STATISTICS WORKSHEET 4

Submitted by Darshik A S

1) The central limit theorem tells us that no matter what the distribution of the population is, the distribution of the sample will approach normality as the sample size (N) increases. It helps in predicting the characteristics of a population.

2) Sampling is the selection of a subset of the population to make statistical inference. Types of sampling methods:
   Probability sampling
   - Random sampling
   - Cluster sampling
   - Stratified sampling

   Non-Probability sampling

3) Type 1 error means rejecting the null hypothesis then it's actually true.
   Type 2 error means failing to reject the null hypothesis when it's actually false.

4) Normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations lie around the peak of the curve.

5) Covariance measures the variability between two random variables whereas correlation measures how strongly two variables are related.

6) Univariate summarizes only one variable at a time. Bivariate compares two variables at a time. Multivariate compare more than two variables at a time.

7) Sensitivity measures the proportion of positives that are correctly identified. Sensitivity is true positives divided by total actual positives.

8) Hypothesis testing tests an assumption regarding a population parameter. H0 is the null hypothesis, H1 is the alternate hypothesis. In two tailed tests, H0 is what is currently stated to be true about the population. H1 is always what is being claimed
   .
9) Quantitative data are numeric data variables. Qualitative data are about categorical variables, i.e types.

10) Range = max value - min value, IQR = 3rd quartile value - 1st quartile value

11) Bell curve depicts a normal probability distribution. Symmetrical in shape. Peak corresponds to the mean of the distribution.

12) Removing samples having z score greater than z-score of |3|.

13) P-value is the probability of finding the observed results when the H0 is true.

14) $P(X) = {}_nC_X\, p^x q^{n-x}$

15) ANOVA is a statistical technique that is used to check if two groups are different from each other. It is used to gain information about the relationship between the dependent and independent variables.