Exercise walk-through

6.1.5
Boxplot Comparisons

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 worked example. Write down the median of group B. [1]



Exercise 2.1

2.1 *(H)* Use the worked example. Write down the median of group B. [1]



2.1

- **8**

Setup for exercise 2.2

2.2 *(H)* A box plot has lower quartile 30 and upper quartile 46. Work out the IQR. [1]



2.3 *(H)* The box plot in **2.2** shows the results of 60 people. How many scored between 30 and 46? [1]



Exercise 2.2

2.4 *(H)* Find the median, quartiles and IQR of these 15 values: 2, 3, 5, 6, 7, 8, 8, 9, 10, 11, 12, 14, 15, 17, 21. [2]



2.5 *(H)* Explain why the IQR is a better measure of spread than the range for data with one very large value. [1]



Exercise 2.2

2.4 *(H)* Find the median, quartiles and IQR of these 15 values: 2, 3, 5, 6, 7, 8, 8, 9, 10, 11, 12, 14, 15, 17, 21. [2]



2.5 *(H)* Explain why the IQR is a better measure of spread than the range for data with one very large value. [1]



2.2

- $46 - 30 = \mathbf{16}$

2.3

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 2.2

2.2

- $46 - 30 = \mathbf{16}$

2.3

- the box holds the middle half: $\frac{1}{2} \times 60 = \mathbf{30}$

2.4

- median: 8th value $= \mathbf{9}$; Q_1: 4th value $= \mathbf{6}$; Q_3: 12th value $= \mathbf{14}$
- IQR $= 14 - 6 = \mathbf{8}$

2.5

- the range uses the very large value, but the IQR uses only the middle half, so one extreme value hardly affects it

Exercise 3.1

3.1 *(H)* Ben says group B in the worked example is more spread out because its range is larger. Comment on his statement. [1]



Exercise 3.1

3.1 *(H)* Ben says group B in the worked example is more spread out because its range is larger. Comment on his statement. [1]



3.1

- the range depends only on the extreme values; B's IQR (4) is smaller than A's (6), so the middle of B is **less** spread out

Setup for exercise 3.2

3.2 *(H)* Boys' test scores: minimum 20, $Q_1 = 32$, median 40, $Q_3 = 48$, maximum 60. Girls' scores: minimum 25, $Q_1 = 38$, median 45, $Q_3 = 50$, maximum 58. Compare the scores of the boys and the girls. [2]



3.3 *(H)* Use the boys' scores in **3.2**. Between which two fractions does the proportion of girls who scored more than the boys' upper quartile lie? Explain. [2]



3.4 *(H)* For group A in the worked example, a value is an outlier if it is more than $1.5 \times$ IQR above Q_3 or below Q_1. A new value of 23 is added. Use the original quartiles to decide whether it is an outlier. [3]



Exercise 3.2

3.5 *(H)* For 120 students in group X, a cumulative frequency graph gives readings of 22 at 30 students, 31 at 60 students and 38 at 90 students. Group Y has median 35 and IQR 10. Work out the IQR of group X, compare the groups, and estimate how many students in X scored more than 38. [4]



Exercise 3.2

3.5 *(H)* For 120 students in group X, a cumulative frequency graph gives readings of 22 at 30 students, 31 at 60 students and 38 at 90 students. Group Y has median 35 and IQR 10. Work out the IQR of group X, compare the groups, and estimate how many students in X scored more than 38. [4]



3.2

- on average the girls scored higher: their median, 45, is above the boys' 40
- the girls' scores are more consistent: IQR $50 - 38 = 12$ against the boys' $48 - 32 = 16$

3.3

- the boys' Q_3 is 48, which lies between the girls' median (45) and Q_3 (50)
- so between **a quarter and a half** of the girls scored more than 48

Full solution on next slide

Full solution 3.2

3.2

- on average the girls scored higher: their median, 45, is above the boys' 40
- the girls' scores are more consistent: IQR $50 - 38 = 12$ against the boys' $48 - 32 = 16$

3.3

- the boys' Q_3 is 48, which lies between the girls' median (45) and Q_3 (50)
- so between **a quarter and a half** of the girls scored more than 48

3.4

- IQR $= 11 - 5 = 6$, and $1.5 \times 6 = 9$
- upper limit $= Q_3 + 9 = 11 + 9 = 20$
- $23 > 20$, so **yes**, 23 is an outlier

3.5

- $n = 120$: Q_1 at 30, median at 60, Q_3 at 90; IQR $= 38 - 22 =$ **16**
- Y scored higher on average: median 35 against 31
- Y is more consistent: IQR 10 against 16
- above 38 is the top quarter of X: $120 - 90 =$ **30** students