



## H20-30min Skewness: 30-minute Pack

EDEXCEL 1ST0 · Calculator allowed · about 30 minutes

Total Marks

Name: \_\_\_\_\_ Date: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

**Answer ALL questions. Show all your working.**

- 1** In a long jump competition, the distances jumped by 30 athletes had a mean of 4.35 m, a median of 4.60 m and a mode of 4.70 m.
- (a) State whether the distribution of jump distances is symmetric, positively skewed or negatively skewed. (1)
- (b) Justify your answer to part (a) by referring to the mean, median and mode. (2)

**(Total for Question 1 is 3 marks)**

**2** The times, in minutes, that 11 customers waited in a queue at a bank are: 2, 3, 3, 4, 4, 4, 5, 6, 7, 9, 25.

**(a)** Work out the mean waiting time, giving your answer correct to 2 decimal places. **(2)**

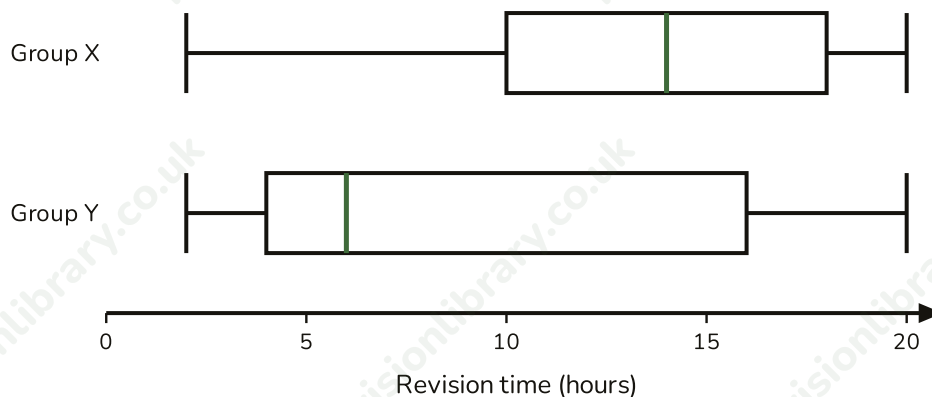
**(b)** Write down the median waiting time. **(1)**

**(c)** Write down the modal waiting time. **(1)**

**(d)** State, with a reason based on your answers to parts (a) to (c), whether the distribution of waiting times is symmetric, positively skewed or negatively skewed. **(2)**

**(Total for Question 2 is 6 marks)**

- 3** The box plots show the amount of time, in hours, that Group X and Group Y spent revising for an exam, drawn on the same scale.



- (a)** Write down the median revision time for Group X. (1)

- .....  
**(b)** Write down the median revision time for Group Y. (1)

- .....  
**(c)** Work out the interquartile range for Group Y. (1)

- .....  
**(d)** Compare the skewness of the two distributions, referring to the shape of each box plot. (2)

**(Total for Question 3 is 5 marks)**

- 4** One measure of skewness is the quartile coefficient of skewness, SK, defined as  $SK = (Q1 + Q3 - 2Q2) / (Q3 - Q1)$ , where Q1 is the lower quartile, Q2 is the median and Q3 is the upper quartile. A value of SK greater than 0 indicates positive skew, a value less than 0 indicates negative skew, and SK = 0 indicates a symmetric quartile spacing. The times, in minutes, taken by 60 runners to complete a park run have lower quartile Q1 = 24, median Q2 = 30 and upper quartile Q3 = 34.

**(a)** Calculate the value of SK for the park run times. **(3)**

---

**(b)** State whether the distribution of park run times is symmetric, positively skewed or negatively skewed. **(1)**

**(c)** The times for a second, larger park run have a quartile coefficient of skewness SK = 0.05. Compare the skew of the two data sets, referring to both the sign and the relative size of SK. **(2)**

**(Total for Question 4 is 6 marks)**

**5** The number of emails received per day by 15 employees, arranged in order, are: 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 40.

**(a)** Find the lower quartile, the median and the upper quartile of this data. (3)

.....  
**(b)** Calculate the quartile coefficient of skewness,  $SK = (Q1 + Q3 - 2Q2) / (Q3 - Q1)$ , (2)  
using your answers to part (a).

.....  
**(c)** Calculate the mean number of emails, giving your answer correct to 2 decimal (2)  
places.

.....  
**(d)** The value of SK suggests the data is symmetric, but the mean is noticeably larger (2)  
than the median. Explain why these two pieces of evidence appear to disagree.

**(Total for Question 5 is 9 marks)**

## Mark scheme · H20-30min Skewness: 30-minute Pack

---

### Question 1

- (a) **B1** Negatively skewed cao
- (a) **Answer: Negatively skewed**
- (b) **B1** identifies mean < median < mode ( $4.35 < 4.60 < 4.70$ ), oe
- (b) **B1** correct interpretation: most athletes jumped distances bunched towards the higher end, with a small tail of shorter jumps pulling the mean down below the median and mode
- (b) **Answer: Since  $4.35 < 4.60 < 4.70$ , most jump distances are bunched near the top of the range, with a tail of shorter jumps pulling the mean down below the median and mode.**

### Question 2

- (a) **M1**  $2+3+3+4+4+4+5+6+7+9+25$  (=72), oe divided by 11
- (a) **A1** 6.55 cao (awrt 6.55)
- (a) **Answer: 6.55**
- (b) **B1** 4 cao
- (b) **Answer: 4**
- (c) **B1** 4 cao
- (c) **Answer: 4**
- (d) **B1** mean (6.55) is greater than the median and mode (both 4)
- (d) **B1** Positively skewed - one extreme value (25 minutes) pulls the mean up while barely affecting the median or mode
- (d) **Answer: Positively skewed**

### Question 3

- (a) **B1** 14 cao
- (a) **Answer: 14**
- (b) **B1** 6 cao
- (b) **Answer: 6**
- (c) **B1** 12 cao
- (c) **Answer: 12**
- (d) **B1** Group X is symmetric (median - LQ = 4 = UQ - median, the two box halves are equal)
- (d) **B1** Group Y is positively skewed (median - LQ = 2 is much smaller than UQ - median = 10, so the higher revision times are more spread out - a long tail towards the higher times)
- (d) **Answer: Group X is symmetric; Group Y is positively skewed, with a tail towards the longer revision times.**

### Question 4

- (a) **M1**  $24 + 34 - 2(30)$  (= -2), oe
- (a) **M1** their numerator divided by  $(34 - 24)$  (=10), dep on first method mark
- (a) **A1** -0.2 cao (allow -1/5)
- (a) **Answer: -0.2**
- (b) **B1** Negatively skewed, since  $SK < 0$  cao
- (b) **Answer: Negatively skewed**
- (c) **B1** this park run is negatively skewed ( $SK = -0.2$ ), while the second park run is very slightly positively skewed, almost symmetric ( $SK = 0.05$ )
- (c) **B1** this distribution shows noticeably more skew, since its SK value is much further from 0 (magnitude 0.2 is greater than 0.05)
- (c) **Answer: This park run is negatively skewed while the second is almost symmetric (very slightly positive); this one shows more skew since its SK is further from 0.**

### Question 5

- (a) **B1** lower quartile = 7 (4th of 15 values)
- (a) **B1** median = 11 (8th of 15 values)
- (a) **B1** upper quartile = 15 (12th of 15 values)
- **(a) Answer: Q1 = 7, median = 11, Q3 = 15**
- (b) **M1**  $7 + 15 - 2(11) (=0)$ , oe divided by  $(15-7)$
- (b) **A1** SK = 0 cao
- **(b) Answer: 0**
- (c) **M1**  $3+5+6+7+8+9+10+11+12+13+14+15+16+18+40 (=187)$ , oe divided by 15
- (c) **A1** 12.47 cao (awrt 12.47)
- **(c) Answer: 12.47**
- (d) **B1** SK only depends on the positions of Q1, Q2 and Q3, so it is not affected by how extreme any values beyond Q3 are (the value of 40 does not change Q1, Q2 or Q3)
- (d) **B1** the mean uses every value in its calculation, so the one very high value (40 emails) pulls it above the median, revealing skew that SK does not detect
- **(d) Answer: SK = 0 only reflects the spread of the middle 50% of the data around the median, which happens to be symmetric here; it ignores the extreme value of 40, which the mean is sensitive to, pulling it well above the median.**