Statistics

Assignment 4

1. **What is the definition of covariance? Create the formula for it.**
    Ans: Covariance is a statistical tool that is used to determine the relationship between the movements of two random variables. The formula for Covariance differs for the population and the sample.
    Population: $Cov(x,y) = (1/N)*\sum((x-\overline{x})*(y-\overline{y}))$
    Sample: $Cov(x,y) = (1/n-1)*\sum((x-\overline{x})*(y-\overline{y}))$

2. **What makes Correlations better than Covariance?**
    Ans: Correlation values are standardized (lies within -1 and +1) while covariance values are not. Also, change in scale affects covariance, but it doesn't affect correlation in any way.

3. **Explain the process as well as Pearson and Spearman Correlation.**
    Ans: Pearson correlation evaluates the linear relationship between two continuous variables. Spearman correlation evaluates the monotonic relationship. The Spearman correlation coefficient is based on the ranked values for each variable rather than the raw data.

4. **What are the advantages of Spearman Correlation over Pearson Correlation?**
    Ans: Pearson Correlation is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship. The advantage of Spearman Rank Correlation is that the X and Y values can be continuous or ordinal, and approximate normal distributions for X and Y are not required. It is easy to compute. It is easy to understand. It can be computed with any kind of variable, be it independent or dependent.

5. **Describe the Central Limit Theorem.**
    Ans: Central Limit Theorem (CLT) establishes that, that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement text annotation indicator, then the distribution of the sample means will be a normal distribution, even if the original variables themselves are not normally distributed.