KEYWORDS

Analyse	**Mean**
Bias	**Median**
Categorical data	**Mode**
Census	**Organisation**
Clustered	**Outlier**
Collection	**Polygon**
Complementary events	**Population sample**
Continuous data	**Probability event outcome**
Data	**Quantitative data**
Discrete data	**Random**
Dot plot	**Range**
Frequency	**Scatter diagram**
Grouped data	**Score**
Histogram	**Spread**
Information	**Stem-and-leaf**
Location	**Tally**

Some Explanations

Statistics involves the collection and organisation of information (**data**) so that:

a Large amounts of information can be easily analysed, and

b Predictions can be made, based on analysis of the data collected.

Tables and graphs allow information to be presented in a clear, concise form. The information can also be readily analysed from the table or graph.

Census versus Sample

A **census** involves surveying every member of the target population, while a **sample** is a limited survey used to get a representative or typical view. The Australian Government conducts a regular census of all Australian households, asking questions about such things as number of residents, ages of residents, occupations etc. The collected census information is analysed and used for future planning by government departments. Things such as new subdivisions, schools, infrastructure and hospitals can be planned in advance, based on information in the census.

Biased and Random Samples

If a sample is **random**, then there is no **bias**. A biased sample is one that does not represent the whole population. For example, if a sample is conducted at a shopping centre on a Thursday morning to find the percentage of Australians who eat Weet Bix for breakfast, it would be a biased sample.

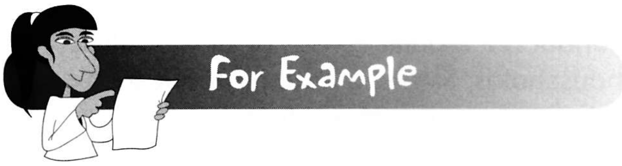
In a random sample participants are chosen 'at random'. For example, to randomly select from 1000 students, a teacher might allocate a three-digit number to each student. She can then use her calculator to generate random numbers to find 20 students to interview.

Data Representation

Data can be **quantitative** or **categorical**. Examples of quantitative data are shoe sizes (**discrete**) or height (**continuous**), while an example of categorical data is hair colour: brown, black, blonde, red. Collected data can be represented in tables or graphs.

Frequency Distribution Table

A table is used to summarise listed data and allows easy analysis of the location and spread of the data.



Consider the shoe sizes of 50 11-year-old boys:

6	$4\frac{1}{2}$	7	4	5
$5\frac{1}{2}$	6	8	5	7
5	$5\frac{1}{2}$	7	$4\frac{1}{2}$	5
6	$5\frac{1}{2}$	4	6	6
7	$4\frac{1}{2}$	5	5	6
5	5	$5\frac{1}{2}$	$4\frac{1}{2}$	5
$7\frac{1}{2}$	$3\frac{1}{2}$	5	$4\frac{1}{2}$	$6\frac{1}{2}$
$4\frac{1}{2}$	$5\frac{1}{2}$	6	$7\frac{1}{2}$	6
6	5	6	5	7
6	6	7	$5\frac{1}{2}$	6

1 Draw up a frequency distribution table for this information.

2 How many boys wear a size $5\frac{1}{2}$ shoe?

3 Which size shoe is most commonly worn?

4 If this is a typical example of 11-year-old boys, what fraction of 11-year-old boys in Australia wear a size 6 shoe?

5 If the population of 11-year-olds in Hambelton is 200, how many would you predict wear a size 6 shoe?

1

Score (x)	Tally	Frequency (f)	
$3\frac{1}{2}$	|	1	
4	||	2	
$4\frac{1}{2}$	卌 |	6	
5	卌 卌 ||	12	
$5\frac{1}{2}$	卌 |	6	← (b)
6	卌 卌 |||	13	← (c)
$6\frac{1}{2}$	|	1	
7	卌 |	6	
$7\frac{1}{2}$	||	2	
8	|	1	
	Σf	50	

[Σ means the sum or running total]

The completed table can then be used to answer the questions:

2 Six boys wear size $5\frac{1}{2}$ shoes.

3 The most common size is 6 as 13 boys wear size 6.

4 Fraction of 11-year-old boys with size 6 shoe:

$$= \frac{13}{50}\left(\frac{\text{Number wearing size 6}}{\text{Total number of boys in survey}}\right)$$

5 Number wearing size 6 $\doteqdot \frac{13}{\cancel{50}_1} \times \frac{\cancel{200}^4}{1}$

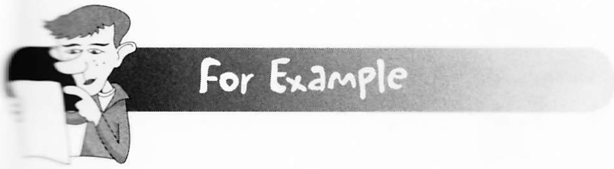
$\doteqdot 52$

[$\doteqdot$ or » means 'approximately equal to']

You would expect about 52 11-year-olds in Hambelton to wear size 6 shoes.

Frequency Distribution Table for Grouped Data

When data is continuous it is sensible to use class intervals.



The mass of forty students was measured and the results listed below:

53	51	62	64	63
71	75	79	56	60
53	48	61	64	63
67	59	63	68	44
53	55	59	52	64
67	68	72	75	79
61	65	48	41	47
48	55	57	56	61

1 Arrange the data in a frequency distribution table using class intervals of 41–45, 46–50 etc. Also, find the class centre for each interval.

2 What was the most common class interval of masses (the **modal** class)?

1

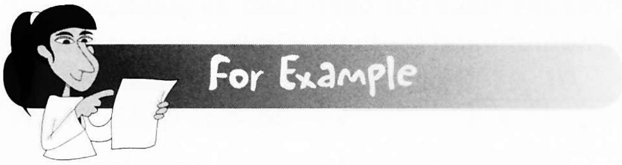
Class intervals	Class centre (x)	Tally	Frequency (f)
41–45	43	𝍷𝍷	2
46–50	48	𝍷𝍷𝍷𝍷	4
51–55	53	𝍸 𝍷𝍷	7
56–60	58	𝍸 𝍷	6
61–65	63	𝍸 𝍸 𝍷	11
66–70	68	𝍷𝍷𝍷𝍷	4
71–75	73	𝍷𝍷𝍷𝍷	4
76–80	78	𝍷𝍷	2
			40

2 Most common class interval of masses was 61–65.

[Note: The scores are **clustered** in the low 60s.]

Stem-and-Leaf Plot

The stem is the *first* digit or digits of a number, whereas the leaf is the *last* digit. The leaf is always a single digit.



1 The results of a mathematics test were recorded:

19	48	36	40	31	22	18	27
18	20	18	36	49	60	45	13
9	17	22	31	39	26	28	30
44	8	23	19	46	33		

Draw a stem-and-leaf plot for this data.

1

Stem	Leaf
0	98
1	9888379
2	2702683
3	6161903
4	809546
5	
6	0

This is sometimes referred to as the initial plot. However, the plot can be refined by ordering each of the leaves:

Stem	Leaf
0	89
1	3788899
2	0223678
3	0113669
4	045689
5	
6	0

An ordered stem-and-leaf plot can be used to analyse the data. A back-to-back stem-and-leaf plot is used to compare two sets of data.

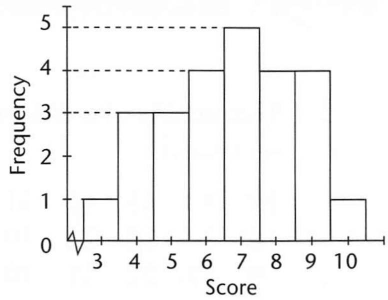
[Note: The score of 60 is called an **outlier**.]

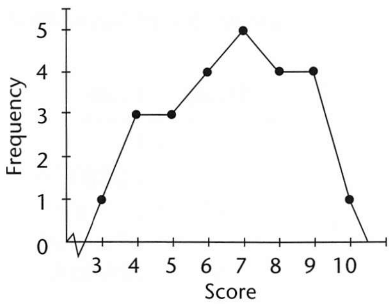
Frequency Histogram and Polygon

When the data are continuous there are two types of frequency graphs that can be used: a **histogram** and a **polygon**.

Here the table has been used to graph a histogram and frequency polygon:

x	f
3	1
4	3
5	3
6	4
7	5
8	4
9	4
10	1



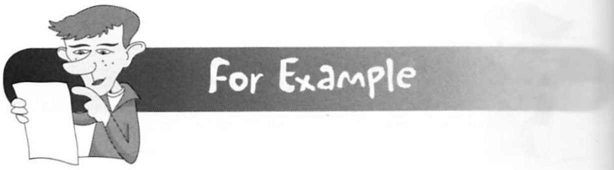


Columns on the histogram are around the scores. Their height is their frequency.

Straight lines on the frequency polygon join dots at the height of the frequency value. The line meets the horizontal axis where the next score would be if there was one. The polygon can also be made by joining the mid-points of the tops of the columns in the histogram.

Dot Plots

A dot plot is an alternative to drawing a histogram and can either be vertical or horizontal.

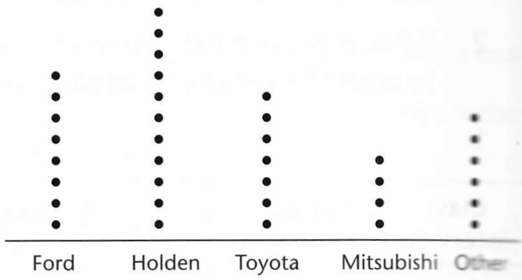


1 The type of cars passing along Honus Avenue over a half-hour period was noted and the results recorded:

Car manufacturer	Frequency
Ford	8
Holden	11
Toyota	7
Mitsubishi	4
Other	6

Use a dot plot to represent this information.

1 Vehicle survey



Scatter Diagram

A **scatter diagram** is formed by using points to represent a pair of results.

The results of an English test and a Science test for ten students are detailed:

Student: A B C D E F G H I J

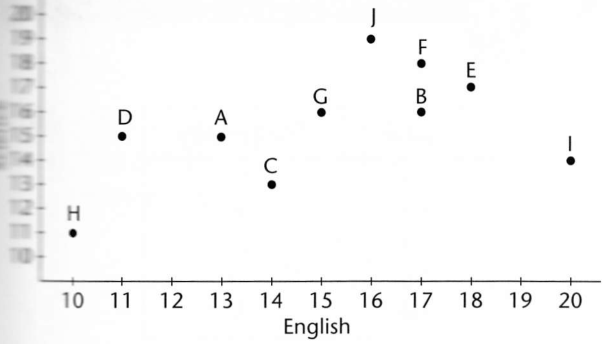
English: 13 17 14 11 18 17 15 10 20 16

Science: 15 16 13 15 17 18 16 11 14 19

1 Draw a scatter diagram for the data.

2 Which students scored the highest and lowest scores in each test?

1 **Test results**

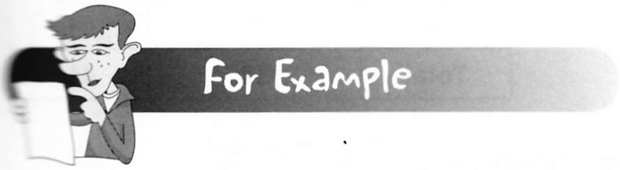


2 English highest: student I; lowest: student H.
Science highest: student J; lowest: student H.

Data Analysis

Measures of Spread

The **range** indicates the size of the distribution of scores; that is, the range = highest score – lowest score.



1 Find the range of the scores 4, 17, 6, –5, 4, 2, 12.

1 Range = 17 – (–5)
= 17 + 5 = 22

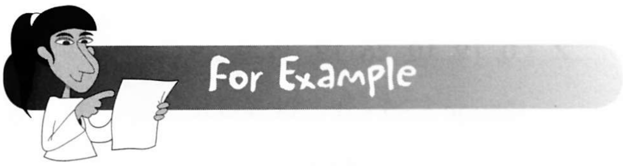
Measures of Location

- **Mean** is the *average score*:

  i.e. $\text{mean} = \dfrac{\text{sum of scores}}{\text{number of scores}}$

  [Notation: $\bar{x}$ = mean]

- **Mode** is most common score—the score with the highest frequency.
- **Median** is the middle score when the scores are arranged in order.



Consider these sets of scores for Alison's first season of cricket. Over the first seven innings she scored 17, 14, 10, 10, 9, 21 and 10. Over the next eight innings she scored 0, 19, 10, 26, 19, 35, 0 and 29.

1 For the first seven innings, calculate the range, mode, mean and median.

2 For the next eight innings, calculate the range, mode, mean and median.

3 Taking all scores together, find the season's range, mode, mean and median.

1 17, 14, 10, 10, 9, 21, 10

Range = 21 – 9
= 12

Mode = 10

$$\text{Mean} = \frac{17 + 14 + 10 + 10 + 9 + 21 + 10}{7} = \frac{91}{7} = 13$$

For the median, rewrite the scores in ascending order:

9 10 10 10 14 17 21

$$\text{3 scores below} \longleftarrow \text{middle score} \longrightarrow \text{3 scores above}$$

Median = 10

[A suggested method to locate the middle is to cross off the first and last numbers, then the remaining first and last numbers etc. until only one or two middle numbers are left.]

2 0, 19, 10, 26, 19, 35, 0, 29

Range = 35 – 0 = 35

Mode = 0 and 19

$$\text{Mean} = \frac{0 + 19 + 10 + 26 + 19 + 35 + 0 + 29}{8}$$
$$= \frac{138}{8}$$
$$= 17.25$$

For the median, consider:
0, 0, 10, 19, 19, 26, 29, 35.

Median is middle of 4th and 5th scores; that is, 19, 19.

Median = 19

3 17, 14, 10, 10, 9, 21, 10, 0, 19, 10, 26, 19, 35, 0, 29

Range = 35 – 0 = 35

Mode = 10

$$\text{Mean} = \frac{17 + 14 + 10 + 10 + 9 + 21 + 10 + 0 + 19 + 10 + 26 + 19 + 35 + 0 + 29}{15}$$
$$= \frac{229}{15}$$
$$= 15.27 \text{ (to 2 decimal places)}$$

For median: 0, 0, 9, 10, 10, 10, 10, 14, 17, 19, 19, 21, 26, 29, 35

Median is the middle score.

Median = 14

Using a Calculator to Find the Mean

A calculator can be placed in 'statistics' mode to find a number of statistical measures. Refer to your calculator manual for instructions.

Measures of Location and Spread from a Table or Graph

Mean, mode, median and range can be found when data is arranged in a table or graph.

For Example

Score	Frequency
12	5
13	7
14	8
15	7
16	3

1 Find the mean, mode, median and range.

1 To find the mean, another column is added to the table. In this column, we multiply each score by its frequency, i.e. the $x \times f$ column.

Score (x)	Frequency (f)	fx
12	5	60
13	7	91
14	8	112
15	7	105
16	3	48
Total	**30**	**416**

- **Mean:** $(\bar{x}) = \dfrac{\text{Sum of scores}}{\text{Number of scores}}$
  $= \dfrac{\text{Sum of } fx}{\text{Sum of } f} = \dfrac{\Sigma fx}{\Sigma f}$

  i.e. $\bar{x} = \frac{416}{30}$
  $= 13.87$ (to 2 decimal places)
- **Mode:** 14 (has highest frequency of 8)
- **Median:** As there are 30 scores, then the median is the average of the two middle scores:

  i.e. the average of 15th and 16th scores

  i.e. the average of 14 and 14

  $\therefore$ Median = 14
- **Range:** 16 – 12 = 4

For Example

1 The divided stem-and-leaf plot shows the results for the boys and girls in a maths test:

Boys	Stem	Girls
42	0	9
98430	1	35689
85441	2	555899
62	3	0025

Find the mean, mode, median and range for boys and girls.

1 ■ **Mean**—Boys: $(2 + 4 + 10 + 13 + \ldots) \div 14$

[or use calculator in statistics mode]

$= 19.29$ (to 2 decimal places)

—Girls: $(9 + 13 + 15 + \ldots) \div 16$

$= 23.625$

■ **Mode**—Boys: 24; Girls: 25

■ **Median**

Boys	Stem	Girls
~~42~~	0	~~9~~
9~~8430~~	1	~~35689~~
~~8544~~1	2	55~~5899~~
~~62~~	3	~~0025~~

Boys: Middle (average) of 19 and 21

i.e. $\dfrac{19 + 21}{2} = 20$

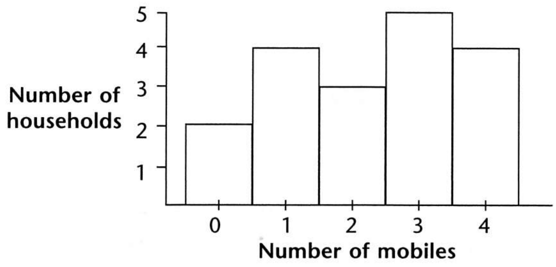
Girls: Middle of 25 and 25 $\therefore$ 25

■ **Range**—Boys: $36 - 2 = 34$

Girls: $35 - 9 = 26$

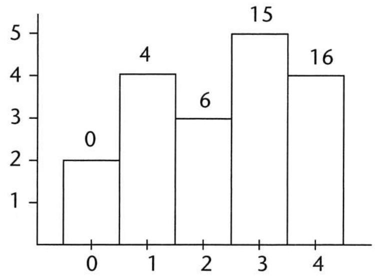
For Example

1 A class is surveyed to find the number of mobile phones per household:



Find the mean, mode and median mobiles.

1 **Mean:** To help find the mean students can either complete a frequency table for the histogram or note the $x \times f$ data on the existing histogram:



Scores (x)	f	fx
0	2	0
1	4	4
2	3	6
3	5	15
4	4	16
	18	41

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{0 + 4 + 6 + 15 + 16}{2 + 4 + 3 + 5 + 4} = \frac{41}{18} = 2.2\dot{7}$$

Mode: 3 (highest frequency—tallest column)

Median: As there are 18 scores:

$\therefore$ median is middle of 9th and 10th score

i.e. 0, 0, 1, 1, 1, 1, 2, 2, 2, 3…

$\therefore$ median = 2.5

Probability

A coin is tossed once. There is one chance in two that it will land with a head facing up. We say that the probability of throwing a head is one out of two, or:

$P(\text{Head}) = \frac{1}{2}$

Similarly, if a die (singular of dice) is thrown once, there is one chance in 6 of throwing a two:

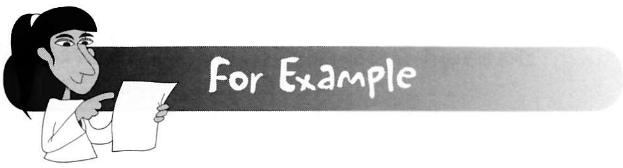
i.e. $P(2) = \frac{1}{6}$

The probability of an event occurring is defined as:

$$\text{Probability (Event)} = \frac{\text{Number of favourable outcomes}}{\text{Number of possible outcomes}}$$

$$\text{or} \quad P(\text{E}) = \frac{n(\text{E})}{\text{Total number possible}}$$

Note: A certainty has probability of one. An impossibility has probability of zero. Therefore probability is expressed as a fraction between and including 0 and 1.



A die is rolled once. Find the probability of rolling:

1 3

2 5

3 An odd number

4 An even number

5 A number larger than 4.

1 $P(3) = \frac{1}{6}$

[one 3 can occur out of 6 possible numbers]

2 $P(5) = \frac{1}{6}$

3 $P(\text{Odd}) = \frac{3}{6}$ [odd numbers = 3]

$= \frac{1}{2}$

4 $P(\text{Even}) = \frac{3}{6}$

$= \frac{1}{2}$

5 $P(>4) = \frac{2}{6}$

[numbers greater than 4 are 5 and 6
i.e. 2 numbers out of 6]

$= \frac{1}{3}$

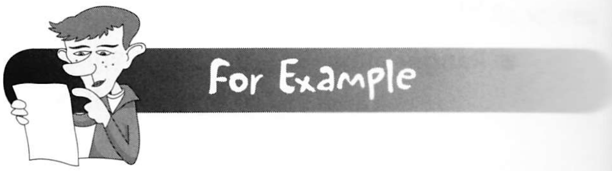
Total Probability

If there are two possible things that can happen, then the sum of their probabilities is 1. When a coin is tossed, for example, the two possible outcomes are head or tail:

$P(\text{Head}) = \frac{1}{2} \qquad P(\text{Tail}) = \frac{1}{2}$

Now, $P(\text{H}) + P(\text{T}) = \frac{1}{2} + \frac{1}{2} = 1$

This is called total probability. It is true for any number of possible outcomes. If you add up the probabilities of all possible outcomes they will total one.



1 When a die is thrown there are 6 possible outcomes: a one, two, three, four, five or six.

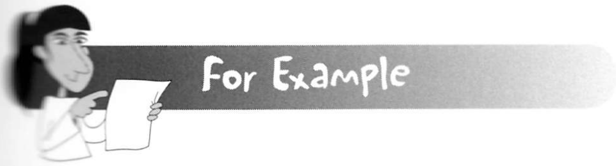
1 Now, $P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$,

$P(4) = \frac{1}{6}$, $P(5) = \frac{1}{6}$, $P(6) = \frac{1}{6}$

Total $P = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$

Complementary Events

Also, because the total of the probabilities of all possible outcomes is 1, then if there are two possible outcomes and we know the probability of one outcome, then we also know the probability of the other.



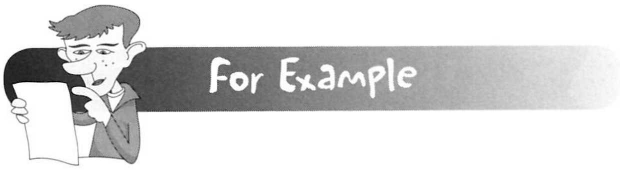
If the probability of rain on a Monday is $\frac{2}{5}$,

then the probability of no rain on a Monday:

$= 1 - \frac{2}{5}$

$= \frac{3}{5}$

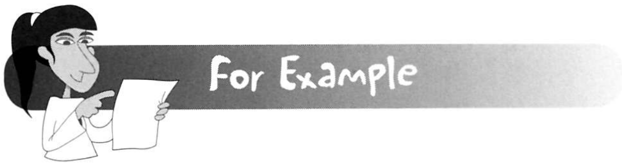
These are called **complementary events.**



1 If the probability of an archer hitting a tree is 0.8, calculate the probability of her not hitting the tree.

1 $P(\text{Not hitting}) = 1 - 0.8$

$= 0.2$



A die is thrown. Calculate the probability that the score on the uppermost face is:

1 Not 6

2 Greater than 2.

1 $P(6) = \frac{1}{6}$

$\therefore P\ (\text{Not } 6) = 1 - \frac{1}{6}$

$= \frac{5}{6}$

2 Prob. (1 or 2) $= \frac{2}{6}$

$= \frac{1}{3}$

$P(>2) = 1 - P(1 \text{ or } 2)$

$= 1 - \frac{1}{3}$

$= \frac{2}{3}$