

STATISTICS

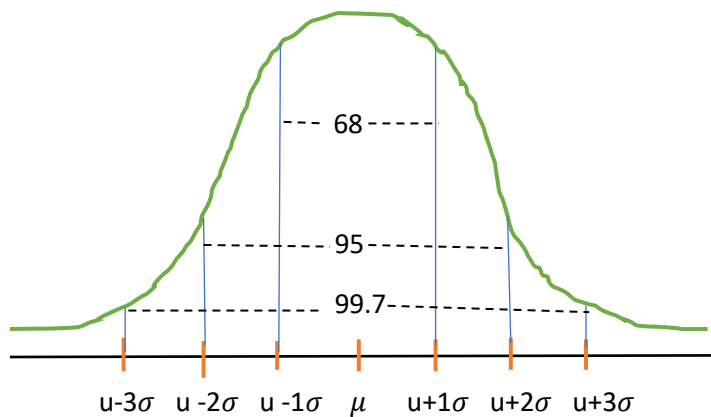
(SESSION-5)

EMPIRICAL RULE : (68 – 95 – 99.7)

Standard deviation = σ

Population mean = μ

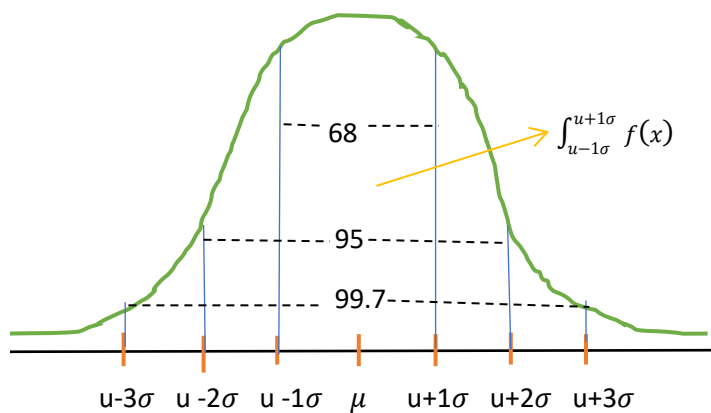
➤ If data follows the normal distribution :



- There is 68% of data covered between : $u - 1\sigma$ to $u + 1\sigma$
- There is 95% of data covered between : $u - 2\sigma$ to $u + 2\sigma$
- There is 99.7% of data covered between : $u - 3\sigma$ to $u + 3\sigma$
- The Maximum data coverage will happen between -3σ to $+3\sigma$ only

➤ Area under the curve :

➤ $\int_{u-1\sigma}^{u+1\sigma} f(x)$



❖ $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-u}{\sigma}\right)^2 \frac{1}{2}}$

- Assume that in India petrol rates are on of average 100rs , It will be vary state to state by 2rs.

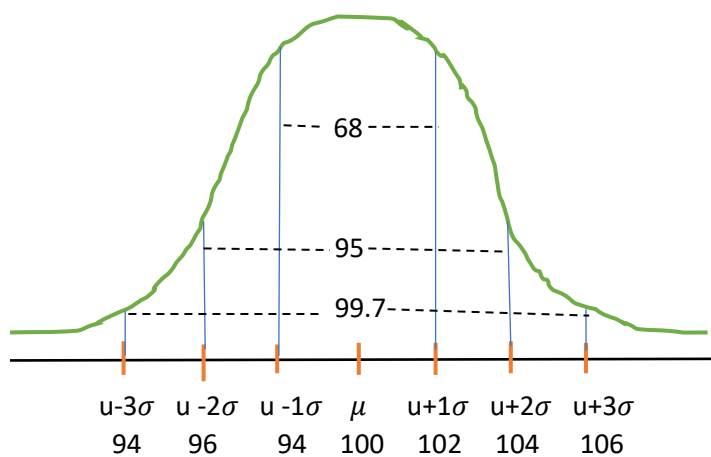
From empirical rule :

$$\mu = 100, \sigma = 2$$

- $\mu - 1\sigma$ to $\mu + 1\sigma = 100 - 1(2)$ to $100 + 1(2)$
 $= 100 - 2$ to $100 + 2$
 $= 98$ to 102
- $\mu - 2\sigma$ to $\mu + 2\sigma = 100 - 2(2)$ to $100 + 2(2)$
 $= 100 - 4$ to $100 + 4$
 $= 96$ to 104
- $\mu - 3\sigma$ to $\mu + 3\sigma = 100 - 3(2)$ to $100 + 3(2)$
 $= 100 - 6$ to $100 + 6$
 $= 94$ to 106

Where ,

- $\mu - 1\sigma$ to $\mu + 1\sigma$ is 68%
- $\mu - 2\sigma$ to $\mu + 2\sigma$ is 95%
- $\mu - 3\sigma$ to $\mu + 3\sigma$ is 99.7%



- In India 68% of states having petrol rates between 98rs to 102rs.
- In India 95% of states having petrol rates between 96rs to 104rs.
- In India 99.7% of states having petrol rates between 94rs to 106rs.
 - Which means the Minimum petrol rates in India is 94rs
 - The maximum petrol rates in India is 106rs

❖ Empirical rule is used only if data follows Normal distribution.

- if data doesnot follow normal distribution , Then we use the chebyshev's rule.

Chebyshev’S Inequality :

- Chebyshev’s inequality is used , If data does not follows Normal Distribution.

❖ The percentage of data coverage is $\left(1 - \frac{1}{k^2}\right)$

- For Example , If data does not follows the Normal Distribution or Normality and you want to find out how much data coverage happens between 2 standard deviation.

- Where K=2

$\left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{2^2}\right) = \left(1 - \frac{1}{4}\right) = \frac{3}{4} = 75\%$

Therefore , Data coverage between 2 standard deviations is 75%.

Normal (Empeirical)	No Normality (Chebyshev’s) $\left(1 - \frac{1}{k^2}\right)$
u -1σ to u +1σ = 68%	u -1σ to u +1σ = $\left(1 - \frac{1}{1^2}\right) = 1 - 1 = 0$ = Not Valid
u -2σ to u +3σ = 95%	u -2σ to u +3σ = $\left(1 - \frac{1}{2^2}\right) = \left(1 - \frac{1}{4}\right) = \frac{3}{4}$ = 75%
u -3σ to u +3σ = 99.7%	u -3σ to u +3σ = $\left(1 - \frac{1}{3^2}\right) = \left(1 - \frac{1}{9}\right) = \frac{8}{9}$ = 90%