Statistics session − 3:

Data measurements

Data measurements divide into two parts

1) *Central tendency*
   - *Mean*
   - *Median*
   - *Mode*
2) *Data dispersion*
   - *Range*
   - *Mean diviation*
   - *Absolute mean deviation*
   - *Variance*
   - *Standard deviation*

- *Central tendency*:
  - *Mean*:

VK average in ODI is 50
- Assume, *If you give bat to VK, he will make* 50 *runs*

$Telugu = \quad 91$
$Hindi = 81$
$Eng = 94$
$Maths = 89$
$Science = 90$
$Social = 94$

$$\frac{91 + 81 + 94 + 89 + 90 + 94}{6} = \frac{537}{6} = 89.5$$

He can make 90 *marks in every sub*

$$Average = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

There are N observations

$$Average = \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_N}{N}$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{N}$$

$\mu = population\ mean$

$\bar{x} = sample\ mean$

- ***Median:***
  - *Median also a metric tell about mid point of Data*
  - *50percentile of data*
  - *keep the order acsending or decending*

$1,\ 2,\ \ 3,\ \ 4,\ \ 5,\ \ 6,\ \ \ 7 = 4$

$1,\ 2,\ \ 3,\ \ 4,\ \ 5,\ \ 6,\ \ \ 7,8\ = \frac{4+5}{2} = 4.5$

- ***Mean vs Medain***
  - *USA is asking indain what is Avg indian income*
  - $1l, 2l, 3l, 4l, 5l$
  - $Avg = \frac{1+2+3+4+5}{5} = 3l$
  - $Median = 1,2,3,4,5 ====> 3l$

$1l, 2l, 3l, 4l, 5l, 200crs$

$Avg = \frac{1+2+3+4+5+200crs}{5} = 20crs$

$Median = 1,2,3,4,5,200cr ====> \frac{3+4}{2} = 3.5$

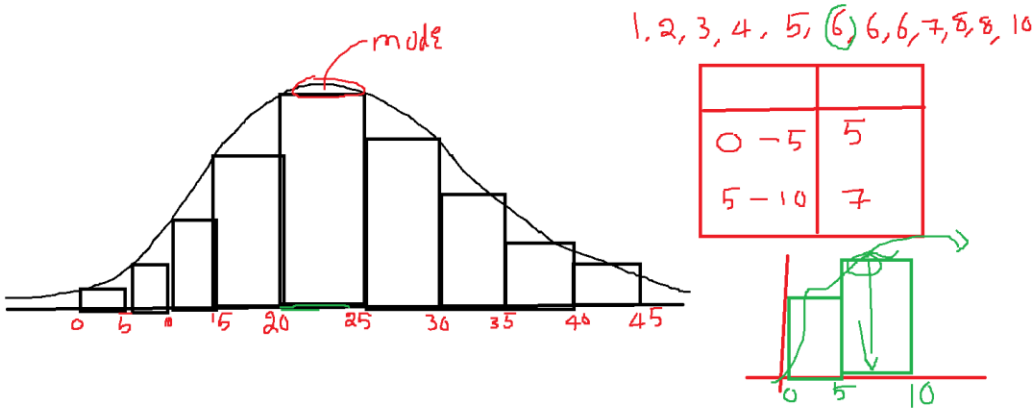
If a data has very very huge value, very very less value

Mean will affect, Median does not affect

These unusual obervations called as : ***Outliers***

- ***Mode:***
  - Most repeated value
  - *Most frequently occured value*
  - *raw data*: 1,5,6,7,1,6,1,5,1,8,1,3,1
  - $1,1,1,1,1,3,5,5,6,6,7,8 ==== Mode = 1$
  - *Data distribution* ===

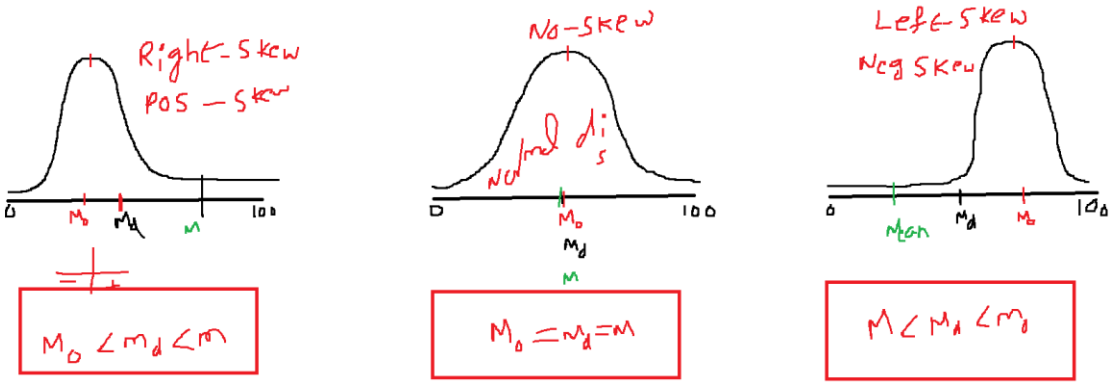
CI	CIF
0-3	6
3-6	4
6-9	2



- *Data distribution Highest peack is called as Mode*
- *Mode is avaiable at that point*
- *we know that distribution forms from histogram*
- *Histogram form from interval*
- *If you are seeing highest peack in the distribution means*
- *That corresponding interval has mode value*

Mean – Median – Mode

1) *Mean will give average value of the data*
2) *Median wil give middile vale* 50 *percentile data*
3) *Model will give highest peack in the distribution*



$avg = 1,1,1,1,1,1,1,1,1,1,1,1,1,3,5,5,6,6,7,8, === > 5marks$

$$avg = 1,1,1,1,1,1,1,1,1,1,1,1,3,5,5,6,6,7,8,99 === 20$$

Left skewed or Negative Skewed:

- *Because of Negative outliers*
- *Mode > Median > Mean*
- *Assume that data ranges from* 0 *to* 100
- *Negative side or left side data is pulling which mean* 0 *side*
- *So that Mean value is low*

Right skewed or Postive Skewed:

- *Because of Postive outliers*
- *Mode < Median < Mean*
- *Assume that data ranges from* 0 *to* 100
- *Postive side or Right side data is pulling which mean* 100 *side*
- *So that Mean value is High*

No skew or Normal distribution:

- *No outliers*
- *Mode = Median = Mean*
- *Bell shape curve*
- 50% *data in left side and* 50% *in Right side*

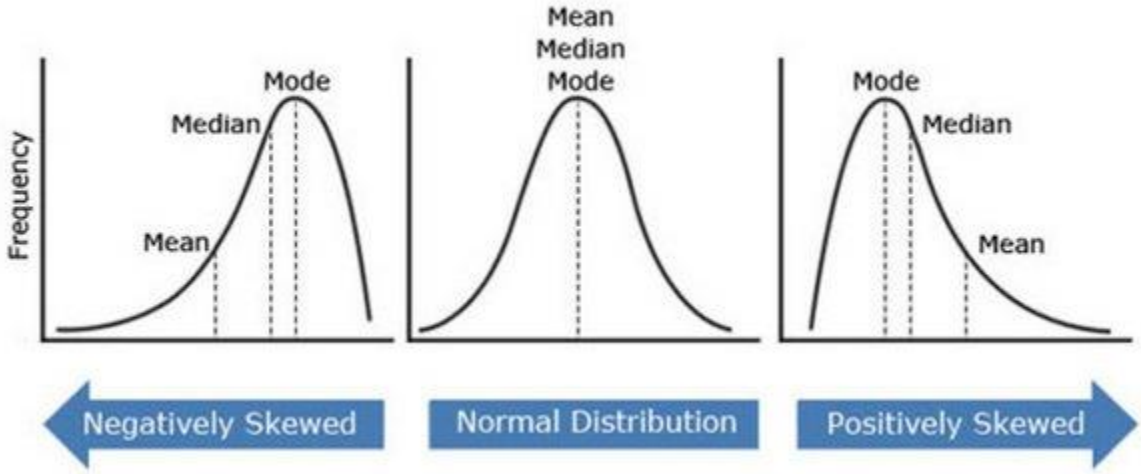
What is the meaning skew: Pulling

who is the reason for skew : Outliers

The oulier is which side:

Right side means (Max value)(Based on Coordinat)

Left side means (Min value)


Frequency
Mode
Median
Mean
Mean
Median
Mode
Mode
Median
Mean
Negatively Skewed
Normal Distribution
Positively Skewed