Worksheet 1 Statistics

1. a) True
2. a) Central Limit Theorem
3. b) Modeling bounded count data
4. d) All of the mentioned above
5. d) Poisson Distribution
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. Normal distribution means the distribution that is normal around mean. Meaning that the data around the mean of all the cumulative value is more recurring than the data that is far away from the mean.
11. Missing data is one of the most severe problems while analyzing the data. It sometimes ruins the data as a whole giving unvaried values that are not good for model analysis. We can deal with this missing data in multiple ways. Some of the techniques i personally used are,
    - Substituting a value according to the given data that will give you a better reading
    - Mean of the values and then substituting that particular value
    - Regression, regressing other values and getting a value that can be implemented instead of the missing values
    - Stochastic regression is nothing but regressing other values plus a mean value and together getting a data that will be beneficial in analysis
12. In A/B testing the A here is the original model whereas B is the newer model with some made with a different testing technique. Both the models are considered and the model

that is more preferable by the business and better helps in solving the business problem. That model is considered to be the winner.

13. The practice of mean imputation is a practice that is done and is okay if you want to build a model but the problem here is making a model with the mean imputed data will make severely biased models because the missing data are means and there is a risk of not getting a centralized value thereby causing a biased model.
14. Linear regression means that you are predicting value of a variable using another already existing variable. The variable you re predicting the value for is called the dependent variable and the variable you are using to predict this value is known as the independent variable.
15. The three major branches of statistics are
    - Data collection - the collection of data to build a statistical model
    - Descriptive statistics - organization, summarisation and display of data
    - Inferential statistics - takes a sample as example to draw conclusion about a mass of data