Statistics – Data Dispersion

Range:

- Range will give how data is distributed from starting point to ending point.
- It is the difference between the highest and lowest values in a data set.
- **Example:** Consider, students with marks: 1 mark is starting point and 99 marks is ending point
- So, that the data is flowing from 1 mark to 99 marks
- 1 = Lowest, 99 = Highest
- So, Range = Highest – Lowest = 99 – 1 = 98
- **Drawback**: - It will not consider middle values
- So, to consider, middle values we use mean deviation

Mean Deviation:

- It measures the average of the differences between each data point and the mean of the data set.
- It provides an idea of how spread out the values are around the mean.
- Where x_i are the data points, $\bar{x}$ is the mean, and *N* is the number of data points.
- **Example**: Assume that 1,2,3,4,5, are the values
- Mean value = $\bar{x}$ = 3
- $x_1 = 1,\ \ x_2 = 2,\ \ x_3 = 3, x_4 = 4,\ \ x_5 = 5$

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

- Mean deviation = $\frac{1}{N} * \sum_{i=1}^{N} (x_i - \bar{x})$
- Mean deviation = $\frac{1}{5} * (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x})$
- $\text{Mean deviation} = \frac{(1-3)+(2-3)+(3-3)+(4-3)+(5-3)}{5}$
- $\text{Mean deviation} = 0$ -------→ which is not correct
- **Drawback**: - We are seeing individual observations has deviations
- But when we add all the deviation it might becomes zero
- So, we have to use Absolute Mean Deviation

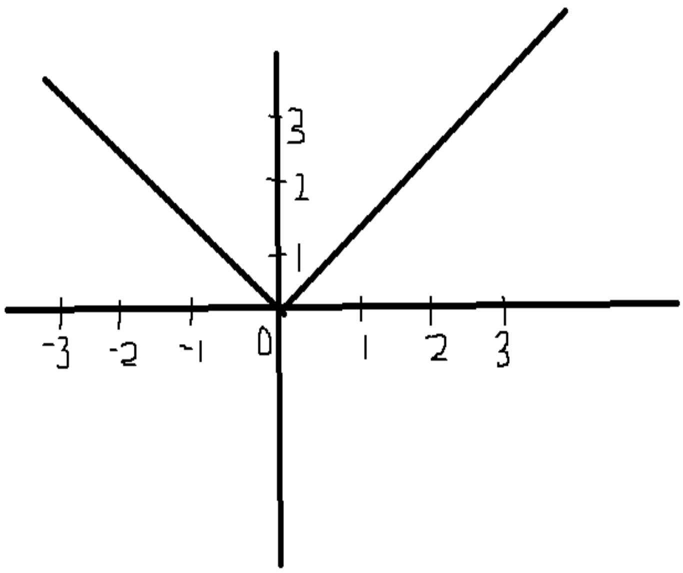
Absolute Mean Deviation:

- $x\ becomes\ |x|$
- The deviation of x_1 from mean $\bar{x}$ = $|x_1 - \bar{x}| = |1 - 3| = 2$
- The deviation of x_2 from mean $\bar{x}$ = $|x_2 - \bar{x}| = |2 - 3| = 1$
- The deviation of x_3 from mean $\bar{x}$ = $|x_3 - \bar{x}| = |3 - 3| = 0$
- The deviation of x_4 from mean $\bar{x}$ = $|x_4 - \bar{x}| = |4 - 3| = 1$
- The deviation of x_5 from mean $\bar{x}$ = $|x_5 - \bar{x}| = |5 - 3| = 2$
- Absolute Mean Deviation = $\frac{1}{N} * \sum_{i=1}^{N} |x_i - \bar{x}|$
- Absolute Mean deviation = $\frac{1}{5} * |x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + |x_4 - \bar{x}| + |x_5 - \bar{x}|$
- $\text{Absolute Mean deviation} = \frac{|1-3|+|2-3|+|3-3|+|4-3|+|5-3|}{5}$
- $\text{Absolute Mean deviation} = \frac{6}{5}$

Table for $|\bar{x}|$:

x	$\|\bar{x}\|$
-3	3
-2	2
-1	1
0	0
1	1
2	2
3	3

Graph for $|\bar{x}|$:



$|\bar{x}|: Power = 1 - -\rightarrow means\ it\ is\ a\ stright\ line == Linear$

Key Points:

- $|\bar{x}|$ graph is not continuous at point 0
- $It\ is\ stopping\ at\ 0$
- Any math equation fails here because it is non continuous
- If you take the differentiation of $|\bar{x}| = \frac{x}{|\bar{x}|}$, at $x = 0$
- It is called indeterminant form (Not defined)

Variance:

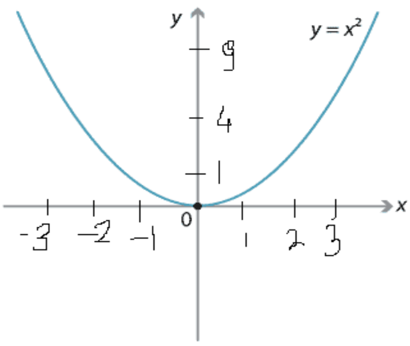
- $x\ becomes\ |x^2|$
- The deviation of x_1 from mean $\bar{x}$ = $(x_1 - \bar{x})^2 = (1-3)^2 = 4$
- The deviation of x_2 from mean $\bar{x}$ = $(x_2 - \bar{x})^2 = (2-3)^2 = 1$
- The deviation of x_3 from mean $\bar{x}$ = $(x_3 - \bar{x})^2 = (3-3)^2 = 0$
- The deviation of x_4 from mean $\bar{x}$ = $(x_4 - \bar{x})^2 = (4-3)^2 = 1$
- The deviation of x_5 from mean $\bar{x}$ = $(x_5 - \bar{x})^2 = (5-3)^2 = 4$
- Variance= $\frac{1}{5} * (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2$
- Variance= $\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5}$
- Suppose N numbers are there,

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

- Table for x^2:

x	x^2
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9

Graph for x^2:



$x^2 : Power = 2 --\rightarrow means\ it\ is\ not\ a\ stright\ line, it\ is\ a\ parabola == Non - Linear$

Drawback:

For example,

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

- Variance= $\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5}$
- Variance = $\frac{10}{5}$
- Variance = $2km^2$
- In above calculation, the Variance = $2km^2$
- **It is not only changing the values but also it is changing the units into square terms**
- We will not able to do proper interpretation

Standard Deviation:

- Denoted with ‘σ’
- Standard Deviation = σ = $\sqrt{Variance}$

- Standard Deviation = σ = $\sqrt{\frac{1}{N} * \sum_{i=1}^{N} (x_i - \bar{x})^2}$
- Standard Deviation means how much a datapoint is deviated from the mean point.
- We have to avoid negative values
- From previous problem calculation, Variance = 2
- Standard Deviation = σ = $\sqrt{Variance}$
- Standard Deviation = σ = $\sqrt{2}$
- Standard Deviation = σ = 1.414
- Standard Deviation is always less than Variance

Important Formulas:

1) Range = Highest – Lowest
2) Mean deviation = $\frac{1}{N}*\sum_{i=1}^{N}(x_i - \bar{x})$
3) Absolute Mean Deviation = $\frac{1}{N}*\sum_{i=1}^{N}|x_i - \bar{x}|$
4) Variance = = $\frac{1}{N}*\sum_{i=1}^{N}(x_i - \overline{x})^2$
5) Standard Deviation = σ = $\sqrt{\frac{1}{N} * \sum_{i=1}^{N}(x_i - \overline{x})^2}$