

32.1 Averages and Ranges

Averages and ranges are useful ways of summarising a data set — by looking at them you can get a good idea of the data set as a whole without needing to see every piece of data within it.

Learning Objectives — Spec Ref 94:

- Find the mode, median, mean and range from raw data.
- Find the mode, median, mean and range from a frequency table.

An **average** is a way of **representing** a whole set of data using a **single value** — it is the **central** or **typical** value within the data set. The **mode**, **median** and **mean** are three common averages that are used.

- The **mode** (or modal value) is the **most common value** — the value that appears in the data set **more often** than any other. Find the mode by counting **how many times** each value appears in the data set.
- The **median** is the **middle value** — the value that is in the **middle** of the data set when all the numbers are put in **ascending order**. For a data set containing n pieces of data, the median is the **value** in position $(n + 1) \div 2$. If n is an **even number**, the median will be **halfway** between two values.
- The **mean** is the **total** of all the values divided by the **number** of values. To find the mean, **add up** all the values in the data set and **divide** by the number of values.

The **range** is the difference between the **largest value** and the **smallest value** in a data set — it tells you how **spread out** the values are within a data set. Data sets with a **small range** are more **consistent** than those with a **large range** — this means there is less variation in the values.

Averages and ranges are useful because they allow us to **compare** one set of data to another.

Example 1

20 customers each gave a restaurant a mark out of 10. Their marks are shown on the right.

Find: a) the modal mark b) the median mark
c) the mean mark d) the range

3	7	4	8	3	7	5	2	8	9
9	6	1	3	4	5	6	5	7	7

1. Put the numbers in order first.

From smallest to largest, the numbers are:

1, 2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 9, 9

2. The mode is the most common number.

a) 4 customers gave a mark of 7, so the mode is 7.

3. There are an even number of values, so the median will be halfway between the two middle numbers.

b) There are 20 values, so the median is in position $(20 + 1) \div 2 = 10.5$, which is halfway between the 10th value = 5 and 11th value = 6.

So the median = $\frac{5+6}{2} = 5.5$

4. Divide the total of the values by the number of values to find the mean.

c) Total = $1 + 2 + 3 + 3 + 3 + 3 + 4 + 4 + 5 + 5 + 5 + 5 + 6 + 6 + 7 + 7 + 7 + 7 + 8 + 8 + 9 + 9 = 109$
Mean = $109 \div 20 = 5.45$

5. Subtract the smallest value from the largest.

d) Range = $9 - 1 = 8$

Exercise 1

- Q1 For the following sets of data: (i) Find the mode. (ii) Find the range.
a) 6, 9, 2, 7, 7, 6, 5, 9, 6 b) 16, 8, 12, 13, 13, 8, 8, 17 c) 8.2, 8.1, 8.1, 8.2, 8.1, 8.2, 8.2
- Q2 Find the median of the following sets of data.
a) 3, 3, 3, 3, 3, 4, 4, 4, 4, 4 b) 2, 4, 7, 1, 5, 9, 2, 7, 8, 0 c) 5.85, 6.96, 2.04, 7.45, 6.9, 7.8
- Q3 The times (to the nearest second) of athletes running the 400 m hurdles are:
78, 78, 84, 81, 90, 79, 84, 78, 95
a) Find the range of times taken to run the 400 m hurdles. b) Find the median time.
- Q4 Nine students score the following marks on a test: 34, 67, 86, 58, 51, 52, 71, 65, 58
Find the mean score. Give your answer to 3 significant figures.
- Q5 Abdul counts the number of crisps in 28 packs. His results are shown below.
a) Find the modal number of crisps in a pack.
b) Calculate: (i) the range of the data,
(ii) the median number of crisps in a pack
(iii) the mean number of crisps in a pack to 3 s.f.
- | | | | | | | |
|----|----|----|----|----|----|----|
| 12 | 20 | 21 | 15 | 18 | 20 | 21 |
| 20 | 15 | 9 | 22 | 16 | 18 | 19 |
| 20 | 18 | 13 | 15 | 18 | 20 | 17 |
| 16 | 15 | 16 | 16 | 18 | 21 | 20 |
- Q6 Look at this data set: 6, 5, 8, 8, 5, ?
a) If the range was 6, what are the two possible values for the missing number?
b) If the mean was 7, find the missing value.
- | | | | | |
|------|------|------|------|------|
| 1.68 | 1.45 | 1.70 | 1.30 | 1.72 |
| 1.80 | 1.29 | 1.40 | 1.42 | 1.60 |
| 1.65 | 1.75 | 1.67 | 1.69 | 1.72 |
| 1.72 | 1.63 | 1.63 | 1.78 | 1.70 |
| 1.50 | 1.65 | 1.40 | 1.36 | 1.69 |
- Q7 The heights (in metres) of a class of students are given on the right. Which is greater, the mean or the median height of the students?



Large data sets can be put into **frequency tables** to make them **more manageable**. The frequency is the **number of times** a particular value appears within a data set. When **calculating averages** from frequency tables you treat the data the **same way** as if it had all been written out normally.

Example 2

This frequency table shows the number of mobile phones owned by a group of people.

Number of mobile phones	0	1	2	3	4
Frequency	4	8	5	2	1

a) Find the modal number of mobile phones.

Find the number of mobile phones with the highest frequency.

Mode = 1 mobile phone

Tip: The raw data is:
0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
1, 1, 2, 2, 2, 2, 2, 3, 3, 4.

b) What is the median number of mobile phones?

- Work out the total frequency and the position of the median.
- The data values in the 10th and 11th positions are both 1.

Total frequency = $4 + 8 + 5 + 2 + 1 = 20$

The median is the position $(20 + 1) \div 2 = 10.5$

It's halfway between the 10th value = 1 and

11th value = 1, so the median = $\frac{1+1}{2} = 1$ mobile phone

c) Find the range for this data.

Subtract the smallest value from the largest value.

$$\text{Range} = 4 - 0 = 4$$

d) Find the mean number of mobile phones owned by a person.

1. Work out the total number of mobile phones by multiplying the number of mobile phones by the frequency of each column.

Number of mobile phones	0	1	2	3	4
Frequency	4	8	5	2	1
Phones \times frequency	0	8	10	6	4

2. Divide the total number of mobile phones by the total number of people.

$$\text{Total number of phones} = 0 + 8 + 10 + 6 + 4 = 28$$

$$\text{Mean} = 28 \div 20 = 1.4 \text{ mobile phones}$$

Exercise 2

- Q1 The table shows the number of people living in each of 30 houses.

- Write down the modal number of people per house.
- Find the median number of people per house.
- Calculate the mean number of people per house.
- Work out the range of the data.

Number of people	Frequency
1	4
2	5
3	8
4	10
5	3

- Q2 This table shows the number of goals scored one week by 18 teams in the premier division.

- Write down the modal number of goals.
- Find the mean number of goals.
Give your answer to 3 significant figures.
- The mean number of goals scored in the same week last year was 2.4.
How do these results compare?

Number of goals	0	1	2	3	4	5
Number of teams	1	3	4	5	3	2

- Q3 A student wrote down the temperature in his garden in Aberdeen every day at noon during the summer. His results are shown in the table.

- Find the median noon temperature.
- Find the mean temperature.
- The mean midday temperature during the summer in the UK is approximately 18.5°C . What does this suggest about the temperature in the student's garden?

Temperature ($^{\circ}\text{C}$)	Frequency
16	18
17	27
18	7
19	15
20	12
21	11

- Q4 A survey asked 200 people, 'How many televisions do you own?'. The results are shown in this table.

- Show that $p + q = 109$.
- The mean number of televisions is 1.88. Show that $2p + 3q = 240$.
- Find the number of people who own 2 televisions and 3 televisions.

No. of Televisions	1	2	3	4	5
Frequency	77	p	q	11	3



Measures of Spread — Interquartile Range

Learning Objective — Spec Ref 94:

Find the upper and lower quartiles and the interquartile range for a data set.

The **lower quartile**, the **median** and the **upper quartile** are values that **divide** a data set into **four groups** of **equal size**. The quartiles are positioned **25%**, **50%** and **75%** of the way through the data. For a data set in **ascending order** with n values, you can work out the position of the quartiles using these formulas:

- Lower quartile (Q_1) — $Q_1 = (n + 1) \div 4$
- Median (Q_2) — $Q_2 = (n + 1) \div 2$
- Upper quartile (Q_3) — $Q_3 = 3(n + 1) \div 4$

The lower and upper quartiles can be used to work out the **interquartile range**. This gives you the range of the **middle 50%** of data and can be used to measure how **dispersed** (spread out) the data is.

To find the interquartile range, **subtract** the lower quartile (Q_1) from the upper quartile (Q_3): $IQR = Q_3 - Q_1$

Example 3

The data on the right shows the goals scored by Team A at each hockey game in a season.

Game	1	2	3	4	5	6	7	8	9
Goals scored	2	4	0	3	4	2	2	3	2

a) Calculate the interquartile range for this data.

- Put the data in ascending order.
- Work out the positions of the lower and upper quartiles.
- Q_1 will be halfway between the 2nd and 3rd values. Q_3 will be halfway between the 7th and 8th values.
- Subtract to find the interquartile range.

0 2 2 2 2 3 3 4 4

Q_1 position = $(9 + 1) \div 4 = 2.5$

Q_3 position = $3(9 + 1) \div 4 = 7.5$

2nd value = 2, 3rd value = 2, so $Q_1 = 2$

7th value = 3, 8th value = 4,

so $Q_3 = (3 + 4) \div 2 = 3.5$

$IQR = Q_3 - Q_1 = 3.5 - 2 = 1.5$

b) Team B scores the same total number of goals as Team A, with an interquartile range of 0.5. Which team scored the most consistent number of goals per game?

The smaller the interquartile range, the less spread out and more consistent the data values are.

Team B — their interquartile range is much smaller than for Team A, so their scores were more consistent.

Exercise 3

Q1 Calculate the interquartile range for each of the following data sets.

- a) 8, 9, 9, 9, 10, 10, 12, 15, 16, 17, 19
- b) 80, 70, 34, 21, 21, 56, 75, 89, 84, 20, 17, 45, 87
- c) 1.5, 1.5, 1.3, 1.4, 1.6, 1.8, 1.2
- d) 1, 9, 3, 9, 3, 4, 5, 6, 9, 0, 1, 9, 9, 5, 9, 2, 5

Q2 The table shows the number of spots on 79 ladybirds. Find the interquartile range for the data.

No. of spots	2	3	4	5	6	7	8	9	10
Frequency	3	16	9	18	9	6	9	7	2

- Q3 The list on the right shows the exam results for the students in class 3A.
- Only students with a mark in the upper quartile of the class results passed the exam. What was the pass mark for this exam?
 - Calculate how many people passed the exam.
 - Find the interquartile range for this data.
 - Class 3B's results for the same exam have an interquartile range of 20. Comment on how the spread of the results for the two classes differs.

30	36	89	92	76
20	57	89	23	55
56	56	98	35	20
86	38	24	13	90
54	67	67	34	78
72	53	88	24	40
76	20	24		

Choosing the Right Average and Range

Learning Objectives — Spec Ref 94:

- Choose which measure of central tendency or spread to use.
- Understand the advantages and disadvantages of each one.

Averages are a good way to **summarise** and **represent** data, but using the **wrong average** can lead to data being **misinterpreted**. The mode, median and mean averages all have different **advantages** and **disadvantages**, which means that some are better suited to certain data sets than others. For example, some averages are more affected by extreme values (known as **outliers**) than others.

	Advantages	Disadvantages
Mode	<p>It's easy to find.</p> <p>It doesn't get distorted by outliers.</p> <p>Can be used for non-numeric data.</p> <p>It's always a value in the data set.</p>	<p>It doesn't always exist, or there are several modes.</p> <p>Not always a good representation of the data — it doesn't use all the data values.</p>
Median	<p>It's easy to find for ungrouped data.</p> <p>It doesn't get distorted by outliers.</p>	<p>Not always a good representation of the data — it doesn't use all the data values.</p> <p>It isn't always a value in the data set.</p>
Mean	<p>It is usually the most representative average — it uses all the data values.</p>	<p>It can be distorted by outliers.</p> <p>It isn't always a value in the data set.</p>

Exercise 4

- Q1 The clothes sizes of 20 women are shown on the left.
- Find the mode.
 - Find the median clothes size.
 - Calculate the mean clothes size.

18	4	10	10	16
12	14	14	6	18
14	12	12	8	16
14	12	12	14	14

- Suggest why the mode might represent this data better than the mean or median.

Q2 The ages of the 30 audience members at a McBeetle concert are shown on the right.

60	12	17	60	14	16	29	15	12	17
60	60	16	17	16	14	13	14	10	16
14	17	19	17	18	19	60	16	19	16

- a) (i) Find the modal age.
 (ii) Work out the median age.
 (iii) Calculate the mean age. Give your answer to 3 significant figures.
- b) A slightly embarrassed 29-year-old says, 'The mean age is the most representative of the ages of the people at the concert.' Do you agree? Explain your answer.

Q3 A company asked 190 people to test their latest wrinkle cream and give it a mark out of 10. The table below shows their results.

Mark	1	2	3	4	5	6	7	8	9	10
Frequency	31	34	35	34	4	6	36	7	2	1

- a) Find the mode, median and mean for the data.
 Give your answers to 3 significant figures where necessary.
- b) (i) The company claims, 'On average, people gave our product 7 out of 10'. Which average has it used in its claim?
 (ii) Do you think this average represents the data well? Explain your answer.

The **range** and **interquartile range** also have advantages and disadvantages, but the interquartile range is generally **more reliable** as it isn't affected by **outliers**.

	Advantages	Disadvantages
Range	It's easy to find.	It can be distorted by outliers.
Interquartile range	It doesn't get distorted by outliers.	Not always a good representation of the data — it doesn't use all the data values .

Exercise 5

Q1 Suggest whether the range is a good measure of spread for the following data sets and explain why.

- a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 b) 14, 14, 20, 22, 24, 27, 29, 30, 35, 93
 c) 2222, 750, 725, 55, 800, 783, 880 d) 1.8, 1.7, 2.0, 1.0, 1.2, 1.6, 1.5

Q2 Look at the data on the right.

- a) (i) Calculate the range for the data.
 (ii) Calculate the interquartile range.
- b) Which of the measures of spread you calculated in part a) do you think best represents the data? Explain your answer.

13	28	893	192	89	98	67	45	78
90	34	56	78	20	783	60	33	45
12	67	54	58	101	56	89	708	9
82	76	90	55	89	347	104	43	76
38	13	27	51	92	18	44	87	71

32.2 Averages for Grouped Data

Data in grouped frequency tables have to be treated differently as you don't know the actual data values.

Learning Objective — Spec Ref 94:

Find the mode, median, mean and range for data in a grouped frequency table.

If you're given a data set in the form of a **grouped frequency table** then you don't know the **exact** data values — so you can't find exact values for the averages or range. You can only **identify** the **modal group** and **group containing the median**, and **estimate** the **mean** and the **range**.

- The **modal group** is the group (sometimes called a **class**) that has the **highest frequency**.
- The **group containing the median** is found by working out the **position** of the median in the usual way. You then use the **group frequencies** to identify which group the median falls into.
- The **estimated mean** is found by **multiplying** the **frequencies** by the **midpoints** for each group, **adding** up the results and **dividing** by the **total frequency**. The midpoints of each group are used as **estimate** for all the data values within a group. To find the **midpoint** of each group, add the **lower** and **upper** bounds together and **divide by 2**.
- The **estimated range** is found by **subtracting** the **lower bound** of the smallest group from the **upper bound** of the largest group — this gives you the **largest possible range** for the data set.

Example 1

The table shows a summary of the times taken by 15 people to eat three crackers.

Time (<i>t</i>) in s	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$	$80 \leq t < 90$
Frequency	2	3	6	4

a) Write down the modal group.

The modal group is the one with the highest frequency.

Modal group is $70 \leq t < 80$.

b) Which group contains the median?

Find the position of the median and which group it's in. The $2 + 3 + 1 = 6$ th to $2 + 3 + 6 = 11$ th values are all in the group $70 \leq t < 80$.

There are 15 values, so the median is in the $(15 + 1) \div 2 = 8$ th position. The group containing the median is $70 \leq t < 80$.

c) Find an estimate for the mean.

1. Find the midpoint for each group — add together the upper and lower bounds and divide by 2.

Time (<i>t</i>) in s	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$	$80 \leq t < 90$
Frequency	2	3	6	4
Midpoint	$(50 + 60) \div 2 = 55$	$(60 + 70) \div 2 = 65$	$(70 + 80) \div 2 = 75$	$(80 + 90) \div 2 = 85$

2. Now find the mean as before, using the midpoints instead of actual data values.

Midpoint \times freq.	$2 \times 55 = 110$	$3 \times 65 = 195$	$6 \times 75 = 450$	$4 \times 85 = 340$
-------------------------	---------------------	---------------------	---------------------	---------------------

Total frequency = $110 + 195 + 450 + 340 = 1095$, so the estimated mean = $\frac{1095}{15} = 73$ seconds.

d) Find an estimate for the range.

Subtract the lower bound of the smallest group from the upper bound of the largest group.

Estimated range = $90 - 50 = 40$ seconds

Exercise 1

Q1 The table on the right shows some information about the weights of some tangerines in a supermarket. Find an estimate for the mean tangerine weight.

Weight (w) in grams	Frequency
$0 \leq w < 20$	1
$20 \leq w < 40$	6
$40 \leq w < 60$	9
$60 \leq w < 80$	24

Q2 Troy collected some information about the number of hours students spent watching television over a week. His results are shown in the table.

- a) Write down the modal group.
- b) Which group contains the median?
- c) Find an estimate for the mean to 3 significant figures.

Time (t) in hours	Frequency
$0 \leq t < 5$	3
$5 \leq t < 10$	8
$10 \leq t < 15$	11
$15 \leq t < 20$	4

Q3 This table shows information about the heights of 200 people.

- a) Write down the modal group.
- b) Which group contains the median?
- c) Find an estimate for the mean to 3 significant figures.
- d) Find an estimate for the range.

Height (h) in metres	Frequency
$1.50 \leq h < 1.60$	27
$1.60 \leq h < 1.70$	92
$1.70 \leq h < 1.80$	63
$1.80 \leq h < 1.90$	18

Q4 The table below shows the times taken to deliver pizzas in one week.

Time (t) in minutes	$0 \leq t < 5$	$5 \leq t < 10$	$10 \leq t < 15$	$15 \leq t < 20$	$20 \leq t < 25$	$25 \leq t < 30$
Frequency	40	64	89	82	34	18

- a) Write down the modal group.
- b) Which group contains the median time taken to deliver a pizza?
- c) Estimate the mean time taken to deliver a pizza to 3 significant figures.
- d) Estimate the range of times taken to deliver a pizza.
- e) The pizza company guarantee to deliver your pizza in less than 15 minutes or your pizza is free. What percentage of the pizzas delivered that week were free? Give your answer to 3 s.f.

Q5 The table below shows the number of different species of bird seen by visitors to a nature reserve on one particular day.

No. of bird species (x)	$0 \leq x < 3$	$3 \leq x < 7$	$7 \leq x < 10$	$10 \leq x < 15$	$15 \leq x < 20$	$20 \leq x < 30$
Frequency	2	5	19	16	7	1

- a) Write down the modal group.
- b) Which group contains the median number of different species seen?
- c) Estimate the mean number of different species seen.
- d) Estimate the range of the number of different species seen.
- e) On the next day, the average number of different species seen was 11. How do these results compare to the previous day?

Review Exercise

Q1 A company asks 40 people to taste their new peanut butter and give it a mark out of 10. Their results are shown in the box.

a) (i) Find the modal score.

(ii) Find the median score.

(iii) Calculate the mean score.

(iv) Calculate the range.

5	2	8	9	5	5	7	8	1	2
3	4	6	8	1	5	3	2	4	5
7	7	8	6	2	7	2	8	9	2
5	7	8	9	8	5	7	8	2	6

b) The company want to quote an average score in their next advertising campaign. Which average best represents the data? Explain your answer.

c) Give one reason why the company may choose to use a different average.

Q2 This table shows the shoe sizes of 30 school pupils.

Shoe size	1	2	3	4	5	6	7
Frequency	2	9	4	6	5	0	1

a) Find the mode, median and mean size, rounding to 3 significant figures where necessary.

b) Explain why the mean might not be the best average for this data set.

Q3 A biologist recorded the following weights for 9 hippos in kilograms:

195, 1525, 1340, 245, 1950, 1750, 1600, 1600, 1400

a) Calculate (i) the mode, (ii) the median and (iii) the mean to the nearest kilogram.

b) Calculate (i) the range and (ii) the interquartile range for the weights.

c) The biologist thinks she may have accidentally included the weights of several pygmy hippos in her list. How might this have affected the range and interquartile range?

Q4 A group of 17 pantomime horses were timed as they ran 100 m. The finishing times for the pantomime horses (in seconds) are shown on the right.

15.3	18.9	40.2	20.5	14.0	17.3
17.8	30.5	60.2	56.0	32.1	34.2
22.7	36.2	19.2	41.1	26.4	

17 kittens ran the same course. The mean time it took a kitten to run 100 m was 24.6 seconds. The interquartile range of the kittens' times was 48.6 seconds. Use this data to investigate the hypothesis 'Pantomime horses run faster than kittens'.

Q5 The table shows the number of potatoes some fish and chip shops go through a day.

a) Write down the modal group.

b) Which group contains the median?

c) Find an estimate for the mean.

d) Find an estimate for the range.

Number of potatoes (p)	Frequency
$500 \leq p < 600$	2
$600 \leq p < 700$	14
$700 \leq p < 800$	25
$800 \leq p < 1000$	9

Exam-Style Questions

- Q1** Work out the median for the following numbers.

17 83 19 28 29 106

[1 mark]

- Q2** Gertrude records the number of eggs laid by hens on her free range farm in a month. Calculate the mean number of eggs laid to 2 decimal places.

No. of eggs laid	24	25	26	27	28	29	30	31
Frequency	7	23	45	109	541	1894	3561	2670

[3 marks]

- Q3** The masses of 25 marrows entered into a vegetable growing competition are shown below.

Mass (m kg)	$3.0 \leq m < 4.0$	$4.0 \leq m < 5.0$	$5.0 \leq m < 6.0$	$6.0 \leq m < 6.5$
Frequency	11	9	4	1

Using the table, work out the percentage of marrows that were above the mean mass.



[3 marks]

- Q4** The seven scores given to a competitor in a diving competition were:

7 7 8 8.5 8.5 8.5 9.5

Find the ratio of the interquartile range to the range giving your answer in the form $a : b$, where a and b are both integers.

[3 marks]

- Q5** The table shows the number of motorcycles sold each day by a salesman over the last six days.

Number of motorcycles sold	Frequency
0	1
1	3
3	2

The salesman sold the same number of motorcycles each day for the four days prior to these six days. The mean number of motorcycles sold each day over the last ten days was 3.3. Work out the modal number of motorcycles sold each day over the last ten days, explaining why your answer is the mode.



[3 marks]

- Q6** Four integers have a median and mode of 7. The range of the integers is 10 and the mean is 9. Work out the four integers, giving them in ascending order.



[3 marks]