STATISTICS WORKSHEET – 1

Answers

1. A. True
2. A. Central Limit Theorem
3. B. Modeling bounded count data
4. A. The exponent of a normally distributed random variable follows what is called the Long normal distribution
5. C. Poisson
6. A. True
7. B. Hypothesis
8. A. 0
9. C. Outliers cannot conform to the regression relationship

10. **Normal Distribution**: Normal Distribution is also known as probability or error distribution. An error is the difference between an actual value and a statistical estimate like the sample mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

11. By understanding the given data, first we need to understand what data is missing and why it is missing also the impact of missing data on our analysis. Based on type of missing data we can decide the imputation techniques...

    Some missing data can be resolved by using mean, medium and mode techniques

    We can also use next or previous values

    Another way is using minimum or maximum value of database

    Hot or Cold deck imputation method or regression imputation

12. The A/B testing is the most common and basic randomized experiment. So, there are two treatments, and one acts as control for the other. It is a way to compare two versions of dataset to figure out which performs better.

13. In the random missing data, using the mean imputation method remains unbiased. Where in the missing data case the mean imputation decreases the variance of database while increasing the biasness.

14. By using the independent variable's value to predict the dependent variable's value is called linear regression. It gives us a linear relationship between an independent variable and a dependent variable to predict the outcome of future events.

15. There are two branches of Statistics

    a. **Descriptive Statistics**: This analysis deals with collection and presentation of data. We can define as an explanatory means to describe the given sample/population, trying to extract summary and basic information about the sample/population using tools like graphs, tables, calculations of absolute and relative frequency, position and variability indices.
    b. **Inferential Statistics:** Once the data has been collected, the use of these stats is analyzed and summarized to draw conclusions that depend on random variation such as sampling error or variation, observation errors, etc. Based on sample, with this method we can make decisions, predictions or inferences from a population