

Statistics and Data Analysis

Unit 03 – Lecture 06 Notes

Tofik Ali

February 14, 2026

Topic

Rank-based tests and p-value interpretation; statistical vs practical significance.

Learning Outcomes

- Explain why non-parametric tests are used
- Choose Mann-Whitney / Wilcoxon / Kruskal-Wallis
- Interpret p-values carefully
- Discuss statistical vs practical significance
- Explain multiple testing risk

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.

When to Use

- Skewed data/outliers
- Ordinal scales
- Small sample and doubtful normality

Common Tests

- Two independent groups: Mann-Whitney U
- Paired samples: Wilcoxon signed-rank
- 3+ groups: Kruskal-Wallis

Exercises (with Solutions)

Exercise 1: Choose test

Same students before/after training (skewed). Which test?

Solution

- Wilcoxon signed-rank

Exercise 2: Practical vs statistical

Very small p-value but tiny difference: what should you report?

Solution

- Report effect size and context; significance \neq importance.

Exercise 3: Multiple testing

20 tests at $\alpha=0.05$: expected false positives?

Solution

- About 1 on average.

Exit Question

Give one reason to prefer a rank-based test over a mean-based test.

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.