

Statistics and Data Analysis

Unit 04 – Lecture 01 Notes

Tofik Ali

February 14, 2026

Topic

Correlation vs regression concepts; causation warning; residual idea.

Learning Outcomes

- Differentiate correlation and regression
- Explain why correlation does not imply causation
- Interpret a scatter plot (trend, outliers)
- Define residual and why residuals matter

Detailed Notes

These notes are designed to be read alongside the slides. They expand each slide bullet into plain-language explanations, small worked examples, and common pitfalls. When a formula appears, emphasize (1) what each symbol means, (2) the assumptions needed to use it, and (3) how to interpret the final number in the problem context.

Concepts

- Correlation measures linear association
- Regression models Y as a function of X
- Regression has roles: predictors vs response

Causation Warning

- Confounding can create misleading correlation
- Reverse causality is possible
- Causal claims need design or strong assumptions

Exercises (with Solutions)

Exercise 1: Pick response variable

Predict house price using size and location. What is the response variable?

Solution

- House price is the response (Y).

Exercise 2: Interpret r

If $r=0.7$ between study hours and score, what does it mean?

Solution

- Strong positive linear association.
- Not proof of causation.

Exercise 3: Residual sign

If $y=74$ and $\hat{y}=80$, what is residual?

Solution

- Residual = $y - \hat{y} = -6$ (over-prediction).

Exit Question

Give one example of a confounder that can create a misleading correlation.

Demo (Python)

Run from the lecture folder:

```
python demo/demo.py
```

Output files:

- `images/demo.png`
- `data/results.txt`

References

- Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers*, Wiley.
- Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, Cengage.
- McKinney, W. *Python for Data Analysis*, O'Reilly.