STATISTICS WORKSHEET-1

1.(a)

2. (a)

3.(b)

4.(d)

5.(c)

6.(b)

7.(b)

8.(a)

9.(c)

10.Normal Distribution is an arrangement of a dataset in which most of the values cluster at the middle of the range and tapper off symmetrically at either extreme.It features a symmetric Bell shape.This curve is symmetric around the mean.

11.Inorder to handle the missing data we have a technique called Imputation.This technique helps in replacing the missing data with a substitute value because removing the data from the dataset is always not feasible.Some of the following imputation techniques can be used:

i) Arbitrary Value Imputation

In this method, the missing values are replaced with some arbitrary values like 0,9,99,999 or negative values if the variable distribution is negative.

ii) Start/End of Distribution Imputation

In this method the imputation is done at the end of the distribution.For example , if the distribution is normally distributed, then we can use plus/minus 3 standard deviations from the mean to determine the ends.

(iii) Mean/Median/Mode Imputation

Here, in this method, the missing values can be substituted with the mean if the data distribution is uniform and with the median, if the data distribution is skewed.

(iv) KNN Imputation

This method is advanced, that uses an algorithm that makes predictions based on the defined number of neighbours.KNN stands for K-nearest neighbours and it is a distance based algorithm.

12.A/B testing is a way to compare the two versions of the variable to find the best performed variable in the controlled environment.This really helps in the growth of the business by the outcome of best performed version of variable through the A/B testing.

13. Mean imputation is not a good practice in general because the mean imputation just preserves the mean in the course of estimation.This leads to the underestimation of the standard deviation.Also, mean substitution leads to bias in multivariate estimates such as correlation or regression coefficients.

14.Linear regression is a commonly used type of predictive analysis.Basically, it attempts to model the relationship between the two variables by fitting a linear equation to the observed data.The equation for the linear regression is as follows:

$y = mx + c$

x: Score of the independent variable

m: Regression coefficient

c: Constant

x: Independent variable

15.There are two major branches in statistics:

(i)Descriptive Statistics

(ii)Inferential Statistics

Descriptive Statistics-This branch of statistics involves organization, summarization and display of data.

Inferential Statistics-This branch of statistics involves the sample to draw a conclusion about the population.The basic tool in the study of inferential statistics is Probability.