15 Scatterplots, model residuals and predictions

Name: _____ Class: _____ Date: _____ **Total: 13 marks**

KEY WORDS extrapolation • residual • evidence

Objective

- Interpret slope and intercept of a fitted model in context
- Calculate a residual and distinguish interpolation from extrapolation
- Choose linear or exponential patterns from equal-step data

Original public practice; marks below are classroom self-check points, not official criteria.

1 Key knowledge and worked examples

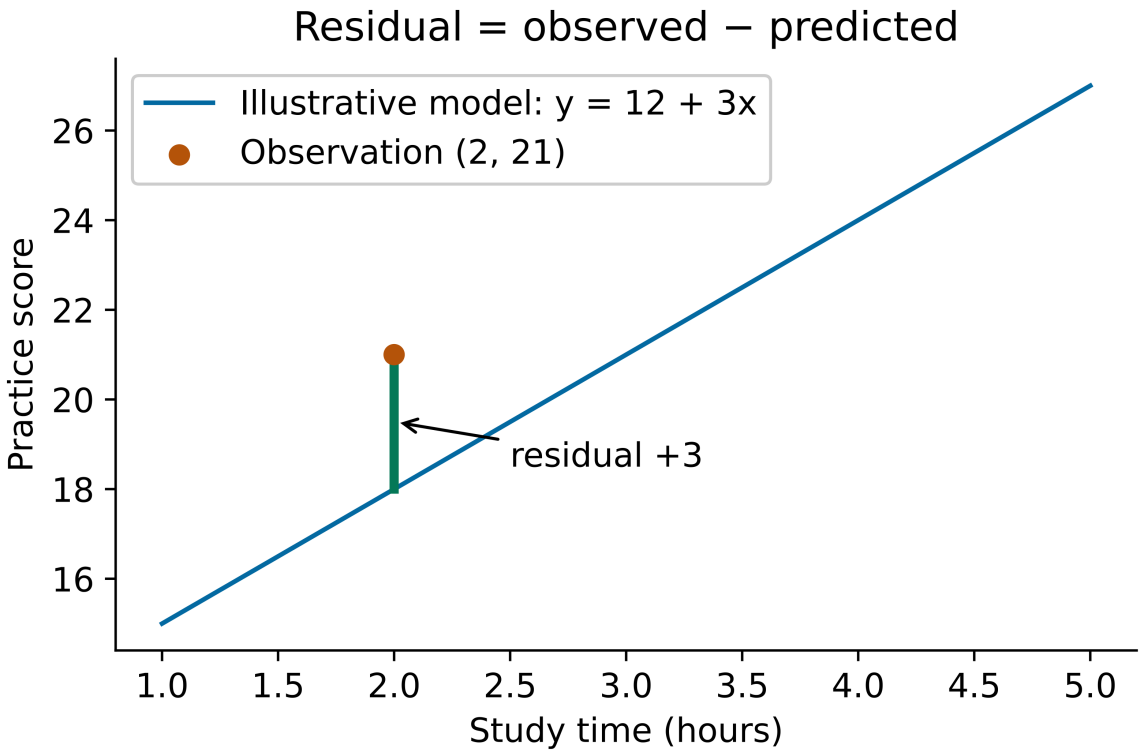
A scatterplot pairs two quantitative variables, one point per paired observation. A fitted line summarises a trend rather than connecting each point. In a model y=a+bx, b predicts the change in y for one-unit increase in x; its units are y-units per x-unit. The intercept a predicts y when x=0, which may lie outside the data range.

A residual is observed y minus predicted y. Positive means the observation lies above the model; negative means below. Substitute the observed x into the model before subtracting. A large residual marks a poor prediction for that point, not automatically an error in measurement or proof that the trend is absent.

Interpolation predicts inside the observed x-range; extrapolation goes beyond it and needs stronger caution because the relationship may change. Use a model's contextual domain as well as its algebraic form. A negative predicted travel time, for example, can reveal an unjustified extrapolation rather than a physically possible outcome.

For equal input steps, nearly constant output differences suggest a linear model; nearly constant output ratios suggest an exponential model. Scatter can make either pattern approximate. Association does not identify a cause: both variables may reflect a third factor. A fitted trend is evidence about prediction, not by itself about intervention.

A residual is observed minus predicted output. A journey model $T = 5 + 2d$ uses minutes for T and kilometres for d. $T(4) = 5\,\text{min} + (2\,\text{min/km})(4\,\text{km}) = 13\,\text{min}$. For observed time 16 minutes, $e = T_{obs} - T_{pred} = 16\,\text{min} - 13\,\text{min} = 3\,\text{min}$.



Residual = observed − predicted = 21 − 18 = 3

Worked example. For study time x hours and predicted practice score y=12+3x, the slope is 3 score points/hour. At x=2 the prediction is 18. If the observed score is 21, residual=21-18=3. With x-values observed from 1 to 5, predicting at x=3 interpolates and x=12 extrapolates. Outputs 5,10,20,40 at equal input steps suggest factor-two exponential growth.

2 Practice

2.1 Model predicts 30; observation is 26. Residual: [1]

- **A** -4
- **B** 4
- **C** 56

2.2 Observed x ranges from 2 to 8. Predicting at x=5 is: [1]

- **A** Interpolation
- **B** Extrapolation
- **C** Proof of causation

2.3 For y=4+2x, prediction at x=7 is _____. [1]

2.4 A strong scatterplot association alone proves x causes y. [1]

3 Exam-style questions

Use the data stated in each original task.

3.1 A fitted delivery model is $T = 8 + 1.5d$ minutes for observed distances 2–10 km. Interpret both parameters. At 6 km the observed time is 15 minutes; find its residual. [3]

3.2 Predictions are requested at 7 km and 20 km. Classify each, and explain why a fitted slope does not prove distance alone causes all time variation. [3]

3.3 At equally spaced inputs 0,1,2,3, outputs are 6,18,54,162. Choose a simple linear or exponential rule and predict at 4, stating the assumption. [3]

Solutions

Accept equivalent correct methods and supported interpretations. The marks check reasoning; they are not an official grade.

2.1 A – 26-30=-4. B reverses observed minus predicted; C adds values that must be subtracted.

Correct response linked to the given case.

2.2 A – 5 is inside the observed range. B requires an input outside 2–8; C makes a causal claim that interpolation cannot establish.

Correct response linked to the given case.

2.3 18 – 4+2 · 7=18.

Correct response linked to the given case.

2.4 False – Confounding or other explanations can produce association.

Correct response linked to the given case.

3.1 The intercept predicts 8 minutes at zero distance; zero lies outside the observed range, so this interpretation is extrapolated. The slope predicts an extra 1.5 minutes per kilometre. $T(6) = 8\,\text{min} + (1.5\,\text{min/km})(6\,\text{km}) = 17\,\text{min}$. Residual $e = 15\,\text{min} - 17\,\text{min} = -2\,\text{min}$, below the line.

Sound conclusion or proposed method. Relevant supporting steps or textual evidence. Accurate conditions, checks or qualification.

3.2 Seven is inside 2–10, so interpolation. Twenty is outside, so extrapolation with weaker support. Traffic, loading and other conditions can influence time; a fitted association does not isolate their effects.

Sound conclusion or proposed method. Relevant supporting steps or textual evidence. Accurate conditions, checks or qualification.

3.3 Ratios are all 3, while differences vary. The exponential rule is $y = 6(3)^x$. Under continued factor-three growth, $y(4) = 6(3)^4 = 486$. This is a model-based extrapolation, not a guaranteed next observation.

Sound conclusion or proposed method. Relevant supporting steps or textual evidence. Accurate conditions, checks or qualification.