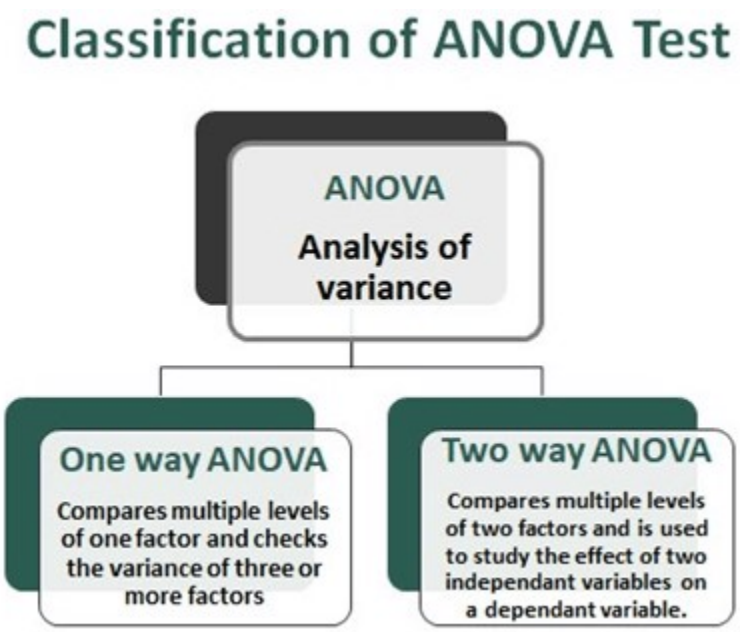
# ANAVO- Analysis of Variance




## One way –

```
import scipy.stats as stats
stats.f_oneway(dataset["ssc_p"],dataset["hsc_p"],dataset["degree_p"])
```

```
F_onewayResult(statistic=0.695991304348204, pvalue=0.49895574725815317)
```

The hypothesis statement for the ANOVA test can be formulated as follows:

Null Hypothesis (H0): There are no significant differences in the means of "ssc_p", "hsc_p", and "degree_p" across the groups.

Alternative Hypothesis (H1): There are significant differences in the means of "ssc_p", "hsc_p", and "degree_p" across the groups.

# Two-Way ANOVA

**ANOVA (Analysis of Variance)** is a statistical test used to analyze the difference between the means of more than two groups. Use a two-way ANOVA when you want to know how two independent variables, in combination, affect a dependent variable.

```
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

```
formula = 'degree_p ~ C(gender) + C(hsc_s) + C(gender):C(hsc_s)'
model = ols(formula, data=dataset).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
```

```
anova_table
```

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| C(gender) | 408.905009 | 1.0 | 7.984433 | 0.005176 |
| C(hsc_s) | 388.809949 | 2.0 | 3.796024 | 0.024024 |
| C(gender):C(hsc_s) | 39.975847 | 2.0 | 0.390292 | 0.677352 |
| Residual | 10703.471564 | 209.0 | NaN | NaN |

Code -

formula = 'degree_p ~ C(gender) + C(hsc_s) + C(gender):C(hsc_s)': This line defines the formula for the ANOVA analysis. The dependent variable is degree_p, and the independent variables are gender, hsc_s, and the interaction between gender and hsc_s. The C() function is used to specify that the variables should be treated as categorical.

model = ols(formula, data=dataset).fit(): This line fits the ANOVA model using the specified formula and the provided dataset. The ols() function performs ordinary least squares regression, and the .fit() method is called to estimate the model parameters.

anova_table = sm.stats.anova_lm(model, typ=2): This line computes the ANOVA table using the fitted model. The anova_lm() function calculates the sums of squares, degrees of freedom, F-statistic, and p-values for each term in the model.

the output:

The C(gender) term has a sum of squares of 408.905009, a degree of freedom of 1, an F-statistic of 7.984433, and a p-value of 0.005176. This indicates that gender has a statistically significant effect on the degree_p variable.

The C(hsc_s) term has a sum of squares of 388.809949, a degree of freedom of 2, an F-statistic of 3.796024, and a p-value of 0.024024. This suggests that the hsc_s variable also has a statistically significant effect on degree_p.

The C(gender):C(hsc_s) term, representing the interaction between gender and hsc_s, has a sum of squares of 39.975847, a degree of freedom of 2, an F-statistic of 0.390292, and a p-value of 0.677352. Since the p-value is relatively high, it suggests that the interaction term is not statistically significant in explaining the variation in degree_p.

The Residual term represents the unexplained variation or error in the model. It has a sum of squares of 10703.471564 and 209 degrees of freedom.

Overall, the ANOVA table provides information on the significance of the categorical variables (gender and hsc_s) and their interaction (gender:hsc_s) in explaining the variation in the degree_p variable.