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.

Answer: - 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?

Answer: - a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Answer:- b) Modeling bounded count data

4. Point out the correct statement.

Answer: -d) All of the mentioned

5. _____ random variables are used to model rates.

Answer: - c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

Answer: - b) False

7. 1. Which of the following testing is concerned with making decisions using data?

Answer:-b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

Answer: -a) 0

9. Which of the following statement is incorrect with respect to outliers?

Answer: -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?**

Answer: -Normal Distribution is also called as Gaussian Distribution. It is an arrangement of data set in which most value cluster in the middle of range and the rest tapers off symmetrically towards either extreme. If we plot this with the help of histogram, it will form bell curve. In normal distribution mean, median and mode are same.

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

Answer: - Handling missing is very important, if data set is small, I use mean, median and mode method and if data sets are big, we can drop the records entirely. Imputation techniques like mean, median and mode are helpful when dealing with less observations. Deleting the records and making a model for missing values.

12. **What is A/B testing?**

Answer: -A/B testing is also known as split testing. It is version of comparing two webpage, ad campaign or app against each other to determine which one performs well.

e.g.:-for social media ad campaign, budget is around 30k, so deploy 5k for each ad initially .so first we make separate two campaign of different design and layout and will run both ads, so through we can figure out which ad campaign is doing best

13. **Is mean imputation of missing data acceptable practice?**

Answer: -No mean imputation is not acceptable practice as it ignores the features correlation and data variance ais also reduced. Due to this, our model accuracy will be reduced.

14. **What is linear regression in statistics?**

Answer: -Linear regression is commonly used type predictive analysis. It attempts to model the relationship between two variables by fitting linear equation to observed data.

Linear equation Y + a + bX, where X=explanatory variable, Y=dependent variable, a=intercept, y=slope of line

15. **What are the various branches of statistics**

Answer: -Main branches of statistics are Inferential and Descriptive statistics. In inferential, we try to reach to a conclusion using data. In descriptive, using data we describe what's going on in our data