CHAPTER 10

LINEAR CORRELATION

10.1 The Concept of Correlation

In some research problems, several variables or characteristics of a population are studied simultaneously to determine whether a relationship exists and if so how close or how significant the relationship might be. Correlation is a statistical tool to measure the association of two or more quantitative variables.

10.2 The Scatterpoint Diagram

To estimate roughly if a relationship exists between two variables, a scatterpoint diagram is made. Draw a straight line intersecting as many points as possible in the graph. If the diagram suggests roughly the existence of a linear relationship then compute for the coefficient of correlation r. The following illustrations show the scatterpoint diagram for different types of correlation

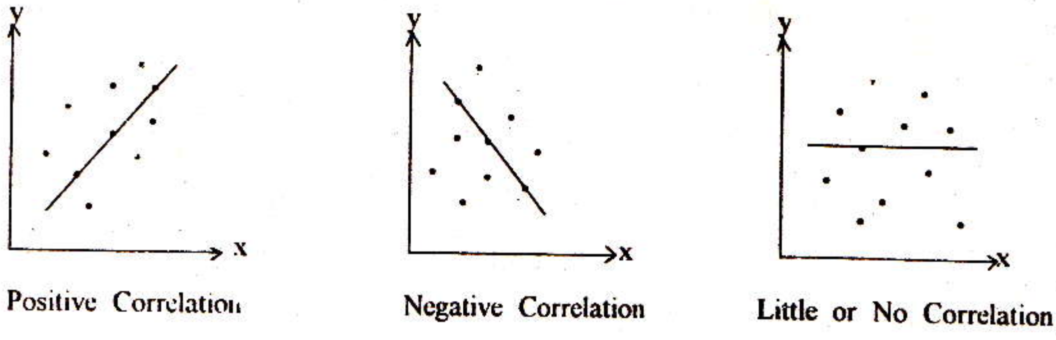


Figure 10.1

10.3 The Pearson Product Moment Correlation Coefficient

The most widely used computational formula for correlation is the **Pearson Prouuct Moment Correlation Coefficient.** In computing for the Pearson's, there are two basic assumptions namely, the presence of linear relationship and the interval or ratio level of measurement of the data. The formula for the Pearson's r is:

$$\boxed{r = \frac{N\,\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}}}$$

where: X = the observed data for the indepentent variable

Y = the observed data for the dependent variable

N = sample size

r = degree of relationship between X and Y

The qualitative interpretation of the degree of linear relationship existing is shown in the following range of values:

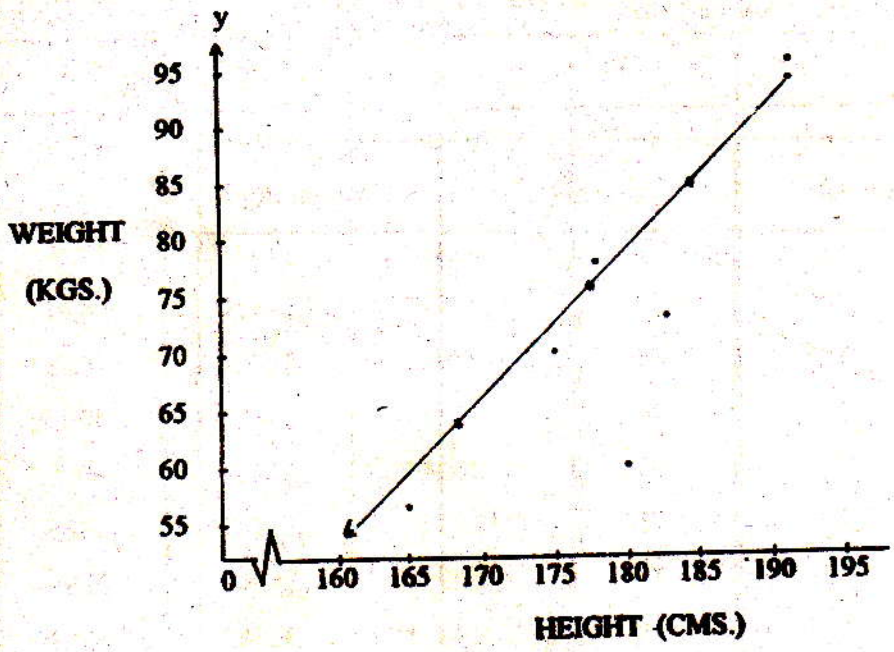
±1.00	perfect positive (negative) correlation
±0.91 - ±0.99	very high positive (negative) correlation
±0.71 - ±0.90	high positive (negative) correlation
±0.51 - ±0.70	moderate positive (negative) correlation
±0.31 - ±0.50	low positive (negative) correlation
±0.01 - ±0.30	negligible positive (negative) correlation
0.00	no correlation

Illustrative Example:

Suppose we want to find out if a relationship exists between the heights and the weights of 15 college students.

Subject	X Height (cm.)	Y Weight (kg.)
1	165	56
2	185	84
3	178	73
4	173	75
5	168	64
6	175	70
7	191	95
8	178	75
9	163	57
10	183	78
11	180	60
12	168	68
13	162	55
14	170	70
15	191	94

The scatterpoint diagram of the given data is as follows:



Interpretation:

The diagram shows a strong positive linear correlation. This means that as the height increases, the weight also increases or vice versa.

For the heights and weight of 15 college students the computation for r is tabulated as follows:

Subject	X	Y	XY	X^2	Y^2
1	165	56	9240	27225	3136
2	185	84	15540	34225	7056
3	178	73	12994	31684	5329
4	173	75	12975	29929	5625
5	168	64	10752	28224	4096
6	175	70	12250	30625	49000
7	191	95	18145	36481	9025
8	178	751	3350	31684	5625
9	163	57	9291	26569	3249
10	183	78	14274	33489	6084
11	180	60	10800	32400	3600
12	168	68	11424	28224	4624
13	162	55	8910	26244	3025
14	170	70	11900	28900	4900
15	191	94	17954	36481	8836
	$\Sigma X = 2630$	$\Sigma Y = 1074$	$\Sigma XY = 189799$	$\Sigma X^2 = 462384$	$\Sigma Y^2 = 79110$

$$r = \frac{N\ (\Sigma XY)\ -\ (\Sigma X)\ (\Sigma Y)}{\sqrt{[N(\Sigma X^2)\ -\ (\Sigma X)^2][N(\Sigma Y^2)\ -\ (\Sigma Y)^2]}}$$

$$r = \frac{15(189799) - (2630)(1074)}{\sqrt{[15(462384) - (2630)^2][15(79110) - (1074)^2]}}$$

$$r = \frac{2846985 - 2824620}{\sqrt{[6935760 - 6916900][1186650 - 1153476]}}$$

$$r = \frac{2846985 - 2824620}{\sqrt{(18860)(33174)}}$$

$$r = \frac{22365}{\sqrt{625661640}}$$

$$r = \frac{22365}{25013.2293}$$

$$r = 0.894$$

Referring to our table of qualitative interpretation of r, we see that the heights and weights of the 15 college students in the sample are highly correlated.

10.4 Other Interpretations of Pearson's r

10.4.1 Coefficient of Determination

The verbal interpretation of r may not be enough to answer the requirements of a certain study. Sometimes, the coefficient of determination r^2 , has to be computed also. This value (r^2) shows the percentage of the variation of the dependent variable y that can be attributed to the variation of the independent variable x. The rest of the variation is due to chance.

Illustrative Example:

In our previous example, our computed r = 0.89. The coefficient of determination r^2 is 0.79. This means that for the 15 college students in the sample 79% of the variation in weights could be attributed to the variation in heights. The rest (21%) is due to chance.

10.4.2 Testing the Significance of r

Although the computed r in the previous example is high (89%), still we cannot be sure if this figure is statistically significant. Therefore a test for the significance of r is needed.

I. H_o : $P = 0$ (There is no relationship between heights and weights of the 15 students)

H_a : $P \neq 0$ (There is a relationship between heights and weights of the 15 college students)

II. Level of significance

$$\alpha = 0.05 \text{ (2 tailed)}$$

$$df = n - 2 = 13$$

$$C.V.\,.t_{.025,\ 13} = \pm 2.16$$

Decision Rule: Reject H_o if t is 2.16 or < -2.16 otherwise accept H_o.

IV. V.