B.A 1st Year PSYCHOLOGY

UNIT 2nd:
CORRELATION AND SIGNIFICANCE

Correlation refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a “scatter plot”. While there are many measures of association for variables which are measured at the ordinal or higher level of measurement, correlation is the most commonly used approach.

Methods of correlation summarize the relationship between two variables in a single number called the correlation coefficient. The correlation coefficient is

usually represented using the symbol r, and it ranges from −1 to +1.A correlation coefficient quite close to 0, but either positive or negative, implies little or no relationship between the two variables. A correlation coefficient close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

A correlation coefficient close to −1 indicates a negative relationship between two variables, with an increase in one of the variables being associated with a decrease in the other variable. A correlation coefficient can be produced for ordinal, interval or ratio level variables, but has little meaning for variables which are measured on a scale which is no more than nominal.

Types of Correlation

The scatter plot explains the correlation between the two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

- Positive Correlation – when the values of the two variables move in the same direction so that an increase/decrease in the value of one variable is followed by an increase/decrease in the value of the other variable.

- Negative Correlation – when the values of the two variables move in the opposite direction so that an increase/

decrease in the value of one variable is followed by decrease/increase in the value of the other variable.

- No Correlation – when there is no linear dependence or no relation between the two variables.

Correlation Formula

Correlation shows the relation between two variables. Correlation coefficient shows the measure of correlation. To compare two datasets, we use the correlation formulas.

Pearson Correlation Coefficient Formula

The most common formula is the Pearson Correlation coefficient used for linear

dependency between the data sets. The value of the coefficient lies between −1 to +1. When the coefficient comes down to zero, then the data is considered as not related. While, if we get the value of +1, then the data are positively correlated, and −1 has a negative correlation.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where n = Quantity of Information

Σx = Total of the First Variable Value

Σy = Total of the Second Variable Value

Σxy = Sum of the Product of first & Second Value

Σx2 = Sum of the Squares of the First

Value

Σy2 = Sum of the Squares of the Second Value

Linear Correlation Coefficient Formula

The formula for the linear correlation coefficient is given by;

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}$$

Sample Correlation Coefficient Formula

The formula is given by:

rxy = Sxy/SxSy

Where Sx and Sy are the sample standard deviations, and Sxy is the sample covariance.