1. **Covariance** :Covariance is a statistical term that determine relation between two qualitative variables

**Correlation** : Correlation is similar to covariance but difference in measuring units, correlation has specific metric called correlation coefficient to determine relationship between two independent numeric variables. The metric ranges from -1 to +1 indicates strong negative correlation to strong positive correlation.

Values in column A: 25,35,21,67,98,27,64

Values in column B: 52,10,5,98,52,36,69

Mean of Column A :

X1 = sum of values/no.of values

sum= 25+35+21+67+98+27+64

no.of.values = 7

**x1** = 337/7

**=48.1428**

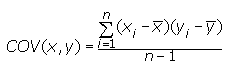
Mean of column B:

X2 = sum of values/no.of values

= 52+10+5+98+52+36+69

=322/7

X2 = **46**

**Covariance** = 

COV(A,B) = (25-48.14)(52-46)+(35-48.14)(10-46)+(21-48.14))(5-46)+(67-48.14)(98-46)+

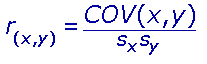
(98-48.14)(52-46)+(27-48.14)(36-46)+(64-48.14)(69-46) /7-1

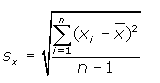
=(-23.14\*6 +(-13.14\*-36)+(-27.14\*-41)+(18.86\*52)+49.86\*6+ (-21.14\*10)+15.86\*13)/6

=-138+473.04+1112.74+980.72+299.16-211.14+206.18/6

= 2722.7/6

=**453.7833**

**Correlation = **

**Sx =**

**=**sqrt (( (-23.14)^2 +(-13.14)^2 +(-27.14)^2+(18.86)^2+(-21.14)^2+(15.86)^2)/7-1))

**=**sqrt((535.4596+172.6596+736.57+355.699+441.24+225.56)/6)

**=**sqrt(411.1)

**=20.27**

**Sy =** sqrt(5986/6)

**=**sqrt(997.6)

**=**31.58

Cor(A,B) = 453.78/(20.27)\*(31.58)

= 453.78/640.12

=**0.7088**

1. **Multicollinearity :**  When two independent variables are highly correlated each other than those variables are said to multi collinear

Multicollinearity can be determined by VIF(Variance Inflation Factor) and by calculating P-values

If multicollinearity exists that means two variable are carry same information, so keep both variables in model is redundant. We are supposed to drop one of the variable in case Multicollinearity exists.

1. Correlation coefficient is a metric used to determine correlation between two independent variables.

Correlation Coefficient Ranges from -1 to +1

-1 ----------🡪 Strong negative correlation between two numerical variables

0-----------🡪 No Correlation

+1---------🡪Strong Correlation

1. **ANOVA :** Analytic variance test is part of hypothesis testing

ANOVA test determines the relation between **Numeric variable** and **categorical variable**

It measures the mean of values based on that we can check whether it leads to NULL Hypothesis or Alternative Hypothesis

1. **Null and alternate hypothesis in chi-square test :**

Chi-square test defines relationship between two Categorical variables

**Null Hypothesis :**  If two variables are independent on each other

**Alternative Hypothesis:**  If two variables are dependent on each other

Example : Consider **temperature** and Interest to **play cricket** variables

If temperature effects interest of play of cricket they it is called Alternative Hypothesis

If temperature doesn’t effects interest of play of cricket then it is called Null Hypothesis