Statistics and Data Analysis
Unit 03 – Lecture 02 Notes
Hypothesis Testing (t-test): Concepts and Setup

Tofik Ali

February 17, 2026

Topic

Null vs alternative; one-sample and two-sample t-tests; p-values 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

- t-test Basics
- p-values
- Exercises
- Demo
- Summary

Learning Outcomes

- Define null and alternative hypotheses clearly
- Compute a one-sample t statistic (given summary)
- Explain p-value and significance level alpha
- Distinguish one-tailed vs two-tailed tests
- State key assumptions behind the t-test

Why these outcomes matter. **Hypothesis testing** is a decision framework. You start with a default claim (the null hypothesis H_0) and ask whether the sample evidence is strong enough to reject it in favor of an alternative H_1. The key idea is to control error rates: a test can be wrong, so we choose a significance level α and interpret results with that risk in mind. A **one-tailed** test is used when only one direction matters (only increase or only decrease). A **two-tailed** test is used when deviations in both directions matter. Choosing one-tailed after seeing the data is not valid; decide the tail based on the research question before looking at results.

t-test Basics: Key Points

- H0/H1 setup
- Test statistic measures how far the sample is from H0
- Assumptions: independence, outliers, normality/CLT

Explanation. The **null hypothesis** H_0 usually represents 'no effect' or a baseline value (e.g., $\mu = 60$). The **alternative** H_1 represents the effect you are looking for (e.g., $\mu \neq 60$ or $\mu > 60$). We compute a test statistic and a p-value assuming H_0 is true. **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). The **Central Limit Theorem (CLT)** explains why many methods work for large samples: sample means tend to become approximately normal even when the population is not perfectly normal. With small samples, you must be more careful about outliers and skewness.

t-test Basics: Key Formula

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad \mathrm{df} = n - 1$$

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).

p-values: Key Points

- p-value = probability of data (or more extreme) assuming H0
- Small p-value -> evidence against H0
- p-value is not effect size

Explanation. The **null hypothesis** H_0 usually represents 'no effect' or a baseline value (e.g., $\mu = 60$). The **alternative** H_1 represents the effect you are looking for (e.g., $\mu \neq 60$ or $\mu > 60$). We compute a test statistic and a p-value assuming H_0 is true. 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.

Exercises (with Solutions)

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

Exercise 1: Write hypotheses

Claim: mean score is 60. Write H0 and H1 for a two-sided test.

Solution

- H0: mu = 60
- H1: mu != 60

Walkthrough. **Hypothesis testing** is a decision framework. You start with a default claim (the null hypothesis H_0) and ask whether the sample evidence is strong enough to reject it in favor of an alternative H_1. The key idea is to control error rates: a test can be wrong, so we choose a significance level α and interpret results with that risk in mind. The **null hypothesis** H_0 usually represents 'no effect' or a baseline value (e.g., $\mu = 60$). The **alternative** H_1 represents the effect you are looking for (e.g., $\mu \neq 60$ or $\mu > 60$). We compute a test statistic and a p-value assuming H_0 is true.

Exercise 2: Compute t

Given n=25, xbar=53, s=10, test H0: mu=50. Compute t.

Solution

- SE = 10/sqrt(25) = 2
- t = (53-50)/2 = 1.5
- df = 24

Walkthrough. The **null hypothesis** H_0 usually represents 'no effect' or a baseline value (e.g., $\mu = 60$). The **alternative** H_1 represents the effect you are looking for (e.g., $\mu \neq 60$ or $\mu > 60$). We compute a test statistic and a p-value assuming H_0 is true. **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).

Exercise 3: Tail choice

You want to show a new method increases mean score. One-tailed or two-tailed?

Solution

- One-tailed (right): H1: mu > mu0

Walkthrough. A **one-tailed** test is used when only one direction matters (only increase or only decrease). A **two-tailed** test is used when deviations in both directions matter. Choosing one-tailed after seeing the data is not valid; decide the tail based on the research question before looking at results.

Mini Demo (Python)

Run from the lecture folder:

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

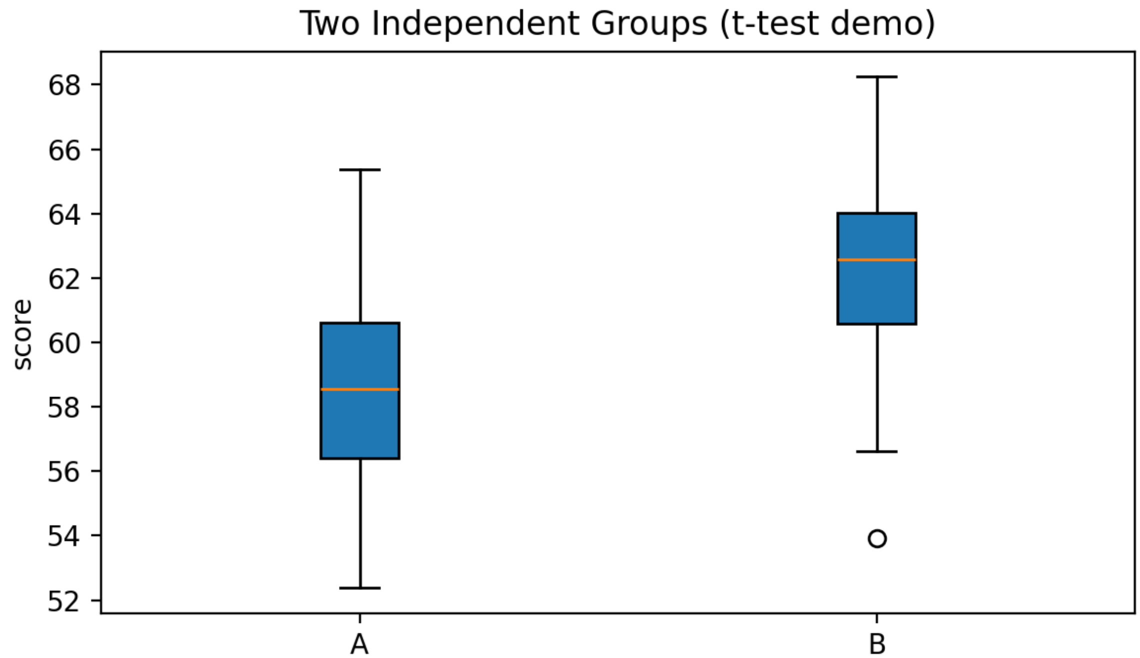
Output files:

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

What to show and say.

- Generates two independent groups and runs a Welch two-sample t-test.
- Shows a boxplot so students can connect the test to the distribution shape/outliers.
- Open data/results.txt to see the t-statistic and p-value and interpret them.

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 a very small p-value still be unimportant in practice?

Suggested answer (for revision). With large n, even tiny differences can yield very small p-values; always check effect size and practical impact.

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 t-test Basics p-values Exercises Demo Summary

Agenda

- Overview
- t-test Basics
- p-values
- Exercises
- Demo
- Summary

Learning Outcomes

- Define null and alternative hypotheses clearly
- Compute a one-sample t statistic (given summary)
- Explain p-value and significance level alpha
- Distinguish one-tailed vs two-tailed tests
- State key assumptions behind the t-test

t-test Basics: Key Points

- H0/H1 setup
- Test statistic measures how far the sample is from H0
- Assumptions: independence, outliers, normality/CLT

t-test Basics: Key Formula

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad \mathrm{df} = n - 1$$

p-values: Key Points

- p-value = probability of data (or more extreme) assuming H0
- Small p-value -> evidence against H0
- p-value is not effect size

Exercise 1: Write hypotheses

Claim: mean score is 60. Write H0 and H1 for a two-sided test.

Solution 1

- H0: mu = 60
- H1: mu != 60

Exercise 2: Compute t

Given n=25, xbar=53, s=10, test H0: mu=50. Compute t.

Solution 2

- SE = 10/sqrt(25) = 2
- t = (53-50)/2 = 1.5
- df = 24

Exercise 3: Tail choice

You want to show a new method increases mean score. One-tailed or two-tailed?

Solution 3

- One-tailed (right): H1: mu > mu0

Mini Demo (Python)

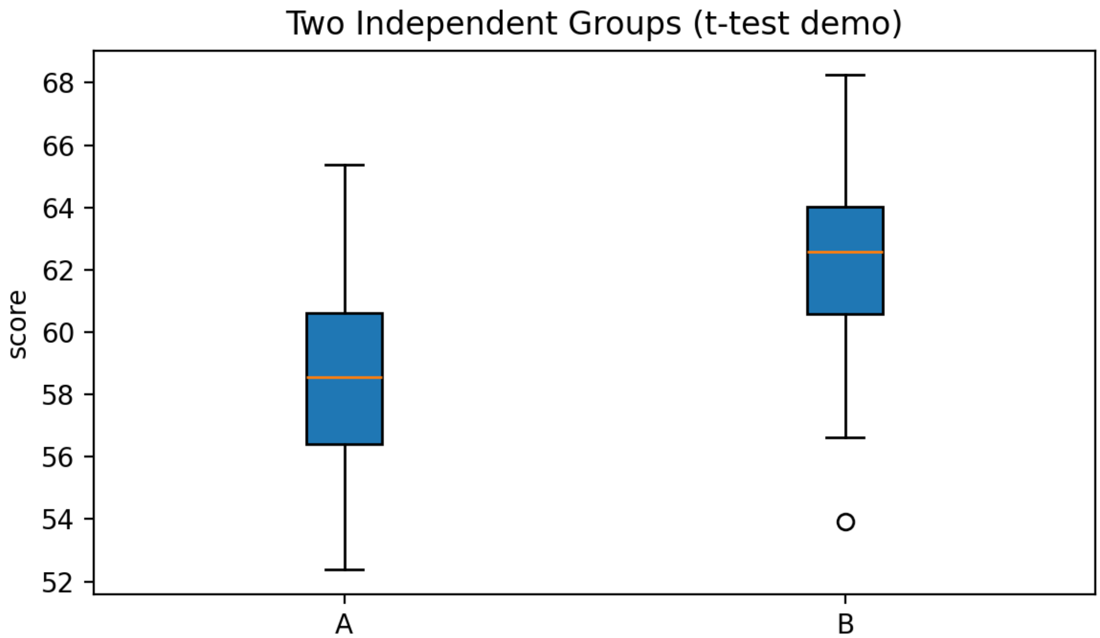
Run from the lecture folder:

```
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 a very small p-value still be unimportant in practice?