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

a) True
b) False

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?
a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned

Ans: **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: **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: **d)All of the mentioned**

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

Ans: **c) Poisson**

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

Ans: **b) False**

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

b) Hypothesis
c) Causal
d) None of the mentioned

Ans: **b) Hypothesis**

8. 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: **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: **c) Outliers cannot conform to the regression relationship**

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

Ans: Normal distribution is data is distributed symetrically across mean, without any skewness. In this, mean and median are equal, so the graph is symetrical bell shaped curve.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans: There are various techniques we follow to handle missing data,

a) Drop the missing values in dataframe - drop the missing values using df.dropna method
b) Mean imputation - Calculate the mean of the column data and replace the missing value with mean value.
c) Hot Deck Imputation - Substitute the missing value with other comparable values. If value expected is within some range, the value can be subsituted from value within same range.
d) Regression Imputation - The missing variable is substitued by a predicted value, based other factors

12. What is A/B testing?
Ans: A/B testing on the new and old models can be an effective final step in validating a new model. It is used for improvement in data models. A/B testing is a strategy for determining how a change in one variable impacts audience or user engagement.

13.Is mean imputation of missing data acceptable practice?
Ans: Mean imputation of missing data is one of the imputation technique to substitute the missing data values. In this technique, the mean is calculated for a column data and the mean value is subsituted to any of the missing values, this ensures the value substitued is average of all values and it would satisfy the condition of normal distribution.

14.What is linear regression in statistics?

Ans: Linear regression is the model that estimates the relation between one independent and one dependent variable to be a straight line. It is predicting the value of one variable based on the value of other variable. The variable which we want to predict is called a dependent variable.

15.What are the various branches of statistics?
Ans: There are two main types of branches of statistics

a) Descriptive Statistics - It summarize data from a sample using indexes such as the mean, mode, standard deviation, variance, count etc. Though, it is not used widely for decision making, it still has importance of providing critical summarized information of various statistical parameters.

b) Inferential Statistics - This type of statistics is used to interpret the meaning of descriptive statistics data. once the data has been collected, analysed and summarised then, this data is used to draw conclusions.