1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False
Answer-a

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
Answer-a

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
Answer-b

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
Answer-d

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

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

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
Answer-b

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
Answer-a

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
Answer-c

10. What do you understand by the term Normal Distribution?
Answer-A normal distribution is the continuous probability distribution with a probability density function that gives you a symmetrical bell curve. In simple words, it is a plot of the probability function of a variable that has maximum data concentrated around one point and a few points taper off symmetrically towards two opposite ends.The data near the mean are more frequent in occurence than data far from the mean.

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

Answer- Missing data (NAN Values).This missing data will impact on the prediction of model. In order to Handle missing value's we should use mean for continues data and mode for Discrete data. But not all the time mean or mode method is recommended. If data set is small we can always use FillNa method, But while working on BigData set's we should use following Imputers:-

- Knn Imputer
- Iterative Imputer
- Mean Imputation

12. What is A/B testing?
Answer- A/B testing is also known as split testing, we split data in two groups, and show two different versions of data sets with the goal of comparing the results to find the more successful version.

13. Is mean imputation of missing data acceptable practice?

Answer- The process of replacing the NAN (NULL) values in data set with the data's mean is known as mean imputation. Typically Mean Imputation is considered a terrible pracitice as it ignores the features correlation. This method also decreases the variance of out data while increasing bias.As a result of reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Answer- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

15. What are the various branches of statistics?

Answer- The main two branches of statistics are as follows:-

1. Inferential Statistics.

2. Descriptive Statistics.