STATISTICS WORKSHEET-1 Assignment

1. . Bernoulli random variables take (only) the values 1 and 0.- True
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=A=Central Limit Theorem
3. Which of the following is incorrect with respect to use of Poisson distribution=B=Modeling bounded count data
4. Point out the correct statement=D=) All of the mentioned
5. _____ random variables are used to model rates=C=) Poisson
6. . Usually replacing the standard error by its estimated value does change the CLT.=false
7. Which of the following testing is concerned with making decisions using data=B=Hypothesis
8. Normalized data are centered at_____and have units equal to standard deviations of the original data=0
9. Which of the following statement is incorrect with respect to outliers=C=Outliers cannot conform to the regression relationship

10. . What do you understand by the term Normal Distribution
    Answer= Normal Distribution is a distribution in which the data points more frequently occurs to mean or center point and it is symmetric about the mean position more frequently as compared to than the farther values from mean position.
    Our aim is to be that about skweness of data is skweness=o ideally by normal distribution we can achiev that it is also called Gaussian distribution or probability bell curve in this curve center is center point is zero and data distributed in positive and negative axix as per given data points.
    If it is observed that skew is not equal to zero that means the data is not normally distributed then it would not given a correct model in machine learning also as data not distributed normally.
    Normal distribution is symmetric about mean, the area under one curve should be equal to 1 also it has only 1 peak value in a curve , it is defined by mean and std deviation
11. How do you handle missing data? What imputation techniques do you recommend?
    Answer= if we want to handle missing data in dataset then we have to either replace it by zeros or by mean or median value because it is very impractical to do analysis using missing data , we want to find out outliers in the data and remove the outliers by 5 way summary and some methods
12. what is A/B testing?
    Answer= A/B testing is used for comparing the both to evaluate the data
13. is mean imputation of missing data acceptable practice= Yes mean is also used to replaced the missing values in dataset.
14. what is linear regression in statistics?

Answer= not known in stats but in ML linear regression is used to identify the independent variable as per dependant variable it also used straight line for best fit these two variables dependant and independent variable.

15. what are the various branches of statistics
    Answer-1. Descriptive statistics2. Inferential statistics,3. Predictive statistics,4. Prescriptive statistics.