Five Number Summary

Minimum | 1st quartile | Median | 3rd quartile | Maximum

1st quartile ↓ 75 Percentile

10 11 12 [25] 27 31 [33] 34 34 35 [36] 43 50 59

3rd Quartile

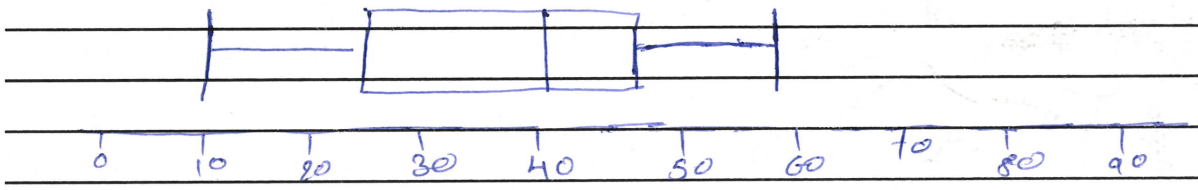
Position Median = Mid Value from data

Position of 1st quartile = $\frac{n+1}{2}$

$= \frac{7+1}{2} = [4]$

Position of 3rd quartile = $\frac{n+1}{2}$

$= \frac{7+1}{2} = [4]$



020

Day 4 - Stats :-

1. Central limit Theorem
2. Probability
3. Permutation and Combination
4. Covariance, Pearson Correlation, Spearman Rank Correlation
5. Bernoulli Distribution
6. Binomial Distribution
7. Power Law pareto Distribution
8. Histogram (pdf, pmf, cdf, univariant, bivariant)

① Central limit theorem

Population data → Gaussian/normal distribution

← Lognormal distribution

← Right log skewed

Sample size $n \geq 30$

Whether population data is normal/Gaussian distributed or not, but if we take sample data size of ≥ 30 & takes mean of all the sample data & plot in the histogram we will getting a gaussian/normal distribution.

② Probability

Probability is a measure of the likelihood of an event.

Ankur 8076153944
Niraj mada 9911913297 160

(A, B, C, D, E, F, G, H, I, J)

Arrang this 10 letters in (taken 4 at a time) into 4 boxes

Permutation $= {}^{n}P_r$ $\quad n = 10,\ r = 4$ $= \frac{n!}{(n-r)!} = \frac{n!}{(n-r)!}$

$= \frac{10 \times 9 \times 8 \times 7 \times 6!}{(10-4)!}$

$= \frac{10 \times 9 \times 8 \times 7 \times 6!}{6!}$

$= 5040$ ∴ ways we can arrange

CLT Population data is normally distributed or not

→ log normal distribution

→ Right skewed normal dist.

→ normal dist / Gaussian dist.

pop. sample. mean ($\bar{x}$) $\bar{x}$ = at least 30 with reapl.

① 5 + 12 + 7 + 9 + 11 / 5 = 8.8

② 2 + 8 + 9 + 11 + 3 / 5 = 27/5 = 5.4

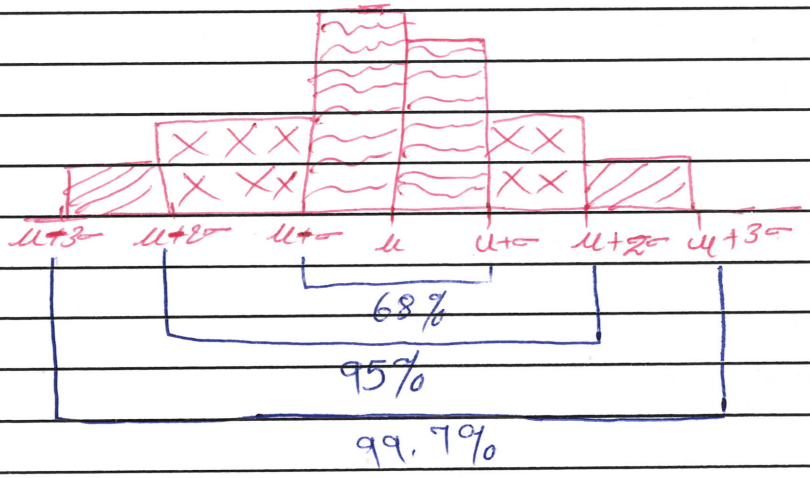
(i) Normal distribution
(ii) Standard normal distribution
(iii) Z-score

① Normal / Gaussian distribution

Every distribution create a bell curve distribution with their present data called normal or Gaussian distribution.

② Emperical rule

* Emperical distribution for normal distribution.



③ Q.Q plot

* Q-Q plot distribution is Gaussian or not.

④ Standard normal distribution

ex. x is Gaussian distribution (μ, σ)

$\Downarrow$

y would be is SND $(\mu = 0, \sigma = 1)$

when $n = 1$ $\left(Z\text{score} = \frac{x_1 - \mu}{\sigma}\right)$

standard error = $\frac{\sigma}{\sqrt{n}}$ (Inferential statistics)

over fitting & underfitting

When our training

* Training dataset gives high accuracy & test dataset gives low accuracy called over fitting.

Low accuracy in Training dataset & low accuracy in test dataset called ~~low~~ accuracy. under fitting