

# Statistics and Data Analysis

## Unit 04 – Lecture 05: Multicollinearity

Tofik Ali

School of Computer Science, UPES Dehradun

February 14, 2026

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

# Quick Links

Overview

What and Why

Detection

Exercises

Demo

Summary

# Agenda

- 1 Overview
- 2 What and Why
- 3 Detection
- 4 Exercises
- 5 Demo
- 6 Summary

# Learning Outcomes

- Define multicollinearity (high correlation among predictors)

# Learning Outcomes

- Define multicollinearity (high correlation among predictors)
- Explain why it harms interpretation (unstable coefficients)

# Learning Outcomes

- Define multicollinearity (high correlation among predictors)
- Explain why it harms interpretation (unstable coefficients)
- Recognize symptoms (large SEs, unstable signs)

# Learning Outcomes

- Define multicollinearity (high correlation among predictors)
- Explain why it harms interpretation (unstable coefficients)
- Recognize symptoms (large SEs, unstable signs)
- List common fixes (drop/combine/regularize)

# What and Why: Key Points

- Predictors overlap in information



# What and Why: Key Points

- Predictors overlap in information
- Coefficients become unstable

# What and Why: Key Points

- Predictors overlap in information
- Coefficients become unstable
- Prediction may still be OK but interpretation suffers

# Detection: Key Points

- Correlation matrix/heatmap (screening)

# Detection: Key Points

- Correlation matrix/heatmap (screening)
- VIF (next)

# Detection: Key Points

- Correlation matrix/heatmap (screening)
- VIF (next)
- Condition number (advanced)

# Exercise 1: Identify

If  $\text{corr}(x_1, x_2) = 0.98$ , what risk do you expect?

# Solution 1

- High multicollinearity; unstable coefficients.

## Exercise 2: Fix

Name one fix for multicollinearity.



## Solution 2

- Drop one feature, combine features, or use ridge/PCA.

## Exercise 3: Prediction vs interpretation

Can multicollinearity still allow good prediction?

# Solution 3

- Yes, but individual coefficients are unreliable.

# Mini Demo (Python)

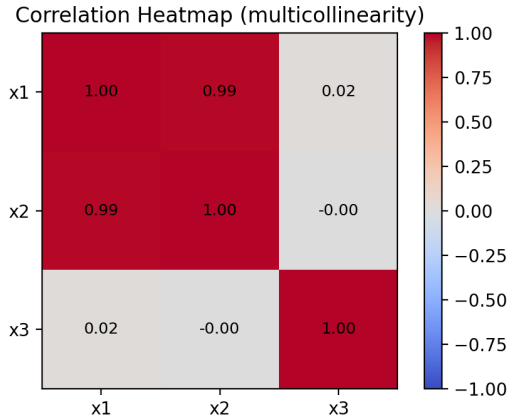
Run from the lecture folder:

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

Outputs:

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

# Demo Output (Example)



# Summary

- Key definitions and the main formula.

# Summary

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

# Summary

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



# Exit Question

What does multicollinearity break first: prediction or interpretation (and why)?