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
Solution: a)True

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
Solution: a)Central Limit Theorem

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
Solution: a) Modeling event/time 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
Solution: d) All of the mentioned

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

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Solution: 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
Solution: 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
Solution: a) 0

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
Solution: c) Outliers cannot conform to the regression relationship

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

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

Sol: A distribution wherein the data is distributed in such a way that it is symmetric about the mean forming a bell shaped curve when plotted is called as normal distribution. Thus data near mean is occurring more frequently than away from the mean in such distribution.

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

Solution. First we check if there are any missing values in rows and columns. If there are missing values, we replace them by either mean, median or mode depending on the type of data. If more than half of the row has missing data, we delete entire row. The algorithms such as k-nearest neighbors, Naïve Bayes and RandomForest can be used as imputation technique for handling missing data.

12. What is A/B testing?

Sol: It is a method used to compare two versions of a variable by recording the responses of two versions and finding out which is more effective than the other. Examples include email marketing, driving political campaigns etc.

14. What is linear regression in statistics?

Solution: It is a predictive analysis technique to determine how good predictor variables are in predicting the outcome and who amongst them have higher impact on the outcome than the others. Using it, we can provide future estimates based on given input

15. What are the various branches of statistics?

Solution: Statistics is mainly divided into two parts

1. Descriptive- Organization, Analysis, Summarization and display of data
2. Inferential-Using sample to draw conclusions on population