8. Bivariate Frequecy Distribution

The set of observations on two variables X and Y that are expressed as ordered pairs x1, y1, x2, y2, x3, y3, ... is called **bivariate data**.

Tabulation of Bivariate Data

If we have bivariate data for a large number of observations, then we require to classify this data and prepare a frequency table. Such a table is called bivariate frequency table.

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. It can vary from +1 to -1.

Types of Correlation

The correlation can be of two types:
1) Positive Correlation
2) Negative Correlation

Methods of Studying Correlation:

i) Scatter Diagram (graphical method)
ii) Karl Pearson's coefficient of correlation (algebraic method)

Scatter Diagram

It is a graphical tool showing relationship between two variables.
The values of the two variables are plotted on a graph paper. The manner in which these points are scattered, suggest the degree and the direction of correlation. The degree of correlation is denoted by '*r*' and its direction is given by the signs positive (+) and negative (−).

1) If all points lie on a rising straight line the correlation is perfectly positive and $r = +1$.
2) If all points lie on a falling straight line the correlation is perfectly negative and $r = -1$.
3) If the points lie in a narrow strip, rising upwards, the correlation is high degree of positive.
4) If the points lie in a narrow strip, rising downwards, the correlation is high degree of negative.
5) If the points are spread widely over a broad strip, rising upwards, the correlation is low degree of positive.
6) If the points are spread widely over a broad strip, falling downwards, the correlation is low degree of negative.
7) If the points are spread without any specific pattern, then there is no correlation between the two variables, i.e. $r = 0$.

Covariance:

Covariance is a measure of joint variation between two variables X and Y.

Cov X, Y = 1n∑i = 1nxi yj - x y

Karl Pearson's Coefficient of Correlation

It gives the numerical expression for the measure of correlation. It is denoted by *r* or corr (X, Y). The value of *r*

gives the magnitude of correlation and the sign denotes whether the direction is positive or negative.

It is defined as the ratio of covariance between X and Y to the product of standard deviations of X and Y, i.e. r = corr X, Y = Cov X, YσX· σY

This coefficient measures only the linear relationship between the two variables.

Interpretation of value of *r*, correlation coefficient :

i) If $r = 1$, there is perfect positive correlation between the two variables.
ii) If $r = -1$, there is perfect negative correlation between the two variables.
iii) If $r = 0$, there is no linear relation between the two variables.
iv) If $r > 0$, there is positive correlation between the two variables.
v) If $r < 0$, there is negative correlation between the two variables.
vi) If $r > 0.8$, there is high correlation.
vii) If $0.3 < r < 0.8$, there is moderate correlation.
viii) If $r < 0.3$, the correlation is marginal.