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: A
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
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: 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: The term Normal means different types of common pattern. Normal Distribution stands for symmetric curve. All the normal curves are bell shaped and have different center and spread. The data near the mean are more frequent.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans: Imputing numerical values using mean/avg. Categorical values can be imputed with frequently occurring values. Deleting rows can be used when huge dataset is available. If more than half of the data is null and column is not an important feature we can eliminate it from dataset.

12. What is A/B testing?
Ans: A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates values.

13. Is mean imputation of missing data acceptable practice?
Ans: Yes, for numeric values like count, age we can easily impute mean values while preventing data loss. There will not be drastic change in prediction.

14. What is linear regression in statistics?
Ans: It is a linear way of modelling dependent and independent variables. Depending on the context, output variables might also be referred to as dependent variables, outcomes, or simply Y variables, and input variables might be referred to as predictors or X variables.

15. What are the various branches of statistics?
Ans: Descriptive statistics and Inferential Statistics. Descriptive statistics describes sample data and Inferential is used to test hypotheses.