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. A - The normal distribution is a continuous probability distribution that is symmetrical around its mean. Most of the observations cluster around the central peak. Probabilities for values further away from the mean taper off equally in both directions. The normal distribution describes how the values of a variable are distributed. It is the most important probability distribution in statistics because it fits many natural phenomena. Eg: blood pressure, heights, etc. It is also known as the bell curve.

11. A - One of the ways we can handle missing data is by eliminating missing data using deletion methods. When data is missing at random, we can use listwise or pairwise deletion of the missing observations. We can use regression analysis to systematically eliminate data. We can also use imputation techniques like mean or median imputation, multivariate imputation by chained equations & random forest.

12. A - An A/B test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not, therefore it is an analytical method for making decisions that estimates population parameters based on sample statistics.

13. A - Mean imputation is not a good solution. The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation.

14. A - Linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. 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).

15. A - There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.
    - Data collection is all about how the actual data is collected.
    - Descriptive statistics is the part of statistics that deals with presenting the data we have. This can take two basic forms presenting aspects of the data either visually or numerically.
    - Inferential statistics is the aspect that deals with making conclusions about the data.