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

Ans. (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?

Ans. a) Central Limit Theorem

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

Ans. b) Modelling bounded count data

4. Point out the correct statement.

Ans. d) All of the mentioned

5. _____ random variables are used to model rates.

Ans. c) Poisson

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

Ans. b) False

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

Ans. b) Hypothesis

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

Ans. a) 0

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

Ans. 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?

Ans. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric 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.

KEY TAKEAWAYS:

- A normal distribution is the proper term for a probability bell curve.

- In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
- Normal distributions are symmetrical, but not all symmetrical distributions are normal.
- In reality, most pricing distributions are not perfectly normal.

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

Ans Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields

  Use regression analysis to systematically eliminate data. ...

  Data scientists can use data imputation techniques.

12. What is A/B testing?

Ans. A/B testing is **a** basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. For instance, let's say you own a company and want to increase the sales of your product.

13. Is mean imputation of missing data acceptable practice?

Ans. The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

14. What is linear regression in statistics?

Ans. Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modelling.

15. What are the various branches of statistics?

Ans. Two branches, descriptive statistics and inferential statistics, comprise the field of statistics