Data Related Concepts

Data is divided in to two parts

Numerical

Categorical

1 Numerical

* Continues
* Discreate

Data Distribution:

1 Center

2 Data is floatation

Central tendency

* Mean
* Median
* Mode

Data Distribution/varies/flows/deviation

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

Every one clear

Central tendency

1. Mean:

Average

Virat Kohli average in ODI 60

Avg he scores 60

On every match then Min 60 runs

600

89.5% ==============?

Average= sum of observation/ Total number of observation

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

= 89.5 marks

On of average in each subject I got approximately 89

Summation

Pi

Mean

/n

Median

Ascending order or descending order first

(@)================ 50 percentage of data

Middle point of data

If the raw data observation is even number?

Mean VS median

Both will explain about same point

This is central tendency

Both will never explain about the min and maximum points

Is both concept effect by min and max value?

Consider

Assume that Indian income

* 50k
* 60k
* 70k
* 100k

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

Indian avg software employee is 70

Median= 70k

Adding Ambani 100crs

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

Indian avg software employee is 7crores

That is called is inflation

The money value is less

1. Median = 70+80/2=75K

Median not effects

This huge observation = outliers

Ambani is an outlier.

Outlier can be as a min and max value

House brokers 3 houses

2cr 2.5cr 1.5cr

Data Analysis

To develop any model ============== good data

How to deal outliers?

Separate session is required

1. Mode

Most occurrence of number

1,1,1,1,2,2,23,3,3,5,5,5,6,6,6

Mode= 3

Bar graph and histogram

Graduation

It provides quick single point idea

When the data is numerical data is fine and categorical data is fine.

Any data highest point is called mode

Mode = more curve median = 50 % mean = average

Skewed

Right skewed data

+ve skewed data

Mode < median < mean

Left skewed data

-ve skewed data

Skewed= pilling, outlier, mean will affects

Normal distribution.

Mean, median, mode is same point there is no skewed data

Bell shop

We have 3 distribution

* Left skewed
* Imagined real axis left side: negative
* Skewed means pulling data
* Data pull or skew happen of outliers
* Mean will affect by outliers
* Median always 50 %
* Mode is highest peak
* Mean<median<mode
* Right skewed or Positive
* Mode<median<mean
* Normal distribution
* Mode=median=mean
* Bell shaped curve
* 50% of data both sides
* All the math follows Normal Distribution
* They will apply math
* Any exam: GMAT/GATE/CAT