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) Poission
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. Normal Distribution it is also called as “bell shaped curve” or Symmetrical curve. Where the mean, median, and mode all are equal to 0 and standard deviation =±1. Normal distribution is also symmetry in nature. It’s an ideal distribution for example let’s says that suppose all the employee in a company are taking the same salary which is ideally not possible.
11. Missing data can be handled in many ways, most common way is to ignore it. Another way is to remove all the missing data from the data set but it is not a good practice. So we handle the missing data by using mean imputation.

    In mean imputation we are taking the sum of the data from the missing column and divide by total number of data in a column, so that we get a mean value, this mean value will be within the range of that particular column so that we can fill this mean value in the missing data column.
12. A/B testing also known as split testing or bucket testing is a methodology for comparing two version of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to user at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

    In an A/B test, we take a webpage or app screen and modify it to create a second version of the same page.
13. Yes, whenever a data is missing we can go with a mean imputation which is general practice, using mean imputation the mean value will be with the range of that sample data.
14. Linear regression is used to predict the value of a variable based on the value of another variable. The variable that we want to predict is called dependent variable and the variable that we are using to predict the other variable value is called independent variable.

    For example: the weight of the person is linearly related to their height, increase in the height, the weight of the person will also increases.

    Linear regression shows the linear relationship between the independent variable and the dependent variable.

15. There are two main branches in statistics
    1. Descriptive statistics and
    2. Inferential statistics

    Descriptive statistics summarizes or describes the characteristics of a dataset.

    Descriptive statistics if further divided into three categories:

    1. Measure of central tendency – describe the center of the data set (mean, median, mode).
    2. Measure of variability (or spread) – describe the dispersion of the data set (standard deviation, variance).
    3. Frequency distribution – describe the occurrence of data within the data set

    Inferential statistics is the practice of using sample data to draw conclusions or make predictions on a large sample data or population.

    Inferential statistics is divided into two main types :

    1. Hypothesis testing – Z test, T test and F test
    2. Regression analysis – linear regression, multiple regression.