$$m_4 = m_4'(4) - 4m_3'(4)m_1'(4) + 6m_2'(4)m_1'^2(4) - 3m_1'^4(4)$$

$$= 45 - 4\times 10\times 1 + 6\times 4\times 1 - 3\times 1^4 = 45 - 40 + 24 - 3$$

$$= 26$$

Hence $b_1 = \frac{m_3^2}{m_2^3} = 0$, $b_2 = \frac{m_4}{m_2^2} = \frac{26}{3^2} = 2.89$

Example 6.2 The lower quartile and upper quartiles of a distribution are 14.6 and 25.2 respectively and the coefficient of skewness is 0.5. Find the median of the distribution.

Solution: Here we have $Q_1 = 14.6$ and $Q_3 = 25.2$ and $S_k = 0.5$. Now on the basis of quartiles, the coefficient of skewness is $S_k = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$

or, $0.5 = \frac{25.2 - 2Q_2 + 14.6}{25.2 - 14.6} = \frac{39.8 - 2Q}{10.6}$

or, $39.8 - 2Q = 10.6 \times 0.5$ or, $39.8 - 2Q = 5.3$

or, $Q_2 = \frac{39.8 - 5.3}{2} = 17.25$.

Hence the median of the distribution $= 17.25$.

Example 6.3 Find Pearson's coefficient of skewness for a frequency distribution having mean $= 50$, mode $= 56$ and s.d $= 15$.

Solution: Pearson's coefficient of skewness is given by $S_k = \frac{\text{mean} - \text{mode}}{\text{s.d}} = \frac{50-56}{15} = \frac{-6}{15} = -0.4$

7. Bivariate distribution

7.1 Bivariate data : Data on two variables recorded simultaneously for a group of individuals are called bivariate data. Examples of Bivariate data are height and weights of the students in a class, the rainfall and yield of paddy in a state for several consecutive years, the marks obtained by students in the test and final examination, the income and expenditure of a number of families etc.

When we have bivariate data, we can consider the values of each variable separately to know the different measures like the mean and standard deviation of the variable. But here we are mainly concerned with two other problems. Firstly, we want to study the nature and extent of association, if any, between the variables. Secondly if the variables are found to be associated, we express one of them (regarded as dependent variable) as a mathematical function of other (considered as independent variable) so that we can predict the value of the dependent variable, when the value of the independent variable is known. The first problem is called correlation analysis and the second is called regression analysis.

When there are data for a considerably large number of individuals, they are summarised as a two-way frequency table. Suppose we are given n pairs of values of variables x and y. If there be k classes for x and l classes for y, the frequency table will have $k \times l$ cells. With the help of tally marks, we can find frequencies of different cells. The whole set of of class frequencies define the bivariate frequency distribution of variables x and y. The general form of a bivariate frequency distribution is shown in Table 7.1

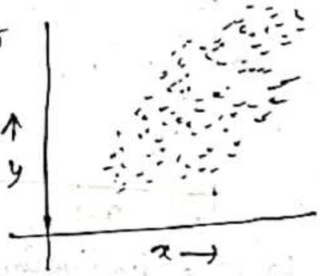
TABLE 7.1

x \ y	y_0-y_1	y_1-y_2	...	$y_{j-1}-y_j$	...	$y_{l-1}-y_l$	Total
x_0-x_1	f_{11}	f_{12}	...	f_{1j}	...	f_{1l}	f_{10}
x_1-x_2	f_{21}	f_{22}	...	f_{2j}	...	f_{2l}	f_{20}
...	...	...	...	...	...	...	...
$x_{i-1}-x_i$	f_{i1}	f_{i2}	...	f_{ij}	...	f_{il}	f_{i0}
...	...	...	...	...	...		...
$x_{k-1}-x_k$	f_{k1}	f_{k2}	...	f_{kj}	...	f_{kl}	f_{k0}
Total	f_{01}	f_{02}	...	f_{0j}	...	f_{0l}	n

Here $f_{i0} = \sum_j f_{ij}$ and $f_{0j} = \sum_i f_{ij}$

7.2 Scatter diagram (or dot diagram)

A scatter diagram is a diagramatic representation of bivariate data. Suppose we are given n pairs of values of variables x and y. Taking two mutually perpendicular straight lines as axes of reference for x and y, each pair of given values can be plotted as a point in the graph paper. The figure obtained, when all the n pair of values have been plotted, is called a scatter diagram (or dot diagram)



From a scatter diagram one can know the nature and the intensity of the association, between the variables under study

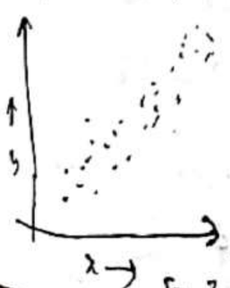


Fig. 7.1(a)

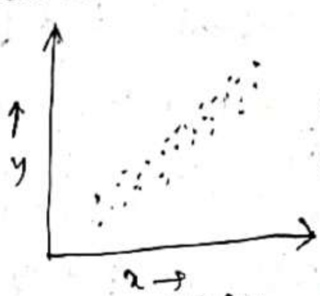


Fig. 7.1(b)

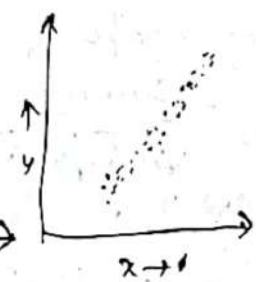


Fig. 7.1(c)

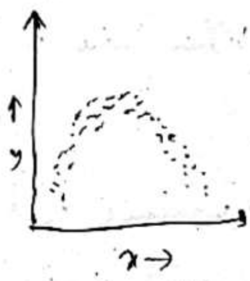


Fig. 7.1(d)

Fig. 7.1(a) to Fig. 7.1(d) are scatter diagrams for four different types of data. For data in first three figures, association between the variables is linear, while it is non-linear for data in the last figure. Again the intensity of linear association in three cases is not the same, it gradually increases as we move over the cases in Fig. 7.1(a), 7.1(b) and 7.1(c).

7.3 Correlation : By Correlation, we mean the association between two variables. If two variables are so related that a change in magnitude of one of them is accompanied by a change in the magnitude of the other, they are said to be correlated. Correlation may be linear or non-linear. Here we restrict ourselves where the correlation is linear. Now, if one variable is found to increase, on the average, as the other increases, the variables are said to be positively correlated. Again, if one variable ~~increases~~ decreases, on the average, as the other increases, they are said to be negatively correlated. A third situation may occur where one variable increases, the other remains constant, on the average. This is the case of zero correlation and variables are said to be uncorrelated.

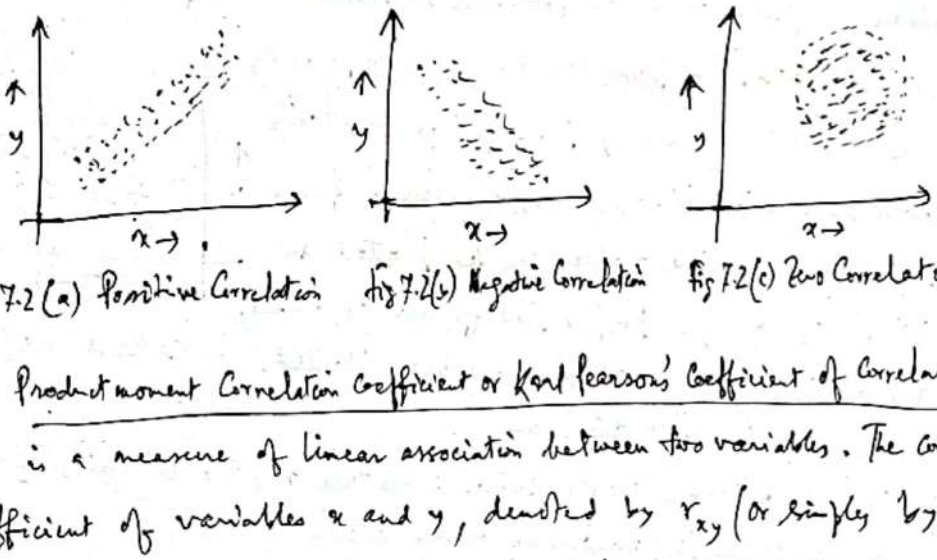


Fig 7.2 (a) Positive Correlation   Fig 7.2(b) Negative Correlation   Fig 7.2(c) Zero Correlation

7.4 Product moment Correlation Coefficient or Karl Pearson's Coefficient of Correlation

It is a measure of linear association between two variables. The correlation coefficient of variables x and y, denoted by r_{xy} (or simply by r when there is no scope of confusion), is defined as

$$r_{xy} = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$ where $Cov(x,y)$ denotes the covariance of x and y. If we are given n pair of values (x_i, y_i), $i = 1, 2, \ldots, n$ of variables x and y.