# Statistics and Data Analysis
# Unit 04 – Lecture 03 Notes
# Multiple Linear Regression

Tofik Ali

February 17, 2026

## Topic

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

## How to Use These Notes

These notes are written for students who are seeing the topic for the first time. They follow the slide order, but add the missing 'why', interpretation, and common mistakes. If you get stuck, look at the worked exercises and then run the Python demo.

Course repository (slides, demos, datasets): `https://github.com/tali7c/Statistics-and-Data-Analysis`

## Time Plan (55 minutes)

- 0–10 min: Attendance + recap of previous lecture
- 10–35 min: Core concepts (this lecture's sections)
- 35–45 min: Exercises (solve 1–2 in class, rest as practice)
- 45–50 min: Mini demo + interpretation of output
- 50–55 min: Buffer / wrap-up (leave 5 minutes early)

## Slide-by-slide Notes

### Title Slide

State the lecture title clearly and connect it to what students already know. Tell students what they will be able to do by the end (not just what you will cover).

### Quick Links / Agenda

Explain the structure of the lecture and where the exercises and demo appear.

- Overview

- Model
- Interpretation
- Exercises
- Demo
- Summary

## 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)

**Why these outcomes matter.** **Regression** models a response variable $Y$ as a function of predictor(s) $X$. It has direction (predictors -> response), produces a fitted equation, and lets you predict and explain. Regression is not automatically causal; causality needs design or strong assumptions. **Adjusted** $R^2$ adds a penalty for extra predictors. It is better than $R^2$ for comparing models with different numbers of features, but it is still an in-sample measure and should be complemented with validation/test performance.

## Model: Key Points

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

## Model: Key Formula

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \epsilon$$

## Interpretation: Key Points

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

**Explanation.** $R^2$ is the fraction of variance in $Y$ explained by the model (in-sample). It increases when you add predictors, even useless ones, so it is not a guarantee of a good model. Use residual diagnostics and out-of-sample evaluation to judge model quality. **Adjusted** $R^2$ adds a penalty for extra predictors. It is better than $R^2$ for comparing models with different numbers of features, but it is still an in-sample measure and should be complemented with validation/test performance. A **dummy variable** (one-hot encoding) converts a category into 0/1 indicators so regression models can use it. Interpret coefficients relative to the reference (dropped) category, and avoid the dummy-variable trap by not including all categories at once when an intercept is present.

## Exercises (with Solutions)

Attempt the exercise first, then compare with the solution. Focus on interpretation, not only arithmetic.

## 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).

**Walkthrough.** A **dummy variable** (one-hot encoding) converts a category into 0/1 indicators so regression models can use it. Interpret coefficients relative to the reference (dropped) category, and avoid the dummy-variable trap by not including all categories at once when an intercept is present.

## 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.

**Walkthrough.** $R^2$ is the fraction of variance in $Y$ explained by the model (in-sample). It increases when you add predictors, even useless ones, so it is not a guarantee of a good model. Use residual diagnostics and out-of-sample evaluation to judge model quality. **Adjusted** $R^2$ adds a penalty for extra predictors. It is better than $R^2$ for comparing models with different numbers of features, but it is still an in-sample measure and should be complemented with validation/test performance.

### Mini Demo (Python)

Run from the lecture folder:

`python demo/demo.py`

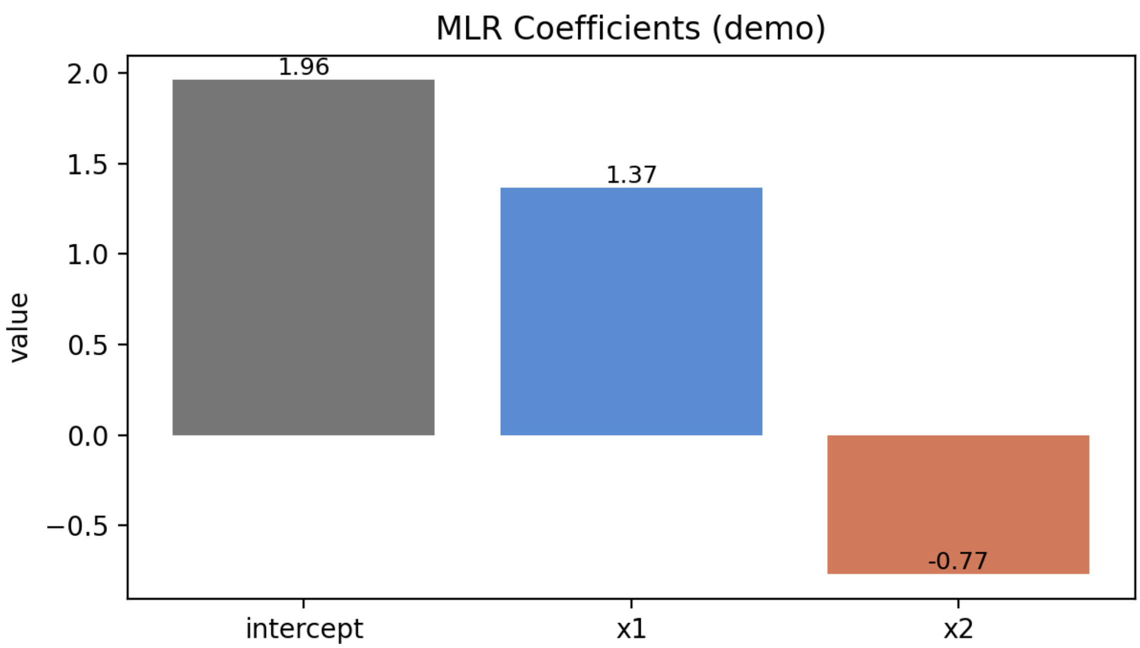
Output files:

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

#### What to show and say.

- Fits multiple linear regression with numeric + categorical (dummy) features.
- Shows how coefficients change when controlling for other variables.
- Use it to discuss adjusted $R^2$ and interpretation vs prediction goals.

### Demo Output (Example)




### Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

### Exit Question

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

**Suggested answer (for revision).** $R^2$ never decreases when adding predictors because the model can always set a new coefficient to 0 and keep the same fit (or improve it).

## 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.

## Appendix: Slide Deck Content (Reference)

The material below is a reference copy of the slide deck content. Exercise solutions are explained in the main notes where applicable.

### Title Slide

https://github.com/tali7c/Statistics-and-Data-Analysis

## Quick Links

**Overview Model Interpretation Exercises Demo Summary**

## Agenda

- Overview
- Model
- Interpretation
- Exercises
- Demo
- Summary

## 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)

## Model: Key Points

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

## Model: Key Formula

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \epsilon$$

## Interpretation: Key Points

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

## Exercise 1: Partial effect

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

## Solution 1

- 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 2

- 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 3

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

## Mini Demo (Python)

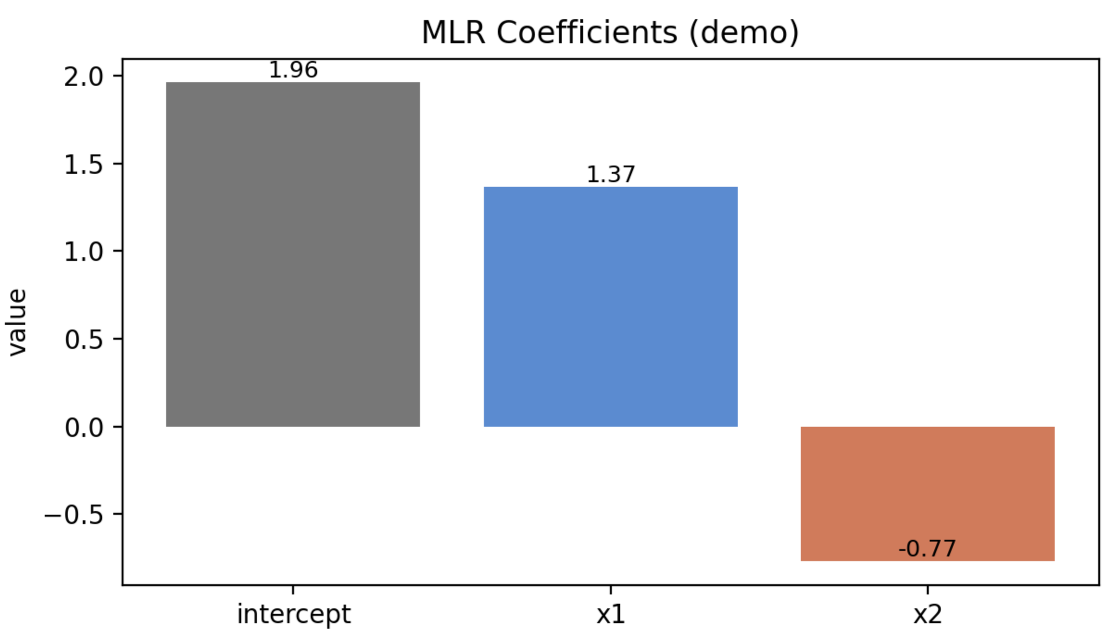
Run from the lecture folder:

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

Outputs:

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

## Demo Output (Example)



## Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

## Exit Question

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