Statistics and regression - Practice test 1

10 questions | 55 marks | 70 minutes

Read paired data, fit and assess a line, then use normal probabilities and cutoffs.

Equivalent correct methods earn the corresponding marks.

1. A class records weekly revision time and a quiz score for six volunteers. [5 marks]

Paired measurements

Time (hours/week)	Score (%)
1	42
2	47
3	51
4	50
5	61
6	65

(a) For predicting score from revision time, identify the explanatory and response variables. [2 marks]

The input is time; score is the predicted outcome.

Answer: Explanatory: revision time. Response: quiz score.

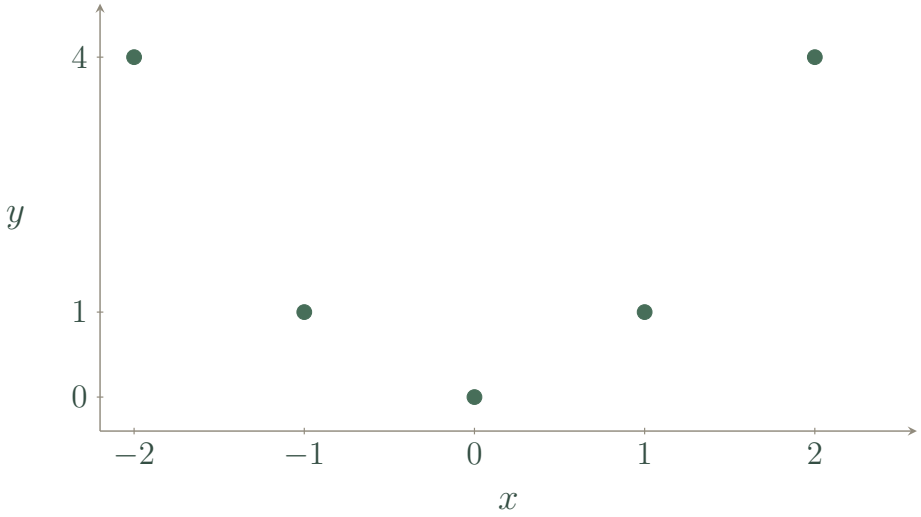
(b) Plot the paired observations on labelled axes and describe the broad direction. [3 marks]

Plot each row as (time,score), preserving its pairing.

Larger times generally accompany higher scores.

Answer: A positive association.

2. Consider points $(-2,4),(-1,1),(0,0),(1,1),(2,4)$. [4 marks]



(a) Describe their form. [2 marks]

The points follow $y = x^2$.

Answer: A U-shaped nonlinear relationship.

(b) Their correlation is zero. Explain why that does not mean the variables are unrelated. [2 marks]

The descending and ascending halves balance the linear correlation.

Answer: Pearson r measures linear association; this relationship is curved.

3. An observational survey reports that students who sleep longer tend to report better concentration. [5 marks]

(a) Give a cautious conclusion. [2 marks]

Avoid a causal claim from an observational result.

Answer: Longer sleep was associated with better reported concentration in this sample.

(b) Identify a possible confounder and describe a design improvement. [3 marks]

Measure likely confounders and improve comparability.

Random allocation of a suitable intervention can reduce confounding.

Answer: Workload may affect both. Use a controlled, ethically appropriate study and account for baseline differences.

4. For a paired sample, $S_{xx} = 40, S_{yy} = 90$ and $S_{xy} = 48$. [5 marks]

(a) Calculate r. [2 marks]

$r = \frac{48}{\sqrt{40 \times 90}} = \frac{48}{60}$.

Answer: 0.8.

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

Square 0.8.

Do not interpret this as causation.

Answer: 0.64; 64% of observed response variation is accounted for by the fitted linear relationship.

5. A sample has $\bar{x}=8, \bar{y}=30, s_x=2, s_y=5$ and $r=0.8$. [5 marks]

(a) Find the least-squares line predicting y from x. [3 marks]

$b=0.8\times\frac{5}{2}=2.$

$a=30-2\times 8=14.$

Answer: $\hat{y}=14+2x.$

(b) Predict y at $x=10$. [2 marks]

$14+2\times 10=34.$

Answer: 34.

6. A courier fits $C=7+2.4m$, where cost C is in dollars and parcel mass m is in kg. Data covered 1-12 kg. [6 marks]

(a) Interpret the slope. [2 marks]

Slope units are dollars per kilogram.

Answer: Predicted cost increases by 2.40 dollars per additional kg.

(b) Predict cost at 8 kg and classify the prediction. [2 marks]

$7+2.4\times 8=26.2.$

Eight lies inside the observed mass range.

Answer: 26.20 dollars; interpolation.

(c) Explain why the intercept is a model coefficient, not proof that a zero-mass parcel can be sent. [2 marks]

The intercept is the model’s value at zero mass.

Answer: Zero is outside the observed range and may not be a practical service.

7. The model $\hat{y} = 5 + 3x$ is used with observations $(1, 9), (2, 10), (3, 16)$. [7 marks]

(a) Calculate the three residuals. [3 marks]

Predictions are 8,11,14.

Observed minus predicted gives 9-8,10-11,16-14.

Answer: $1, -1, 2$.

(b) Calculate the sum of squared residuals. [2 marks]

$1^2 + (-1)^2 + 2^2 = 6.$

Answer: 6.

(c) Why need these residuals not sum to zero? [2 marks]

The zero-sum property requires the fitted intercept model.

Answer: The stated line was not necessarily the least-squares line fitted to these same points.

8. One observation is unusually far above a scatter pattern. A report deletes it and announces a higher r^2. [5 marks]

(a) Explain why the deletion is not justified by the higher r^2 alone. [2 marks]

Investigate the observation’s validity and relevance.

Answer: Deleting genuine data to improve a statistic can bias the report.

(b) Describe a transparent analysis. [3 marks]

Report the sensitivity and keep an audit trail.

Answer: Check the original entry and compare models with and without the observation, explaining any exclusion.

9. Packet mass X is normal with mean 250 g and standard deviation 3 g. [8 marks]

(a) Find $P(X < 247)$. [3 marks]

$z = \frac{247 - 250}{3} = -1$.

Use $\Phi(-1)$.

Answer: 0.1587.

(b) Find $P(247 < X < 256)$. [3 marks]

The upper z is 2.

Subtract $\Phi(-1)$ from $\Phi(2)$.

Answer: 0.8186.

(c) For 400 packets, find the expected number below 247 g. [2 marks]

Multiply the probability by 400.

An expected count is not a guaranteed observed count.

Answer: 63.46, approximately 63 packets.

10. Scores follow a normal model with mean 60 and SD 9. [5 marks]

(a) Find the lower cutoff for the highest 15%. [3 marks]

Use lower-tail proportion 0.85.

Calculate $60 + 9\Phi^{-1}(0.85)$.

Answer: 69.33.

(b) State one reason this cutoff should not be used for a markedly skewed score distribution. [2 marks]

Check the shape before using the model.

Answer: Its calculation assumes a normal shape; severe skewness can give inaccurate tail estimates.