

Q1. Bernoulli random variables take (only) the values 1 and 0.

Ans :- a) 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?

Ans :- a) Central Limit Theorem

Q3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans :- b) Modeling bounded count data

Q4. Point out the correct statement.

Ans :- d) All of the mentioned

Q5. \_\_\_\_\_ random variables are used to model rates

Ans :- c) Poisson

Q6. 10. Usually replacing the standard error by its estimated value does change the CLT.

Ans :- b) False

Q7. Which of the following testing is concerned with making decisions using data?

Ans :- b) Hypothesis

Q8. Normalized data are centered at \_\_\_\_\_ and have units equal to standard deviations of the original data.

Ans :- a) 0

Q9. Which of the following statement is incorrect with respect to outliers?

Ans :- c) Outliers cannot conform to the regression relationship

Q10. What do you understand by the term Normal Distribution?

Ans :- toward the middle of the range, while the rest taper off symmetrically toward either extreme  
A normal distribution is a type of continuous probability distribution in which most data points cluster.

Q11. How do you handle missing data? What imputation techniques do you recommend?

Ans :- Single or Multiple Imputation

- Single and multiple imputation are the two forms of imputation. When people say imputation, they usually mean single.
- The term "single" refers to the fact that you only use one of the seven methods to estimate the missing number outlined above.
- It's popular since it's simple to understand and generates a sample with the same number of observations as the complete data set.
- When listwise deletion eliminates a considerable amount of the data set, single imputation appears to be a tempting option. It does, however, have certain restrictions.
- Unless the data is Missing Completely at Random, certain imputation processes, such as means, correlations, and regression coefficients, result in skewed parameter estimations. The bias is frequently worse than with listwise deletion, which is most software's default.
- The level of the bias is determined by a number of factors, including the imputation technique, the missing data mechanism, the fraction of missing data, and the information in the data set.

Q12. What is A/B testing?

Ans :- [A/B testing](#), also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (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.

Q13. Is mean imputation of missing data acceptable practice?

**Ans :-** The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation.

**Q14.** What is linear regression in statistics?

**Ans :-** Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.

**Q15.** What are the various branches of statistics?

**Ans :-** The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypotheses and draw conclusions