

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

Ans: 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: Central Limit Theorem

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

Ans: Modeling bounded count data

4. Point out the correct statement.

Ans: All of the mentioned

5. _____ random variables are used to model rates.

Ans: Poisson

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

Ans: False

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

Ans: Hypothesis

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Ans: 0

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

Ans: Outliers cannot conform to the regression relationship

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

Ans: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

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

Ans: Missing data appear when no value is available in one or more variables of an individual.

First, detect the percentage of missing values in every column of the dataset will give an idea

about the distribution of missing values. Once we get the location of the missing data, we can

easily find out the type of missing data. If a column holds a lot of missing values, say more than

80%, and the feature is not meaningful, that time we can drop the entire column. I recommend K_Nearest Neighbor Imputation

12. What is A/B testing?

Ans: It is also known as split testing. A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs. variant B and concluding which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?

Ans: No, Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations.

14. What is linear regression in statistics?

Ans: Linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line

15. What are the various branches of statistics?

Ans: Data collection, descriptive statistics and inferential statistics