Statistics Worksheet-1

1. A
2. A
3. B
4. C
5. C
6. B
7. B
8. A
9. D

10. Normal Distribution curve is also known as the bell curve because of its shape. It is a continuous probability distribution in which most of the points are towards the centre that is the mean. With the mean as the centre it as three standard deviation value on each side of the mean with left side as negative standard deviation and right side positive.
    The general rule of Normal distribution state that :-
    a) Between +1 and –1 standard deviation lies 68% of the data
    b) Between +2 and –2 standard deviation lies 95% of the data
    c) Between +3 and –3 standard deviation lies 99.7% of the data

11. There are various techniques to handle Missing data :-
    a) Deleting missing values – This technique can only be used if we have a very large data. Using this on a small data can give you false result
    b) Statistical approach- In this we replace the missing values with the mean or median of the available data.
    c) Regression – Use regression technique such as linear regression to predict the missing values based on the values of other features. This method can only be used when there is a clear correlation between the features.
    Based on the type of data provided we can used any of the three following techniques.

12. A/B testing is most widely use testing tool to compare two or more variations of a particular feature to determine which performs better. It involves randomly dividing the samples into two or more groups and exposing each group with a different version of features. By comparing the outcome of the different group we come to a conclusion. It is popularly sued by marketing, product designing and website designing.

13. Mean imputation can be considered as an acceptable practice but it has its own limitation and drawbacks.
    a) Uncertainty: - Mean imputation ignores the possibility of uncertainty. It assumes that the values are as acute as the observed value, which may not be true.
    b) Loss of variable: - It means that the all the missing variables are replaced by the same variable.
    c) Biasness: - If the missing data is not random but has a relationship with a dependence variable then it can create biasness into the data.
14. Linear regression is a statistical tool which help us understanding the relation between two or more variables. It helps in determining the relationship between the dependent variable and the independent variable. It is widely used technique which is used to find the value of

dependent variable depending upon the independent variable.

Formula:-

$Y=mX+b$

Y= dependent variable
m= slope of the regression line
X= represents the independent variable
b= it represent the interception

15. The various branches of statistics are:-
    a) Descriptive statistic - It is method of organizing, summarizing and presenting the data into informative way. It involves techniques such as central tendency, measure of dispersion and graphical representation
    b) inferential statistic- It is method which is used to determine something about a population baes of sample. It uses techniques such as hypothesis testing, estimation etc. to draw a conclusion
    c) Statistical Modelling- It focuses on developing statistical methods and techniques and explain the relationship between two variables.
    d) Data Science and Machine learning- It involves using stats methods to analyse and interpret large and complex data

    These are the some of the example of branches of statistics