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. a)

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) 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. d) All of the mentioned

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

Ans. c) Poisson

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

Ans. b) False

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) Hypothesis

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. c)

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

10. What do you understand by the term Normal Distribution?

Ans.) In a data set when most of the data points are symmetric to data set mean and in graph its forms shape a bell curve then its normal distribution.

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

Ans.) To handle missing data I have different approach with different data set, sometimes we have to delete data and sometimes we can impute missing value with SimpleImputer using feature (mean, median, mode).

12. What is A/B testing?

Ans.) A/B testing is to compare 2 different version of a variable to check which once is better, in data science we can perform A/B testing on our models and conclude which one is performing well.

13. Is mean imputation of missing data acceptable practice?

Ans.) mean imputation is not acceptable practice in data science it can lead to mislead our model as it doesn't check relationship between feature and label.

14. What is linear regression in statistics?

Ans.) In linear regression we check correlation between dependent and independent variable and we use it to predict an outcome of any given dependent variable on basis of its intercept & slop. Eq-Y=mx+c

15. What are the various branches of statistics?

Ans.) Two main branches of statistics are Descriptive and Inferential.