STATISTICS WORKSHEET-4

1. The CLT is a statistical theory that states that - if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean. It helps to balance the time and cost of collecting all the data needed to draw conclusions about the population.
2. Sampling is a process in statistical analysis where researchers take a predetermined number of observations from a larger population.
   * Simple random sampling.
   * Systematic sampling.
   * Stratified sampling.
   * Cluster sampling.
   * Convenience sampling.

3. A type I error (false-positive) occurs if a null hypothesis is rejected that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

4.A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

5. Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related

6. Univariiate consists of only one variable. Bivariate depending on two variates. Multivariate analysis is were multiple measurements are made on each experimental unit and where the relationships among multivariate measurements and their structure are important.

7. Sensitivity is the probability of a positive test, conditioned on truly being positive. Sensitivity = TP TP + FN.

8. Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution.H0 is Null Hypothesis and H1 is Alternate hypothesis.

9. Quantitative data are data about numeric variables Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10. The range is calculated by subtracting the lowest value from the highest value. In IQR method of outliers detection we need to find the 25 th percentile which is Q1 and the 75th percentile which is Q3. IQR =Q3-Q1.Lower fence is Q1-IQR. This will give the points lying below -3 SD and Upper fence =Q3+IQr will give points lying above+3 Standard Deviation. These points –Lower fence & upper fence are the outliers under IQR method.

11. A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

12. IQR method of outliers detection we need to find the 25 th percentile which is Q1 and the 75th percentile which is Q3. IQR =Q3-Q1.Lower fence is Q1-IQR. This will give the points lying below -3 SD and Upper fence =Q3+IQr will give points lying above+3 Standard Deviation. These points –Lower fence & upper fence are the outliers under IQR method.

13. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.

15. Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests. ANOVA is helpful for testing three or more variables