

GCSE MATHEMATICS

HANDLING DATA COURSEWORK

Glossary of statistical terms.

When you write up your coursework, try to use the correct terminology to describe various aspects of your work. The following list might help.

N.B. ** denotes elements of higher level GCSE only.

- Bias** Something to be avoided at all costs!
An investigation will typically be biased towards one outcome or another if it is poorly planned. For example, consider investigating whether people think sweets are bad for your teeth and then including only dentists in your investigation! This will lead to bias.
To avoid bias, you must plan your *sample* very carefully.
- Continuous data** Numerical data which can take ANY value within a given range. Such data is usually obtained by measurement and usually has to be rounded up in some way.
E.g. a persons height can, theoretically, take any value within a certain range and would be rounded up to, say, the nearest cm etc.
Compare with *discrete data*.
- Correlation** The word used to describe a relationship between 2 sets of measurements taken from members of the same sample. Plotting a *scatter graph* is the best way of determining which type of correlation the data exhibit, namely: **positive** correlation, **negative** correlation or **no** correlation depending upon the shape of the scatter graph.
See also *line of best fit* and *scatter graphs*.
- Cumulative frequency graph**
A graph used to represent *continuous data*. Used to estimate the *median* of grouped data. Can also be used to estimate the *inter-quartile range*.
- Data** Observations taken from the members of a sample.
E.g. the heights of 30 randomly chosen year 11 girls comprises a set of data.
- Discrete data** Numerical data which can only take certain pre-determined values.
E.g. the number of brothers each person has can only take values 0, 1, 2, ... etc. Compare with *continuous data*.
- ** Frequency density** Used in plotting *histograms*. Equal to frequency \div width of group.
- Frequency polygon** A graph used to represent grouped data. Basically a bar-chart where the mid-points of the tops of each bar are connected with straight lines.

Grouped frequency table

A table showing how many items of data fall within certain groups.

E.g. the following grouped frequency table shows the IQ's of 10 individuals.

IQ	Frequency
70-80	1
80-90	4
90-100	4
100-110	1
110-120	0

****Histogram**

A grouped diagram, similar to a bar-chart, but where we plot **frequency densities** in place of frequencies. The area of each bar (not the height) is proportional to the frequency etc.

Hypothesis

A question or assertion concerning the members of a population.

E.g. *are year 11 girls better at estimating angles than year 7 girls ?*

The whole point of many statistical enquiries is determining whether various hypotheses are true or false.

Inter-quartile range

The difference between the **upper-quartile** and the **lower-quartile**. The width of the interval between which the central 50% of observations lie.

Line of best fit

A straight line drawn on a **scatter graph** and used to estimate one measurement given the value of another measurement.

E.g. on a scatter graph of weights against heights, a line of best fit could be used to estimate a persons height if we knew their weight etc.

Only to be used when the scatter graph shows either **positive** or **negative** correlation.

Lower-quartile

Similar to the **median**. The observation lying one-quarter of the way into the sample. E.g. in a class of 20 students, the lower-quartile test score would be the score of the 5th person from the bottom of the class.

See also **cumulative frequency graph** and **inter-quartile range**.

Mean

The '**average**' obtained by adding up all the data and dividing by the number of observations. The only average which uses every data item.

Median

The '**average**' obtained by selecting the 'middle' observation.

N.B. the data must first be placed in order.

Mode

The 'most common' data item.

Population

The collection of ALL items or people under investigation.

E.g. all secondary school students.

Qualitative data

Data which consists of categories.

E.g. the gender of an individual would be recorded as *female* or *male* etc.

Quantitative data

Numerical data. E.g. heights or weights etc.

Random sample

A sample chosen so that each member of the population has an equal chance of being included in the sample.

Range	The difference between the largest data item and the smallest. Quite a crude measure of dispersion.
Sample	Any subset of the population. E.g. if our population comprises all secondary school students, then a typical sample could simply be 20 secondary school students chosen from year 7 etc.
Scatter graph	A graph used to determine if there is any link between 2 sets of measurements obtained from the same sample. E.g. the heights and weights of a sample of individuals could be plotted in a scatter graph where each point represents the height and weight of a single individual. See also <i>correlation</i> and <i>line of best fit</i> .
**Standard deviation	A numerical quantity obtained from a set of data, related to the mean, which measures how ' <i>spread out</i> ' the data items are. Also known as the <i>mean square deviation from the mean</i> !
Statistics	The science that studies the collection and interpretation of numerical data.
**Stratified sample	The population is divided into different (naturally arising) groups called strata , and then random samples are taken from each group in a way which reflects the underlying population. For example, if the population of a certain town comprised 30% female and 70% male, then a random sample of ten people from the town should consist of 3 females and 7 males etc.
Unbiased	The absence of any <i>bias</i> ! Difficult to achieve completely.
Upper-quartile	Similar to the <i>median</i> . The observation lying three-quarters of the way into the sample. E.g. in a class of 20 students, the upper-quartile test score would be the score of the 15 th person from the bottom of the class. See also <i>cumulative frequency graph</i> and <i>inter-quartile range</i> .