

HYPOTHESIS TESTING

By Team 3 and Team 4



HYPOTHESIS AND HYPOTHESIS TESTING

What is a Hypothesis?

A hypothesis is a tentative assumption, guess, or prediction about an outcome. Hypothesis is always tested with data.

Example: Average marks of students in our class is 60.

What is Hypothesis Testing?

Hypothesis testing is a method to decide whether the claim (hypothesis) about a population is true or not, using sample data.

Example: Is the average weight of students in a class equal to 60 kg?

Hypothesis Testing



Formulating Hypothesis

First things first, we need to come up with our main suspects, or in other words, our hypotheses.



Determining Confidence and Significance Levels

Next, we need to decide how confident we need to be before we can make our final decision.



Selecting a Statistical Test

Now, it's time to choose our detective tool-our statistical test. Different tests are used for different kinds of data.



Applying the Statistical Test

Next, we run our evidence-our data-through our chosen tool. This is like putting the clues together to see what they reveal.



Analyzing the Outputs

Finally, we look at our results.

1. Why We Use Hypothesis Testing

- Testing the entire population is not practical (too costly/time-consuming).
- Helps us make **scientific decisions** with limited data.
- Avoids guesswork → relies on probability and statistics.

2. Where We Use (Applications)

- Medicine: Testing if a new drug lowers blood pressure.
- Education: Checking if a new teaching method improves exam scores.
- Business: Comparing sales before and after a discount scheme.

3. Purpose

- To verify claims/assumptions about populations using data.
- To decide whether observed differences are real or just due to chance.
- Provides a **systematic framework** for making evidence-based decisions.

4. Key Idea (Analogy)

- Think of it as a courtroom trial:
- Null Hypothesis (H_0): "The accused is innocent."
- Alternate Hypothesis (H_1): "The accused is guilty."
- Evidence (data): Examined by the jury.
- Decision: Only reject H_0 if there is strong evidence.

BASICS OF HYPOTHESIS TESTING

Null Hypothesis(H_0):

- It assumes there is no effect, no difference, or no change.
- It acts as the *default or baseline* assumption.
- Example: "The average marks of the class = 60."



IN SIMPLE WORDS:

H_0 : NOTHING SPECIAL IS HAPPENING.

BASICS OF HYPOTHESIS TESTING

Alternative Hypothesis(H_1 or H_a):

- It assumes there is an effect, difference, or change.
- This is what we want to prove with data.
- Example: "The average marks of the class \neq 60."



IN SIMPLE WORDS:

H_1 : SOMETHING IS HAPPENING.

SELECTING A SIGNIFICANCE LEVEL

Definition:

The significance level (α) is the probability of making a Type I Error (rejecting H_0 when it is actually true)

Common Choices:

- $\alpha = 0.05 \rightarrow$ 5% chance of error (most used)
- $\alpha = 0.01 \rightarrow$ very strict (1% chance of error)
- $\alpha = 0.10 \rightarrow$ more relaxed (10% chance of error)

Procedure of Hypothesis Testing

1.State Hypotheses (H_0 and H_1).

Example: H_0 = “Average marks = 60”, H_1 = “Average marks \neq 60”.

2.Set Significance Level (α).

Common choice: $\alpha = 0.05$ (5% chance of error).

3.Collect Sample Data.

Example: Take 30 students' scores.

4.Calculate Test Statistic & p-value.

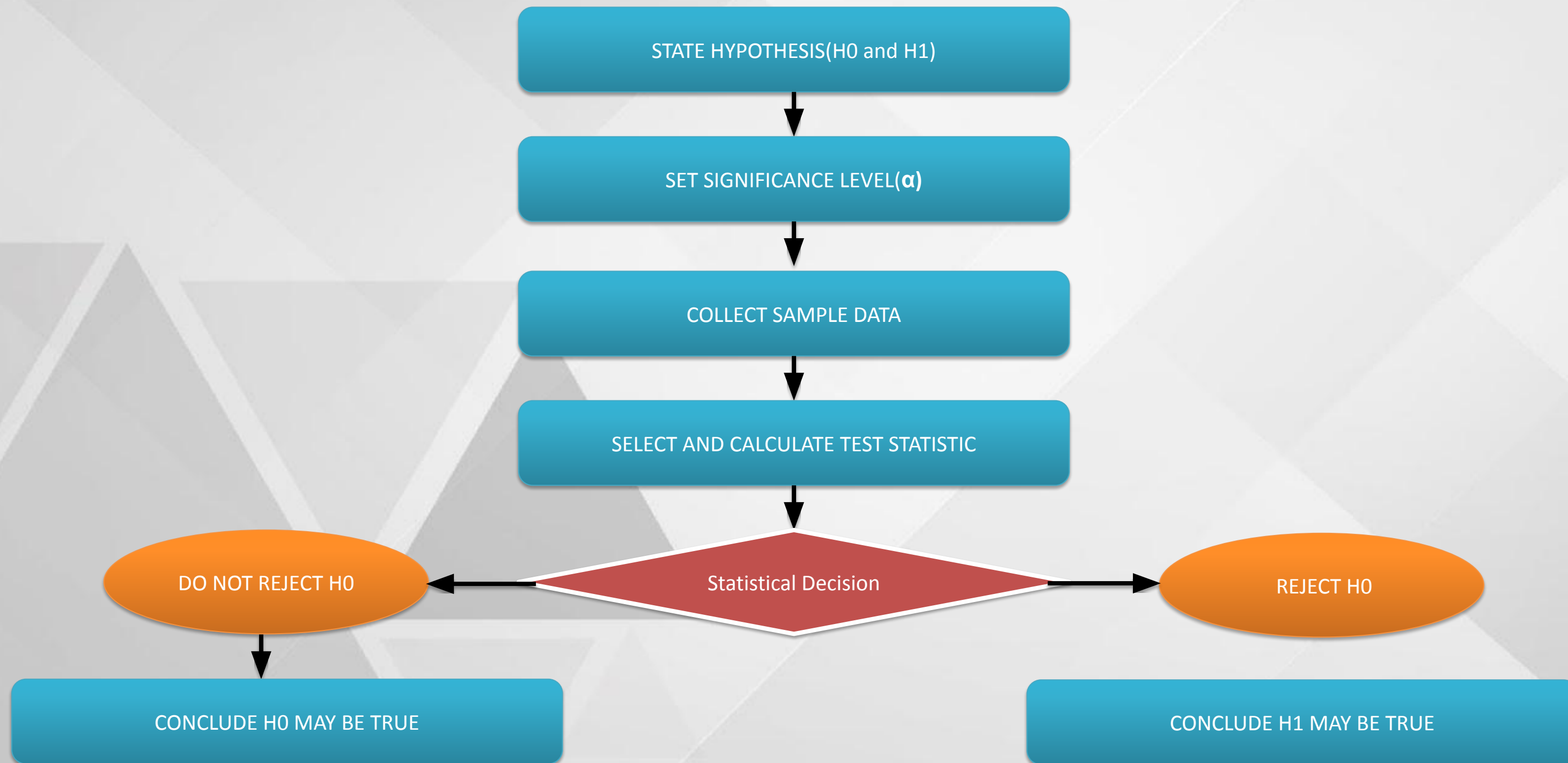
p-value = Probability of getting results as extreme as observed if H_0 is true.

5.Make Decision:

If $p \leq \alpha \rightarrow$ Reject H_0 (enough evidence).

If $p > \alpha \rightarrow$ Fail to reject H_0 (not enough evidence).

Procedure of Hypothesis Testing



SELECTING A SIGNIFICANCE LEVEL

Interpretation:

- Smaller $\alpha \rightarrow$ more cautious (less chance of false alarm, but higher chance of missing real effect).
- Larger $\alpha \rightarrow$ easier to reject H_0 (but higher risk of false alarm).

Example:

If $\alpha = 0.05$, we accept a **5% risk** of wrongly rejecting H_0

ERRORS IN HYPOTHESIS TESTING

When we make a decision using sample data, two types of mistakes are possible:

Type I Error (α -error):

- Rejecting H_0 when H_0 is actually true.
 - False Alarm.
 - Example: Saying "average marks $\neq 60$ " when actually it is 60.
- Probability of making this error = **Significance Level (α)**, usually

ERRORS IN HYPOTHESIS TESTING

Type II Error (β):

- Failing to reject H_0 when H_1 is actually true.
- Missed Detection.
- Example: Saying “average marks = 60” when actually it is not 60.



There's a trade-off: If we decrease α (become strict), the chance of β (missing real effect) increases.

ERRORS IN HYPOTHESIS TESTING

	H_0 True	H_0 False
Accept H_0 /Reject H_1	True $P(\text{Type I}) = \beta$	False: Type II error Power = $1 - \beta$ (sensitivity)
Reject H_0 /Accept H_1	False: Type I error Confidence level = $1 - \alpha$ (specificity)	True $P(\text{Type II}) = \beta$

TYPE I AND TYPE II ERRORS – EXAMPLE

Your null hypothesis is that the battery for a heart pacemaker has an average life of 300 days, with the alternative hypothesis that the average life is more than 300 days. You are the quality control manager for the battery manufacturer.

(a) Would you rather make a Type I error or a Type II error?

(b) Based on your answer to part (a), should you use a high or low significance level?

TYPE I AND TYPE II ERRORS – EXAMPLE

Hypotheses

- H_0 : Average life = 300days
- H_A : Average life > 300days

(a) Which error would you rather make?

It is better to make a **Type II error**.

- **Type I error**: Reject H_0 when it is true → incorrectly conclude battery life > 300days. This overstates battery life and risks patient safety and regulatory problems.
- **Type II error**: Fail to reject H_0 when H_0 is false → miss a true improvement (the battery actually lasts >300 days). This is conservative (missed opportunity), but safer than overstating life.

Therefore, being the quality control manager, you prefer the safer mistake of **Type II** (understating an Improvement) rather than risking a harmful **Type I** (overstating life).

TYPE I AND TYPE II ERRORS – EXAMPLE

Hypotheses

- H_0 : Average life = 300days
- H_A : Average life > 300days

(b) High or low significance level?

Choose a **low significance level** (α).

- Increasing α raises the chance of a Type I error.
- Since we want to avoid Type I errors here, we set α **low** (e.g., 0.01) even though that increases the chance of a Type II error. That trade-off is acceptable because patient safety and regulatory correctness are priority.

POWER OF A TEST

Definition: The probability of correctly rejecting H_0 when H_1 is actually true.

Formula: $\text{Power} = 1 - \beta$.

Interpretation: Higher power means our test is better at detecting real effects.

Example: If a new medicine works, a powerful test will almost always detect the improvement.

Ways to increase Power:

- Increase sample size.
- Use a larger significance level (α).
- Reduce variability in data.

MINI EXAMPLE OF HYPOTHESIS TESTING

Problem Statement

A teacher claims that the **average marks of students = 60**. We want to test if this claim is true based on a **sample of 10 students**.

- ◆ Step 1: State the Hypotheses

Null Hypothesis (H_0): $\mu = 60$ (average marks are 60).

Alternate Hypothesis (H_1): $\mu \neq 60$ (average marks are not 60).

MINI EXAMPLE OF HYPOTHESIS TESTING

- ◆ **Step 2: Set Significance Level**

Choose $\alpha = 0.05$ (5% risk of being wrong).

- ◆ **Step 3: Collect Sample Data**

Marks of 10 students: 62, 65, 58, 63, 67, 64, 61, 59, 66, 60

Sample Mean (\bar{x}): 62.5

Sample Size (n): 10

Population Mean (μ): 60

MINI EXAMPLE OF HYPOTHESIS TESTING

Step 3a: Calculate Sample Mean (\bar{x})

$$\bar{x} = \frac{\text{Sum of all marks}}{n}$$

Sum = 62 + 65 + 58 + 63 + 67 + 64 + 61 + 59 + 66 + 60 = **625**
n = 10

$$\bar{x} = \frac{625}{10} = 62.5$$

Step 3b: Calculate Sample Standard Deviation (s)

Formula:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Differences from mean: (62-62.5), (65-62.5), (58-62.5), ...
Squared differences sum = **82.5**

$$s = \sqrt{\frac{82.5}{9}} \approx 3.03$$

MINI EXAMPLE OF HYPOTHESIS TESTING

◆ Step 4: Perform Test (t-test calculation)

Formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where:

$$\mu = 60$$

$$s = 3.03$$

$$n = 10$$

$$\bar{x} = 62.5$$

$$t = \frac{62.5 - 60}{3.03/\sqrt{10}} = \frac{2.5}{0.958} \approx 2.61$$

Step 4b: Find p-value

Degrees of freedom (df) = $n - 1 = 9$

From t-distribution table, p-value \approx **0.03**

MINI EXAMPLE OF HYPOTHESIS TESTING

- ◆ **Step 5: Make Decision**

Since **p-value (0.03) < α (0.05) → Reject H_0**

Conclusion

There is **enough evidence** to say that the average marks are **not equal to 60**.

The teacher's claim is **not supported** by the data.

T-Value

T-test is fundamentally a tool for hypothesis testing. We can think of it as a formal method to answer the question: **"Is the difference I'm observing between two groups meaningful, or is it just due to random chance?"**

It's used when we have a limited sample of data and want to infer something about the larger "population".

- Analogy: **Signal-to-Noise Ratio**
 - Signal = Difference between group means.
 - Noise = Variability within groups.
 - $t = \text{Signal} / \text{Noise}$

Assumptions

1. The data collected must follow a continuous or ordinal scale, such as the scores for an IQ test.
2. The data is collected from a randomly selected portion of the total population(unbiasedness)
3. The data will result in a normal distribution of a bell-shaped curve.
4. Equal or homogenous variance exists when the standard variations are equal.

Null hypothesis: The differences are statistically insignificant.

Types of T-Test

1. Independent Samples: Compare two separate groups (Class A vs. Class B scores).
2. Paired Samples: Compare the same group at two times (Before/After scores)
3. One-Sample: Compare one group's mean to a known value (Is average bolt length 10 cm?)

Eg. 1: One-Sample t-test

A coffee company claims its “large” cup contains 300 ml of coffee on average.

- Null Hypothesis (H_0): $\mu = 300$ ml
- Alternative Hypothesis (H_1): $\mu \neq 300$ ml
- Sample: 10 randomly chosen cups \rightarrow mean = 295 ml, standard deviation = 4 ml

Calculate t-value:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{295 - 300}{4/\sqrt{10}} = \frac{-5}{1.26} \approx -3.97$$

- Interpretation: The sample mean is almost 4 standard errors below the claimed 300 ml.
- This difference is large relative to natural variation.
- Conclusion: The cups likely contain less than advertised. (Null hypothesis rejected.)

Eg.2: Independent Two-Sample t-test

A school wants to compare math scores between two teaching methods.

- $H_0: \mu_1 = \mu_2$ (no difference)
- Group A (new method): 20 students, mean = 82, std. dev. = 5
- Group B (traditional): 22 students, mean = 78, std. dev. = 6

Calculate std. Pooled error(SE):

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{25}{20} + \frac{36}{22}} \approx \sqrt{1.25 + 1.64} \approx 1.68$$

Calculate t-value:

$$t = \frac{82 - 78}{1.68} \approx 2.38$$

- The gap of 4 points is 2.4 times larger than expected random fluctuation.
- Conclusion: The new method appears to improve scores.(Null hypothesis rejected)

Eg.3: Paired t-test

A gym introduces a 6-week strength program. They test members before and after.

- H_0 : mean improvement = 0
- Data (15 members):
- Average improvement in squat = +12 kg
- Standard deviation of improvements = 8 kg

Calculate Std. Error:

$$SE = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{15}} \approx 2.07$$

Calculate t-value:

$$t = \frac{12}{2.07} \approx 5.8$$

- The improvement is almost 6 times larger than expected noise.
- Conclusion: The training program clearly increases strength. (Again, Null hypothesis is rejected)

P-VALUE

The p-value tells us how likely it is to see the results we got if the null hypothesis is true.

It helps determine the strength of evidence against the null hypothesis

Interpretation:

- Small p-value (< 0.05): Strong evidence against null \rightarrow reject null hypothesis.
- Large p-value (≥ 0.05): Weak evidence against null \rightarrow fail to reject null hypothesis.

0.05 = significance threshold (α)

- If we decrease this threshold , it means , we need a stronger evidence to reject the null hypothesis

EXAMPLE 1

(COIN TOSS TEST – FAIR COIN?)

- Null Hypothesis (H_0): The coin is fair ($p = 0.5$).
- Experiment: Toss a coin 10 times, get 9 heads.
 - p-value Calculation:
- Probability of getting 9 or more heads in 10 tosses if $p = 0.5$.
 - $p = P(X \geq 9) = C(10,9)(0.5)^{10} + C(10,10)(0.5)^{10}$
 - $p = (10 + 1) / 1024$
 - $p \approx 0.0107$

INTERPRETING : EXAMPLE 1

- Calculated p-value = 0.0107
- Since $0.0107 < 0.05 \rightarrow$ we reject H_0
- Conclusion: The coin is likely biased
- Key takeaway: A small p-value means the observed result is very unlikely if the null hypothesis were true

USING P VALUE – EXAMPLE 2

A medicine manufacturer claims the new drug has an average recovery time = 8 days.

- Null hypothesis (H_0): $\mu = 8$
- Alternative hypothesis (H_1): $\mu \neq 8$
- Sample: 30 patients → sample mean = 7.2 days, std. dev. = 1.5
- Significance level: $\alpha = 0.05$

USING P VALUE – EXAMPLE 2

Step 1 — Standard Error (SE):

$$SE = \frac{s}{\sqrt{n}} = \frac{1.5}{\sqrt{30}} \approx 0.274$$

Step 2 — t-statistic:

$$t = \frac{\bar{x} - \mu}{SE} = \frac{7.2 - 8}{0.274} \approx -2.92$$

Step 3 — p-value (df = 29):

$$p \approx 0.007$$

Decision: $p = 0.007 < 0.05 \rightarrow \text{Reject } H_0$.

Interpretation: Evidence shows the true mean \neq 8 days.

ANOVA

- Analysis of Variance (ANOVA), is a technique that determines whether the averages of three or more independent groups differ significantly from one another.
- Unlike a t-test, which only compares two groups, ANOVA can handle multiple groups in a single analysis.

It does this by comparing two types of variation: (F-statistics)

1. Differences BETWEEN groups (how much group averages differ from each other)
2. Differences WITHIN groups (how much individuals in the same group vary naturally).

If the between-group differences are significantly larger than within-group variation, ANOVA tells us: At least one group is truly different. Otherwise, it concludes: The differences are likely due to random chance.

Key Concepts and Formulas in ANOVA

- Sum of Squares (SS): This measures the overall variability in the dataset.
 - SS Total: Total variability across all observations.
 - SS Between (SSB): Variability due to the differences between group means.
 - SS Within (SSW): Variability within each group, showing how scores differ within individual groups.
- Mean Square (MS): The average of squared deviations, calculated for both between-group and within-group variability.

- MS Between (MSB): $MSB = \frac{SS \text{ Between}}{df \text{ Between}}$

- MS Within (MSW): $MSW = \frac{SS \text{ Within}}{df \text{ Within}}$

- Degrees of Freedom (df): The number of values that are free to vary when calculating statistics.

- df Between: $df \text{ Between} = k - 1$, where k is the number of groups.

- df Within: $df \text{ Within} = N - k$, where N is the total number of observations.

- F-Ratio: The ratio of MSB to MSW, used to test the null hypothesis. $F = MSB/MSW$
- p-Value: This probability value helps determine if the F-ratio is significant. A small p-value (e.g., <0.05) suggests significant differences between groups.

Assumptions of ANOVA

- Normality: The data within each group should be normally distributed.
- Homogeneity of Variance: The variance among the groups should be approximately equal.
- Independence: Observations must be independent of each other.
- Randomness: The sample should be randomly selected.

Types of ANOVA

1. One-Way ANOVA

- Compares means of 3 or more groups based on one independent variable (factor).
- Example: Comparing exam scores across three teaching methods.

2. Two-Way ANOVA

- Compares means with two independent variables (factors) and test for interaction effect (whether factors influence each other).
- Example: Effect of diet and exercise on weight loss.

3. Repeated Measures ANOVA

- Used when the same subjects are measured multiple times under different conditions.
- Example: Testing memory scores of the same group at 3 different time intervals.

4. MANOVA (Multivariate ANOVA) (advanced)

Advantages of ANOVA

- Can compare 3 or more groups at once (better than multiple t-tests).
- Helps identify if significant differences exist between group means.
- Reduces Type I Error: By testing all groups together, it minimizes the likelihood of finding false significance.
- Flexibility: Handles different data types and experimental designs.

Limitations of ANOVA

- Only tells if a difference exists, not which groups differ (needs Post Hoc tests).
- Assumes normality and equal variances – violations reduce accuracy.
- Sensitive to outliers, which can distort results.
- Requires relatively balanced sample sizes for reliability.

Steps for Conducting one-way ANOVA

Compare plant growth under 3 fertilizers (A, B, C):

- Fertilizer A: [10, 11, 12]
- Fertilizer B: [7, 8, 9]
- Fertilizer C: [4, 5, 6]

1. State the Hypotheses

- Null Hypothesis (H_0): All group means are equal, indicating no significant difference.
- Alternative Hypothesis (H_1): At least one group mean differs significantly from the others.

- Null Hypothesis (H_0): $\mu_A = \mu_B = \mu_C$
- Alternative Hypothesis (H_a): At least one μ differs.
 - Group Means: $\bar{X}_A, \bar{X}_B, \text{ and } \bar{X}_C$
 - Grand Mean: \bar{X}_{grand}

2.

$$\bar{X}_A = \frac{10+11+12}{3} = 11$$

$$\bar{X}_B = \frac{7+8+9}{3} = 8$$

$$\bar{X}_C = \frac{4+5+6}{3} = 5$$

$$\bar{X}_{\text{grand}} = \frac{10+11+12+7+8+9+4+5+6}{9} = \frac{72}{9} = 8$$

Contd.

3. Compute Sum of Squares (SS):

SSB (Sum of Squares Between Groups): Accounts for variation due to the treatment or independent variable.

$$SSB = \sum n_i(\bar{X}_i - \bar{X}_{\text{grand}})^2$$

SSE (Sum of Squares Error or Within Groups): Accounts for variation within groups (random error or residuals).

$$SSE = \sum (x_i - \bar{X})^2$$

SST (Total Sum of Squares): Accounts for total variation from overall mean.

$$SST = SSB + SSW$$

$$SSB = 3(11 - 8)^2 + 3(8 - 8)^2 + 3(5 - 8)^2 = 3(9) + 3(0) + 3(9) = 54$$

SSE:

- Fertilizer A: $(10 - 11)^2 + (11 - 11)^2 + (12 - 11)^2 = 1 + 0 + 1 = 2$
- Fertilizer B: $(7 - 8)^2 + (8 - 8)^2 + (9 - 8)^2 = 1 + 0 + 1 = 2$
- Fertilizer C: $(4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 = 1 + 0 + 1 = 2$

$$SSW = 2 + 2 + 2 = 6$$

$$SST = 54 + 6 = 60$$

4. Calculate Degrees of Freedom (df):

df1 (Between Groups) = $k - 1$, where k is number of groups.

df2 (Within Groups) = $N - k$, where N is the total observations.

$$df3 (Total) = N - 1$$

Contd.

5. Calculate Mean Squares (MS):

- MSB (Mean Square Between Groups) = $SSB / df1 = 54/2 = 27$
- MSE (Mean Square Error) = $SSE / df2 = 6/6 = 1$

6. F-statistic:

- The F-statistic is calculated as the ratio of MSB to MSE:
- $F = MSB / MSE$
 - $F = 27/1 = 27$

7. P-value:

The p-value is used to decide whether differences among groups are statistically significant. When the p-value is smaller than the significance level (α), the null hypothesis is rejected.

- If $F > F_{critical} \rightarrow p < 0.05$: Null Hypothesis Rejected
- Use the F-distribution table or software with: Numerator $df1 = 2$, Denominator $df2 = 6$, $\alpha = 0.05$
- Critical F-value, $F_{critical}$: 5.14 (From F-distribution table)
- $F > F_{critical} : 27 > 5.14 \rightarrow p < 0.05$; Reject null hypothesis



**DOES ANYONE
have questions?**

**THANK
YOU**

TEAM-3

2201AI20

2201AI21

2201AI24

2201AI25

2201AI27

2201AI34

2201CS22

2201CS27

2201CS28

2201CS43

2201CS44

2201CS62

TEAM-4

2201AI48

2201CS01

2201CS19

2201CS25

2201CS30

2201CS31

2201CS35

2201CS59

2201CS73

2201CS86

2101CS86

REFERENCES

1. <https://imai.fas.harvard.edu/teaching/files/tests.pdf>
2. <https://academicweb.nd.edu/~rwilliam/stats1/x24.pdf>
3. <https://www.geeksforgeeks.org/data-science/analysis-of-variance-anova/>
4. <https://researchmethod.net/anova/>