📘 Variance Explained for Beginners

What is Variance?

Variance is a measure of how much the data points **differ from the average (mean)** value.

> It shows how **spread out** the values are in a dataset.

Example Student Marks

Let's compare marks of two classes:

Class A (Low Variance)

Student	Marks
A1	69
A2	70
A3	71

- Mean = (69 + 70 + 71) / 3 = **70**
- All marks are **close to the mean**
- **Low Variance**

Class B (High Variance)

Student	Marks
B1	50
B2	70
B3	90

- Mean = (50 + 70 + 90) / 3 = **70**
- Marks are **spread far from the mean**
- **High Variance**

🧮 Variance Formula (Simplified)

$$ \text{Variance} = \frac{1}{N} \sum (x_i - \mu)^2 $$

- x_i : each data value
- μ: mean
- N : number of data points

📌 It's the **average of the squared differences** from the mean.

📘 What is Standard Deviation?

Standard Deviation (SD) is the **square root of variance**.

> It tells you **how much the data varies from the mean** in the **original units** (like marks, rupees, etc.).

$$ \text{Standard Deviation} = \sqrt{\text{Variance}} $$

Example: score={10,20,30,40,50}

🔷 Step 1: Calculate the Mean (μ)

$$ \text{Mean} = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30 $$

✅ So, the average score is **30**.

🔷 Step 2: Find Deviations from the Mean

Score (X)	Deviation (X − μ)	Squared Deviation ((X - μ)^2)
10	10 − 30 = -20	$(-20)^2 = 400$
20	20 − 30 = -10	$(-10)^2 = 100$
30	30 − 30 = 0	$0^2 = 0$
40	40 − 30 = 10	$10^2 = 100$
50	50 − 30 = 20	$20^2 = 400$

🔷 Step 3: Calculate Variance (σ²)

$$ \text{Variance} = \frac{\sum (X - \mu)^2}{N} = \frac{400 + 100 + 0 + 100 + 400}{5} = \frac{1000}{5} = 200 $$

📌 Variance tells us how **spread out** the data is (in squared units).

🔹 Step 4: Calculate Standard Deviation (σ)

$$ \text{Standard Deviation} = \sqrt{\text{Variance}} = \sqrt{200} \approx 14.14 $$
📌 This means, on average, each score is about **14.14 points away** from the mean.

- **Mean** is the average score: 30
- **Variance** shows the average of the squared differences from the mean: 200
- **Standard Deviation** is the square root of variance: ≈ 14.14
- A **low standard deviation** means scores are tightly clustered around the mean.
- A **high standard deviation** means the scores are more spread out.

📊 Step-by-Step Quartile Calculation

Dataset:[12, 18, 25, 26, 30, 34, 40, 45, 50]

Number of elements, **n = 9**

$$ position = (n - 1) * q $$
Where:

- `n` is the number of elements
- `q` is the desired quantile (e.g., 0.25, 0.5, 0.75)

Step-by-Step Example

Q1 (25th percentile):

- $ Position = (9 - 1) * 0.25 = 2.0 \rightarrow index 2 $
- Value at index 2 = **25.0**

Q2 (50th percentile / Median):

- $ Position = (9 - 1) * 0.50 = 4.0 \rightarrow index 4 $
- Value at index 4 = **30.0**

Q3 (75th percentile):

- $ Position = (9 - 1) * 0.75 = 6.0 \rightarrow index 6 $

- Value at index 6 = **40.0**

IQR (Interquartile Range)

$ IQR = Q3 - Q1 = 40.0 - 25.0 = 15.0 $

$ Lower Bound = Q1 - 1.5 \times IQR = 25.0 - 1.5 \times 15.0 = 25.0 - 22.5 = **2.5** $

$ Upper Bound = Q3 + 1.5 \times IQR = 40.0 + 1.5 \times 15.0 = 40.0 + 22.5 = **62.5** $

Metric	Value
Q1	25.0
Q2 (Median)	30.0
Q3	40.0
IQR	15.0
Lower Bound	2.5
Upper Bound	62.5

Any data point outside **[2.5, 62.5]** would be considered an **outlier.**

🎓 Z-Score and Probability Examples (Markdown)

We'll explore how to calculate:

- Z-scores
- Probabilities from Z-scores
- Number of values above or below a certain score in a dataset

---

📘 Dataset Used:

Student	Score
A	60
B	70
C	80
D	90
E	100

- Mean (μ) = 80
- Standard Deviation (σ) ≈ 14.14

- Total Data Points = 5

Example 1: How many values are **below 80**?

Step 1: Z-score for X = 80

$Z = \frac{X - \mu}{\sigma} = \frac{80 - 80}{14.14} = 0$

📊 Step 2: Z-table Probability

$P(Z= < 0) = 0.5$

👉 This means **50%** of values are less than 80.

🧮 Step 3: Apply to Dataset

$0.5 \times 5 = 2.5 \approx 2~\text{or}~3~values$

✅ Manual Check:

Score	< 80?
60	✅
70	✅
80	❌ (equal)
90	❌
100	❌

✔️ Matches: 2 values below 80.

✅ Example 2: How many values are **greater than 90**?

🔢 Step 1: Z-score for X = 90

$Z = \frac{90 - 80}{14.14} \approx 0.71$

📊 Step 2: Z-table Probability

$P(Z < 0.71) \approx 0.7611$

$P(Z > 0.71) = 1 - 0.7611 = 0.2389$

So **23.89%** of values are greater than 90.

Step 3: Apply to Dataset

$$ 0.2389 \times 5 = 1.1945 \approx 1~\text{value} $$

✅ Manual Check:

Score	> 90?
60	❌
70	❌
80	❌
90	❌ (equal)
100	✅

✔️ Matches: 1 value greater than 90.

Summary Table

Query	Z-score	Probability	Approx. Count in Dataset
X < 80	0	0.5	2 or 3 values
X > 90	0.71	0.2389	~1 value

Probability Density Curve (PDF) for Continuous Data – Normal Distribution

---

What is a Probability Density Function (PDF)?

A **Probability Density Function (PDF)** describes the **likelihood** of a **continuous random variable** taking on a specific range of values.

For continuous data:

- The **probability at a single point is zero**
- Probability is calculated **over an interval**
- The **area under the curve** between two values gives the probability of falling in that range

The **total area under the curve = 1**, representing 100% probability.

Example: Heights of Students

Suppose we collected the **heights of 200 students**, and they follow a **normal distribution**.

- Mean height (μ) = 165 cm
- Standard deviation (σ) = 10 cm

We want to understand:

- The shape of the distribution
- What percentage of students are between 155 and 175 cm

Graph: Normal Distribution Curve

Normal Distribution Curve

> 🟦 The curve represents the **probability density** of students' heights.
>
> The shaded area between 155 cm and 175 cm gives the **probability that a randomly selected student falls in that range**.

📊 Graph Explanation

🧭 Curve Properties

- **Peak (Center)**: At μ = 165 cm
- **Symmetry**: The curve is symmetrical around the mean
- **Spread**: Controlled by standard deviation (σ = 10 cm)

🔎 Interpretation

Using the normal distribution rule:

- ~68% of students fall within **1 standard deviation** (155 cm to 175 cm)
- ~95% fall within **2 standard deviations** (145 cm to 185 cm)

So, **probability that height is between 155 and 175 cm ≈ 68%**

Formula for Normal Distribution PDF

$$ f(x) = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 } $$

Where:

- $ \mu $: Mean

- $ \sigma $: Standard deviation
- $ x $: Value of the random variable

Feature	Value
Variable	Height of students
Mean (μ)	165 cm
Std. Deviation (σ)	10 cm
Distribution Type	Normal
Probability (155–175)	≈ 68%

- PDF is used to understand **how values are distributed** in continuous data.
- The **area under the curve** between two points gives the **actual probability**.

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 your 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.

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.

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.

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.

🔹 It's like **punishing an innocent person**.
🔹 The chance of making this error is called **alpha (α)**, usually 5%.

❌ 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.

🔹 It's like **letting a guilty person go free**.
🔹 The chance of this error is called **beta (β)**.

🔁 Quick Summary

Decision	Reality (H_0 True)	Reality (H_0 False)
Reject H_0	❌ Type I Error	✅ Correct Decision
Fail to Reject H_0	✅ Correct Decision	❌ Type II Error

What Is a Correct Decision?

In hypothesis testing, a **correct decision** happens when your conclusion **matches the actual truth** about the claim.

Example: Apple-Eating at School

The school canteen says:

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

You test this claim using data from your classmates.

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!**

Summary Table

Your Decision	Reality	Outcome
Reject H_0	H_0 is False	✅ Correct Decision
Fail to Reject H_0	H_0 is True	✅ Correct Decision

A correct decision is like giving the **right answer** in a multiple-choice question after checking the facts!

25-july-2025 Hypothesis Testing

Hypothesis Testing

Hypothesis testing is a method used to **make decisions using data**. It helps us decide whether a claim about a **population** is likely to be true.

Hypothesis Testing Process – With Simple Examples

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

📦 1. Population

A **population** is the **entire group** you want to study.

> Example: All students in your school.

Since it's hard to ask everyone, we collect data from a **sample** and use that to make conclusions about the **population**.

---

2. Hypotheses: H_0 and H_1

In every hypothesis test, we start with two ideas:

Null Hypothesis (H_0)

This is the **default assumption** — nothing has changed or no effect exists.

> Example: H_0 — Students eat **2 apples per day**.

Alternative Hypothesis (H_1)

This is what you're **trying to prove** — that something **has changed**.

> Example: H_1 — Students do **not** eat 2 apples per day.

---

3. Significance Level (α)

This is the **threshold** for deciding whether a result is rare or not.
It's usually set to **0.05** (or 5%).

> It means you're okay with a 5% chance of being wrong if you reject H_0.

---

4. p-value

The **p-value** is the probability of getting your results (or more extreme) **if H_0 is actually true**.

- If the **p-value is small** (usually < 0.05), the result is rare.
- So, we **reject H_0** — meaning we support H_1.

---

5. Decision Rule

- If **p-value ≤ α → Reject H_0** (support H_1)
- If **p-value > α → Fail to reject H_0**

Apple Example Summary

- **Population**: All students in school
- **H_0**: Students eat 2 apples per day
- **H_1**: Students do NOT eat 2 apples per day
- **α (significance level)**: 0.05
- **p-value**: Calculated from sample data
- **Decision**: Compare p-value with α to decide if H_0 should be rejected

---

Hypothesis Testing — Step-by-Step Example with Real Data

Let's test a claim about how many apples students eat daily using real numbers.

The Scenario

The school canteen says:

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

You want to test if this claim is correct using a sample of students.

Step 1: Collect Sample Data

You ask **10 students** how many apples they eat per day.
Here's what they say:

`[1, 2, 2, 3, 1, 2, 1, 2, 3, 1]`

Step 2: Calculate the Sample Mean

Mean = (1 + 2 + 2 + 3 + 1 + 2 + 1 + 2 + 3 + 1) / 10 = **1.8 apples**

🧪 Step 3: Set Up the Hypotheses

- **H_0 (Null Hypothesis):** $\mu = 2$ (students eat 2 apples per day)
- **H_1 (Alternative Hypothesis):** $\mu \neq 2$ (students do not eat 2 apples per day)

📏 Step 4: Choose Significance Level

Let's choose **α = 0.05** (5%)
This means we are okay with a 5% risk of being wrong when rejecting H_0.

📉 Step 5: Calculate the p-value (Using t-test)

Let's assume:

- Sample mean ($\bar{x}$) = 1.8
- Population mean under H_0 (μ) = 2
- Sample standard deviation (s) ≈ 0.79
- Sample size (n) = 10

🔍 t-score formula:

$ t = (x̄ - μ) / (s / √n)$

$ t = (1.8 - 2) / (0.79 / √10) ≈ -0.8 $

Using a t-distribution table or calculator:

- **Degrees of freedom (df) = 9**
- **Two-tailed test**
- The **p-value ≈ 0.44**

✅ Step 6: Make a Decision

- p-value = **0.44**
- α = **0.05**

Since **0.44 > 0.05**, we **fail to reject H_0**.

Final Conclusion

> We **do not have enough evidence** to say students eat a different number of apples than 2 per day.
> The school canteen's claim seems reasonable **based on this data**.

Summary Table

Element	Value
Sample Mean ($\bar{x}$)	1.8
Assumed Mean (μ)	2
Sample Std Dev (s)	~0.79
Sample Size (n)	10
t-score	-0.8
p-value	0.44
Significance Level α	0.05
Decision	❌ Fail to Reject H_0
Conclusion	Not enough evidence to reject the canteen's claim

Example :

🎯 One-Tailed Test (Right Tail Example)

Scenario:

A pencil manufacturer claims the average pencil length is **15 cm**.
You think it is **more than 15 cm**.

Step-by-Step:

- H_0: μ = 15
- H_1: μ > 15 (Right-tailed)

Let's say:

- Sample mean ($\bar{x}$) = 15.5 cm
- Sample standard deviation (s) = 0.4
- n = 25
- α = 0.05

Use **T-test** because population SD is unknown and n < 30:

$T = \frac{15.5 - 15}{0.4 / \sqrt{25}} = \frac{0.5}{0.08} = 6.25$

Look up t critical for df = 24 at $\alpha = 0.05$ → around **1.711**

Since **6.25 > 1.711**, reject H_0 → the pencil is longer than 15 cm

✅ **Conclusion**: Evidence shows the pencil is longer.

🎯 Two-Tailed Test Example

Scenario:

You want to check if pencil length is **not equal to** 15 cm.

- H_0: $\mu = 15$
- H_1: $\mu \neq 15$ (Two-tailed)

Same data as above:

- Sample mean = 15.5
- s = 0.4
- n = 25
- $\alpha = 0.05$ (split into 0.025 in both tails)

Critical value for two-tail at df = 24 is approx **±2.064**

$T = 6.25$

Since **6.25 > 2.064**, reject H_0

✅ **Conclusion**: Pencil length is significantly different from 15 cm.

Z-Test Example (When population SD is known)

- Population SD (σ) = 0.3
- Sample mean = 15.2
- Population mean = 15
- n = 100

$Z = \frac{15.2 - 15}{0.3 / \sqrt{100}} = \frac{0.2}{0.03} = 6.67$

Critical z value at $\alpha = 0.05$ (two-tailed) is ±1.96

Since **6.67 > 1.96**, reject H_0

🎓 Easy Analogy for Students:

- **Z-test**: You know the population well (like candle lengths in a factory)

- **T-test**: You have only small samples (like measuring pencils from one box)

Summary Table

Test Type	Use When	Population SD	Sample Size	Example
Z-Test	Population SD known	Known	Any	Known candle size
T-Test	Population SD unknown	Unknown	< 30	Pencil lengths
One-tail	Directional hypothesis	-	-	Pencil > 15 cm
Two-tail	Non-directional (≠) hypothesis	-	-	Pencil ≠ 15 cm

Perfect! Let's walk through **one simple example each** for:

1. **Independent Two-Sample t-Test**
2. **Dependent (Paired) t-Test**

🎓 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:** $\mu_d = 0$ (No improvement)
- **H_1:** $\mu_d > 0$ (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 ($\alpha = 0.05$, df = 4, one-tailed) ≈ 2.132

4️⃣ Decision:

- 21.97 > 2.132 → ✅ **Reject H_0**

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

🧠 Summary:

Type	Groups	Example	Test Used
Independent	Different (boys vs girls)	Math test comparison	Independent t-test
Dependent	Same students (before vs after)	Coaching class	Paired t-test

---

Would you like this in **markdown or PDF** with formulas shown visually?

📊 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

🧮 Step 1: Calculate Group Means

- Mean A = (40 + 45 + 50) / 3 = 45
- Mean B = (42 + 41 + 44) / 3 ≈ 42.33
- Mean C = (55 + 60 + 65) / 3 = 60
- **Grand Mean (GM)** = Total sum / 9 = 442 / 9 ≈ 49.11

🧮 Step 2: Sum of Squares

🔹 Between Groups (SSB)

$$ SSB = n \cdot \sum (\bar{X}_{group} - GM)^2 $$$$ = 3 \cdot [(45 - 49.11)^2 + (42.33 - 49.11)^2 + (60 - 49.11)^2] = 544.56 $$

🔸 Within Groups (SSW)

$$ SSW = \sum (X_{ij} - \bar{X}_{group})^2 = 50 (A) + 4.66 (B) + 50 (C) = 104.66 $$

🔷 Total Sum of Squares (SST)

$$ SST = SSB + SSW = 544.56 + 104.66 = 649.22 $$

🧾 Step 3: Degrees of Freedom

Source	Formula	Value
Between Groups	(k - 1)	2
Within Groups	(N - k)	6
Total	(N - 1)	8

🧮 Step 4: Mean Squares

- $$ MSB = \frac{SSB}{df_{between}} = \frac{544.56}{2} = 272.28 $$
- $$ MSW = \frac{SSW}{df_{within}} = \frac{104.66}{6} \approx 17.44 $$

🧪 Step 5: F-Ratio

$$ F = \frac{MSB}{MSW} = \frac{272.28}{17.44} \approx 15.61 $$

✅ Final ANOVA Table

Source	SS	df	MS	F
Between Groups	544.56	2	272.28	15.61
Within Groups	104.66	6	17.44	
Total	649.22	8		

📌 Conclusion

Compare the F-value (15.61) with the F-critical value from the F-table at α = 0.05.

- ($F_{critical} \approx 5.14$) for (df1 = 2, df2 = 6)
- Since **15.61 > 5.14**, we **reject the null hypothesis**.

✅ **At least one group has a significantly different average score.**

🧮 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.

There are two main types:

1. **Chi-Square Goodness-of-Fit Test** – Tests how well observed data fits an expected distribution.
2. **Chi-Square Test of Independence** – Tests if two variables are related in a contingency table.

🧪 Example: Chi-Square Test of Independence

🎯 Situation:

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

📊 Data (Observed):

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

🧮 Step 1: Find Expected Values

Use the formula:

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

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.

📌 Summary

Step	Result
Observed Values	Given in table
Expected Values	Calculated using row × col / total

Step	Result
χ^2 Value	20
Degrees of Freedom	1
Decision	Reject H_0 (significant relationship)