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
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
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
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
5. _____ random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   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
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
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

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?

Ans 10

Nornal distribution means a symmetrical data distribution, where most of the results lie near to the mean value.

Ans 11

Deletion method is used to handle the missing data.

There are some techniques which is recommended as mentioned below.
1) Complete case analysis(CCA)
2) Arbitary value imputation
3)Frequent category iimputation

Ans 12

A/B tesing means a randomized controlled environment.
i.e If i have two versions of variable and if i want to find out which one performs better in a controlled environement.

Ans 13

Yes, it is good practise. It saves the data.

Ans 14

It is a predictive analysis. It is used to predict the value of a variable based on the value of another variable.

There are two types of variable.
1) Dependent variable
2) Independent variable

Equation for it.

Y=C+B*X

where, Y= Dependent variable (Predicted result)
x= independent variable(Input data)
c= slope
B= Coefficient (regression value)

Ans 15

There are two branches of statastics
1) Descriptive 2) Inferential