[Summary]

Correlation

Subject:	Business Economics
Course:	B. A. (Hons.), 1st Semester, Undergraduate
Paper No. & Title:	Paper – 102 Statistics for Business Economics
Unit No. & Title:	Unit – 3 Multivariate Analysis
Lecture No. & Title:	Lecture – 1 Correlation

Summary

- A group of information regarding two or more characteristics of an object collected at same point of time is called multivariate-variate data.
- The Direct or indirect cause and effect relationship between two variables of a bi-variate data is called correlation.
- When the correlation between one variable and the linear combination of other variable is to be studied then it is called multiple correlation.
- When the correlation between one variable and other variable is to be studied by ignoring the effect of third variable then it is called partial correlation.
- A numerical measure which shows the degree and direction of the relationship between the correlated variable is called correlation coefficient.
- Karl Pearson's correlation coefficient is a ratio of covariance between two variables to the product of their standard deviations.
- Karl Pearson's method will not give reliable value of correlation coefficient in the case of non-linear relationship between the variables.
- For the qualitative data Spearman's method is used to find correlation coefficient.
- An average of the absolute differences between population correlation coefficient and all possible sample correlation coefficients is called probable error.

Formulae

Some important formulae for Karl Pearson's correlation coefficient

1. $r = \dfrac{\operatorname{cov}(x,y)}{Sx \cdot Sy}$;

2. $r = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\ \sqrt{\sum (y-\bar{y})^2}}$

3. $r = \dfrac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$

4. $r = \dfrac{n\sum uv - (\sum u)(\sum v)}{\sqrt{n\sum u^2 - (\sum u)^2}\ \sqrt{n\sum v^2 - (\sum v)^2}}$; *Where* $u = \dfrac{x-A}{C_x}$, $v = \dfrac{y-B}{C_y}$

For bi-variate frequency distribution the following formula is used.

5. $r = \dfrac{n\sum fuv - (\sum ufx)(\sum vfy)}{\sqrt{n\sum u^2 fx - (\sum ufx)^2}\ \sqrt{n\sum v^2 fy - (\sum vfy)^2}}$

Spearman’s rank correlation coefficient

1. $r = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$ (When ranks are not repeated.)

2. $r = 1 - \dfrac{6\left[\sum d^2 + \frac{m}{12}(m^2-1) + \frac{m}{12}(m^2-1) + ...\right]}{n(n^2-1)}$ (Whenranks are repeated.)

Probable error

$$PE = \frac{0.6745(1-r^2)}{\sqrt{n}}$$

Limits for population correlation coefficient $(r-PE,\ r+PE)$