Cumulative frequency and box plots

S3, S4 · Statistics · revision notes

Cumulative frequency tables and graphs, using them to estimate the median, quartiles and interquartile range, and drawing and comparing box plots. This subtopic is Higher tier only. Expect to complete a table, draw the graph, read estimates from it and compare two distributions.

Key facts

- Cumulative frequency = running total of the frequencies
- From a graph: median at $\frac{n}{2}$ · LQ at $\frac{n}{4}$ · UQ at $\frac{3n}{4}$
- IQR = UQ − LQ (the spread of the middle 50%)
- Number more than a value = total − cumulative frequency at that value
- Plot at the upper class boundary; the last value = total frequency n
- From an ordered list: LQ = $\frac{n+1}{4}$th, median = $\frac{n+1}{2}$th, UQ = $\frac{3(n+1)}{4}$th value
- Box plot: minimum, LQ, median, UQ, maximum
- Compare: medians (on average) and IQRs (consistency), in context

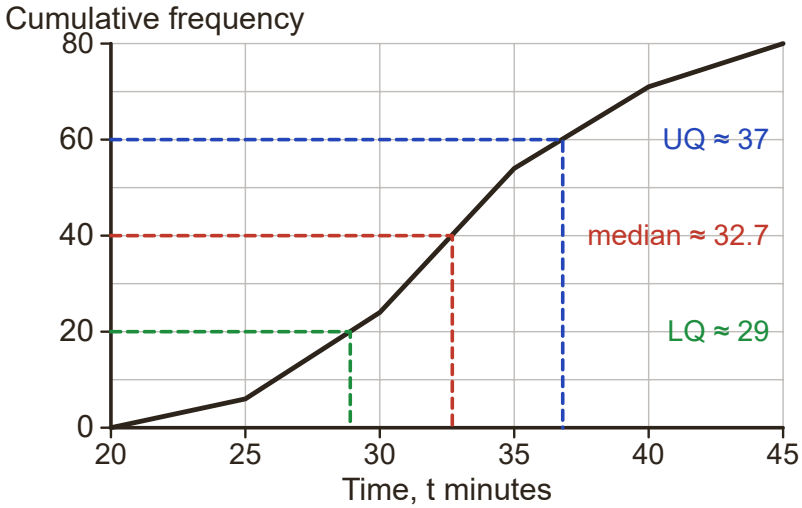
Notes

Cumulative frequency tables and graphs

- **Cumulative frequency** is a running total: add each frequency to the total so far.
- The last cumulative frequency is the total frequency, n. Check it.
- Plot each cumulative frequency at the **upper end** of its class (the upper class boundary), not the midpoint.
- If the first class starts at, say, 20, also plot (20, 0), because no values are below 20.
- Join the points with a smooth curve or with straight lines. The graph never goes down.

Reading estimates from the graph

- Median: go across from $\frac{n}{2}$ on the cumulative frequency axis to the curve, then down to read the value.
- Lower quartile (LQ): across from $\frac{n}{4}$. Upper quartile (UQ): across from $\frac{3n}{4}$
- **Interquartile range** (IQR) = UQ − LQ. It is the spread of the middle half of the data, so extreme values do not affect it.

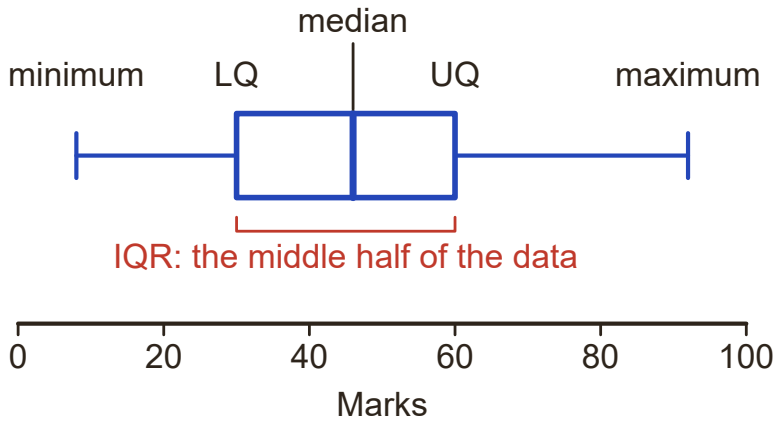


n = 80: read across from 20 (LQ), 40 (median) and 60 (UQ). IQR ≈ 37 − 29 = 8 minutes.

- How many are less than a value: go up from the value and read across. How many are more than it: total − that reading.

Box plots

- A box plot shows five values: minimum, LQ, median, UQ and maximum.
- Draw the box from the LQ to the UQ with a line at the median, and whiskers out to the minimum and maximum, against a scale.



The box is the middle half of the data; each whisker holds about a quarter.

- **Where the quartiles are**, for n values:

Quartile	From a graph	From a list
LQ	$\frac{n}{4}$	$\frac{n+1}{4}$th value
Median	$\frac{n}{2}$	$\frac{n+1}{2}$th value
UQ	$\frac{3n}{4}$	$\frac{3(n+1)}{4}$th value

- About half of the data lie inside the box, and about a quarter lie along each whisker.
- To check a box plot against a cumulative frequency graph or table: its median, LQ and UQ must match the readings at $\frac{n}{2}$, $\frac{n}{4}$ and $\frac{3n}{4}$, and its minimum and maximum must lie in the first and last classes that have any values. Grade 8+ only

Comparing distributions

- Compare the **medians**: which group is higher (or lower) on average.
- Compare the **IQRs**: the smaller IQR means the data are more consistent.
- Always write in context, e.g. 'On average, the journeys on Route B are quicker'.
- Box plots drawn on the same scale make two groups easy to compare.

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own, C = a communication mark.

Complete a cumulative frequency table and draw the graph

3 marks | Grade 6

1. Add up the frequencies as you go down the table.
2. Plot each running total at the upper end of its class.
3. Also plot the lower end of the first class with a cumulative frequency of 0.
4. Join the points with a smooth curve or straight lines.

Example. The table shows information about the times, t minutes, taken by 80 runners to finish a race.

Time (t minutes)	Freq.	Cum. freq.
$20 < t \leq 25$	6	
$25 < t \leq 30$	18	
$30 < t \leq 35$	30	
$35 < t \leq 40$	17	
$40 < t \leq 45$	9	

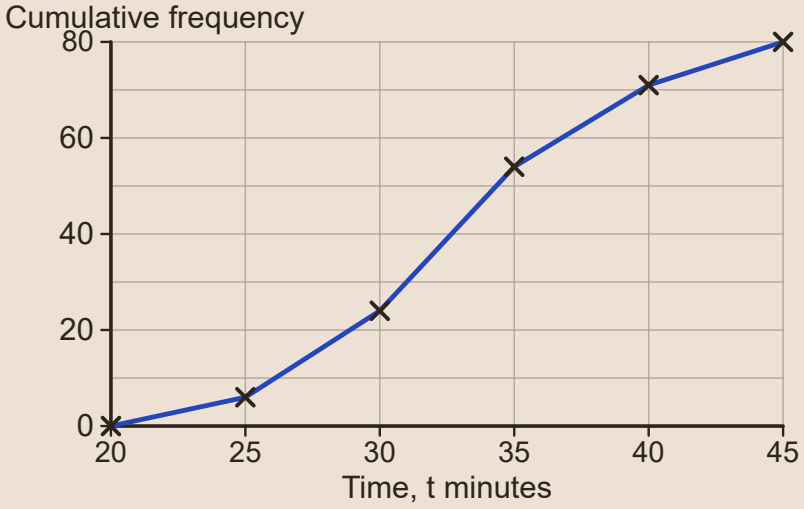
(a) Complete the cumulative frequency column.
(b) Draw a cumulative frequency graph.

Model answer.

(a) 6, 24, 54, 71, 80 (B1)

(b) Points plotted at (25, 6), (30, 24), (35, 54), (40, 71) and (45, 80) (M1 for at least 4 points plotted correctly at the upper ends of the classes)

A smooth curve or straight lines through all the points, starting at (20, 0) (A1)



Estimate the median and interquartile range from the graph

3 marks | Grade 6

1. Work out $\frac{n}{2}$, $\frac{n}{4}$ and $\frac{3n}{4}$
2. Draw lines across from each of these to the curve, then down to the horizontal axis.
3. IQR = UQ − LQ. Leave your lines on the graph to show your method.
4. Asked if a box plot fits the graph? Its median, LQ and UQ must match your readings, and its whiskers must end inside the first and last classes.

Example. Use the cumulative frequency graph for the 80 runners from the first example.
(a) Find an estimate for the median time.
(b) Find an estimate for the interquartile range.

Model answer.

(a) Read across from 40: median ≈ 32.7 minutes (B1, accept readings from about 32 to 33.5)

(b) Read across from 20 and 60: LQ ≈ 29 minutes and UQ ≈ 37 minutes (M1)

IQR ≈ 37 − 29 = 8 minutes (A1, accept about 7 to 9)

Draw a box plot from a list

3 marks | Grade 6

1. Make sure the data are in order.
2. Find the median and quartiles from their positions: $\frac{n+1}{2}$, $\frac{n+1}{4}$ and $\frac{3(n+1)}{4}$
3. Draw the box from the LQ to the UQ, a line at the median, and whiskers to the smallest and largest values.

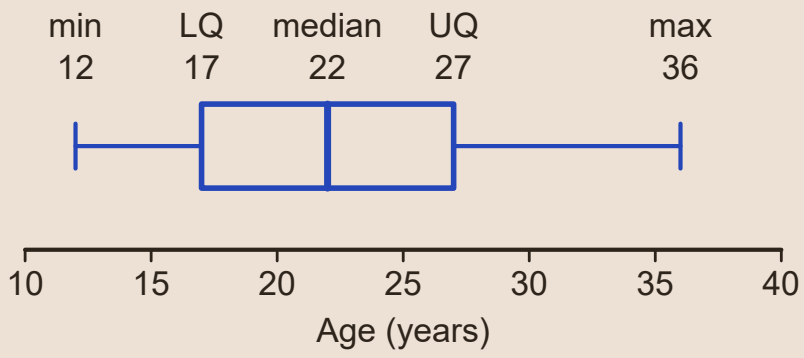
Example. Here are the ages, in years, of 11 people at a party.
12, 15, 17, 18, 20, 22, 23, 25, 27, 30, 36
Draw a box plot for this information.

Model answer.

Median = 6th value = 22 (B1)

LQ = 3rd value = 17 and UQ = 9th value = 27 (B1)

A box from 17 to 27 with a line at 22, and whiskers to 12 and 36, drawn accurately against a scale (B1)



Estimate how many are more than a value

2 marks **Grade 7**

1. Find the value on the horizontal axis and go up to the curve.
2. Read across to the cumulative frequency: this is how many are less than the value.
3. For 'more than', subtract this reading from the total.

Example. Use the cumulative frequency graph for the 80 runners from the first example.
Work out an estimate for the number of runners who took more than 38 minutes.

Model answer.

Reading up from 38 minutes gives a cumulative frequency of about 64 (M1)

80 − 64 = 16 runners (A1, accept about 14 to 18)

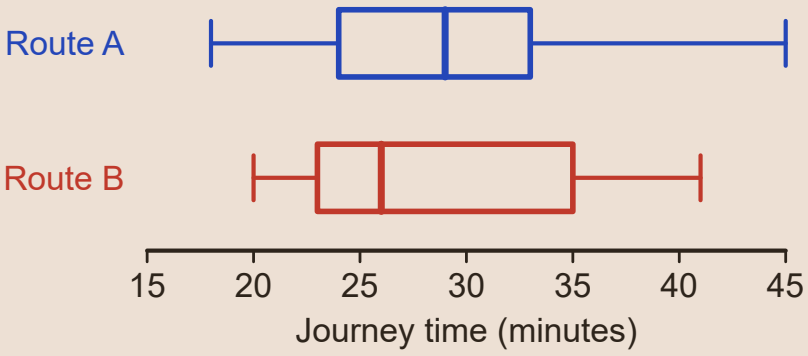
Compare two distributions from box plots

2 marks | Grade 7

1. Read each median and work out each IQR.
2. Compare the medians: say which group is higher or lower on average, in context.
3. Compare the IQRs: the smaller IQR is more consistent.

Example. Box plots are drawn to show the journey times, in minutes, for two routes to school. The table gives the values shown on them.

	Route A	Route B
Min	18	20
LQ	24	23
Median	29	26
UQ	33	35
Max	45	41



Compare the journey times for the two routes.

Model answer.

On average the journeys on Route B are quicker, because its median (26 minutes) is lower than Route A's (29 minutes) (C1)

The times on Route A are more consistent, because its IQR (33 − 24 = 9 minutes) is smaller than Route B's (35 − 23 = 12 minutes) (C1)

Don’t lose marks

- Plotting the cumulative frequencies at the midpoints of the classes instead of the upper ends.
- Giving a cumulative frequency (such as 40) as the median instead of reading the value on the horizontal axis.
- Forgetting to subtract from the total in a 'more than' question.
- Comparing ranges when the question asks you to use box plots or the interquartile range.
- Comparing without context, e.g. 'A has a higher median' without saying what that means for the data.

Misreading the scale: check what one small square is worth on each axis.

More tips

Memory tricks

- Quarters of n: for 80 values read across at 20, 40 and 60; for 120 values at 30, 60 and 90.
- Cumulative frequency only ever climbs. If a running total goes down, you have made an adding slip.
- Box plot = five numbers: minimum, LQ, median, UQ, maximum.
- Sense check: the median must lie between the quartiles, and the IQR is never bigger than the range.
- Percentages: 'the fastest 10%' of 80 runners means read across at 10% of 80 = 8.

Exam technique

- Leave your reading lines on the graph. They show your method and can earn marks even if your reading is slightly out.
- Plot points accurately, to within half a small square, and join them with a single smooth curve or straight lines.
- In comparisons, one mark is usually for comparing the medians and one for comparing the IQRs, each in context.
- For a box plot from a graph or table, look for the minimum and maximum in the question: they are often given in words (e.g. 'the shortest time was 21 minutes').

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade 5 — **Complete a cumulative frequency table**
Keep a running total of the frequencies; the last value must equal the total frequency.

Grade 6 — **Find quartiles and IQR from a list**
In an ordered list of n values, the quartiles are the $\frac{n+1}{4}$th and $\frac{3(n+1)}{4}$th values.

Grade 6 — **Draw a cumulative frequency graph**
Plot each running total at the upper end of its class and join the points with a smooth curve or straight lines.

Grade 6 — **Estimate the median and quartiles from a graph**
Read across from $\frac{n}{2}$, $\frac{n}{4}$ and $\frac{3n}{4}$ on the cumulative frequency axis to the curve, then down.

Grade 6 — **Draw and interpret a box plot**
Mark the minimum, lower quartile, median, upper quartile and maximum against a scale.

Grade 7 — **Estimate how many are above a value**
Read up from the value to the curve and across; for 'more than', subtract this reading from the total.

Grade 7 — **Compare distributions using median and IQR**
Compare the medians (on average) and the interquartile ranges (consistency), in context.

Grade 8 — **Check a box plot against a graph**
Test whether the median, quartiles and extreme values of a box plot agree with a cumulative frequency graph or table.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.