

M.K. HOME TUITION

Mathematics Revision Guides
 Level: GCSE Higher Tier

AVERAGES

1, 2, 5, 3, 4, 2, 5, 4, 3, 4 mode 4

10, 12, 15, 19, 22, 26, 28 median 19

12, 15, 19, 27, 32, 35, 41, 47 range $47 - 12 = 35$

$$\text{mean } \frac{12+15+19+27+32+35+41+47}{8} = 28\frac{1}{2}$$

Battery Life (h)	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
Frequency	2	3	4	7	13	11	6	4

modal class 40-44 hours

median class containing the 25th and 26th items 40-44 hours

Life (hrs), x	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	TOTAL
Frequency, f	2	3	4	7	13	11	6	4	50
Midpoint, m	22	27	32	37	42	47	52	57	
Mid. × Freq. mf	44	81	128	259	546	517	312	228	2115

$$\text{estimated mean} = \frac{2115}{50} \text{ or } 42.3 \text{ hours.}$$

Finding averages – the mode, median and mean.

There are three different averages involving **numeric** data that feature heavily in statistics – the mode, the median and the mean.

The **mode** is simply the data item with the highest frequency, i.e. the one which occurs most often. The mode of the set of numbers 1, 2, 5, 3, 4, 2, 5, 4, 3, 4 is thus 4, since it occurs three times in the list.

The mode is particularly suited when ordering items for a shop, for instance - think mode = fashion.

Alone of the three averages, the mode can apply to non-numeric data as well.

As an example, a sports shop's replica football shirt sales by size on a certain week were 12 small, 15 medium, 19 large and 16 extra large. The modal size in terms of sales that week was 'large'.

The **median** can be worked out by sorting the data items in order, and finding the middle number.

The median of the (sorted) set of numbers 10, 12, 15, 19, 22, 26, 28 is therefore 19, because there are 7 numbers in all, and 19 is the fourth one, having exactly as many below it as above it.

Note: If the sample size is even, there will be two 'middle numbers'. The median is then found by adding the two middle numbers and halving the result.

Thus, the median of the set 12, 15, 19, 27, 32, 35, 41, 47 is taken as half of $(27 + 32)$ or $29\frac{1}{2}$.

More succinctly, if the list contains n data items (where n is odd), the median's position is $\frac{1}{2}(n + 1)$. For even n , the two 'middle numbers' are in positions $\frac{1}{2}(n)$ and $\frac{1}{2}(n) + 1$.

The **mean** can be worked out by adding the data items and dividing the result by the size of the sample. Thus, the mean of the set 12, 15, 19, 27, 32, 35, 41, 47 can be found by summing the numbers. The sum is 228 and there are 8 numbers in all, so the mean is equal to $228 \div 8 = 28\frac{1}{2}$.

One more feature of a set of data is the **range** – this is simply the largest number in the group minus the smallest one.

Example (1): Dixie received the following scores over an 11-week period on *Strict Dance*:

27, 35, 34, 32, 35, 33, 36, 38, 37, 38, 35

Find her modal, median and mean scores, as well as the range.

Mode.

The mode is simply the data item with the highest frequency, and Dixie had received a score of 35 more often than any other. Her modal score is therefore 35.

Median.

We place the scores in numerical order as 27, 32, 33, 34, 35, **35**, 35, 36, 37, 38, 38. The middle number is the 6th one in the list, having exactly as many below it as above it. Hence Dixie's median score is 35.

Mean.

The total of all the 11 scores is 380, so the mean is equal to $380 \div 11 = 34.5$.

Range.

The highest score is 38, and the lowest is 27, so the range is $38 - 27 = 11$.

Example (2): A survey was carried out to find out how many portions of fruit and vegetables each pupil in a class of 32 was eating per day. The results were given in the table below.

Portions	0	1	2	3	4	5	6	7	8
No. of pupils	2	3	6	10	5	2	1	0	3

Find the mode, median, mean and range of the number of portions of fruit eaten.

Mode.

The mode is simply the data item with the highest frequency. More pupils (namely 10) have 3 portions of fruit and vegetables a day than any number of portions, so 3 is the mode.

Median.

To find the median, we could write out the data in full from the table as a list of 2 '0's, 3 '1's, 6 '2's and so forth:

0, 0, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 6, 8, 8, 8

There are 32 numbers in all, i.e. $n = 32$, so the positions of the 'middle numbers' are $\frac{1}{2}(n)$ or 16, and $\frac{1}{2}(n) + 1$ or 17.

There are two middle numbers in the list - the 16th and 17th.
 Each is a 3, so 3 is the median.

This method was a little long-winded given the fact that the data was already grouped and sorted for us.

We could reason as follows;

- Pupils 1-2 (2 of them) had no fruit and vegetables
- Pupils 3-5 (3 of them) had one portion of fruit and vegetables
- Pupils 6-11 (6 of them) had two portions
- Pupils 12-21 (10 of them) had three portions, and so on..

The 16th and 17th pupils both fall into the '3 portions' category, so the median is 3.

Mean.

To find the mean, we need an extra row for subtotals and an extra column for totals as shown below.

Portions (Number) x	0	1	2	3	4	5	6	7	8	TOTAL
Pupils (Frequency) f	2	3	6	10	5	2	1	0	3	32
Portions × Pupils, xf	0	3	12	30	20	10	6	0	24	105

(Note: some books have the tables of data shown with columns and rows interchanged. The preference for the row format is purely arbitrary.)

We have multiplied the number of portions (x) by the number of pupils (f) to work out the total number of portions of fruit and vegetables eaten by the class.

As an example, there are 5 pupils who eat 4 portions of fruit and vegetables. They therefore contribute 5×4 or 20 portions towards the total.

The total number of portions of fruit and vegetables is 105, and so the mean is $\frac{105}{32}$ or 3.3.

The range of the number of portions of fruit and vegetables eaten is $8 - 0$ or 8.

Example (3): A football club has a squad of 23 players, and each player is currently on one of five different weekly wage structures.

Wage in £K	11	12	14	17	21
No. of players	4	4	6	5	4

Find the mode, median, mean and range of the players' wages.

Mode: The modal wage is £14 K, as more players are on that wage than any other.

Median: Eight players earn under £14 K, and the 9th to the 14th players earn £14 K.

There are 23 players in all, so the middle player's position is $\frac{1}{2}(23+1)$ or 12. This 12th player in the list is in the £14 K wage band, and so the median wage is also £14 K.

Mean: We need an extra row and extra column for subtotals and totals as shown below.

Wage in £K, x	11	12	14	17	21	TOTAL
No. of players, f	4	4	6	5	4	23
Sub-total, xf	44	48	84	85	84	345

There are 23 players earning a total weekly wage of £345 K, so the mean wage = $\frac{345}{23} = 15$ or £15 K.

Finally, the range of the footballers' wages is £21 K - £11 K, or £10 K.

Example (3a): The football club in Example 3 has lost one of their players on a wage of £17 K through retirement. To replace him, they have signed a Brazilian, Rondinho, and put him on a £155 K weekly wage. How will Rondinho's signing affect each average wage, and the range ?

The table of players' wages will now look like this:

Wage in £K	11	12	14	17	21	155
No. of players	4	4	6	4	4	1

Mode: There are still more players on £14 K than on any other wage, so the mode is unaffected. (Had one of the players on a weekly wage of £14 K, there would have been **two** modal weekly wages, £14 K and £17 K.)

Median: The 9th to the 14th players still earn £14 K, and so the median is still unaffected at £14 K.

Wage in £K, x	11	12	14	17	21	155	TOTAL
No. of players, f	4	4	6	4	4	1	23
Sub-total, xf	44	48	84	68	84	155	483

The 23 players are now earning a total weekly wage of £483 K, so the mean wage = $\frac{483}{23}$ or £21 K.

This is a considerable difference from the earlier value of £15 K.

The range of the footballers' wages is now £155 K - £11 K, or £144 K.

Example (2a) showed how different averages were differently affected by a value outside the 'normal' range, or an **outlier** – here it was the Brazilian footballer's wage.

The mode is not generally affected by outliers, and neither is the median. The mean and the range, however, are more sensitive, which is why care needs to be taken in selecting the most appropriate average.

Example (4): The heights, x , of 25 sunflower plants grown in the same garden from the same seed packet were measured (in centimetres, rounded to the nearest cm) on one day and sorted in order.

183, 189, 192, 193, 194, 196, 202, 203, 204, 207, 207, 208, 209,
212, 214, 215, 216, 216, 218, 219, 219, 222, 225, 228, 234.

Find the mode, median, mean and range.

Mode: There are three heights vying for the mode here – 207, 216 and 219 cm. These values each occur twice whilst all the others occur only once.

Median: There are 25 plants here, i.e. $n = 25$. The middle number's position in the sorted list is $\frac{1}{2}(25+1)$ or 13. The 13th value in the list is 209, so the median height is 209 cm.

Mean: The sum of the heights (in cm) is 5225, so the mean is $\frac{5225}{25}$ cm or 209 cm.

This time, it seems that the mode is unsuitable because the range of values is large, the sample fairly small, and there is no one highly popular value. The median and mean are both suitable, because there are no outlying values and the distribution is uniform.

Range: The range is 234 – 183 cm, or 51cm.

Example (5): (Non-cricket fans: there are two innings in a Test match, and a duck is a score of zero.)

Ashok batted in five Test cricket matches. After three Tests and six innings, his batting average (i.e. the mean) was 66.

He had a disastrous fourth Test, scoring a duck in his first innings and only single figures in his second, which brought his batting average down to 50.

After completing his fifth Test, he returned to top form and brought his batting average back up to what it was at the end of the third Test, scoring 96 in his first innings.

Calculate Ashok's second innings scores in the fourth and fifth Tests.

Ashok's mean score (i.e. his batting average) was 66 runs per innings after three Tests, so his total run score was $66 \times 6 = 396$ runs after six innings.

After eight innings, his batting average fell to 50, implying a total run score of 400 runs. Hence he had only scored $400 - 396$, or 4, runs in the fourth Test in total. As his first innings was a duck (zero), he had scored **4 runs** in his second innings in the fourth Test.

After five tests and ten innings, his batting average was restored to 66, implying a total run score of $66 \times 10 = 660$ runs in all ten innings.

Carrying on from his total of 400 after eight innings, Ashok had scored $660 - 400$, or 260 runs overall in the fifth Test. His first innings was 96, so he had scored $260 - 96$, or **164 runs** in his second innings in that Test.

Data Grouped by Class Interval.

When we have a large set of data, it is often more convenient to arrange it into class interval groups.

The last example featured the individual heights of 25 sunflower plants, but grouping them helps illustrate the distribution more plainly. The following example has the same data, but grouped by **class intervals** of 10cm height.

Example (6): Find the modal and median classes from the grouped heights of sunflower plants below:

Height Range (cm)	180-189	190-199	200-209	210-219	220-229	230-239
Frequency	2	4	7	8	3	1

We can no longer find the actual mode or the median, but we can still find the class intervals in which they lie.

This is a different case from Examples (1) and (2), because, although they featured grouped data, there was no loss of actual values.

The **modal class** is 210-219 cm, since it has the highest frequency.

To find the **median class**, we now look for the class interval where the 13th data item occurs. The 13th data item occurs in the 200-209 cm class, so 200-209 cm is the median class.

Estimating the mean from grouped data.

Estimating a mean from grouped data needs a little more work. Because we no longer have actual data, the best we can do is assume that the data items are distributed evenly within their class intervals.

We therefore assume that all the data items take a class midpoint value, taken to be the mean of the end values of the class interval.

Sometimes, due to the nature of the class groupings, the midpoint value would not seem 'exact', and have a messy ".5" in it. This is especially true for discrete or rounded data, where there are gaps between class intervals.

Thus, if a class interval of rounded data is quoted as '30-39', the effective class interval is 29.5 to 39.5, and the midpoint would be 34.5 and not 35.

Similarly a range quoted as '61-80' would have a midpoint of 70.5, not 70.

Examination paper questions however tend to assume **continuous** data – see Examples 8 and 9.

Example (6a): Estimate the mean from the grouped heights of sunflower plants below.
(Values have been rounded to the nearest cm.)

Height Range (cm) (Class Interval)	180-189	190-199	200-209	210-219	220-229	230-239
Frequency	2	4	7	8	3	1

Notice the gaps between the class intervals here at 189 to 190, 199 to 200 and all the others. In other words, the end of one class interval is not the same as the start of the next one.

The '180-189' class interval is actually '179.5 - 189.5', the '190-199' actually '189.5 - 199.5', and so forth, due to the use of rounded discrete data.

Firstly we work out the class interval midpoints and place them in another row:

Height Range (cm) (Class Interval)	180-189	190-199	200-209	210-219	220-229	230-239
Frequency	2	4	7	8	3	1
Midpoint	184.5	194.5	204.5	214.5	224.5	234.5

Finally we add a subtotals row and a totals column:

Height Range (cm) (Class Interval)	180-189	190-199	200-209	210-219	220-229	230-239	TOTAL
Frequency, f	2	4	7	8	3	1	25
Midpoint, m	184.5	194.5	204.5	214.5	224.5	234.5	
Midpoint × Frequency, mf	369	778	1431.5	1716	673.5	234.5	5202.5

We then sum the “Midpoint × Frequency” values and divide them by the sample size.

The estimated mean is $\frac{5202.5}{25}$ cm, or 208.1 cm.

Note how this is slightly different from the actual mean of 209 cm found in Example (4).

Example(7): The lifetimes (in hours, rounded to the nearest hour) of 50 Alko batteries were tested under controlled conditions in a simulation lab.

The results were as follows:

Life (hrs)	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
Frequency	2	3	4	7	13	11	6	4

Note that this is discrete data, hence the gaps between 24 and 25, 29 and 30 and so on.

Find the modal and median classes, and estimate the mean.

The **modal class** is 40-44 hours lifetime, as it has the highest frequency.

The **median class** is the one containing the 25th and 26th items.
 This is the 40-44 hours class, actually containing the 17th-29th items.

To estimate the **mean**, we firstly find the class midpoints:

Life (hrs), x	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
Frequency, f	2	3	4	7	13	11	6	4
Midpoint, m	22	27	32	37	42	47	52	57

Next, we add the subtotals row and totals columns:

Life (hrs), x	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	TOTAL
Frequency, f	2	3	4	7	13	11	6	4	50
Midpoint, m	22	27	32	37	42	47	52	57	
Mid. \times Freq. , mf	44	81	128	259	546	517	312	228	2115

The estimated mean is $\frac{2115}{50}$ or 42.3 hours.

If the source data is taken to be continuous, then the class interval midpoints are easier to find. The vast majority of examination questions assume continuous data, as in the example below.

Note how the class intervals are delimited by inequality symbols.

Example(8): An electrical retailer has analysed the money spent by customers in the first 500 transactions in the January sales, excluding purchases greater than £1000.

The results were tabulated as follows:

Spending, x , in £	$x < 50$	$50 \leq x < 100$	$100 \leq x < 200$	$200 \leq x < 400$	$400 \leq x < 600$	$600 \leq x \leq 1000$
Frequency, f	70	110	100	80	60	80

Estimate the mean amount of money spent per customer transaction by completing the table below:

Spending, x , in £	$x < 50$	$50 \leq x < 100$	$100 \leq x < 200$	$200 \leq x < 400$	$400 \leq x < 600$	$600 \leq x \leq 1000$
Frequency, f	70	110	100	80	60	80
Class midpoint, m	25	75	150			
Mid. \times Freq., mf						

The completed data table looks like this:

Spending, x , in £	$x < 50$	$50 \leq x < 100$	$100 \leq x < 200$	$200 \leq x < 400$	$400 \leq x < 600$	$600 \leq x \leq 1000$	TOTAL
Frequency, f	70	110	100	80	60	80	500
Class midpoint, m	25	75	150	300	500	800	
Mid. \times Freq., mf	1750	8250	15000	24000	30000	64000	143000

The estimated mean spending is therefore $\pounds \frac{143000}{500}$, or £286 per customer.

Example (9): Sam is a keen gardener and his two most productive apple trees are a Cox and a Gala. The details of his harvest last autumn are as below.

Cox	
Weight of apple, w , (g)	Frequency, f
$100 \leq w < 110$	25
$110 \leq w < 120$	60
$120 \leq w < 130$	90
$130 \leq w < 140$	180
$140 \leq w < 150$	140
$150 \leq w < 160$	80
$160 \leq w < 170$	25
TOTAL	600

Gala	
Weight of apple, w , (g)	Frequency, f
$120 \leq w < 130$	15
$130 \leq w < 140$	30
$140 \leq w < 150$	140
$150 \leq w < 160$	170
$160 \leq w < 170$	120
$170 \leq w < 180$	25
TOTAL	500

- i) Write down the modal weight class for each variety of apple, and also the weight class in which the median lies.
- ii) Estimate the total yield of apples (in kg), and hence the mean weight of an apple, from each tree.
- iii) Sam reckons that the Cox tree had produced a heavier crop than the Gala, even though the individual apples from the Cox tree were of lower weight.

Comment on Sam's statement, giving supporting evidence or otherwise.

i) The modal weight class for the Cox apples is the $130 \leq w < 140$ class, i.e. apples weighing between 130g and 140g each. The corresponding class for the Gala apples is the $150 \leq w < 160$ class, or apples in the 150g-160g weight range.

ii) The total number of Cox apples is 600, so the median class interval is the one containing the 300th and 301st apples. This is the " $130 \leq w < 140$ " weight interval, as it contains the 176th to 355th apples.

By similar reckoning, the total number of Gala apples is 500, so the median class includes the 250th and 251st apples. This is the " $150 \leq w < 160$ " weight interval, containing the 186th to 355th apples.

Note how the modal and median classes are the same for both varieties of apple.

iii) We need to add class midpoint and sub-total columns in order to estimate the total crop weight of the apples from each tree, and thus the mean weight of a single apple.

Cox			
Weight of apple, w , (g)	Frequency, f	Class midpoint, m	Mid. \times Freq. , mf
$100 \leq w < 110$	25	105	2625
$110 \leq w < 120$	60	115	6900
$120 \leq w < 130$	90	125	11250
$130 \leq w < 140$	180	135	24300
$140 \leq w < 150$	140	145	20300
$150 \leq w < 160$	80	155	12400
$160 \leq w < 170$	25	165	4125
TOTAL	600		81900

The estimated total yield of Cox apples is **81.9 kg**, and as 600 apples were harvested in all, the

estimated mean weight of a single Cox apple is $\frac{81900}{600}$ g or **136.5 g**.

Gala			
Weight of apple, w , (g)	Frequency, f	Class midpoint, m	Mid. \times Freq. , mf
$120 \leq w < 130$	15	125	1875
$130 \leq w < 140$	30	135	4050
$140 \leq w < 150$	140	145	20300
$150 \leq w < 160$	170	155	26350
$160 \leq w < 170$	120	165	19800
$170 \leq w < 180$	25	175	4375
TOTAL	500		76750

The estimated total yield of Gala apples is **76.75 kg**, and as 500 apples were harvested in all, the

estimated mean weight of a single Gala apple is $\frac{76750}{500}$ g or **153.5 g**.

iii) From the results, the individual mean weight of a Gala apple is 17g more than that of a Cox, but as Sam harvested more Cox apples, the total estimated yield was 5 kg greater for the Cox apples than for the Gala apples. Sam's assertions were therefore correct.

More on Median, Ranges and Quartiles.

We have seen how to calculate the median and overall range for a list of numeric data. Sometimes it is useful to find the range within the “middle half” of the data.

Recall : the median is the middle number in the sorted data list.

The median of the (sorted) set of numbers 10, 12, 15, 19, 22, 26, 28 is therefore 19, because there are 7 numbers in all, and 19 is the fourth one, having exactly as many below it as above it.

Note: If the sample size is even, there will be two ‘middle numbers’. The median is then found by adding the two middle numbers and halving the result.

Thus, the median of the set 12, 15, 19, 27, 32, 35, 41, 47 is taken as half of $(27 + 32)$ or $29\frac{1}{2}$.

More succinctly, if the list contains n data items (where n is odd), the median’s position is $\frac{1}{2}(n + 1)$. For even n , the two ‘middle numbers’ are in positions $\frac{1}{2}(n)$ and $\frac{1}{2}(n) + 1$.

We can extend the idea to include **quartiles**, which are entries one-quarter of the way along the list (the **lower quartile**) and three-quarters of the way along the list (the **upper quartile**).

The difference between the upper and lower quartiles is known as the **inter-quartile range**.

One method of finding quartiles is to find the median, split the data into two halves to either side of it, and then find the median of each half.

The lower quartile is the median of the lower half of the data; the upper quartile is the median of the upper half of the data. If the number of data items is odd, and there is an actual median, then this median is not included in either half.

Note : There seems to be an inconsistency here. Some schools of thought reckon the positions of the quartiles as $\frac{1}{4}(n + 1)$ and $\frac{3}{4}(n + 1)$ where there are n data items. **Check with your exam board !**

In actual exam questions, the values of n are usually chosen to avoid such difficulties. Typical examples are a list of 15 data items, where the lower quartile, median and upper quartile are taken to be the 4th, 8th and 12th items in the list.

Example (10):

i) Find the lower quartile, median and upper quartile from this list:
14, 16, 19, 22, 25, 29, 32, 34, 37, 41.

ii) Find the overall range and the inter-quartile range.

i) There are 10 numbers in the list, i.e. $n = 10$; the position of the median is $\frac{1}{2}(n + 1)$ or $5\frac{1}{2}$. There are two “middle numbers”, 25 and 29, so the median is halfway between 25 and 29, or 27.

After chopping the data in two, the lower half contains 14, 16, 19, 22, 25 and the upper half contains 29, 32, 34, 37, 41. Note that the calculated median of 27 is not included in either half.

The lower quartile is the median of 14, 16, 19, 22, 25, or 19.

The upper quartile is the median of 29, 32, 34, 37, 41, or 34.

ii) The overall range is $41 - 14$, i.e. 27 ; the inter-quartile range is $34 - 19$, or 15.

Example (11): i) Find the lower quartile, median and upper quartile from this sorted set of numbers; 10, 12, 14, 15, 19, 22, 26, 28, 30, 33, 34. (Exclude the actual median from the two data halves)

ii) Find the overall range and the inter-quartile range.

i) The median is the middle number in the list, namely 22.

The lower half of the data contains 10, 12, 14, 15 and 19.

The lower quartile is the median of *that* set – i.e. 14.

The upper half of the data contains 26, 28, 30, 33 and 34.

The upper quartile is the median of *that* set – i.e. 30.

ii) The overall range is $34 - 10$, or 24.

The inter-quartile range is $30 - 14$, or 16.

Example (12): Legolamb and Gimlet played 11 rounds of video games, and their tally of enemy kills was as follows:

Legolamb: 40, 44, 42, 44, 45, 47, 46, 47, 51, 49, 52.

Gimlet : 37, 35, 41, 39, 33, 38, 44, 50, 46, 57, 53

Compare the distributions of their scores, using the median, quartiles and ranges.

(Exclude the actual median when ‘halving’ the data to obtain the quartiles).

i) We need to sort the scores in ascending order first:

Legolamb: 40, 42, 44, 44, 45, 46, 47, 47, 49, 51, 52.

Gimlet : 33, 35, 37, 38, 39, 41, 44, 46, 50, 53, 57

The median scores were 46 for Legolamb and 41 for Gimlet, so Legolamb had a better median score.

Legolamb also had a better lower quartile score of 44 against Gimlet’s 37, and the same was true for the minimum score, i.e. 40 against 33.

On the other hand, Gimlet’s upper quartile score of 50 was higher than Legolamb’s 49, and his maximum score of 57 was well up on the 52 achieved by Legolamb.

The ranges also tell a contrasting story. For Legolamb, the inter-quartile range was 5, as opposed to Gimlet’s 13, so Legolamb’s performances were more consistent in that area. The same applied for the total ranges of their scores, namely 12 for Legolamb and 24 for Gimlet.

In summary. we can deduce that

a) Legolamb was more consistent in his scoring than Gimlet, and

b) Gimlet was able to score several higher individual rounds than Legolamb.