Statistics Worksheet-1

Answers 1 to 9

1. a)
2. a)
3. b)
4. d)
5. c)
6. b)
7. b)
8. a)
9. c)

Answers 10 to 15

10. In a Normal Distribution data is symmetrically distributed with no skewness. When plotted on a graph it will shape a bell curve. Normal distribution is also known as Gaussian Distribution or Bell curve. In Normal Distribution the mean, median and mode are exactly the same. It can be described by two values mean and the Standard Deviation.
11. i)To handle a missing data I first analyse why the data is missing and accordingly apply imputation techniques.
    ii) Being a beginner in Data Science world I was first introduced with Statistical imputation techniques which is Mean, Median and Mode and later I learned about few more scikit-learn imputation techniques like Knn and Iterative Imputer and I think they works best in handling missing values.
12. A/B testing is also known as split testing or bucket testing. It is a simple randomized controlled technique to compare two versions of a single variable. It includes application of Statistical hypothesis testing or two-sample hypothesis testing. Through A/B testing its's easy to get a clear idea of what users' preference as it's directly comparing one variable to another.
13. Mean imputation of missing data is not a good practice as it's ignoring feature correlation. Sometime we can use it when data set is small and we need to retain as many information as possible but in big datasets if we use mean imputation technique and the standard error will be more which can lead to incorrect predictions in Machine Learning.
14. Linear Regression is one of the most fundamental and widely known Machine Learning model to predict label based on features. Linear regression can be used when the label is continues value.

15. The Various branches of Statistics are Descriptive and Inferential Statistics. Descriptive Statistics represents what's going on in a population or a data set and Inferential Statistics helps to come to conclusions and make prediction based on data.
    a. Descriptive Statistics (Tools Using):
        i) Measure of Central Tendency (Mean, Median and Mode).
        ii) Measure of Dispersion or Spread (Variance and Standard Deviation).
    b. Inferential Statistics (Tools Using):
        i) Regression Analysis.
        ii) Hypothesis Testing.