STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.
a) 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?
a) Central Limit Theorem

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

4. Point out the correct statement
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

4. _____ random variables are used to model rates.
c) Poisson

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

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

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

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

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

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

The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side.

The area under the normal distribution curve represents probability and the total area under the curve sums to one.

Most of the continuous data values in a normal distribution tend to cluster around the mean, and the further a value is from the mean, the less likely it is to occur. The tails are asymptotic, which means that they approach but never quite meet the horizon .

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

a.Dropping rows with null values

b.Dropping features with high nullitys

c.Mean or median or other summary statistic substitution

d.Predicting The Missing Values

e.Impute missing values for categorical variable

f.Impute missing values for continuous variable

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?

Mean imputation of missing data acceptable in observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing.Plus, by imputing the mean, you are able to keep your sample size up to the full sample size.

If all you are doing is estimating means and if the data are missing completely at random, mean imputation will not bias your parameter estimate.

14. What is linear regression in statistics?

14. What is linear regression in statistics?

It is used to predict the data based on the value of another data.The data you want to predict is called the dependent.

15. What are the various branches of statistics?

1.Descriptive Statistics

2.Descriptive Statistics