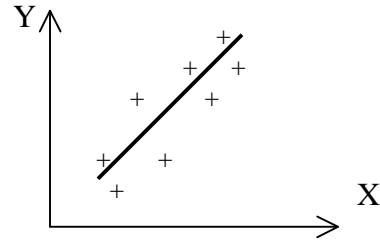


**CHAPTER FIVE: CORRELATION AND REGRESSION****Correlation**

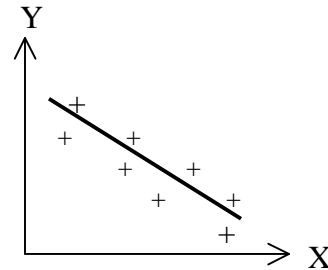
- Correlation analysis is the study of relationship that exists between any two variables, x and y.
- Scatter graph is a graph that shows the degree of correlation between the independent variable, x and dependent variable, y.

**a) Positive Correlation**

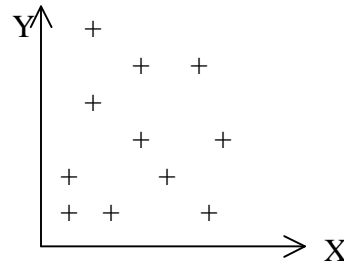
- Indicates that when x increases / decreases, the variable y also increases / decreases.

**b) Negative Correlation**

- Indicates that when x increase / decrease, the variable y decrease / increase.

**c) No Correlation**

- Indicates that there is no direct relationship between the variables x and y.

**Regression Analysis**

$$Y = a + b.x$$

$$b = \frac{n \sum (x.y) - \sum x \sum y}{n \cdot \sum x^2 - (\sum x)^2}$$

$$\text{and } a = \frac{\sum y}{n} - b \frac{\sum x}{n}$$

**Example** The following figures relate to the income and expenditure of six families. Determine the regression line showing the relationship between Income and Expenditure.

Income (\$'000)	Expenditure (\$'000)
15	10
20	15
30	25
40	35
50	45
60	50

x	y	x.y	x <sup>2</sup>
15	10	150	225
20	15	300	400
30	25	750	900
40	35	1400	1600
50	45	2250	2500
60	50	3000	3600
$\Sigma x = 215$	$\Sigma y = 180$	$\Sigma (x.y) = 7850$	$\Sigma x^2 = 9225$

$$b = \frac{n \sum (x.y) - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$= \frac{6(7850) - (215)(180)}{6(9225) - (215)^2}$$

$$= 0.92$$

$$a = \frac{\sum y}{n} - b \frac{\sum x}{n}$$

$$= \frac{180}{6} - 0.92 \left( \frac{215}{6} \right)$$

$$= -2.97$$

The linear equation can now be written as:

$$Y = -2.97 + 0.92X$$

***Learning Outcomes***

- Students should be able to identify the dependent variable (y) and independent variable (x).
- Students should be able to identify the correlation of variables x and y.
- Students should be able to draw scatter diagram.
- Students should be able to calculate the regression line.

**Basic Reading**

1. Saravanan Kullandavelli (1994) LCCI Business Statistics; 5<sup>th</sup> ed. Malaysia; Stamford College Group Publishing.
2. A Francis (1995) Business Mathematics and Statistics; 4<sup>th</sup> ed. London DP Publications Ltd.

**Revision Questions**

1. The following data shows the relationship between the number of hours spent in a week by students to revise and practice for a particular subject and the average marks scored by them in the examination.

Number of hours spent	Marks (%)
2	40
3	43
5	55
8	68
10	77
15	90

- a) Draw a scatter graph for the above data,
  - b) Obtain the least squares regression equation of the number of hours spent against average exam marks.
2. For the following data, obtain the least squares regression equation.

x	20	21	17	12	23	16	24
y	4	8	9	20	13	8	14