2.2 Regression, financial models and residuals

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

KEY WORDS residual

Objective

Regression, financial models and residuals. Original classroom practice for IB HL; not an official IB paper.

1 Key knowledge and worked examples

Is the prediction supported by the data?

- A line predicts bus travel time from distance. A small calculation error matters less than using a model outside its evidence.
- This lesson studies **residual** : The observed value minus the value predicted by a fitted model.

The method

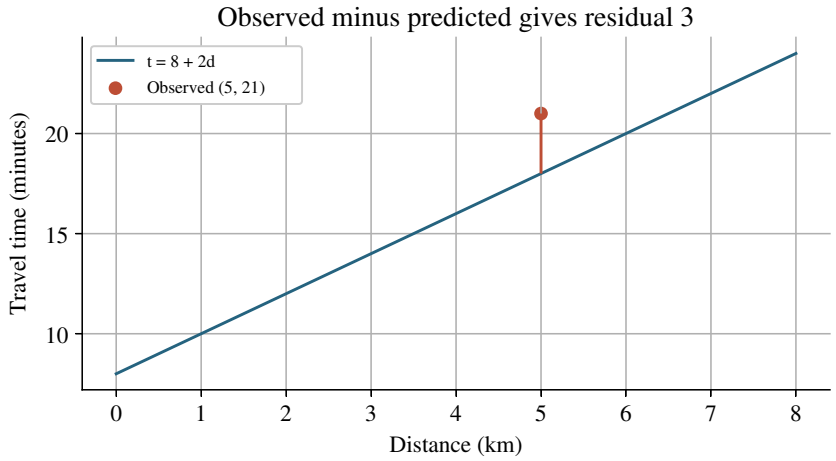
- A linear model y=a+bx has intercept a and slope b with contextual units.
- Inspect residuals and the data range.
- For repeated percentage change use a geometric model; for a loan distinguish principal, rate, repayment and period.

A worked case

$$\widehat{t} = 8 + 2d, \qquad e = t - \widehat{t}$$

Example:

The fitted model t=8+2d predicts 18 minutes at d=5 km. If the observed time is 21, the residual is 3 minutes. A ¥1000 deposit at 5% compound annual interest becomes 1000×1.05^3=1157.625 after 3 years.



Where it goes wrong

- A good fit does not prove a causal mechanism.
- A correlation coefficient measures linear association, not the gradient.
- Calculator output must be translated into a model, checked and interpreted.

Not true: A high correlation permits reliable extrapolation to any distance. A model fitted to data is reliable only near that data; extrapolating far outside it is unreliable.

Applying it

- In IB AI, record the data source, domain, assumptions and calculator method.
- Compare an alternative model and judge predictions against residuals.
- A financial answer must state payment timing and whether interest is compounded.

In your course

- Applications and Interpretation HL, the course first assessed in 2021.

Key:

- **Residual**: The observed value minus the value predicted by a fitted model.
- A linear model y=a+bx has intercept a and slope b with contextual units.
- Watch for: A good fit does not prove a causal mechanism.

■ **Write from a prediction of 32**

- The model t = 2 + 6d predicts the time at d = 5.
- An observed time of 41 leaves a residual. A high correlation does not license extrapolation.

2 Practice

2.1 For t=8+2d, find t at d=5. Show the relation, substitution and units. [2]

2.2 Find the residual if observed t=21 and predicted t=18. Show the relation, substitution and units. [2]

2.3 Find the value of 1000 after 3 years at 5% compound annual interest. Give it to 3 significant figures. Show the relation, substitution and units. [2]

2.4 Decide whether this claim is true and justify: A high correlation permits reliable extrapolation to any distance. [2]

2.5 The model t=8+2d fits observed journey times. An observed 21 minutes against a predicted 18 means what? Choose one answer.

- **A** The observation should be discarded because it does not fit the model
- **B** A residual of +3 minutes: the model under-predicted this observation
- **C** The model is exact, so the 3-minute gap must be a measurement error [1]

3 Exam-style questions

3.1 A fictional journey model is t=8+2 d. An observed trip at d=5 takes 21 minutes. Find the residual, predict at d=50 and explain why that prediction needs caution. [4]

Solutions

2.1 Substitute distance 5: $8+2\times5=18$ minutes.

2.2 Residual=observation-prediction=21-18=3 minutes.

2.3 Compound annually: $1000\times1.05^3=1157.625$.

2.4 False. A model fitted to data is reliable only near that data; extrapolating far outside it is unreliable.

2.5 B Residual = observed - predicted = 21-18 = 3. Residuals reveal systematic model weakness; a good fit on average can still hide poor regions.

3.1 Prediction 18, residual+3 minutes; at 50 prediction 108. If the fitted distance range was short, this is extrapolation; traffic, stops and model form may change. Regression does not establish a causal coefficient by itself.