

Multivariable Data Analysis and Model Critique

g11_u09_w35

Purpose and boundary

Use original fictional spreadsheet-style tables only. No data collection, personal-data analysis, experiment, or real-world decision is requested.

Targets: I can compare patterns with more than one variable, identify confounding, and critique a model without overstating association. **Vocabulary:** association, confounder, residual, predictor, response, multivariable. **Prerequisites:** graphs, averages, residuals, causal reasoning.

For items 4–12 use fictional data cards with columns X, Z, and response Y. Tables are instructional models, not evidence about people or real systems.

1. Distinguish association from causation.

2. What is a confounder?

3. Why can adding a second variable change an apparent trend?

4. Card A has rows $(X, Z, Y) = (1, 1, 4), (2, 1, 6), (3, 1, 8)$. Describe the X–Y pattern while Z is fixed.

5. Card B has rows $(1, 2, 5), (2, 2, 7), (3, 2, 9)$. Compare it with Card A.

6. What model question could distinguish an X effect from a Z effect?

7. A prediction model gives 6.0 and observed Y is 6.4. Find residual observed minus predicted.

8. A second residual is -0.3. Which observation has larger residual magnitude?

9. Why does one residual not prove a model is good or bad?

10. A residual pattern becomes larger as X increases. What model concern does that suggest?

11. A table has one extreme row. Name one question before removing it.

12. Explain why removing a row after seeing results can introduce bias.

13. Two groups have different average Y values but also different average Z. Why is a direct group cause claim limited?

14. A student adds many predictors and claims certainty. Name one risk of overfitting.

15. Why should a model be evaluated on conditions beyond its fitting rows?

16. State one supported and one unsupported claim from a fictional association table.

17. Propose one additional column that could test a confounder.

18. Explain why a graph scale can influence pattern perception.

19. Write a bounded model critique that names residual evidence and one alternative explanation.

20. Write a two-sentence conclusion that separates a data pattern from a causal claim.
