

1.4 Regression, financial models and residuals

Name: _____ Class: _____ Date: _____

Total: 13 marks

KEY WORDS residual

Objective

Regression, financial models and residuals. Supported HL concept preparation; original classroom practice, not an official IB paper or assessed learner work.

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.
-

■ **Choose the mathematical structure**

- 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.
 - State the allowed inputs and units before calculating. An equation should express the relationship, not just record a calculator entry.
-

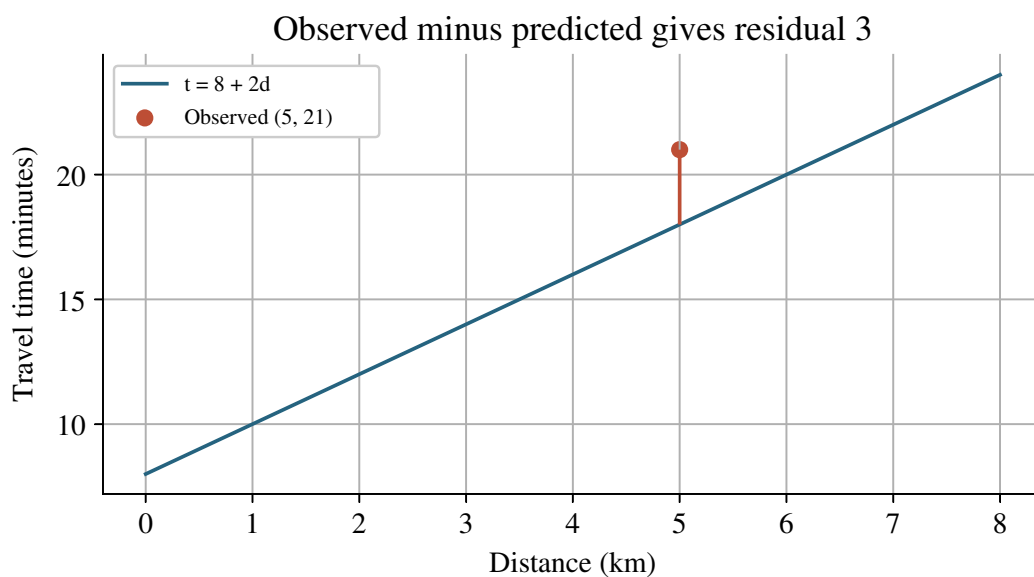
■ **Work through a checked case**

$$\hat{t} = 8 + 2d, \quad e = t - \hat{t}$$

- Check the result against the starting quantities. Substitute into the original relation, or compare the graph and numerical answer where appropriate.

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 $1000 \times 1.05^3 = 1157.625$ after 3 years.



■ Test a tempting shortcut

- 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.
- When a shortcut fails, identify the assumption it breaks. Keep an exact value until the requested final rounding.

Warn:

A high correlation permits reliable extrapolation to any distance. This claim is false. Explain which definition or assumption it violates.

■ Interpret a new situation

- 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.
 - A complete solution gives the mathematical result and explains what it means. Check that it is possible in the stated context.
-

■ Use this in your course

- Current first-assessment-2021 Applications and Interpretation HL. This is authored concept support; the full guide is needed to certify every objective.
- Give the method before the final answer, and use the paper's calculator and formula rules. Review a wrong answer by locating the first invalid step.

Key:

The observed value minus the value predicted by a fitted model. Choose the relationship, show the method, check its assumptions and interpret the result.

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. 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 Which description correctly defines residual? Choose one answer.

- **A** The observed value minus the value predicted by a fitted model.
- **B** A numerical answer that needs no assumptions or interpretation
- **C** A rule that remains valid even when its required conditions fail

[1]

3 Exam-style questions

3.1 A fictional journey model is $t=8+2d$. 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: $8+2\times 5=18$ minutes.

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

2.3 Compound annually: $1000\times 1.05^3=1157.625$.

2.4 False. 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.

2.5 A The observed value minus the value predicted by a fitted model.

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.