Box plot (IQR)

This Graph is used find the Outliers

3/11/25

Different types of Distributions:

→ To understand Data patterns.
→ To summarize the data Easily.
→ To calculate the probabilities.
→ To make prediction and Decision.
→ To choose right statistical test.

There are 2 categories of distribution:-

→ Continous Distribution [Numerical Distribution]
→ Discreate Distribution. [categorical ,,]

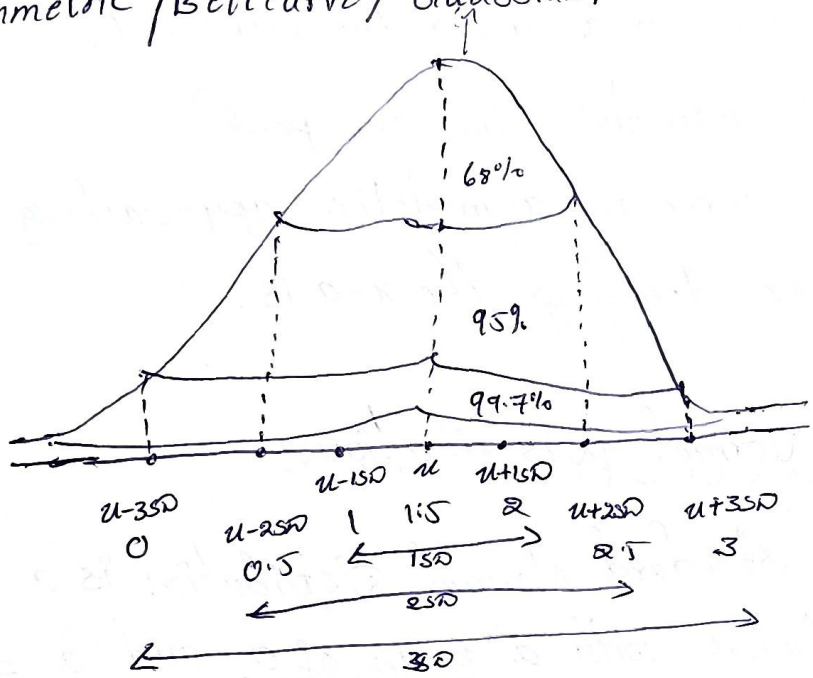
Top 5 Distributions:

1) Normal Distribution
2) Standard Normal Distribution } CD [Numerical]
3) Bernoulli Distribution
4) Binomial Distribution
5) Poisson Distribution } DD [categorical]

1) Normal Distribution:-

→ Mean = median = mode

symmetric / Bell curve / Gaussian / Normal Distribution



$u = 1.5$

$S = 0.5$

[68%-95%-99.7%] are called as Confidence Interval.

Empherical Rules:- SD = standard Deviation, u = mean

68% of data will present in 1SD.

95% of data will present in 2SD

99.7% of data will present in 3SD

Defination:-

→ In a normal Distribution, the mean, median and mode are Equal. The total area under the curve should be Equal to 1. The Normally distributed curve should be symmetric at the centre.

Properties :-

→ It is symmetric around its center
→ mean, median, and mode are Equal
→ The total area under the curve is 1.
→ It is unimodal (has one peak)
→ the curve is asymptotic, approaching but never touching the x-axis.

Standard Normal Distribution :-

The Standard Normal Distribution is a specific normal distribution with a mean of 0 and a SD of 1 also known as the z-distribution.

$\mu = 0$
$SD(\sigma) = 1$

Formula :-

$$z\text{score} = \frac{x_i - \mu}{\sigma}$$

→ Any normal distribution can be converted to the standard normal distribution by calculating its z-scores

Key characteristics :-

- mean (μ) : 0
- SD(σ) : 1
- Shape :- A symmetric, bell-shaped curve, with peak at the mean (0)
- z-scores :- A z-scores represents a specific values distance from the mean, measured in SD.

→ A +ve z-score means the value above the mean.

→ A -ve score means it's below the mean.

→ A score of zero means the value is Equal to the mean.

Ex:-

Normal dist Data	standard Normal dist Data
2	-0.9
7	1.57
5	0.57
4	0.07
1	-1.43
3	-0.43
5	0.57

$$\text{mean} = \frac{\text{Sum of obs}}{\text{No. of obs}} = \frac{2+7+5+4+1+3+5}{7}$$

$$\mu = \frac{27}{7} = 3.86$$

standard deviation (σ) = 2.

$$\text{z-score} = \frac{x_i - \mu}{\sigma}$$

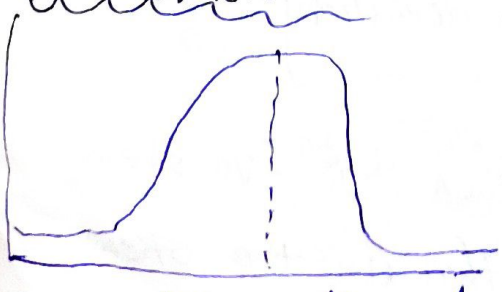
$$= \frac{2 - 3.86}{2} = -0.9$$

$$\text{z-score} = \frac{7 - 3.86}{2} = 1.57$$

$$\mu = 0.02 = 0$$

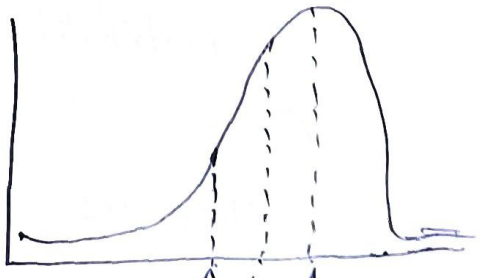
$$\sigma = 0.93 = 1$$

Normal distribution



mean = median = mode

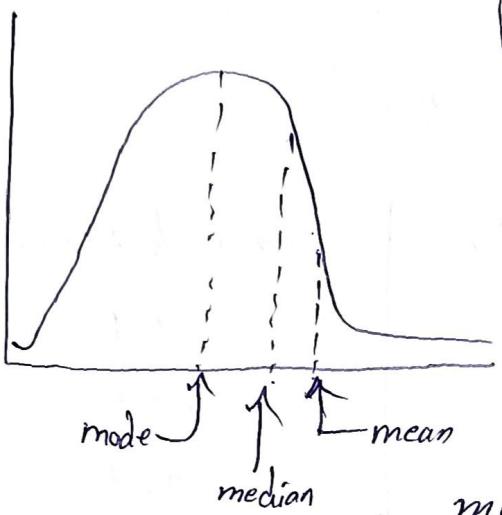
symmetric



Skewed Left (Negatively)

mean < median

Dist, skew are all these used for individual purpose



mean > median

Skewed Right
[Positively]

→ Positive Skew (Right skew):-

Tail on the right side is longer, most data are on the left side.

→ Negative Skew (Left skew):-

Tail on the left side is longer, most data are on the right side.

→ Zero skew (symmetric):-

The data is evenly distributed around the Mean (like a Normal distribution).

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

If value is near to -1 then it is -ve skew

If value is near to +1 then it is +ve skew

If value is lies in -0.5 to 0.5 then it is zero skew