Statistics Worksheet-1

Q.1- **a**

Q.2- **a**

Q.3- **b**

Q.4- **d**

Q.5- **c**

Q.6- **b**

Q.7- **b**

Q.8- **a**

Q.9- **c**

Q.10-Normal Distribution?

Normal distribution also known as the Gaussian distribution is a probability distribution that is symmetric about the mean.

Also knowing as Three sigma Rule or empirical Rule

(68.26-95.44-99.7)

The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side. The area under the normal distribution curve represents probability and the total area under the curve sums to one.

The shape of the distribution changes as the parameter values changes

Mean. The mean is used by researchers as a measure of central tendency.

Standard Deviation.

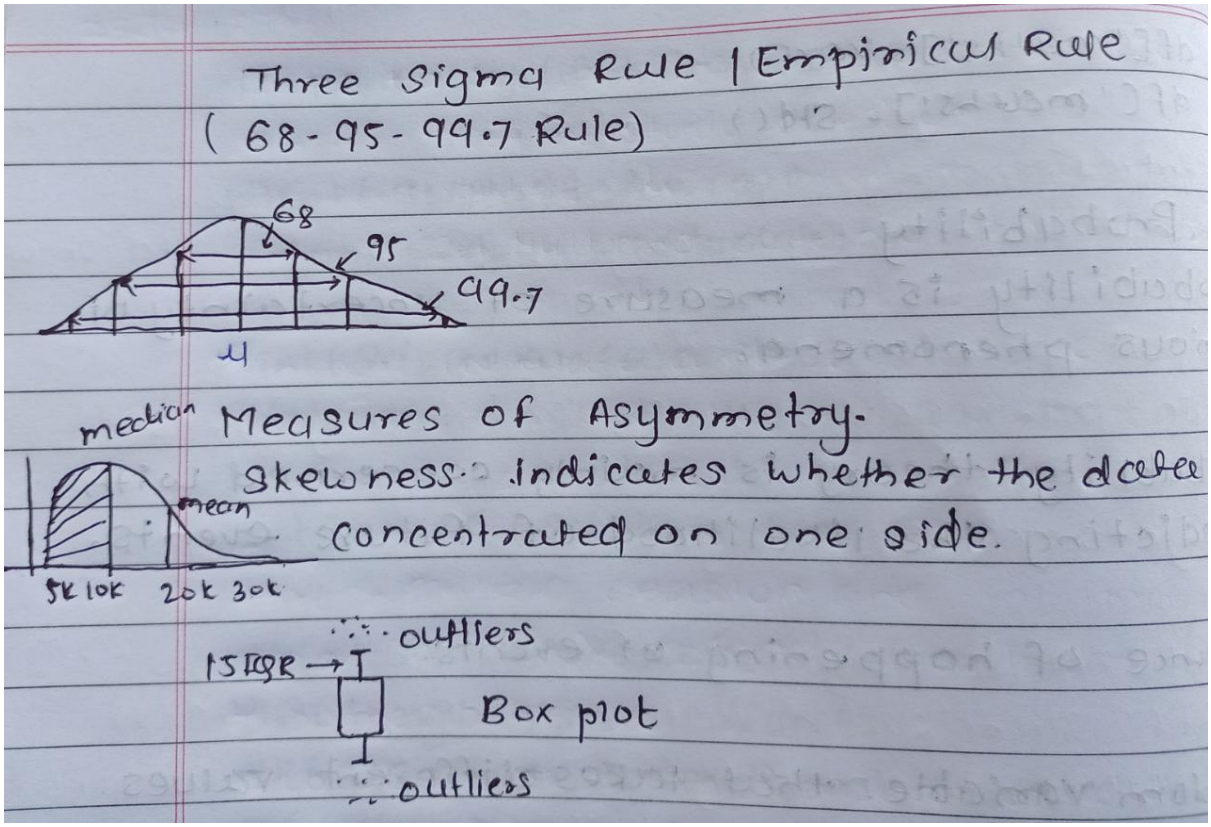
It is symmetric.

The mean, median, and mode are equal.

Empirical rule.

Skewness and kurtosis.

Statistics Worksheet-1



Q.12-Handle Missing Data?

In statistics missing data, occur when no data value is stored for any variable or when data has been stored incompletely.

Missing data are a common occurrence and can have a significant effect on the conclusion that can be from any data.

<u>Data Imputation</u>

in statistics imputation is the process of replacing missing data with substituted values. we will use expectation maximization method (Em).

the Em algorithm is an iterative method to find maximum likelihood estimates of parameters in statistical model

Various techniques to handle missing data:

Delete the record\observation missing value

Create a seperate mode to handle missing value

Statistics methods mean,median or mode

Statistics Worksheet-1

Q.12-A/B Testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

Q.13-is mean imputation of missing data acceptable practice?

Yes, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing. ... Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

Q.14-Linear Regression?

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. In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables

Q.15-what are the varies branches of statistics?

Types of statistics-1.Descriptive Statistics 2.Inferential Statistics

Further again Descriptive Statistics is again divide in 2 part and then subpart

1.Measure of central Tendency-Mean,Mode,Median

2.Measured of variability(Spread of Data)-Range,Variance,Dispersion