

Probability & Statistics

Chapter 2

Measures of central tendency

Cambridge AS level



2.3 THE MEDIAN



KEY POINT:

For n ordered values, the median is at the $\left(\frac{n+1}{2}\right)$ th value.

For even values of n , the median is the mean of the two middle values.



Example 1:

To find the median number of customers served on each of these days.



Monday (12)		Saturday (15)	Key: 0 2 2
8 6 3 1 0 0	2	2 3 4 6	represents 20 customers
4 3 1 1 0	3	5 5 6 8 9 9	on Monday and 22
1	4	0 1 3 7 9	customers on Saturday

For Saturday, $n = 15$

$$\text{Median} = \left(\frac{n+1}{2}\right) \text{th} = \left(\frac{15+1}{2}\right) \text{th} = 8 \text{th value}$$

$$\text{Median} = 38$$

For Monday, $n = 12$

$$\text{Median} = \left(\frac{n+1}{2}\right) \text{th} = \left(\frac{12+1}{2}\right) \text{th} = 6.5 \text{th value}$$

$$\text{Median} = \frac{28+30}{2} = 29$$



Example 2:

The following table shows 65 ungrouped readings of x . Cumulative frequencies and the positions of the readings are also shown. Find the median value of x .



x	f	cf	Positions
40	11	11	1 st to 11 th
41	23	34	12 th to 34 th
42	19	53	35 th to 53 rd
43	8	61	54 th to 61 st
44	4	65	62 nd to 65 th

$$n = 65$$

$$\text{Median} = \left(\frac{n + 1}{2}\right) \text{th} = \left(\frac{65 + 1}{2}\right) \text{th} = 33\text{th value}$$

$$\text{Median} = 41$$



2.3 ESTIMATING THE MEDIAN



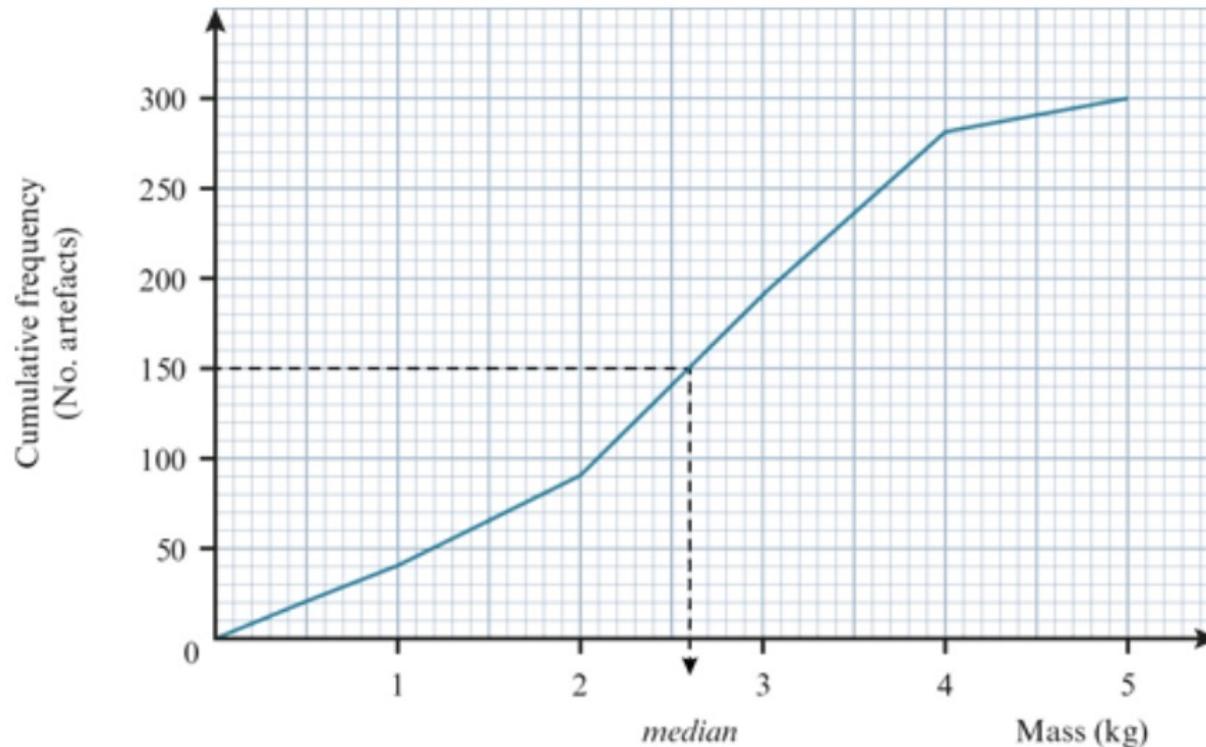
In large datasets and in sets of continuous data, values are grouped and the actual values cannot be seen. This means that we cannot find the exact value of the median but we can estimate it. The method we use to estimate the median for this type of data is by reading its value from a cumulative frequency graph.

KEY POINT:

On a cumulative frequency graph with total frequency $n = \sum f$, the median is at the $\frac{n}{2}$ th value.



2.3 ESTIMATING THE MEDIAN



$$n = 300$$

$$\begin{aligned} \text{Median} &= \frac{n}{2} \\ &= \frac{300}{2} \\ &= 150\text{th value} \end{aligned}$$

TIP: The graph is only an estimate, so we use $\frac{n}{2}$ rather than $\frac{n+1}{2}$ to estimate the median. This ensures that we arrive at the same position for the median whether we count up from the bottom or down from the top of the cumulative frequency axis.

2.3 CHOOSING AN APPROPRIATE AVERAGE



	Advantages	Disadvantages
Mode	Unlikely to be affected by extreme values . Useful to manufacturers that need to know the most popular styles and sizes. Can be used for all sets of qualitative data.	Ignores most values. Rarely used in further calculations.
Mean	Takes all values into account. Frequently used in further calculations. The most commonly understood average. Can be used to find the sum of the data values.	Cannot be found unless all values are known. Likely to be affected by extreme values.
Median	Can be found without knowing all of the values. Relatively unaffected by extreme values.	Only takes account of the order of the values and so ignores most of them.



Example 3:

Consider the number of days taken by a courier company to deliver 100 packages, as given in the following table and represented in the bar chart.



No. days	1	2	3	4	5	6
No. packages (f)	10	40	25	14	8	3

$$\text{Mode} = 2 \text{ days}$$

$$\text{Median} = \frac{100 + 1}{2} = 50.5\text{th}$$

$$\text{Median} = 2.5 \text{ days}$$

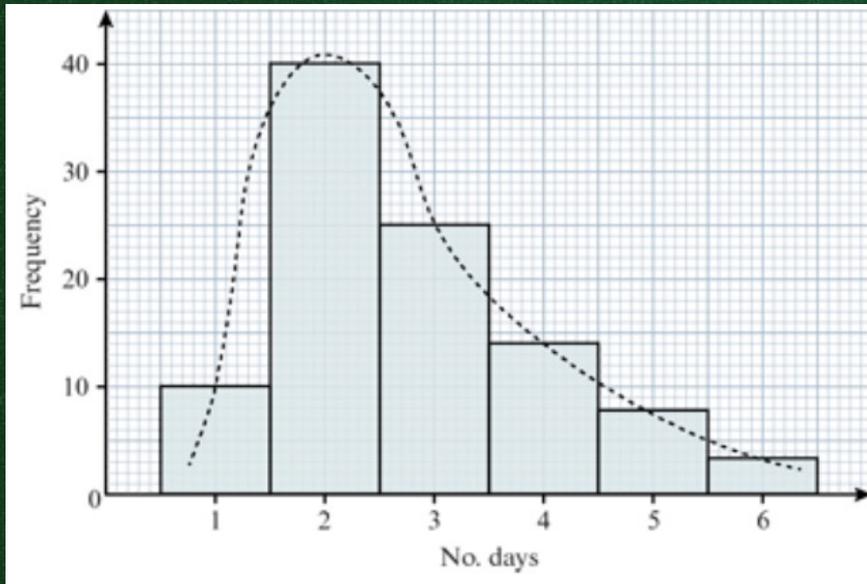
$$\text{Mean} = \frac{1 \times 10 + 2 \times 40 + 3 \times 25 + 4 \times 14 + 5 \times 8 + 6 \times 3}{100}$$

$$\text{Mean} = 2.79 \text{ days}$$



Example 3:

Consider the number of days taken by a courier company to deliver 100 packages, as given in the following table and represented in the bar chart.



Mode = 2 days

Median = 2.5 days

Mean = 2.79 days

A set of data that is not symmetrical is said to be **skewed**. When the curve's longer tail is to the side of the larger values, as in the previous bar chart, the data are said to be positively skewed. When the longer tail is to the side of the smaller values, the data are said to be negatively skewed.



2.3 CHOOSING AN APPROPRIATE AVERAGE



Generally, we find that:

Mode < median < mean when the data are **positively skewed.**

Mean < median < mode when the data are **negatively skewed.**



EXERCISE 2E

- 1 The number of patients treated each day by a dentist during a 20-day period is shown in the following stem-and-leaf diagram.

0		4	4	5	6	6	6	7	Key: 1 5			
1		4	5	5	6	6	7	7	8	8	9	represents
2		0	1									15 patients

- a Find the median number of patients.
 b On eight of these 20 days, the dentist arrived late to collect their son from school. If they decide to use their average number of patients as a reason for arriving late, would they use the median or the mean? Explain your answer.
 c Describe a situation in which it would be to the dentist's advantage to use a mode as the average.
- 2 a Find the median for the values of t given in the following table.

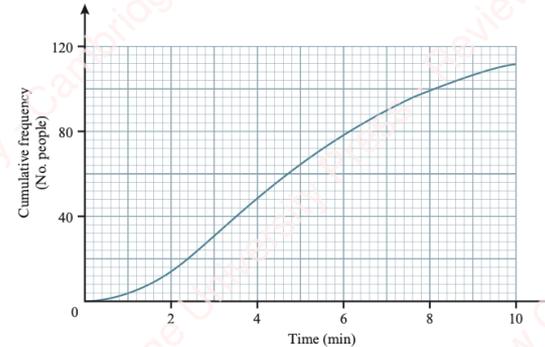
t	7	8	9	10	11	12	13
f	4	7	9	14	16	41	9

- b What feature of the data suggests that \bar{t} is less than the median? Confirm whether or not this is the case.
- 3 a Find the median and the mode for the values of x given in the following table.

	4	5	6	7	8
f	14	13	4	12	15

- b Give one positive and one negative aspect of using each of the median and the mode as the average value for x .
 c Some values in the table have been incorrectly recorded as 8 instead of 4. Find the number of incorrectly recorded values, given that the true median of x is 5.5.

- 4 The following graph illustrates the times taken by 112 people to complete a puzzle.



- a Estimate the median time taken.
 b The median is used to divide these people into two groups. Find the median time taken by each of the groups.
- 5 The masses, m kilograms, of 148 objects are summarised in the following table.

Mass (m kg)	$m < 0$	$m < 0.2$	$m < 0.3$	$m < 0.5$	$m < 0.7$	$m < 0.8$
cf	0	16	28	120	144	148

Construct a cumulative frequency polygon on graph paper, and use it to estimate the number of objects with masses that are:

- a within 0.1 kg of the median
 b more than 200 g from the median.
- 6 A teacher recorded the quiz marks of eight students as 11, 13, 15, 15, 17, 18, 19 and 20. They later realised that there was a typing error, so they changed the mark of 11 to 1. Investigate what effect this change has on the mode, mean and median of the students' marks.
- 7 The following table shows the lifetimes, to the nearest 10 days, of a certain brand of light bulb.

Lifetime (days)	90–100	110–120	130–140	150–160	170–190	200–220	230–260
No. light bulbs (f)	12	28	54	63	41	16	6

- a Use upper class boundaries to represent the data in a cumulative frequency graph and estimate the median lifetime of the light bulbs.
 b How might the manufacturer choose a value to use as the average lifetime of the light bulbs in a publicity campaign? Based on the figures in the table, investigate whether it would be to the manufacturer's advantage to use the median or the mean.

Homework

Page 46 – Exercise 2E



- 8 It is claimed on the packaging of a brand of battery that they can run a standard kitchen clock continuously for 'at least 150 days on average'. Tests are carried out to find the length of time, t hours, that a standard kitchen clock runs using one of these batteries. The results are shown in the following table.

Time (t hours)	$3000 \leq t < 3096$	$3096 \leq t < 3576$	$3576 \leq t < 3768$	$3768 \leq t < 3840$
No. batteries (f)	34	66	117	33

What could the words on the packaging mean? Test the claim by finding the mean, the median and the modal class. What conclusions, if any, can you make about the claim?

- 9 Homes in a certain neighbourhood have recently sold for \$220 000, \$242 000, \$236 000 and \$350 000. A potential buyer wants to know the average selling price in the neighbourhood. Which of the mean, median or mode would be more helpful? Explain your answer.

- 10 A study was carried out on 60 electronic items to find the currents, x amperes, that could be safely passed through them at a fixed voltage before they overheat. The results are given in the two tables below.

Current (x amperes)	0.5	1.5	2.0	3.5	5.0
No. items that do not overheat	60	48	20	6	0

Current (x amperes)	0.5	1.5	2.0	3.5	5.0
No. items that overheat	0	p	q	r	60

- a Find the value of p , q and of r .
- b Cumulative frequency graphs are drawn to illustrate the data in both tables.
- Describe the transformation that maps one graph onto the other.
 - Explain the significance of the point where the two graphs intersect.
- 11 The lengths of extra-time, t minutes, played in the first and second halves of 100 football matches are summarised in the following table.
- | Extra-time (t min) | $t \leq 1$ | $t \leq 2$ | $t \leq 4$ | $t \leq 5$ | $t \leq 7$ | $t \leq 9$ |
|-------------------------|------------|------------|------------|------------|------------|------------|
| First halves (f_1) | 24 | 62 | 80 | 92 | 97 | 100 |
| Second halves (f_2) | 6 | 17 | 35 | 82 | 93 | 100 |
- a Explain how you know that the median extra-time played in the second halves is greater than in the first halves.
- b The first-half median is exactly 100 seconds.
- Find the upper boundary value of k , given that the second-half median is k times longer than the first-half median.
 - Explain why the mean must be greater than the median for the extra-time played in the first halves.
- 12 Eighty candidates took an examination in Astronomy, for which no candidate scored more than 80%. The examiners suggest that five grades, A, B, C, D and E, should be awarded to these candidates, using upper grade boundaries 64, 50, 36 and 26 for grades B, C, D and E, respectively. In this case, grades A, B, C, D and E, will be awarded in the ratio 1:3:5:4:3.
- a Using the examiners' suggestion, represent the scores in a cumulative frequency polygon and use it to estimate the median score.
- b All of the grade boundaries are later reduced by 10%. Estimate how many candidates will be awarded a higher grade because of this.

- 13 The values of x shown in the following table are to be represented in a bar chart.

x	5	6	7	8	9	10	11
Frequency	2	5	9	10	9	5	2

- Sketch a curve that shows the shape of the data.
 - Find the mode, mean and the median of x .
- b The two smallest values of x (i.e. 5 and 5) are changed to 21 and 31. Investigate the effect that this has on the mode, the mean, the median and on the shape of the curve.
- c If, instead, the two largest values of x (i.e. 11 and 11) are changed to -9 and b , so that the mean of x decreases by 1, find the value of b and investigate the effect that this has on the mode, the median and the shape of the curve.
- 14 A histogram is drawn to illustrate a set of continuous data whose mean and median are equal. Make sketches of the different types of curve that could be drawn to represent the shape of the histogram.

- 15 Students' marks in a Biology examination are shown by percentage in the following table.

Marks (%)	20–	30–	40–	50–	60–	70–	80–90
Frequency (%)	5	10	20	30	20	10	5

- a Without drawing an accurate histogram, describe the shape of the set of marks. What does the shape suggest about the values of the mean, the median and the mode?
- b Information is provided about the marks in examinations in two other subjects:
- Chemistry: mode > median > mean Physics: mean > median > mode
- Sketch a curve to show the shape of the distribution of marks in each of these exams.