Mean, median, mode and range are all different ways of describing certain qualities of a data set.

Median

The median is the middle value of the data set. To find out what it is, we put the values in order from smallest to largest, and then find the middle value (you can remember this by thinking *median* sounds a lot like *medium* which is in the *middle*).

For example, we have recorded the ages of a group of baby pandas (in months). They are 5, 7, 2, 3, 9, 1 and 3.

We find the median by first putting them in order from smallest to largest: 1, 2, 3, 3, 5, 7, 9. Since there are 7 values, we want to find the fourth value, since it will be in the middle, so we have that the median age is 3 months old.

If we have an even number of values, we have two values in the middle. For example, a group of children have all written down how many siblings they have. They have written 1, 3, 0, 0, 2, 4.

We put them in order to get: 0, 0, 1, 2, 3, 4. We now have that 1 and 2 are the two numbers in the middle, so we find the point that is halfway between the two: $\frac{1+2}{2} = 1.5$. So, the median number of siblings that the children have is 1.5 siblings.

Mode

Mode is the most common value (you can remember this because *mode* sounds like *most*). We can find this by writing down each value, and then write down next to it how many times it occurs. For example, a class records their eye colours. They write down: blue, brown, green, brown, brown, blue, hazel, blue, green, brown.

We write them in a table as follows:

Eye colour	Frequency
Blue	3
Brown	4
Green	2
Hazel	1

So, the mode is brown eyes.

In the case where there are two most common values, for example if another person joined the class and they had blue eyes, then we simply have two modes, blue and brown.

Range

The range is the largest value minus the smallest value. For example, if we record the number of miles each person in a small office has to travel to get to work, we could have something like 1.5 miles, 3 miles, 0.75 miles, 4 miles, 10 miles. The range of these values is the largest distance (10 miles) minus the smallest distance (0.75 miles), so the range is 9.25 miles.

Mean

The mean is the average of the data set. The formula for calculating it for a (discrete) data set $\{x_i : i = 1, 2, \dots, n\}$ is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

This essentially means: add up all the values, and divide by the number of values.

For example, we want to find the mean average height of 5 students who are doing a group project together. Their heights are 155cm, 162cm, 161.5cm, 170cm and 157cm. We calculate the mean:

$$\frac{155\text{cm} + 162\text{cm} + 161.5\text{cm} + 170\text{cm} + 157\text{cm}}{5} = 161.1\text{cm}$$

Estimated mean

Sometimes you will be given data that falls into certain categories, instead of giving exact values. Let's look at a frequency table for a supermarket that is researching how much time customers spend shopping in their stores:

Time that a customer spends shopping (minutes)	Frequency
0 minutes ≤ time < 10 minutes	10
10 minutes ≤ time < 20 minutes	3
20 minutes ≤ time < 30 minutes	7
30 minutes ≤ time < 50 minutes	2

Finding a mean from this table seems a bit more difficult. Often people calculate $\frac{10+3+7+2}{4}$, but this doesn't actually tell you anything about the amount of time the customers are shopping for, so be careful not to get confused.

In order to calculate the mean, we find the *estimated* mean, by following these steps:

1. Add another column to the table. Calculate the midpoint of each grouping of times.

Time that a customer spends shopping (minutes)	Midpoint	Frequency
0 minutes ≤ time < 10 minutes	5 minutes	10
10 minutes ≤ time < 20 minutes	15 minutes	3
20 minutes ≤ time < 30 minutes	25 minutes	7
30 minutes ≤ time < 50 minutes	40 minutes	2

2. Add another column to the table. In this one, calculate midpoint multiplied by frequency.

Time that a customer spends shopping (minutes)	Midpoint	Frequency	Midpoint × frequency
0 minutes ≤ time < 10 minutes	5 minutes	10	50 minutes
10 minutes ≤ time < 20 minutes	15 minutes	3	45 minutes
20 minutes ≤ time < 30 minutes	25 minutes	7	175 minutes
30 minutes ≤ time < 50 minutes	40 minutes	2	80 minutes

3. Calculate the total of the frequency column, and the total of the midpoint × frequency column.

Time that a customer spends shopping (minutes)	Midpoint	Frequency	Midpoint × frequency
0 minutes ≤ time < 10 minutes	5 minutes	10	50 minutes
10 minutes ≤ time < 20 minutes	15 minutes	3	45 minutes
20 minutes ≤ time < 30 minutes	25 minutes	7	175 minutes
30 minutes ≤ time < 50 minutes	40 minutes	2	80 minutes
	Total:	22	350 minutes

4. Divide the total of the midpoint × frequency column by the total of the frequency column.

    This gives you the estimated mean.

    350 minutes ÷ 22 = 15.91 minutes

Note: It is quite easy to find the median, mode and range using this table. For the median, we simply find which group the middle person or people are in. In this case they are both in the 10 minutes ≤ time < 20 minutes group, so this is the median. To find the mode, we just look for the group with the highest frequency, so that's 0 minutes ≤ time < 10 minutes. The range of values is the highest possible value (50 minutes) minus the lowest possible value (0 minutes), so it is 50 minutes.

YORK
ST JOHN
UNIVERSITY
Est. 1841

Student Life
Library and
Learning Services

Median, Mode, Range, Mean and Estimated Mean
Study Development Factsheet

When can each one be used?

When doing an exam, you will normally be told what the examiner would like you to find, however there are a lot of times when it would be useful in the real world to calculate some of these values, so it is useful to know when it's appropriate to find each one.

Median can only be found for numerical values. You can assign numerical values to non-numerical data, for example, if you were surveying mood and people chose from options like 'very good', 'good', 'okay', 'not good' and 'really bad', you could rewrite the options to say '1. very good', '2. good', '3. okay' etc, so that you could find the median number. Ethically, in experiments and trials, you must put the numbers in before the participants choose their option, not after.

Median is often not good at showing the average value for data- even though people tend to assume it will be. A good example of this is when companies calculate their gender pay gap. Some companies use the median hourly wage for female employees and compare it with the median hourly wage for male employees, which means they can have a case like this:
Male hourly wages: £14.70, £14.70, £14.74, £15, £22, £30- giving a median wage of £14.87
Female hourly wages: £10, £14.50, £14.50, £15- giving a median wage of £14.50.
This gives a wage gap of around 2.89%, which would be considered a very good gender pay gap, however we can see that 75% of the female employees are paid less than the male employees.

Mode is most useful for data that is given in categories. Things like eye colour, type of pet, or favourite fruit are all good to find the mode of, since we cannot find the mean, median, or range of these data points. We can also find the mode of numerical data. Finding the mode is only useful when you want to know the most common result. Do not use this to find an 'average' result.

Library and Learning Services
Study Development
Email: studydevelopment@yorksj.ac.uk

Est. 1841 | YORK
ST JOHN
UNIVERSITY

Mean is a very good measure of average in most cases. Mean can only be found for numerical data. It can be skewed heavily by anomalous results, in ways that mode and median can't be, however- so it is sometimes best to remove those results before calculating a mean, if you are able to justify the removal.

If we take another look at the gender pay gap example, and use the mean average instead, we now find that male employees have a mean average wage of £18.52, and the female employees mean average wage is £13.50, giving a gender pay gap of 27.12%, which seems much closer to an accurate representation of the figures shown.

Range can only be found for numerical data. It is a way of showing the spread of data, though it is not as useful as some other measures of spread, such as standard deviation, or interquartile range.

Support: Study Development offers workshops, short courses, 1 to 1 and small group tutorials.

- Join a tutorial or workshop on the Study Development tutorial and workshop webpage or search 'YSJ study development tutorials.'
- Access our Study Success resources on the Study Success webpage or search 'YSJ study success.'