STATISTICS WORKSHEET-1

1) Bernoulli random variables take (only) the values 1 and 0.

Ans :- b) False

2) Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans :- a) Central Limit Theorem

3) Which of the following is incorrect with respect to use of Poisson distribution?

Ans :- b) Modeling bounded count data

4) Point out the correct statement.

Ans :- d) All of the mentioned

5) _____ random variables are used to model rates.

Ans :- c) Poisson

6) Usually replacing the standard error by its estimated value does change the CLT.

Ans :- a) True

7) Which of the following testing is concerned with making decisions using data?

Ans :- b) Hypothesis

8) Normalized data are centered at_____and have units equal to standard deviations of the original data.

Ans :- a) 0

9) Which of the following statement is incorrect with respect to outliers?

Ans :- c) Outliers cannot conform to the regression relationship

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

Ans :- Normal Distribution is given example, in this example given a curve represent human height measurement

    For example people height short, average height , and tall any one is between, and then in statistics curve chart Y-axis represent the relative probability of observing , and X-axis are represent is human height group(short, average, tall) someone who is really short or tall and who has an average height they provide data(human height) to giving X- axis criteria to arrange. And provided data(human height) are in statistical curve format spread out data.

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

Ans :-   In statistics missing data handle before using imputation technique , I have to understand the reason why data goes missing.

Basically three main reason are giving

                        1) Missing at random ;- means that the propensity for a data point to be missing is not related to the missing data, but it is related to some of the observed data

                        2) Missing completely at random :- certain value is missing has nothing to do with its hypothetical value and with the value of other variables.

                        3) Missing not at random :- in the other two cases not giving then third case is removed the data .

And the find the reason what are the reason of missing data then basically I am using then given huge amount of data then I am Deletion technique method using.

 And its given data is important then using imputation method because find reason of missing data and imputation method are two types are taken 1) time-series problem :- its using simple statistical method random value imputation , linear interpolation , seasonal adjustment this type of problem can execute.

                        2) General problem :- its using also simple statistical method categorical , Continuous this methos logical and linear regression problem can execute.

And that type of problem giving then I am recommend General problem imputation technique, because its generally missing data related .

12) What is A/B testing?

Ans :- A/B testing is a statistical way of comparing two or more version , such a version A or version B to determine not only which version perform better. But also to understand if a difference between two version is statistically significant.

To understand what A/B testing is about , lets consider two alternative design: A and B. Visitors of a website are randomly served one of the two. Then ,data about their activity is collected by web analytics. Given this data, one can apply statistical tests to determine whether one of the two design has better efficacy.

13) Is mean imputation of missing data acceptable practice?

Ans :- No , Because imputing the mean preserves the mean of the observed data. So if the data are missing completely at random. Mean imputation reduces the variance of the imputed variable. Mean imputation shrinks standard error, which invalidates most hypothesis test and calculation of confidence interval. Mean imputation does not preserve relationships between variables such a correlations .

Mean imputation are not correct value to calculate ,then mostly people not recommended mean imputation . and does not support and acceptable for missing data for practice.

14) What is linear regression in statistics?

Ans :- A Statistical way of measuring the relationship between variables, so what is used linear regression are predict the future !

For example . House price by Square footage 1000sqft = 200rs

Then what is value of 2500sqft = _____? , then his find create a graph. Better understand given mathematical example , $Y = mx + b$

Y is what we are predicting , m is the slope or constant (1000sqft = 200rs)
x is the input(they not given) , b is the bias

Then implement the math $Y = mx + b$

Y = Housing price

M = 200 average

X = Square footage input

B = 0

Plugin the value   Y = mx + b

$$Y = 200x + b$$

$$Y = 200x + 0$$

$$Y = 200 * 1100 + 0$$

$$Y = 22000rs$$

Its simple linear regression

15) <u>What are the various branches of statistics?</u>

Ans :- In statistics have two major branches i) Descriptive Statistics ii) inferential Statistics

- **Descriptive Statistics : are involves the organization , summarization , and display of data mainly showing for data use chart , graph , table and is consider as the first part of statistics analysis , a data set can either represent a sample of a population or the entire population. Descriptive statistics can be categorized into 1. Measures of central tendency 2. Measure of variability**
- **Inferential statistics : are involves using a sample to draw conclusion about a population and inferential statistics based on probability. That enable**

statistics to use the gathered info from a sample to make inference ,

decision or prediction about a given population .