STATISTICS WORKSHEET-1_

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

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)

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)

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 random variables are used
   Ans-d)

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

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

7. 1. 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)

8. 4. 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)

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- a)

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10 Ans) A Normal Distribution, also called a Gaussian distribution, is like a bell-shaped curve. In a normal distribution, the majority of data clusters around the mean, and the spread of values is determined by the standard deviation. There is one 68-95-99.7 rule which says about (about 68%) falls within one standard deviation, even more (95%) within two standard deviation, and almost all (99.7%) within three standard deviation.

11 Ans) To handle missing data, you can delete rows or use imputation techniques. Imputation methods include mean/median imputation, mode imputation for categorical data, regression imputation, and K-nearest neighbors imputation.

12Ans) A/B testing is a method for comparing two versions (A and B) of a webpage or app to determine which performs better based on user responses. It helps make data-driven decisions for optimizing user experience and conversion rates.

13 Ans) Mean imputation for missing data is common but has limitations, potentially introducing bias. Its acceptability depends on context, and alternative methods like multiple imputation might be preferred for robustness.

14 Ans) Linear regression is like drawing the best-fitting straight line through scattered points on a graph. It helps us understand and predict how changes in one thing (independent variable) relate to changes in another (dependent variable)

15 Ans) Statistics has two main branches:
Descriptive Statistics: Involves summarizing and presenting data.
Inferential Statistics: Deals with making predictions or inferences about a population based on a sample of data.