Statistics and regression - Practice test 5

10 questions | 54 marks | 70 minutes

Combine full regression analysis, defensible reporting and normal-model decisions.

Equivalent correct methods earn the corresponding marks.

1. A small study records weekly production hours and output. [5 marks]

Paired measurements

Hours	Output (units)
1	9
2	11
3	12
4	15
5	17
6	20

(a) Draw a scatter plot and describe its form. [3 marks]

Plot corresponding hours and output values.

The responses increase with a fairly steady trend.

Answer: A positive, approximately linear association.

(b) State a limitation of the sample for predicting a different factory’s output. [2 marks]

A numerical fit need not transfer to another population.

Answer: The relationship may depend on equipment, staffing or conditions at this factory.

2. Use the production data from Question 1. [8 marks]

Paired measurements

Hours	Output (units)
1	9
2	11
3	12
4	15
5	17
6	20

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

Use least squares on the paired table.

Answer: $\hat{y} \approx 6.4 + 2.1714x$.

(b) Find r^2 and interpret it. [3 marks]

Square the unrounded Pearson r.

This is not a causal percentage.

Answer: $r^2 \approx 0.9823$; approximately 98.23% of response variation is accounted for by this fitted line.

(c) Predict output at 4.5 hours. [2 marks]

Substitute $x = 4.5$ using the unrounded coefficients.

Answer: 16.17 units.

3. For the production fit, the observation at $x = 3$ has output 12. [5 marks]

(a) Calculate its residual. [3 marks]

$\text{Residual} = 12 - (a + 3b)$.

Answer: -0.9143 units.

(b) What is the sum of residuals for the full least-squares fit before rounding? [2 marks]

The fitted model includes an intercept and uses these same observations.

Answer: Zero, apart from numerical precision.

4. A separate model has $r^2 = 0.81$ and a negative fitted slope. [5 marks]

(a) Find r. [2 marks]

The magnitude is $\sqrt{0.81} = 0.9$.

Choose the negative sign from the slope.

Answer: -0.9.

(b) Explain why a line with $r^2 = 0.81$ can still be a poor model. [3 marks]

Inspect the scatter and residual plots as well as the statistic.

Answer: A curved pattern, influential point or inappropriate domain may invalidate a simple linear interpretation.

5. For $y = 50(1.08)^t$, where t is in years: [5 marks]

(a) Write the natural-log form. [3 marks]

Apply the logarithm to the product and power.

Answer: $\ln(y) = \ln(50) + t\ln(1.08)$.

(b) Find the modelled percentage increase over two years. [2 marks]

$\text{Two} - \text{year multiplier} = 1.08^2 = 1.1664.$

$\text{Increase} = 100(1.1664 - 1).$

Answer: 16.64%.

6. A positive data set is fitted by $\ln(\hat{y}) = 2 + 0.3x$. [6 marks]

(a) Predict y at $x = 2$. [3 marks]

Compute $e^{2 + 0.3 \times 2}$.

Answer: 13.464.

(b) The original response there is 15. Find its original-scale residual. [3 marks]

Use $15 - e^{2.6}$, not $\ln(15) - 2.6$.

Answer: 1.536.

7. A normal product length has mean 80 mm. Its 90th percentile is 86.4078 mm. [6 marks]

(a) Estimate the standard deviation. [4 marks]

$86.4078 = 80 + z_{0.90}\sigma.$

Divide the difference by $\Phi^{-1}(0.90)$.

Answer: 5 mm, approximately 5 mm.

(b) Why must the standard deviation be positive here? [2 marks]

Both numerator and denominator are positive.

Answer: The 90th percentile lies above the mean and $z_{0.90}$ is positive.

8. Normal task time has mean 24 min and SD 3 min. [5 marks]

(a) Find the probability of finishing between 21 and 27 min. [3 marks]

The standardised limits are -1 and 1.

Answer: 0.6827.

(b) Find the expected number in that interval among 180 tasks. [2 marks]

Multiply the unrounded probability by 180.

Answer: 122.88, about 123.

9. A study combines two groups. Each group shows no clear within-group association, but the combined scatter plot has two clusters and a positive r. [5 marks]

(a) Why should the combined r be treated cautiously? [3 marks]

An aggregate correlation can conceal different within-group patterns.

Answer: Group membership may create the apparent association between the clusters.

(b) Give a better presentation. [2 marks]

Make the grouping visible rather than reporting one coefficient alone.

Answer: Plot and analyse the groups separately as well as together, with group identity explained.

10. A manager uses a fitted model as an exact target for every individual. [4 marks]

(a) Explain why the fitted mean tendency is not an exact individual prediction. [2 marks]

A regression summarises a tendency, not a deterministic law.

Answer: Residual variation means individuals can fall above or below the predicted response.

(b) List two items that should accompany a published prediction. [2 marks]

Readers need context to judge applicability and precision.

Answer: The data/prediction range and a statement about fit or uncertainty, with units.