Statistics

Hypothesis Testing

Hypothesis Testing

🎯 What Is Hypothesis Testing?

Hypothesis testing is a way for scientists or researchers to **test ideas or claims** using data.
Think of it like a **trial**: you make a claim (the hypothesis), collect evidence (data), and then decide if your claim makes sense based on the evidence.

Real-Life Example

Let's say school canteen says:

> "Students eat **an average of 2 apples a day**."

You and your friends think, "That can't be right. We don't eat that many apples!"

So you decide to **test their claim** by collecting data.

Important Terms

- **Hypothesis:** A guess or claim you test.
- **Null Hypothesis (H_0):** The claim you're testing (usually that nothing has changed).
- **Alternative Hypothesis (H_1):** The opposite of the null (something **has** changed).
- **Significance level (α):** A cutoff (like 5%) to decide if the result is rare or not.

Steps in Hypothesis Testing

1. **State the Hypotheses**
    - **Null Hypothesis (H_0):** The canteen is correct. Students eat 2 apples a day.
    - **Alternative Hypothesis (H_1):** The canteen is wrong. Students eat **less** (or more) than 2 apples a day.
2. **Collect Data**
   Ask 30 students how many apples they eat per day and find the average.

3. **Analyze the Data**
   Use math (like calculating the mean and standard deviation) to check if the new average is **significantly different** from 2.

4. **Make a Decision**
    - If your results show a big enough difference, you can **reject the null hypothesis**.
    - If not, you **fail to reject it** — meaning the canteen's claim might still be true.

Type I and Type II Errors

When we make a decision in hypothesis testing, there's always a chance we could be wrong.
These wrong decisions are called **Type I and Type II errors**.

Continuing Our Apple Example

The school canteen claims:

> **"Students eat 2 apples per day."**

You test this claim using hypothesis testing.

- **Null Hypothesis (H_0):** Students eat 2 apples per day.
- **Alternative Hypothesis (H_1):** Students do **not** eat 2 apples per day.

❌ Type I Error (False Positive)

You **reject H_0** even though it's **actually true**.

> Example: You say the canteen is wrong (students *don't* eat 2 apples), but in reality, they **do** eat 2 apples per day.

❌ Type II Error (False Negative)

You **fail to reject H_0** even though it's **actually false**.

> Example: You say the canteen is correct (students eat 2 apples), but in reality, they **don't** — maybe they eat only 1 apple per day.

Two Types of Correct Decisions

1. **Fail to Reject H_0 — and H_0 is True**

- You say:
  "The canteen's claim seems correct — students eat 2 apples."

- And in reality, this is **true**.

✔ You **accepted a true claim → Correct Decision!**

---

2. **Reject H_0 — and H_0 is False**

- You say:
  "The canteen's claim is wrong — students do NOT eat 2 apples."
- And in reality, they actually **don't** (maybe only 1 apple per day).

✔ You **rejected a false claim → Correct Decision!**

Hypothesis Testing: Significance Value (α) and Confidence Interval (CI)

1. Significance Value (α)

The **significance value** (α) is the **threshold probability** we set **before hypothesis testing**.
It determines **when to reject the null hypothesis (H_0)**.

1.1 What is α?

α = 0.05 (5%), **α = 0.01 (1%)**, **α = 0.10 (10%)**

Formula:

$$\boxed{\alpha = 1 - \text{Confidence Level}}$$

Example:

- Confidence level = **95%**

$$\alpha = 1 - 0.95 = 0.05$$

P-value

P-Value

- The **p-value** tells us **how likely our sample result (or something more extreme) is, if the null hypothesis (H_0) were true**.
- **Small p-value → evidence against H_0**.
- Rule of thumb:
  - If **$p < 0.05$** → Reject H_0 (significant).
  - If **$p \geq 0.05$** → Fail to reject H_0 (not significant).

🎯 P-value: One-Tailed vs Two-Tailed Tests

1️⃣ One-Tailed Test

- Hypothesis checks only **one direction** (greater than or less than).
- Example: Test if students eat **more than 1 apple/day** (H_1: $\mu > 1$).
- The p-value is the **area in one tail** (right side for $\mu >$, left side for $\mu <$).

2️⃣ Two-Tailed Test

- Hypothesis checks for **any difference** (not equal).
- Example: Test if students eat **different from 1 apple/day** (H_1: $\mu \neq 1$).
- The p-value is the **area in both tails** (extreme low and extreme high).

✅ Rule:

- **One-tailed** → Use when you care about only one direction.
- **Two-tailed** → Default choice when testing for “difference.”

Step-by-Step Process

1. **Set up Null and Alternate Hypotheses**

    - **H_0 (Null Hypothesis)**: The statement we test (no change/no effect)
    - **H_1 (Alternative Hypothesis)**: What we want to prove (there is a change/effect)

2. **Decide the Significance Level (α)**

    - Common choices: 0.05 (5%) or 0.01 (1%)

3. **Select the Right Test**

    - Use **Z-test** if population standard deviation is known
    - Use **T-test** if population standard deviation is unknown and sample size < 30

4. **Calculate the Test Statistic (Z or T score)**

    - Formula for Z-score:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

    - Formula for T-score:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

5. **Find the p-value**

6. **Compare p-value and α**

    - If **p < α**, reject the null hypothesis
    - If **p > α**, accept the null hypothesis

Example Scenario

A machine produces **bolts** with a target diameter of **5 mm**.
We want to check if the **average diameter** has **increased**.

Given:

- Population standard deviation, $\sigma = 0.4$ mm
- Sample size, $n = 25$
- Sample mean, $\bar{x} = 5.3$ mm
- Significance level, $\alpha = 0.05$

Step 1 — Define Hypotheses

Case A: Right-Tailed Test (Check if diameter has increased)

$$H_0 : \mu = 5 \quad \text{(machine is producing correct size)}$$

$$H_1 : \mu > 5 \quad \text{(machine produces larger bolts)}$$

Step 2 — P-value Approach

Formula:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Calculation:

$$Z = \frac{5.3 - 5}{0.4/\sqrt{25}} = \frac{0.3}{0.08} = 3.75$$

Find P-value

Using **Z** = **3.75** and right-tailed test:

$p = P(Z > 3.75) \approx 0.00009$

Decision Rule:

- If $p < \alpha$ → **Reject** \\(H_0\\)
- Here, $p = 0.00009 < 0.05$ ✅

Conclusion:

> The bolts are **significantly larger** than 5 mm.

Independent Samples t-test

What it does

- Compares the **means** of two *independent groups*.
- Helps to check if the difference in means is **significant** or just due to chance.

🎓 1. **Independent t-test Example (Unpaired Groups)**

📘 Scenario:

You want to compare **marks of boys and girls** in a math test.

- Group 1 (Boys): [55, 60, 65, 70, 75]
- Group 2 (Girls): [65, 70, 72, 68, 74]

Step-by-Step:

1️⃣ **Hypotheses**:

- **H_0**: $\mu_1 = \mu_2$ (No difference in scores)
- **H_1**: $\mu_1 \neq \mu_2$ (There is a difference)

2️⃣ **Calculate sample means:**

- Mean (Boys) = 65
- Mean (Girls) = 69.8

3️⃣ **Use independent t-test formula or calculator:**

- Test statistic **t** ≈ **-2.02**
- Degrees of freedom ≈ 8
- Critical t-value ($\alpha = 0.05$, two-tailed) ≈ ±2.306

4️⃣ **Decision:**

- $|t| = 2.02 < 2.306$ → ❌ **Fail to reject H_0**

✅ **Conclusion:** No significant difference between boys' and girls' scores.

📗 2. Dependent t-test Example (Paired Groups)

📘 Scenario:

You want to test if a **coaching class improved scores**. You take scores of **same students** before and after the class.

- **Before:** [50, 55, 52, 60, 58]
- **After:** [55, 60, 58, 65, 64]

Step-by-Step:

1️⃣ Hypotheses:

- **H_0:** μ_d = 0 (No improvement)
- **H_1:** μ_d > 0 (Improved after class)

2️⃣ Find differences:

Student	After	Before	Difference (d)
1	55	50	5
2	60	55	5
3	58	52	6
4	65	60	5
5	64	58	6

- Mean of d = 5.4
- SD of d ≈ 0.55
- n = 5

3️⃣ t-statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{5.4}{0.55/\sqrt{5}} \approx 21.97$$

- Critical t (α = 0.05, df = 4, one-tailed) ≈ 2.132

4️⃣ Decision:

- 21.97 > 2.132 → ✅ **Reject H_0**

✅ **Conclusion:** The class **significantly improved** scores.

ANOVA (Analysis of Variance)

Full Form

ANOVA = **Analysis of Variance**

Introduction

- A statistical method to compare the **means of 3 or more groups**.
- Checks whether differences in sample means are **statistically significant** or just due to random variation.

📊 One-Way ANOVA Example (Manual Calculation)

🎯 Objective:

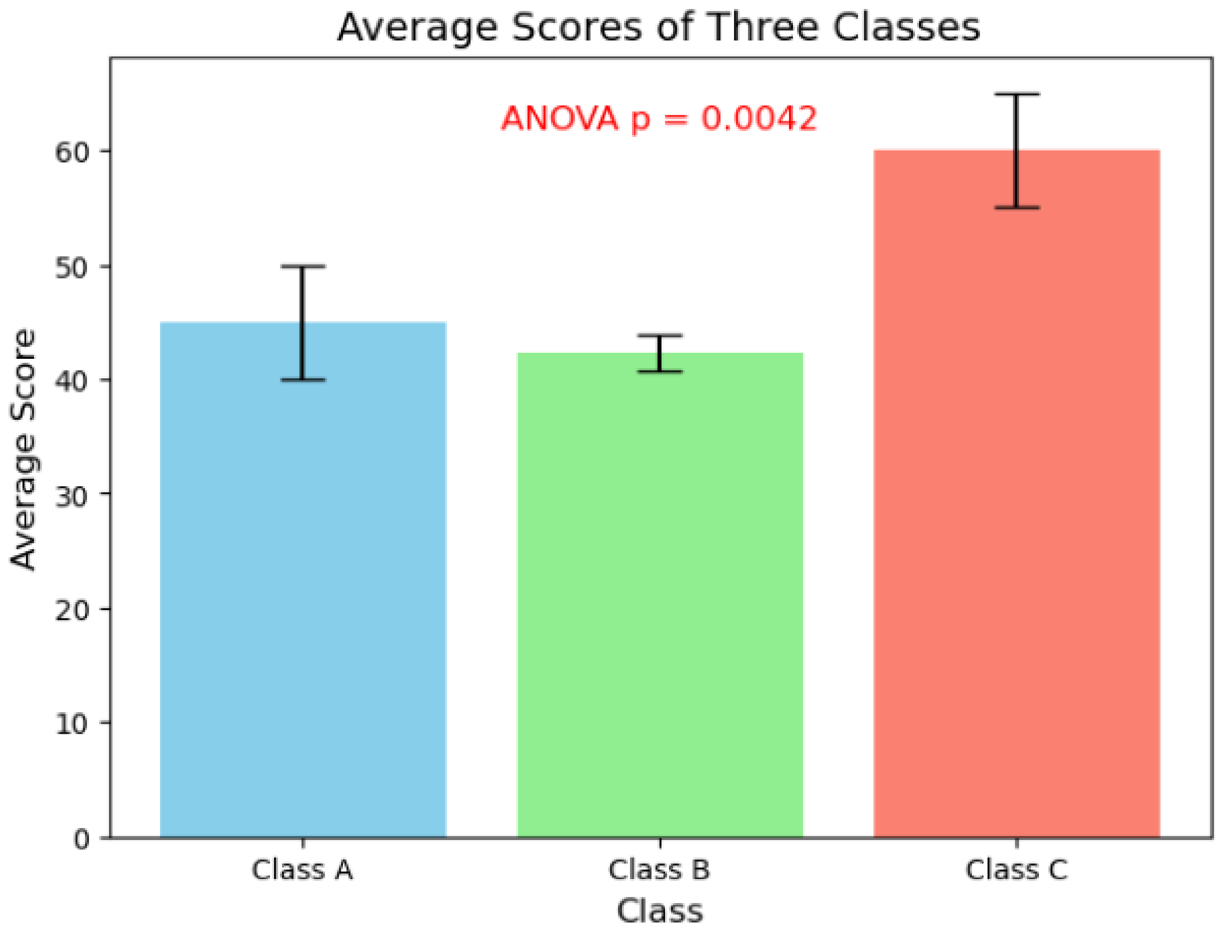
Test if there is a significant difference in average scores among **three different classes** of students.

🧩 Data Table:

Class A	Class B	Class C
40	42	55
45	41	60
50	44	65

- **k** = **3** groups
- **n** = **3** observations per group
- **N** = **9** total observations

```
F-statistic = 15.60, p-value = 0.0042
```



ince p-value = 0.0009 < 0.05, we reject H_0.

That means: Not all class averages are equal. At least one class differs significantly.

🧮 Chi-Square Test (χ^2 Test)

📘 What is a Chi-Square Test?

The **Chi-Square (χ^2) Test** is a statistical method used to:

- Compare **observed** values with **expected** values.
- Check if there is a **significant association** between two categorical variables.

🧪 Example: Chi-Square Test of Independence

🎯 Situation:

We want to know whether **gender** and **preference for a subject** are related.

Step 1 — Hypotheses

- H_0: Gender and subject choice are **independent** (no relation).
- H_1: Gender and subject choice are **not independent** (there is a relation)

📊 Data (Observed):

	Science	Arts	Total
Boys	30	10	40
Girls	10	30	40
Total	40	40	80

🧮 Step 1: Find Expected Values

$$E_{ij} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

Example:

Expected value for Boys–Science:

$$E = \frac{40 \times 40}{80} = 20$$

	Science (E)	Arts (E)
Boys	20	20
Girls	20	20

🧮 Step 2: Apply Chi-Square Formula

Formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Calculations:

- For Boys–Science: (\frac{(30 - 20)^2}{20} = 5)
- For Boys–Arts: (\frac{(10 - 20)^2}{20} = 5)
- For Girls–Science: (\frac{(10 - 20)^2}{20} = 5)
- For Girls–Arts: (\frac{(30 - 20)^2}{20} = 5)

Total χ^2 = 5 + 5 + 5 + 5 = 20

📏 Step 3: Degrees of Freedom

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$$

✅ Step 4: Compare with χ^2 Table

- At **α** = **0.05** and **df** = **1**, critical value ≈ **3.84**
- Our χ^2 = **20**, which is **greater** than 3.84

✅ **Conclusion**: Reject H_0

There is a significant relationship between gender and subject preference.