

WORKSHEET 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

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

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

b) Modeling bounded count data

4. Point out the correct statement.

d) All of the mentioned

5. _____ random variables are used to model rates.

c) Poisson

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

b) False

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

b) Hypothesis

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

a) 0

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

c) Outliers cannot conform to the regression relationship

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans- A normal distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency.

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

Ans- the best way to handle missing data is imputation.

the most advanced method for performing missing data imputation is multiple imputation. In multiple imputation we generate missing values from the dataset many times.

12. What is A/B testing?

Ans- A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Ans- It is acceptable when the missing value proportion is not large enough. But, when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been.

14. What is linear regression in statistics?

Ans- Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables. Linear regression looks at various data points and plots a trend line. Linear regression can create a predictive model on apparently random data, showing trends in data, such as in cancer diagnoses or in stock prices.

15. What are the various branches of statistics?

Ans- There are two branches in statistics

- 1- Descriptive statistics- **Descriptive statistics** summarize and organize characteristics of a data set. A data set is a collection of responses or observations from a sample or entire population.
- 2- Inferential statistics- **Inferential statistics** help you come to conclusions and make predictions based on your data .When you have collected data from a sample, you can use inferential statistics to understand the larger population from which the sample is taken.