Outliers and justified treatment

6 questions | 18 marks

Equivalent correct methods earn the corresponding marks.

1. For $(1, 2), (2, 4), (3, 6), (4, 8), (5, 20)$, calculate the fitted slope using all observations and compare with the first four. [3 marks]

Paired measurements

x	y
1	2
2	4
3	6
4	8
5	20

For all five, $\frac{S_{xy}}{S_{xx}}$ gives the fitted slope.

The first four are exactly $y = 2x$.

Answer: All-data slope 4; first-four slope 2.

2. An unusually high score is confirmed as correct. Is this alone a reason to delete it? [3 marks]

Investigate whether it belongs to the target population.

Report justified sensitivity rather than silently improving the fit.

Answer: No. A genuine score should not be excluded merely because it is unusual.

3. An isolated point has a very large x-value but lies close to a regression line. Could it be influential? [3 marks]

Influence is about changing the fit, not just distance vertically from it.

Answer: Yes. Its extreme predictor location can strongly affect the slope even with a small residual.

4. Two reports use the same data, but one silently omits three low outcomes. State the problem. [3 marks]

A report must disclose exclusions and give a defensible criterion.

Answer: Selective omission can bias the model and mislead readers.

5. An all-data model predicts 40 and the model without one point predicts 55 for the same x. What does this show? [3 marks]

The difference is 15 response units.

Do not hide the sensitivity behind a single prediction.

Answer: The prediction is sensitive to that observation; further investigation and both results should be reported.

6. Suggest a justified reason to exclude a correctly recorded point from a particular analysis. [3 marks]

The population criterion should be scientifically relevant and disclosed, not chosen to maximise correlation.

Answer: It may be from outside the explicitly defined target population, such as a different machine type.