicate the actions you will take to prevent these mistakes.

Possible mistakes	Prevention actions

The use of the scatter plot continues in Mathematics as well as in Mathematical literacy in the FET Phase. As Senior Phase teachers we must ensure that we lay a solid foundation for the graphical representation of data. In the Senior Phase we will not require learners to name the correlation, but this can be done for enrichment.

Here is something to help you guide your learners in choosing the best type of graph for a specific case: Line graphs can be used to compare changes over the same period of time for more than one group of data. Pie charts are best to use when you are trying to compare parts of a whole. They do not show changes over time. Bar graphs are used to compare things between different groups of data or to track changes over time.

3.7 Conclusion

The graphical representation of data is done to make a visual summary of the data in order to compare different data sets. The graphical representation must help solve the survey problem or answer the survey question. Often the graphs are also used to make predictions based on known information.

We can use computer programs like Excel to draw graphical representations that are colourful, accurate and visual stimulating. Remember to provide suitable headings (titles), axes labels and a suitable scale.

To represent data graphically is an excellent opportunity to accommodate learners with different learning styles. We have three types of learning styles namely visual, auditory and tactile/kinaesthetic. The visual learners will enjoy drawing graphs in different colours, especially the pie charts and bar graphs. For the auditory learners, read the information aloud and repeat. Use different voice tones and loudness to accommodate these learners. What about the *tactile/kinaesthetic* learners? Look at this picture:

http://creativestarlearning.co.uk/

This is an interesting idea to accommodate these tactile/kinaesthetic learners: Let them draw a graph on the school grounds. They will enjoy the physical drawing and construction of the graph on the ground or tarred surface. We must be creative and unafraid to try new techniques to help our learners construct meaning. Choose teaching strategies where your learners are actively involved in learning. We must help them to develop conceptual knowledge of how and when to draw the different graphs.

In the next unit, we will expand our knowledge on the interpretation, analysis and reporting of data.

At the end of this unit I can guide my learners to			Reason
✓ draw, compile, understand and represent various graphical representations of univariate and bivariate data; and			
✓ develop the skill to compare different graphical representations, including their advantages and disadvantages.			

Glossary

Auditory learner	a person that prefers to hear in order to learn mathematical concepts
Bivariate data	data that has two variables
Continuous data	data that has an infinite number of possible values within a selected range, e.g. temperature range
General trend	what the graph tells you about the data
Learning styles	the method a person prefers to use when learning
Scatter plot	represents the relationship between two sets of data
Tactile/kinaesthetic learner	a person that prefers doing in order to learn mathematical concepts
Univariate data	data with only one variable
Visual learner	a person that prefers to write and draw in colours in order to learn mathematical concepts

References

Groenewald, M., Otto, H., Roos, H. & Van der Westhuizen, G. (2012). *Mathematics Today Grade 8 Learner's Book* Maskew Miller Longman: Cape Town.

Morrison, K., Slamang, M., Greenstein, L. & Carter, P. (2013). *Ken & Verstaan Wiskunde, Leerdersboek Graad 7.* Cambridge University Press: Cape Town.

Phillips, M.D., Basson, J. & Odendaal, J. (2014). *Mind Action Series Mathematics Textbook Grade 9*. Allcopy publishers: Sanlamhof.

Press, K., Morrison, K. (2012). *Study & Master Mathematical Literacy Learner's Book Grade 12.* Cambridge University Press: Cape Town.

Yule, D., Harris, L., Pretorius, J., Kriel, D. & Ebrahim, R. (2013). *Oxford Successful Mathematics Grade 8 Learner's Book*. Oxford University Press: Cape Town.

Unit 4: Interpret, analyse and report data

Outcomes

At the end of this unit, you will be able to guide your learners to

- ✓ develop the ability and skills to interpret, analyse and report data; and
- ✓ develop the ability and skills to identify and comment on misleading graphs.

4.1 Introduction

In the previous unit we discussed and explored the different graphical representations of data. Yes, we discussed how and when to use bar graphs, histograms, pie charts as well as line and broken-line graphs. Can you remember why we draw graphs of collected data sets?

4.2 Interpret data

To answer the question asked in the previous section, we need to represent data in the form of graphs as it makes the data visual, easier to understand and possible to *interpret*. That brings us to the important concept of the meaning of interpreting data.

Graphical representation is a useful way to gain an instant picture of the distribution of data and to identify any relationships in the data that require further investigation. Patterns in data can be more easily *discerned* when displayed in graphs. A range of graphical techniques can be used to present data in a *pictorial format*, for example bar graphs, histograms, pie charts or line graphs.

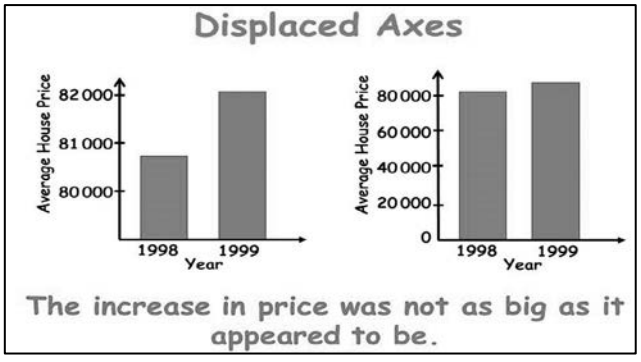
You need to read data carefully to interpret the information, because data can be *manipulated* and presented in a misleading way.

Activity 1: Interpreting data

Work alone

You should use about 30 minutes for this activity.

1. Study the two graphs, showing the average increase in house prices for 1998 and 1999.



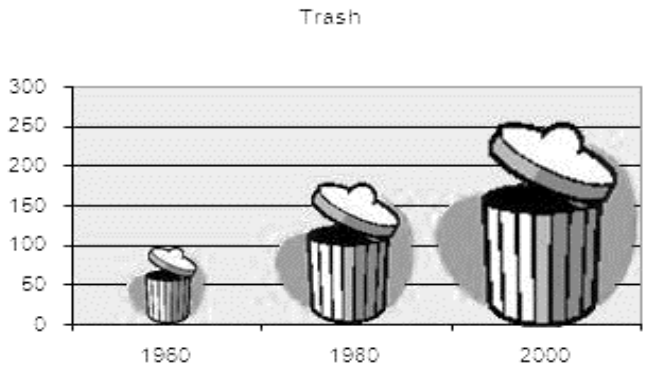
Explain why the data represented in these graphs can be considered as misleading.

2. Analyse the two graphs above and answer the following questions:
   a) You are selling houses. You want your clients to buy a house rather than to rent a house. Which of the two graphs will you show your clients? Explain.

   b) You want to show that the inflation in average house prices is far below the annual inflation rate of a country.

   c) How were the misleading graphs created?

3. Study the following graph and answer the questions that follow:

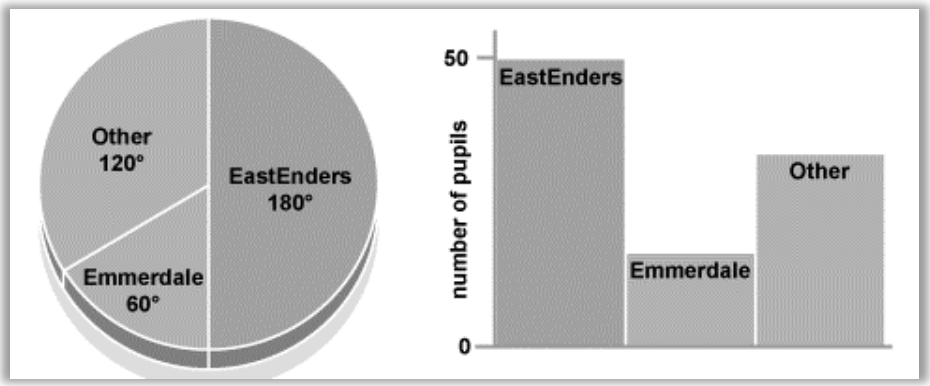


a) By what factor did the amount of trash increase from 1960 to 2000?

b) What misleading technique was used in this graph?

c) What are the real facts indicated in the graph?

4. Study the graphs and answer the questions that follow:



a) Which of the two representations will you use to show that most learners in the school live in the suburb EastEnders? Explain.

b) Which of the two representations will you use to show that many learners come from other suburbs that are not specified? Explain.

5. Comment on the following statement: People use the same information and represent it differently to *mislead* people.

Poor quality interpretation can lead you to draw incorrect and inappropriate *conclusion*s. Data can be manipulated to emphasise a specific message to the reader. You can use data either to support a statement or to mislead people. An easy way to do this is to make the vertical axis create the effect you want the viewer to see on your graph.

Here are some more ideas:

- If you want to make data values look far apart, do this by starting the vertical axis close to the minimum value of the data (not zero).
- If you want the data values to look very close together, use a much bigger range of values on the vertical axis.

4.3 Analyse data

Graphical analysis is a useful way to gain an instant picture of the distribution of data and to identify any relationships in the data that might require further investigation. Patterns in data can be more easily discerned when displayed in graphs. A range of graphical techniques can be used to present data in a visual and pictorial format.

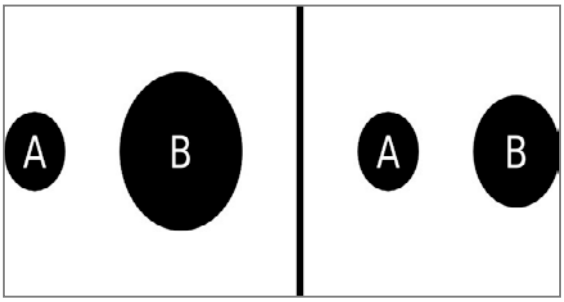
Activity 2: Analysis

Work in groups

You should use about 30 minutes for this activity.

1. What do you understand under the idea of "analysing data"?

2. A teacher asked her learners the following question: A company wants to illustrate that the sales of product B is double the sales of product A. Which of the following representations will be accurate? Explain your answer.



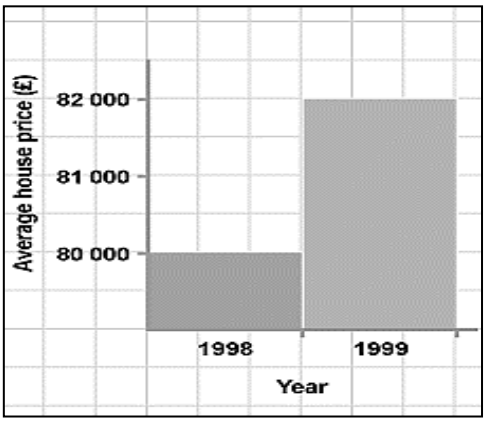
https://schoolofdata.org/handbook/courses/common-misconceptions/

Answer of a learner:

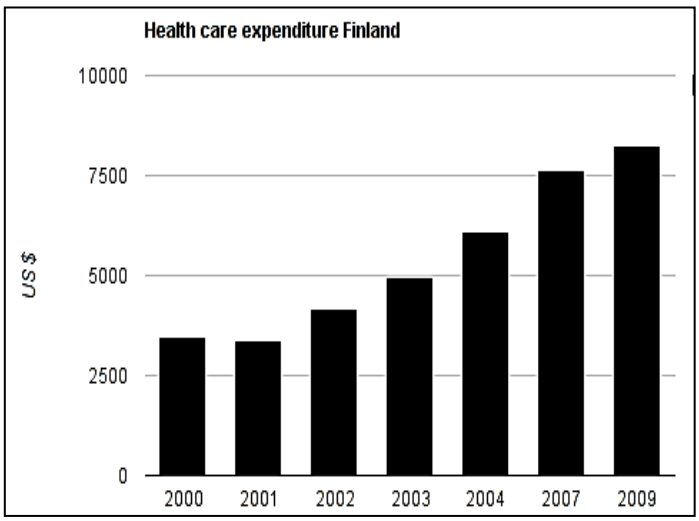
I will use the representation on the left as circle B looks double the size of circle A.

Evaluate the answer and give feedback to the learner. (**Tip:** Use the area of a circle to determine if the learner's answer is correct.) Study **Appendix B**, p. 81, for more ideas.

3. Study the following graphs and indicate reasons why each graph is misleading. Study **Appendix B**, p. 81, for more ideas.



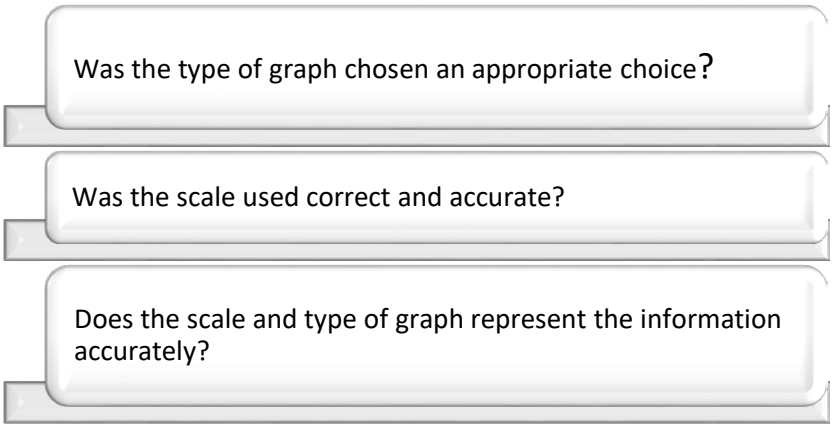
https://bbc.in/2LxuHxw



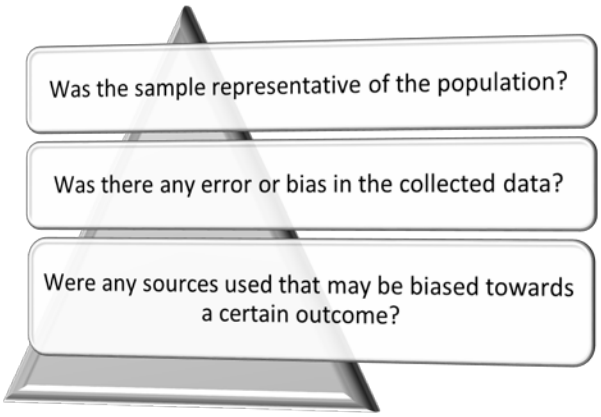
https://schoolofdata.org/handbook/courses/common-misconceptions

Handy tip: Always examine data and data representations critically as graphs in newspapers, magazines and on the internet can be presented in different ways. Think about whether the information is represented accurately or whether there is something missing or hidden.

When we analyse represented data we can ask ourselves the following questions:



The sources used to collect the information can also lead to misleading data. Ask yourself the following questions when collecting data:



4.4 Reporting on data

When you are doing a statistical study, you need to state the purpose of data collection and describe your collection method. Then the data must be organised, summarised, represented and analysed. The entire data collection process must be *valid* and free of bias. When you *analyse data*, be on the lookout for any misleading data or representations. After you have done all these things, you can *report* on the data and make informed, unbiased *conclusions*.

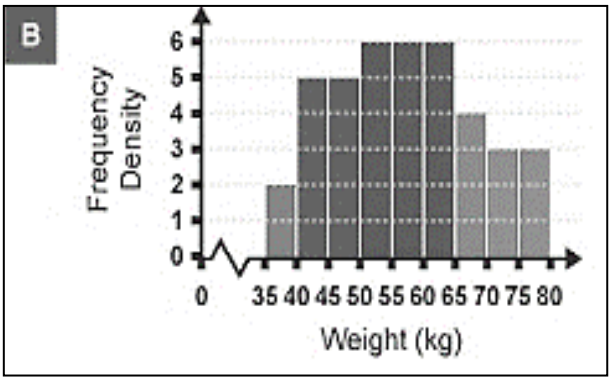
A report must be written based on the findings of the investigation. You can summarise the questions posed in the analysis of the data and the conclusions. The report can contain written conclusions about current attitudes as well as the way forward. In the report, you can include recommendations to interested parties.

Activity 3: Reporting

Work in pairs

You should use about 35 minutes for this activity.

The following graph represents the weight of learners at a school.



1. Compile a report based on the information represented in the histogram.

2. Reflect on your teaching of data handling in the past. What will you do differently?

3. Why must we teach data handling to our learners?

We must compile accurate and complete reports on information collected. The report must answer the survey question or problem. In the report we can also do predictions based on the information.

Whenever you conduct a statistical investigation, you should follow the steps of the data cycle :

- **Select a topic of research**
  This can be a problem in the community, or attitudes of people towards certain topics.
- **Collect the data**
  The researcher collects data in the form of questionnaires.
- **Record the data**
  Once the data has been collected, the researcher should record the data captured from the questionnaires.
- **Organise the data**
  The data must be organised into frequency tables.
- **Represent the data**
  The researcher must decide which graphs will be most suitable for representing the data.
- **Analyse the data**
  The researcher studies the graphs and look for trends in the data. Measures of central tendency and measures of dispersion are calculated.
- **Summarise the data**
  The findings from the data analysis must now be summarised in a sensible manner so that conclusions can be made in response to the question investigated.
- **Report the data**
  A written report on the findings is drawn up. A final conclusion can be made and recommendations can be given to the interested parties.

4.5 Conclusion

The first four units of this module have shown how the process of data handling is a continuous process. Do not make the teaching of data it formula driven; rather emphasise the process of finding answers to problems or questions. Your learners must understand the purpose and form a holistic picture of data handling. Choose investigation topics that will interest them in finding out more information and keep them motivated until they write a report. In the last unit, we will discuss probability.

Note: Bring a die to the next contact session.

At the end of this unit I can guide my learners to			Reason
✓ develop the ability and skills to interpret, analyse and report data; and			
✓ develop the ability and skills to identify and comment on misleading graphs.			

Glossary

Analyse data	to examine in detail in order to discover the meaning or essential features
Conclusion	to give an answer to the survey question or to solve a problem
Discerned	to perceive or recognise something
Interpret	to explain the meaning of something
Manipulated	to control something to your advantage, often unfairly or dishonestly
Mislead	to cause someone to get a wrong idea or impression
Pictorial format	illustrated or expressed as pictures
Report	to create a statement or report on your findings after doing an investigation on the information given
Valid	indicates how sound or accurate your information is

References

Groenewald, M., Otto, H., Roos, H. & Van der Westhuizen, G. (2012). *Mathematics Today Grade 8 Learner's Book.* Maskew Miller Longman: Cape Town.

Morrison, K., Slamang, M., Greenstein, L. & Carter, P. (2013). *Ken & Verstaan Wiskunde, Leerdersboek Graad 7.* Cambridge University Press: Cape Town.

Phillips, M.D., Basson, J. & Odendaal, J. (2014). *Mind Action Series Mathematics Textbook Grade 9*. Allcopy publishers: Sanlamhof.

Yule, D., Harris, L., Pretorius, J., Kriel, D. & Ebrahim, R. (2013). *Oxford Successful Mathematics Grade 8 Learner's Book*. Oxford University Press: Cape Town.

Unit 5: Probability – what is the chance?

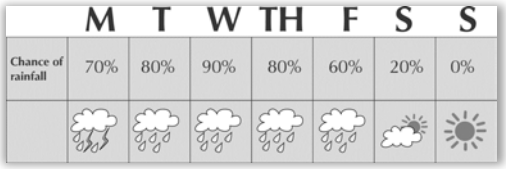
Outcomes

At the end of this unit, you will be able to guide your learners to

- ✓ use the vocabulary to define, calculate and list the possible outcomes of events; and
- ✓ *predict*, compare, explain and calculate the relative frequency of events.

5.1 Introduction

What do the activities shown in the following images have in common?



You can calculate the *chance* for each event to happen: you can determine your chance to win the National Lottery, or the chance for rain tomorrow or the chance that Bafana Bafana will win their next match. In Mathematics we refer to this chance as *probability.*

5.2 What is probability?

Probability is about calculating, estimating or predicting what might happen in the future. How do we measure it? We already mentioned that probability is about the chance that something will happen. Let us investigate this in more depth in the following activity.

Activity 1: What is probability?

Work alone

You should use about 15 minutes for this activity.

1. Use any suitable example and explain to learners the meaning of probability.

2. Differentiate between *theoretical probability* and experimental probability? Use the throwing of a die as an example to illustrate the difference.

3. The following question is asked: You have an apple, a sandwich and an energy bar in your bag for lunch. Calculate the probability of picking the energy bar. Write down possible mistakes learners will make in answering the question. How will you prevent these mistakes?

4. When we use the formula to calculate probability, we call it theoretical probability. How will you explain this term to the learners in your class?

5. Study **Appendix A**, p. 78, and explain what *empirical* probability is. Give the formula.

The topic of probability has its own language and teachers must explain and help our second and third language learners to understand the language of probability. Now that you know what probability is, let us look at the probability scale.

5.3 The probability scale

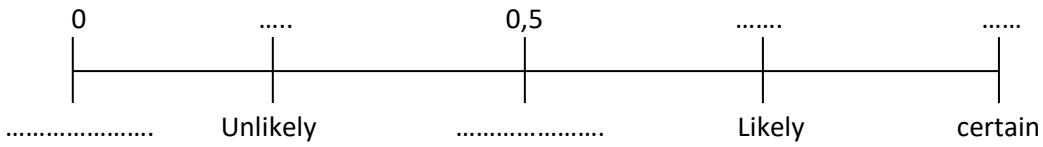
Probability always has a value between 0 and 1. Yes, if you look at your ruler between 0 and 1, you will find the probability of all possible *events*. I suppose your first reaction to this will be to say there is not a large variety in values for probability possible if it must always be between 0 and 1. You will be amazed at all the possible values, although most of them will be decimal numbers.

Activity 2: Probability

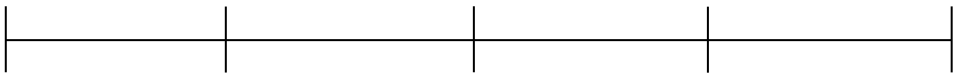
Work in groups

You should use about 20 minutes for this activity.

1. Below is an incomplete probability scale. Use **Appendix D**, p. 91, to complete the scale.



2. Indicate the probability of the following events on the probability scale given below. Only indicate the letter on the scale.



Event letter	Event
A	I will be younger tomorrow than today.
B	It will rain next week.
C	A die will land on 3.
D	There are 32 days in December.
E	A new-born baby will be a boy.
F	To pick a picture card from a full set of 52 cards.
G	A rainbow has 7 colours.

3. Explain the need for learners to practice the probability scale.

Learners must understand that probability is always a value between 0 and 1, as shown on the probability scale. We can calculate the chance that something will happen, although we sometimes cannot calculate the exact probability, but only determine the chance of something happening in terms of terminology like "likely" or "unlikely".

5.4 Relative frequency vs. probability

We have already distinguished between theoretical and experimental probability. Probability is one of the sections in mathematics that we can practically do with learners. Learners can relate to the topic as it is based on real-life situations and usually fun for learners. Here is an activity you can do with learners to help them differentiate between the two concepts of theoretical probability and experimental probability (*relative frequency*).

Activity 3: Frequency vs. probability

Work in pairs

You should use about 25 minutes for this activity.

1. In this experiment you will use a die.
   a) Calculate the probability to throw a 2 with a die. We call this theoretical probability.

b) Let us discover empirical, or experimental, probability. Take a die and throw it 30 times each. You and your peer can take turns. Count how many times the die lands on 2. Record it.

Person 1: how many times the die landed on 2: ……
Person 2: how many times the die landed on 2: ……

c) Each person can calculate the *relative* frequency of the *outcome*.

d) Compare your relative frequency with that of your peer.

e) Repeat the calculation for relative frequency, but this time take the sum total of the number of times the die landed on 2 (add person 1 and 2 together).

2. Compare the answers in a), c) and e). What do you notice?

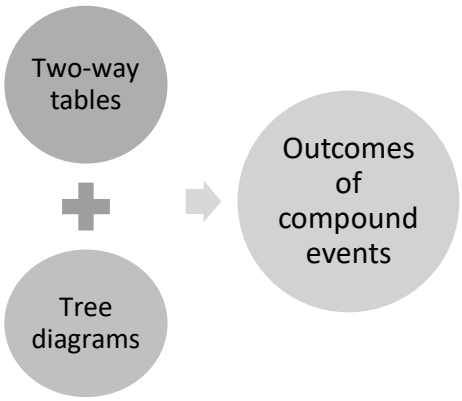
3. Discuss possible mistakes of learners in understanding the difference between theoretical and experimental probability.

The frequency of an outcome is the number of times the outcome occurs when an activity is performed again and again. In the example of the die, the number of times the die lands on 2 is called the frequency. The outcome is for the die to land on 2. The larger the number of trials in an experiment, the closer the relative frequency of an outcome will get to the theoretical probability of the outcome.

5.5 Compound events

What will be the effect if one die is red and another is green? If you throw the red die, you will get the following possible outcomes: {1; 2; 3; 4; 5; 6}. If you now throw the green die, you will get the same possible outcomes. We can call the throw of the red die a simple event. When you throw the green die, you have now created a compound event. How can we determine the total number of outcomes for the compound event (the red and green die being thrown together)?

The outcomes of a compound event can be represented in one of two ways:



Activity 4: Events

Work in groups

You should use about 25 minutes for this activity.

The following question is asked: Two coins are tossed. Use a tree diagram to determine the chance for both coins to land on heads.

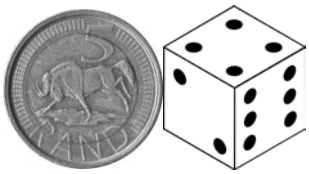
1. What prior knowledge do learners need to answer this question?

2. Draw the tree diagram to find the answer to the problem.

3. Write down possible mistakes of learners when working with tree diagrams and suggest prevention strategies.

Common mistakes	Possible prevention actions

4. A coin is tossed and die is thrown.



Below is an incomplete two-way table to determine the possible outcomes.

	1	2	3	4	5	6
H	H 1				H 5	
T			T 3			T 6

a) Where do you start your teaching of two-way tables?

b) In this question, the teacher starts the teaching of two-way tables by giving learners the opportunity to complete an incomplete two-way table. Do you think this is a good idea? Explain your answer.

c) Calculate the probability for the die to land on an even number and then for the coin to land on heads. How will you guide learners to answer this question?

5. Write down possible mistakes of learners when working with two-way tables and suggest prevention strategies.

Common mistakes	Possible prevention actions

6. Why do you think we must teach learners a topic like probability?

Two-way tables and tree diagrams make it easy for us to determine the possible outcomes of compound events. Practise easy compound events to help learners master the skill of working with two-way tables and tree diagrams.

5.6 Conclusion

Why must we teach our learners about probability? Firstly, there are several major industries that often make use of probability. The first of these is the gambling, sports and games industry. All gambling, sports and many other games rely heavily on the use of probabilities to predict winning combinations and the number of winners. Poker players know the odds of any particular hand winning over any other hand. Probabilities are used to determine how much and how often a slot machine pays out.

The second major industry using probabilities is the *insurance industry*. All insurance rates are determined by the probability that a certain person of a certain age doing a certain activity will have an accident or a heart attack or get diabetes or die at a certain age.

The third big user is *meteorology* and weather forecasting. Every time the weatherman says there is a 80% chance of snow, he has studied the weather patterns, applied the appropriate Mathematics and Science formulas and made the prediction.

A fourth area where probabilities are used, but which may not be widely known, is in the area of research. Research is done in almost every field of study from agriculture to business trends to education to *psychology* to medicine. The Mathematics will get harder, as will calculating the probability, so your learners will need to learn the fundamentals at an early age.

Probability is a topic that continues in both Mathematics FET and Mathematical Literacy. We must, therefore, ensure that our learners understand the concepts and terminology regarding probability. Make this topic practical and not only formula-driven to help learners understand the concepts better.

At the end of this unit I can guide my learners to			**Reason**
✓ use the vocabulary to define, calculate and list the possible outcomes of events; and			
✓ predict, compare, explain and calculate the relative frequency of events.			

Glossary

Chance	the possibility that something will happen
Frequency	the number of times the outcome shows when you repeat the activity numerous times
Insurance industry	a sector-following fund that invests primarily in insurance companies, so as to obtain investment results that closely track an underlying index of insurers
Meteorology	the branch of science concerned with the processes and phenomena of the atmosphere, especially as a means of forecasting the weather
Outcome	the actual result of a trial
Predict	to say or estimate that (a specified thing) will happen in the future
Probability	a numerical measure of the likelihood of an outcome
Psychology	the scientific study of the human mind and its functions, especially those affecting behaviour in a given context
Relative frequency	the relative frequency of an outcome is the number of times the event takes place divided by the number of the times the activity was performed.
Theoretical probability	when we use a theoretical formula to calculate probability; the calculation "on paper"

References

Groenewald, M., Otto, H., Roos, H. & Van der Westhuizen, G. (2012). *Mathematics Today Grade 8 Learner's Book.* Maskew Miller Longman: Cape Town.

Phillips, M.D., Basson, J. & Odendaal, J. (2014). *Mind Action Series Mathematics Textbook Grade 9*. Allcopy publishers: Sanlamhof.

Yule, D., Harris, L., Pretorius, J., Kriel, D. & Ebrahim, R. (2013). *Oxford Successful Mathematics Grade 8 Learner's Book*. Oxford University Press: Cape Town.

Module Conclusion

Data handling and probability are two topics of Mathematics that can be taught in a practical and constructive manner. Choose teaching strategies that are learner-centred, and that will allow learners to be actively involved in learning. These topics are excellent opportunities to accommodate all the learning styles namely visual, auditory and tactile/kinaesthetic.

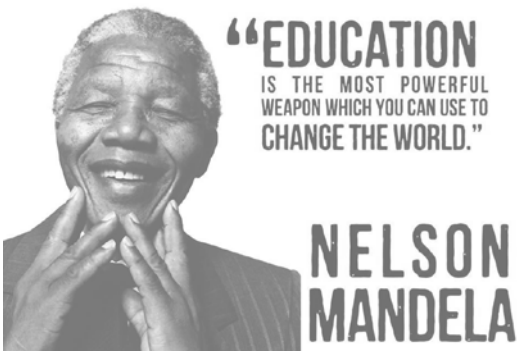
A big problem is that many teachers make these topics formula driven, meaning the focus is on procedural knowledge. You must focus on conceptual knowledge, meaning learners must understand the different mathematical concepts and the relationship between these concepts.

As agents of change, we want to encourage our learners to make meaning and understand mathematical concepts. We want learners to apply mathematical concepts, to reason and to solve problems in real-life contexts. We must help learners to answer higher cognitive level questions, and not simply stay at lower level cognitive questions.

In Mathematics in the Senior Phase, our responsibility is to build a solid foundation of the basic calculations that our learners will need in the FET Phase. We must develop the unique language associated with data handling and probability. This can be a challenge as we often have second and third English language speaking learners in our classes. Bring resources into the classroom to help learners make sense of the terminology.

Be creative and try out new teaching techniques and don't be afraid to make mistakes. You must make sure that learning takes place during each lesson, but as agents of change, we know that we want effective learning to take place. We will use creative and innovative ways to test for effective teaching and learning. We can use interesting and creative assessment ideas, such as quizzes. Data handling and probability are real-life applications of Mathematics and it can be fun and enjoyable to learners. Choose topics that learners can relate to in the specific age groups.

I want to conclude with these inspiring words by Nelson Mandela.



https://bit.ly/2VxOVMc

Assessment

Assessment 1

Collecting and summarising data are important skills that we practise in Mathematics.

1. Write three prior knowledge and skills required before doing data collection with your learners.
   - Give a reason why this is needed.
   - Indicate how you will assess learners if they have the necessary prior knowledge and skills.
2. Do a survey on any problem at school that you want to solve/investigate.
   - You must be able to collect two sets of numerical data. (See number 4.)
   - Give the purpose of your research. Clearly, state the problem at school.
   - How long has it been a problem at school?
   - What has been done in the past about the problem?
   - Is it a new problem?
3. Identify your sample/population to collect information to solve the problem.
   - Clearly, indicate which learners will be used to collect information.
   - Give a reason for your choice.
   - Name three challenges that you foresee in collecting the data?
   - Name prevention strategies to limit challenges.
   - What can you do to ensure an honest response from learners?
4. Design a questionnaire to collect numerical data (learners must give answers in number form).
   - You must collect only two sets of data from the learners.
   - Give the survey question on your questionnaire.
   - Choose a creative topic that learners will find interesting.
   - Your questionnaire must be in the form of a fill-in written form that learners must complete.
   - Submit a copy of the questionnaire, and ten learners' responses.
5. Use the ten learners' two data sets and do the following calculations:
   - Calculate the measures of central tendency and dispersion for each set.
   - In each case, explain the meaning of each measure.
   - Indicate which measures can be influenced by the influence of outliers.
6. Give one of the two data sets (of the ten learners you submitted) to your learners. Let them calculate the mean, median, mode and range of each.
   - Mark their responses using your calculations in number 5.
   - Submit ten learners' answers.
   - Reflect on their performance. Give reasons for performance.
   - Indicate how you can improve their performance in data handling.

What to submit:

- Cover page
- Prior-knowledge
- Survey
- Definitions
- Questionnaire
- Calculations
- Marked learners' scripts
- Rubric

Assessment 1: Planning

SURNAME:													INITIALS:				
										D	D	M	M	Y	Y	Y	Y
STUDENT NUMBER										DATE							

1. State two prior knowledge skills learners need to collect and summarise data.

2. State a problem at school to organise a survey for.

3. Write one fill-in question for the questionnaire.

Assessment 1: Rubric

Criteria	0	1	2	3	4	Mark
• **Prior knowledge/skills (3)** • **Correct** • **Relevant** • **Appropriate reason**	None	1 criterion	2 criteria	3 criteria	4 criteria	
• **Assessing clear** • **Correct** • **Appropriate** • **Detailed**	None	1 criterion	2 criteria	3 criteria	4 criteria	
Problem/Purpose stated	None	Vague; no actions	Vague, previous actions	Clear, no previous actions	Clear, previous actions	
Sample or population: • **Clearly stated** • **Correct reason**	None	1 criterion	Both criteria			
Three challenges	None	1 relevant	2 relevant	3 relevant		
• **Prevention clear** • **Relevant to challenges** • **Practical**	None	1 criterion	2 criteria	3 criteria		
Questionnaire: • **Fill-in** • **2 data sets** • **Correct survey question** • **Numerical data**	None	1 criterion	2 criteria	3 criteria	4 criteria	
Ten learners' response	None	Incomplete	Completed			
• **Mean: formula correct** • **Calculations correct** • **Correct definition** • **Effect of outlier**	None	1 criterion	2 criteria	3 criteria	4 criteria	
• **Median: formula correct** • **Calculations correct** • **Correct definition** • **Effect of outlier**	None	1 criterion	2 criteria	3 criteria	4 criteria	
• **Mode: formula correct** • **Calculations correct** • **Correct definition** • **Effect of outlier**	None	1 criterion	2 criteria	3 criteria	4 criteria	
• **Range: formula correct** • **Calculations correct** • **Correct definition** • **Effect of outlier**	None	1 criterion	2 criteria	3 criteria	4 criteria	
Learners' work marked (10)	None	Not all	All			
• **Reflection clear** • **Correct** • **Relevant**	None	1 criterion	2 criteria	3 criteria		
• **Improvement clear** • **Correct** • **Relevant**	None	1 criterion	2 criteria	3 criteria		
					Total	**/50**

COVER PAGE FOR ASSESSMENT
ADVANCED CERTIFICATE IN TEACHING

MODULE CODE:	**EMBC2521**								
ASSESSMENT NAME:	**ASSESSMENT 1**								
SURNAME:			**INITIALS:**						
PHONE NUMBER:									
EMAIL ADDRESS:									
DUE DATE:									
CENTRE:		**FACILITATOR:**							
SIGNATURE									

										D	**D**	**M**	**M**	**Y**	**Y**	**Y**	**Y**
STUDENT NUMBER										**DATE**							

DECLARATION BY STUDENT

I have read the University's Policy on the Prevention of Plagiarism and Dealing with Academic Writing Misconduct: http://www.ufs.ac.za/docs/default-source/all-documents/plagiarism-prevention-policy-364-eng.pdf?sfvrsn=0

I understand that I must:

1. abide by all the directives of this Policy, the Assessment Policy, course guides, specific and general regulations and assessment requirements;
2. seek assistance if I am unsure about appropriate citation and referencing techniques;
3. accept responsibility for having full knowledge of the Policy;
4. submit only my own work for any form of assessment, except where:
   the work of others is appropriately acknowledged; and
   the assessor/moderator has required, or given prior permission for, group or collaborative work to be submitted;
5. refrain from intentionally or negligently deceiving the reader by preventing my own work from being copied by another student, who may or may not have an intentional or negligent aim to deceive the reader;
6. be aware that according to the Policy, measures for all Level Four violations and repeated Levels One, Two and Three violations are reported and investigated in accordance with the *UFS Statute* on Student Discipline (see Chapter XIII, pp. 50 – 58); and
7. include with my assignment an electronic report of the software programme if required/where applicable.

COMMENTS FROM ASSESSOR:	
MARK:	**SIGNATURE:**

Assessment 2

1. Develop a learning activity on data graphs.
   - List five mathematical terminology and skills needed as prior knowledge.
   - Give two reasons why pre-knowledge is necessary.

2. Make a list of resources needed for the activity.
   - Explain the need for each of the resources.

3. Compile an interesting introduction to the activity.
   - Make four PowerPoint slides to use.
   - It must be creative and stimulate the learner's interest.
   - Include pictures and written text.

4. Design a lesson activity to represent data on two different types of graphs.
   a) Collection of the data.
      - Choose a survey topic for you to use.
      - Collect data from your learners.
      - Differentiate between boys and girls when you collect data.
      - Indicate how you collected the data. Explain why you chose that specific method.
   b) Representing the data.
      - Draw two different types of graphs to compare the data between boys and girls.
      - Ensure all labels and headings are indicated.

5. Write a report on your findings of the graphs.
   - Compare boys and girls.
   - Give possible reasons for differences.

6. Compile a homework activity for your learners.
   - Supply learners with data.
   - Learners must draw two different graphs.
   - Learners must interpret the data.
   - Supply a memorandum for the activity.
   - Submit five completed, marked learner homework activities.

7. Reflect on your lesson.
   - What worked well and what needs improvement?
   - How did the learners react to the lesson?

What to submit:

- Cover page
- Frequency table with data
- Homework activity
- Reflection
- Prior-knowledge and resources
- Two graphs of data
- Memorandum
- Rubric
- PowerPoint slides
- Report on findings
- Learners' copies

Assessment 2: Planning

SURNAME:													INITIALS:				
										D	D	M	M	Y	Y	Y	Y
STUDENT NUMBER										DATE							

Answer the following questions to prepare for the assessment.

1. List two terminology and skills required for the activity on data graphs.	
2. List two resources you will use for the activity.	
3. State one way of collecting data.	

Assessment 2: Rubric

Criteria	0	1	2	3	4	Mark
Topic	None	Relevant				
Maths terminology	None	2 correct	3 correct	4 correct	5 correct	
Skills	None	Indicated; improve it	Correct & relevant			
Prior-knowledge: Reasons	None	1 incorrect	1 correct; relevant	2; only 1 relevant	2 correct; relevant	
Resources & explanation	None	Incomplete not applicable	Complete & applicable			
PowerPoint slides	None	1 slide	2 slides	3 slides	4 slides	
Introduction	None	Not relevant; no pictures	Incomplete; no pictures	Text relevant; some pictures	Relevant & interesting; with pictures	
Topic for survey	None	Relevant				
Data population: • **Boys & Girls**	None	Any 1	Both			
Data collection method	None	Incorrect for learners	Vague; needs improvement	Correct idea; suitable; reliable	Correct; suitable; reliable; valid	
Reason for method	None	Not clear	Detailed			
Graphs 1&2: • **2 types** • **Labels** • **Headings** • **Boys & girls**	None	1 criteria	Any 2 criteria	Any 3 criteria	All criteria	
Quality of graphs 1 & 2	None	Not clear	Inaccuracies	Few inaccuracies	Accurate; represent real data	
Report: • **Based on graphs** • **Compares boys & girls** • **Clear explanation**	None	1 criteria	Any 2 criteria	Any 3 criteria		
Homework & memo	None	1 graph **or** memo incomplete	Data with 2 graphs; memo correct			
Marked scripts	None	2	3	4	5	
Reflection: • **What worked** • **Improvement** • **Learner's reaction**	None	1 criterion correct; relevant	2 criteria correct; relevant	All criteria; correct; relevant		
					Total	**/50**

COVER PAGE FOR ASSESSMENT
ADVANCED CERTIFICATE IN TEACHING

MODULE CODE:	**EMBC2521**										
ASSESSMENT NAME	**ASSESSMENT 2**										
SURNAME:							**INITIALS:**				
PHONE NUMBER:											
EMAIL ADDRESS:											
DUE DATE:											
CENTRE:							**FACILITATOR:**				
SIGNATURE											
							D	**D**	**M**	**M**	**Y**	**Y**	**Y**	**Y**
STUDENT NUMBER							**DATE**							

DECLARATION BY STUDENT

I have read the University's Policy on the Prevention of Plagiarism and Dealing with Academic Writing Misconduct: http://www.ufs.ac.za/docs/default-source/all-documents/plagiarism-prevention-policy-364-eng.pdf?sfvrsn=0

I understand that I must:

1. abide by all the directives of this Policy, the Assessment Policy, course guides, specific and general regulations and assessment requirements;
2. seek assistance if I am unsure about appropriate citation and referencing techniques;
3. accept responsibility for having full knowledge of the Policy;
4. submit only my own work for any form of assessment, except where:
   the work of others is appropriately acknowledged; and
   the assessor/moderator has required, or given prior permission for, group or collaborative work to be submitted;
5. refrain from intentionally or negligently deceiving the reader by preventing my own work from being copied by another student, who may or may not have an intentional or negligent aim to deceive the reader;
6. be aware that according to the Policy, measures for all Level Four violations and repeated Levels One, Two and Three violations are reported and investigated in accordance with the *UFS Statute* on Student Discipline (see Chapter XIII, pp 50 – 58); and
7. include with my assignment an electronic report of the software programme if required/where applicable.

COMMENTS FROM ASSESSOR:	
MARK:	**SIGNATURE:**

Resources

Appendix A: Collect and organise data

Anon. (2015). *Data.* Viewed on 6 June 2017. Available from:
https://bit.ly/1yU85v7

Appendix B: Misleading graphs

Anon. (2013). Common *misconceptions*. Viewed on 6 June 2017. Available from:
https://bit.ly/2VnsFEI

Appendix C: Summarise data

Lund, A. (2013a). *Measures of central tendency.* Viewed on 18 May 2017. Available from:
https://bit.ly/1zX41bK

Lund, A. (2013b). *Measures of Spread.* Viewed on 18 May 2017. Available from:
https://bit.ly/2Sf53zQ

Appendix D: Probability

Anon. (2014) *Probability.* Viewed on 7 June 2017. Available from: https://bit.ly/1yU85v7

Appendices

Appendix A: Collect and organise data

Anon. (2015). *Data.* Viewed on 6 June 2017. Available from: https://bit.ly/1yU85v7

What is Data?

Data is a collection of facts, such as numbers, words, measurements, observations or even just descriptions of things.

Qualitative vs Quantitative

Data can be qualitative or quantitative.

- **Qualitative data** is descriptive information (it *describes* something)
- **Quantitative data**, is numerical information (numbers).

And **Quantitative data** can also be Discrete or Continuous:

- **Discrete data** can only take certain values (like whole numbers)
- **Continuous data** can take any value (within a range)

Put simply: **Discrete data** is counted, **Continuous data** is measured.

More Examples

Qualitative:

- Your friends' favorite holiday destination
- The most common given names in your town
- How people describe the smell of a new perfume

Quantitative:

- Height (Continuous)
- Weight (Continuous)
- Petals on a flower (Discrete)
- Customers in a shop (Discrete)

Types of questions

A survey question can be:

- **Open-ended** (the person can answer in any way they want), or
- **Closed-ended** (the person chooses from one of several options)

Closed ended questions are much easier to total up later on, but may stop people giving an answer they really want.

Census or Sample

A **Census** is when we collect data for **every** member of the group (the whole "population").
A **Sample** is when we collect data just for **selected members** of the group.

- Example: there are 120 people in your local football club.
- You can ask everyone (all 120) what their age is. That is a census.
- Or you could just choose the people that are there this afternoon. That is a sample

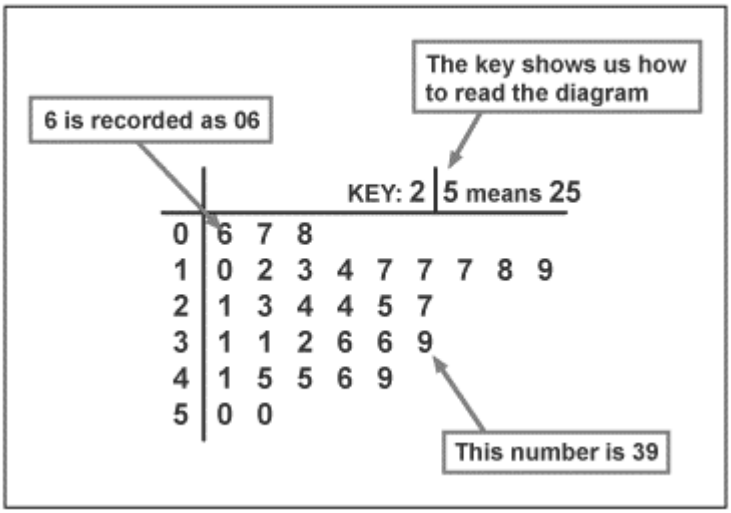
A census is accurate, but hard to do. A sample is not as accurate, but may be good enough, and is a lot easier.

When you are **Sampling** you should be careful who you ask.

To be a good sample, each person should be chosen **randomly.**
If you only ask people who look friendly, you will only know what friendly people think!
If you go to the swimming pool and ask people "Can you swim?" you will get a biased answer ... maybe even 100% will say "Yes".
So be careful not to bias your survey. Try to choose randomly.

Stem-and-leaf display

A **stem-and-leaf display or diagram** is a table used to display data. The '**stem**' is on the left displays the first digit or digits. The '**leaf**' is on the right and displays the last digit. For **example**, 543 and 548 can be displayed together on a **stem and leaf** as 54. Below is an example.

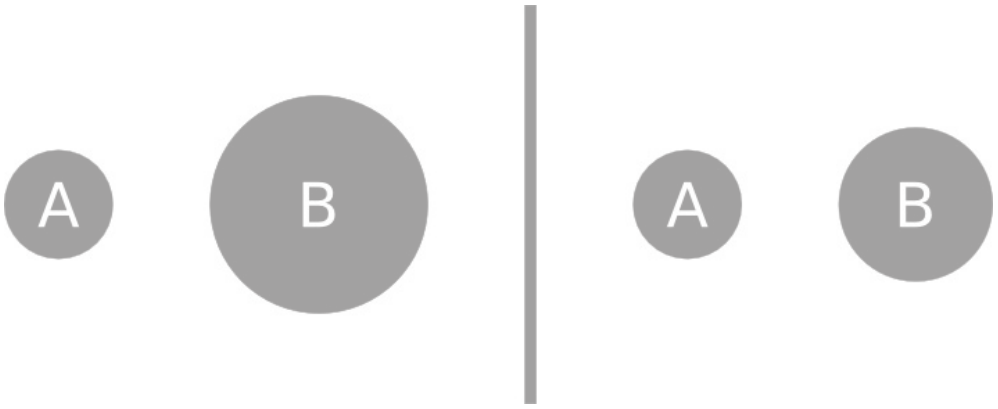


http://www.bbc.co.uk/bitesize/ks3/maths/images/stem_leaf1.gif

Appendix B: Misleading graphs

Anon. (2013). Common *misconceptions*. Viewed on 6 June 2017. Available from: https://bit.ly/2VnsFEI

The simple "reflect the size" rule becomes even more difficult in 2 dimensions, when you have to worry about the total area. At one point, news outlets started to replace columns with pictures, and then continue to scale the dimensions of pictures up in the old way. The problem? If you adjust the height to reflect the change and the width automatically increases with it, the area increases even more and will become completely wrong! Confused? Look at these bubbles:

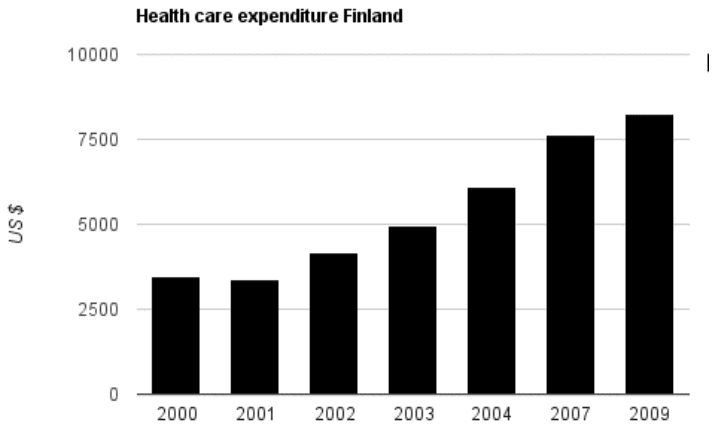


Task: We want to show that B is double the size of A. Which representation is correct? Why?
Answer: The diagram on the right.

Remember the formula for calculating the area of a circle? (Area = πr^2 If this doesn't look familiar, see here). In the left hand diagram, the radius of A (r) was doubled. This means that the total area goes up by a scale factor of four! This is wrong. If B is to represent a number twice the size of A, we need the area of B to be double the area of A. To correctly calculate this, we need to adjust the length of the radius by $\sqrt{2}$. This gives us a realistic change in size.

Time will tell?

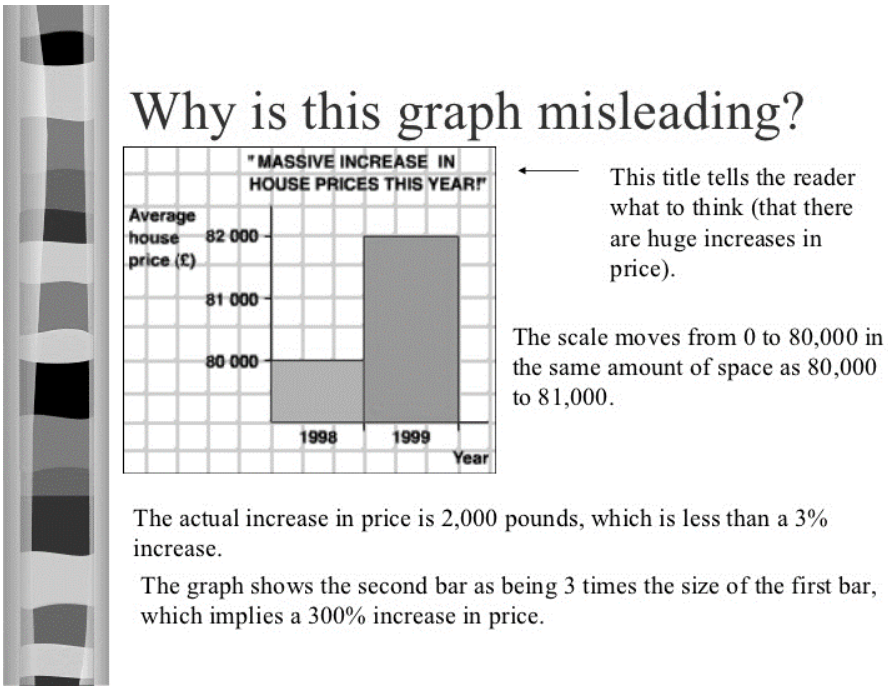
Time lines are also critical when displaying data. Look at the chart below:

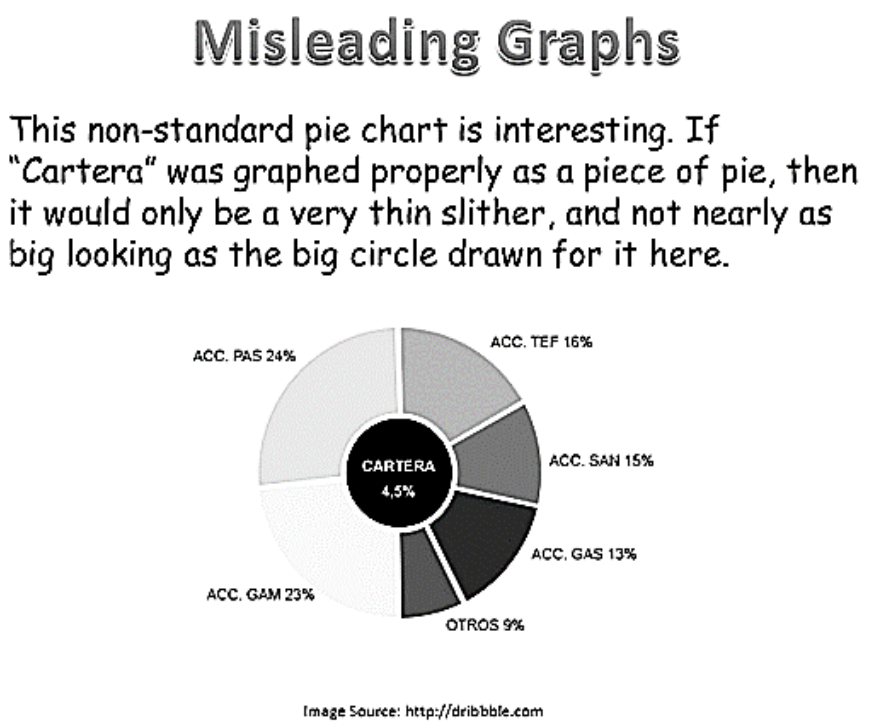


A clear stable increase in health care costs since 2002? Not quite. Notice how before 2004, there are 1 year steps. After, there is a gap between 2004 and 2007, and 2007 and 2009. This presentation makes us believe that healthcare expenditure increases continuously at the same rate since 2002 – but actually it doesn't. So if you deal with time lines: make sure that the spacing between the data points are correct! Only then will you be able to see the trends correctly.

Context, context, context

One thing incredibly important for data is context: A number or quality doesn't mean a thing if you don't give context. So explain what you are showing – explain how it is read, explain where the data comes from and explain what you did with it. If you give the proper context the conclusion should come right out of the data.





http://www.passyworld.com/passyImagesThirteen/MissNine540x431JPG.jpg

Appendix C: Summarise data

Lund, A. (2013a). *Measures of central tendency.* Viewed on 18 May 2017. Available from: https://bit.ly/1zX41bK
Lund, A. (2013b). *Measures of Spread.* Viewed on 18 May 2017. Available from: https://bit.ly/2Sf53zQ Leard Statistics. Lund Research Ltd. Bath, United Kingdom

Measures of Central Tendency

Introduction

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

Mean (Arithmetic)

The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have n values in a data set and they have values x_1, x_2, ..., x_n, the sample mean, usually denoted by $\bar{x}$ (pronounced x bar), is:

$$\bar{x} = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$$

This formula is usually written in a slightly different manner using the Greek capitol letter, $\sum$, pronounced "sigma", which means "sum of...":

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

You may have noticed that the above formula refers to the sample mean. So, why have we called it a sample mean? This is because, in statistics, samples and populations have very different meanings and these differences are very important, even if, in the case of the mean, they are calculated in the same way. To acknowledge that we are calculating the population mean and not the sample mean, we use the Greek lower case letter "mu", denoted as μ:

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

The mean is essentially a model of your data set. It is the value that is most common. You will notice, however, that the mean is not often one of the actual values that you have observed in your data set. However, one of its important properties is that it minimises error in the prediction of any one value in your data set. That is, it is the value that produces the lowest amount of error from all other values in the data set.

An important property of the mean is that it includes every value in your data set as part of the calculation. In addition, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero.

When not to use the mean

The mean has one main disadvantage: it is particularly susceptible to the influence of outliers. These are values that are unusual compared to the rest of the data set by being especially small or large in numerical value. For example, consider the wages of staff at a factory below:

Staff	1	2	3	4	5	6	7	8	9	10
Salary	15k	18k	16k	14k	15k	15k	12k	17k	90k	95k

The mean salary for these ten staff is $30.7k. However, inspecting the raw data suggests that this mean value might not be the best way to accurately reflect the typical salary of a worker, as most workers have salaries in the $12k to 18k range. The mean is being skewed by the two large salaries. Therefore, in this situation, we would like to have a better measure of central tendency. As we will find out later, taking the median would be a better measure of central tendency in this situation.

Another time when we usually prefer the median over the mean (or mode) is when our data is skewed (i.e., the frequency distribution for our data is skewed). If we consider the normal distribution - as this is the most frequently assessed in statistics - when the data is perfectly normal, the mean, median and mode are identical. Moreover, they all represent the most typical value in the data set. However, as the data becomes skewed the mean loses its ability to provide the best central location for the data because the skewed data is dragging it away from the typical value. However, the median best retains this position and is not as strongly influenced by the skewed values. This is explained in more detail in the skewed distribution section later in this guide

Median

The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data. In order to calculate the median, suppose we have the data below:

65	55	89	56	35	14	56	55	87	45	92

We first need to rearrange that data into order of magnitude (smallest first):

14	35	45	55	55	**56**	56	65	87	89	92

Our median mark is the middle mark - in this case, 56 (highlighted in bold). It is the middle mark because there are 5 scores before it and 5 scores after it. This works fine when you have an odd number of scores, but what happens when you have an even number of scores? What if you had only 10 scores? Well, you simply have to take the middle two scores and average the result. So, if we look at the example below:

65	55	89	56	35	14	56	55	87	45

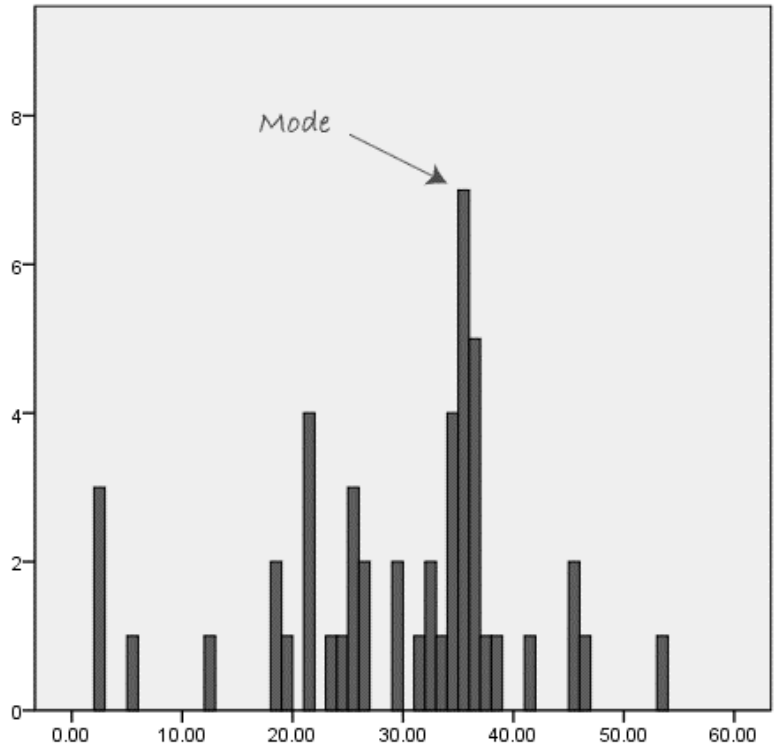
We again rearrange that data into order of magnitude (smallest first):

14	35	45	55	**55**	**56**	56	65	87	89

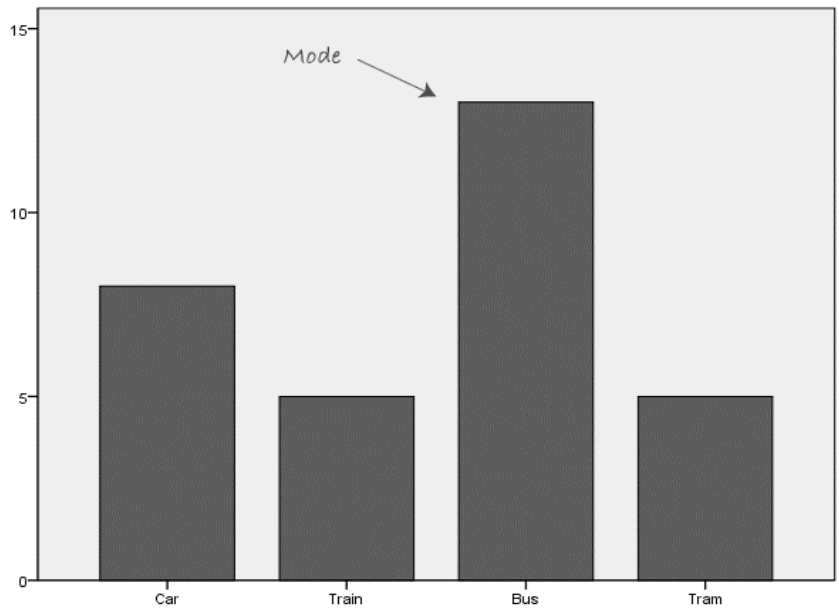
Only now we have to take the 5th and 6th score in our data set and average them to get a median of 55.5.

Mode

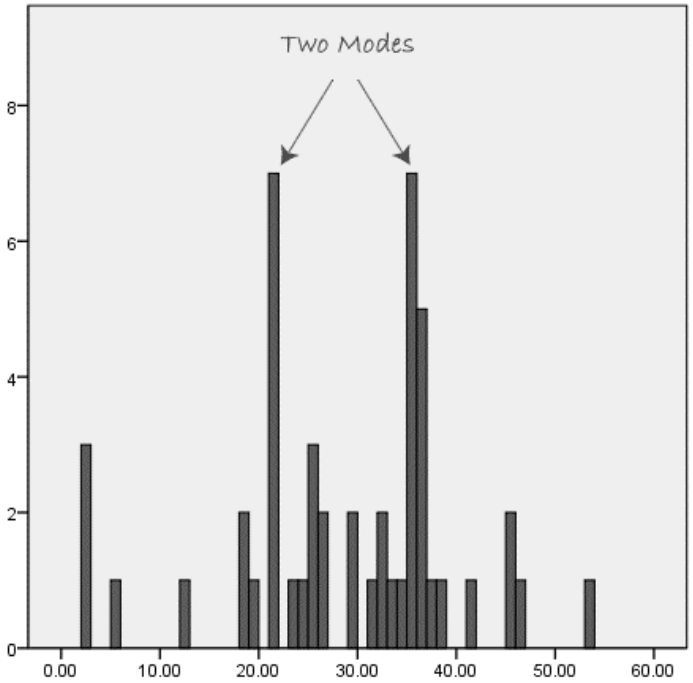
The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:



Normally, the mode is used for categorical data where we wish to know which the most common category, as illustrated is below is:

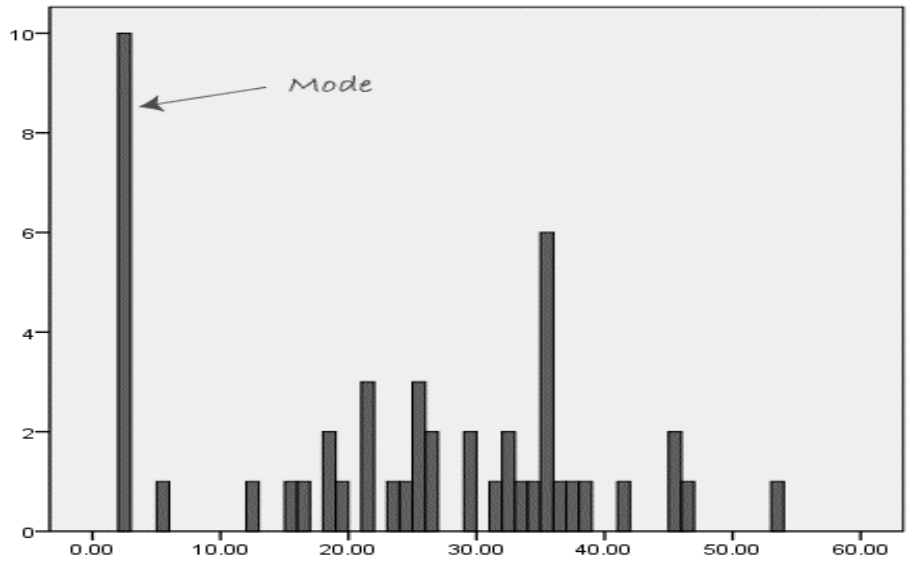


We can see above that the most common form of transport, in this particular data set, is the bus. However, one of the problems with the mode is that it is not unique, so it leaves us with problems when we have two or more values that share the highest frequency, such as below:



We are now stuck as to which mode best describes the central tendency of the data. This is particularly problematic when we have continuous data because we are more likely not to have any one value that is more frequent than the other. For example, consider measuring 30 peoples' weight (to the nearest 0.1 kg). How likely is it that we will find two or more people with **exactly** the same weight (e.g., 67.4 kg)? The answer, is probably very unlikely - many people might be close, but with such a small sample (30 people) and a large range of possible weights, you are unlikely to find two people with exactly the same weight; that is, to the nearest 0.1 kg. This is why the mode is very rarely used with continuous data.

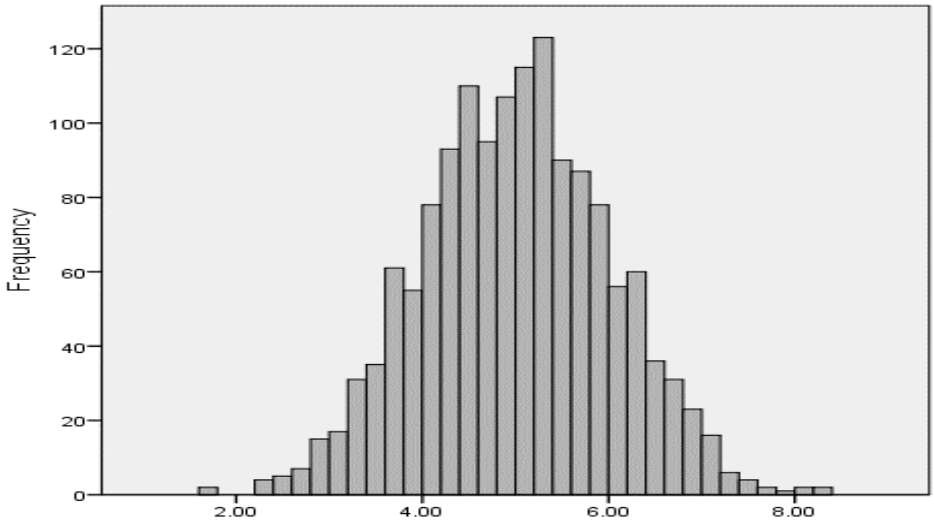
Another problem with the mode is that it will not provide us with a very good measure of central tendency when the most common mark is far away from the rest of the data in the data set, as depicted in the diagram below:



In the above diagram the mode has a value of 2. We can clearly see, however, that the mode is not representative of the data, which is mostly concentrated around the 20 to 30 value range. To use the mode to describe the central tendency of this data set would be misleading.

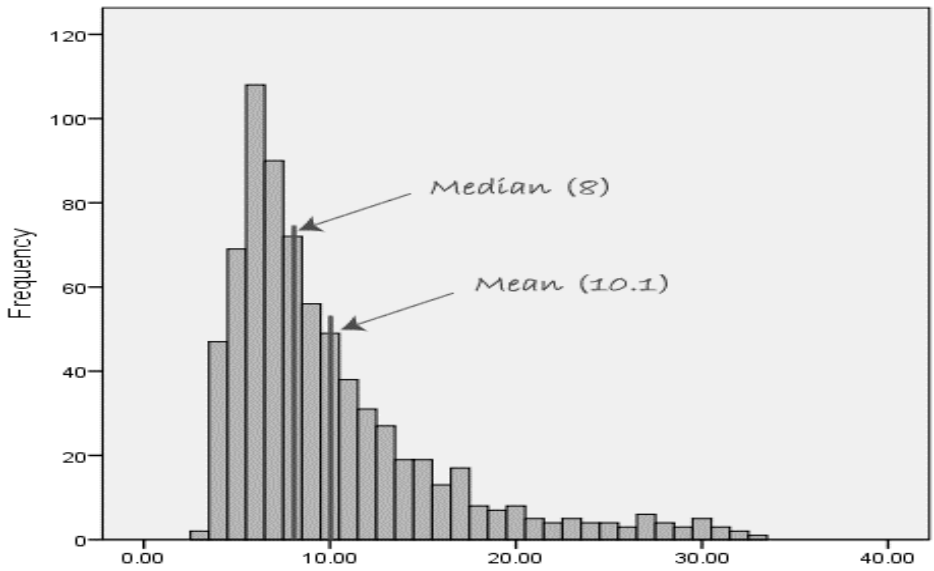
Skewed Distributions and the Mean and Median

We often test whether our data is normally distributed because this is a common assumption underlying many statistical tests. An example of a normally distributed set of data is presented below:



When you have a normally distributed sample you can legitimately use both the mean or the median as your measure of central tendency. In fact, in any symmetrical distribution the mean, median and mode are equal. However, in this situation, the mean is widely preferred as the best measure of central tendency because it is the measure that includes all the values in the data set for its calculation, and any change in any of the scores will affect the value of the mean. This is not the case with the median or mode.

However, when our data is skewed, for example, as with the right-skewed data set below:



We find that the mean is being dragged in the direct of the skew. In these situations, the median is generally considered to be the best representative of the central location of the data. The more skewed the distribution, the greater the difference between the median and mean, and the greater emphasis should be placed on using the median as opposed to the mean. A classic example of the

above right-skewed distribution is income (salary), where higher-earners provide a false representation of the typical income if expressed as a mean and not a median.

If dealing with a normal distribution, and tests of normality show that the data is non-normal, it is customary to use the median instead of the mean. However, this is more a rule of thumb than a strict guideline. Sometimes, researchers wish to report the mean of a skewed distribution if the median and mean are not appreciably different (a subjective assessment), and if it allows easier comparisons to previous research to be made.

Summary of when to use the mean, median and mode

Please use the following summary table to know what the best measure of central tendency is with respect to the different types of variable.

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

Measures of Spread

Introduction

A measure of spread, sometimes also called a measure of dispersion, is used to describe the variability in a sample or population. It is usually used in conjunction with a measure of central tendency, such as the mean or median, to provide an overall description of a set of data.

Why is it important to measure the spread of data?

There are many reasons why the measure of the spread of data values is important, but one of the main reasons regards its relationship with measures of central tendency. A measure of spread gives us an idea of how well the mean, for example, represents the data. If the spread of values in the data set is large, the mean is not as representative of the data as if the spread of data is small. This is because a large spread indicates that there are probably large differences between individual scores. Additionally, in research, it is often seen as positive if there is little variation in each data group as it indicates that the similar.

Range

The range is the difference between the highest and lowest scores in a data set and is the simplest measure of spread. So we calculate range as:

Range = maximum value - minimum value

For example, let us consider the following data set:

23 56 45 65 59 55 62 54 85 25

The maximum value is 85 and the minimum value is 23. This results in a range of 62, which is 85 minus 23. Whilst using the range as a measure of spread is limited, it does set the boundaries of the scores. This can be useful if you are measuring a variable that has either a critical low or high threshold (or both) that should not be crossed. The range will instantly inform you whether at least one value broke these critical thresholds. In addition, the range can be used to detect any errors when entering data. For example, if you have recorded the age of school children in your study and your range is 7 to 123 years old you know you have made a mistake!

Appendix D: Probability

Anon. (2014) *Probability.* Viewed on 7 June 2017. Available from: https://bit.ly/1yU85v7

Probability

How likely something is to happen.

Many events can't be predicted with total certainty. The best we can say is how likely they are to happen, using the idea of probability.

Tossing a Coin

When a coin is tossed, there are two possible outcomes:

- heads (H) or
- tails (T)

We say that the probability of the coin landing H is ½
And the probability of the coin landing T is ½

Probability

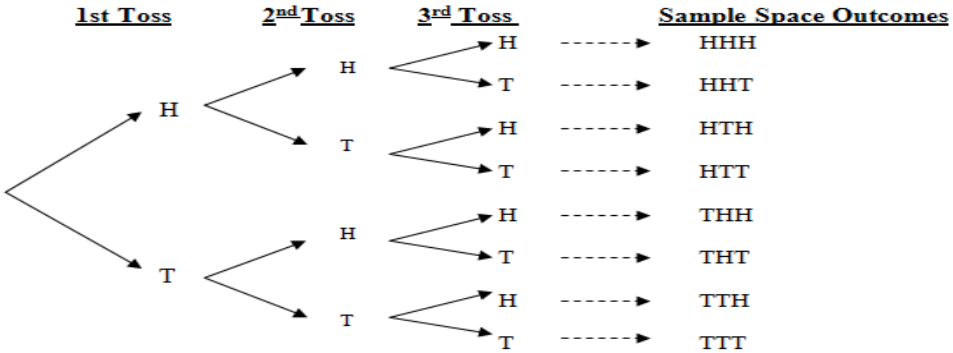
In general:

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

Tree Diagrams

Calculating probabilities can be hard, sometimes we add them, sometimes we multiply them, and often it is hard to figure out what to do ... tree diagrams to the rescue!

Here is an example of a tree diagram to determine possible outcomes when a coin is tossed for three times:



Throwing Dice

When a single die is thrown, there are six possible outcomes: 1, 2, 3, 4, 5, 6.
The probability of any one of them is $\frac{1}{6}$.

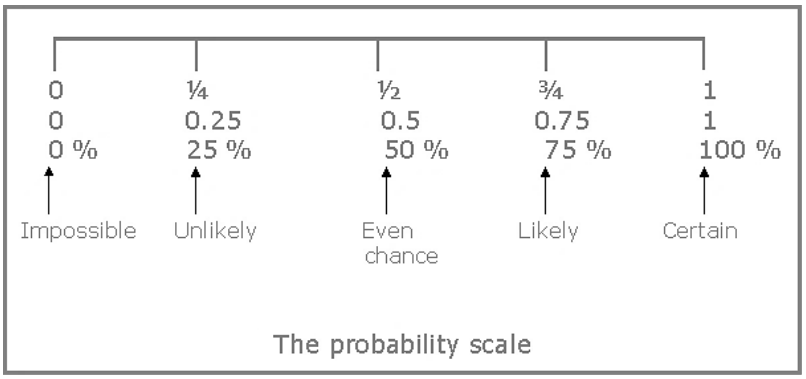
Example: the chances of rolling a "4" with a die

Number of ways it can happen: 1 (there is only 1 face with a "4" on it)
Total number of outcomes: So the probability = $\frac{1}{6}$

Example: there are 5 marbles in a bag: 4 are blue, and 1 is red. What is the probability that a blue marble gets picked?

Number of ways it can happen: 4 (there are 4 blues)
Total number of outcomes: 5 (there are 5 marbles in total)
So the probability = $\frac{4}{5}$ = 0.8

We can represent probability on the probability scale:



Probability is always a value between 0 and 1.

Probability is Just a Guide. Probability does not tell us exactly what will happen, it is just a guide.
Example: toss a coin 100 times, how many Heads will come up?

Probability says that heads have a ½ chance, so we can expect 50 Heads.
But when we actually try it we might get 48 heads, or 55 heads ... or anything really, but in most cases it will be a number near 50.

Some words have special meaning in Probability:
Experiment or Trial: an action where the result is uncertain.
Tossing a coin, throwing dice, seeing what pizza people choose are all examples of experiments.
Sample Space: all the possible outcomes of an experiment

Example: choosing a card from a deck
There are 52 cards in a deck (not including Jokers)
So the Sample Space is all 52 possible cards: {Ace of Hearts, 2 of Hearts, etc... }

The Sample Space is made up of Sample Points:
Sample Point: just one of the possible outcomes

Two-way tables

Two–way tables are used to determine the possible outcomes. In the example below a coin is tossed. A coin can land on H (head) or T (tail). A die is thrown and can land on 1, 2, 3 ,4, 5 or 6. When the coin and die are thrown simultaneously, we can use the two-way table to find the different possible outcomes.

		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

http://www.utas.edu.au/__data/assets/image/0004/622732/m6.1.png

EvaSys	**ACT Student Feedback**	Electric Paper EVALUATIONSSYSTEME
South Campus	EMBC2521	UFS UV

Mark as shown: ☐ ☒ ☐ ☐ ☐ Please use a ball-point pen or a thin felt tip. This form will be processed automatically.

Correction: ☐ ■ ☐ ☒ ☐ Please follow the examples shown on the left hand side to help optimize the reading results.

1. Instructions

The UFS is committed to assist and support you during your learning journey by providing a quality service and learning experience. We would like to invite you to share some feedback on your learning experience. This survey is confidential and no identifying information is asked. You can also complete the survey using the link sent to you by SMS. When answering the questions make sure you indicate your selection with an X. If you made a mistake, colour in the block the mistake was made in, and indicate your new choice with an X.

2. Study guide and content

2.1	The study guide is user-friendly.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
2.2	Information and instructions in the study guide are helpful.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
2.3	I understood the purpose and outcomes of this module clearly.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
2.4	The study guide is professional and of good quality.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly

2.5 If you could change anything relating to the content of the module, what would it be?

3. My learning experience

3.1	The study guide's design and how it is presented helped me reach the outcomes.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
3.2	I was able to do everything that was expected of me in this module in the allocated time.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
3.3	The reading material in the study guide supported my learning in this module.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
3.4	I can apply what I have learnt in my teaching.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
3.5	This module contributed to my development as an agent of change.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly

3.6 Describe your learning experience in this module.

4. Continuous assessment

4.1	Instructions for assessments were clear.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.2	Assessment tasks included the assessment criteria.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.3	I understood the assessment criteria provided.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.4	I used the assessment criteria to complete the assessments.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.5	Assessment tasks covered the outcomes of the module.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.6	I had enough time to complete my assessments.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
4.7	I received feedback on my assessments.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly

4.8 How did you experience the assessment in the module? What challenges did you experience? How did you overcome these?

5. The facilitator

5.1	The facilitator was well-prepared for sessions.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
5.2	I felt motivated to learn because of the facilitator.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly
5.3	The facilitator treated all students fairly.	☐ Not at all ☐ Always	☐ Sometimes	☐ Mostly

5.4 Describe your experience with the facilitator.