## Hypothesis Testing:

**Definition of Hypothesis Testing:** Hypothesis testing is a method used to make statistical decisions using experimental data. It involves testing an assumption (hypothesis) by determining the likelihood that a sample statistic could have occurred, given the null hypothesis (H0) is true.

**Null and Alternative Hypotheses:**

- The null hypothesis (H0) is a statement of no effect or no difference and is assumed true until evidence indicates otherwise.
- The alternative hypothesis (H1) is what a researcher wants to prove. It is a statement that contradicts the null hypothesis and indicates the presence of an effect or difference.

**Significance Level (α):** This is the threshold for rejecting the null hypothesis. A common choice for α is 0.05, which means there's a 5% risk of rejecting the null hypothesis when it's actually true (Type I error).

**P-Value:** The p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so it is rejected in favor of the alternative hypothesis.

**Type I and Type II Errors:**

**Type I Error:** Rejecting the null hypothesis when it is true.

**Type II Error:** Failing to reject the null hypothesis when it is false.

**Test Statistic:** This is a standardized value that is calculated from sample data during a hypothesis test. It's used to decide whether to reject the null hypothesis.

**Decision Rule:** Based on the test statistic and the p-value, a decision is made to either reject or not reject the null hypothesis.

## Student's t-Test

**Independent (Unpaired) t-Test:** Used when comparing the means of two independent groups (e.g., control group vs. experimental group).

**Paired t-Test:** Used when comparing means from the same group at different times (e.g., before and after a treatment in the same subjects).

### Assumptions

**Normality:** The data in each group should be approximately normally distributed.

**Homogeneity of variances:** The variances of the two groups should be equal. When this assumption is not met, a variation of the t-test called Welch's t-test can be used.

**Independence:** The data points in each group should be independent of each other.

### Procedure

**State the Hypotheses:**

- Null Hypothesis (H0): The means of the two groups are equal.
- Alternative Hypothesis (H1): The means of the two groups are not equal (or one mean is greater/less than the other, depending on the test type).

**Calculate the Test Statistic:** The t-value is calculated using the sample data. The formula varies depending on whether it's an independent or paired t-test.

**Determine the p-Value:** The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis. It is obtained from the t-distribution.

**Make a Decision:** If the p-value is less than the chosen significance level (α, often set at 0.05), reject the null hypothesis. This suggests that there is a statistically significant difference between the group means.

### Interpretation

**Reject H0:** There is enough evidence to suggest a significant difference between the group means.

**Fail to Reject H0:** There is not enough evidence to suggest a significant difference between the group means.

independent (unpaired) t-test

## Independent (Unpaired) t-Test example & kinda Welch's t-Test:

Imagine a study to test the effectiveness of a new teaching method. The test scores of students taught using the traditional method are compared to those taught with the new method. Here are the scores:

**Traditional Method:** 70,75,70,73,68,74

**New Method:** 76,81,79,74,77,73,

### Hypotheses

**Null Hypothesis (H0):** There is no difference in mean scores between the two teaching methods. μ_trad = μ_new

**Alternative Hypothesis (H1):** There is a difference in mean scores between the two teaching methods μ_trad != μ_new

**Formula for Independent Two-Sample t-Test:**

For an independent two-sample t-test (comparing two separate groups), the test statistic (t-value) is calculated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where:

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means of the two groups.
- $s_p$ is the pooled standard deviation of the two samples.
- $n_1$ and $n_2$ are the sample sizes of the two groups.

The pooled standard deviation, $s_p$, is calculated as:

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

Where $s_1^2$ and $s_2^2$ are the variances of the two samples.

### Calculations

**Calculate the Mean and Standard Deviation for Each Group:**

- Traditional Method: Mean (M1), Standard Deviation (SD1)
- New Method: Mean (M2), Standard Deviation (SD2)

**Calculate the Test Statistic (t-value):**

$$t = \frac{M1 - M2}{\sqrt{\frac{SD1^2}{n1} + \frac{SD2^2}{n2}}}$$

Where $n1$ and $n2$ are the sample sizes of the two groups.

**Degrees of Freedom (df):** n1+n2−2

**Determine the p-Value:** Using the calculated t-value and df, the p-value can be found using a t-distribution table or statistical software.
**Make a Decision:** Compare the p-value to the significance level (typically 0.05). If p-value < 0.05, reject H0.
**Mean of Traditional Method:** 71.67
**Standard Deviation of Traditional Method:** 2.73
**Mean of New Method:** 76.67
**Standard Deviation of New Method:** 3.01
**t-value:** -3.012
**p-value:** 0.0132

Count, N: 6
Sum, Σx: 430
Mean, x̄: 71.666666666667
Variance, $s^2$: 7.4666666666667

**Steps**

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{N - 1}$$
$$= \frac{(70 - 71.666666666667)^2 + ... + (74 - 71.666666666667)^2}{6 - 1}$$
$$= \frac{37.333333333333}{5}$$
$$= 7.4666666666667$$
$$s = \sqrt{7.4666666666667}$$
$$= 2.7325202042559$$

Interpretation

The t-value of -3.012 and a p-value of 0.0132 suggest that there is a statistically significant difference between the two teaching methods. Since the p-value (0.0132) is less than the typical alpha level of 0.05, we reject the null hypothesis. This indicates that the new teaching method has a significantly different effect on student test scores compared to the traditional method.

## Paired t-Test example:

Suppose we want to evaluate the effectiveness of a new diet program. We measure the weights of 5 individuals before and after completing the program. The goal is to determine if there is a significant change in weight due to the program.
**Weights Before Program:** 200,220,215,210,205 pounds
**Weights After Program:** 190,210,205,205,200 pounds

Hypotheses

**Null Hypothesis (H0):** The diet program does not significantly change weight.
**Alternative Hypothesis (H1):** The diet program significantly changes weight.

Calculations

**Calculate the Differences** for each individual's weight before and after the program.
**Compute the Mean and Standard Deviation** of the Differences.
**Calculate the Test Statistic (t-value)**:

$$t = \frac{\text{Mean of Differences}}{\text{Standard Deviation of Differences}/\sqrt{n}}$$

Where $n$ is the number of pairs (individuals in this case).

**Degrees of Freedom (df)**: n−1
**Determine the p-Value** using a t-distribution table or statistical software, based on the calculated t-value and df.
**Make a Decision:** If the p-value is less than the chosen significance level (typically 0.05), reject H0.

Results of Calculations

**Mean of Differences (Before - After):** 8.0 pounds
**Standard Deviation of Differences:** 2.74
**t-value:** 6.532
**p-value:** 0.0028

Interpretation

The t-value of 6.532 and a p-value of 0.0028 indicate that there is a statistically significant difference in weights before and after the diet program. Since the p-value (0.0028) is much less than the typical alpha level of 0.05, we reject the null hypothesis. This suggests that the diet program has a significant effect on reducing weight.

## Analysis of Variance (ANOVA):

Analysis of Variance is a statistical method used to compare the means of three or more groups to determine if at least one group mean is significantly different from the others. It's particularly useful when dealing with multiple groups where multiple t-tests would increase the risk of a Type I error (incorrectly rejecting a true null hypothesis).

- ANOVA tests the null hypothesis that all group means are the same against the alternative hypothesis that at least one group mean is different.

Types of ANOVA

**One-way ANOVA:** Used when comparing the means of three or more groups based on one independent variable. For example, testing the effect of different diets on weight loss.
**Two-way ANOVA:** Used when comparing groups based on two independent variables. For instance, studying the effect of diet and exercise on weight loss.

Assumptions

**Independence of Observations:** Each group's observations should be independent of the others.
**Normality:** The data in each group should be approximately normally distributed.
**Homogeneity of Variances**: The variances among the groups should be approximately equal. This is known as the assumption of homoscedasticity.

Procedure

- Null Hypothesis (H0): All group means are equal.
- Alternative Hypothesis (H1): At least one group mean is different.

**Partition the Total Variation:**

- Between-Group Variation: Variation due to the interaction between the groups.
- Within-Group Variation: Variation due to differences within individual groups.

**Calculate the F-Statistic:**
F = (Variation Between Groups / Degrees of Freedom Between) /
(Variation Within Groups / Degrees of Freedom Within)
This ratio compares the between-group variance to the within-group variance.
**Determine the p-Value:** Using the calculated F-statistic and degrees of freedom, the p-value can be obtained from an F-distribution table or statistical software.
**Make a Decision:** If the p-value is less than the chosen significance level (usually 0.05), reject H0.
Interpretation
**Reject H0:** Indicates that at least one group mean is significantly different from the others.
**Fail to Reject H0:** Suggests that there is no evidence of a significant difference in means across the groups.

**ANOVA example:**
Imagine an experiment to test the effectiveness of three different fertilizers on plant growth. We have three groups of plants, each treated with a different fertilizer, and we measure their growth in centimeters after a fixed period
**Fertilizer A:** 15,12,14,16
**Fertilizer B:** 22,20,23,21
**Fertilizer C:** 14,15,13,16
**Null Hypothesis (H0):** The mean growth of plants is the same for all fertilizers.
**Alternative Hypothesis (H1):** At least one fertilizer leads to a different mean growth.
Calculations
- Calculate the Mean for Each Group and for the overall data.
- Calculate the Sum of Squares Between Groups (SSB) and the Sum of Squares Within Groups (SSW).
- Calculate the Mean Squares:
  - **Mean Square Between (MSB) =** SSB / (number of groups - 1)
  - **Mean Square Within (MSW)** = SSW / (total number of observations - number of groups)
- Calculate the F-Statistic: **F** = MSB / MSW
- Determine the p-Value using an F-distribution table or statistical software based on the F-statistic and degrees of freedom.
- **Make a Decision:** If the p-value is less than the significance level (usually 0.05), reject H0.

Results of Calculations
**Means for Each Group:**
- Fertilizer A: 14.25 cm
- Fertilizer B: 21.5 cm
- Fertilizer C: 14.5 cm
- Overall Mean = 16.75

$$\text{SSB} = 4(14.25 - 16.75)^2 + 4(21.5 - 16.75)^2 + 4(14.5 - 16.75)^2$$

$$\text{SSW} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

SSW:
**Calculate the Mean of Each Group**: First, find the mean (average) of each group. For instance, if Group 1 has data points [15, 12, 14, 16], its mean would be (15 + 12 + 14 + 16) / 4 = 14.25.
**Calculate the Deviations for Each Data Point**: For each data point in each group, calculate the deviation from the group mean. This is done by subtracting the group mean from each data point. For example, the deviations for Group 1 would be [15 - 14.25, 12 - 14.25, 14 - 14.25, 16 - 14.25].
**Square Each Deviation**: Next, square each of these deviations. Continuing with Group 1, the squared deviations would be [$(15 - 14.25)^2$, $(12 - 14.25)^2$, $(14 - 14.25)^2$, $(16 - 14.25)^2$].
**Sum the Squared Deviations for Each Group**: Add up all the squared deviations within each group. For Group 1, it would be the sum of all the squared deviations calculated in the previous step.
**Sum Across All Groups**: Finally, add up these sums for each group to get the total SSW. This is the sum of the squared deviations for all groups.
**SSB:** 135.5
**SSW:** 18.75
**Mean Square Between (MSB):** 67.75
**Mean Square Within (MSW):** 2.08
**F:** 32.52
The calculation of the F-statistic, which is 32.52 in this case, is based on comparing the variance between the groups (MSB) to the variance within the groups (MSW). The large F-value suggests that the variability among group means is more than what would be expected by chance. Given the low p-value obtained earlier (0.0000761), we can conclude that there is a statistically significant difference in plant growth among the different fertilizer treatments. This suggests that the type of fertilizer used has a significant impact on plant growth

**Degrees of Freedom Between Groups (DFB):** the number of groups - 1 (2 in the example)
**Degrees of Freedom Within Groups (DFW):** the total number of observations across all groups - the number of groups (12 - 3 = 9 in the example)
**Total DF=** $N$−1 (11)
**Total Sum of Squares (SST) =** SSB+SSW (calculated as whole the dataset as merged, one mean)