STATISTICS WORK SHEET-1

Answers for 1 to 9 MCQ'S

1)option –(a) true

2)option-(a) central limit theorm

3)option-(b) modelling bounded count data

4)option-(d) All of above

5)option-(c) poison

6)option-(b) false

7)option-(b)hypothesis

8)option-(a)0

9)option-(c)cannot confirm the regression relationship

10. What do you understand by the term Normal Distribution?

Ans:- It is a probability function used in statistics that tells about how the data values are distributed It is the most important probability distribution function used In statistics because of its advantages in real case scenarios

It is generally observed that data distribution is normal when there is a random collection of data from independent sources, the graph produced after plotting the value of the variable on x-axis and count of the value on y-axis is bell-shaped curve graph

11. How do you handle missing data? What imputation techniques do you recommend?

Ans:- I can use imputation technique, imputation is the process of replacing missing values with substituted data

There are many ways in which we can handle missing data

- **NORMAL IMPUTATIPON**
- **MEAN( )**:- if the data is numerical, we can use mean and median values to replace.
- **MEDIAN( )**:- it is arranging data into lower to higher order and choose the middle value of data.
- **MODE( )**:- is use if the data is categorical, we can use mode which is frequently occurring value.

- KNN-IMPUTER:-KNN imputer will try to find the relation with other columns and impute the data according the relation with other columns.
- Iterative imputer:- iterative imputation refers to a process where each feature is modeld as a function of other features.

12.what is A/B testing?

Ans:- it is a kind of testing that is used to test estimators performance, comparison between estimators.

13. Is mean imputation of missing data acceptable practice?

Ans:- yes, if the data is numerical we can use mean imputation, but if the data is categorical we don't use mean imputation technique, because it produces continuous values.

14. What are the various branches of statistics?

Ans:- There are two branches of statistics.

1) **Descriptive statistics**
2) **Inferential statistics**

1)Descriptive statistics:- descriptive statistics uses data that provides a description of the population either through numerical calculations or graph or table. It provides graphical summery of data. It is simply used for summarizing objects.

There are two categories in Descriptive statistics.

1)Measure of central tendency

a)mean( ):- It is measure of average of all values in sample set

formula=total no of observations /no of observations

b)median( ):- it is measure of central value of sample set, in these data set is ordered from lowest to highest value and then finds exact middle value.

c)mode( ):- it is most frequently occurring value.

2) Measure of variability:-

a)Range:- max – min

b)variance:- it is simply describes how much a random variable defers from expected value and it is also computed as square of deviation

3)Dispersion :- it is measure of dispersion of data from its mean

2)Inferential statistics:- Inferential statistics makes inference and prediction about population based on a sample of data taken from population

Types of inferential statistics:-

- **One sample test difference**
- **Confidence interval**
- **Chi-square statistic**
- **T-test or ANOVA**