STATISTICS WORKSHEET-1

ANSWERS

1. Bernoulli random variables take (only) the values 1 and 0.
Ans: a

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; a

3. Which of the following is incorrect with respect to use of Poisson distribution?
Ans: b

4. Point out the correct statement.
Ans:d

5. _____ random variables are used to model rates
Ans: c

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

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

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the
original data.
Ans: a

9. Which of the following statement is incorrect with respect to outliers?
Ans: c

10. What do you understand by the term Normal Distribution?
Ans:Normal distribution also known as the Gaussian distribution is a probability distribution that is symmetric about the mean,showing that data near near the mean are more frequent in occurrence than data far from the mean.In graphical form, the normal distribution appears as a bell curve.
In normal distribution, the mean is 0 and standard deviation(std) is 1.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans:There are two ways of handling the missing data:
1. Deleting the missing data(this approach is not recommended)
2. Imputing the missing data.
Most common imputation techniques are:
1. Simple Imputer
2. KNN Imputer
3. Iterative Imputer
Some times we use mean, mode and median depending upon the data.

12. What is A/B testing?

Ans:A/B testing is a user experience methodology. It consist of a randomized experiment with two variants A and B. It includes application of statistical hypothesis testing which means making a wild guess(not wild guess) based on a assumption without scientific proof or explaining the situation based on reasonable assumption.It consist of :

1. Null Hypothesis(Ho)-Decision always leads to status quo. Current status assumption doesn't change.
2. Alternate Hypothesis(Ha)-Decision leads to opposite of Null Hypothesis(Ho).

It also contains one tail test,two tail test,left tail test and right tail test.

13. Is mean imputation of missing data acceptable practice?

Ans: We can use mean for imputation of missing data but using mean values for imputation of missing data may not create a great model and hence gets ruled out.

Again while using mean, the data should be continuous data only than we will be able use that.so instead of using that we can use other imputing techniques such as Simple imputer,KNN Imputer and Iterative imputer.

14. What is linear regression in statistics?

Ans. Regression in statistics is the process of predicting a label(dependent variables) based on the features (independent variables) at hand.

Regression is used for time series modelling and finding the casual effect relationship between the variables and forecasting.

Regression analysis is an important tool for analysis and modelling data.

The case of one explanatory variable is called Simple linear regression and for more than one is called Multiple linear regression.

15. What are the various branches of statistics?

Ans. The various branches of statistics are

1. **Descriptive**: when the population or data is less we will be able to describe it.
2. **Inferential:** when the population or data is more.collecting information from everyone and than making decision.