STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False
Ans: a) 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: a) 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: b) 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: a) The exponent of a normally distributed random variables follows what is called the log-normal distribution

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

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: b) 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: a) 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: c) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?
Ans: The Normal Distribution is a type of continuous probability distribution for a real-valued random variable. In simple terms, it's a way to represent data using a bell-shaped curve.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans: Missing data can be handled by Dropping/Deleting the missing value type and imputing/substituting the missing value with some other data type.
Few techniques for imputation are: Mean/Median/Mode, Regression, K-Nearest Neighbors.

12. What is A/B testing?
Ans: It refers to a randomized experimentation process where two or more versions of a variable are shown to different segments of website visitors at the same time.

13. Is mean imputation of missing data acceptable practice?
Ans: It is a simple and easy method to apply, but it has many drawbacks that can affect the quality and validity of the data analysis. Mean imputation is not a recommended practice for handling missing data.

14. What is linear regression in statistics?
Ans: linear regression is a method to find the straight line that best fits the data points. It is used to predict the value of one variable based on the value of another variable.

15. What are the various branches of statistics?
Ans: There are 2 branches of statistics as follows:

1. Descriptive Statistics – Mean, Median, Mode, Standard Deviation.
2. Inferential Statistics – Hypothesis Testing, Regression Analysis.