Insider tip…

You will be allowed to use a calculator in the exam – just make sure you have one. However, most mathematical skills won't require any calculation.

SOME BASIC MATHEMATICAL CONCEPTS

You are likely to have encountered at least some of the concepts outlined below in your maths course; therefore only brief explanations are provided.

Fractions

A **fraction** is a part of a whole number such as $\frac{1}{2}$ or $\frac{3}{4}$. We may want to present the results from a study as a fraction. For example, if there were 120 participants in a study and 40 of them were in condition A, what fraction of the participants is this?

To calculate a fraction we divide 40 by $120 = \frac{40}{120}$.

To make a fraction more comprehensible we reduce a fraction by dividing the top number (the numerator) and the bottom number (the denominator) by the lowest number that divides evenly into both (the lowest common denominator).

In this case that number is 40, which results in a fraction of $\frac{1}{3}$.

Percentages

The term 'per cent' means 'out of 100' (cent means 100). Therefore 5% essentially means 5 out of 100 or $\frac{5}{100}$. We have converted the fraction to a **percentage**.

We can reduce this fraction to $\frac{1}{20}$.

Or we can write $\frac{5}{100}$ as a decimal $= 0.05$, because the first decimal place is out of 10 and the second is out of 100.

The decimal 0.5 would be 5 out of 10, not 5 out of 100.

To change a fraction to a percentage, divide the numerator by the denominator. For example, for the fraction $\frac{19}{36}$, we divide 19 by 36 (using a calculator) and get 0.52777778.

To make this into a percentage we multiply by 100 (move the decimal point two places to the right) and get 52.777778%.

Ratios

A **ratio** says how much there is of one thing compared to another thing.

Ratios are used in betting, so if you are a betting man or woman you will be at home. Odds are given as 4 to 1 (4:1), meaning that out of a total of five events you would be expected to lose four times and win once.

There are two ways to express a ratio. Either the way above, which is called a part-to-part ratio; or a part-to-whole ratio, which would be expressed as 4:5, meaning four losses out of five occurrences.

A part-to-whole ratio can easily be changed to a fraction: 4:5 is $\frac{4}{5}$.

Ratios can be reduced to a lowest form in the same way that fractions are, so 10:15 would more simply be 2:3 (both parts of the fraction have been divided by 5).

Estimate results

When doing any calculations it helps to estimate what the result is likely to be because then you can detect if you make a mistake.

Consider the fraction $\frac{19}{36}$. It is fairly close to $\frac{18}{36}$, which is the same as half (50%), therefore my answer should be slightly more than half.

The same thing could be done when dealing with big numbers. For example, to estimate the product of 185,363 times 46,208 I could round up 185,363 to 200,000 and round up 46,208 to 50,000.

Then I multiple 5×2 and add nine zeros = 10,000,000,000.

I know the actual answer will be smaller because I rounded both numbers up. The actual answer is 8,565,253,504.

Significant figures

In the example above there are a lot of digits, many of which are distracting! It would be a lot simpler if I told you that the answer was about eight billion (8,000,000,000). In this case I have given the answer to one **significant figure** and all the rest are zeros for less distraction.

Except that's not quite right. We can't just remove all the remaining figures without considering whether we have to round up. In our example, 8,500,000,000 would be half way between eight and nine billion and 8,565,253,504 should be rounded up to nine billion (1 significant figure). Two significant figures would be 8,600,000,000.

Let's consider the percentage on the left, 52.777778%, another awkward number. We might give that to two significant figures, which would be 53% (removing all but two figures and rounding up because the third figure is more than five). If we wanted to give this number to three significant figures it would be 52.8%. If the number was 52.034267% then three significant figures would be 52.0% – we have to indicate three figures.

Order of magnitude and standard form

When dealing with very large numbers it is sometimes clearer to just give two significant figures and then say how many zeros there are, thus focusing on the **order of magnitude**. The convention for doing this for 8,600,000,000 is 8.6×10^9 where 9 represents how many places we have moved the decimal point. To convert 0.0045 we write 4.5×10^{-3} (this is **standard form**).

Mathematical symbols

And finally, you deserve a reward if you have got this far! The symbols you need to be able to use are in the table below.

= and ~	< and <<	> and >>	≤	∝
Equal and approximately equal	*Less than and much less than*	*More than and much more than*	*Less than or equal to*	*Proportional to*

KEY TERMS

Fraction, percentage, ratio Methods of expressing parts of a whole.

Order of magnitude A means of expressing a number by focusing on the overall size (magnitude). This is done by expressing the number in terms of powers of 10.

Significant figure Refers to the number of important single digits used to represent a number. The digits are 'important' because, if removed, the number would be quite different in magnitude.

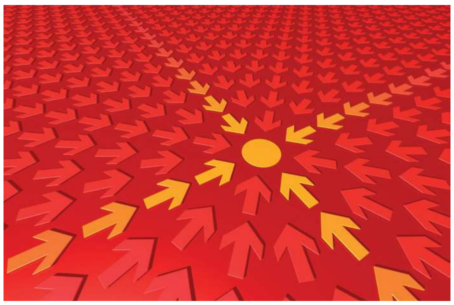
Standard form A means of expressing very large or very small numbers, a number between 1 and 10 multiplied by 10 (to the power of a number).

CAN YOU? **No. 7.17**

1. Represent $\frac{3}{8}$ as a percentage. Give your answer to two significant figures. (2 marks)
2. A researcher wants to divide 4,526 by 42. Estimate what the result would be, explaining how you arrived at your answer. (2 marks)
3. Express 0.02 as a fraction. (1 mark)
4. Briefly explain what the following expression means: 'The number of girls < number of boys'. (1 mark)

Measures of central tendency and dispersion

The information collected in any study is called data or, more precisely, a 'data set' (a set of items). Data are not necessarily numbers; they could be words used to describe how someone feels. For the moment we are going to focus on numerical data, called **quantitative data**. Once a researcher has collected such data, the data needs to be analysed in order to identify trends or to see the 'bigger picture'. One of the ways to do this is *describing* the data, for example by giving an average score for a group of participants. For this reason such statistics are called descriptive statistics – they identify general patterns and trends.



▲ Finding the centre of your data – a measure of the centre or 'central tendency'.

Levels of measurement

Distinctions are made between different kinds of data.

- **Nominal** Data are in separate categories, such as grouping people according to their favourite football team (e.g. Liverpool, Inverness Caledonian Thistle, Oxford United, etc.).
- **Ordinal** Data are ordered in some way, for example asking people to put a list of football teams in order of liking. Liverpool might be first, followed by Inverness Caledonian Thistle, and so on. The 'difference' between each item is not the same, i.e. the individual may like the first item a lot more than the second, but there might only be a small difference between the items ranked second and third.
- **Interval** Data are measured using units of equal intervals, such as when counting correct answers or using any 'public' unit of measurement. Many psychological studies use 'plastic interval scales' where the intervals are arbitrarily determined and we can't therefore know for certain that there are equal intervals between the numbers. However, for the purposes of analysis, such data may be accepted as interval.
- **Ratio** There is a true zero point as in most measures of physical quantities.

MEASURES OF CENTRAL TENDENCY

Measures of central tendency inform us about central (or middle) values for a set of data. They are 'averages' – ways of calculating a typical value for a set of data. The average can be calculated in different ways, each one appropriate for a different situation.

Mean

The **mean** is calculated by adding up all the data items and dividing by the number of data items. It is properly called the arithmetic mean because it involves an arithmetic calculation. It can only be used with ratio and interval level data.

Median

The **median** is the middle value in an ordered list. All data items must be arranged in order and the central value is then the median. If there are an even number of data items there will be two central values. To calculate the median add the two data items and divide by two. The median can be used with ratio, interval and ordinal data.

Mode

The **mode** is the value that is the most common data item. With nominal data it is the category that has the highest frequency count. With interval and ordinal data it is the data item that occurs most frequently. To identify this the data items need to be arranged in order. The modal group is the group with the greatest frequency.

If two categories or data items have the same frequency the data have two modes, i.e. are bi-modal.

MEASURES OF DISPERSION

A set of data can also be described in terms of how dispersed or spread out the data items are. These descriptions are known as **measures of dispersion**.

Range

The **range** is the arithmetic distance between the top and bottom values in a set of data. It is customary to add 1, so, for example, with the first data set below the range would be 15–3+1. The addition of 1 is because the bottom number of 3 could represent a value as low as 2.5 and the top number of 15 could represent a number as big as 15.5.

Consider the data sets below:

3, 5, 8, 8, 9, 10, 12, 12, 13, 15 mean = 9.5; range = 13 (15–3+1)

1, 5, 8, 8, 9, 10, 12, 12, 13, 17 mean = 9.5; range = 17 (17–1+1)

The two sets of numbers have the same mean but a different range, so the range is helpful as a further method of *describing* the data. If we just used the mean, the data would appear to be the same.

Standard deviation

There is a more precise method of expressing dispersion, called the **standard deviation**. This is a measure of the average distance between each data item above and below the mean, ignoring plus or minus values. It is usually worked out using a calculator. The standard deviations for the two sets of numbers above are 3.69 and 4.45 respectively (worked out using a calculator). You won't be asked to calculate a standard deviation in the exam.

▶ The mean number of legs that people have is 1.999. It would be better to use the mode to describe the average number of legs.

NOIR – an acronym to help remember the four levels of measurement of data: nominal, ordinal, interval and ratio.

EVALUATION OF MEASURES OF CENTRAL TENDENCY

The mean

Strengths

- The mean is the most sensitive measure of central tendency because it takes account of the exact distance between all the values of all the data.

Limitations

- This sensitivity means that it can be easily distorted by one (or a few) extreme values and thus end up being misrepresentative of the data as a whole.
- It cannot be used with nominal data.
- It does not make sense to use it when you have discrete values such as average number of legs.

Therefore, the mean is not always representative of the data as a whole and should always be considered alongside the standard deviation.

The median

Strengths

- The median is not affected by extreme scores.
- It is appropriate for ordinal (ranked) data.
- It can be easier to calculate than the mean.

Limitations

- The median is not as 'sensitive' as the mean because the exact values are not reflected in the final calculation.

The median therefore has strengths in that it can be used to describe a variety of data sets, including skewed data and non-normal distributions.

The mode

Stengths

- The mode is also unaffected by extreme values.
- It is much more useful for discrete data.
- It is the only method that can be used when the data are in categories, i.e. nominal data.

Limitations

- It is not a useful way of describing data when there are several modes.
- It also tells us nothing about the other values in a distribution.

As with all three measures of central tendency, the key is to use the mode only with data sets for which it is appropriate.

EVALUATION OF MEASURES OF DISPERSION

Range

Stengths

- The range is easy to calculate.

Limitations

- It is affected by extreme values.
- It fails to take account of the distribution of the numbers. For example, it doesn't indicate whether most numbers are closely grouped around the mean or spread out evenly.

The range is useful for ordinal data or with highly skewed data or when making a quick calculation.

Standard deviation

Strengths

- The standard deviation is a precise measure of dispersion because it takes all the exact values into account.
- It is not difficult to calculate if you have a calculator.

Limitations

- It may hide some of the characteristics of the data set (e.g. extreme values).

Therefore, the standard deviation is best used, together with the mean, to describe interval or ratio data which is normally distributed.

▶ In an American study 66% of students rated themselves as having a better sense of humour than the average person (Kruger and Dunning, 1999). This is called the better-than-average effect – other findings include 93% of a US sample rating their driving as better than average (Svenson, 1981) and 85% of students rating themselves as above average in their ability to get on with others (Alicke and Govorun, 2005).

APPLY YOUR KNOWLEDGE

1. For each of the following data sets, where appropriate calculate the mean, the median and/or the mode. (3 marks for each data set)
   a. 2, 3, 5, 6, 6, 8, 9, 12, 15, 21, 22
   b. 2, 3, 8, 10, 11, 13, 13, 14, 14, 29
   c. 2, 2, 4, 5, 5, 5, 7, 7, 8, 8, 8, 10
   d. cat, cat, dog, budgie, snake, gerbil
2. For each of the data sets (a–d) in question 1, state which of the three measures of central tendency would be most suitable to use and why. (2 marks for each data set)
3. Estimate the mean and standard deviation for the following data sets. (2 marks for each data set)
   a. 119, 131, 135, 142, 145, 147, 155, 156, 161, 163
   b. 0.15, 0.23, 0.28, 0.34, 0.34, 0.34, 0.36, 0.46
4. For each of the data sets (a–b) in question 3 explain what measure of central tendency might be preferable to use and explain why. (2 marks for each data set)
5. For each of the data sets (a–b) in question 3 explain what measure of dispersion might be preferable to use and explain why. (2 marks for each data set)
6. Look at the following two data sets. Which one do you think would have the smaller standard deviation? (1 mark)
   Data set A: 2 2 3 4 5 9 11 14 18 20 21 22 25
   Data set B: 2 5 8 9 9 10 11 12 14 15 16 20 25

KEY TERMS

Mean The arithmetic average of a data set. Takes the exact values of all the data into account.

Measure of central tendency A descriptive statistic that provides information about a 'typical' value for a data set.

Measure of dispersion A descriptive statistic that provides information about how spread out a set of data are.

Median The middle value of a data set when the items are placed in rank order.

Mode The most frequently occurring value or item in a data set.

Quantitative data Data measured in numbers.

Range The difference between the highest and lowest item in a data set. Usually 1 is added as a correction.

Standard deviation shows the amount of variation in a data set. It assesses the spread of data around the mean.

CAN YOU?

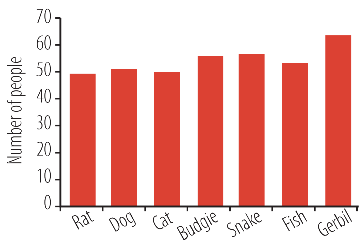
No. 7.18

1. Identify **one** measure of central tendency and explain how to calculate it for a set of data. (3 marks)
2. Briefly explain **one** strength and **one** limitation of using the mean to work out the central tendency of a data set. (2 marks + 2 marks)
3. Identify **one** measure of dispersion and explain how to calculate it for a set of data. (3 marks)
4. Explain why it is sometimes preferable to use the mode rather than the mean as a measure of central tendency. (3 marks)
5. Explain why it might be better to know the standard deviation of a data set rather than the range. (3 marks)

Display of quantitative data and data distributions

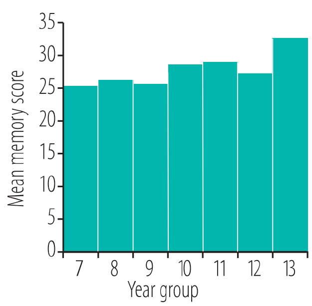
A picture is worth 1,000 words! Graphs and tables provide a means of 'eyeballing' your data and seeing the findings at a glance. Using graphs and tables are a way of describing data and therefore are also descriptive statistics, like measures of central tendency and dispersion. In fact we often display measures of central tendency and dispersion in a graph because it is easier to grasp the significance of the statistics in visual form.

Graph A



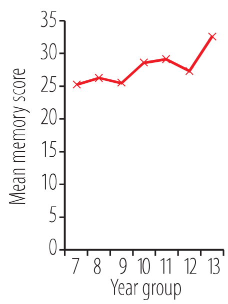
▲ A bar chart showing the students' favourite pets.

Graph B



▲ A histogram showing the mean memory scores for each year group in a school (maximum score is 40).

Graph C



▲ A line graph showing the same data as Graph B.

KEY TERMS

Bar chart A graph used to represent the frequency of data; the categories on the *x*-axis have no fixed order and there is no true zero.

Histogram Type of frequency distribution in which the number of scores in each category of continuous data are represented by vertical columns. There is a true zero and no spaces between the bars.

Negative skewed distribution Most of the scores are bunched towards the right. The mode is to the right of the mean because the mean is affected by the extreme scores tailing off to the left.

Normal distribution A symmetrical bell-shaped frequency distribution. This distribution occurs when certain variables are measured, such as IQ or the life of a light bulb. Such 'events' are distributed in such a way that most of the scores are clustered close to the mid-point; the mean, median and mode are at the mid-point.

Positive skewed distribution Most of the scores are bunched towards the left. The mode is to the left of the mean because the mean is affected by the extreme scores tailing off to the right.

Skewed distribution A distribution is skewed if one tail is longer than another, signifying that there are a number of extreme values to one side or the other of the mid-score.

DISPLAY OF QUANTITATIVE DATA

Graphs and tables should be simple so they can be read easily.

- They should clearly show the findings from a study.
- There should be a short but informative title.
- In a graph both axes should be clearly labelled. The *x*-axis goes across the page. In the case of a bar chart or histogram, it is usually the independent variable. The vertical, or *y*-axis, is usually frequency.
- Always use squared paper if you are hand-drawing graphs.

Tables

The measurements collected in a research study are referred to as 'raw data' – numbers that haven't been before any descriptive statistics have been carried out. These data can be set out in a table and/or summarised using measures of central tendency and dispersion. Such summary tables are more helpful for interpreting findings.

Bar chart

The height of each bar represents the frequency of each item. **Bar charts** are especially suitable for data that is not continuous, i.e. has no particular order such as Graph A on the left which is categorical or nominal data. In a bar chart a space is left between each bar to indicate the lack of continuity.

Histogram

A **histogram** is similar to a bar chart except that the area within the bars must be proportional to the frequencies represented (see Graph B). In practice this means that the vertical axis (frequency) must start at zero. In addition the horizontal axis must be continuous (therefore you can't draw a histogram with data in categories). Finally, there should be no gaps between the bars.

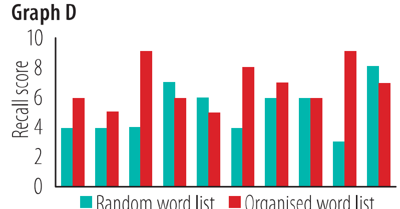
Line graph

A line graph, like a histogram, has continuous data on the *x*-axis and there is a dot to mark the middle top of where each bar would be and each dot is connected by a line (see Graph C).

Scattergram

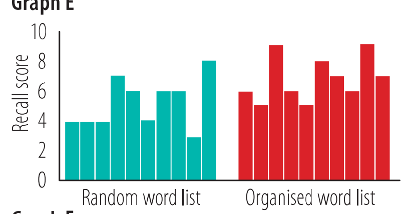
A scattergram is a kind of graph used when doing a correlational analysis (see page 206).

Graph D



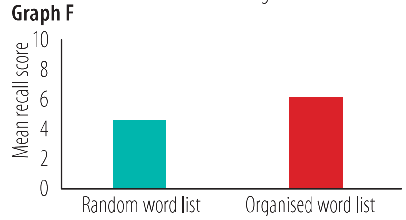
◀ Participant number 1 in the random word group is placed next to participant number 1 in the organised word group. Students like to draw 'participant charts', but they are totally meaningless.

Graph E



◀ The findings from each participant are shown in this graph. They are grouped together so that you can see all the scores from participants in the random word group and all the scores from the participants in the organised word group. This is slightly better than Graph D because we can just about tell that the random word list led to better recall – but a glance at the means (as in Graph F) shows this effortlessly.

Graph F



◀ This graph shows the mean scores for each group. The findings are immediately obvious, which is the point of using a graph.