

# Assignment-Statistics

Q1.(a)

Q2.(a)

Q3.(d)

Q4.(d)

Q5.(d)

Q6.(b)

Q7.(b)

Q8.(a)

Q9.(c)

Q10. Normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. In normal distribution, data are symmetrically distributed with no skew.

Q11. One way of handling missing values is the deletion of the rows or columns having null values. If any columns have more than half of the values as null then you can drop the entire column. In the same way, rows can also be dropped if having one or more columns values as null.

Imputation techniques:-

The simplest imputation method is replacing missing values with the mean or median values of the dataset at large.

Others are:-

- Next or Previous Value.
- K Nearest Neighbors.
- Maximum or Minimum Value.
- Missing Value Prediction.
- Most Frequent Value

12.) A/B testing compares the performance of two versions of the same thing, often a webpage or a digital product.

Q13.) Mean imputation is not very preferred/acceptable practice since it ignores feature correlation, it changes the variance of the dataset, making it a less accurate estimate for the variance of the actual population. More generally, it will make the dataset a less accurate reflection of the population, in many ways

Q14.) Linear regression is the most basic and commonly used predictive analysis. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Q15.) The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data