

STATISTICS WORKSHEET-1

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 the mentioned above. (a) The exponent of a normally distributed random variables follows what is called the log- normal distribution , b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent , c) The square of a standard normal random variable follows what is called chi-squared distribution)

5. _____ random variables are used to model rates.

ANS: poisson

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

ANS: True

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: It is symmetrical distribution where mean, median and mode are equal. It is also called bell shaped distribution or bell shaped curve. Here most of its observations are around the mean.

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

ANS: 1. Most common way to handle missing data is by deleting the row, column which have large missing value.

2. Replace the missing values with mean for numeric data, mode for categorical data, median.

12. What is A/B testing?

ANS: It is an technique often used to understand how an altered variable affects audience or user engagement. It's commonly used in web design, marketing

13. Is mean imputation of missing data acceptable practice?

ANS: In most cases mean imputation is a bad approach. Mean imputation of missing data may lead to bias because we can't use mean in all type of dataset. It can't be used if the dataset contains large difference between 75th percentile and the maximum data, in this case outliers will be present.

14. What is linear regression in statistics?

ANS: It is used to model relationship between two variables(dependent and independent variable). Its objective is to predict the output variable based on the input variable.

15. What are the various branches of statistics?

ANS: Descriptive Statistics, Inferential Statistics.