Statistics Worksheet

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

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

4. Point out the correct statement.
Answer: d) All of the mentioned

5. _____ random variables are used to model rates.
Answer: c) Poisson

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

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

8. Normalized data are centered at_____ and have units equal to standard deviations of the original data.
Answer: a) 0

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

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

Answer: Normal distribution is a type of distribution which has a bell-shaped curve with equal values of mean, median and mode. It is a very important distribution in statistics and probability because it is observed in a lot of natural processes and the Central Limit Theorem confirms that.

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

Answer: Handling missing data is a data preprocessing step which can't be uniformly applied to all the datasets. It needs to be handled based on data type, dataset size and case to case basis. For example, if we have a numerical column in a dataset which has some missing values, then mean imputation is a commonly used technique, but if the percentage of missing data is very high then it is recommended to remove that column or apply some other technique.

Imputation Techniques:

1. Iterative Imputation.

2. KNN Imputation.
3. Mean/Median/Mode Imputation

12. What is A/B testing?

Answer: A/B testing is a way to statistically determine the better product or version of a product with respect to performance, customer demand, etc. It is a widely used testing methodology in today's industries. In A/B testing, A refers to the original testing variable and B refers to the new version of it. Both the versions are then tested to capture their respective data on performance. This enables industries to make decisions backed by data.

13. Is mean imputation of missing data acceptable practice?

Answer: Mean imputation is acceptable or not depends on case to case basis, in my opinion. Many people consider it a bad practice because it reduces variance and also affects correlation among the variables. But in some scenarios, like, if we know that the data is normally distributed and there are only few missing values then mean imputation is a good way to tackle it.

14. What is linear regression in statistics?

Answer: Linear regression is a type of predictive analysis which tries to predict the value of a continuous variable based on values of one or more variables related to it. We try to find a best fit linear equation based on available data and then use it for prediction.

15. What are the various branches of statistics?

Answer: There are two main branches of statistics.

1. Descriptive Statistics: It involves describing the data using measures like mean, median, mode, variance, etc.
2. Inferential Statistics: It basically tries to infer results or conclusions from the data.