Exercise walk-through

6.1.2 Cumulative

AQA · GCSE · Mathematics

Histograms, cumulative frequency and box plots (H)

- **Histograms.** In a **histogram** the area of each bar, not its height, shows the frequency.
- When the **class intervals** have different widths, the height is the **frequency density**:
- **Worked example 6 (H).** Classes $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 40$ have frequencies 20, 30 and 20.
- The **class width** of the last class is 20, so the frequency densities are $20 \div 10 = 2$, $30 \div 10 = 3$ and $20 \div 20 = 1$.

$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}$$

Sources for this lesson

- AQA, *GCSE Mathematics (8300) specification*, subject content section 3.6 (Statistics).
- "Landon, 1,293,669; Roosevelt, 972,897", *The Literary Digest* (31 October 1936).
- P. Squire, "Why the 1936 Literary Digest poll failed", *Public Opinion Quarterly* 52 (1988), pages 125 to 133.
- AQA, *GCSE Mathematics* Higher Papers 1, 2 and 3 (June 2024), the **pie chart**, frequency table and **box plot** questions, adapted.

Common errors (2)

- (H) Drawing **histogram** bars with height equal to frequency when the class widths differ.
- (H) Plotting **cumulative frequency** at the midpoint instead of the upper end of the class.
- Comparing only averages; a comparison also needs a measure of spread.
- Saying that **correlation** proves cause.

Exercise 2.1

2.1 *(H)* Use the box plot in the worked example. Work out the range of the journey times. [1]



Exercise 2.1

2.1 *(H)* Use the box plot in the worked example. Work out the range of the journey times. [1]



2.1

- $28 - 6 = \mathbf{22}$ minutes

Setup for exercise 2.2

2.2 *(H)* Use the box plot. Estimate how many of the 80 students took between 12 and 16 minutes. [1]



2.3 *(H)* Find the median, quartiles and IQR of 3, 5, 6, 8, 9, 11, 12, 14, 15, 18, 20. [2]



Exercise 2.2

2.4 *(H)* The cumulative frequency table gives the times of 40 runners. How many took more than 30 minutes? [1]

Time t (min)	$t \leq 10$	$t \leq 20$	$t \leq 30$	$t \leq 40$	$t \leq 50$
Cumulative frequency	4	15	31	38	40



2.5 *(H)* Use the table in **2.4**. What percentage of the runners took 20 minutes or less? [2]



Exercise 2.2

2.4 *(H)* The cumulative frequency table gives the times of 40 runners. How many took more than 30 minutes? [1]

Time t (min)	$t \leq 10$	$t \leq 20$	$t \leq 30$	$t \leq 40$	$t \leq 50$
Cumulative frequency	4	15	31	38	40



2.5 *(H)* Use the table in **2.4**. What percentage of the runners took 20 minutes or less? [2]



Full solution on next slide

Full solution 2.2

2.2

- Q_1 to the median is a quarter: $\frac{1}{4} \times 80 = \mathbf{20}$

2.3

- $n = 11$: median is the 6th value, **11**; Q_1 the 3rd, **6**; Q_3 the 9th, **15**
- IQR $= 15 - 6 = \mathbf{9}$

2.4

- $40 - 31 = \mathbf{9}$ runners

2.5

- $\dfrac{15}{40} \times 100\% = \mathbf{37.5}\%$

Exercise 3.1

3.1 *(H)* Explain why cumulative frequency is plotted at the upper end of each class. [1]



Exercise 3.1

3.1 *(H)* Explain why cumulative frequency is plotted at the upper end of each class. [1]



3.1

- a cumulative frequency counts everything **up to the end** of the class; only at the upper boundary have all values in that class been included

Setup for exercise 3.2

3.2 *(H)* Route A has median 16 minutes and IQR 9 minutes. Route B has median 20 minutes and IQR 5 minutes. Compare the two routes. [2]



3.3 *(H)* Draw a box plot for data with minimum 4, lower quartile 9, median 13, upper quartile 18 and maximum 25, on a scale from 0 to 30. [2]

Exercise 3.2

3.4 *(H)* Use the box plot in the worked example. Estimate how many of the 80 students took between 12 and 28 minutes, and how many took more than 16 minutes. [3]



3.5 *(H)* Use the table in **2.4**. Assuming the times are spread evenly within each class, estimate the median time. [3]



Exercise 3.2

3.4 *(H)* Use the box plot in the worked example. Estimate how many of the 80 students took between 12 and 28 minutes, and how many took more than 16 minutes. [3]



3.5 *(H)* Use the table in **2.4**. Assuming the times are spread evenly within each class, estimate the median time. [3]



3.2

- route A is quicker on average: its median, 16 minutes, is lower than 20

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 3.2

3.2

- route A is quicker on average: its median, 16 minutes, is lower than 20
- route B is more consistent: its IQR, 5 minutes, is smaller than 9

3.3

- box from 9 to 18 with the median line at 13
- whiskers to 4 and 25, on an accurate scale

3.4

- 12 to 28 is from Q_1 to the maximum: three quarters, $\frac{3}{4} \times 80 = \mathbf{60}$
- more than the median is half: **40**

3.5

- the median is the 20th time, in the class $20 < t \leq 30$, where the cumulative frequency goes from 15 to 31
- it is $20 - 15 = 5$ of the 16 values into the class: $20 + \frac{5}{16} \times 10 = \mathbf{23.1}$ minutes