STATISTICS

(SESSION-4)

DATA DISPERSION :

1.Range

2.Mean deviation

3.Absolute Mean Deviation

4.Variance

5.Standard Deviation

Range :

- Range will give how a data distributed from starting point to ending point.

- Let us , consider student marks starting with 1 mark and ending with 99 marks.

- so that data is flowing from 1mark to 99marks.

- lowest = 1 , Highest = 99

- **Range = high – low**
  = 99 – 1
  = 98

Therefore , the Range of data flowing from 1mark to 99marks is '98'.

Drawback :

- It will not consider middle values.

Mean Deviation :

- Mean Deviation means how much an observation is deviated from mean point.

- Assume that , we have points 1 , 2 , 3 , 4 , 5.

- Mean value $\bar{x}$ = 3

Let $x_1 = 1 , x_2 = 2, x_3 = 3, x_4 = 4 , x_5 = 5$

- The deviation of x_1 from $\bar{x}$ = $(x_1 - \bar{x})$ = 1 – 3 = -2
- The deviation of x_2 from $\bar{x}$ = $(x_2 - \bar{x})$ = 2 – 3 = -1
- The deviation of x_3 from $\bar{x}$ = $(x_3 - \bar{x})$ = 3 – 3 = 0
- The deviation of x_4 from $\bar{x}$ = $(x_4 - \bar{x})$ = 4 – 3 = 1
- The deviation of x_5 from $\bar{x}$ = $(x_5 - \bar{x})$ = 5 – 3 = 2
- - 2 of x_1 has 2 units below from the mean point
- 2 of x_5 has 2 units ahead from the mean point

$$\text{Total deviation} = \frac{\text{sum of all deviations}}{\text{Total no . of deviations}}$$

- sum of all deviations = $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x})$

- Total number of deviations = 5

Total Deviation = $\frac{1}{5} \times \sum_{i=1}^{5} (x_i - \bar{x})$

- if there are N points

Mean deviation = $\boldsymbol{\frac{1}{N} \times \sum_{i=1}^{N} (x_i - \bar{x})}$

Drawback :

- we are seeing individual observations has deviation
- But , when we add all the deviations it might becomes **'zero'**
- $= \frac{1}{5} \times [(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x})]$

  $= \frac{1}{5} \times [-2 - 1 + 0 + 1 + 2] = 0$

- Here , we are seeing the total deviation is 0(zero).
- Maths fails.

Absolute Mean Deviation :

- Doing Module to the 'X'
- Where X becomes |X|
  - If X = -2 becomes |-2| = 2
- Absolute Mean deviation shows how much a point is deviated from Mean.
- Assume that , 1,2,3,4,5 are the values
  - Mean value $\bar{x}$ = 3

Let $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5$

- The deviation of x_1 from $\bar{x}$ = $|(x_1 - \bar{x})|$ = |1 – 3| = 2
- The deviation of x_2 from $\bar{x}$ = $|(x_2 - \bar{x})|$ = |2 – 3| = 1
- The deviation of x_3 from $\bar{x}$ = $|(x_3 - \bar{x})|$ = |3 – 3| = 0
- The deviation of x_4 from $\bar{x}$ = $|(x_4 - \bar{x})|$ = |4 – 3| = 1
- The deviation of x_5 from $\bar{x}$ = $|(x_5 - \bar{x})|$ = | 5 – 3| = 2

- $= \frac{1}{5} \times [|(x_1 - \bar{x})| + |(x_2 - \bar{x})| + |(x_3 - \bar{x})| + |(x_4 - \bar{x})| + |(x_5 - \bar{x})|]$

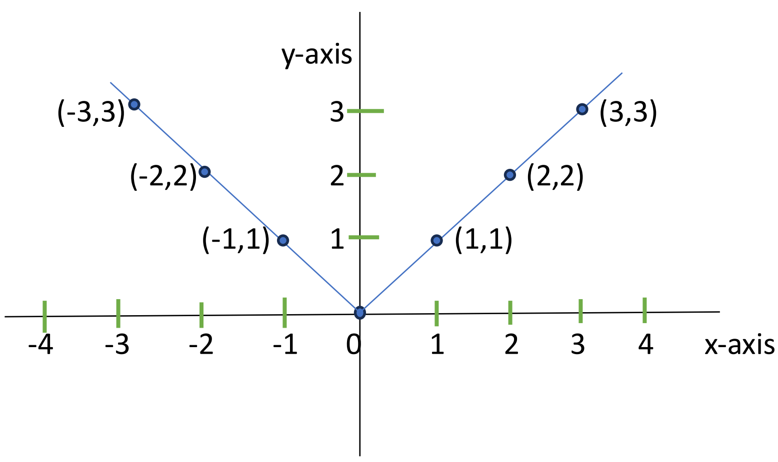
Total Deviation = $\frac{1}{5} \times \sum_{i=1}^{5} |(x_i - \bar{x})|$

- if there are N points

Absolute Mean deviation = $\boldsymbol{\frac{1}{N} \times \sum_{i=1}^{N} |(x_i - \bar{x})|}$

Drawback :

X	Y=|x|	points
-3	3	(-3,3)
-2	2	(-2,2)
-1	1	(-1,1)
0	0	(0,0)
1	1	(1,1)
2	2	(2,2)
3	3	(3,3)



- |x| : power = -1 ---> line as straight line : linear

Thumb rule :

- |x| is not continuous at point = 0 , we are stopping at zero.
    - At any math fails because of non continuous.
    - If you take differentiate of |x| = x / |x|.
    - X = 0 / |0| is indterminant (not defined).
- Continuous equations are good
- Differentiate is not possible on discontinuous curves / graphs.

Variance :

- Squaring the x ==> x becomes x^2.

- Assume that the values are 1 , 2 , 3 , 4 , 5.

- Mean value $\bar{x}$ = 3

Let $x_1 = 1 , x_2 = 2, x_3 = 3, x_4 = 4 , x_5 = 5$

- The deviation of x_1 from $\bar{x}$ = $(x_1 - \bar{x})^2$= $(1 – 3)^2$ = 4
- The deviation of x_2 from $\bar{x}$ = $(x_2 - \bar{x})^2$= $(2 – 3)^2$ = 1
- The deviation of x_3 from $\bar{x}$ = $(x_3 - \bar{x})^2$= $(3 – 3)^2$ = 0
- The deviation of x_4 from $\bar{x}$ = $(x_4 - \bar{x})^2$ = $(4 – 3)^2$ = 1
- The deviation of x_5 from $\bar{x}$ = $(x_5 - \bar{x})^2$ = $(5 – 3)^2$ = 4

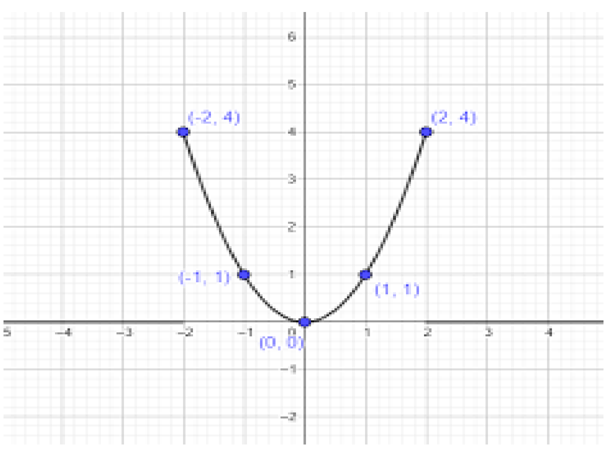
- $= \frac{1}{5} \times [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2]$

Total Deviation = $\frac{1}{5} \times \sum_{i=1}^{5} (x_i - \bar{x})^2$

Suppose we have N points.

Variance = $\frac{1}{N} \times \sum_{i=1}^{N} (x_i - \bar{x})^2$

X	Y= x^2	points
-3	9	(-3,9)
-2	4	(-2,4)
-1	1	(-1,1)
0	0	(0,0)
1	1	(1,1)
2	4	(2,4)
3	9	(3,9)



- X^2 : Power = 2 ===> Parabola : Non linear.
- Graph is parabola , which means Continuous.

Drawback :

- The values are increasing because of square terms.

- Not only values , but also units are increasing.

EXAMPLE:

Taking some distances 1km , 2km , 3km , 4km , 5km.

Distance (x)	$x - \bar{x}$	$(x - \bar{x})^2$
1km	1km – 3km = -2km	$(-2km)^2 = 4km^2$
2km	2km – 3km = -1km	$(-1km)^2 = 1km^2$
3km	3km – 3km = 0km	$(0km)^2 = 0km^2$
4km	4km – 3km = 1km	$(1km)^2 = 1km^2$
5km	5km – 3km = 2km	$(2km)^2 = 4km^2$
Average = 3km ($\bar{x}$)		$10km^2$

Variance = $\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})^2$

$= \frac{1}{5} \times$ [(4km^2 + 1km^2 + ^{2} + 0km^2 + 1km^2 + 4km^2)]

$= \frac{10\text{km}^2}{5}$ = 2km^2

- In the above calculation the variance = 2km^2.

- It is not only increasing the values , but also it is changing the units into square terms.

- we will not able to do proper Interpretation.

Standard Deviation :

- Standard deviation is denoted with 'σ'

$$\boldsymbol{\sigma = \sqrt{variance}}$$

Variance = $\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})^2$

Standard deviation $\boldsymbol{\sigma = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})^2}}$

- Let the Distance Hyderabad to Banglore = 600km

How much a data point is deviated from mean point.

Standard deviation $\sigma = \sqrt{2\text{km}^2}$ = 1.41km

Drawbacks in Data Dispersion :

Range : It will not consider the middle value.

Mean deviation : The sum of total deviations becomes zero.

Absolute Mean Deviation : |x| has discontinuous at zero.

Variance : The values are increasing and units are also increasing because of square.

Formulas covered :

- Range = high – low
- Mean deviation = $\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})$
- Absolute Mean deviation = $\frac{1}{N} \times \sum_{i=1}^{N} |(x_i - \bar{x})|$
- Variance = $\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})^2$
- Standard deviation $\sigma = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N}(x_i - \bar{x})^2}$