STATISTICS WORKSHEET- 4

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

   The central limit theorem states that the sampling distribution of the mean approaches a normal distribution, as the sample size increases.

   The central limit theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size increases.

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

   The method of drawing inferences based on samples of a population is called sampling.

   There are two types of sampling:

   1. Probabilty Sampling
   2. Non Probabilty Sampling

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

   Type 1 error is false positive when we reject the correct null hypothesis, whereas type 2 error is false negative when we reject the true alternate hypothesis,

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

   A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

5. What is correlation and covariance in statistics?

   Correlation is the analysis of how 2 different variables are related.
   If the correlation is 0, that means the 2 variables are not at all related, where as if the value is between 0-+1 then there is positive correlation and if the value is between 0-(-1) then there is negative correlation.

   Covariance is a measure of the relationship between two random variables and to what extent, they change together.

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

   Univariatae Analysis is the analysis of a single variable.

   Biavariate Analysis is the relationship derived between 2 different variables

   Multivariate Analysis is the relationship derived between more than 2 different variables

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

The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as **sensitive analysis**.

The sensitivity is calculated by dividing the percentage change in output by the percentage change in input

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

There are 2 arguments about the value of a population parameter , one in favor and the other against. H0 is the null hypothesis and H1 is the alternae Hypothesis, he null hypothesis is assumed to be true unless there is strong evidence to the contrary.

Both the null and alternative hypothesis should be stated before any statistical test of significance is conducted

Two-tailed Test H0 : $\mu = k$ H1 : μ 6= k P-value = 2P(z > |t|) If P-value   α, we reject H0. If P-value > α, we do not reject H0

9. What is quantitative data and qualitative data?

Quantitative data are measures of values or counts and are expressed as numbers.

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.

Qualitative data are data about categorical variables

10. How to calculate range and interquartile range?

1.Range- The difference Max and min value

2.Inter Quartile Range- The difference between 75th Quartile and 25th Quartile

11. What do you understand by bell curve distribution ?

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. Mention one method to find outliers.

ZSCORE is a powerful method to detect outliers. If the ZSCORE score is beyond +- 3 , then its an outlier

13. What is p-value in hypothesis testing?

The p-value is defined as the 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.

ANOVA is helpful for testing three or more variables.