

# Statistics and Data Analysis

## Unit 03 – Lecture 05 Notes

Tofik Ali

February 14, 2026

### Topic

One-way ANOVA; F statistic; assumptions; post-hoc comparisons.

### Learning Outcomes

- Explain why ANOVA is used for comparing 3+ means
- Describe between-group vs within-group variation
- Interpret F statistic at a high level
- State main assumptions of one-way ANOVA
- Explain what a post-hoc test is

### 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.

### ANOVA Concept

- One global test for equality of means
- Avoids inflating Type I error vs many t-tests
- If significant, follow with post-hoc

### Assumptions

- Independent observations
- Rough normality within groups (or robust with n)
- Similar variances across groups

## Exercises (with Solutions)

### Exercise 1: Write H0

Compare 3 group means. What is H0?

### Solution

- $H_0: \mu_1 = \mu_2 = \mu_3$

### Exercise 2: Within variance

If within-group variance increases, what happens to F (all else equal)?

### Solution

- F tends to decrease; harder to detect differences.

### Exercise 3: Next step

ANOVA p-value is 0.01 at alpha=0.05. What next?

### Solution

- Reject H0.
- Run post-hoc to find which pairs differ.

## Exit Question

Why are several pairwise t-tests not equivalent to one ANOVA?

## 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.