



S08-test Comparing Data Sets: Test Pack

EDEXCEL 1ST0 · Calculator allowed · about 35 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** Two swimming clubs each timed their members over 50m. The table shows summary statistics for their times.

	Mean time (s)	Range (s)
Club A	34.2	5.1
Club B	35.6	8.7

- (a) Compare the mean swim times of Club A and Club B. (1)
- (b) Compare the ranges of the two clubs. (1)
- (c) Using both the mean and the range, write a conclusion comparing the two swimming clubs. (2)

.....
(Total for Question 1 is 4 marks)

- 2** For each statement about comparing data sets, write down whether it is True or False.

- (i) The range only uses the smallest and largest values in a data set. (1)
-
- (ii) The mean is always the best average to use when comparing two data sets. (1)
-
- (iii) Two data sets can have the same mean but different ranges. (1)
-

.....
(Total for Question 2 is 3 marks)

- 3** Team A has a mean score of 24 and a range of 6. Team B has a mean score of 19 and a range of 14. Write down which team has the higher average score.

(Total for Question 3 is 1 mark)

- 4** Class 9A has a mean test score of 58 with a standard deviation of 9. Class 9B has a mean test score of 64 with a standard deviation of 9. Give one similarity and one difference between the two classes' results.

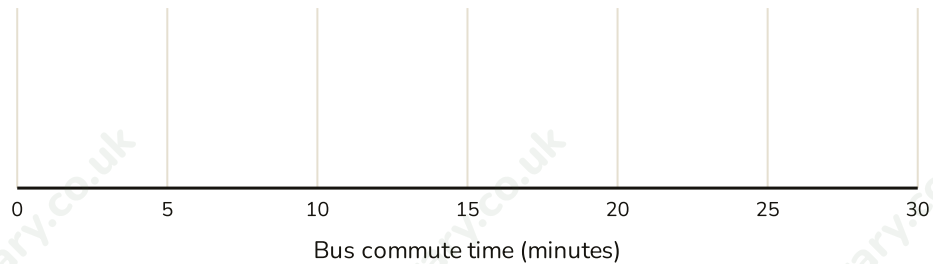
(Total for Question 4 is 2 marks)

- 5** Two shifts at a warehouse pack parcels. Shift 1 has a mean of 145 parcels packed per day and a standard deviation of 12. Shift 2 has a mean of 145 parcels packed per day and a standard deviation of 20. Work out which shift has more variable daily output, and by how much the standard deviations differ.

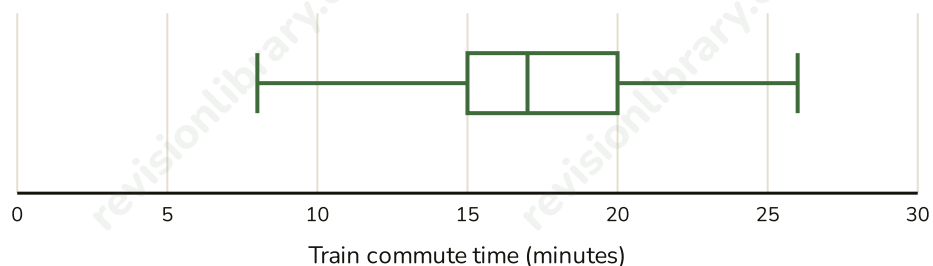
(Total for Question 5 is 2 marks)

6 A student is comparing commute times for two routes to college.

- (a)** On the grid, draw a box plot for the Bus route using this five-number summary: **(3)**
minimum 5 minutes, lower quartile 12 minutes, median 18 minutes, upper quartile 24 minutes, maximum 29 minutes.



- (b)** The box plot for the Train route is shown. Find the interquartile range for the Train route. **(2)**



-
- (c)** Compare the two routes' commute times using the median and the interquartile range, and suggest which route would give a more reliable journey time. **(2)**

(Total for Question 6 is 7 marks)

- 7 Two call centres record how long, in minutes, calls take to resolve. Centre 1: mean 6.8, standard deviation 1.4. Centre 2: mean 5.9, standard deviation 2.6. A manager wants calls to be resolved quickly and consistently.
- (a) Work out the difference between the two means.
- (b) Explain, using both the means and standard deviations, which centre is performing better overall.

(Total for Question 7 is 3 marks)

- 8 A quality inspector at a drinks factory samples 4 bottles from each of two filling machines and measures the volume (in ml).

Machine A: 496, 498, 502, 504

Machine B: 480, 500, 500, 520

- (a) Show that the mean volume is 500 ml for both Machine A and Machine B. (1)

- (b) Calculate the standard deviation of Machine A's volumes. (3)

-
- (c) Calculate the standard deviation of Machine B's volumes. (3)

-
- (d) Compare the two machines using the standard deviations calculated, and state which machine fills bottles more reliably. (2)

(Total for Question 8 is 9 marks)

Mark scheme · S08-test Comparing Data Sets: Test Pack

Question 1

- (a) **B1** Club A has the lower mean time (34.2s compared with 35.6s), so Club A's swimmers were faster on average, oe
- (a) **Answer: Club A is faster on average (34.2s < 35.6s).**
- (b) **B1** Club A has the smaller range (5.1s compared with 8.7s), so Club A's times were more consistent, oe
- (b) **Answer: Club A is more consistent (range 5.1s < 8.7s).**
- (c) **B1** Club A was faster on average than Club B (lower mean)
- (c) **B1** Club A was also more consistent than Club B (smaller range)
- (c) **Answer: Club A was faster on average and more consistent than Club B.**

Question 2

- (i) **B1** True cao
- (i) **Answer: True**
- (ii) **B1** False cao
- (ii) **Answer: False**
- (iii) **B1** True cao
- (iii) **Answer: True**

Question 3

- **B1** Team A cao
- **Answer: Team A**

Question 4

- **B1** correct similarity, e.g. the two classes have the same spread, since the standard deviation is 9 for both oe
- **B1** correct difference, e.g. Class 9B scored higher on average, since its mean (64) is higher than 9A's (58) oe
- **Answer: Similarity: both classes have the same standard deviation (9), so the same spread of scores. Difference: Class 9B has the higher mean (64 vs 58), so 9B performed better on average.**

Question 5

- **M1** correct method, 20 - 12
- **A1** Shift 2, differ by 8, both cao
- **Answer: Shift 2 is more variable; the standard deviations differ by 8**

Question 6

- (a) **B1** whiskers drawn correctly to minimum = 5 and maximum = 29
- (a) **B1** box drawn correctly from lower quartile = 12 to upper quartile = 24
- (a) **B1** median line drawn correctly at 18
- (a) **Answer: Box plot with whiskers to 5 and 29, box from 12 to 24, median line at 18.**
- (b) **M1** 20 - 15, oe
- (b) **A1** 5 cao
- (b) **Answer: 5 minutes**
- (c) **B1** the medians are similar (Bus 18 minutes, Train 17 minutes) but the Bus has a much bigger IQR (12 compared with 5 minutes)
- (c) **B1** so the Train would give a more reliable/predictable journey time, since its middle 50% of times are much less spread out, oe
- (c) **Answer: The Train is more reliable, since the Bus route's IQR (12 minutes) is much bigger than the Train's (5 minutes).**

Question 7

- **M1** correct method, 6.8 - 5.9
- **A1** 0.9 minutes cao
- **B1** valid overall judgement using both statistics, e.g. Centre 2 resolves calls faster on average, but Centre 1 is more consistent since its standard deviation is smaller, so neither centre is better on both measures oe
- **Answer: (a) 0.9 minutes. (b) Centre 2 has the faster average resolve time (lower mean), but Centre 1 is more consistent (lower standard deviation), so the two centres trade off speed against consistency.**

Question 8

- (a) **B1** $496+498+502+504 = 2000$ and $2000/4 = 500$; $480+500+500+520 = 2000$ and $2000/4 = 500$
- (a) **Answer: Both means equal 500 ml.**
- (b) **M1** finds deviations from the mean: -4, -2, 2, 4
- (b) **M1** squares and sums the deviations, then divides by 4: $(16+4+4+16)/4 = 10$
- (b) **A1** $SD = \sqrt{10} = 3.16$ (3 s.f.), awrt 3.16
- (b) **Answer: 3.16 ml (3 s.f.)**
- (c) **M1** finds deviations from the mean: -20, 0, 0, 20
- (c) **M1** squares and sums the deviations, then divides by 4: $(400+0+0+400)/4 = 200$
- (c) **A1** $SD = \sqrt{200} = 14.1$ (3 s.f.), awrt 14.1
- (c) **Answer: 14.1 ml (3 s.f.)**
- (d) **B1** Machine A has the much smaller standard deviation (3.16 ml compared with 14.1 ml)
- (d) **B1** so Machine A fills bottles more reliably/consistently, as its volumes vary less around the mean of 500 ml, oe
- (d) **Answer: Machine A is more reliable (smaller SD of 3.16 ml vs 14.1 ml).**