



H21-test Box Plots from Cumulative Frequency and Outliers: Test Pack

EDEXCEL 1ST0 · Calculator allowed · about 15 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** In this topic, a value in a data set is classed as an outlier if it lies more than $1.5 \times \text{IQR}$ below the lower quartile, or more than $1.5 \times \text{IQR}$ above the upper quartile, where IQR is the interquartile range.

A small data set has lower quartile = 10 and upper quartile = 22.

- (a)** Calculate the interquartile range. (1)

- (b)** Use the rule given above to calculate the lower and upper boundaries for outliers. (2)

(Total for Question 1 is 3 marks)

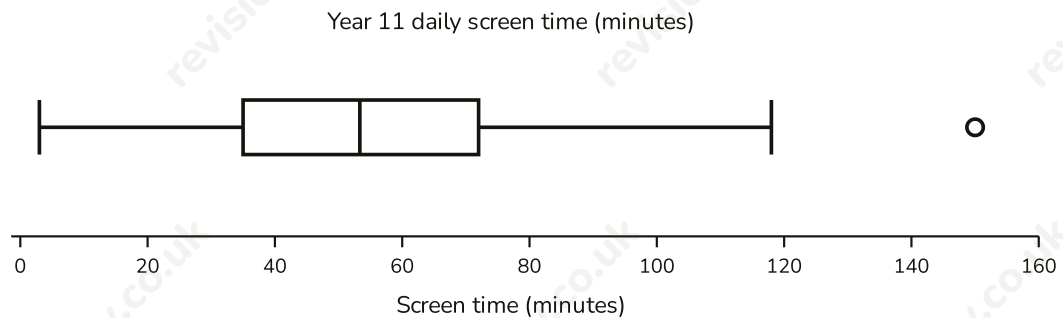
- 2** A student records a five-number summary for a data set:

Minimum = 8, Lower quartile = 22, Median = 19, Upper quartile = 27, Maximum = 33

Explain, giving a reason, why this five-number summary cannot be correct.

(Total for Question 2 is 2 marks)

- 3** Using the box plot for the Year 11 screen-time data shown below, determine whether the distribution of screen times is skewed, and justify your answer. Comment on what effect the outlier at 150 minutes has on the mean compared with the median.



(Total for Question 3 is 3 marks)

- 4** For a set of delivery times, the lower quartile is 20 minutes. The upper boundary for outliers, found using the $1.5 \times \text{IQR}$ rule, is 50 minutes.

(a) Work out the upper quartile. (3)

(b) Hence state the interquartile range. (1)

(Total for Question 4 is 4 marks)

- 5** A shop manager records the amount, in GBP, spent by 40 customers in one hour. Most customers spent between 5 and £25, but two customers made much larger purchases of £180 and £210, which are clear outliers. The manager wants to summarise the 'typical' spend and the spread of spending.
- (a) State, giving a reason, whether the mean and standard deviation, or the median and interquartile range, would be more appropriate summary statistics for this data set. (2)
- (b) Explain what would happen to the mean if the two outlying purchases were removed from the data set, compared with what would happen to the median. (2)

(Total for Question 5 is 4 marks)

Mark scheme · H21-test Box Plots from Cumulative Frequency and Outliers: Test Pack

Question 1

- (a) **B1** 12 cao
- (a) **Answer: 12**
- (b) **M1** $1.5 \times 12 = 18$, oe
- (b) **A1** lower boundary = -8 and upper boundary = 40 cao
- (b) **Answer: Lower boundary = -8, upper boundary = 40**

Question 2

- **B1** identifies that the median (19) is smaller than the lower quartile (22), which is impossible
- **B1** correct reasoning, e.g. by definition the lower quartile can never be greater than the median, since at least 25% of the data lies at or below the lower quartile and the median is the middle value oe
- **Answer: Incorrect: the median (19) is less than the lower quartile (22); the lower quartile can never be greater than the median.**

Question 3

- **B1** valid skew judgement with correct comparison, e.g. the lower whisker (3 to 35, a gap of 32) is shorter than the upper whisker (72 to 118, a gap of 46), so the distribution is positively (right) skewed oe
- **B1** recognises the outlier at 150 reinforces/increases this positive skew, extending the long tail even further to the right
- **B1** correct comparison of mean and median, e.g. the outlier would pull the mean up/increase it, since it is an extreme high value, but would have little or no effect on the median, so the mean would end up higher than the median oe
- **Answer: Positively (right) skewed; the outlier increases the skew and would inflate the mean but not the median.**

Question 4

- (a) **M1** forms a correct equation, e.g. $q + 1.5(q - 20) = 50$, where q is the upper quartile
- (a) **M1** correctly simplifies to $2.5q = 80$, oe
- (a) **A1** 32 cao
- (a) **Answer: 32 minutes**
- (b) **B1** ft: 12 cao
- (b) **Answer: 12 minutes**

Question 5

- (a) **B1** median and interquartile range cao
- (a) **B1** valid reason, e.g. the mean and standard deviation are strongly affected/distorted by extreme values, whereas the median and IQR are resistant to (not affected by) outliers oe
- (a) **Answer: Median and interquartile range, because they are not distorted by the two outliers, unlike the mean and standard deviation.**
- (b) **B1** the mean would decrease/drop noticeably, since it currently includes two very large values which inflate the total oe
- (b) **B1** the median would change little or not at all, since it depends only on the middle value(s) of the ordered data, not the extreme values oe
- (b) **Answer: The mean would drop noticeably, but the median would stay about the same.**