STATISTICS WORKSHEET-4

1. What is central limit theorem and why is it important?

   The samples will be normally distributed and the sample mean will be close to the population mean when the high sample size has a finite variance.

2. What is sampling? How many sampling methods do you know?

   Sampling is method of taking records from population. It is method used to get model for population based on Sample. There are two main sampling methods, those are Probability sampling and non-Probability sampling. Then again random sampling, systematic Sampling etc.

3. What is the difference between type1 and typeII error?

   A type I error occurs if an investigator rejects a null hypothesis 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. Although type I and type II errors can never be avoided entirely, the investigator can reduce their likelihood by increasing the sample size (the larger the sample, the lesser is the likelihood that it will differ substantially from the population).

4. What do you understand by the term Normal distribution?

   Normal distribution is Statistical technique, that explains all data follows -+ 1SD, -+ 2SD and -+ 3SD. All data lies withing 3 standard deviation.

5. What is correlation and covariance in statistics?

   Covariance – It represents how two variables are related to each other.
   Correlation – It show how strong they are corelated.

6. Differentiate between univariate ,Biavariate,and multivariate analysis.

   In Univariate – There is only one bell shaped curve in normal distribution.
   In Bivariate – There are two bell shaped curves in Normal distribution curve.
   Multivariate – There are more than two features, on which bell shape is normally distributed.

7. What do you understand by sensitivity and how would you calculate it?

   It is ratio of True Positive values to sum of True positive and True negative values.

8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

   It is method is ensuring that our sample represents whole population. It is majorly used in Research and analysis. This explains, is our

assumption based on previous result is OK or NOT. There are two types, Null and Alternate Hypothesis.

9. What is quantitative data and qualitative data?

Quantitative data is Numerical type data, divided into two sub categories such as continuous and discrete. Does not require label encoding for model building.

Qualitative data is Categorical type of data used to represent variable names. These require label encoding for model building.

10. How to calculate range and interquartile range?

Range = Max-Min

Interquartile range (Q2) = Q3-Q1

Q1 = $1/4(n+1)^{th}$ term

Q3 = $3/4(n+1)^{th}$ term

11. What do you understand by bell curve distribution ?

A bell curve is a graph depicting the normal distribution, which has a shape reminiscent of a bell. The top of the curve shows the mean, mode, and median of the data collected. Its standard deviation depicts the bell curve's relative width around the mean.

12. Mention one method to find outliers.

IQR method or Z score method

13. What is p-value in hypothesis testing?

probability of obtaining the result at least as extreme as the observed result of a statistical hypothesis test, assuming that the null hypothesis is true.

14. What is the Binomial Probability Formula?

$P(r) = {}^{n}C^{r} \cdot p^{r} (1 - p)^{n-r}$.

15. Explain ANOVA and it's applications.

It is similar to variance. But in variance, we measure variance related to two variables only. But in ANOVA, we can measure variance between two or more than two features. determine the influence that independent variables have on the dependent variable in a regression study