Statistics and Data Analysis
Unit 03 – Lecture 03 Notes
Hypothesis Testing (t-test): Paired Test and Effect Size

Tofik Ali

February 17, 2026

Topic

Paired t-test, mean difference, effect size, and interpretation.

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

- Paired Design
- Effect Size
- Exercises
- Demo
- Summary

Learning Outcomes

- Differentiate paired vs independent designs
- Compute within-pair differences di
- Run a paired t-test (conceptually)
- Explain effect size and why we report it
- Interpret results in context (not only p-value)

Why these outcomes matter. 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.

Paired Design: Key Points

- Same unit measured twice (before/after)
- Analyze differences di = after - before
- Pairing reduces noise from individual differences

Explanation. **Degrees of freedom (df)** roughly represent how much independent information is available to estimate variability. For a one-sample t-test, df $= n - 1$ because one constraint is used to estimate the sample mean. df affects the critical values and the shape of the t-distribution (small df -> heavier tails). **Differencing** transforms a series by subtracting the previous value: $y_t - y_{t-1}$. It removes trend and can help achieve stationarity. Over-differencing can add noise, so use the smallest differencing order that works.

Paired Design: Key Formula

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \quad \text{df} = n - 1$$

How to read the formula. **Degrees of freedom (df)** roughly represent how much independent information is available to estimate variability. For a one-sample t-test, df $= n - 1$ because one constraint is used to estimate the sample mean. df affects the critical values and the shape of the t-distribution (small df -> heavier tails).

Effect Size: Key Points

- p-value answers: evidence?
- Effect size answers: how big?
- Large n can make tiny effects significant

Explanation. 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.

Effect Size: Key Formula

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}}$$

How to read the formula. **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.

Exercises (with Solutions)

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

Exercise 1: Compute differences

Before/After: (10,12), (12,12), (11,14), (9,10). Compute di and dbar.

Solution

- di: 2,0,3,1
- dbar = 1.5

Walkthrough. **Differencing** transforms a series by subtracting the previous value: $y_t - y_{t-1}$. It removes trend and can help achieve stationarity. Over-differencing can add noise, so use the smallest differencing order that works.

Exercise 2: CI idea

If the 95% CI for mean difference excludes 0, what does it suggest?

Solution

- Evidence of a change (difference likely non-zero).
- Check magnitude and context.

Walkthrough. **Differencing** transforms a series by subtracting the previous value: $y_t - y_{t-1}$. It removes trend and can help achieve stationarity. Over-differencing can add noise, so use the smallest differencing order that works.

Exercise 3: Interpret d

If Cohen's d=0.3, what does it suggest (rule of thumb)?

Solution

- Small effect (context dependent).
- Still may matter if cheap/safe to adopt.

Mini Demo (Python)

Run from the lecture folder:

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

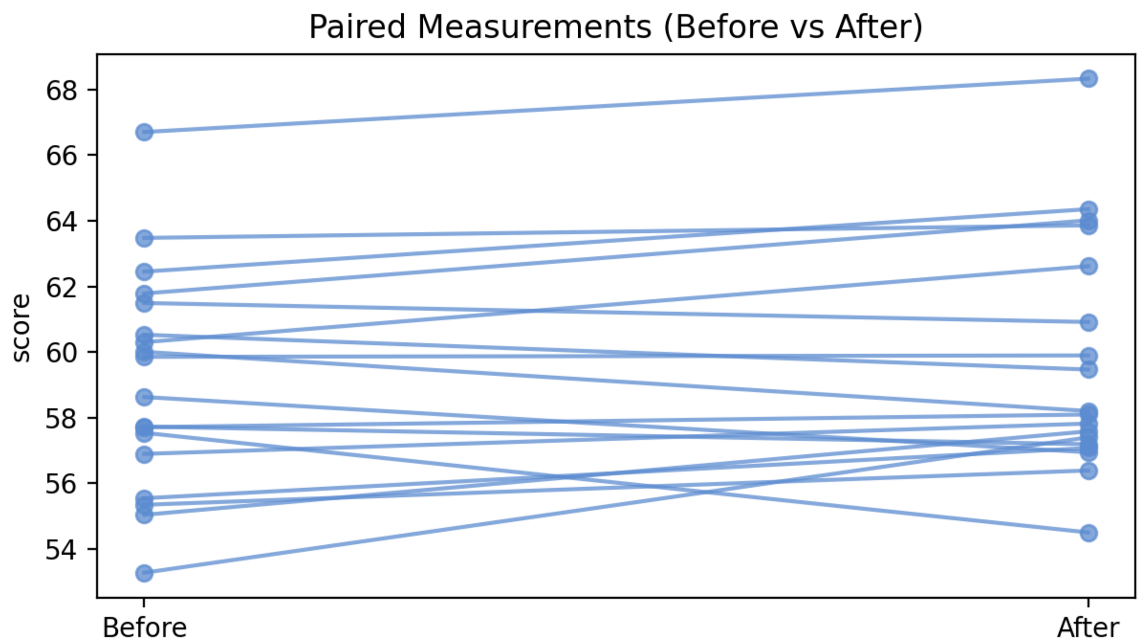
Output files:

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

What to show and say.

- Generates before/after scores for the same individuals and runs a paired t-test.
- Shows paired lines so students see why pairing reduces noise.
- Reports Cohen's d (paired) to emphasize effect size, not only p-value.

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 can paired designs be more powerful than independent designs?

Suggested answer (for revision). Pairing removes between-subject variation by comparing each person to themselves, so the test has higher power.

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 Paired Design Effect Size Exercises Demo Summary

Agenda

- Overview
- Paired Design
- Effect Size
- Exercises
- Demo
- Summary

Learning Outcomes

- Differentiate paired vs independent designs
- Compute within-pair differences di
- Run a paired t-test (conceptually)
- Explain effect size and why we report it
- Interpret results in context (not only p-value)

Paired Design: Key Points

- Same unit measured twice (before/after)
- Analyze differences di = after - before
- Pairing reduces noise from individual differences

Paired Design: Key Formula

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \quad \mathrm{df} = n - 1$$

Effect Size: Key Points

- p-value answers: evidence?
- Effect size answers: how big?
- Large n can make tiny effects significant

Effect Size: Key Formula

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}}$$

Exercise 1: Compute differences

Before/After: (10,12), (12,12), (11,14), (9,10). Compute di and dbar.

Solution 1

- di: 2,0,3,1
- dbar = 1.5

Exercise 2: CI idea

If the 95% CI for mean difference excludes 0, what does it suggest?

Solution 2

- Evidence of a change (difference likely non-zero).
- Check magnitude and context.

Exercise 3: Interpret d

If Cohen's d=0.3, what does it suggest (rule of thumb)?

Solution 3

- Small effect (context dependent).
- Still may matter if cheap/safe to adopt.

Mini Demo (Python)

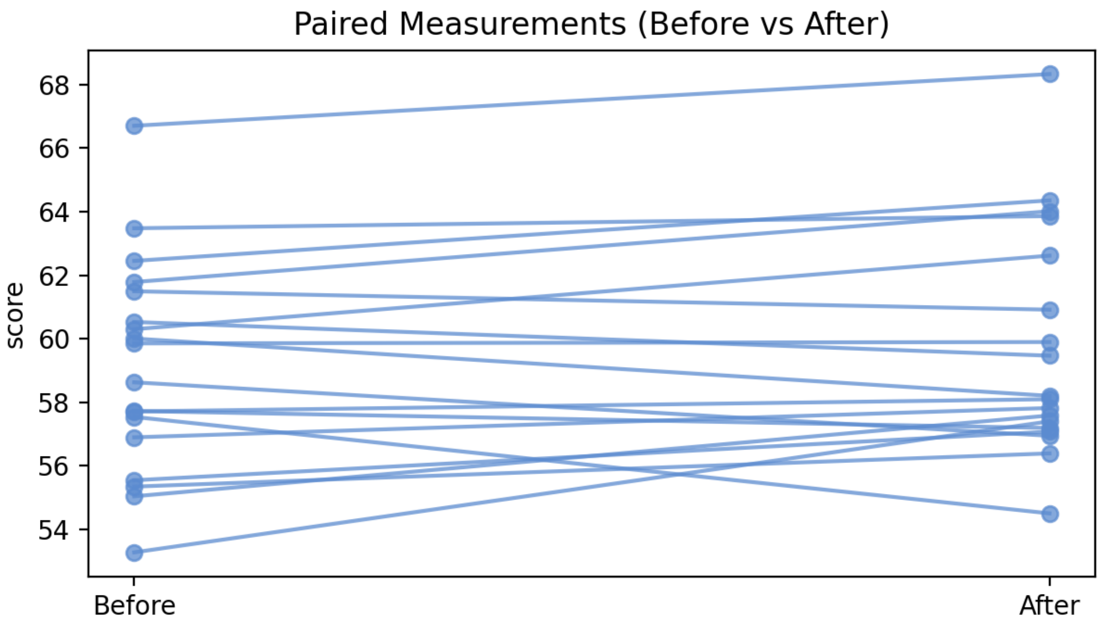
Run from the lecture folder:

```
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 can paired designs be more powerful than independent designs?