STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False
   1.   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

   b) Central Mean Theorem

   c) Centroid Limit Theorem

   d) All of the mentioned

2.   Central Limit Theorem

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

   a) Modeling event/time data

   b) Modeling bounded count data

   c) Modeling contingency tables

   d) All of the mentioned

3. Modelling bounded count data

4.Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution

b) Sums of normally distributed random variables are again normally distributed even if

the variables are dependent

c) The square of a standard normal random variable follows what is called chi-squared distribution

d) All of the mentioned

4.All of the above

5. _____ random variables are used to model rates.

 a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

5.Poisson

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

a) True

b) False

6. False

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

7. Hypothesis

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

the original data.

a) 0 b) 5 c) 1 d) 10
8. 0

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

1. a) Outliers can have varying degrees of influence
2. b) Outliers can be the result of spurious or real processes
3. c) Outliers cannot conform to the regression relationship
4. d) None of the mentioned

9. c- Outliers cannot confirm to the regression relationship

10. What do you understand by the term Normal Distribution?
The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean which is equally in both directions.

The normal distribution describes how the values of a variable are distributed. It is the most important probability distribution in statistics because it fits many natural phenomena. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution. It is also known as the Gaussian distribution and the bell curve.

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

11. When dealing with missing data, we can use two primary methods to solve the error: imputation or the removal of data.

1. **IMPUTATION METHOD:**

The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.

2.Remove Data:

The other option is to remove data. When dealing with data that is missing at random, related data can be deleted to reduce bias. Removing data may not be the best option if there are not enough observations to result in a reliable analysis.

Imputation Techniques-:

Missing at Random (MAR):
Missing at Random means the data is missing relative to the observed data. It is not related to the specific missing values. The data is not missing across all observations but only within sub-samples of the data.

Missing Completely at Random (MCAR):

In the MCAR situation, the data is missing across all observations regardless of the expected value or other variables. Using a t-test, if there is no difference between the two data sets, the data is characterized as MCAR.

Missing Not at Random (MNAR):

The MNAR category applies when the missing data has a structure to it. In other words, there appear to be reasons the data is missing. In a survey, perhaps a specific group of people – say women ages 45 to 55 – did not answer a question.

12.What is A/B testing?
12.A/B testing is that testing in which comparison of two variables of two versions to find out which performs better in controlled environment.

13.Is mean imputation of missing data acceptable practice?
13. Mean imputation of missing data is not good practice , it just preserved the mean of observed data but leads to underestimate of standard deviation as very much big values exists and moreover everytime correlation leads to zero which provides unpredictable model sometimes.

14.What is linear regression in statistics?
14.Linear Regression states the relationship between two variables by fitting a linear equation of observed data. A Linear Regression has an equation $Y=a + bX$ where X is explanatory and Y is dependent variable.

15.What are the various branches of statistics?
15.There are two branches of statistics-

1. Descriptive Statistics:- It describes properties of sample and population data.
2. Inferential Statistics:- It uses those properties to analysis and test the hypothesis on that sample of data.