Correlation

Ms. Pallavi V. Ransubhe

Assistant Professor

Vivekanand College, Kolhapur.

- Correlation is a statistical measure that quantifies the relationship between two or more variables.
- It indicates how changes in one variable are associated with changes in another variable.
- Correlation measures the strength and direction of the relationship between variables.
- It is used to understand patterns, make predictions, and inform decision-making in various fields.

Types of Correlation

1. Positive Correlation

- Indicates that as one variable increases (decreases), the other variable also tends to increase (decrease).
- Example: Height and weight, income and expenditure of group of people, etc.

2. Negative Correlation

- Indicates that as one variable increases (decreases), the other variable tends to decrease (increase).
- Example: Supply and price of the commodities, volume and pressure of a perfect gas, etc.

3. Perfect Correlation

- If the change in one variable is followed by a proportional change in the other variable.
- Example: Circumference and diameter of a circle, temperature in Celsius an Fahrenheit

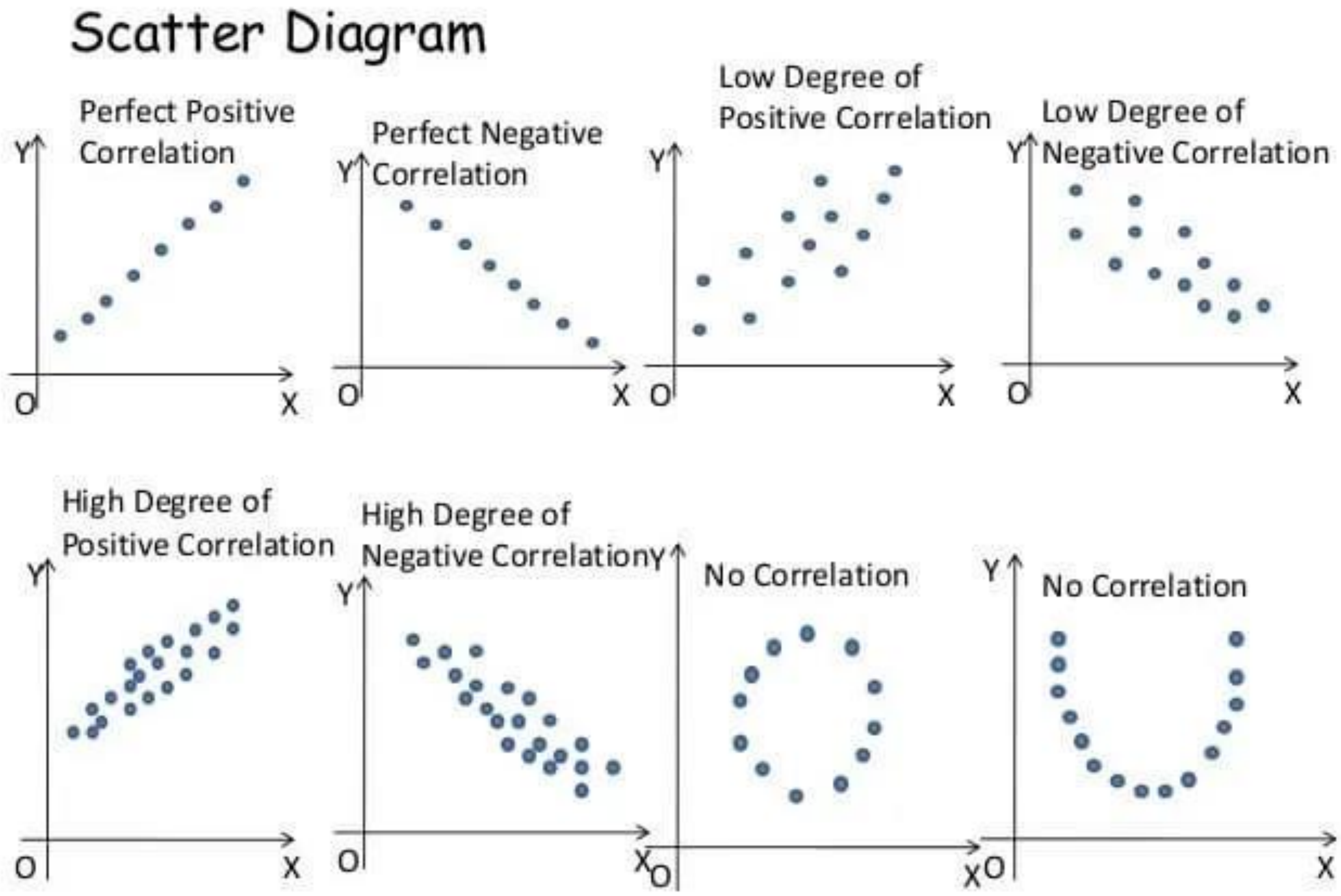
4. Zero Correlation

- Indicates that change in one variable does not affect the other variable.
- Example: Height and study hours, income and IQ level of group of people, etc.

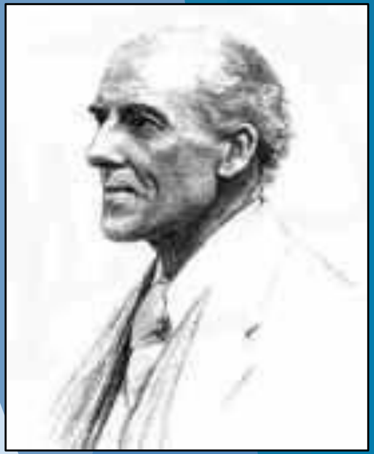
Methods of Studying Correlation

1. **Scatter Diagram**

- Simplest method of studying the nature of correlation between the two variables.
- Scatter diagram is a diagrammatic representation of bivariate data.
- Consider n pairs of values (X_1, Y_1), (X_2, Y_2), ……..., (X_n, Y_n) of two variables X and Y.
- To draw a scatter diagram plot one variable on X-axes and other variable on Y-axes on a graph paper.
- The whole set of points taken together is called Scatter diagram.
- It is also known as dot diagram.


Scatter Diagram
Perfect Positive Correlation
Y
O
X
Perfect Negative Correlation
Y
O
X
Low Degree of Positive Correlation
Y
O
X
Low Degree of Negative Correlation
Y
O
X
High Degree of Positive Correlation
Y
O
X
High Degree of Negative Correlation
Y
O
X
No Correlation
Y
O
X
No Correlation
Y
O
X

2. Karl Pearson’s Correlation Coefficient(r):



- Karl Pearson gave the mathematical formula to measure the correlation between two variables.
- Karl Pearson's coefficient of correlation between two variables X and Y, denoted by r(X,Y) or simply r and is given by,

$$\boxed{r = \frac{Cov(X,Y)}{\sigma_x.\sigma_y}}$$

Where Cov(X,Y) = Covariance between X and Y

$$= \frac{\sum_{i=0}^{n}(Xi-\bar{X})(Yi-\bar{Y})}{n}$$

σ_x = Standard deviation of X

$$= \sqrt{\frac{\sum_{i=0}^{n}(Xi-\bar{X})^2}{n}}$$

σ_y = Standard deviation of Y

$$= \sqrt{\frac{\sum_{i=0}^{n}(Yi-\bar{Y})^2}{n}}$$

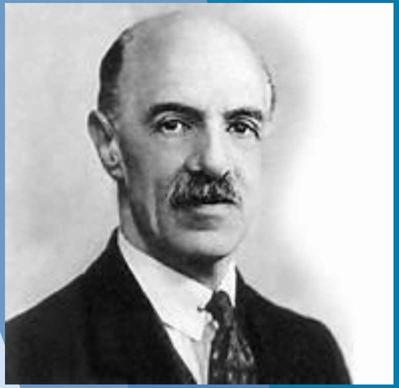
n = Number of pairs of observations

- Karl Pearson’s correlation coefficient lies between -1 and +1 i.e. $-1 < r < 1$

Interpretation of r :

1. If $r = +1$: There is a perfect positive correlation between X and Y. In this case, a scatter diagram will be a straight line rising towards the right.
2. If $r = -1$: There is a perfect negative correlation between X and Y. In this case, a scatter diagram will be a straight line falling towards the right.
3. If $r = 0$: There is no correlation between X and Y.
4. If $r > 0$: There is positive correlation between X and Y. ($0.8 < r < 1$ indicates high degree positive correlation while $0 < r < 0.4$ indicates low degree positive correlation.)
5. If $r < 0$: There is negative correlation between X and Y. ($-1 < r < -0.8$ indicates high degree negative correlation while $-0.4 < r < 0$ indicates low degree negative correlation.)

3. Spearman's Rank Correlation Coefficient (R):



- We Cannot use Karl Pearson's coefficient of correlation in case of qualitative characteristics such as honesty, beauty, morality, etc.
- We can arrange these characteristics serially and use the formula given by British Psychologist Charles Edward Spearman in 1904.
- Which is based on the ranks of n individuals in two characteristic under study thus, the correlation coefficient between the ranks is called Rank Correlation Coefficient.
- It is denoted by R and is given by,

$$\boxed{R = 1 - \left\{ \frac{6 \sum_{i=0}^{n} D^2}{n(n^2-1)} \right\}}$$

where, D = $R_x - R_y$ = Difference between the ranks of the pairs of

observations.

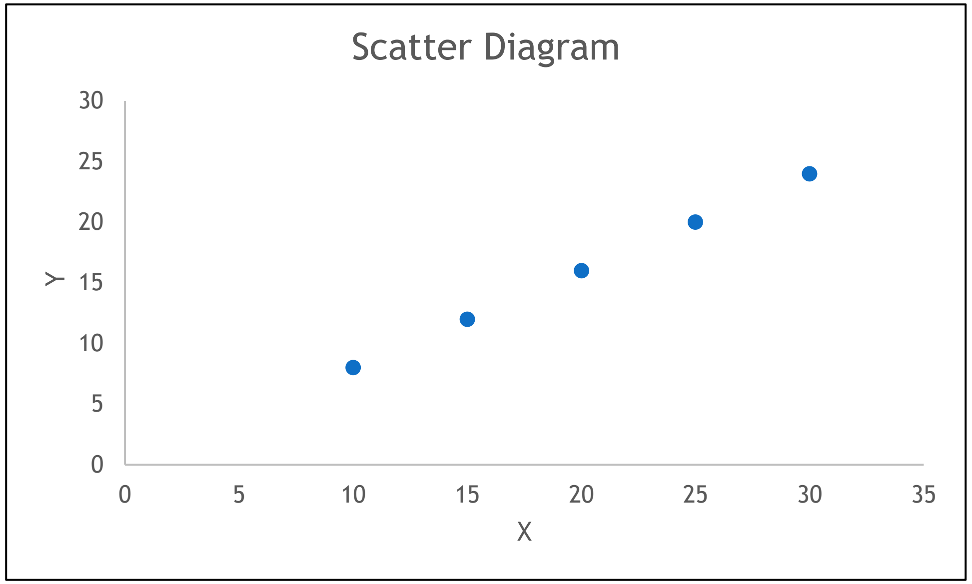
n= Number of pairs of observations

- Spearman's rank correlation coefficient lies between -1 and +1i.e. $-1 < R < 1$
- Interpretation of R is same as that of r.

Q.1. Draw a scatter diagram for the following data and indicate the nature of correlation.

X	10	15	20	25	30
Y	8	12	16	20	24

Ans: Plot the values of X on X-axes and values of Y on Y-axes.



Conclusion: As all the points are lying in the straight line rising towards right, there is perfect positive correlation between X and Y.

Q.2. Find the correlation coefficient between X and Y from the following data.

Sale (X)	17	18	19	19	20	20	21	21	22	23
Profit (Y)	12	16	14	11	15	19	22	16	15	20

Comment on your result.

Ans: Here n=10

- Formulae:

1. $r = \frac{Cov(X,Y)}{\sigma_x.\sigma_y}$

2. Cov(X,Y) = $\frac{\sum_{i=0}^{n}(Xi-\bar{X})(Yi-\bar{Y})}{n}$

3. $\sigma_x = \sqrt{\frac{\sum_{i=0}^{n}(Xi-\bar{X})^2}{n}}$

4. $\sigma_y = \sqrt{\frac{\sum_{i=0}^{n}(Yi-\bar{Y})^2}{n}}$

- **Observation Table:**

X	Y	$(Xi-\bar{X})$	$(Yi-\bar{Y})$	$(Xi-\bar{X})^2$	$(Yi-\bar{Y})^2$	$(Xi-\bar{X})(Yi-\bar{Y})$
17	12	-3	-4	9	16	12
18	16	-2	0	4	0	0
19	14	-1	-2	1	4	2
19	11	-1	-5	1	25	5
20	15	0	-1	0	1	0
20	19	0	3	0	9	0
21	22	1	6	1	36	6
21	16	1	0	1	0	0
22	15	4	-1	4	1	-2
23	20	9	4	9	16	12
200	**160**	-	-	**30**	**108**	**35**

❖ **Calculations:**

1. $\bar{X} = \frac{\sum_{i=0}^{n} Xi}{n} = \frac{200}{10} = 20$

2. $\bar{Y} = \frac{\sum_{i=0}^{n} Yi}{n} = \frac{160}{10} = 16$

3. Cov(X,Y) = $\frac{\sum_{i=0}^{n}(Xi-\bar{X})(Yi-\bar{Y})}{n} = \frac{35}{10} = 3.5$

4. $\sigma_x = \sqrt{\frac{\sum_{i=0}^{n}(Xi-\bar{X})^2}{n}} = \sqrt{\frac{30}{10}} = 1.73$

5. $\sigma_y = \sqrt{\frac{\sum_{i=0}^{n}(Yi-\bar{Y})^2}{n}} = \sqrt{\frac{108}{10}} = 3.28$

6. $r = \frac{Cov(X,Y)}{\sigma_x.\sigma_y} = \frac{3.5}{1.73*3.28} = 0.61$

❖ **Conclusion:** Here, r = 0.61, which indicates there is positive correlation between sales and profit. i.e. sales and profit change in same direction.

Q. 3. The ranks of 8 students given by two judges in a voice test are as follows:

Student No.	1	2	3	4	5	6	7	8
Ranks by 1^{st} Judge	4	1	7	8	2	5	3	6
Ranks by 2^{nd} Judge	7	5	2	3	8	1	4	6

Find rank correlation coefficient and comment on liking of the judge.

Ans:

Given, n=8

❖ **Formulae:**

1. $R = 1 - \left\{ \frac{6 \sum_{i=0}^{n} D^2}{n(n^2-1)} \right\}$

where, D = $R_1 - R_2$

n= Number of pairs of observations

- **Observation Table:**

Student No.	R_1	R_2	D = R_1 - R_2	D^2
1	4	7	-3	9
2	1	5	-4	16
3	7	2	5	25
4	8	3	5	25
5	2	8	-6	36
6	5	1	4	16
7	3	4	-1	1
8	6	6	0	0
-	-	-	-	**128**

- **Calculations:**

1. $R = 1 - \left\{ \frac{6 \sum_{i=0}^{n} D^2}{n(n^2-1)} \right\} = 1 - \frac{6*128}{8(64-1)} = 1 - 1.52 = -0.52$

- **Conclusion:** Here R = -0.52, which indicates negative correlation. i.e. liking of the Judges about voice is not same.