



S08-30min Comparing Data Sets: 30-minute Pack

EDEXCEL 1ST0 · Calculator allowed · about 30 minutes

Total Marks

--

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** A gym's morning class has a mean attendance of 22 with a range of 8. Its evening class has a mean attendance of 22 with a range of 15. Write down which class has more variable attendance.

(Total for Question 1 is 1 mark)

- 2** Bakery A sold loaves with a mean price of £2.40 and Bakery B sold loaves with a mean price of £2.65. Work out the difference between the two mean prices.

(Total for Question 2 is 2 marks)

- 3** Class A has 8 students; the range of their test scores is 45 marks. Class B has 32 students; the range of their test scores is 20 marks. A student claims: "Class A's test scores are more spread out than Class B's, because Class A has the bigger range."

(a) Explain one reason why comparing the ranges of Class A (8 students) and Class B (32 students) directly might be misleading. **(1)**

(b) Suggest a more appropriate statistic to compare the spread of the two classes' scores, and explain why it would give a fairer comparison. **(2)**

(Total for Question 3 is 3 marks)

- 4 A call centre records the number of calls answered per hour by staff on two different shifts.

Calls per hour	3	4	5	6	7
Frequency (Shift A)	2	5	8	4	1
Frequency (Shift B)	1	3	6	7	3

- (a) Find the median number of calls answered per hour for Shift A. (There are 20 staff-hours recorded for Shift A.) (2)

-
- (b) Find the median number of calls answered per hour for Shift B. (There are 20 staff-hours recorded for Shift B.) (2)

-
- (c) Compare the two shifts using the medians, and write a conclusion in context. (2)

(Total for Question 4 is 6 marks)

-
- 5 Write down whether the following statement is True or False.
When two data sets have the same median, they must also have the same spread.

(Total for Question 5 is 1 mark)

-
- 6 A survey of two book clubs recorded the number of books read last year. Club A: median 14 books, IQR 5 books. Club B: median 11 books, IQR 5 books. Give one similarity and one difference between the two clubs' data.

(Total for Question 6 is 2 marks)

7 Two small companies each employ 5 staff. Their annual salaries are shown below.

Company A: £22000, £23000, £24000, £25000, £26000

Company B: £20000, £21000, £22000, £23000, £95000 (the owner's salary)

(a) Calculate the mean salary for each company. (2)

(b) Calculate the median salary for each company. (2)

(c) Explain why the median is a more appropriate measure than the mean to compare typical salaries at the two companies, referring to your answers. (2)

(Total for Question 7 is 6 marks)

- 8** A student is investigating whether reaction times differ between two age groups at a science fair: Under 16s and Adults. Raw reaction times (in milliseconds) collected from a sample of each group were:

Under 16s: 210, 225, 198, 240, 215, 205

Adults: 180, 175, 190, 185, 170, 300

- (a) Suggest a suitable method the student could use to select a sample of participants for each age group, and give one reason why this method would help make a fair comparison. (2)
- (b) Calculate the mean and the range of the reaction times for each age group, using all the data given. (4)
-
- (c) One adult recorded a reaction time of 300 ms, much higher than the rest of that group. State what this value is called, and recalculate the mean and the range for the Adults group with this value removed. (3)
- (d) Using the most appropriate statistics, write a conclusion comparing the reaction times of the two age groups. (2)

(Total for Question 8 is 11 marks)

Mark scheme · S08-30min Comparing Data Sets: 30-minute Pack

Question 1

- **B1** Evening class cao
- **Answer: Evening class**

Question 2

- **M1** correct method, $2.65 - 2.40$
- **A1** 0.25 (25p) cao
- **Answer: £0.25 (25p)**

Question 3

- (a) **B1** with a much smaller sample, the range is more likely to be affected by just one or two extreme/unusual scores, so it may not fairly represent the typical spread, oe
- (a) **Answer: The range from a small sample (Class A) is more easily distorted by one extreme score.**
- (b) **B1** suggests the interquartile range (or standard deviation)
- (b) **B1** explains that it is less affected by extreme values / describes the spread of the middle of the data, giving a fairer comparison regardless of sample size, oe
- (b) **Answer: The interquartile range (IQR) would give a fairer comparison, as it ignores extreme values at either end.**

Question 4

- (a) **M1** identifies the median lies between the 10th and 11th values ($n = 20$)
- (a) **A1** 5 cao
- (a) **Answer: 5 calls per hour**
- (b) **M1** identifies the median lies between the 10th and 11th values ($n = 20$)
- (b) **A1** 5.5 cao
- (b) **Answer: 5.5 calls per hour**
- (c) **B1** Shift B has the higher median (5.5 compared with 5 calls per hour)
- (c) **B1** so Shift B typically answered slightly more calls per hour than Shift A, oe
- (c) **Answer: Shift B typically answers slightly more calls per hour (median 5.5 vs 5).**

Question 5

- **B1** False cao
- **Answer: False**

Question 6

- **B1** correct similarity, e.g. both clubs have the same spread, since the IQR is 5 for both oe
- **B1** correct difference, e.g. Club A read more books on average, since its median (14) is higher than Club B's (11) oe
- **Answer: Similarity: both clubs have the same IQR (5), so the same spread. Difference: Club A has the higher median (14 vs 11), so Club A members typically read more books.**

Question 7

- (a) **B1** Company A mean = £24000
- (a) **B1** Company B mean = £36200
- (a) **Answer: Company A: £24000; Company B: £36200**
- (b) **B1** Company A median = £24000
- (b) **B1** Company B median = £22000
- (b) **Answer: Company A: £24000; Company B: £22000**
- (c) **B1** Company B's mean (£36200) is distorted/inflated by the one very high salary (£95000), an outlier
- (c) **B1** the median is not affected by this outlier and better represents a typical employee's salary at Company B, oe
- (c) **Answer: The median is a fairer typical value, since the mean at Company B is skewed by the owner's much higher salary.**

Question 8

- (a) **B1** names a suitable sampling method (e.g. random sampling, or stratified sampling by age group)
- (a) **B1** gives a valid reason (e.g. avoids bias in who is chosen, so the two samples can be fairly compared), or
- (a) **Answer: Use random sampling within each age group (e.g. numbering participants and using a random number generator).**
- (b) **B1** Under 16s mean = 215.5 ms
- (b) **B1** Under 16s range = 42 ms
- (b) **B1** Adults mean = 200 ms
- (b) **B1** Adults range = 130 ms
- (b) **Answer: Under 16s: mean 215.5 ms, range 42 ms. Adults: mean 200 ms, range 130 ms.**
- (c) **B1** identifies the value as an outlier
- (c) **B1** new mean = 180 ms
- (c) **B1** new range = 20 ms
- (c) **Answer: 300 ms is an outlier; without it, mean = 180 ms and range = 20 ms.**
- (d) **B1** recognises the Under 16s figures (mean 215.5 ms, range 42 ms) are not affected by an outlier and are representative of that group
- (d) **B1** concludes that, once the outlier is removed, the Adults reacted faster (180 ms vs 215.5 ms) and more consistently (range 20 ms vs 42 ms) than the Under 16s, so the original Adults data was misleading, or
- (d) **Answer: Once the outlier is removed, Adults reacted faster and more consistently than Under 16s.**