Assignment 4

Qun1 What is the definition of covariance? Create formula for it.

Ans Covariance is a measure of the relationship b/w two random variables

Calculate upto what extend variables change together

+ve Covariance ⟹ 2 variable move in same direction
-ve Covariance ⟹ " " " " inverse direction

Population $$\left\{ Cov(X, Y) = \frac{\sum (X_i - \bar{X})(Y_j - \bar{Y})}{n} \right\}$$

Sample $$\left\{ Cov(X, Y) = \frac{\sum (X_i - \bar{X})(Y_j - \bar{Y})}{n-1} \right\}$$

correlation

Qun2 What makes ~~covariance~~ better than covariance?

Ans Correlation measures the strength of the relationship b/w variables.

It is scaled measure of covariance. It is dimensionless and can indicate the strength of the relationship, and dependency b/w variable which is not possible for covariance

$$\left\{ P(X, Y) = \frac{Cov(X, Y)}{\sigma_X \, \sigma_Y} \right\}$$

Ques 3 Explain the proven as well as Pearson & Spearman Correlation.

Ans Pearson correlation evaluates the linear Relationship b/w 2 continuous Variable

Spearman correlation evaluates the monotonic relationship. Based on the Ranked values for each variable rather than Raw data

$$(\text{Pearson}) \; \rho(x,y) = \frac{cov(x,y)}{\sigma_x \, \sigma_y}$$

$$(\text{Spearman}) \; R_s = \frac{cov(R_x, R_y)}{\sigma_{R_x} \, \sigma_{R_y}} \qquad \left\{ \begin{array}{l} R_x = Rank\, of\, x \\ R_y = \text{''} \quad \text{''} \quad y \end{array} \right\}$$

Ques4 What are the advantages of spearman over Pearson?

Ans Advantage of Spearman Rank correlation coefficient is that X and Y values can be continuous or ordinal & approximate normal distribution of X&Y is not required.

It can also capture non linear Properties

Ques5 Describe the central Limit theorem?

Ans It states that regardless of the shape of population distribution The distribution of sample means will be approximately normal, &

Given Sample size (n) $\geq$ 30