Statistics and Data Analysis
Unit 03 – Lecture 07 Notes
Case Exercise: Interpreting Hypothesis Testing Results

Tofik Ali

February 17, 2026

Topic

Interpret published hypothesis testing results; emphasize CI, effect size, and pitfalls.

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

- Reading Results
- Pitfalls
- Exercises
- Demo
- Summary

Learning Outcomes

- Interpret p-values and confidence intervals correctly
- Compute a simple effect size from summary statistics
- Identify common red flags (only p-values, many tests, no effect size)
- Write a cautious conclusion in plain language
- Avoid correlation-causation confusion

Why these outcomes matter. A **confidence interval (CI)** is an interval estimate, not a single number. The correct interpretation is long-run: if we repeated the same sampling procedure many times, about 95% of the computed 95% intervals would contain the true population value. It is **not** correct to say there is a 95% probability the parameter lies in a particular computed interval. A **p-value** is computed assuming the null hypothesis H_0 is true. It measures how surprising the observed data (or something more extreme) would be under H_0. A small p-value suggests the data is hard to explain by H_0 alone, but it does not tell you how large the effect is or whether it is practically important.

Reading Results: Key Points

- Check n, center, spread
- Prefer CI + effect size
- Ask: what does it mean in the real world?

Explanation. **Effect size** quantifies *how big* a difference/relationship is (e.g., Cohen's d, correlation r). With large samples, even tiny effects can be statistically significant, so reporting effect size prevents over-claiming.

Pitfalls: Key Points

- Multiple comparisons
- Selective reporting (p-hacking)
- Over-claiming causation

Exercises (with Solutions)

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

Exercise 1: Interpret CI

95% CI for (new-old) is (1.2, 3.8). What does it suggest?

Solution

- Likely positive effect (CI above 0).
- Magnitude between 1.2 and 3.8 units.

Exercise 2: Compute d

A: n=20 mean=72 SD=10; B: n=20 mean=68 SD=10. Compute Cohen's d.

Solution

- Pooled SD=10
- d=(72-68)/10=0.4

Exercise 3: Cautious conclusion

p-value=0.03 but effect size is tiny. What should you conclude?

Solution

- Evidence of difference, but small magnitude.
- May not justify action without cost/benefit.

Walkthrough. A **p-value** is computed assuming the null hypothesis H_0 is true. It measures how surprising the observed data (or something more extreme) would be under H_0. A small p-value suggests the data is hard to explain by H_0 alone, but it does not tell you how large the effect is or whether it is practically important. **Effect size** quantifies *how big* a difference/relationship is (e.g., Cohen's d, correlation r). With large samples, even tiny effects can be statistically significant, so reporting effect size prevents over-claiming.

Mini Demo (Python)

Run from the lecture folder:

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

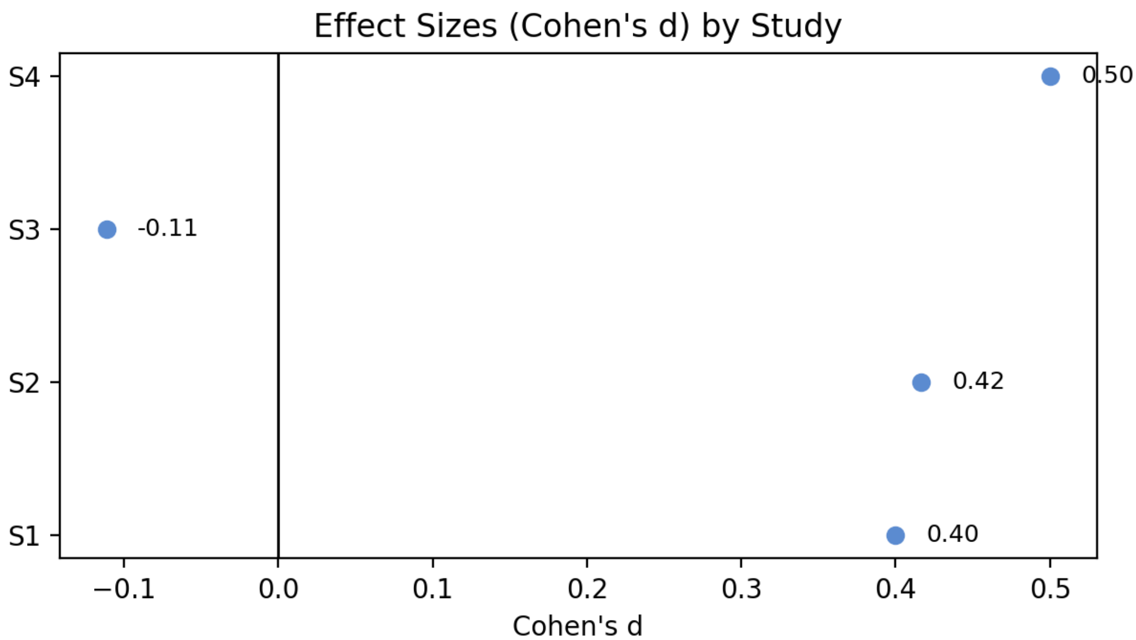
Output files:

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

What to show and say.

- Computes simple effect sizes (Cohen's d) from multiple study summaries.
- Plots effect sizes around 0 to practice interpretation and cautious conclusions.
- Use it to discuss why we want CI/effect size, not only p-values.

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

What is one red flag when a paper reports only p-values and no effect sizes?

Suggested answer (for revision). Only reporting p-values hides the magnitude; without effect sizes/CIs you cannot judge whether a result matters in practice.

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

Quick Links

Overview **Reading Results** **Pitfalls** **Exercises** **Demo** **Summary**

Agenda

- Overview
- Reading Results
- Pitfalls
- Exercises
- Demo
- Summary

Learning Outcomes

- Interpret p-values and confidence intervals correctly
- Compute a simple effect size from summary statistics
- Identify common red flags (only p-values, many tests, no effect size)
- Write a cautious conclusion in plain language
- Avoid correlation-causation confusion

Reading Results: Key Points

- Check n, center, spread
- Prefer CI + effect size
- Ask: what does it mean in the real world?

Pitfalls: Key Points

- Multiple comparisons
- Selective reporting (p-hacking)
- Over-claiming causation

Exercise 1: Interpret CI

95% CI for (new-old) is (1.2, 3.8). What does it suggest?

Solution 1

- Likely positive effect (CI above 0).
- Magnitude between 1.2 and 3.8 units.

Exercise 2: Compute d

A: n=20 mean=72 SD=10; B: n=20 mean=68 SD=10. Compute Cohen's d.

Solution 2

- Pooled SD=10
- d=(72-68)/10=0.4

Exercise 3: Cautious conclusion

p-value=0.03 but effect size is tiny. What should you conclude?

Solution 3

- Evidence of difference, but small magnitude.
- May not justify action without cost/benefit.

Mini Demo (Python)

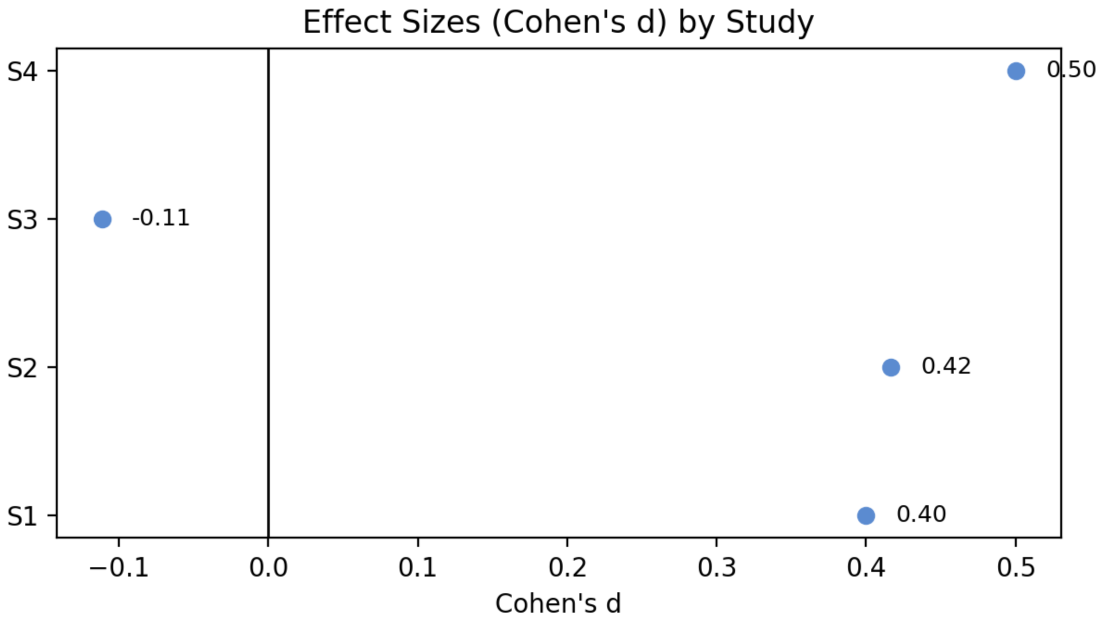
Run from the lecture folder:

```
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

What is one red flag when a paper reports only p-values and no effect sizes?