Hypothesis Testing

Hypothesis Testing

- In essence, hypothesis testing is a procedure to compute a probability that reflects the strength of the evidence (based on a given sample) for rejecting the null hypothesis.
- The techniques for hypothesis testing depend on
  - The type of outcome variable being analyzed (continuous, dichotomous, discrete)
  - the number of comparison groups in the investigation
  - whether the comparison groups are independent (i.e., physically separate such as men versus women) or dependent (i.e., matched or paired such as pre- and post-assessments on the same participants).

Hypothesis Testing

- Hypothesis testing is a statistical method used to make decisions about a population based on sample data. It involves formulating two competing hypotheses:
  1. Null Hypothesis (H_0): Assumes no effect or no difference (e.g., "There is no difference between the means of two groups").
  2. Alternative Hypothesis (H_1 or Ha): Suggests there is an effect or a difference (e.g., "There is a significant difference between the means of two groups").
- The test determines whether there is enough statistical evidence in a sample to reject the null hypothesis in favor of the alternative hypothesis.

- The Centers for Disease Control (CDC) reported on trends in weight, height and body mass index from the 1960's through 2002.
- The general trend was that Indians were much heavier and slightly taller in 2002 as compared to 1960.
- In 2002, the mean weight for men was reported at 191 pounds.
- Suppose that an investigator hypothesizes that weights are even higher in 2022 (i.e., that the trend continued over the subsequent 20 years).
- The ***research hypothesis*** is that the mean weight in men in 2022 is more than 191 pounds.
- The ***null hypothesis*** is that there is no change in weight, and therefore the mean weight is still 191 pounds in 2022

- In order to test the hypotheses, we select a random sample of Indian males in 2022 and measure their weights. Suppose we have resources available to recruit n=100 men into our sample.
- The sample mean of 197.1 is numerically higher than 191.
- Do the sample data support the null or research hypothesis?
- is this difference more than would be expected by chance?
- We therefore need to determine the likelihood of observing a sample mean of 197.1 or higher when the true population mean is 191.
- We can compute this probability using the Central Limit Theorem.
- Suppose the probability is 0.0087

- Thus, there is less than a 1% probability of observing a sample mean as large as 197.1 when the true population mean is 191.
- we might infer, from our data, that the null hypothesis is probably not true.
- Suppose that the sample data had turned out differently.
- And, sample mean is 192.1
- How likely it is to observe a sample mean of 192.1 or higher when the true population mean is 191 (i.e., if the null hypothesis is true)?
- Again, we compute the probability and it comes 0.3336
- There is a 33.4% probability of observing a sample mean as large as 192.1 when the true population mean is 191.
- Do you think that the null hypothesis is likely true?

- Neither of the sample means that we obtained allows us to know with certainty whether the null hypothesis is true or not.
- In essence, hypothesis testing is a procedure to compute a probability that reflects the strength of the evidence (based on a given sample) for rejecting the null hypothesis.

- In hypothesis testing, we determine a threshold or cut-off point (called the critical value) to decide when to believe the null hypothesis and when to believe the research hypothesis.
- What we need is a threshold value such that if $\bar{x}$ is above that threshold then we believe that H1 is true and if $\bar{x}$ is below that threshold then we believe that H0 is true. The difficulty in determining a threshold for $\bar{x}$ is that it depends on the scale of measurement
- First, to address the issue of scale in determining the critical value, we convert our sample data (in particular the sample mean) into a Z score.

- the center of the Z distribution is zero and extreme values are those that exceed 2 or fall below -2. Z scores above 2 and below -2 represent approximately 5% of all Z values.
- In hypothesis testing, we select a critical value from the Z distribution
- The level of significance is the probability that we reject the null hypothesis (in favor of the alternative) when it is actually true and is also called the Type I error rate.
- **α =Level of significance = P(Type I error) = P(Reject H0 | H0 is true).**

Hypothesis Testing: Upper-, Lower, and Two Tailed Tests

Upper-tailed, Lower-tailed, Two-tailed Tests

The research or alternative hypothesis can take one of three forms. An investigator might believe that the parameter has increased, decreased or changed. For example, an investigator might hypothesize:

1. H_1: $\mu > \mu_0$, where μ_0 is the comparator or null value (e.g., $\mu_0 = 191$ in our example about weight in men in 2006) and an increase is hypothesized - this type of test is called an **upper-tailed test**;
2. H_1: $\mu < \mu_0$, where a decrease is hypothesized and this is called a **lower-tailed test**; or
3. H_1: $\mu \neq \mu_0$, where a difference is hypothesized and this is called a **two-tailed test**.

The exact form of the research hypothesis depends on the investigator's belief about the parameter of interest and whether it has possibly increased, decreased or is different from the null value. The research hypothesis is set up by the investigator before any data are collected.

Hypothesis testing procedure

The procedure can be broken down into the following five steps.

- **Step 1.** Set up hypotheses and select the level of significance α.
  - H0: Null hypothesis (no change, no difference);
  - H1: Research hypothesis (investigator's belief); $\alpha = 0.05$
- **Step 2.** Select the appropriate test statistic.
  - The test statistic is a single number that summarizes the sample information.
  - Example: Z statistic, T statistic

Hypothesis testing procedure

- **Step 3.** Set up decision rule.
- The decision rule is a statement that tells under what circumstances to reject the null hypothesis.
- The decision rule depends on whether an upper-tailed, lower-tailed, or two-tailed test is proposed.
  - In an **upper-tailed** test the decision rule has investigators reject H0 if the test statistic is larger than the critical value.
  - In a **lower-tailed** test the decision rule has investigators reject H0 if the test statistic is smaller than the critical value.
  - In a **two tailed test** the decision rule has investigators reject H0 if the test statistic is extreme, either larger than an upper critical value or smaller than a lower critical value.

- **Step 4:** Compute the test statistic

Step 5: Compare and Conclude

Types of Hypothesis Tests

- Hypothesis tests can be broadly categorized into **parametric** and **non-parametric** tests.

Parametric Tests

- These tests assume that the data follows a specific probability distribution (usually normal distribution) and that certain conditions (like homogeneity of variance) are met.

Test	Purpose	Example Application
Z-test	Compares sample mean with population mean (large samples, known variance)	Testing if the average IQ of students in a school differs from the national average
t-test	Compares means between two groups (small samples, unknown variance)	Comparing test scores of two classes
ANOVA	Compares means across more than two groups	Examining the effect of different diets on weight loss
Chi-Square Test	Tests the association between categorical variables	Checking if gender and voting preference are related
F-test	Compares variances between two or more groups	Comparing variance in scores across multiple schools

Non-Parametric Tests

- These tests do not assume a specific distribution and are useful for small sample sizes or ordinal data.

Test	Purpose	Example Application
Mann-Whitney U test	Compares differences between two independent groups (alternative to t-test)	comparing customer satisfaction ratings between two brands
Wilcoxon Signed-Rank Test	Compares paired samples (alternative to paired t-test)	Before-and-after comparison of blood pressure in patients
Kruskal-Wallis Test	Compares more than two independent groups (alternative to ANOVA)	Comparing customer ratings of three different restaurants
Friedman Test	Compares repeated measures (alternative to repeated-measures ANOVA)	Evaluating the performance of students across multiple subjects
Spearman's Rank Correlation	Measures correlation between two ordinal variables (alternative to Pearson's correlation)	Relationship between student rankings in math and science

Z-Test

Z-test

- A Z-Test is a statistical test used to determine if there is a significant difference between a sample mean and a population mean.
- It is applied when the population variance is known and the sample size is large ($n \geq 30$).
- Based on the standard normal distribution (Z-distribution).

Types of Z-test

- One-Sample Z-Test: Compares a sample mean to a known population mean.
- Two-Sample Z-Test: Compares the means of two independent samples.
- Z-Test for Proportions: Compares proportions of a sample and population or two samples.

Z-test formulas

- **one-sample Z-test**:
  - $z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$
- Where:
  - $\bar{X}$ = Sample mean
  - μ = Population mean
  - σ = Population standard deviation
  - n= Sample size

Z-test

- **two-sample Z-test**:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

- Where subscripts 1 and 2 represent two different samples.

Z-test

- **Assumptions of Z-Test**

1. Data follows a **normal distribution** (or sample size is large enough for the Central Limit Theorem to apply).
2. The sample observations are **independent**.
3. The **population standard deviation (σ) is known**.

Z-test Example 1

- **Problem Statement**
  - A researcher claims that the average height of students in a university is **170 cm**. A random sample of **50 students** is taken, with an average height of **168 cm** and a population standard deviation of **5 cm**. Test the claim at a **5% significance level** (α=0.05).
- **Solution Step 1: Identify Given Data**
  - Sample mean ($\bar{X}$) = 168 cm
  - Population mean (μ) = 170 cm
  - Population standard deviation (σ) = 5 cm
  - Sample size (n) = 50
  - Significance level (α) = 0.05

Z-test

- **Step 2: Set Hypotheses**
  - Null Hypothesis (H_0): $\mu=170$ cm (No significant difference)
  - Alternative Hypothesis (H_1): $\mu\neq170$ cm(There is a significant difference)
- **Step 3: Calculate Z-Value**
  - **Z=-2.83**

Z-test

- **Step 4: Find Critical Value**
  - For **α=0.05 (two-tailed test)**, the critical values from the **Z-table** are **±1.96**.
- **Step 5: Compare and Conclude**
  - **Observed Z-value (-2.83) is beyond -1.96** (in the rejection region).
  - Hence, we **reject H_0** and conclude that there is a significant difference in student heights.
- **Final Conclusion**: The claim that the average height is **170 cm is not supported** by the data at a 5% significance level. i.e. we reject the null hypothesis.

Z-Test Example 2

Problem statement: Suppose the jewelry of exams has a population that is normally distributed with a standard deviation of 5. You are walking down the street and sample 9 exams from this population and obtain a mean jewelry of 28.95 and a standard deviation of 6.3802. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected jewelry of 27?

Z-Test Example 2

Two-tailed Z-test

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{5}{\sqrt{9}} = 1.6667$$

$$z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(28.95 - 27)}{1.6667} = 1.17$$

z_{crit} for $\alpha = 0.01$ (Two tailed) is ± 2.58

We fail to reject H_0.

The jewelry of exams (M = 28.95) is not significantly different than 27, z=1.17, p = 0.242.

Z-Test Example 3

- **Problem statement:** Suppose the courage of psychologists has a population that is normally distributed with a standard deviation of 10. You decide to sample 57 psychologists from this population and obtain a mean courage of 34.81 and a standard deviation of 9.0579. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly greater than an expected courage of 34?

Z-Test Example 3

- One-tailed Z-test

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{10}{\sqrt{57}} = 1.3245$$

$$z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(34.81 - 34)}{1.3245} = 0.61$$

z_{crit} for $\alpha = 0.05$ (One tailed) is 1.64

We fail to reject H_0.

- The courage of psychologists (M = 34.81) is not significantly greater than 34, z=0.61, p = 0.2709.

Two sample Z-test

- Example: A company wanted to compare the performance of its call center employees in two different centers located in two different parts of the country – Hyderabad, and Bengaluru, in terms of the number of tickets resolved in a day (hypothetically speaking). The company randomly selected 30 employees from the call center in Hyderabad and 30 employees from the call center in Bengaluru. The following data was collected:
- Hyderabad: $\bar{x}1 = 750$, $\sigma1 = 20$
- Bengaluru: $\bar{x}2 = 780$, $\sigma2 = 25$
- The company wants to determine if the performance of the employees in Hyderabad is different from the performance of the employees in the Bengaluru center.

- we will use a two-sample z-test for means. We will perform two-tailed test.
- First, we will formulate the null and alternate hypotheses and set the level of significance for the test.
  - H_0: There is no difference between the performance of employees at different call centers.
  - H_1: There is a difference in the performance of the employees.
  - The level of significance is set as 0.05.
- Next, the mean and standard deviation for each sample will need to be determined.
  - Hyderabad: $\bar{x}_1 = 750$, $\sigma_1 = 20$
  - Bengaluru: $\bar{x}_2 = 780$, $\sigma_2 = 25$
- Next, we will calculate the hypothesized difference between the two population means.

- In this case, the company is hypothesizing that the mean performance in Hyderabad

  is the same as that of Bengaluru. So, $(\mu 1 - \mu 2) = 0$

- Finally, we will use the formula for two-sample z-test for means to calculate the test statistic.
  - $z = (\bar{x}1 - \bar{x}2) / \sqrt{((\sigma 1)^2/n1 + (\sigma 2)^2/n2)}$
  - $z = (-30) / \sqrt{((20)^2/30 + (25)^2/30)}$
  - $z = -5.13$
- At a significance level of 0.05, the Z statistic is less than critical value -1.96.
- hence you can reject the null hypothesis.
- Hence, the performance of Hyderabad's team is considered to be not equal to the performance of Bengaluru's team.

Two sample Z-test

- compare the height of two male populations from the United States and Sweden, a sample of 30 males from each country is randomly selected and the measured heights are provided in Table

	Height (Inches)			Height (Inches)			Height (Inches)	
ID	US	Swedish	ID	US	Swedish	ID	US	Swedish
1	69.12	74.56	11	69.18	68.00	21	72.12	72.00
2	66.88	71.89	12	66.18	72.00	22	66.88	70.00
3	74.82	73.00	13	64.94	73.56	23	73.82	69.22
4	67.00	67.78	14	71.76	72.56	24	74.00	74.44
5	69.12	72.22	15	70.12	75.00	25	71.18	68.00
6	65.00	68.00	16	71.00	68.33	26	67.88	73.89
7	71.00	73.56	17	71.88	71.67	27	65.94	70.00
8	66.76	75.00	18	65.24	72.44	28	68.88	70.44
9	72.12	68.22	19	70.06	75.00	29	68.00	70.22
10	72.94	69.00	20	71.94	71.89	30	75.12	73.33

Two sample Z-test

- Currently, the mean and standard deviation for the US and Swedish populations are known as provided in Table below

Population parameter	US Male	Swedish Male
Mean, μ	69.98	70.43
Standard Deviation, σ	3.12	2.44

(Hanson, Sperling et al. 2009; Salvendy 2012)

Two sample Z-test

$$Null\ Hypothesis, H_0: \mu_1 - \mu_2 = 0$$

$$Alternative\ Hypothesis, H_A: \mu_1 - \mu_2 \neq 0$$

$$\text{Where, } \mu_1 = mean\ height\ of\ US\ male\ population\ \&$$

$$\mu_2 = mean\ height\ of\ Swedish\ male\ population$$

Two sample Z-test

Calculation	US male	Swedish Male
Sample mean $\overline{x}$,	69.70	71.51
Population Standard Deviation	3.12	2.44
Sample Size, n	30	30
$Hypothesized\ Difference = \boldsymbol{\mu_1} - \boldsymbol{\mu_2} =$	0	
Z-value	-2.51	
p-value (one sided) =NORMSDIST(-2.51)	0.0061	P-value is Calculated
p-value (two sided) =	0.0122	Using the MS Excel

Two sample Z-test

- We reject the null hypothesis
- Statistically, US and Swedish male populations are significantly different with respect to the height.
- The next question would be then who is taller or shorter. Both the sample and the population data shows that the Swedish male population is taller than the US male population. However, the alternative hypothesis was written as “Not Equal.” Therefore, to test “the Swedish male population is taller than the US male” or “the US male population is shorter than the Swedish male population,” the hypothesis is written as below.

Two sample Z-test

$$Null\ Hypothesis, H_0: \mu_1 - \mu_2 = 0$$

$$Alternative\ Hypothesis, H_A: \mu_1 - \mu_2 < 0$$

Where, $\mu_1 = mean\ height\ of\ US\ male\ population$ &

$$\mu_2 = mean\ height\ of\ Swedish\ male\ population$$

Two sample Z-test

- Now the alternative hypothesis become one-sided. As the one-sided probability is the half of the two-sided probability (*p*-value), we would still reject the null hypothesis.
- The new contextual conclusion would be “Statistically, the US male population is significantly shorter than the Swedish male population.” However, making this contextual conclusion for the original “not equal” alternative hypothesis would be wrong.

T-test

T-Test: Definition & Purpose

- A **T-test** is a statistical test used to compare the means of two groups to determine if they are significantly different from each other.
- It is based on **Student's T-distribution** and is used when the sample size is small ($n < 30$).
- The T-test assumes that the data follows a **normal distribution** and that variances of the groups are equal.

Types of T-Test

1. One-Sample T-Test

- **Purpose:** Compares the mean of a single sample to a known population mean.
- **Formula:**

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

  Where:
  - $\bar{X}$ = Sample mean
  - μ = Population mean
  - s = Sample standard deviation
  - n = Sample size
- **Example Use Case:** Checking if the average test score of a class differs from the national average.

Types of T-Test

2. Independent (Unpaired) T-Test

- **Purpose:** Compares means between two independent groups to check if they are significantly different.
- **Assumptions:**
  - Both groups are independent of each other.
  - Data is normally distributed.
  - Equal variance in both groups (for standard T-test).
- **Formula:**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Where s_p^2 is the pooled variance.
- Example Use Case: Comparing exam scores of students from two different schools.

Types of T-Test

3. Paired (Dependent) T-Test

- **Purpose:** Compares means of the same group before and after an intervention or treatment.
- **Assumptions:**
  - The same subjects are measured twice.
  - Data differences are normally distributed.
- **Formula**

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- Where:
  - $\bar{d}$ = Mean of differences
  - s_d= Standard deviation of differences
- **Example Use Case:** Measuring blood pressure before and after administering a drug in the same patients.

Assumptions of T-Test

- **Normality:** Data should be approximately normally distributed.
- **Homogeneity of Variance:** For independent T-tests, both groups should have similar variances.
- **Independence:** Observations should be independent of each other.
- **Scale of Measurement:** The dependent variable should be continuous (interval/ratio scale).

Steps to Perform a T-Test

- State the Null Hypothesis (H_0): No significant difference between the means.
- State the Alternative Hypothesis (H_A): Significant difference exists.
- Choose a significance level (α), typically 0.05.
- Calculate the T-statistic using the appropriate formula.
- Find the critical value from the T-distribution table.
- Compare the T-statistic with the critical value or use the p-value approach.
- Make a decision:

$$\text{If } |t| > t_{critical} \text{ or } p < \alpha, \text{ reject } H_0$$

Otherwise, fail to reject H_0.

One-Sample T-Test Example

- A teacher claims that the average math test score is **75**. A random sample of **10 students** gives the following scores:
  70, 78, 80, 65, 72, 74, 69, 81, 77, 73
- Test the teacher's claim at a **5% significance level (α = 0.05).**
- **Step 1 - Define Hypotheses**
  - **Null Hypothesis (H_0):** The average test score is **75**
  - **Alternative Hypothesis (H_a):** The average test score is **not 75**
- Since we are checking for any difference, this is a **two-tailed test.**

One-Sample T-Test Example

- **Step 2 - Set Significance Level**
  - Given **$\alpha = 0.05$**, for a **two-tailed test**
  - Degrees of Freedom **$df = n - 1 = 10 - 1 = 9$**
  - Critical **t-value from T-table** (for df = 9 at $\alpha = 0.05$) is **± 2.262**
- **Step 3 - Compute Sample Statistics**
  - **Calculate Sample Mean ($\bar{X}$)**

  - **Calculate Sample Standard Deviation (s)**
    (computed using formula)

One-Sample T-Test Example

- **Step 4 - Compute T-Statistic**
  - Formula:

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

  - Substituting values: t=-0.678
- **Step 5 - Compare with Critical Value**
  - Computed **t-statistic = -0.678**
  - **Critical t-value = ±2.262**
  - Since **|t| = 0.678 < 2.262**, we **fail to reject H_0**

One-Sample T-Test Example

- **Conclusion**
  - There is **not enough evidence** to reject the teacher’s claim.
  - The sample data **does not show a significant difference** from the assumed population mean of **75**.
  - **Final Decision:** The teacher's claim is reasonable based on the given data.

Independent (unpaired) T-Test Example

- **Problem Statement**
  - A researcher wants to test whether there is a significant difference in exam scores between **male and female students**. Two independent samples are selected:
  - **Male Students' Scores:** 85, 88, 90, 75, 78, 82, 87, 92, 85, 80
    Female Students' Scores: 79, 83, 85, 88, 84, 81, 86, 90, 82, 80
  - Test at a **5% significance level (α = 0.05).**

Independent (unpaired) T-Test Example

- **Step 1 - Define Hypotheses**
  - **Null Hypothesis (H_0):** There is no difference in mean exam scores between male and female students.
  - **Alternative Hypothesis (H_a):** There is a significant difference in mean exam scores.
  - Since we are checking for any difference, this is a **two-tailed test.**
- **Step 2 - Set Significance Level**
  - Given **$\alpha = 0.05$**, for a **two-tailed test**
  - Degrees of Freedom calculated using **$df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$**
  - Critical **t-value from T-table** (for df = 18 at $\alpha = 0.05$) is **±2.101**

Independent (unpaired) T-Test Example

- **Compute Sample Statistics**
- **Calculate Mean for Each Group ($\bar{X}_1$ and $\bar{X}_2$)**
  - Male Students Mean: **84.7**
  - Female Students Mean: **83.8**
- **Calculate Standard Deviations (s_1 and s_2)**
  - Male Students Std Dev: **5.84**
  - Female Students Std Dev: **3.67**

Independent (unpaired) T-Test Example

- **Compute T-Statistic**
- Formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where pooled variance is calculated as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Independent (unpaired) T-Test Example

Substituting values

$$t = \frac{84.7 - 83.8}{\sqrt{(4.74)\left(\frac{1}{10} + \frac{1}{10}\right)}} = 0.428$$

- **Compare with Critical Value**
  - Computed **t-statistic = 0.20**
  - **Critical t-value = ±2.101**
  - Since **|t| = 0.20 < 2.101**, we **fail to reject H_0**

Independent (unpaired) T-Test Example

Conclusion

- There is **not enough evidence** to suggest a significant difference between male and female students' scores.
- The sample data **does not show a statistically significant difference**.
- **Final Decision:** The exam scores for male and female students are **statistically similar**.

Paired T-Test Example

Problem Statement

- A teacher wants to know if a **new teaching method** improves student performance. The **same group of students** takes a test **before and after** the method is applied.
- **Pre-Test Scores:** 70, 75, 80, 65, 72, 78, 74, 81, 77, 73
  Post-Test Scores: 75, 78, 85, 70, 76, 82, 78, 86, 80, 76
- Test at a **5% significance level (α = 0.05).**

Paired T-Test Example

- **Define Hypotheses**
  - **Null Hypothesis (H_0):** There is no difference in mean scores before and after the new method.
  - **Alternative Hypothesis (H_a):** There is a significant difference in mean scores.
- Since we are checking for any difference, this is a **two-tailed test.**
- **Set Significance Level**
  - Given **α = 0.05**, for a **two-tailed test**
  - Degrees of Freedom **df = n - 1 = 10 - 1 = 9**
  - Critical **t-value from T-table** (for df = 9 at α = 0.05) is **±2.262**

Paired T-Test Example

- **Compute Sample Statistics**
  - **Calculate the Difference (D) for Each Pair**

    - Differences: **5, 3, 5, 5, 4, 4, 4, 5, 3, 3**
  - **Calculate Mean Difference ($\bar{X}_{a-e}$)**
    - Mean Difference: **4.1**
  - **Calculate Standard Deviation of Differences (s_{a-e})**
    - Standard Deviation: **0.99**

Paired T-Test Example

- **Compute T-Statistic**

    Formula:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

t=13.1

Paired T-Test Example

- **Compare with Critical Value**
  - Computed **t-statistic = 13.1**
  - **Critical t-value = ±2.262**
  - Since **$|t|$ = 13.1 > 2.262**, we **reject H_0**
- **Conclusion**
  - There is **strong evidence** to suggest a significant improvement in student scores.
  - The new teaching method **significantly increased** student performance.
  - **Final Decision:** The teaching method **has a positive impact** on student learning.

Questions

- Based on field experiments, a new variety green gram is expected to given an yield of 12.0 quintals per hectare. The variety was tested on 10 randomly selected farmers fields. The yield (quintals/hectare) were recorded as 14.3,12.6,13.7,10.9,13.7,12.0,11.4,12.0,12.6,13.1. Do the results conform the expectation?

- Independent t-test; We accept the null hypothesis H0 We conclude that the new variety of green gram will give an average yield of 12 quintals/hectare. (t=1.836)

- In certain food experiment to compare two types of baby foods A and B, the following results of increase in weight (lbs) we observed in 8 children as follows

Food A(x)	49	53	51	52	47	50	52	53
Food B(y)	52	55	52	53	50	54	54	53

- Examine the significance of increase in weight of children due to food B.

- Paired t-test; $t=4.32$; $t_{critical} = 2.365$; We reject the null hypothesis H_0 (i.e) there is significant difference between the two foods A and B.

F-Test

F-Test

- Definition: The F-test is a statistical test used to compare the variances of two or more groups to determine if they are significantly different.
- Purpose: It helps in testing whether two population variances are equal or whether a regression model is significant.
- Common Applications:
  - ANOVA (Analysis of Variance): Compares multiple group means by analyzing variances.
  - Regression Analysis: Tests the overall significance of the regression model.
  - Equality of Two Variances: Checks if two populations have the same variance.

F-Test

- Hypotheses in F-Test:
  - Null Hypothesis (H_0): The variances are equal ($\sigma_1^2 = \sigma_2^2$).
  - Alternative Hypothesis (H_1): The variances are different ($\sigma_1^2 \neq \sigma_2^2$).
- Formula for F-Statistic:

$$F = \frac{Variance\ of\ sample\ 1}{Variance\ of\ sample\ 2}$$

  - The larger variance is usually placed in the numerator to ensure $F \geq 1$.
- Degrees of Freedom (df):
  - Numerator df: $n_1 - 1$ (for the first sample).
  - Denominator df: $n_2 - 1$ (for the second sample).

F-Test

- **Assumptions of F-Test:**
  - The samples are drawn from normally distributed populations.
  - The observations are independent.
  - The variances are tested for homogeneity.
- **Decision Rule:**
  - Compare the calculated F-statistic with the critical value from the F-distribution table at a given significance level (α).
  - If $F_{calculated} > F_{critical}$, reject the null hypothesis.
  - Otherwise, fail to reject the null hypothesis.

F-Test example

- **Scenario:** A teacher wants to compare the effectiveness of two teaching methods (A and B) on student test scores. They believe Method B might lead to more *variable* scores (some students excelling, others struggling), while Method A might produce more consistent results.
- **Data:**
  - **Method A:** Scores: 75, 80, 85, 90, 95
  - **Method B:** Scores: 60, 70, 85, 95, 100

F-Test example

- **Null Hypothesis (H0):** The variances of the test scores for both teaching methods are equal. ($\sigma^2 A = \sigma^2 B$)
- **Alternative Hypothesis (H1):** The variances of the test scores for both teaching methods are not equal. ($\sigma^2 A \neq \sigma^2 B$)
- **Calculate Sample Variances:**
- **Method A:**
  - Mean (xA) = (75+80+85+90+95)/5 = 85
  - Variance (s^2A) = $\Sigma(xi - xA)^2 / (n - 1) = [(10)^2 + (5)^2 + (0)^2 + (5)^2 + (10)^2] / 4 = 250/4 = 62.5$
- **Method B:**
  - Mean (xB) = (60+70+85+95+100)/5 = 82
  - Variance (s^2B) = $\Sigma(xi - xB)^2 / (n - 1) = [(22)^2 + (12)^2 + (3)^2 + (13)^2 + (18)^2] / 4 = 1010/4 = 252.5$

F-Test example

- **Calculate the F-statistic:**
  - F = (Larger Sample Variance) / (Smaller Sample Variance)
  - $F = s^2B / s^2A = 252.5 / 62.5 = 4.04$
- **Degrees of Freedom:**
  - Numerator df (df1) = nB - 1 = 5 - 1 = 4
  - Denominator df (df2) = nA - 1 = 5 - 1 = 4
- critical F-value is approximately 6.39

F-Test example

Decision:

- Our calculated F-statistic (4.04) is *less than* the critical F-value (6.39).
- Therefore, we *fail to reject* the null hypothesis.

- Conclusion: There is not enough evidence to conclude that the variances of the test scores for the two teaching methods are significantly different at the 0.05 significance level. While Method B appears to have a larger variance, the difference is not statistically significant.

Chi Square Test (χ^2)

Chi Square Test (χ^2)

- This test is a test of Independence or association between two categorical variables
- It is used to analyse the frequencies of two variables with multiple categories to determine whether the two variables are independent
- Qualitative variables
- Nominal data

Note: If your categorical variables represent "pre-test" and "post-test" observations, then the chi-square test of independence **is not appropriate**. This is because the assumption of the independence of observations is violated. In this situation, McNemar's Test is appropriate.

Chi Square Test (χ^2)

- Example: Investment data
  - Is there any dependency between geographical regions and the investment type
  - A survey was conducted, where following questions were asked:
    - In which region of the country do you reside?
    - A. Northeast B.Midwest C.South D.West
    - Which type of financial investement you prefer?
    - E. Stocks F.Bonds G.Fixed Deposits

Chi Square Test (χ^2)

- Based on above data we create a contingency table

	E	F	G	
A				n_A
B				n_B
C				n_C
D				n_D
	n_E	n_F	n_G	N

Chi Square Test (χ^2)

- Based on above data we create a contingency table

If A and F are independent,
$P(A \cap F) = P(A) \cdot P(F)$

$$P(A) = \frac{n_A}{N} \qquad P(F) = \frac{n_F}{N}$$

$$P(A \cap F) = \frac{n_A}{N} \cdot \frac{n_F}{N}$$

$$e_{AF} = N \cdot P(A \cap F) = N\left(\frac{n_A}{N} \cdot \frac{n_F}{N}\right) = \frac{n_A \cdot n_F}{N}$$

	E	F	G	
A				n_A
B				n_B
C				n_C
D				n_D
	n_E	n_F	n_G	N

Chi Square Test (χ^2)

- Expected Frequency:

$$e_{ij} = \frac{n_i \times n_j}{N}$$

- Where n_i ➔ the total of row i
- n_j ➔ the total of row j
- N ➔ the total of all frequencies

Chi Square Test (χ^2)

- Calculated χ^2 (observed χ^2)

$$\chi^2 = \sum\sum\frac{(f_o - f_e)^2}{f_e}$$

- Here: degree of freedom = (r-1)(c-1)
- r→no. of rows, c→ no. of columns

Chi Square Test (χ^2)

- Example:

  H0: Type of fuel is independent of income

  H1: Type of fuel is not independent of income

Income\Type of fuel	Regular	Premium	Extra Premium	
Less than 30000	85	16	6	107
30000 to 49999	102	27	13	142
50000 to 99000	36	22	15	73
At least 100000	15	23	25	63
	238	88	59	385

Chi Square Test (χ^2)

E11= (107 × 238)/385=66.15
E12=(107× 88)/385 = 24.46
.
.
E21=(238×142)/385 = 87.78

- Example:

  H0: Type of fuel is independent of income

  H1: Type of fuel is not independent of income

Income\Type of fuel	Regular	Premium	Extra Premium	
Less than 30000	85 (66.15)	16 (24.46)	6	107
30000 to 49999	102 (87.78)	27	13	142
50000 to 99000	36	22	15	73
At least 100000	15	23	25	63
	238	88	59	385

Chi Square Test (χ^2)

$$\chi^2 = \sum\sum \frac{(f_o - f_e)^2}{f_e}$$

$$= \frac{(85-66.15)^2}{66.15} + \frac{(16-24.46)^2}{24.46} + \frac{(6-16.40)^2}{16.40} +$$

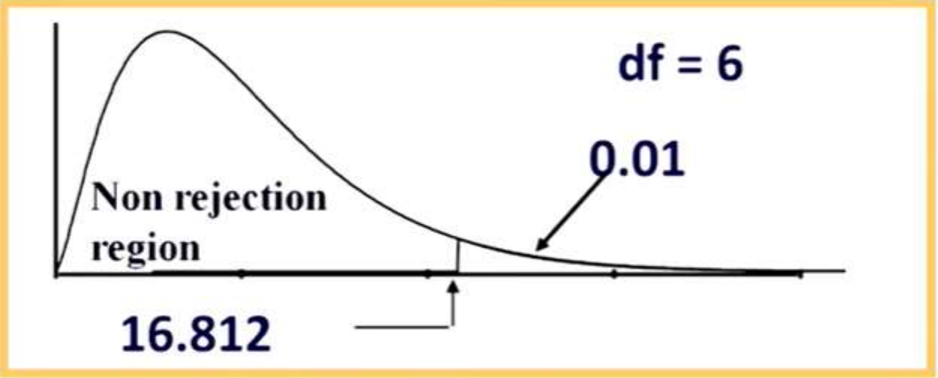
$$\frac{(102-87.78)^2}{87.78} + \frac{(27-32.46)^2}{32.46} + \frac{(13-21.76)^2}{21.76} +$$

$$\frac{(36-45.13)^2}{45.13} + \frac{(22-16.69)^2}{16.69} + \frac{(15-11.19)^2}{11.19} +$$

$$\frac{(15-38.95)^2}{38.95} + \frac{(23-14.40)^2}{14.40} + \frac{(25-9.65)^2}{9.65}$$

$$= 70.75$$

Chi Square Test (χ^2)



$$\chi^2_{Cal} = 70.78 > 16.812, \text{ reject } H_o.$$