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
   Answer: A

2. Which of the following theorem states that the distribution of averages of id 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
   Answer: 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
   Answer: 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
   Answer: D

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

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

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
   Answer: 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
   Answer: 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?
Answer: Normal distribution is basically a distribution of data which shows that the data is symmetrically distributed about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve

11. How do you handle missing data? What imputation techniques do you recommend?
Answer: We can use the mean value method or we can drop the missing data. I believe changing the missing values to the mean value is the best way as data is important and we can't drop the data as this might hamper the output.

12. What is A/B testing?
Answer: A/B testing is type of testing use to compare two versions and identify which is working better.

13. Is mean imputation of missing data acceptable practice?
Answer: Yes, as this replaces the missing values in a data set with the mean of that particular column and helps the model to perform more accurately and give a good result.

14. What is linear regression in statistics?
Answer: Linear regression is a approach to identify the relation between the input or independent variables and the labels or output variables. It shows that how the two variables are related linearly.

15. What are the various branches of statistics?
Answer: The various branches of statistics are: Data Collection, Checking the description of the data and analyzing the data, making assumptions and fitting the data for various models.