Statistics and Data Analysis
Unit 04 – Lecture 03 Notes

Tofik Ali

February 14, 2026

Topic

Multiple predictors; partial effects; dummy variables; adjusted R-squared (overview).

Learning Outcomes

- Write the multiple linear regression model
- Interpret a coefficient as a partial effect
- Explain dummy variables for categories (basic)
- Explain adjusted R-squared (intuition)

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.

Model

- y = b0 + b1 x1 + b2 x2 + ...
- Each coefficient is a partial effect (others fixed)
- Scaling helps when using regularization

Interpretation

- Dummy variables encode categories
- Adjusted R^2 penalizes unnecessary predictors
- Multicollinearity can harm interpretability

Exercises (with Solutions)

Exercise 1: Partial effect

Model: yhat=5 + 0.8x1 + 2.0x2. Interpret coefficient 2.0.

Solution

- Holding x1 fixed, +1 in x2 increases yhat by 2.0 units.

Exercise 2: Dummy variable

Urban=1, Rural=0. If coef(Urban)=10, interpret.

Solution

- Urban has predicted y about 10 units higher than Rural (all else equal).

Exercise 3: Adjusted R^2

Why use adjusted R^2 when comparing models with different number of predictors?

Solution

- Because R^2 never decreases, adjusted R^2 penalizes extra predictors.

Exit Question

Why does adding a useless feature still increase (or keep) R^2?

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.