Statistics Worksheet

1. Bernoulli random variable take(only) the value 1 and 0.
   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?
   Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   Modeling bounded count data

4. Point out the correct statement.
   All of the mentioned

5. _____ random variables are used to model rates.
   Poisson

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

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

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   0

9. Which of the following statement is incorrect with respect to outliers?
   Outliers cannot conform to the regression relationship

10. Normal Distribution : symmetric bell shape
    mean and median are equal; both located at the center of the distribution
    Empirical rule:
    68.2 percent of the data falls within 1 standard deviation of the mean
    95.4 percent of the data falls within 2 standard deviations of the mean
    99. 7, percent of the data falls within 3 standard deviations of the mean

11. How do you handle missing data? What imputation techniques do you recommend?
    First of all we check how many null does are dataset contain either we delete the rows with the

Statistics Worksheet

Null or try some imputation techniques. Using fillna() we can fill those values with different methods like pad and bfill or using max, min, or mean. We can also drop the rows who has missing values.

12. What is A/B testing :
    A/B testing is statistical way of comparing 2 or more versions to determine which version is more impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?
    Mean imputation data practice is considered terrible because
- mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. A linear regression is a method to determine the relationship between one variable with other variable by fitting a linear equation to observed data by fitting a linear equation to observed data. One variable is called independent and other dependent.

15. What are the various branches of statistics?
    There are 2 branches of statistics :
    Descriptive Statistics:
    The branch of statistics that focuses on collecting, summarizing, and presenting a set of data. By using this we can have the data in numerical and graphical form

    Inferential Statistics:
    The branch of statistics that analyzes sample data to draw conclusions about a population. These stats are more useful when its not easy or possible to examine each member of the population.