Statistics and regression - Practice test 4

10 questions | 49 marks | 70 minutes

Critique claims, compare models and investigate how units and observations affect a fit.

Equivalent correct methods earn the corresponding marks.

1. A survey links weekly club participation with wellbeing scores. Students choose whether to participate. [5 marks]

(a) Identify a possible confounder. [2 marks]

Identify a third variable linked to both measurements.

Answer: Existing social support, health or free time could affect both participation and wellbeing.

(b) Does a large positive r establish that joining a club causes improved wellbeing? [3 marks]

Self-selection and confounding remain possible.

Answer: No. The observational design does not isolate a causal effect.

2. Two sets of paired data have $r = 0.85$ and $r = -0.85$. [4 marks]

(a) Compare their linear strengths and directions. [2 marks]

Compare $|r|$ for strength and sign for direction.

Answer: They have equal linear strength but opposite directions.

(b) Compare their coefficients of determination. [2 marks]

Squaring removes the sign.

Answer: Both have $r^2 = 0.7225$, or 72.25%.

3. A least-squares line is $\hat{y} = 12 + 0.4x$, where x is in centimetres. A new predictor u measures the same lengths in metres. [5 marks]

(a) Rewrite the line using u. [3 marks]

$x = 100u$.

$0.4 \times 100 = 40$.

Answer: $\hat{y} = 12 + 40u$.

(b) What happens to Pearson r? [2 marks]

The unit conversion is multiplication by a positive constant.

Answer: It is unchanged.

4. For paired data, $\bar{x} = 15, \bar{y} = 40, S_{xx} = 80$ and $S_{xy} = -120$. [5 marks]

(a) Find the fitted equation. [3 marks]

$b = -\frac{120}{80} = -1.5$.

$a = 40 - (-1.5 \times 15) = 62.5$.

Answer: $\hat{y} = 62.5 - 1.5x$.

(b) At $x = 18$, an observed y is 33. Find the residual. [2 marks]

Predicted $y = 62.5 - 27 = 35.5$.

$\text{Residual} = 33 - 35.5$.

Answer: -2.5.

5. An analyst removes a genuine extreme observation because it “makes the graph ugly”. [5 marks]

(a) Explain why this is not an acceptable statistical criterion. [2 marks]

Treatment should follow data validity and population relevance.

Answer: Appearance is not a defensible reason to discard genuine relevant data.

(b) Describe a sensitivity check. [3 marks]

The comparison reveals influence without concealing information.

Answer: Refit with and without the point and compare slopes, predictions and residual patterns, disclosing the result.

6. A fitted line predicts battery life from charge-cycle count. Its residuals are near zero early but increasingly negative later. [5 marks]

(a) Interpret the later negative residuals. [2 marks]

$$\text{Residual} = \text{observed} - \text{predicted}.$$

Answer: Later observed lifetimes are below the predictions; the model overestimates them.

(b) What does the pattern suggest? [3 marks]

A systematic residual trend indicates information left unexplained.

Answer: The model may miss a changing degradation pattern and should be reconsidered.

7. The exponential model $\hat{y} = 30e^{0.25x}$ was fitted on $0 \le x \le 6$. [5 marks]

(a) Find the multiplier and percentage increase for a one-unit rise in x. [3 marks]

Use $e^{0.25}$ as the multiplier.

Answer: Multiplier 1.284; increase 28.4%.

(b) Classify a prediction at $x = 4.5$ and one at $x = 10$. [2 marks]

Compare each x with the observed range.

Answer: 4.5: interpolation; 10: extrapolation.

8. A normal thickness model has $\mu = 2.0$ mm and $\sigma = 0.08$ mm. [5 marks]

(a) Find $P(X < 1.9)$. [3 marks]

$z = \frac{1.9 - 2.0}{0.08} = -1.25$.

Answer: 0.1056.

(b) Find the symmetric upper limit u with $P(X > u) = P(X < 1.9)$. [2 marks]

The lower limit is 0.1 mm below the mean.

By symmetry the matching upper limit is 0.1 mm above it.

Answer: $u = 2.1$ mm.

9. A normal variable has mean 200 and SD 20. [6 marks]

(a) Find its 10th and 90th percentiles. [4 marks]

Use $\mu + \sigma\Phi^{-1}(p)$ for $p = 0.10$ and 0.90.

Answer: 174.37 and 225.63.

(b) What proportion lies between these two percentiles? [2 marks]

Subtract the cumulative proportions: 0.90-0.10.

Answer: 0.80, or 80%.

10. Using the notation $N(\mu, \sigma^2)$, a report says $N(100, 25)$ has 95% of values between 50 and 150. [4 marks]

(a) Identify the parameter error. [2 marks]

The second parameter is the variance under the stated notation.

Answer: The SD is $\sqrt{25} = 5$, not 25.

(b) Give the approximate central 95% interval. [2 marks]

Use $\mu \pm 2\sigma = 100 \pm 10$.

Answer: 90 to 110.