# ANOVA – ONE INDEPENDENT VARIABLE

## DEFINITION:

ANOVA, which stands for Analysis of Variance, is a statistical test used to analyse the difference between the means of more than two groups.

A one-way ANOVA uses one independent variable, while a two-way ANOVA uses two independent variables.

## When to use a one-way ANOVA?

- It is used when there are more than two groups, and you want to determine whether there is a significant difference between the means of those groups.

One-way ANOVA is appropriate when the following conditions are met:

1. The dependent variable is continuous.
2. The independent variable has three or more levels i.e. groups.
3. The observations are independent and come from normal distributions.
4. Homogeneity of variance: The variances of the dependent variable are equal across all levels of the independent variable.

- If these conditions are met, one-way ANOVA can be used to test whether there is a significant difference between the means of the groups. If the test is significant, post-hoc tests can be used to identify which groups differ significantly from each other.

## Assumptions of one-way ANOVA:

The assumptions of ANOVA include:

- **Independence:** The observations within each group must be independent of each other. This means that the value of one observation should not be related to the value of another observation in the same group.
- **Normality:** The dependent variable should be normally distributed within each group. This means that the distribution of the values should be symmetrical and bell-shaped.

- **Homogeneity of variance:** The variance of the dependent variable should be equal across all levels of the independent variable. This means that the spread of the values should be the same for each group.
- **Random sampling:** The observations in each group should be randomly selected from the population.

## Code for placement.csv sample dataset:

### ANAVO - Analysis of Variance

```
#One independent variable
import scipy.stats as stats
stats.f_oneway(dataset['ssc_p'],dataset['hsc_p'],dataset['degree_p'])
```

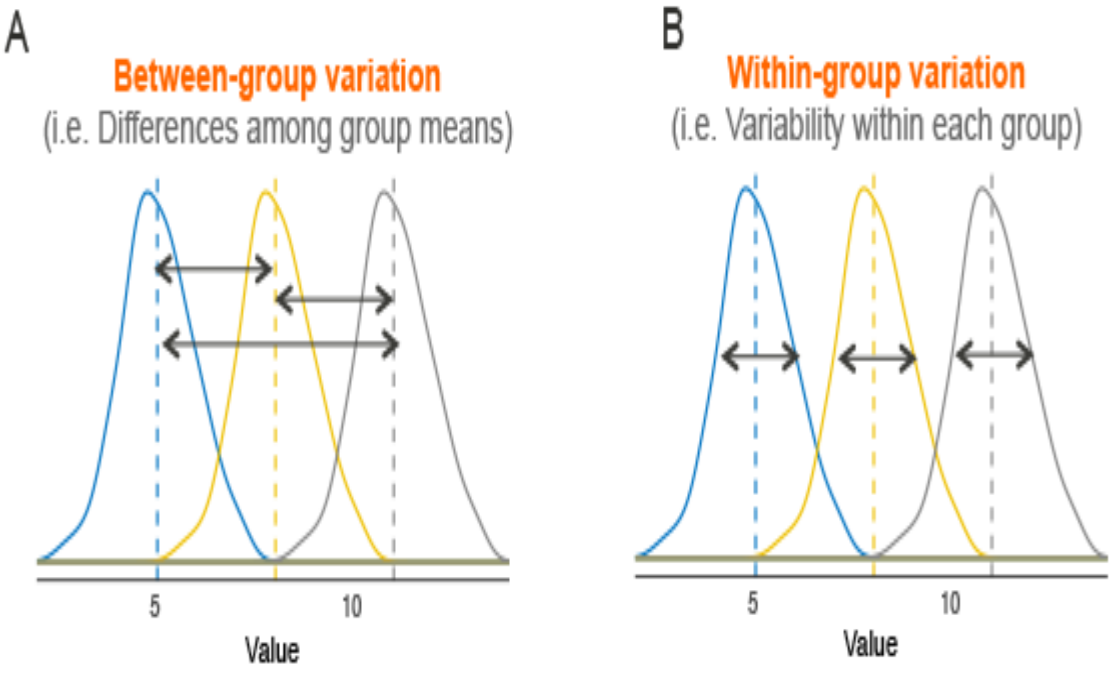
```
F_onewayResult(statistic=0.6719700864663097, pvalue=0.5110602818995302)
```

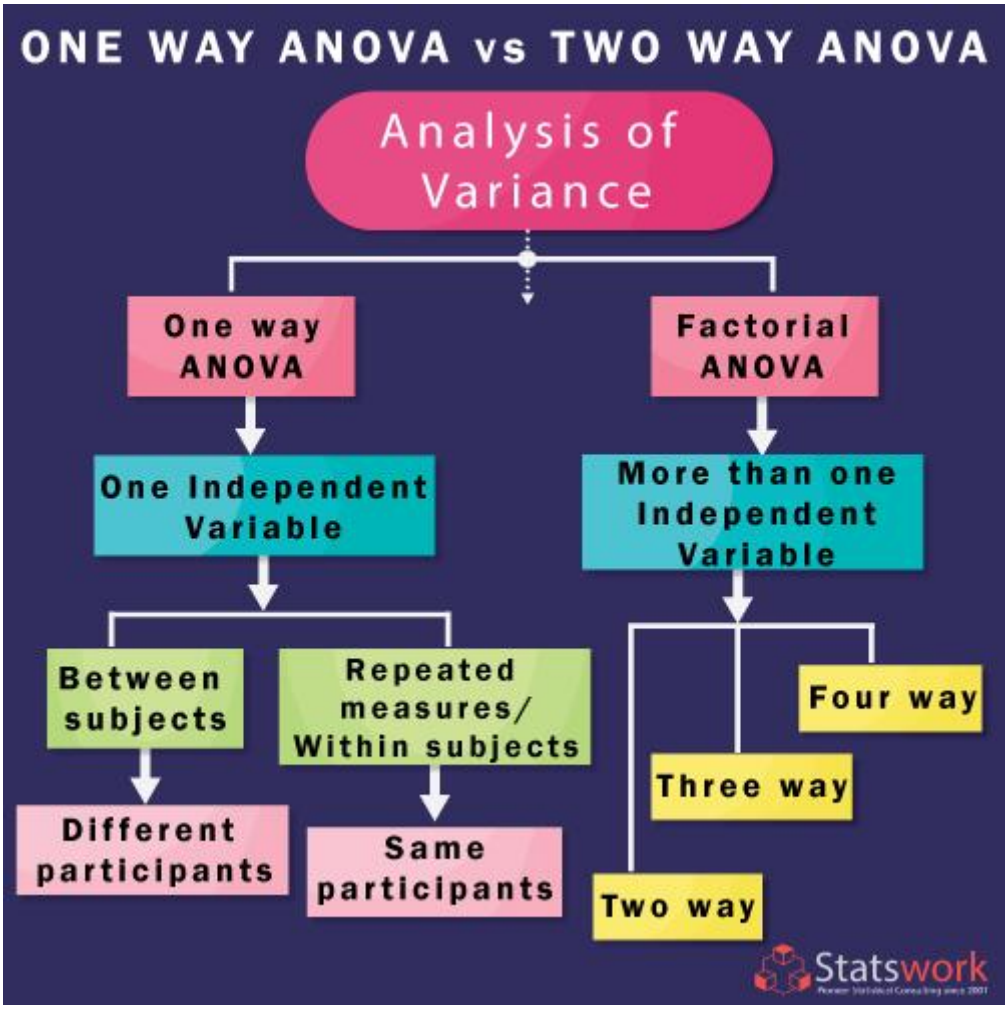
H0- Null Hypothesis - There is no significant difference between these columns H1- Alternate Hypothesis - There is significant difference between these columns

#Reject Null hypothesis when p<0.05 and accept alternate hypothesis #Accept Null hypothesis when p>0.05

```
Accepting Null hypothesis H0 because p>0.05
```

## ONE WAY ANOVA:






# TWO INDEPENDENT VARIABLE

**DEFINITION:**

- ANOVA is a statistical test used to analyse the difference between the means of more than two groups.
- A two-way ANOVA is a statistical method used to analyze the effects of two independent variables on a continuous dependent variable. It is a type of ANOVA that allows us to examine how the interaction between two independent variables affects the dependent variable.
- In a two-way ANOVA, each independent variable can have two or more levels, and the dependent variable must be continuous. The two independent variables are called factors.

## When to use a two-way ANOVA?

- A two-way ANOVA is used when you have two independent variables or factors and you want to examine how each of these variables affects the dependent variable, as well as the interaction between the two factors.
- For example, let us investigating the effect of a new drug on blood pressure, and you want to know if the effect differs between men and women. In this case, you have two independent variables: the drug treatment and gender, and your dependent variable is blood pressure. A two-way ANOVA would allow you to test for main effects of the drug and gender, as well as the interaction between them.

## Assumptions of two-way ANOVA:

**Normality:** The data should be normally distributed within each group and for each combination of the two independent variables.

**Homogeneity of variances:** The variances of the dependent variable should be equal across all groups and for each combination of the two independent variables.

**Independence:** The observations should be independent of each other.

**Random sampling:** The sample should be selected randomly from the population of interest.

**Additivity:** The effect of one independent variable should be independent of the effect of the other independent variable.

**No multicollinearity:** The two independent variables should not be highly correlated with each other.

**Sphericity:** The variances of the differences between all pairs of conditions should be equal.

## Code for placement.csv sample dataset:

```python
#two independent variable

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


# Fit the model
model = ols('salary ~ C(gender) + C(mba_p) + C(gender):C(mba_p)', data=dataset).fit() #ols is ordinary least square
#ols is a function of statsmodels.formula.api is used to fit linear regression model to data using ols estimation.

# Perform two-way ANOVA
anova_table = sm.stats.anova_lm(model, typ=2) # to perform the two-way ANOVA and store the results in anova_table
print(anova_table)
```

```
                       sum_sq     df         F    PR(>F)
C(gender)          8.248839e+08    1.0  0.034660  0.857590
C(mba_p)           5.586650e+12  204.0  1.150684  0.461936
C(gender):C(mba_p) 5.189954e+12  204.0  1.068976  0.518493
Residual           1.665955e+11    7.0       NaN       NaN
```

```
C:\Users\abina\Anac\envs\aiml\lib\site-packages\statsmodels\base\model.py:1873: ValueWarning: covariance of constraints does no
t have full rank. The number of constraints is 204, but rank is 45
  'rank is %d' % (J, J_), ValueWarning)
C:\Users\abina\Anac\envs\aiml\lib\site-packages\statsmodels\base\model.py:1873: ValueWarning: covariance of constraints does no
t have full rank. The number of constraints is 204, but rank is 132
  'rank is %d' % (J, J_), ValueWarning)
```

```
Accepting Null hypothesis H0 because p>0.05 i.e (pR(>F)) for two independent variable interaction
```

## TWO WAY ANOVA: