Statistics is a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data.

**Annova test:-**

**Source of Variation:**

**"Between Groups":** This represents the variation between the different groups (in this case, the classes or categories you are comparing, e.g., Class A, Class B, Class C).

"Within Groups": This represents the variation within each group.

Sum of Squares (SS):

**"Between Groups SS":** This is the sum of squares between the group means and the overall mean. It quantifies the variability between the group means.

**"Within Groups SS":** This is the sum of squares within each group, measuring the variability of data points within each group.

"Total SS": The total sum of squares, which is the sum of the between groups and within groups sums of squares. It represents the total variability in the data.

**Degrees of Freedom (df):**

**"Between Groups df":** The degrees of freedom associated with the between groups variation. It's typically the number of groups minus one (k - 1), where k is the number of groups.

"Within Groups df": The degrees of freedom associated with the within groups variation. It's the total number of observations minus the number of groups (N - k), where N is the total number of observations and k is the number of groups.

**Mean Square (MS):**

"Between Groups MS": This is the mean square for the between groups variation. It's calculated by dividing the between groups sum of squares by the between groups degrees of freedom (MS = SS / df).

"Within Groups MS": This is the mean square for the within groups variation. It's calculated by dividing the within groups sum of squares by the within groups degrees of freedom (MS = SS / df).

**F-statistic (F):**

The F-statistic is the ratio of the between groups mean square to the within groups mean square (F = MS(Between Groups) / MS(Within Groups)). It measures whether there are significant differences among the group means. A larger F-value suggests greater differences.

**P-value (P-value):**

The p-value associated with the F-statistic. It tells you the probability of observing such an extreme F-statistic if there were no significant differences among the group means. A smaller p-value (typically less than your chosen significance level, e.g., 0.05) indicates significance.

**F Critical (F crit):**

The critical F-value, which is used to compare with the calculated F-statistic. If the calculated F-statistic is greater than the critical F-value, it suggests that there are significant differences among the group means.

In your specific analysis:

The "Between Groups" variation has an F-statistic of approximately 3.846 with a p-value of approximately 0.0356.

The "Within Groups" variation does not have an associated F-statistic or p-value because it is part of the background information for the analysis.

**SS CALCULATION FOR ANNOVA SINGLE FACTOR :-**

Group A: 10, 12, 14

Group B: 15, 18, 20

Group C: 22, 25, 28

Here's how we calculate SSB:

Step 1: Calculate the mean for each group (mean\_i):

Mean for Group A (mean\_A) = (10 + 12 + 14) / 3 = 12

Mean for Group B (mean\_B) = (15 + 18 + 20) / 3 = 17.67 (rounded to two decimal places)

Mean for Group C (mean\_C) = (22 + 25 + 28) / 3 = 25

Step 2: Calculate the Grand Mean (mean of all data points):

Grand Mean = (10 + 12 + 14 + 15 + 18 + 20 + 22 + 25 + 28) / 9 = 18.11 (rounded to two decimal places)

Step 3: Calculate the squared differences for each group:

For Group A: (mean\_A - Grand Mean)^2 = (12 - 18.11)^2 = 37.05 (rounded to two decimal places)

For Group B: (mean\_B - Grand Mean)^2 = (17.67 - 18.11)^2 = 0.20 (rounded to two decimal places)

For Group C: (mean\_C - Grand Mean)^2 = (25 - 18.11)^2 = 47.40 (rounded to two decimal places)

Step 4: Calculate SSB:

SSB = Σ (ni \* (mean\_i - Grand Mean)^2)

SSB = (3 \* 37.05) + (3 \* 0.20) + (3 \* 47.40)

SSB = 111.15 + 0.60 + 142.20

SSB = 253.95 (rounded to two decimal places)

Group A: 10, 12, 14

Group B: 15, 18, 20

Group C: 22, 25, 28

**how we calculate SSE:(that is sum of square error within groups)**

Step 1: Calculate the mean for each group (mean\_i):

Mean for Group A (mean\_A) = 12

Mean for Group B (mean\_B) = 17.67

Mean for Group C (mean\_C) = 25

Step 2: Calculate the squared differences for each data point within each group:

For Group A:

(10 - mean\_A)^2 = (10 - 12)^2 = 4

(12 - mean\_A)^2 = (12 - 12)^2 = 0

(14 - mean\_A)^2 = (14 - 12)^2 = 4

For Group B:

(15 - mean\_B)^2 = (15 - 17.67)^2 = 7.08 (rounded to two decimal places)

(18 - mean\_B)^2 = (18 - 17.67)^2 = 0.12 (rounded to two decimal places)

(20 - mean\_B)^2 = (20 - 17.67)^2 = 5.38 (rounded to two decimal places)

For Group C:

(22 - mean\_C)^2 = (22 - 25)^2 = 9

(25 - mean\_C)^2 = (25 - 25)^2 = 0

(28 - mean\_C)^2 = (28 - 25)^2 = 9

Step 3: Sum the squared differences within each group:

For Group A: 4 + 0 + 4 = 8

For Group B: 7.08 + 0.12 + 5.38 = 12.58 (rounded to two decimal places)

For Group C: 9 + 0 + 9 = 18

Step 4: Calculate SSE:

SSE = Σ Σ (Xi - mean\_i)^2

SSE = (8 in Group A) + (12.58 in Group B) + (18 in Group C)

SSE = 8 + 12.58 + 18

SSE = 38.58 (rounded to two decimal places)

So, in this example, the Sum of Squares Within (SSE) is approximately 38.58. This value represents the variation within each group, showing how individual data points within each group differ from their respective group means.

**Annova two factor with replication:-**

Sample (Food Species): You have different plant species, such as Species A, Species B. Each of these food species represents a different "Sample" or group in your analysis.

Columns (column Conditions): fish weight and fish height.

Interaction:-Significant interaction between sample and columns

**Source of Variation:** This column describes the different factors or sources of variation that are being considered in the analysis.

**SS (Sum of Squares):** This is a measure of the total variation attributed to each factor. In ANOVA, the SS for each factor indicates how much variation in the data can be attributed to that factor.

**df (Degrees of Freedom):** Degrees of freedom represent the number of values in the final calculation of a statistic that are free to vary. For the factors Sample, Columns, and Interaction, df is 1. The df for the Within factor is 12. The Total df is 15.

**MS (Mean Square):** MS is calculated by dividing the SS by the df for each factor. It represents the average variability within each source of variation.

**F (F-statistic):** The F-statistic is a ratio of the mean square values for different factors. It's used to test whether there are significant differences between the groups.

**P-value:** The p-value associated with each F-statistic is used to determine the statistical significance of the differences between groups. A small p-value (typically less than 0.05) suggests that the variation between groups is significant.

**F crit (Critical F-value):** This is the critical value of the F-statistic, which is compared to the calculated F-statistic to determine significance. If the calculated F-statistic is greater than the critical F-value, it suggests that there is a significant difference between the groups.

**Sample**: The F-statistic is 0.8575 with a p-value of 0.3727. Since the p-value is greater than 0.05 (the common significance level), there is no significant difference between the sample groups.

**Columns**: The F-statistic is 0.0036 with a p-value of 0.9528. Again, the p-value is much greater than 0.05, indicating no significant difference between the columns.

**Interaction**: The F-statistic is 3.8132 with a p-value of 0.0746. While the p-value is somewhat close to 0.05, it's not below it, so there's not strong evidence of significant interaction between the sample and columns.

**Within**: The table doesn't show an F-statistic or p-value for the Within factor. This is likely because it's often not necessary to perform a hypothesis test within the same group; instead, it's the differences between groups that are typically of interest.

**Annova Two factor test without replication:-**

**Error**: The "Error" component typically represents the variability within the groups that is not explained by the factors being tested.

**Rows**: The F-statistic is 0.3904 with a p-value of 0.9113. Since the p-value is much greater than 0.05 (the common significance level), there is no significant difference between the rows.

**Columns**: The F-statistic is 0.0381 with a p-value of 0.8495. Again, the p-value is much greater than 0.05, indicating no significant difference between the columns.

**Error**: The "Error" component typically represents the variability within the groups that is not explained by the factors being tested. In this case, the Error SS is 9921.45, and it has 9 degrees of freedom. The Mean Square (MS) for Error is 1102.38.

The t-test is used when the population standard deviation is unknown or when working with small sample sizes, relying on sample data to estimate population parameters.

The z-test is applied when the population standard deviation is known or when working with large sample sizes, making it appropriate for testing hypotheses based on known population parameters.

**One-Tail (or One-Sided) T-Test:**

A one-tail t-test is used when you have a specific directional hypothesis. You are interested in testing if the sample mean is significantly greater than or less than a certain value, but not both.

**Two-Tail (or Two-Sided) T-Test:**

A two-tail t-test is used when you have a non-directional or two-sided hypothesis. You are interested in testing if the sample mean is significantly different from a certain value, without specifying whether it is greater or less.

**One sample T-TEST:-**

Male:

mean: 37.85714286

Variance: 657.4761905

Observations: 7

Dummy:

Mean: 0

Variance: 0

Observations: 3

Hypothesized Mean Difference: 30

Degrees of Freedom (df): 6

t Statistic (t): 0.810725166

**P(T<=t) one-tail: 0.224239216**

This represents the probability of observing a t-statistic as extreme as, or more extreme than, 0.810725166 in only one tail of the t-distribution.

In this case, the t-statistic is relatively close to the center of the distribution, and the probability is 0.224239216. It means there's a 22.42% chance of observing a t-statistic as extreme as 0.810725166 in one direction, depending on your test's hypothesis (left or right-tailed).

**t Critical one-tail: 1.943180281**

This is the critical t-value for a one-tailed test at a specific significance level (usually denoted as alpha, such as 0.05). It represents the threshold beyond which you would typically reject the null hypothesis.

In this case, if you're conducting a one-tailed test and your significance level is 0.05, you would compare your t-statistic (0.810725166) to the critical t-value (1.943180281). Since the t-statistic is less extreme than the critical value, you would typically fail to reject the null hypothesis.

**P(T<=t) two-tail: 0.448478432**

This represents the probability of observing a t-statistic as extreme as, or more extreme than, 0.810725166 in both tails of the t-distribution.

In a two-tailed test, this value takes into account both the possibility of a positive or a negative difference from the null hypothesis.

The probability is 0.448478432, which means there's a 44.85% chance of observing a t-statistic as extreme as 0.810725166, regardless of the direction, under a two-tailed test.

**t Critical two-tail: 2.446911851**

This is the critical t-value for a two-tailed test at a specific significance level (usually alpha divided by 2). It represents the threshold beyond which you would typically reject the null hypothesis in a two-tailed test.

In this case, if you're conducting a two-tailed test with a significance level of 0.05 (alpha/2 = 0.025), you would compare your t-statistic to the critical t-value. Since 0.810725166 is less extreme than 2.446911851 (in absolute terms), you would typically fail to reject the null hypothesis in a two-tailed test.

In summary, the key point is to compare calculated t-statistic to the critical t-value based on your chosen significance level and the type of test you are conducting (one-tailed or two-tailed). In your example, the t-statistic (0.810725166) does not exceed the critical values, which suggests that, at a 0.05 significance level, you would typically fail to reject the null hypothesis.

**T-TEST TWO SAMPLE:-**

Male:

Mean: 37.85714286

Variance: 657.4761905

Observations: 7

Female:

Mean: 45.14285714

Variance: 674.1428571

Observations: 7

Hypothesized Mean Difference: 0

This represents the difference in means you are testing against. In this case, you are testing whether the mean difference between "Male" and "Female" is equal to zero, which implies no difference.

Degrees of Freedom (df): 12

The degrees of freedom in this context are typically calculated as (N1 + N2 - 2), where N1 is the number of observations in the "Male" group and N2 is the number of observations in the "Female" group. So, df = 7 + 7 - 2 = 12.

t Statistic (t): -0.528239845

The t-statistic is a measure of how many standard errors the sample mean is away from the hypothesized population mean (0 in this case).

The negative sign indicates that the sample mean for "Male" (37.85714286) is less than the sample mean for "Female" (45.14285714), as expected.

**P(T<=t) one-tail: 0.303481941**

This value represents the probability of observing a t-statistic as extreme as, or more extreme than, -0.528239845 in one tail of the t-distribution.

In this context, you seem to be conducting a one-tailed test, which suggests that you are specifically interested in testing whether the "Male" group has a significantly lower mean than the "Female" group.

A value of 0.303481941 indicates that there's a 30.35% chance of observing a t-statistic as extreme as -0.528239845 in the left tail of the distribution (indicating that "Male" has a lower mean), depending on your chosen significance level.

**t Critical one-tail: 1.782287556**

This is the critical t-value for a one-tailed test at a specific significance level (usually denoted as alpha, such as 0.05). It represents the threshold beyond which you would typically reject the null hypothesis.

In this case, with a one-tailed test, if your calculated t-statistic is more extreme (in this case, more negative) than -1.782287556, you might reject the null hypothesis.

**P(T<=t) two-tail: 0.606963882**

This value represents the probability of observing a t-statistic as extreme as, or more extreme than, -0.528239845 in both tails of the t-distribution. It takes into account both the possibility of a negative and a positive difference in means.

A value of 0.606963882 indicates that there's a 60.70% chance of observing a t-statistic as extreme as -0.528239845, regardless of the direction, under a two-tailed test.

**t Critical two-tail: 2.17881283**

This is the critical t-value for a two-tailed test at a specific significance level (usually alpha divided by 2). It represents the threshold beyond which you would typically reject the null hypothesis in a two-tailed test.

In a two-tailed test, if your calculated t-statistic falls outside the range of -2.17881283 and 2.17881283 (in absolute terms), you might reject the null hypothesis.

In your specific test, you have a one-tailed test, and the t-statistic (-0.528239845) is not more extreme than the critical value (-1.782287556), so you would likely fail to reject the null hypothesis. This suggests that, at the chosen significance level, there isn't enough evidence to conclude that the "Male" group has a significantly lower mean than the "Female" group.