statistics

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) Modeling 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:- a) true
7) . Which of the following testing is concerned with making decisions using data?
   Ans:- b) Hypothesis
8) Normalized data are centered 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

10) What do you understand by the term Normal Distribution?
    Ans:- normal distribution will appear as a bell curve In graph form, is a probability distribution that is symmetric about the mean**.** Normal distributions are symmetric, unimodal, and asymptotic, and the mean, median, and mode are all equal. A normal distribution is perfectly symmetrical around its center. That is, the right side of the center is a mirror image of the left.

11) How do you handle missing data? What imputation techniques do you recommend?
    Ans:- we can handle the missing data by applying the techniques , Deletions. Pairwise Deletion. Listwise Deletion/ Dropping rows. Dropping complete columns.
    K-Nearest Neighbor Imputation
    Use regression analysis to systematically eliminate data.

12) What is A/B testing?
    Ans:- A/B resting also known split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element etc.)

13) Is mean imputation of missing data acceptable practice?
    Ans:- False**!** imputation has gotten a bad rap because early imputation methods, like mean imputation, bias your results pretty badly. And single imputation underestimates standard errors.

14) What is linear regression in statistics?

Ans:- Linear regression is **an attempt to model the relationship between two variables by fitting a linear equation to observed data**, where one variable is considered to be an explanatory variable and the other as a dependent variable. 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.

15) What are the various branches of statistics?

Ans:- The field of statistics is divided into two major divisions:

1. **Descriptive**
2. **inferential**.

Descriptive statistics: -- descriptive statistics describe what is going on in a population or data set. Descriptive statistics mostly focus on the central tendency, variability, and distribution of sample data. Central tendency means the estimate of the characteristics, a typical element of a sample or population, and includes descriptive statistics such as mean , median, and mode. Variability refers to a set of statistics that show how much difference there is among the elements of a sample or population along the characteristics measured, and includes metrics such as range, variance, and standard deviation.

Inferential statistics: -- Inferential statistics by contrast, allow scientists to take findings from a sample group and generalize them to a larger population. In general we can say its large data set and we can take a sample and find the average. Inferential statistics are used to make generalizations about large groups, such as estimating average demand for a product by surveying a sample of consumers' buying habits or to attempt to predict future events, such as projecting the future return of a security or asset class based on returns in a sample period

For example:--we have a 10000 student data and they are study in different different class and find out the average marks so we can take random marks from every class and divided by total no of class.