1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Ans. The correct answer is **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

Ans. The correct answer is **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

Ans. The correct answer is **b) Modeling bounded count 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

Ans. The correct answer is **a) The exponent of a normally distributed random variables follows what is called the log- normal distribution**

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans. The correct answer is **c) Poisson**

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

a) True

b) False

Ans. The correct answer is **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

Ans. The correct answer is **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

Ans. The correct answer is **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

Ans. The correct answer is **c) Outliers cannot conform to the regression relationship**

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

Ans. Normal Distribution, also known as Gaussian Distribution, is a type of continuous probability distribution for a real-valued random variable. It is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a “bell curve” and is sometimes referred to as the bell curve. A large number of random variables are either nearly or exactly represented by the normal distribution, in every physical science and economics

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

Ans. Missing data is a common problem in data analysis and can arise for many reasons, such as non-response, data entry errors, or data corruption. There are several methods for handling missing data, including deletion, imputation, and model-based methods.

Imputation is a popular method for handling missing data. It involves replacing missing values with estimated values based on the observed data. Some common imputation techniques include:

Mean imputation: This method involves replacing missing values with the mean of the observed values for that variable.

Regression imputation: This method involves using a regression model to predict missing values based on the observed data.

Multiple imputation: This method involves creating multiple imputed datasets and then combining the results to obtain final estimates.

The choice of imputation technique depends on the nature of the missing data and the goals of the analysis. It is important to carefully evaluate the assumptions and limitations of each method before choosing an appropriate technique.

12. What is A/B testing?

Ans. A/B testing, also known as split testing or bucket testing, is a randomized experimentation process wherein two or more versions of a variable (such as a web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics. It is a way to compare multiple versions of a single variable, for example by testing a subject's response to variant A against variant B, and determining which of the variants is more effective

13. Is mean imputation of missing data acceptable practice?

Ans. Mean imputation is a simple method for handling missing data, where the missing values are replaced with the mean of the observed values for that variable. While it is a commonly used technique, it has some limitations and is not always considered an acceptable practice.

One of the main issues with mean imputation is that it can lead to biased estimates and distorted relationships between variables. By replacing missing values with the mean, we are essentially assuming that the missing data is missing completely at random (MCAR), which may not always be the case. If the data is not MCAR, then mean imputation can introduce bias into the analysis.

Another issue with mean imputation is that it can reduce the variability in the data, leading to underestimated standard errors and inflated test statistics. This can result in incorrect conclusions and misleading results.

In summary, while mean imputation can be a useful technique in some situations, it is not always considered an acceptable practice and should be used with caution. It is important to carefully evaluate the assumptions and limitations of this method before using it to handle missing data

14. What is linear regression in statistics?

Ans. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The simplest form of linear regression involves two variables: one independent variable and one dependent variable. The goal of linear regression is to find the line of best fit that can accurately predict the value of the dependent variable based on the value of the independent variable. This is done by estimating the coefficients of the linear equation that minimizes the discrepancies between the predicted and actual values of the dependent variable. Linear regression can be used for both simple and multiple regression analysis, depending on whether there is only one independent variable or more than one.

15. What are the various branches of statistics?

Ans. Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. There are two main branches of statistics: descriptive statistics and inferential statistics.

Descriptive statistics deals with the collection, organization, summarization, and presentation of data. It provides simple summaries about the sample and the measures. Descriptive statistics are used to describe the basic features of the data in a study.

Inferential statistics, on the other hand, is used to make inferences and predictions about a population based on a sample of data taken from the population in question. It is used to test hypotheses and make estimations using sample data.

Both branches of statistics follow a particular scientific approach that makes them equally important to every statistical student