

## 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 of the mentioned

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. 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

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 is the proper term for a probability bell curve.

In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.

Normal distributions are symmetrical, but not all symmetrical distributions are normal.

In reality, most pricing distributions are not perfectly normal.

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

Ans. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.

The other option is to remove data. When dealing with data that is missing at random, related data can be deleted to reduce bias. Removing data may not be the best option if there are not enough observations to result in a reliable analysis. In some situations, observation of specific events or factors may be required.

12. What is A/B testing?

Ans. it is a test where the model is fed a large number of examples and it is trained to make prediction.

13. Is mean imputation of missing data acceptable practice?

Ans. bad practice in general. If just estimating means mean imputation preserves the mean of the observed data leads to an underestimate of the standard deviation

14. What is linear regression in statistics?

Ans. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

15. What are the various branches of statistics?

Ans. The two main branches of statistics are descriptive statistics and inferential statistics.