Name: Dishant Doshi

Intership19

STATASTIC WORK SHEET

01) A
02) A
03) B
04) D
05) C
06) A
07) B
08) A
09) C

Q10

Ans: Normal Distribution is a probability function that describe how the values of variable are distributed.it is a symmetric distribution about the mean. That indicated data near the

Mean are more frequently occurred and data far away from mean are less occurred. It is

Also known as Gaussian Distribution Curve and Bell Curve.

In Normal distribution mean and median are equal and 68% data lies with 1 standard deviation. 95% data lies with 2 standard deviation and 99.7 % data lies with 3 standard deviations. It is know as Three sigma rule or empirical rule.

Normal Disribution Curve
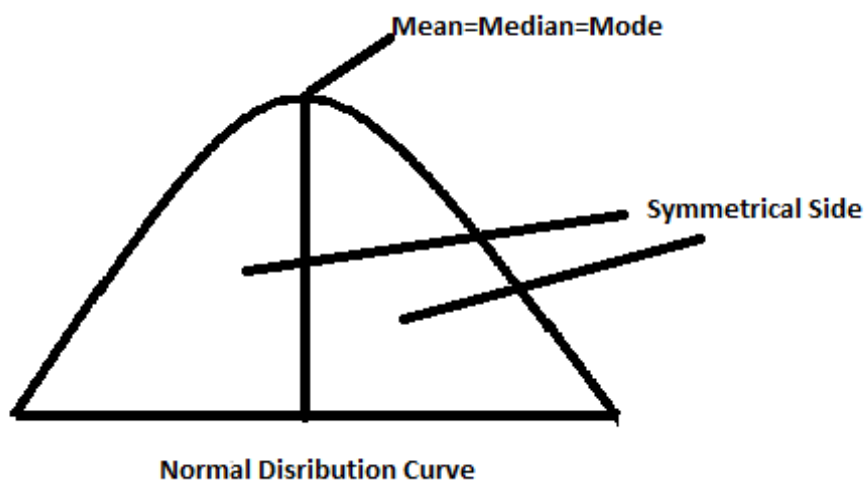
Q11

Ans: When the no value stored in variable during observation its called missing value/data.

There are two way to Handel this missing data

   01) Removing Data
   02) Imputation

When the missing data is at random, related data can be deleted to reduce the bias,but deleting the data is not good practice if there are not enough observation at that time result in analysis may be affect so that we have to fill this missing data For that we have to find out the Mean or Median of the features from the existing data and replace missing values with this Mean or Median.

Q12

Ans. A/B testing means show your current experience to half of your customer and alternat experience to to remaining half of the customer and compare both of them and find the difference of the performance, then either continue with the old one or switch all with the new one.

Q13

Ans: No, mean imputation of missing data is not acceptable practice. Because using mean can significantly reduce the model's accuracy and bias the result.

Q14

Ans. Linear Regression is used for predictive analysis show the relation between the continues variable it shows the relationship between the independent variable and Dependent variable. If

There is one single input variable its simple linear regression and there are more than one input variable than it is called multiple linear regression.

It simply uses traditional slope intercept form.

$Y = mx + b$

Where,

Y= Dependent variable

B=intercept of line

M=Linear Regression Co-efficient

X= independent variable

Q15

Ans. There are mainly two branches of statastics.

01) Dispersive statistic: Descriptive statistics is the first part of statistics that deals with the collection of data. It having two parts.
- Central tendency measures
- Variability measures

02) Inferential statistic: The inference statistics are techniques that enable statisticians to use the information collected from the sample to conclude, bring decisions, or predict a defined population.it include following methods,

- Regression analysis
- Analysis of variance (ANOVA)
- Analysis of covariance (ANCOVA)
- Statistical significance (t-test)
- Correlation analysis