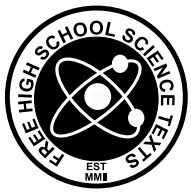


FHSST Authors

The Free High School Science Texts: Textbooks for High School Students Studying the Sciences Mathematics Grades 10 - 12

Version 0
September 17, 2008

Did you notice the FREEDOMS we've granted you?

Our copyright license is **different!** It grants freedoms rather than just imposing restrictions like all those other textbooks you probably own or use.

- We know people copy textbooks illegally but we would LOVE it if you copied our's - go ahead copy to your hearts content, **legally!**
- Publishers revenue is generated by controlling the market, we don't want any money, go ahead, distribute our books far and wide - we DARE you!
- Ever wanted to change your textbook? Of course you have! Go ahead change ours, make your own version, get your friends together, rip it apart and put it back together the way you like it. That's what we really want!
- Copy, modify, adapt, enhance, share, critique, adore, and contextualise. Do it all, do it with your colleagues, your friends or alone but get involved! Together we can overcome the challenges our complex and diverse country presents.
- So what is the catch? The only thing you can't do is take this book, make a few changes and then tell others that they can't do the same with your changes. It's share and share-alike and we know you'll agree that is only fair.
- These books were written by volunteers who want to help support education, who want the facts to be freely available for teachers to copy, adapt and re-use. Thousands of hours went into making them and they are a gift to everyone in the education community.

FHSST Core Team

Mark Horner ; Samuel Halliday ; Sarah Blyth ; Rory Adams ; Spencer Wheaton

FHSST Editors

Jaynie Padayachee ; Joanne Boulle ; Diana Mulcahy ; Annette Nell ; René Toerien ; Donovan Whitfield

FHSST Contributors

Rory Adams ; Prashant Arora ; Richard Baxter ; Dr. Sarah Blyth ; Sebastian Bodenstein ; Graeme Broster ; Richard Case ; Brett Cocks ; Tim Crombie ; Dr. Anne Dabrowski ; Laura Daniels ; Sean Dobbs ; Fernando Durrell ; Dr. Dan Dwyer ; Frans van Eeden ; Giovanni Franzoni ; Ingrid von Glehn ; Tamara von Glehn ; Lindsay Glesener ; Dr. Vanessa Godfrey ; Dr. Johan Gonzalez ; Hemant Gopal ; Umeshree Govender ; Heather Gray ; Lynn Greeff ; Dr. Tom Gutierrez ; Brooke Haag ; Kate Hadley ; Dr. Sam Halliday ; Asheena Hanuman ; Neil Hart ; Nicholas Hatcher ; Dr. Mark Horner ; Mfandaidza Hove ; Robert Hovden ; Jennifer Hsieh ; Clare Johnson ; Luke Jordan ; Tana Joseph ; Dr. Jennifer Klay ; Lara Kruger ; Sihle Kubheka ; Andrew Kubik ; Dr. Marco van Leeuwen ; Dr. Anton Machacek ; Dr. Komal Maheshwari ; Kosma von Maltitz ; Nicole Masureik ; John Mathew ; JoEllen McBride ; Nikolai Meures ; Riana Meyer ; Jenny Miller ; Abdul Mirza ; Asogan Moodaly ; Jothi Moodley ; Nolene Naidu ; Tyrone Negus ; Thomas O'Donnell ; Dr. Markus Oldenburg ; Dr. Jaynie Padayachee ; Nicolette Pekeur ; Sirika Pillay ; Jacques Plaut ; Andrea Prinsloo ; Joseph Raimondo ; Sanya Rajani ; Prof. Sergey Rakityansky ; Alastair Ramlakan ; Razvan Remsing ; Max Richter ; Sean Riddle ; Evan Robinson ; Dr. Andrew Rose ; Bianca Ruddy ; Katie Russell ; Duncan Scott ; Helen Seals ; Ian Sherratt ; Roger Sieloff ; Bradley Smith ; Greg Solomon ; Mike Stringer ; Shen Tian ; Robert Torregrosa ; Jimmy Tseng ; Helen Waugh ; Dr. Dawn Webber ; Michelle Wen ; Dr. Alexander Wetzler ; Dr. Spencer Wheaton ; Vivian White ; Dr. Gerald Wigger ; Harry Wiggins ; Wendy Williams ; Julie Wilson ; Andrew Wood ; Emma Wormauld ; Sahal Yacoob ; Jean Youssef

Contributors and editors have made a sincere effort to produce an accurate and useful resource. Should you have suggestions, find mistakes or be prepared to donate material for inclusion, please don't hesitate to contact us. We intend to work with all who are willing to help make this a continuously evolving resource!

www.fhsst.org

Contents

Chapter 15

Statistics - Grade 10

15.1 Introduction

Information in the form of numbers, graphs and tables is all around us; on television, on the radio or in the newspaper. We are exposed to crime rates, sports results, rainfall, government spending, rate of HIV/AIDS infection, population growth and economic growth.

This chapter demonstrates how Mathematics can be used to manipulate data, to represent or misrepresent trends and patterns and to provide solutions that are directly applicable to the world around us.

Skills relating to the collection, organisation, display, analysis and interpretation of information that were introduced in earlier grades are developed further.

15.2 Recap of Earlier Work

The collection of data has been introduced in earlier grades as a method of obtaining answers to questions about the world around us. This work will be briefly reviewed.

15.2.1 Data and Data Collection

Data

Definition: Data
Data refers to the pieces of information that have been observed and recorded, from an experiment or a survey. There are two types of data: primary and secondary. The word "data" is the plural of the word "datum", and therefore one should say, "the data are" and not "the data is".

Data can be classified as *primary* or *secondary*, and primary data can be classified as *qualitative* or *quantitative*. Figure 15.1 summarises the classifications of data.

Primary data describes the original data that have been collected. This type of data is also known as *raw* data. Often the primary data set is very large and is therefore summarised or processed to extract meaningful information.

Qualitative data is information that cannot be written as numbers.

Quantitative data is information that can be written as numbers.

Secondary data is primary data that has been summarised or processed.

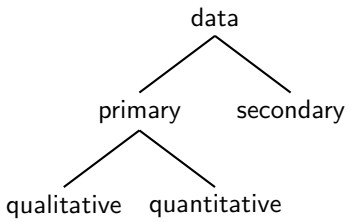


Figure 15.1: Classes of data.

Purpose of Data Collection

Data is collected to provide answers that help with understanding a particular situation. For example:

- The local government might want to know how many residents have electricity and might ask the question: "Does your home have a safe, independent supply of electricity?"
- A supermarket manager might ask the question: "What flavours of soft drink should be stocked in my supermarket?" The question asked of customers might be "What is your favourite soft drink?" Based on the customers' responses, the manager can make an informed decision as to what soft drinks to stock.
- A company manufacturing medicines might ask "How effective is our pill at relieving a headache?" The question asked of people using the pill for a headache might be: "Does taking the pill relieve your headache?" Based on responses, the company learns how effective their product is.
- A motor car company might want to improve their customer service, and might ask their customers: "How can we improve our customer service?"
- A cell phone manufacturing company might collect data about how often people buy new cell phones and what factors affect their choice, so that the cell phone company can focus on those features that would make their product more attractive to buyers.
- A town councillor might want to know how many accidents have occurred at a particular intersection, to decide whether a robot should be installed. The councillor would visit the local police station to research their records to collect the appropriate data.

However, it is important to note that different questions reveal different features of a situation, and that this affects the ability to understand the situation. For example, if the first question in the list was re-phrased to be: "Does your home have electricity?" then if you answered yes, but you were getting your electricity from a neighbour, then this would give the wrong impression that you did not need an independent supply of electricity.

15.2.2 Methods of Data Collection

The method of collecting the data must be appropriate to the question being asked. Some examples of data collecting methods are:

1. Questionnaires, surveys and interviews
2. Experiments
3. Other sources (friends, family, newspapers, books, magazines and the Internet)

The most important aspect of each method of data collecting is to clearly formulate the question that is to be answered. The details of the data collection should therefore be structured to take your question into account.

For example, questionnaires, interviews or surveys would be most appropriate for the list of questions in Section 15.2.1.

15.2.3 Samples and Populations

Before the data collecting starts, an important point to decide upon, is how much data is needed to make sure that the results give an accurate reflection to the answers that are required for the study. Ideally, the study should be designed to maximise the amount of information collected while minimising the effort. The concepts of *populations* and *samples* is vital to minimising effort.

The following terms should be familiar:

Population describes the entire group under consideration in a study. For example, if you wanted to know how many learners in your school got the flu each winter, then your population would be all the learners in your school.

Sample describes a group chosen to represent the population under consideration in a study. For example, for the survey on winter flu, you might select a sample of learners, maybe one from each class.

Random sample describes a sample chosen from a population in such a way that each member of the population has an equal chance of being chosen.

Choosing a representative sample is crucial to obtaining results that are unbiased. For example, if we wanted to determine whether peer pressure affects the decision to start smoking, then the results would be different if only boys were interviewed, compared to if only girls were interviewed, compared to both boys and girls being interviewed.

Therefore questions like: "How many interviews are needed?" and "How do I select the subjects for the interviews?" must be asked during the design stage of the interview process.

The most accurate results are obtained if the entire population is sampled for the survey, but this is expensive and time-consuming. The next best method is to *randomly* select a sample of subjects for the interviews. This means that whatever the method used to select subjects for the interviews, each subject has an equal chance of being selected. There are various methods of doing this but all start with a complete list of each member of the population. Then names can be picked out of a hat or can be selected by using a random number generator. Most modern scientific calculators have a random number generator or you can find one on a spreadsheet program on a computer.

If the subjects for the interviews, are randomly selected then it does not matter too much how many interviews are conducted. So, if you had a total population of 1 000 learners in your school and you randomly selected 100, then that would be the sample that is used to conduct your survey.

15.3 Example Data Sets

The remainder of this chapter deals with the mathematical details that are required to analyse the data collected.

The following are some example sets of data which can be used to apply the methods that are being explained.

15.3.1 Data Set 1: Tossing a Coin

A fair coin was tossed 100 times and the values on the top face were recorded.

15.3.2 Data Set 2: Casting a die

A fair die was cast 100 times and the values on the top face were recorded. The data are recorded in Table 15.3.2.

H	T	T	H	H	T	H	H	H	H
H	H	H	H	T	H	H	T	T	T
T	T	H	T	T	H	T	H	T	H
H	H	T	T	H	T	T	H	T	T
T	H	H	H	T	T	H	T	T	H
H	T	T	T	T	H	T	T	H	H
T	T	H	T	T	H	T	T	H	T
H	T	T	H	T	T	T	T	H	T
T	H	T	T	H	H	H	T	H	T
T	T	T	H	H	T	T	T	H	T

Table 15.1: Results of 100 tosses of a fair coin. H means that the coin landed heads-up and T means that the coin landed tails-up.

3	5	3	6	2	6	6	5	5	6	6	4	2	1	5	3	2	4	5	4
1	4	3	2	6	6	4	6	2	6	5	1	5	1	2	4	4	2	4	4
4	2	6	4	5	4	3	5	5	4	6	1	1	4	6	6	4	5	3	5
2	6	3	2	4	5	3	2	2	6	3	4	3	2	6	4	5	2	1	5
5	4	1	3	1	3	5	1	3	6	5	3	4	3	4	5	1	2	1	2
1	3	2	3	6	3	1	6	3	6	6	1	4	5	2	2	6	3	5	3
1	1	6	4	5	1	6	5	3	2	6	2	3	2	5	6	3	5	5	6
2	6	6	3	5	4	1	4	5	1	4	1	3	4	3	6	2	4	3	6
6	1	1	2	4	5	2	5	3	4	3	4	5	3	3	3	1	1	4	3
5	2	1	4	2	5	2	2	1	5	4	5	1	5	3	2	2	5	1	1

Table 15.2: Results of 200 casts of a fair die.

15.3.3 Data Set 3: Mass of a Loaf of Bread

A loaf of bread should weigh 800g. The masses of 10 different loaves of bread were measured at a store for 1 week. The data is shown in Table 15.3.

"The Trade Metrology Act requires that if a loaf of bread is not labelled, it must weigh 800g, with the leeway of five percent under or 10 percent over. However, an average of 10 loaves must be an exact match to the mass stipulated. - Sunday Tribune of 10 October 2004 on page 10"

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
802.39	787.78	815.74	807.41	801.48	786.59	799.01
796.76	798.93	809.68	798.72	818.26	789.08	805.99
802.50	793.63	785.37	809.30	787.65	801.45	799.35
819.59	812.62	809.05	791.13	805.28	817.76	801.01
801.21	795.86	795.21	820.39	806.64	819.54	796.67
789.00	796.33	787.87	799.84	789.45	802.05	802.20
788.99	797.72	776.71	790.69	803.16	801.24	807.32
808.80	780.38	812.61	801.82	784.68	792.19	809.80
802.37	790.83	792.43	789.24	815.63	799.35	791.23
796.20	817.57	799.05	825.96	807.89	806.65	780.23

Table 15.3: Masses (in g) of 10 different loaves of bread, from the same manufacturer, measured at the same store over a period of 1 week.

15.3.4 Data Set 4: Global Temperature

The mean global temperature from 1861 to 1996 is listed in Table 15.4. The data, obtained from `http://www.cgd.ucar.edu/stats/Data/Climate/`, was converted to mean temperature in degrees Celsius.

`http://lib.stat.cmu.edu/DASL/`

Year	Temperature	Year	Temperature	Year	Temperature	Year	Temperature
1861	12.66	1901	12.871	1941	13.152	1981	13.228
1862	12.58	1902	12.726	1942	13.147	1982	13.145
1863	12.799	1903	12.647	1943	13.156	1983	13.332
1864	12.619	1904	12.601	1944	13.31	1984	13.107
1865	12.825	1905	12.719	1945	13.153	1985	13.09
1866	12.881	1906	12.79	1946	13.015	1986	13.183
1867	12.781	1907	12.594	1947	13.006	1987	13.323
1868	12.853	1908	12.575	1948	13.015	1988	13.34
1869	12.787	1909	12.596	1949	13.005	1989	13.269
1870	12.752	1910	12.635	1950	12.898	1990	13.437
1871	12.733	1911	12.611	1951	13.044	1991	13.385
1872	12.857	1912	12.678	1952	13.113	1992	13.237
1873	12.802	1913	12.671	1953	13.192	1993	13.28
1874	12.68	1914	12.85	1954	12.944	1994	13.355
1875	12.669	1915	12.962	1955	12.935	1995	13.483
1876	12.687	1916	12.727	1956	12.836	1996	13.314
1877	12.957	1917	12.584	1957	13.139		
1878	13.092	1918	12.7	1958	13.208		
1879	12.796	1919	12.792	1959	13.133		
1880	12.811	1920	12.857	1960	13.094		
1881	12.845	1921	12.902	1961	13.124		
1882	12.864	1922	12.787	1962	13.129		
1883	12.783	1923	12.821	1963	13.16		
1884	12.73	1924	12.764	1964	12.868		
1885	12.754	1925	12.868	1965	12.935		
1886	12.826	1926	13.014	1966	13.035		
1887	12.723	1927	12.904	1967	13.031		
1888	12.783	1928	12.871	1968	13.004		
1889	12.922	1929	12.718	1969	13.117		
1890	12.703	1930	12.964	1970	13.064		
1891	12.767	1931	13.041	1971	12.903		
1892	12.671	1932	12.992	1972	13.031		
1893	12.631	1933	12.857	1973	13.175		
1894	12.709	1934	12.982	1974	12.912		
1895	12.728	1935	12.943	1975	12.975		
1896	12.93	1936	12.993	1976	12.869		
1897	12.936	1937	13.092	1977	13.148		
1898	12.759	1938	13.187	1978	13.057		
1899	12.874	1939	13.111	1979	13.154		
1900	12.959	1940	13.055	1980	13.195		

Table 15.4: Global temperature changes over the past x years. Is there a warming of the planet?

15.3.5 Data Set 5: Price of Petrol

The price of petrol in South Africa from August 1998 to July 2000 is shown in Table 15.5.

15.4 Grouping Data

One of the first steps to processing a large set of raw data is to arrange the data values together into a smaller number of groups, and then count how many of each data value there are in each group. The groups are usually based on some sort of interval of data values, so data values that fall into a specific interval, would be grouped together. The grouped data is often presented graphically or in a frequency table. (Frequency means "how many times")

Table 15.5: Petrol prices

Date	Price (R/l)
August 1998	R 2.37
September 1998	R 2.38
October 1998	R 2.35
November 1998	R 2.29
December 1998	R 2.31
January 1999	R 2.25
February 1999	R 2.22
March 1999	R 2.25
April 1999	R 2.31
May 1999	R 2.49
June 1999	R 2.61
July 1999	R 2.61
August 1999	R 2.62
September 1999	R 2.75
October 1999	R 2.81
November 1999	R 2.86
December 1999	R 2.85
January 2000	R 2.86
February 2000	R 2.81
March 2000	R 2.89
April 2000	R 3.03
May 2000	R 3.18
June 2000	R 3.22
July 2000	R 3.36

Worked Example 61: Grouping Data

Question: Group the elements of Data Set 1 to determine how many times the coin landed heads-up and how many times the coin landed tails-up.
Answer
Step 1 : Identify the groups
There are two unique data values: H and T. Therefore there are two groups, one for the H-data values and one for the T-data values.
Step 2 : Count how many data values fall into each group.

Data Value	Frequency
H	44
T	56

Step 3 : Check that the total of the frequency column is equal to the total number of data values.
There are 100 data values and the total of the frequency column is 44+56=100.

15.4.1 Exercises - Grouping Data

1. The height of 30 learners are given below. Fill in the grouped data below. (Tally is a convenient way to count in 5's. We use IIII to indicate 5.)

142	163	169	132	139	140	152	168	139	150
161	132	162	172	146	152	150	132	157	133
141	170	156	155	169	138	142	160	164	168

Group	Tally	Frequency
$130 \leq h < 140$		
$140 \leq h < 150$		
$150 \leq h < 160$		
$160 \leq h < 170$		
$170 \leq h < 180$		

2. An experiment was conducted in class and 50 learners were asked to guess the number of sweets in a jar. The following guesses were recorded.

56	49	40	11	33	33	37	29	30	59
21	16	38	44	38	52	22	24	30	34
42	15	48	33	51	44	33	17	19	44
47	23	27	47	13	25	53	57	28	23
36	35	40	23	45	39	32	58	22	40

A Draw up a grouped frequency table using intervals 11-20, 21-30, 31-40, etc.

15.5 Graphical Representation of Data

Once the data has been collected, it must be organised in a manner that allows for the information to be extracted most efficiently. One method of organisation is to display the data in the form of graphs. Functions and graphs have been studied in Chapter **??**, and similar techniques will be used here. However, instead of drawing graphs from equations as was done in Chapter **??**, bar graphs, histograms and pie charts will be drawn directly from the data.

15.5.1 Bar and Compound Bar Graphs

A bar chart is used to present data where each observation falls into a specific category and where the categories are unrelated. The frequencies (or percentages) are listed along the y-axis and the categories are listed along the x-axis. The heights of the bars correspond to the frequencies. The bars are of equal width and should not touch neighbouring bars.

A compound bar chart (also called component bar chart) is a variant: here the bars are cut into various components depending on what is being shown. If percentages are used for various components of a compound bar, then the total bar height must be 100%. The compound bar chart is a little more complex but if this method is used sensibly, a lot of information can be quickly shown in an attractive fashion.

Examples of a bar and a compound bar graph, for Data Set 1 Table 15.1, are shown in Figure 15.2. According to the frequency table for Data Set 1, the coin landed heads-up 44 times and tails-up 56 times.

15.5.2 Histograms and Frequency Polygons

It is often useful to look at the frequency with which certain values fall in pre-set groups or classes of specified sizes. The choice of the groups should be such that they help highlight features in the data. If these grouped values are plotted in a manner similar to a bar graph, then the resulting graph is known as a histogram. Examples of histograms are shown in Figure 15.3 for Data Set 2, with group sizes of 1 and 2.

Groups	$0 < n \leq 1$	$1 < n \leq 2$	$2 < n \leq 3$	$3 < n \leq 4$	$4 < n \leq 5$	$5 < n \leq 6$
Frequency	30	32	35	34	37	32

Table 15.6: Frequency table for Data Set 2, with a group size of 1.

The same data used to plot a histogram are used to plot a frequency polygon, except the pair of data values are plotted as a point and the points are joined with straight lines. The frequency polygons for the histograms in Figure 15.3 are shown in Figure 15.4.

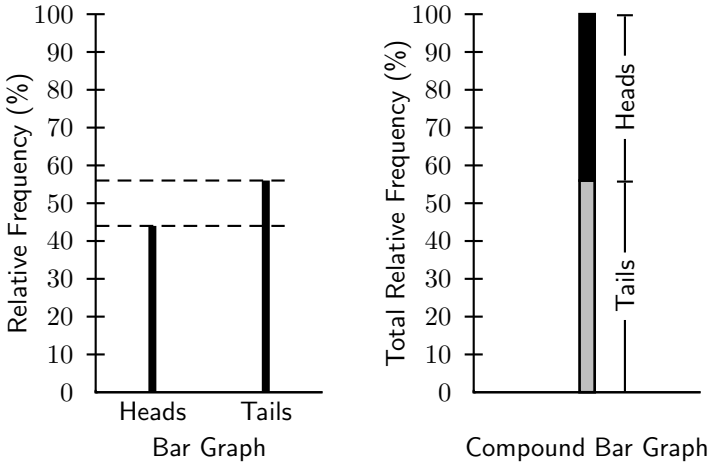


Figure 15.2: Examples of a bar graph (left) and compound bar graph (right) for Data Set 1. The compound bar graph extends from 0% to 100%.

Groups	$0 < n \leq 2$	$2 < n \leq 4$	$4 < n \leq 6$
Frequency	62	69	69

Table 15.7: Frequency table for Data Set 2, with a group size of 2.

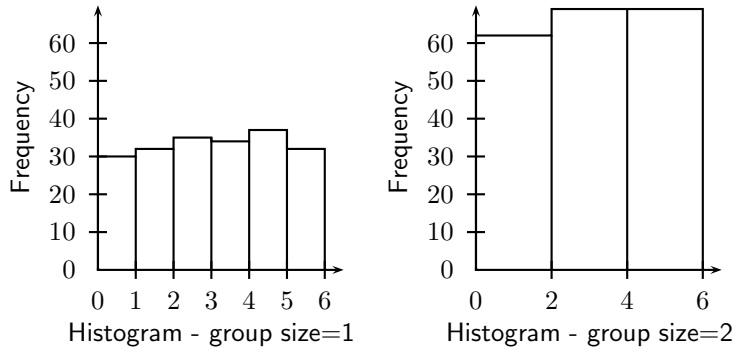


Figure 15.3: Examples of histograms for Data Set 2, with a group size = 1 (left) and a group size = 2 (right). The scales on the y-axis for each graph are the same, and the values in the graph on the right are higher than the values of the graph on the left.

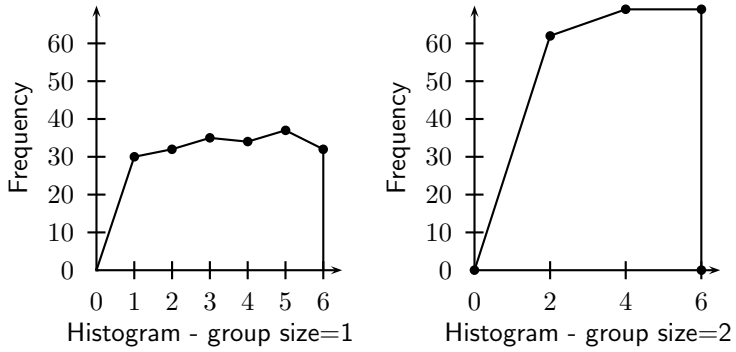


Figure 15.4: Examples of histograms for Data Set 2, with a group size = 1 (left) and a group size = 2 (right). The scales on the y-axis for each graph are the same, and the values in the graph on the right are higher than the values of the graph on the left.

Unlike histograms, many frequency polygons can be plotted together to compare several fre-

quency distributions, provided that the data has been grouped in the same way.

15.5.3 Pie Charts

A pie chart is a graph that is used to show what categories make up a specific section of the data, and what the contribution each category makes to the entire set of data. A pie chart is based on a circle, and each category is represented as a wedge of the circle or alternatively as a slice of the pie. The area of each wedge is proportional to the ratio of that specific category to the total number of data values in the data set. The wedges are usually shown in different colours to make the distinction between the different categories easier.

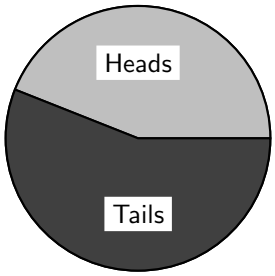


Figure 15.5: Example of a pie chart for Data Set 1. Pie charts show what contribution each group makes to the total data set.

Method: Drawing a pie-chart

1. Draw a circle that represents the entire data set.
2. Calculate what proportion of 360° each category corresponds to according to

$$\text{Angular Size} = \frac{\text{Frequency}}{\text{Total}} \times 360^\circ$$

3. Draw a wedge corresponding to the angular contribution.
4. Check that the total degrees for the different wedges adds up to close to 360°.

Worked Example 62: Pie Chart

Question: Draw a pie chart for Data Set 2, showing the relative proportions of each data value to the total.

Answer

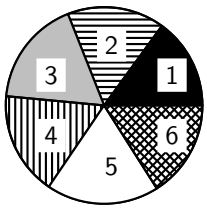
Step 1 : Determine the frequency table for Data Set 2.

							Total
Data Value	1	2	3	4	5	6	–
Frequency	30	32	35	34	37	32	200

Step 2 : Calculate the angular size of the wedge for each data value

Data Value	Angular Size of Wedge
1	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{30}{200} \times 360 = 54^\circ$
2	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{32}{200} \times 360 = 57{,}6^\circ$
3	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{35}{200} \times 360 = 63^\circ$
4	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{34}{200} \times 360 = 61{,}2^\circ$
5	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{37}{200} \times 360 = 66{,}6^\circ$
6	$\frac{\text{Frequency}}{\text{Total}} \times 360^\circ = \frac{32}{200} \times 360 = 57{,}6^\circ$

Step 3 : Draw the pie, with the size of each wedge as calculated above.



Pie Chart for Data Set 2

Note that the total angular size of the wedges may not add up to exactly 360° because of rounding.

15.5.4 Line and Broken Line Graphs

All graphs that have been studied until this point (bar, compound bar, histogram, frequency polygon and pie) are drawn from grouped data. The graphs that will be studied in this section are drawn from the ungrouped or raw data.

Line and broken line graphs are plots of a dependent variable as a function of an independent variable, e.g. the average global temperature as a function of time, or the average rainfall in a country as a function of season.

Usually a line graph is plotted after a table has been provided showing the relationship between the two variables in the form of pairs. Just as in (x,y) graphs, each of the pairs results in a specific point on the graph, and being a LINE graph these points are connected to one another by a LINE.

Many other line graphs exist; they all CONNECT the points by LINES, not necessarily straight lines. Sometimes polynomials, for example, are used to describe approximately the basic relationship between the given pairs of variables, and between these points.

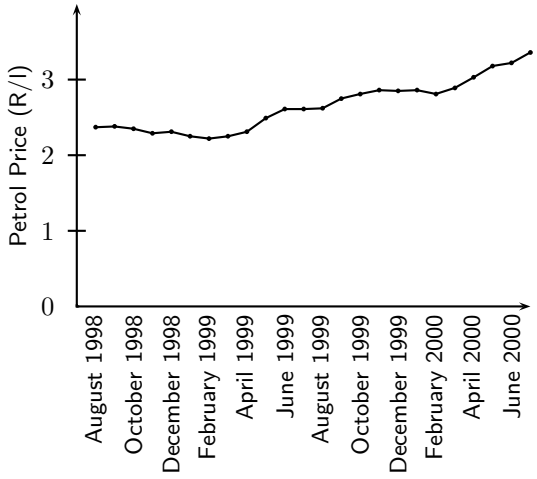


Figure 15.6: Example of a line graph for Data Set 5.

Worked Example 63: Line Graphs

Question: Clawde the cat is overweight and her owners have decided to put her on a restricted eating plan. Her mass is measured once a month and is tabulated

below. Draw a line graph of the data to determine whether the restricted eating plan is working.

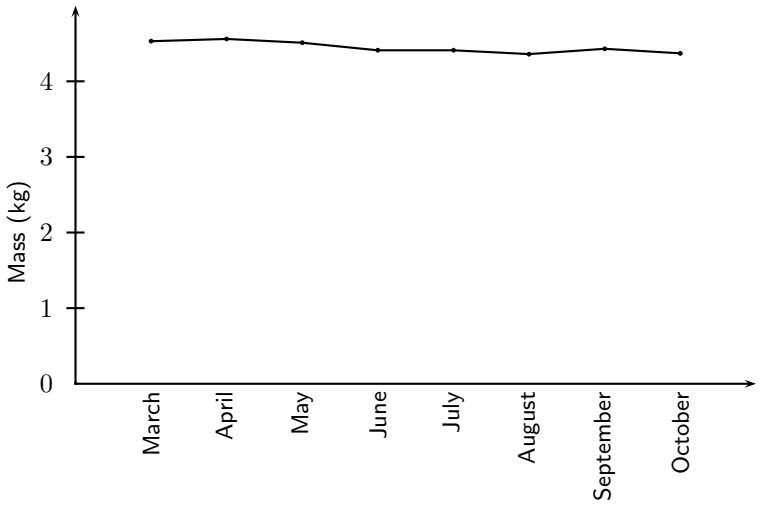
Month	**Mass** (kg)
March	4,53
April	4,56
May	4,51
June	4,41
July	4,41
August	4,36
September	4,43
October	4,37

Answer

Step 1 : Determine what is required

We are required to plot a line graph to determine whether the restricted eating plan is helping Clawde the cat lose weight. We are given all the information that we need to plot the graph.

Step 2 : Plot the graph



Step 3 : Analyse Graph

There is a slight decrease of mass from March to October, so the restricted eating plan is working, but very slowly.

15.5.5 Exercises - Graphical Representation of Data

1. Represent the following information on a pie chart.

Walk	15
Cycle	24
Train	18
Bus	8
Car	35
Total	100

2. Represent the following information using a broken line graph.

Time	07h00	08h00	09h00	10h00	11h00	12h00
Temp (°C)	16	16,5	17	19	20	24

3. Represent the following information on a histogram. Using a coloured pen, draw a frequency polygon on this histogram.

Time in seconds	Frequency
16 - 25	5
26 - 35	10
36 - 45	26
46 - 55	30
56 - 65	15
66 - 75	12
76 - 85	10

4. The maths marks of a class of 30 learners are given below, represent this information using a suitable graph.

82	75	66	54	79	78	29	55	68	91
43	48	90	61	45	60	82	63	72	53
51	32	62	42	49	62	81	49	61	60

5. Use a compound bar graph to illustrate the following information

Year	2003	2004	2005	2006	2007
Girls	18	15	13	12	15
Boys	15	11	18	16	10

15.6 Summarising Data

If the data set is very large, it is useful to be able to summarise the data set by calculating a few quantities that give information about how the data values are spread and about the central values in the data set.

15.6.1 Measures of Central Tendency

An average is simply a number that is representative of a set of data. Specifically, it is a *measure of central tendency* which means that it gives an indication of the main tendency of the set of data. Averages are useful for comparing data, especially when sets of different sizes are being compared.

There are several types of average. Perhaps the simplest and most commonly used average is the *mean* of a set of data. Other common types of average are the *median* and the *mode*.

Mean

The mean, (also known as arithmetic mean), is simply the arithmetic average of a group of numbers (or data set) and is shown using the bar symbol $^-$. So the mean of the variable x is $\bar{x}$ pronounced "x-bar". The mean of a set of values is calculated by adding up all the values in the set and dividing by the number of items in that set. The mean is calculated from the raw, ungrouped data.

Definition: Mean
The mean of a data set, x, denoted by $\bar{x}$, is the average of the data values, and is calculated as:

$$\bar{x} = \frac{\text{sum of all values}}{\text{number of values}} \tag{15.1}$$

Method: Calculating the mean

1. Find the total of the data values in the data set.
2. Count how many data values there are in the data set.
3. Divide the total by the number of data values.

Worked Example 64: Mean

Question: What is the mean of $x = \{10,20,30,40,50\}$?
Answer
Step 1 : Find the total of the data values

$$10 + 20 + 30 + 40 + 50 = 150$$

Step 2 : Count the number of data values in the data set
There are 5 values in the data set.
Step 3 : Divide the total by the number of data values.

$$150 \div 5 = 30$$

Step 4 : Answer
$\therefore$ the mean of the data set $x = \{10,20,30,40,50\}$ is 30.

Median

Definition: Median
The median of a set of data is the data value in the central position, when the data set has been arranged from highest to lowest or from lowest to highest. There are an equal number of data values on either side of the median value.

The median is calculated from the raw, ungrouped data, as follows.

Method: Calculating the median

1. Order the data from smallest to largest or from largest to smallest.
2. Count how many data values there are in the data set.
3. Find the data value in the central position of the set.

Worked Example 65: Median

Question: What is the median of $\{10,14,86,2,68,99,1\}$?
Answer
Step 1 : Order the data set from lowest to highest
1,2,10,14,68,85,99
Step 2 : Count the number of data values in the data set
There are 7 points in the data set.
Step 3 : Find the central position of the data set
The central position of the data set is 4.
Step 4 : Find the data value in the central position of the ordered data set.
14 is in the central position of the data set.
Step 5 : Answer
$\therefore$ 14 is the median of the data set $\{1,2,10,14,68,85,99\}$.

This example has highlighted a potential problem with determining the median. It is very easy to determine the median of a data set with an odd number of data values, but what happens when there is an even number of data values in the data set?

When there is an even number of data values, the median is the mean of the two middle points.

Important: Finding the Central Position of a Data Set

An easy way to determine the central position or positions for any ordered data set is to take the total number of data values, add 1, and then divide by 2. If the number you get is a whole number, then that is the central position. If the number you get is a fraction, take the two whole numbers on either side of the fraction, as the positions of the data values that must be averaged to obtain the median.

Worked Example 66: Median

Question: What is the median of $\{11,10,14,86,2,68,99,1\}$?
Answer
Step 1 : Order the data set from lowest to highest
1,2,10,11,14,68,85,99
Step 2 : Count the number of data values in the data set
There are 8 points in the data set.
Step 3 : Find the central position of the data set
The central position of the data set is between positions 4 and 5.
Step 4 : Find the data values around the central position of the ordered data set.
11 is in position 4 and 14 is in position 5.
Step 5 : Answer
$\therefore$ the median of the data set $\{1,2,10,11,14,68,85,99\}$ is

$$(11 + 14) \div 2 = 12,5$$

Mode

Definition: Mode
The mode is the data value that occurs most often, i.e. it is the most frequent value or most common value in a set.

Method: Calculating the mode Count how many times each data value occurs. The mode is the data value that occurs the most.

The mode is calculated from grouped data, or single data items.

Worked Example 67: Mode

Question: Find the mode of the data set $x = \{1, 2, 3, 4, 4, 4, 5, 6, 7, 8, 8, 9,10,10\}$
Answer
Step 1 : Count how many times each data value occurs.

data value	frequency	data value	frequency
1	1	6	1
2	1	7	1
3	1	8	2
4	3	9	1
5	1	10	2

Step 2 : Find the data value that occurs most often.
4 occurs most often.
Step 3 : Answer
The mode of the data set $x = \{1, 2, 3, 4, 4, 4, 5, 6, 7, 8, 8, 9, 10, 10\}$ is 4.

A data set can have more than one mode. For example, both 2 and 3 are modes in the set 1, 2, 2, 3, 3. If all points in a data set occur with equal frequency, it is equally accurate to describe the data set as having many modes or no mode.

15.6.2 Measures of Dispersion

The mean, median and mode are measures of central tendency, i.e. they provide information on the central data values in a set. When describing data it is sometimes useful (and in some cases necessary) to determine the spread of a distribution. Measures of dispersion provide information on how the data values in a set are distributed around the mean value. Some measures of dispersion are range, percentiles and quartiles.

Range

Definition: Range
The range of a data set is the difference between the lowest value and the highest value in the set.

Method: Calculating the range

1. Find the highest value in the data set.
2. Find the lowest value in the data set.
3. Subtract the lowest value from the highest value. The difference is the range.

Worked Example 68: Range

Question: Find the range of the data set $x = \{1, 2, 3, 4, 4, 4, 5, 6, 7, 8, 8, 9, 10, 10\}$
Answer
Step 1 : Find the highest and lowest values.
10 is the highest value and 1 is the lowest value.
Step 2 : Subtract the lowest value from the highest value to calculate the range.

$$10 - 1 = 9$$

Step 3 : Answer
For the data set $x = \{1, 2, 3, 4, 4, 4, 5, 6, 7, 8, 8, 9, 10, 10\}$, the range is 9.

Quartiles

Definition: Quartiles
Quartiles are the three data values that divide an ordered data set into four groups containing equal numbers of data values. The median is the second quartile.

The quartiles of a data set are formed by the two boundaries on either side of the median, which divide the set into four equal sections. The lowest 25% of the data being found below the first quartile value, also called the lower quartile. The median, or second quartile divides the set into two equal sections. The lowest 75% of the data set should be found below the third quartile, also called the upper quartile. For example:

Data Items										
22	24	48	51	60	72	73	75	80	88	90
		↓ Lower quartile (Q_1)			↓ Median (Q_2)			↓ Upper quartile (Q_3)		

Method: Calculating the quartiles

1. Order the data from smallest to largest or from largest to smallest.
2. Count how many data values there are in the data set.
3. Divide the number of data values by 4. The result is the number of data values per group.
4. Determine the data values corresponding to the first, second and third quartiles using the number of data values per quartile.

Worked Example 69: Quartiles

Question: What are the quartiles of $\{3,5,1,8,9,12,25,28,24,30,41,50\}$?
Answer
Step 1 : Order the data set from lowest to highest
$\{1, 3, 5, 8, 9, 12, 24, 25, 28, 30, 41, 50\}$
Step 2 : Count the number of data values in the data set
There are 12 values in the data set.
Step 3 : Divide the number of data values by 4 to find the number of data values per quartile.

$$12 \div 4 = 3$$

Step 4 : Find the data values corresponding to the quartiles.

1 3 5 $\|$ 8 9 12 $\|$ 24 25 28 $\|$ 30 41 50
Q_1 Q_2 Q_3

The first quartile occurs between data position 3 and 4 and is the average of data values 5 and 8. The second quartile occurs between positions 6 and 7 and is the average of data values 12 and 24. The third quartile occurs between positions 9 and 10 and is the average of data values 28 and 30.
Step 5 : Answer
The first quartile = 6,5. (Q_1)
The second quartile = 18. (Q_2)
The third quartile = 29. (Q_3)

Inter-quartile Range

Definition: Inter-quartile Range
The inter quartile range is a measure which provides information about the spread of a data set, and is calculated by subtracting the first quartile from the third quartile, giving the range of the middle half of the data set, trimming off the lowest and highest quarters, i.e. $Q_3 - Q_1$.

The semi-interquartile range is half the interquartile range, i.e. $\frac{Q_3 - Q_1}{2}$

Worked Example 70: Medians, Quartiles and the Interquartile Range

Question: A class of 12 students writes a test and the results are as follows: 20, 39, 40, 43, 43, 46, 53, 58, 63, 70, 75, 91. Find the range, quartiles and the Interquartile Range.
Answer
Step 1 :

$$\underset{}{20\ \ 39\ \ 40} \underset{Q_1}{\|} \underset{}{43\ \ 43\ \ 46} \underset{M}{\|} \underset{}{53\ \ 58\ \ 63} \underset{Q_3}{\|} \underset{}{70\ \ 75\ \ 91}$$

Step 2 : The Range
The range = 91 - 20 = 71. This tells us that the marks are quite widely spread.
Step 3 : The median lies between the 6th and 7th mark
i.e. $M = \frac{46+53}{2} = \frac{99}{2} = 49{,}5$
Step 4 : The lower quartile lies between the 3rd and 4th mark
i.e. $Q_1 = \frac{40+43}{2} = \frac{83}{2} = 41{,}5$
Step 5 : The upper quartile lies between the 9th and 10th mark
i.e. $Q_3 = \frac{63+70}{2} = \frac{133}{2} = 66{,}5$
Step 6 : Analysing the quartiles
The quartiles are 41,5, 49,5 and 66,5. These quartiles tell us that 25% of the marks are less than 41,5; 50% of the marks are less than 49,5 and 75% of the marks are less than 66,5. They also tell us that 50% of the marks lie between 41,5 and 66,5.
Step 7 : The Interquartile Range
The Interquartile Range = 66,5 - 41,5 = 25. This tells us that the width of the middle 50% of the data values is 25.
Step 8 : The Semi-interquatile Range
The Semi-interquartile Range = $\frac{25}{2}$ = 12,5

Percentiles

Definition: Percentiles
Percentiles are the 99 data values that divide a data set into 100 groups.

The calculation of percentiles is identical to the calculation of quartiles, except the aim is to divide the data values into 100 groups instead of the 4 groups required by quartiles.

Method: Calculating the percentiles

1. Order the data from smallest to largest or from largest to smallest.
2. Count how many data values there are in the data set.

3. Divide the number of data values by 100. The result is the number of data values per group.
4. Determine the data values corresponding to the first, second and third quartiles using the number of data values per quartile.

15.6.3 Exercises - Summarising Data

1. Three sets of data are given:
   A **Data set 1:** 9 12 12 14 16 22 24
   B **Data set 2:** 7 7 8 11 13 15 16 16
   C **Data set 3:** 11 15 16 17 19 19 22 24 27

   For each one find:
   i. the range
   ii. the lower quartile
   iii. the interquartile range
   iv. the semi-interquartile range
   v. the median
   vi. the upper quartile
2. There is 1 sweet in one jar, and 3 in the second jar. The mean number of sweets in the first two jars is 2.
   A If the mean number in the first three jars is 3, how many are there in the third jar?
   B If the mean number in the first four jars is 4, how many are there in the fourth jar?
   C If the mean number in the first n jars is n, how many are there in the n jar?
3. Find a set of five ages for which the mean age is 5, the modal age is 2 and the median age is 3 years.
4. Four friends each have some marbles. They work out that the mean number of marbles they have is 10. One of them leaves. She has 4 marbles. How many marbles do the remaining friends have together?

Worked Example 71: Mean, Median and Mode for Grouped Data

Question:
Consider the following grouped data and calculate the mean, the modal group and the median group.

Mass (kg)	Frequency
41 - 45	7
46 - 50	10
51 - 55	15
56 - 60	12
61 - 65	6
	Total = 50

Answer
Step 1 : Calculating the mean
To calculate the mean we need to add up all the masses and divide by 50. We do not know actual masses, so we approximate by choosing the midpoint of each group. We then multiply those midpoint numbers by the frequency. Then we add these numbers together to find the approximate total of the masses. This is show in the table below.

Mass (kg)	Midpoint	Frequency	Midpt × Freq
41 - 45	(41+45)/2 = 43	7	43 × 7 = 301
46 - 50	48	10	480
51 - 55	53	15	795
56 - 60	58	12	696
61 - 65	63	6	378
		Total = 50	Total = 2650

Step 2 : Answer
The mean $= \frac{2650}{50} = 53$.
The modal group is the group 51 - 53 because it has the highest frequency.
The median group is the group 51 - 53, since the 25th and 26th terms are contained within this group.

Exercise: More mean, modal and median group exercises.
In each data set given, find the mean, the modal group and the median group.

1. Times recorded when learners played a game.

Time in seconds	Frequency
36 - 45	5
46 - 55	11
56 - 65	15
66 - 75	26
76 - 85	19
86 - 95	13
96 - 105	6

2. The following data were collected from a group of learners.

Mass in kilograms	Frequency
41 - 45	3
46 - 50	5
51 - 55	8
56 - 60	12
61 - 65	14
66 - 70	9
71 - 75	7
76 - 80	2

15.7 Misuse of Statistics

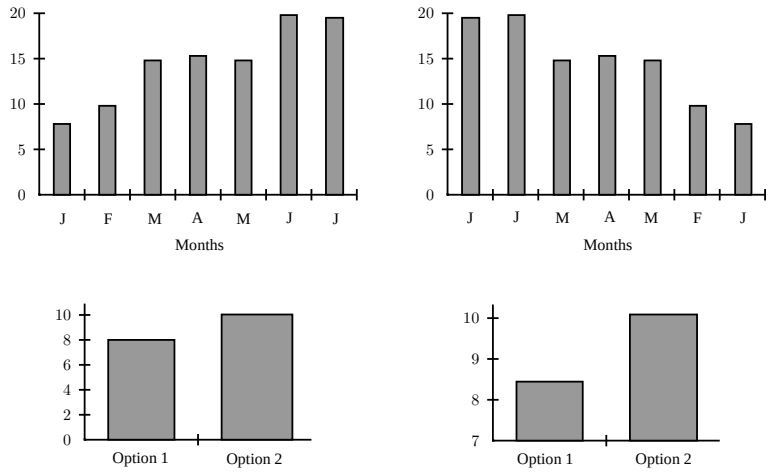
In many cases groups can gain an advantage by misleading people with the misuse of statistics.

Common techniques used include:

- Three dimensional graphs.
- Axes that do not start at zero.

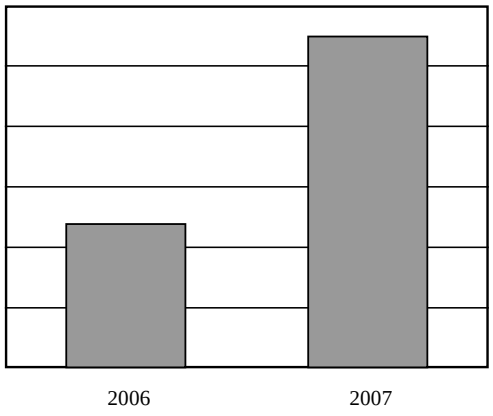
- Axes without scales.
- Graphic images that convey a negative or positive mood.
- Assumption that a correlation shows a necessary causality.
- Using statistics that are not truly representative of the entire population.
- Using misconceptions of mathematical concepts

For example, the following pairs of graphs show identical information but look very different. Explain why.



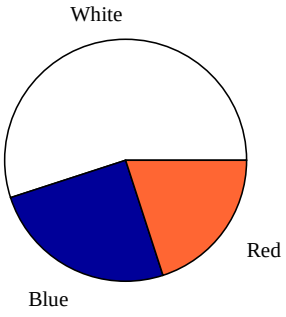
15.7.1 Exercises - Misuse of Statistics

1. A company has tried to give a visual representation of the increase in their earnings from one year to the next. Does the graph below convince you? Critically analyse the graph.



2. In a study conducted on a busy highway, data was collected about drivers breaking the speed limit and the colour of the car they were driving. The data were collected during a 20 minute time interval during the middle of the day, and are presented in a table and pie chart below.

Colour of car	Frequency of drivers speeding
White	22
Blue	10
Red	8



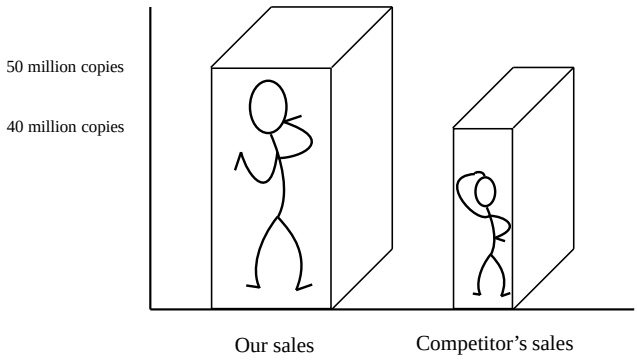
Conclusions made by a novice based on the data are summarised as follows:

"People driving white cars are more likely to break the speed limit."

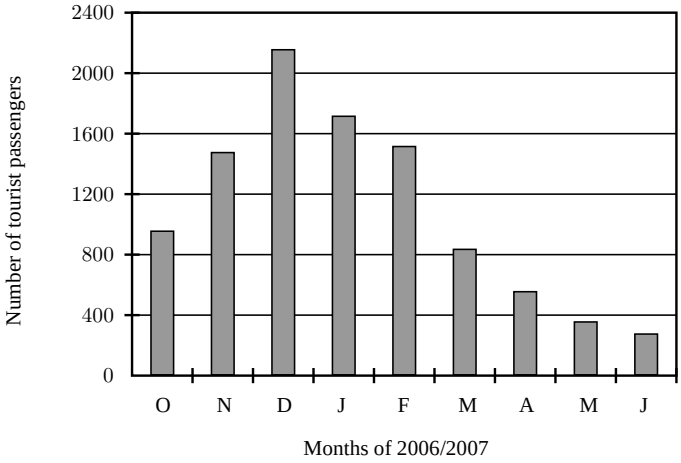
"Drivers in blue and red cars are more likely to stick to the speed limit."

Do you agree with these conclusions? Explain.

3. A record label produces a graphic, showing their advantage in sales over their competitors. Identify at least three devices they have used to influence and mislead the readers impression.



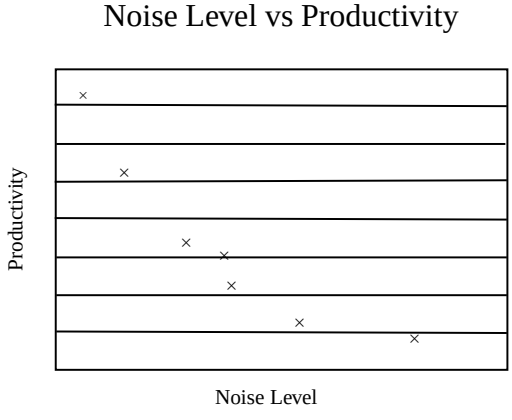
4. In an effort to discredit their competition, a tour bus company prints the graph shown below. Their claim is that the competitor is losing business. Can you think of a better explanation?



5. The caption from a newspaper article quoted below, demonstrates a misuse of statistical concepts. Explain.

"More than 40% of learners in South Africa are below average in mathematics."

6. To test a theory, 8 different offices were monitored for noise levels and productivity of the employees in the office. The results are graphed below.



The following statement was then made:
"If an office environment is noisy, this leads to poor productivity."

Explain the flaws in this thinking.

15.8 Summary of Definitions

mean The mean of a data set, x, denoted by $\bar{x}$, is the average of the data values, and is calculated as:

$$\bar{x} = \frac{\text{sum of values}}{\text{number of values}} \tag{15.2}$$

median The median is the centre data value in a data set that has been ordered from lowest to highest

mode The mode is the data value that occurs most often in a data set.

15.9 Exercises

1. "Using the median size as a reference, you would be able to fit four 1 cent coins and a car into a match box." Explain why this statement is true.
2. Calculate the mean, median, and mode of Data Set 3.
3. The tallest 7 trees in a park have heights in metres of 41, 60, 47, 42, 44, 42, and 47. Find the median of their heights.
4. The students in Bjorn's class have the following ages: 5, 9, 1, 3, 4, 6, 6, 6, 7, 3. Find the mode of their ages.
5. The masses (in kg, correct to the nearest 0,1 kg) of thirty people were measured as follows:

45,1	57,9	67,9	57,4	50,7	61,1	63,9	67,5	69,7	71,7
68,0	63,2	58,7	56,9	78,5	59,7	54,4	66,4	51,6	47,7
70,9	54,8	59,1	60,3	60,1	52,6	74,9	72,1	49,5	49,8

   A Copy the frequency table below, and complete it.

Mass (in kg)	Tally	Number of people
$45{,}0 \leq m < 50{,}0$		
$50{,}0 \leq m < 55{,}0$		
$55{,}0 \leq m < 60{,}0$		
$60{,}0 \leq m < 65{,}0$		
$65{,}0 \leq m < 70{,}0$		
$70{,}0 \leq m < 75{,}0$		
$75{,}0 \leq m < 80{,}0$		

   B Draw a frequency polygon for this information.

   C What can you conclude from looking at the graph?

6. An engineering company has designed two different types of engines for motorbikes. The two different motorbikes are tested for the time it takes (in seconds) for them to accelerate from 0 km/h to 60 km/h.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Average
Bike 1	1.55	1.00	0.92	0.80	1.49	0.71	1.06	0.68	0.87	1.09	
Bike 2	0.9	1.0	1.1	1.0	1.0	0.9	0.9	1.0	0.9	1.1	

   A What kind of average should be used for this information?

   B Calculate the average you chose in the previous question for each motorbike.

C Which motorbike would you choose based on this information? Take note of accuracy of the numbers from each set of tests.

D How far will a motorbike travelling at 60 km/h travel in 1 second?

7. The heights of 40 learners are given below.

154	140	145	159	150	132	149	150	138	152
141	132	169	173	139	161	163	156	157	171
168	166	151	152	132	142	170	162	146	152
142	150	161	138	170	131	145	146	147	160

A Set up a frequency table using 6 intervals.

B Calculate the approximate mean.

C Determine the mode.

D Determin the modal class.

E How many learners are taller than your approximate average in (b)?

8. In a traffic survey, a random sample of 50 motorists were asked the distance they drove to work daily. This information is shown in the table below.

Distance in km	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
Frequency	4	5	9	10	7	8	3	2	2

A Find the approximate mean.

B Find the modal class.

C What percentage of samples drove

i. less than 16 km?

ii. more than 30 km?

iii. between 16 km and 30 km daily?

9. A company wanted to evaluate the training programme in its factory. They gave the same task to trained and untrained employees and timed each one in seconds.

Trained	121	137	131	135	130
	128	130	126	132	127
	129	120	118	125	134
Untrained	135	142	126	148	145
	156	152	153	149	145
	144	134	139	140	142

A Draw a back-to-back stem and leaf diagram to show the two sets of data.

B Find the medians and quartiles for both sets of data.

C Find the Interquartile Range for both sets of data.

D Comment on the results.

10. A small firm employs nine people. The annual salaries of the employers are:

R600 000	R250 000	R200 000
R120 000	R100 000	R100 000
R100 000	R90 000	R80 000

A Find the mean of these salaries.

B Find the mode.

C Find the median.

D Of these three figures, which would you use for negotiating salary increases if you were a trade union official? Why?

11. The marks for a particular class test are listed here:

67	58	91	67	58	82	71	51	60	84
31	67	96	64	78	71	87	78	89	38
69	62	60	73	60	87	71	49		

A Complete the frequency table using the given class intervals.

Class	Tally	Frequency	Mid-point	Freq $\times$ Midpt
30-39		34,5		
40-49		44,5		
50-59				
60-69				
70-79				
80-89				
90-99				
		Sum =		Sum =

Appendix A

GNU Free Documentation License

PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or non-commercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts,

you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections A and A above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

14. Do not re-title any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organisation as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section A above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section A) to Preserve its Title (section A) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sub-license, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sub-license or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.