Correlation

Correlation: A **correlation** is a linear relationship between two variables. Correlation measures the linear association between two variables.
Example:

- Is there any relationship between height and shoe size
- Number of cigarettes smoked per day and lung cancer.
- Supply and demand.

Scatter diagram: The basic idea of correlation analysis is to report the association between two variables. A Scatter Diagram is a chart that portrays the relationship between two variables.

X	2	4	6	8	10	12
Y	5	10	8	12	11	14

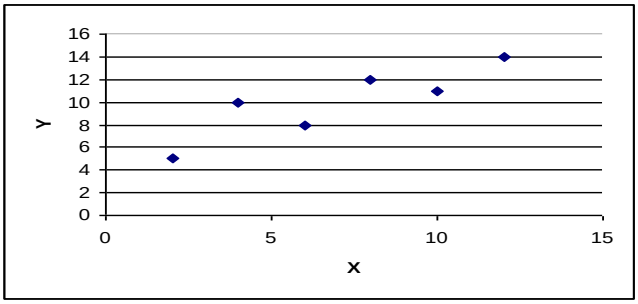


Fig: Scatter plot

Coefficient of correlation: The Coefficient of Correlation (*r*) is a measure of the strength of the linear relationship between two variables.

- **Karl Pearson's coefficient of correlation:**
  We calculate the Karl Pearson's coefficient of correlation from the following formula:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

Properties of Correlation coefficient:

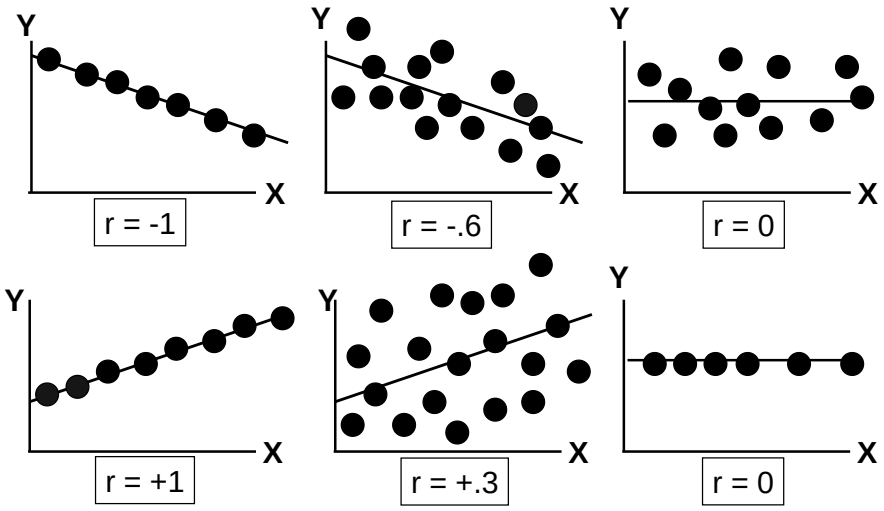
- A correlation coefficient varies from -1 to +1
- -1 indicating a perfect negative relationship (one increase while other decrease),
- 0 indicating no relationship
- +1 indicating a perfect positive relationship.

- The size of the correlation indicates the strength of the relationship; for example, the correlation coefficient -0.89 indicates a stronger relationship than a coefficient of +0.60.
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker any positive linear relationship

Correlation Coefficient Interpretation:

Coefficient Range	Strength of Relationship
0.00 - 0.20	Very Low
0.21 - 0.40	Low
0.41- 0.60	Moderate
0.60 - 0.80	High Moderate
0.81- 1.00	Very High

Scatter Plots of Data with Various Correlation Coefficients



Problem: Mr. Johnson is concerned about the cost to students of textbooks. He believes there is a relationship between the number of pages in the text and the selling price of the book. To provide insight into the problem he selects a sample of eight textbooks currently on sale in the bookstore. Compute the correlation coefficient.

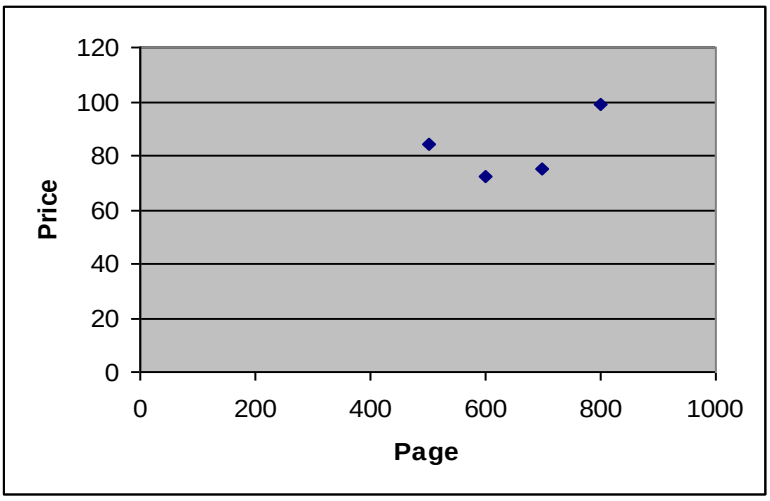
Book	Page(x)	Price($)(y)
Introduction to History	500	84
Basic Algebra	700	75
Business Management	800	99
Introduction to Sociology	600	72

a)Draw Scatter diagram
b) Determine the coefficient of correlation
c) Interpret the result.

Solution:

a) Scatter diagram:

a)



b) Here,
We know

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2600}{4} = 650$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{330}{4} = 82.5$$

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
500	84	-150	1.5	22500	2.25	-225
700	75	50	-7.5	2500	56.25	-375
800	99	150	16.5	22500	272.25	2475
600	72	-50	-10.5	2500	110.25	525
				$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 50000$	$\sum_{i=1}^{n}(y_i - \bar{y})^2 = 441$	$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 2400$

Now,
From the correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$= \frac{2400}{\sqrt{50000 * 441}}$$

$$= 0.511$$

c) **Interpretation:** The correlation between the number of pages and the selling price of the book is 0.511. This indicates a moderate association between the variable.