Premnath C

STATISTICS WORKSHEET – 1

1. a) True

2. a) Central Limit Theorem

3. b) Modeling bounded count data

4. b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

5. c) Poisson

6. b) False

7. b) Hypothesis

8. c) 1

9. c) Outliers cannot conform to the regression relationship

10. A normal distribution is **a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme**. The middle of the range is also known as the mean of the distribution.

11. Handle Missing Data:

- One way of handling missing values is the **deletion of the rows or columns having null values**. If any columns have more than half of the values as null then you can drop the entire column. In the same way, rows can also be dropped if having one or more columns values as null.
- **The following are common methods:**

- Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. ...
- Substitution. ...
- Hot deck imputation. ...
- Cold deck imputation. ...
- Regression imputation. ...

- Stochastic regression imputation. ...
- Interpolation and extrapolation.

12. A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better.

13. Mean imputation is typically considered terrible practice since it ignores feature correlation.

14. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. There are three real branches of statistics: Data collection, descriptive statistics and inferential statistics.