Statistics and regression - Applications test

10 questions | 58 marks | 75 minutes

Evaluate competing interpretations and models using supplied evidence and transparent calculations.

Equivalent correct methods earn the corresponding marks.

1. A monitoring study records time since maintenance x and a vibration measure y. [7 marks]

Paired measurements

Days	Vibration units
2	7
4	8
6	12
8	12
10	16

(a) Fit y on x and state the observed predictor range. [4 marks]

Use the five paired observations.

Record the fitted coefficients without early rounding.

Answer: $\hat{y} \approx 4.4 + 1.1x$; range 2-10 days.

(b) Predict at day 7 and interpret the result as an estimate. [3 marks]

Seven is inside the observed range.

Residual scatter makes it an estimate, not a guaranteed reading.

Answer: 12.1 vibration units; interpolation.

2. Use the monitoring data from Question 1. [6 marks]

(a) Calculate r and r^2. [3 marks]

Use the paired data and square unrounded r.

Answer: $r \approx 0.9648$; $r^2 \approx 0.9308$.

(b) A technician claims time causes exactly $r^2 \times 100$ percent of all vibration. Correct the claim. [3 marks]

The percentage describes variation, not a share of a measured amount.

Other changing conditions may confound the association.

Answer: The fitted linear relationship accounts for about 93.08% of observed vibration variation in this sample; it does not prove causation.

3. The monitoring model is used on day 40. [5 marks]

(a) Find the model prediction. [2 marks]

Substitute $x = 40$.

Answer: 48.4 vibration units.

(b) Explain why this value should not be accepted without further evidence. [3 marks]

The calculation is extrapolation beyond the observations.

Answer: Day 40 is far outside 2-10 days; wear or maintenance conditions may change the relationship.

4. A measurement at day 8 was originally recorded as 120 but the signed instrument log shows 12. [5 marks]

(a) What is the justified treatment? [2 marks]

The source confirms an entry error.

Answer: Correct the verified transcription to 12, document the correction and refit.

(b) How would the response differ if 120 were confirmed as genuine? [3 marks]

Data validity determines treatment, not the desired fitted model.

Answer: Investigate the unusual valid event and report sensitivity; do not silently replace it with a more convenient value.

5. A proposed line has residuals $(2, -1, -2, -1, 2)$, while a curved model has $(0.5, -0.5, 0, 0.5, -0.5)$. [6 marks]

(a) Calculate both sums of squared residuals. [3 marks]

Line:$4 + 1 + 4 + 1 + 4 = 14$.

Curve:$0.25 + 0.25 + 0 + 0.25 + 0.25 = 1$.

Answer: Line:14; curve:1.

(b) Give a numerical and a graphical reason to investigate the curve. [3 marks]

Check that its added complexity is appropriate to the context and supported by data.

Answer: It has lower squared error and removes the U-shaped residual pattern.

6. A decaying response is modelled by $y = 240e^{-0.18t}$, $t \geq 0$. [6 marks]

(a) Find its one-unit multiplier and percentage decrease. [3 marks]

The multiplier is $e^{-0.18}$.

Answer: Multiplier 0.8353; decrease 16.47%.

(b) Find when the model first reaches 120. [3 marks]

$\frac{120}{240} = e^{-0.18t} = \frac{1}{2}$.

Take logs and solve.

Answer: $t = \frac{\ln(2)}{0.18} \approx 3.851$.

7. An analyst has two fitted equations: y-on-x is $\hat{y} = 4 + 2x$, while x-on-y is $\hat{x} = -1 + 0.4y$. [7 marks]

(a) Use the appropriate model to predict y when $x = 5$. [2 marks]

Use the model with y as response: $4 + 2 \times 5$.

Answer: 14.

(b) Use the appropriate model to predict x when $y = 14$. [2 marks]

Use x-on-y: $-1 + 0.4 \times 14$.

Answer: 4.6.

(c) Explain why the two predictions do not reverse exactly. [3 marks]

The fits answer different prediction questions.

Answer: Each regression minimises errors in a different response direction; imperfect association gives different fitted lines.

8. A normal measurement X has mean 30 and SD 2.5. Accepted measurements must lie between 26 and 34. [6 marks]

(a) Calculate the acceptance probability. [3 marks]

Standardise the limits to -1.6 and 1.6.

Answer: 0.8904.

(b) Among 600 measurements, find the expected rejected count. [3 marks]

Rejected probability is one minus acceptance.

Multiply by 600.

Answer: 65.76, about 66.

9. A normal model has mean 120 and SD 10. A threshold is to exceed 97.5% of outcomes. [5 marks]

(a) Find the threshold. [3 marks]

Use lower-tail area 0.975.

Calculate $120 + 10\Phi^{-1}(0.975)$.

Answer: 139.6.

(b) A report instead uses lower-tail area 0.025. Explain the effect. [2 marks]

0.025 is the lower 2.5th percentile, below the mean.

Answer: It obtains the symmetric low cutoff instead of the high threshold.

10. A report presents a regression line and a normal model but omits the sample source, variable units and data range. [5 marks]

(a) Explain two ways this prevents a useful judgement. [3 marks]

A formula without its evidence and context is incomplete.

Answer: Readers cannot assess population relevance or whether predictions leave the observed range; missing units obscure meaning.

(b) State what should be checked before accepting the normal model. [2 marks]

A mean and SD alone do not establish normality.

Answer: Check distribution shape and context, including skewness, clusters and outliers.