Statistics and regression - Practice test 2

10 questions | 53 marks | 70 minutes

Analyse a service-data study, compare fits and explain predictions and uncertainty.

Equivalent correct methods earn the corresponding marks.

1. A delivery study records route distance x and delivery time y. [5 marks]

Paired measurements

Distance (km)	Time (min)
2	18
4	23
6	29
8	31
10	38
12	41

(a) Identify the response when predicting travel time. [2 marks]

Time is the quantity being predicted.

Answer: Delivery time in minutes.

(b) Describe one way to collect comparable observations. [3 marks]

Other conditions can affect travel time.

Answer: Use the same type of vehicle and comparable traffic conditions, recording each route’s paired distance and time.

2. Use the delivery data supplied in Question 1. [8 marks]

Paired measurements

Distance (km)	Time (min)
2	18
4	23
6	29
8	31
10	38
12	41

(a) Use technology to find the least-squares line y on x. [3 marks]

Enter distance and time as paired lists.

Select linear regression with time as response.

Answer: $\hat{y} \approx 13.8 + 2.3143x$.

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

Calculate Pearson correlation on the paired lists.

Square the unrounded r.

Answer: $r \approx 0.9933$; $r^2 \approx 0.9866$.

(c) Predict the time for a 7 km route. [2 marks]

Substitute $x = 7$ in the unrounded fitted line.

Answer: 30 min.

3. For the fitted delivery line from Question 2, consider the 8 km observation with time 31 min. [5 marks]

(a) Find its residual. [3 marks]

Use $e = 31 - (a + 8b)$ with unrounded coefficients.

Answer: -1.3143 min.

(b) Does the model overestimate or underestimate this observation? [2 marks]

The residual is negative.

Answer: It overestimates.

4. A rival report fits x on y and rearranges its line, claiming that the result must equal y on x. [5 marks]

(a) Explain the flaw. [3 marks]

Rearranging a fitted line is not generally refitting with reversed variables.

Answer: The two regressions minimise errors in different response directions.

(b) When would both fits describe the same line? [2 marks]

Then both response directions have zero residuals.

Answer: When all points lie exactly on a nonvertical, nonhorizontal line, so $|r| = 1$.

5. Two candidate delivery models have residuals $A = (-2, 1, 3, -2)$ and $B = (-1, -1, 2, 0)$. [5 marks]

(a) Compute each squared-error sum. [2 marks]

A: $4 + 1 + 9 + 4 = 18$.

B: $1 + 1 + 4 + 0 = 6$.

Answer: $A = 18$; $B = 6$.

(b) Select the smaller-error fit and give a limitation of this criterion. [3 marks]

Compare the residual sums on the same data.

Also inspect form, range and context.

Answer: B; the error comparison alone does not prove valid extrapolation or causation.

6. Adding a distant 50 km route changes a model slope from 2.1 to 3.8 min/km. [5 marks]

(a) What does this suggest about the route? [2 marks]

The fitted relationship changes substantially when it is included.

Answer: It is influential for the fitted slope.

(b) What should be checked before excluding it? [3 marks]

A valid long route may reveal that a single model is inadequate.

Answer: Check its recording, operating conditions and membership of the target population.

7. A machine's response is modelled by $\ln(\hat{y}) = 3.2 + 0.15x$. [5 marks]

(a) Write the model as $\hat{y} = Ab^x$. [3 marks]

$A = e^{3.2}$ and $b = e^{0.15}$.

Answer: $\hat{y} \approx 24.5325(1.1618)^x$.

(b) Interpret the percentage change per x-unit. [2 marks]

Convert $b - 1$ to a percentage.

Answer: An increase of about 16.18%.

8. A model from distances 2-12 km predicts time for routes of 9 km and 60 km. [5 marks]

(a) Classify each prediction. [2 marks]

Compare the inputs with 2-12.

Answer: 9 km is interpolation; 60 km is extrapolation.

(b) Explain a practical reason the long-route relationship might differ. [3 marks]

Conditions beyond the sample need not follow the same linear rule.

Answer: Long routes may use highways or breaks, changing average speed and delays.

9. Delivery time on one fixed route is normal with mean 32 min and SD 4 min. [6 marks]

(a) Find the probability of a time exceeding 38 min. [3 marks]

$z = \frac{38 - 32}{4} = 1.5.$

Use $1 - \Phi(1.5)$.

Answer: 0.0668.

(b) Find a time that 90% of deliveries are expected to finish within. [3 marks]

Use the 90th percentile, $32 + 4\Phi^{-1}(0.90)$.

Answer: 37.13 min.

10. A manager says, “The 90th-percentile time means every group of ten deliveries has exactly one late delivery.” [4 marks]

(a) Explain why this is incorrect. [2 marks]

Counts fluctuate from group to group.

Answer: A percentile specifies a model probability, not a fixed outcome in each group.

(b) For 250 comparable deliveries, state the expected number exceeding that cutoff. [2 marks]

$\text{Expected count} = 250 \times 0.10.$

Answer: 25.