GCSE Maths – Statistics

Graphical Representation of Distributions

Notes

WORKSHEET

Types of data

Primary data is collected by the person carrying out the research. It could be collected by the researcher in a survey or questionnaire.

Secondary data is data collected by somebody else. Data read from a website on the internet or from a book is secondary, as the researcher has not gathered it themselves.

Example: Give an example of primary and secondary data types related to measuring rainfall over a period of days

Measuring rainfall over a period of days.

Primary data*: collecting rainfall using a circular funnel in your garden and recording the depth.*
Secondary data*: looking at the rainfall data for your town online and copying the results.*

Discrete data must take certain values. For example, the number of people in a room cannot be 5.5 - it can only take whole number values such as 5.

Continuous data is numerical data which can take any value in a given range. Age, length and weight are continuous data, as they can be measured to infinite decimal places.

Example: Are the following examples of data discrete or continuous?

1. Shoe size.
Discrete *because shoe size can only take whole or half number values. It cannot take a value such as 5.2.*

2. Time taken to run a race.
Continuous *because the time taken can take any value.*

3. Number of oranges sold in a supermarket.
Discrete *because the number of oranges can only take whole number values. The supermarket cannot sell half an orange.*

4. Weight of each puppy in a litter.
Continuous *because the weight of each puppy can take any value within a reasonable range.*

Qualitative data is non-numerical. Colours are qualitative.
Quantitative data is numerical. Length, age, and time are quantitative.

Example: Give two examples of quantitative and qualitative data

Quantitative: *age, weight (they are numerical values).*
Qualitative: *county, name (they are non-numerical values).*

Grouped data is where data points are given as values within a group. We call this group a **class or category**.

> For example, in the local netball team there are 5 netball players with height in the class $150\text{ cm} < \text{height} \leq 160\text{ cm}$, and 2 netball players with height in the class $160\text{ cm} < \text{height} \leq 170\text{ cm}$.

When a study has a lot of data points, they can be **easier to read** when grouped together. However, the exact data **values are lost** within each group.

Box Plots (Higher Only)

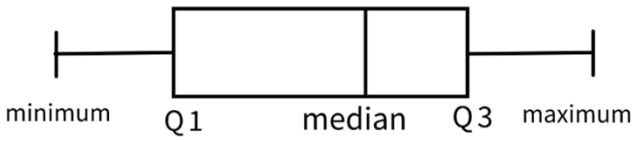
Box plots show the **median** and **quartiles** of a set of data.

When the data is arranged in order of size:

- **Lower quartile (Q1)** is 25% of the way through the data.
- **Median** is the middle point of the data, 50% of the way through (Q2).
- **Upper quartile (Q3)** is 75% of the way through the data.

The median and quartiles are connected by a box.

The minimum and maximum values are also marked on the box plot. They are connected to the box with horizontal lines.



We can use a box plot to calculate the **range** and **interquartile range**, which are two measures of the spread of the data.

The **range** is the difference between the highest and lowest values, including outliers (values very far from the mean). It is a measure of spread.

$$Range = Maximum\ Value - Minimum\ Value$$

The **interquartile range** is the difference between the upper quartile and the lower quartile. It is a measure of spread that reduces or minimises the impact of outliers.

$$Interquartile\ range\ (IQR) = Q3 - Q1$$

When comparing different box plots, you can comment on the **difference** between their medians and interquartile ranges. Comparisons can also been made across the **maximum and minimum values** of the boxplots.

Example: The following data is collected.
Data = {2, 2, 3, 4, 8, 8, 9, 11, 12, 12, 13}
Display the data in a box plot.

Calculate the median and quartiles.

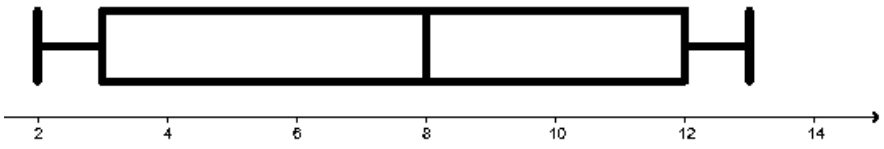
$$Lower\ quartile\ =\ \frac{n+1}{4}th\ term\ =\ \frac{11+1}{4}=3rd\ term\ =\ 3$$

$$Upper\ quartile\ =\ \frac{3(n+1)}{4}th\ term=\frac{3(11+1)}{4}=9th\ term=12$$

$$Median\ =\ Middle\ Value\ =\ 8$$

Construct the box plot.

Draw the axis and label it. Mark points along the axis that spread slightly further than the spread of the data point.
Mark the lowest and highest value on the box plot. Mark the median and quartiles. Connect the median and quartiles with a box. Connect the lowest and highest value with horizontal lines.



More ways to display grouped data

Histograms and cumulative frequency graphs are also used to display grouped data - see the following revision notes for more information:

- *Maths GCSE Revision Notes – Statistics - Grouped Discrete Data and Continuous Data (Higher).*

Graphical Representation of Distributions – Practice Questions

1. State whether the following examples of data are discrete or continuous:

   a) The height of 6 children
   b) The eye colour of 15 dogs
   c) The length of a field of sunflowers
   d) The number of wins for a team in a sports tournament

2. (Higher Only) A class of students took a maths test. The results are summarised below.

   - The lowest mark was 10.
   - The highest mark was 56.
   - The median was 41.
   - The upper quartile was 52.
   - The interquartile range was 35.

   Using the information above, construct a box plot to represent the data.

3. (Higher Only) A group of men and a group of women had their heights measured and recorded. The heights were rounded to the nearest cm and are shown in the table below.

Heights of men (cm)	Heights of women (cm)
165	171
178	155
180	165
175	148
185	160

   a) Draw a box plot for each set of data. Draw both box plots on the same scale.
   b) Using the box plots, make two comparisons about the distributions of the men's heights compared with the distribution of the women's heights.

Worked solutions for the practice questions can be found amongst the worked solutions for the corresponding worksheet file.