1. Data related concepts
2. Data is divided into two types

* Numerical
* Categorical

1. Numerical data

* Continues : height,weight, temp
* Discrete : counting numbers

1. Levels of data

* Nominal
* Ordinal
* Interval
* Ratio level

1. Population (collection of every object)
2. Sample (selection of some objects from corresponding population)
3. Population is never exist
4. Resource / money / time ========= sample
5. How to represent the data
6. Frequency table
7. Bar graph
8. Pie chart
9. Frequency distribution table
10. Histogram
11. Distribution plot

Data distribution:

First will have raw data ===== class intervals ======= how many observations fall in the interval

Histogram

On free line on top of histogram ======= distribution plot

I want understand more about my data

1. Centre of the data
2. How data varies
3. Central tendency

* Mean
* Median
* Mode

1. Data distribution/varies/flows/deviation

* Range
* Mean deviation
* Absolute mean deviation
* Variance
* Standard deviation

1. Central tendency
2. Mean :

Average

Virat kohli average in ODI is 60=====

Approximately on every match he scored 60 runs

If he play next match we can expect 60 runs

600

89.5% in ssc?

Out of 100% get 89.5%

Approximately each subject get above 89.5 marks

Out of 100 he score 89.5 marks

There are 6 subjects

91 92 89 90 94 81

Average = sum of observation / total number of observations

= 91+92+89+90+94+81/6=89.5

On of average in each subject I got 89 marks approximately

1. Median

Mid value

Middle point of the data

1,6,9,20,21,26,13

What is the middle point

1. Make the numbers are in ascending or descending order

1, 6, 9, 13, 20, 21, 26

Median = 50 percentile of data

Middle point of data

If raw data observation are even number?

1, 6, 9 13 20, 21, 26

Median = 13

1, 6 , 9 , 13, 20, 21, 26, 30

Average of two numbers = (13+20)/2 = 33/2 = 16.5

1, 6 , 9 , 13, 20, 21, 26, 30, 30, 30, 30, 30

Average of two numbers = (21+26)/2 = 47/2 = 23.5

Assume consider total data is 100 percentile

50 percentile is mid point

Mean vs Median

Both will explain about centre point

Both will not explain about min and max points

Is both concepts effect by min and max value?

Assume that Indians income

50k, 60k, 70k, 80k, 100k

What is the average : 50+60+70+80+100/5 = 360/5 = 70k

What is the median : 70k

50k, 60k, 70k, 80k, 100k, 100crs

What is the average: 50+60+70+80+100+10000k/5=360/7= 75k

Median : 75k

With huge observation either min or max : mean will affect

: median will not effect

This huge observation = outliers

Amabni is a outlier

India ========== middle class is more

Zim =========== poverty is more

Poverty number is pulling down middle class number when you do average

Rich number is pulling up middled class number when you do average

Outlier can be min value or can be max value

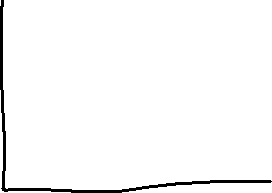
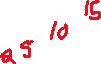
c) mode:

most occurrence of number

most repeated number

1,1,2,2,2,3,3,3,3,4,5,6,7

Mode = 3



There is a number between 15 and 20 repeated 15 times

Mode is : most repeated number

13 ================ 15 times

13 is mode ------ ? from the graph

It provides quick single point idea

When the data has numerical data

When the data has categorical data

10 times tou coin a toss

H, H, H,H,H,H,T,T,H,H

11th time ======== H

1. Mean will affect with huge observation either negative or positive value
2. Median 50 percentile data
3. Mode highest peak of distribution

