Statistics Assignment 4

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

2. What makes Correlations better than Covariance?

3. Explain the process as well as Pearson and Spearman Correlation.

4. What are the advantages of Spearman Correlation over Pearson Correlation?

5. Describe the Central Limit Theorem.

1. Covariance is a measure of the joint variability of two random variables. If the greater values of one variable mainly correspond with the greater value of the other variable and same holds for the lesser values, the covariance is

positive. In the opposite case, when the greater values of one variable mainly correspond to the lesser value of the other, the covariance is negative.

If the random variable pair(x,y) can take on the values (x_i,y_i) for i=1,2,3,.....,n with equal probabilities $P_i=1/n$, then the covariance can be equivalently written as:

$COV(x,y)=1/n\sum(x_i-\mu_x)(y_i-\mu_y)$

2. Both covariance and correlation measure the relationship and the dependency between two variables. Covariance indicates the direction of the linear relationship between variables while correlation measures both the strength and direction of the linear relationship between two variables.

3. Pearson’s correlation coefficient is the test statistics that measures the statistical relationship between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance.

spearman ‘s correlation works by calculating Pearson’s correlation on the ranked values of this data. Ranking from low to high is obtained by assigning a rank of 1 to the lowest value, 2 to the next lowest and so on. If we look at the plot of the ranked data, then we see that they are perfectly linearly related.

4. Pearson’s correlation coefficient measure only linear relationships. spearman ‘s correlation coefficient measures monotonic relationships. So a meaningful relationship can exist even if the correlation coefficient are 0.

5. The central limit theorem states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population distribution. Sample sizes equal to or greater than 30 are often considered sufficient for central limit theorem to hold. In many situation, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.