https://www.youtube.com/watch?v=LZzq1zSL1bs

What is statistics?

Collecting, organizing and analyzing the data for better decision making

Types of Statistics:

1. Descriptive Stats: It consists of organizing and sumarrizing the data

Ex: What is the average marks of students in a class?

2. Inferential Stats: Techiniques where we used the data to measure the conclusions

Ex: Are the marks of the students of this classromm similar to the marks maths classroom

Sample and Population:

Sample is subset of population

Sampling will have multiple techiniques.

1. Simple Random Sampling: Randomly pick the samples. Used in many of the scenarios. Every member of the population has equal chance of being selected.

2. Stratified Sampling: Where the population is split into non overlapping groups. Ex: Gender -> male and Female

3. Systematics Sampling: Nth individual. Ex: Taking the survey for every 8th person.

4. Convience Sampling : Only those people who are expert speficic to that domain

5.

Variables:

Quantitative variables: These are like numarical variables.

1. Descrete variables: Ex: Number of childeren, number of bank accounts etc: We will not have floating values.

2. Continuous variables: Ex: Height, Weight , Tempature etc.

Categorical variables:

Ex: Gender, states etc

Variable manage techniques:

1. Nominal Data : Categorical data

2. oridnal Data : Order of the data is the matters

3. Interval Data: Order and values matters

Frequencey Distribution:

Sample Data sets : Rose, lilly,sunflower, rose, lilly, sunflower, lilly,rose, lilly

Rose: 3

Lilly - 4

Sunflower : 2

Cumative is 3+4+2 =9

Bar chart is used for represent descrete values, and Histogram is used to represent Continous values

Central Tendancy:

It refers to the measure used to dentermine the center of the distribution of the data.

Mean (Average) , Median and Mode

Mean is the average

Median is the central value after sorting the dataset. If the dataset is even, then its avg of those two values

Mode is most frequent element in the data

Measure of Dispersion:

Dispersion means spred

Variance: If we need to identify how two distributions are different even though mean is same

Ex: d1={1,1,2,2,4} mean = 10/5 =2

d2={2,2,2,2,2} mean = 10/5 =2

Population Variance : (Sum of (xi - mean)square)/ N

and Sample Variance

If the variance is more then, data is more distributed.

Standard Deviation : SQRT of Variance.

From mean, one strd deviation right or left called one standard deviation

Percentaile and Quartiles:

Percentalie: It is a value below which a certain percentage of observation lie

Ex: Data set = 2,2,3,4,5,5,5,6,7,8

Percentile rank of 6 is = (number of values below x/n)\*100

x is the number

n is the total

= (7/10)\*100 = 70%

means 70% of entire distribution is less than 6

What value exists at percentalie ranking of 25%

Formula : value\_postition = (percentaile/100)\*(n+1)

value\_position = (25/100)\*11 = 2.75

value = (2+3)/2 = 2.5 is value of 25 percentaile

Five Number Summary:

1. Minimum

2. First Quartile (Q1)

3. Median

4. Third Quartile(Q3)

5. Maximum

Usinng these we will remove outliers

ex: {1,2,2,2,3,4,4,5,6,6,7,9,27}

Define lower fence and higher fence, values in the data set below lower fence are outliers and same as above higher fence are outliers

lower fence = Q1 - 1.5(IQR)

higher fence = Q3 +1.5(IQR)

IQR (Inter Quartile range) = Q3-Q1

Q3 is 75 percentalie = (75/100)\*(13+1) = 10.5 (postition) = (6+7)/2 = 6.5

Q1 is 25 Percentalie = (25/100)\*(13+1) = 3.5 (position) = (2+2)/2 = 2

IQR = 6.5-2 = 3.5

Lower fence = 2-1.5\*3.5 = -3.25

higher fence = 6.5+1.5\*3.5 = 11.75

So 27 is the outlier

after removing outlier the data set is

{1,2,2,2,3,4,4,5,6,6,7,9}

Minimum value is 1

Q1 = 2

Median = 4

Q3 = 6.5

Maximum = 9

Need to draw Box plot

Box plot is used to represent outliers in visulation format.

Distribution:

1. Normal or gaussian Distribution: It is a distribution that appears as a bell curved when graphed.

central line will be mean or median or mode. Left side of the bell curve will be simitrical as right side