Statistics and data analysis - Applications test

16 questions | 46 marks | 60 minutes

Apply statistical tools to clinics, transport and school investigations; critique the limits of claims.

Equivalent correct methods earn the corresponding marks.

1. A clinic has recorded waits 8, 10, 12, 12, 13, 15, 18, 40 minutes. Calculate mean and median. [3 marks]

Total 128; divide by 8.

Middle pair 12 and 13 gives 12.5.

Answer: Mean 16 min; median 12.5 min.

2. Which of those two summaries better describes the centre without the extreme wait dominating? Explain. [2 marks]

The 40 minute wait raises the mean.

The median depends on the middle positions.

Answer: Median 12.5 min.

3. Give the clinic's five-number summary using median-of-halves, and its IQR. [4 marks]

Lower half 8, 10, 12, 12 has median 11.

Upper 13, 15, 18, 40 has median 16.5.

Answer: 8, 11, 12.5, 16.5, 40; IQR 5.5 min.

4. A second clinic has median 15 min and IQR 3 min. Compare the clinics without calling either universally better. [3 marks]

$12.5 < 15$ compares centre.

$5.5 > 3$ compares central spread.

Answer: First has lower typical wait; second has a smaller middle-half spread.

5. The clinic changes every wait by subtracting 3 minutes. State the new mean and what happens to SD. [2 marks]

$16 - 3 = 13.$

A common shift leaves deviations unchanged.

Answer: Mean 13 min; SD unchanged.

6. A delivery firm has population mean 4 kg and variance 2.25 kg². Find its SD and interpret it. [3 marks]

$\sqrt{2.25} = 1.5$.

It is not the range or the mass of every parcel.

Answer: SD 1.5 kg; it measures spread of parcel masses around the mean.

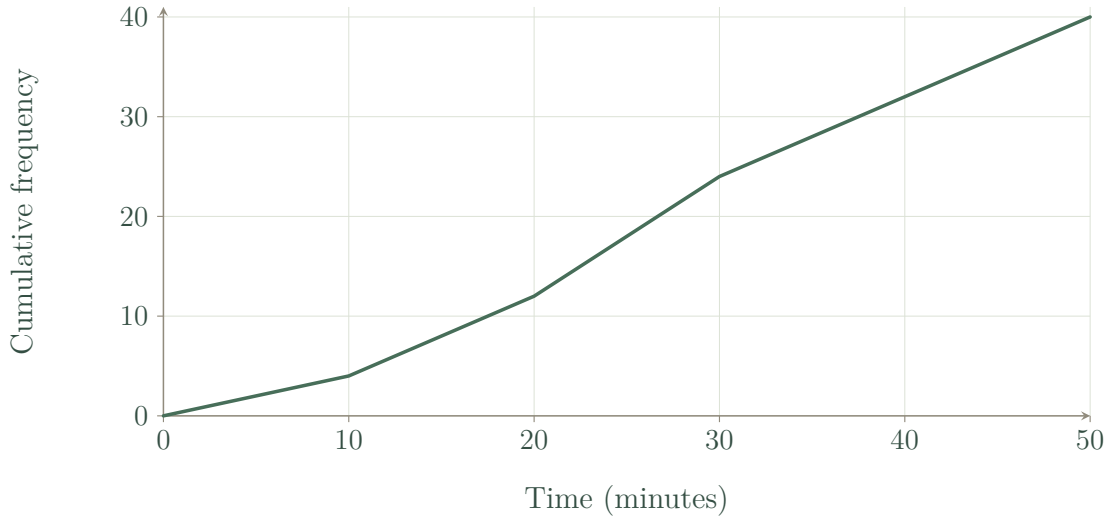
7. A school reports an estimated mean journey time from grouped intervals. Why should it not claim the exact mean to six decimal places? [2 marks]

The estimate depends on unknown within-class values.

Report suitable precision.

Answer: Midpoint replacement loses individual-value information; extra decimal places do not restore it.

8. Forty race times have the supplied cumulative graph. Estimate P_{50} and P_{80}. [4 marks]



P_{50} uses cumulative 20 with linear interpolation in 20-30.

P_{80} uses cumulative 32 at 40.

Answer: $P_{50} \approx 26.7$ min; $P_{80} = 40$ min.

9. A team says '80% of future athletes will finish within 40 minutes'based only on that sample. Explain the overstatement. [2 marks]

A descriptive percentile is not a guaranteed future proportion.

Answer: The sample percentile describes the recorded sample; future/population predictions need representative data and uncertainty.

10. In a study-time investigation, scores rise with reported study. Name two possible limitations. [3 marks]

The variables may not be measured consistently.

Other influences can explain part of the association.

Answer: Self-report error and confounding by previous attainment/motivation, for example.

11. The study contains pairs $(1, 45)$, $(2, 50)$, $(3, 58)$, $(4, 67)$, $(5, 72)$. Predicting a score at 3.5 hours is different from predicting at 20 hours. Explain. [3 marks]

A local trend may help within the data range.

It may fail far beyond that range.

Answer: 3.5 hours is interpolation within the observed range; 20 hours is extrapolation far outside it.

12. A transport survey compares bus use: Year 9 24 of 40; Year 10 30 of 75. Compare fairly. [3 marks]

$\frac{24}{40} = 0.60$; $\frac{30}{75} = 0.40$.

Compare within-year rates.

Answer: Year 9 60%; Year 10 40%.Year 9 has higher bus-use proportion despite fewer users.

13. Of all 54 bus users in that survey, what percentage are Year 10? [2 marks]

The stated group is all bus users.

Use denominator 54, not 75.

Answer: $\frac{30}{54} \times 100 \approx 55.6\%$.

14. A sponsor publishes only the best-performing subgroup after trying many groups. What concern does this raise? [3 marks]

The omitted groups matter.

A preplanned analysis and full reporting are more transparent.

Answer: Selective reporting can make the evidence look stronger or more general than it is.

15. A student's investigation excluded one outlier without checking it. Write a better procedure. [3 marks]

Outliers are evidence until a defensible reason supports exclusion.

Report the decision transparently.

Answer: Check entry/measurement and context; retain genuine data or justify a stated exclusion rule, and compare conclusions with/without it.

16. Write a complete short conclusion for a convenience sample showing a positive association between height and arm span, including two limitations. [4 marks]

Describe direction and the actual sample.

State specific limits rather than a universal claim.

Answer: In this convenience sample, taller participants tended to have larger arm spans; selection and measurement error limit generalisation, and the data do not establish causation.