STATISTICS WORKSHEET-1

Choose the correct option to answer your question.

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

Ans-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

Ans- 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

Ans- Modeling 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

Ans- All of the mentioned

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans-Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

Ans-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

Ans-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

Ans- 0

9. Which of the following statement is incorrect with respect to outliers?
a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned

Ans- Outliers cannot conform to the regression relationship

Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution.
Ans- It is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve. It is also known as Gaussian distribution.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans-We handle missing data by Deletion method, Regression, data Imputation techniques. Average imputation and common point imputation are two important techniques used to handle missing data.

12. What is A/B testing?
Ans-A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?
Ans- Yes, imputing the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimates of the mean remain unbiased.

14.What is linear regression in statistics?

Ans- Linear regression is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable and the variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Ans-There are two branches of Statistics, Descriptive Statistics and Inferential Statistics. Descriptive statistics consist of organizing and summarizing of data. Inferential statistics is a technique where we use the data that we have measured to form conclusions.