Choosing an Appropriate Graph

To choose an appropriate statistical graph, consider the set of data values.  In general, use the following guidelines:

* Use a [bar chart](http://www.mathsteacher.com.au/year8/ch17_stat/05_disp/data.htm#bar) if you are not looking for trends (or patterns) over time; and the items (or categories) are not parts of a whole.
* Use a [pie chart](http://www.mathsteacher.com.au/year8/ch17_stat/06_pie/charts.htm#pie) if you need to compare different parts of a whole, there is no time involved and there are not too many items (or categories).
* Use a [line graph](http://www.mathsteacher.com.au/year8/ch17_stat/07_line/graphs.htm#line) if you need to see how a quantity has changed over time.  Line graphs enable us to find trends (or patterns) over time.

Line Graphs

A **line graph** is often used to represent a set of data values in which a quantity varies with time.  These graphs are useful for finding **trends**.  That is, finding a **general pattern** in data sets including temperature, sales, employment, company profit or cost over a period of time.

Example 10

A cylinder of liquid was heated.  Its temperature was recorded at ten-minute intervals as shown in the following table.





b.  The estimated temperature after 25 minutes of heating is 52°C.

### Bar Charts

**Bar charts** are often used to present data in a pictorial form to illustrate the information collected and highlight important points.  They are especially useful to depict monthly car production, monthly sales, quarterly profit, average annual rainfall etc.  A bar chart provides a useful comparison of data over time.  The height of each bar shows the total amount of the item of interest for each month (or year).

Bar charts are drawn with parallel bars placed vertically (or horizontally).  The **width**of each bar andthe **spacing**between the bars are kept the same to avoid giving a misleading representation.  The height of the bar is drawn to scale to represent the amount of the item.

#### Example 8

The yearly production of cars by a particular company is recorded as follows:



Draw a bar chart to display this information.

##### Solution: for figuring out how to scale each axis



Thus, bars of equal width whose heights are in the ratio of 4 : 5 : 8 will represent the company's yearly production.



###### Note the following:

* It is simple to read a bar chart.  Just look at the required bar and read off the value.  E.g. The bar chart shows that the number of cars produced was 160 000 in 2001, 200 000 in 2002 and 320 000 in 2003.
* Bar charts are used to describe only simple pieces of data but can describe a data set clearly and provide immediate visual impact.  It is clear from the bar chart that the car production by the company had increased more slowly between 2001 and 2002 than between 2002 and 2003.
* Bar charts are useful for presenting data sets consisting of a number of values that are each assigned to different categories such as years, months, quarters etc.

## Pie Charts

**Pie charts** are useful to compare different parts of a whole amount.  They are often used to present financial information.  E.g. A company's expenditure can be shown to be the [sum](http://www.mathsteacher.com.au/year8/ch01_arithmetic/01_basic/oper.htm#add) of its parts including different expense categories such as salaries, borrowing interest, taxation and general running costs (i.e. rent, electricity, heating etc).

A pie chart is a circular chart in which the [circle](http://www.mathsteacher.com.au/year8/ch10_geomcons/03_circles/comp.htm#circle) is divided into [sectors](http://www.mathsteacher.com.au/year8/ch10_geomcons/03_circles/comp.htm#sector).  Each sector visually represents an item in a data set to match the amount of the item as a [percentage](http://www.mathsteacher.com.au/year8/ch08_consumer/01_per/per.htm#percentage) or [fraction](http://www.mathsteacher.com.au/year8/ch02_fracdec/01_frac/frac.htm#frac) of the total data set.

#### Example 9

A family's weekly expenditure on its house mortgage, food and fuel is as follows:



##### Solution:



We can find what percentage of the total expenditure each item equals.

Percentage of weekly expenditure on:


To draw a pie chart, divide the circle into 100 percentage parts.  Then allocate the number of percentage parts required for each item.



###### Note:

* It is simple to read a pie chart.  Just look at the required sector representing an item (or category) and read off the value.  For example, the weekly expenditure of the family on food is 37.5% of the total expenditure measured.
* A pie chart is used to compare the different parts that make up a whole amount.

## Mean, Median and Mode

We use **statistics** such as the [mean](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mean), [median](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#median) and [mode](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mode) to obtain information about a [population](http://www.mathsteacher.com.au/year8/ch17_stat/01_data/surv.htm#pop) from our [sample](http://www.mathsteacher.com.au/year8/ch17_stat/01_data/surv.htm#sample) set of observed values.

### Mean

The **mean** (or average) of a set of data values is the [sum](http://www.mathsteacher.com.au/year8/ch01_arithmetic/01_basic/oper.htm#add) of all of the data values divided by the number of data values.  That is:



#### Example 1

The marks of seven students in a mathematics test with a maximum possible mark of 20 are given below:
     15     13     18     16     14     17     12

Find the mean of this set of data values.

##### Solution:



So, the mean mark is 15.

Symbolically, we can set out the solution as follows:



So, the mean mark is 15.

### Median

The **median** of a set of data values is the middle value of the data set when it has been arranged in ascending order.  That is, from the smallest value to the highest value.

#### Example 2

The marks of nine students in a geography test that had a maximum possible mark of 50 are given below:
     47     35     37     32     38     39     36     34     35

Find the median of this set of data values.

##### Solution:

Arrange the data values in order from the lowest value to the highest value:

     32     34     35     35     36     37     38     39     47

The fifth data value, 36, is the middle value in this arrangement.



###### Note:



###### In general:



If the number of values in the data set is even, then the **median** is the average of the two middle values.

#### Example 3

Find the median of the following data set:
     12     18     16     21     10     13     17     19

##### Solution:

Arrange the data values in order from the lowest value to the highest value:
     10     12     13     16     17     18     19     21

The number of values in the data set is 8, which is even.  So, the median is the average of the two middle values.



###### Alternative way:

There are 8 values in the data set.



The fourth and fifth scores, 16 and 17, are in the middle.  That is, there is no one middle value.


Note:

* Half of the values in the data set lie below the median and half lie above the median.
* The median is the most commonly quoted figure used to measure property prices.  The use of the median avoids the problem of the mean property price which is affected by a few expensive properties that are not representative of the general property market.

### Mode

The **mode** of a set of data values is the value(s) that occurs most often.

The mode has applications in printing.  For example, it is important to print more of the most popular books; because printing different books in equal numbers would cause a shortage of some books and an oversupply of others.

Likewise, the mode has applications in manufacturing.  For example, it is important to manufacture more of the most popular shoes; because manufacturing different shoes in equal numbers would cause a shortage of some shoes and an oversupply of others.

#### Example 4

Find the mode of the following data set:

     48     44     48     45     42     49     48

##### Solution:

The mode is 48 since it occurs most often.

###### Note:

* It is possible for a set of data values to have more than one mode.
* If there are two data values that occur most frequently, we say that the set of data values is **bimodal**.
* If there is no data value or data values that occur most frequently, we say that the set of data values has no mode.

## Analysing Data

The [mean](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mean), [median](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#median) and [mode](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mode) of a data set are collectively known as **measures of central tendency** as these three measures focus on where the data is centred or clustered.  To analyse data using the mean, median and mode, we need to use the most appropriate measure of central tendency.  The following points should be remembered:

* The mean is useful for predicting future results when there are no extreme values in the data set.  However, the impact of extreme values on the mean may be important and should be considered.  E.g. The impact of a stock market crash on average investment returns.
* The median may be more useful than the mean when there are extreme values in the data set as it is not affected by the extreme values.
* The mode is useful when the most common item, characteristic or value of a data set is required.