# Statstics

Q1- Bernoulli random variables take (only) the values 1 and 0.
**Answer-True**

**Q2-** 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?
**Answer- a) Central Limit Theorem**

**Q3-** Which of the following is incorrect with respect to use of Poisson distribution?
**Answer- b ) Modeling bounded count data**

Q4- . Point out the correct statement.
**Answer- d) All of the mentioned**

**Q5-_____** random variables are used to model rate
**Answer- C) Poisson**

**Q6-** Usually replacing the standard error by its estimated value does change the CLT.
**Answer**- b**) False**

**Q7-** Which of the following testing is concerned with making decisions using data?
**Answer- b) Hypothesis**

Q8- Normalized data are centered at_____and have units equal to standard deviations of the original data.
**Answer**- **a) 0**

Q9- Which of the following statement is incorrect with respect to outliers
**Answer- Outliers cannot conform to the regression relationship**

Q10-What do you understand by the term Normal Distribution?
**Answer- Normal Distribution is a proper term for a probability Bell Curve. In a Normal Distribution the mean is 0 and the Standard Deviation is 1.**

Q11-How do you handle missing data? What imputation techniques do you recommend?
**Answer**-We use some imputation Techniques for handling missing data.There are Several Imputation techniques
**Simple Impute**- which the mean of the observed values for each variable is computed and the missing values for that variable are imputed by this mean. X`
**KNNImputer-**Will try to find the relationship with the other columns and impute the data according to relationship.
**Iterative Imputer**-
This methd treat other columns(which doesn't have Null) as features and train on them and Treat Null Columns as label. Finding it will predict the NaN data & impute it's just like Regression problem.

Q12. What is A/B testing?
**Answer-** It is hypothesis testing for two variables A and B .
An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

Q13. Is mean imputation of missing data acceptable practice?
**Answer- No** , its not a good Practice because this method can lead into severely biased estimates and our model will not predict properly

Q14-What is linear regression in statistics?

**Answer- In** statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression

Q15- What are the various branches of statistics?

**Answer- 2 Types of Statistics**

**Descriptive-** When you are able to describe the data OR population is small , it is called Descriptive Statistics.In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or standard deviation.

**Inferential-** This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.