Statistics Worksheet – 1

1. A) True

2. A) Central Limit Theorem

3. B) Modeling bounded count data

4. D) All of the mentioned

5. C) Poisson

6. B) False

7. B) Hypothesis

8. A) 0

9. C) Outliers cannot conform to the regression relationship

10. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

11. Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields, Use regression analysis to systematically eliminate data.

* **Data scientists can use data imputation techniques such as :-**

i) Complete Case Analysis (CCA) :- This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.
ii) Arbitrary Value Imputation.
iii) Frequent Category Imputation.

12. A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

13. Yes, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). Linear regression has many practical uses.

15. There are three real branches of statistics: Data collection, Descriptive statistics and Inferential statistics.