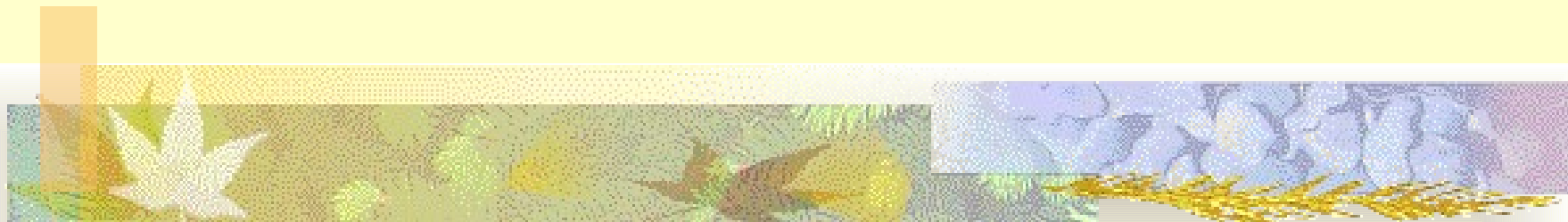# CORRELATION& REGRESSION ANALYSIS

Binod Kumar Singh
Ph.D., Statistics
Department of QT/RM/Operation

Contents

Meaning of Correlation

Types of correlation

Correlation coefficient

Range of correlation coefficient

Interpretation of Correlation Coefficient (r)

Meaning of Regression

Difference between Correlation & Regression

Lines of Regression

Why two lines of regression

Regression coefficient

Correlation analysis Vs Regression analysis

Correlation

- Correlation is a statistical tool that helps to measure and analyse the degree of relationship between two variables.

- Correlation analysis deals with the association between two or more variables.

Cont…

- The measure of correlation called the correlation coefficient .

- The degree of relationship is expressed by coefficient which range from correlation **(-1 ≤ r ≥ +1)**

- The direction of change is indicated by a sign.

- The correlation analysis enable us to have an idea about the degree & direction of the relationship between the two variables under study.

Types of Correlation

Correlation

Positive Correlation

Negative Correlation
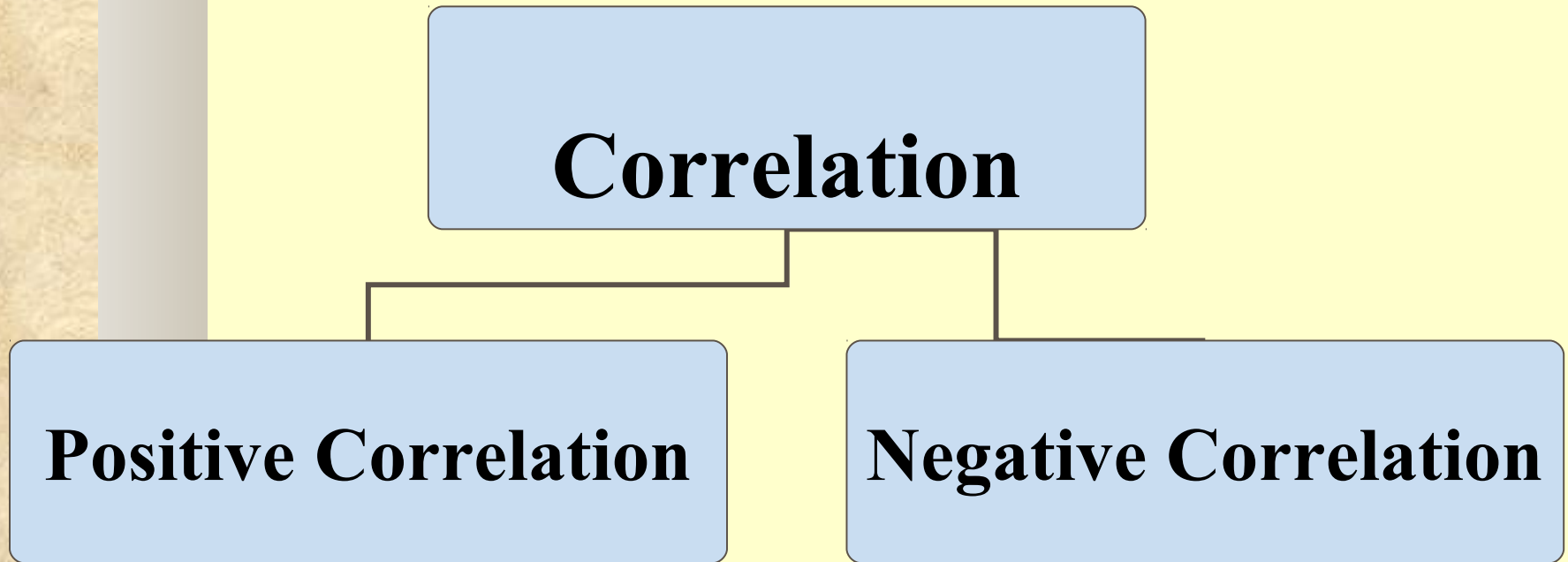
Cont…

- **Positive Correlation:** The correlation is said to be positive correlation if the values of two variables changing with same direction.

  Ex. Pub. Exp. & sales, Height & weight.

- **Negative Correlation:** The correlation is said to be negative correlation when the values of variables change with opposite direction.

  Ex. Price & qty. demanded.

Direction of the Correlation

- **Positive relationship** – Variables change in the same direction.
  - As X is increasing, Y is increasing
  - As X is decreasing, Y is decreasing
  - E.g., As height increases, so does weight.
- **Negative relationship** – Variables change in opposite directions.
  - As X is increasing, Y is decreasing
  - As X is decreasing, Y is increasing
  - E.g., As TV time increases, grades decrease

Indicated by sign; (+) or (-).

More Examples

- **<u>Positive relationships</u>**
- water consumption and temperature.
- study time and grades.

- **<u>Negative relationships</u>:**
- alcohol consumption and driving ability.
- Price & quantity demanded

Karl Pearson's Coefficient of Correlation

- Pearson's 'r' is the most common correlation coefficient.

- Karl Pearson's Coefficient of Correlation denoted by- 'r' The coefficient of correlation 'r' measure the degree of linear relationship between two variables say x & y.

Karl Pearson's Coefficient of Correlation

- Karl Pearson's Coefficient of Correlation denoted by- r

  $$-1 \leq r \leq +1$$

- Degree of Correlation is expressed by a value of Coefficient

- Direction of change is Indicated by sign

(- ve) or (+ ve)

Interpretation of Correlation Coefficient (r)

- The value of correlation coefficient 'r' ranges from -1 to +1

- If r = +1, then the correlation between the two variables is said to be perfect and positive

- If r = -1, then the correlation between the two variables is said to be perfect and negative

- If r = 0, then there exists no correlation between the variables

Regression Analysis

- Regression Analysis is a very powerful tool in the field of statistical analysis in predicting the value of one variable, given the value of another variable, when those variables are related to each other.

Regression Analysis

- Regression Analysis is mathematical measure of average relationship between two or more variables.

- Regression analysis is a statistical tool used in prediction of value of unknown variable from known variable.

Advantages of Regression Analysis

- Regression analysis provides estimates of values of the dependent variables from the values of independent variables.

- Regression analysis also helps to obtain a measure of the error involved in using the regression line as a basis for estimations .

- Regression analysis helps in obtaining a measure of the degree of association or correlation that exists between the two variable.

Regression line

- **Regression line** is the line which gives the best estimate of one variable from the value of any other given variable.

- **The regression line** gives the average relationship between the two variables in mathematical form.

Regression line

- For two variables X and Y, there are always two lines of regression –

- **Regression line of X on Y** : gives the best estimate for the value of X for any specific given values of Y

- **X = a + b Y**   a = X - intercept

                                    b = Slope of the line

                                    X = Dependent variable

                                    Y = Independent variable

Regression line

- For two variables X and Y, there are always two lines of regression –

- **Regression line of Y on X** : gives the best estimate for the value of Y for any specific given values of X

- **Y = a + bx**   a = Y - intercept

-              b = Slope of the line

-              Y = Dependent variable

-              x= Independent variable

Why always two lines of Regression

Properties of the Regression Coefficients

- The coefficient of correlation is geometric mean of the two regression coefficients. $\mathbf{r} = \sqrt{\mathbf{b}_{yx} * \mathbf{b}_{xy}}$

- If b_{yx} is positive than b_{xy} should also be positive & vice versa.

- If one regression coefficient is greater than one the other must be less than one.

- The coefficient of correlation will have the same sign as that our regression coefficient.

- Arithmetic mean of $b_{yx\ \&}\ b_{xy}$ is equal to or greater than coefficient of correlation. $b_{yx +}\ b_{xy/2\ \geq}\ r$

- Regression coefficient are independent of origin but not of scale.

Correlation analysis vs. Regression analysis.

- Regression is the average relationship between two variables

- Correlation need not imply cause & effect relationship between the variables understudy.- R A clearly indicate the cause and effect relation ship between the variables.

- There may be non-sense correlation between two variables.- There is no such thing like non-sense regression.

Cont…

- When the correlation coefficient is zero, the two lines are perpendicular and when it is ±1, then the regression lines are parallel or coincide.