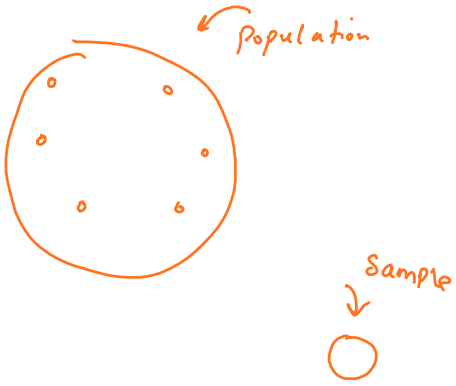
17 June 2023 15:30

Recap :-

- Statistics Concept
- Mean / Median / Mode
- Population and Sample
- Standard deviation
- Variance

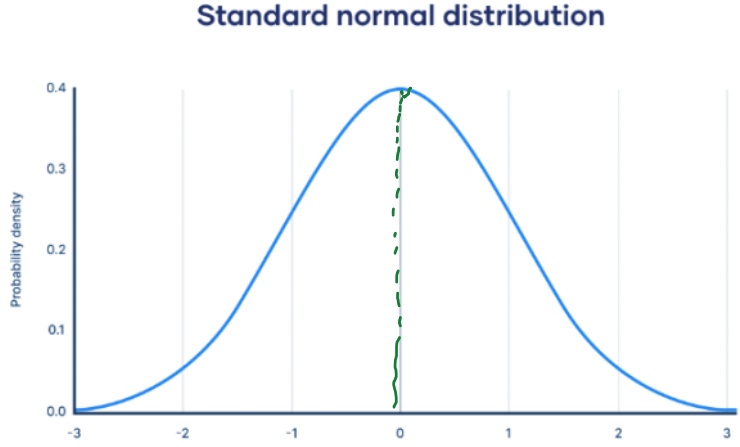


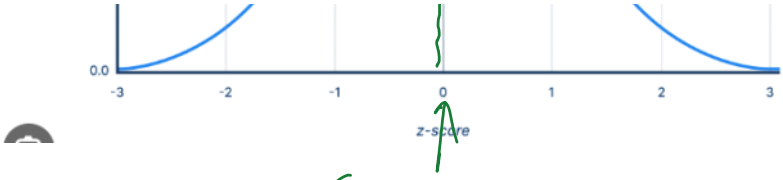
Agenda :-

- Normal distribution
- Z-Score
- Hypothesis testing

Normal distribution

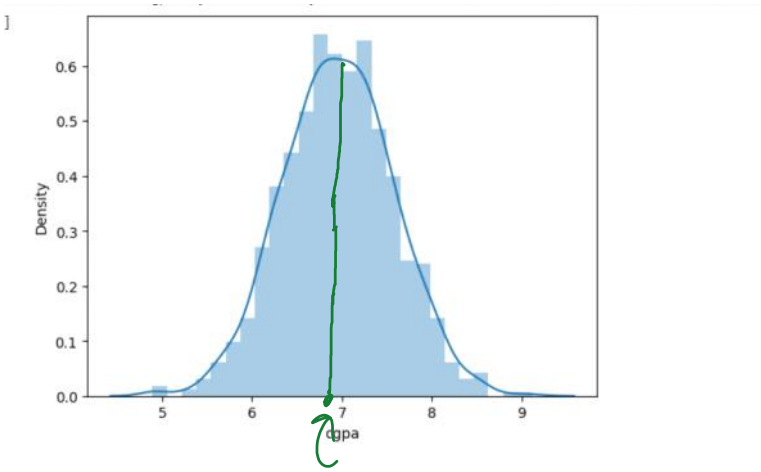
* It is also called as Gaussian Distribution. (Bell Curve)





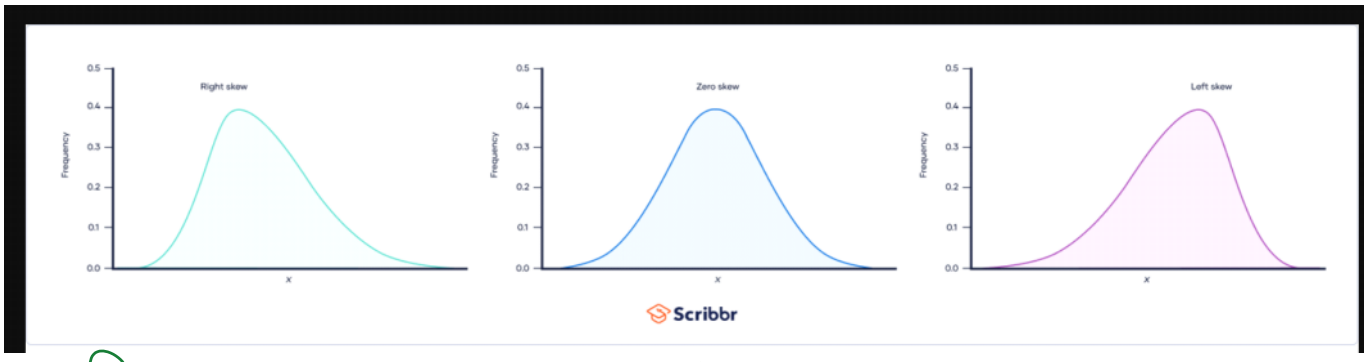
(Mean Value)

* if a column follows Normal distribution it Means most of the data point are Close to the MEAN Value of that column.



Mean

17 June 2023 16:06

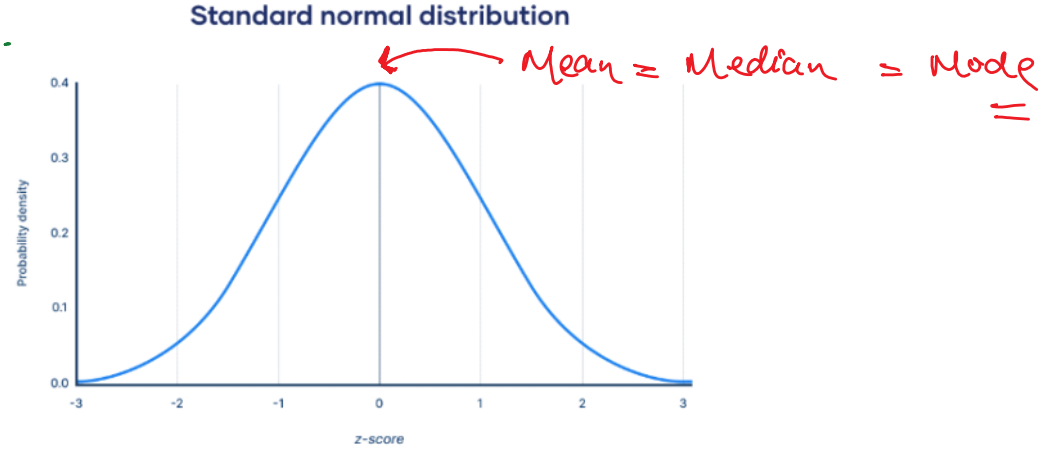


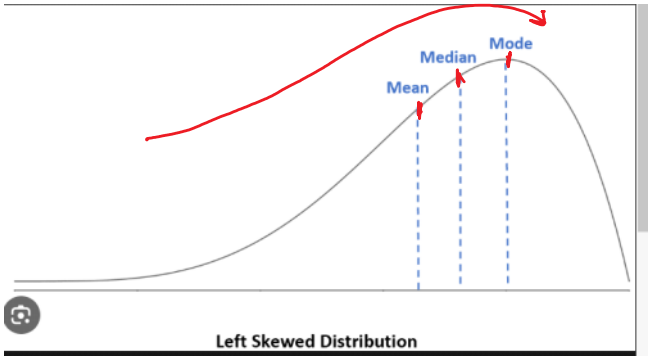
Right skew
(+ve skew)

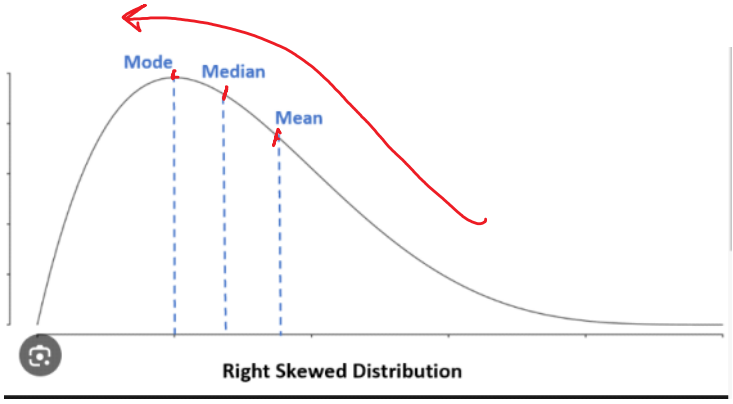
↑
Normal distribution

left skew
(-ve skew)

Note:

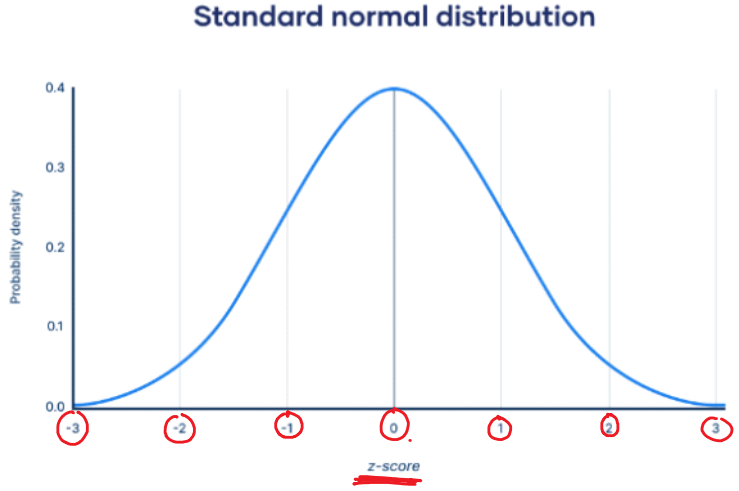






Z-Score :- We use this concept to determine that, how many standard deviation a point is far from Mean value in normal distribution

$$\text{Z-Score} = \frac{X_i - \text{Mean}}{\text{std}}$$

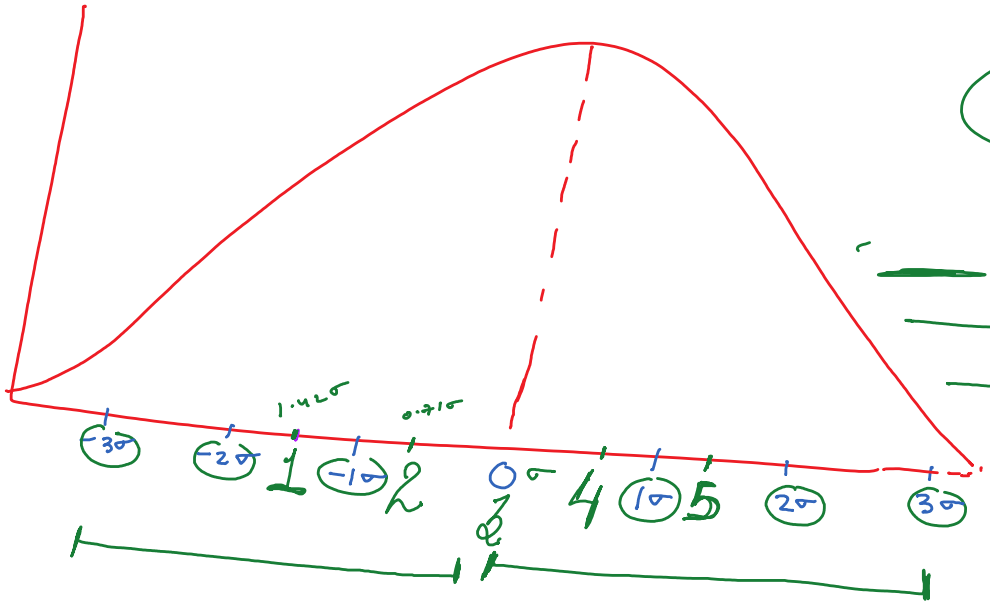


$X = [1, 2, 3, 4, 5]$

Std = 1.414

Mean = 3

[1, 2, 3, 4, 5]



$$\text{Z-Score}(1) = \frac{1-3}{1.414} = -1.42 \text{ std}$$

$$\text{Z-Score}(2) = \frac{2-3}{1.414} = -0.71 \text{ std}$$

$$\text{Z-Score}(3) = \frac{3-3}{1.414} = 0 \text{ std}$$

$$\text{Z-Score}(4) = \frac{4-3}{1.414} = 0.71 \text{ std}$$

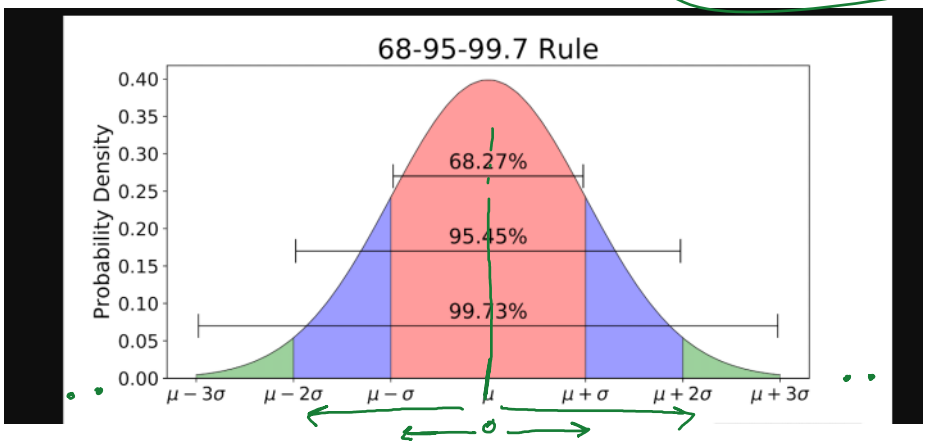
$$\text{Z-Score}(5) = \frac{5-3}{1.414} = 1.42 \text{ std}$$

→ Number Scale

17 June 2023 17:08

Empirical formula:

100%



0

68%

+2 +2

95%

99.7%

age:
12
15
9
7
15

- 750

1900

* It is also called as 68-95-99.7% Rule

* It says if a graph follows Normal distribution then-

68% falls under 1st std

95% falls under 2nd std

99.7% falls under 3rd std.

So- 0.3% data are Outliers