Agenda

1. Query → (n+1) or (n) → Box plot ✓ { by n+1 }
2. Github → Submit the Assignments { Bench's correction }
3. Super way of new support (Monday)
4. Normal Distribution, Standard Normal Distribution, Z-score

5 Number Summary

50 percentile.

18

1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

2.5 (25 percentile) 5 6.5 (75 percentile)

Median

25 percentile

$= \frac{25}{100} \times (n+1) = \frac{25}{100} \times 19 = 4.75 \Rightarrow$ Index

Lower fence ✗

$= \frac{75}{100} \times (19) = 14.25 \Rightarrow$ Index $= 6.5$

$= \frac{25}{100} \times (18) \boxed{4.5}$

{ No need to apply formula }

[2, 3, 4, 6, 6, 6, 8, 9, 11, 13, 14] — Q3

50 percentile

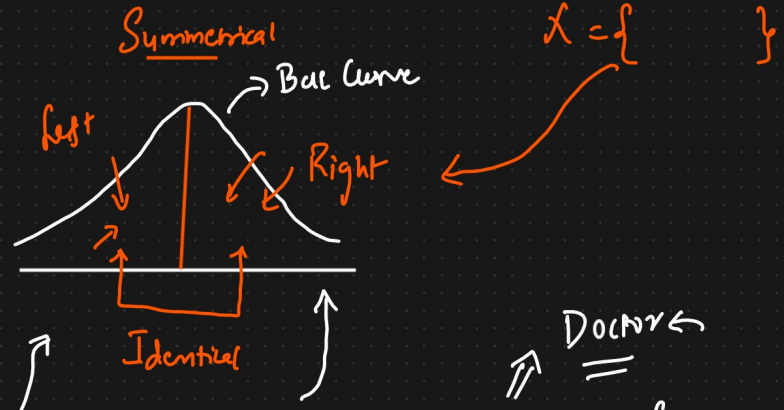
25 percentile 6 => 50th percentile

96 percentile 98.5

99

$\left\{ \frac{99}{100} \times (n+1) \right\}$

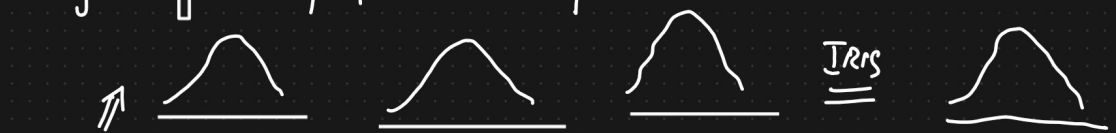
⑧ Gaussian or Normal Distribution



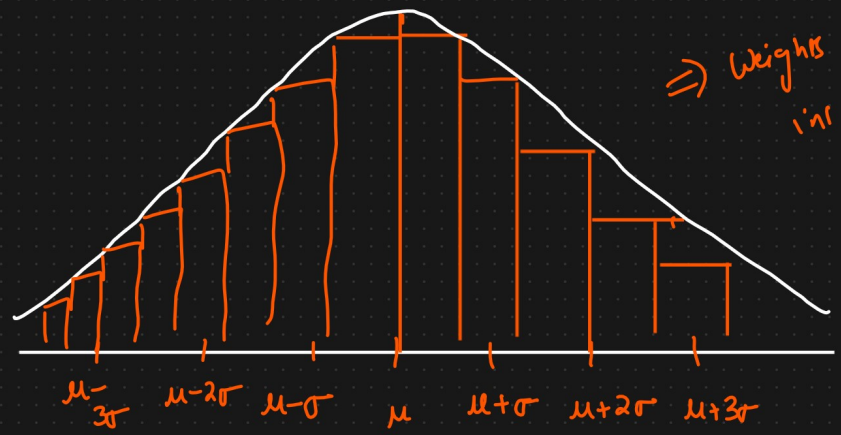
{ IRIS Flower Dataset }

Eg: Weights of the people in a city? ⇐ Doctors ⇐ { Domain Expertise }

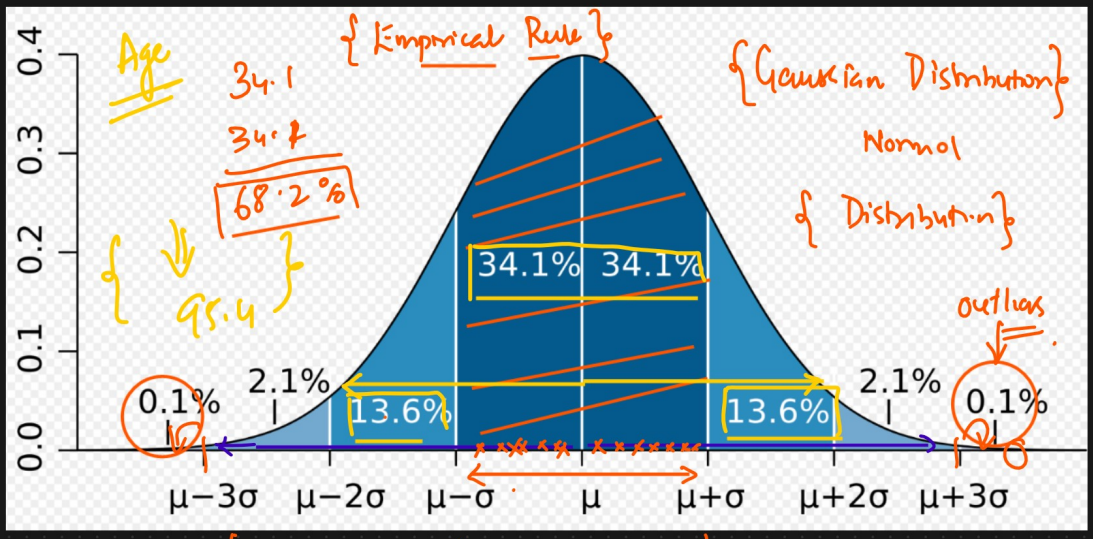
Heights of the people in a city?



Sepal length	Sepal width	Petal length	Petal width	Species
—	—	—	—	Setosa
				Virginica
				Versicolor



=> Weights of the people in the city



$X = \{2, 4, 8, 10, 12 \; \ldots \; 100\}$ → Gaussian Distribution

⇓ 68.2%

⇓ 68.2%

95.4%

1
68.2
13.6
13.6
95.4
2.1
2.1
99.6

[68.2 − 95.4 − 99.7] ⇒ Empirical Formula

[68.2, 95.4, 99.7]

① $\{1, 2, 3, 4, 5\}$ ⇒ Normal Distribution

-1

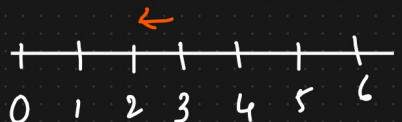
$\{1, 2, 3, 4, 5\}$

+1

1.414

$\mu = \frac{1+2+3+4+5}{5} = 3$

$\sigma = 1.58 \approx \sigma = 1$



Standard Normal Distribution

$X \sim$ Gaussian Distribution (μ, σ)

$\Downarrow \Downarrow$

$Y \sim SND\ (\mu = 0, \sigma = 1)$

$X = \{1, 2, 3, 4, 5\}$

$\mu = 3$

$\sigma = 1$

$y = \{\ \}$

$Z\text{-score} = \dfrac{x_i - \mu}{\sigma}$

$\Rightarrow \sqrt{n}$

$\dfrac{x_i - \mu}{\sigma}$

$\{1, 2, 3, 4, 5\}$

$n = 1$

$\mu = 0$

$\sigma = 1$

$\{1, 2, 3, 4, 5\}$

$\mu = 3 \quad \sigma = 1$

$\dfrac{3-3}{1} = 0$

$\dfrac{1-3}{1} = -2$

$\dfrac{2-3}{1} = -1$

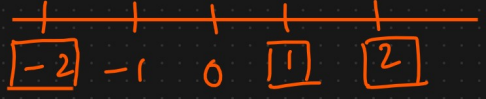
$\dfrac{4-3}{1} = 1$

$\{1, 2, 3, 4, 5\}$

$\Downarrow$

Z score

$\{-2, -1, 0, 1, 2\}$



$\bar{x}$ 10 min break

$n = 1$

$X = \{2, 3, 4, 5, 8\} \Rightarrow y \quad \mu = 0, \sigma = 1$

$\mu = ?$ 4.4

$\sigma = ?$ 2.05

$y = \{-1.17, -0.68, \quad \}$

population

N

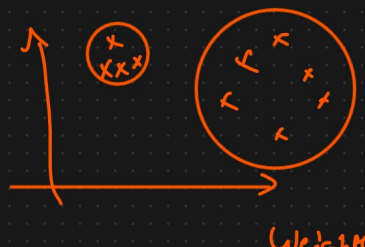
$Z\text{-score} = \dfrac{x - \mu}{\sigma}$

$\dfrac{3 - 4.4}{2.05}$

$\frac{2-4.4}{2.05} =$

$\mu=0, \sigma=1$

Scale down ↓ Same scale

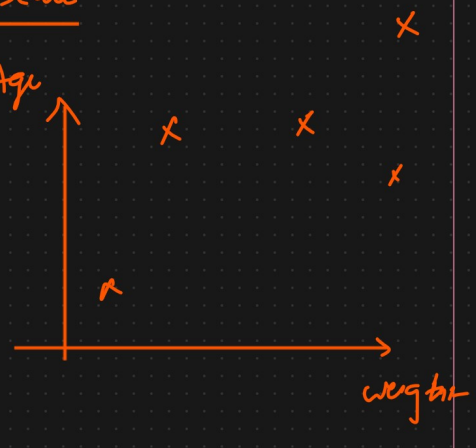


⇓

Why Standard Normal Distribution

(Years) ⟺ (Kg) ⟺ (Km/miles)

Age ∈ Weight ∈ Distance



X { 24 ⟷ 100, 26 ⟷ 120, 28 ⟷ 140, 30 ⟷ 160 } Z

Standardization

Y { $\mu=0, \sigma=1$ } Z' $(\mu=0, \sigma=1)$

Feature Scaling =

Z-score :

Z-test

$X \overset{\sim}{\sim}$ Normal Distribution (μ, σ)

⇓

$Y \overset{\sim}{\sim}$ SND $(\mu=0, \sigma=1)$.

⇓

{ Standardization }.

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

① Gaussian or Normal Distribution

② Empirical Formula

③ Standard Normal Distribution

Percentiles

$\frac{95}{100} \times (n+1)^{15}$

Practical

5 days

Average days

95 percentile

95

Z-score

Logistic Manager { Quickly actions }

{ # of days to deliver a panel } = { 1, 2, 2, 2, 3, 4, 5, 5, 5, 6, 8, 9, 10 }

20

14.25

So many

15

Outliers

{ 95 percentile, 99 percentile }

Quickly rectify ⇐ 20 days 95 percentile

Empirical Rule

Bell Curve

{ Influential stars }

40 50 Age

log [Normal Distribution]

A B

$X =$ $\Rightarrow \log_e(X) =$

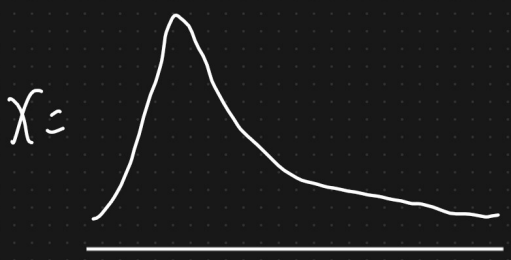
length of comments people writing in my channel

Salary of IT Employees

Elon Musk

Wealth distribution in the world

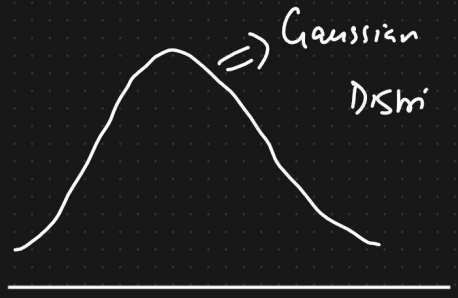
Log Normal Distribution



$\Rightarrow \log_e(x) =$

$\downarrow$

y

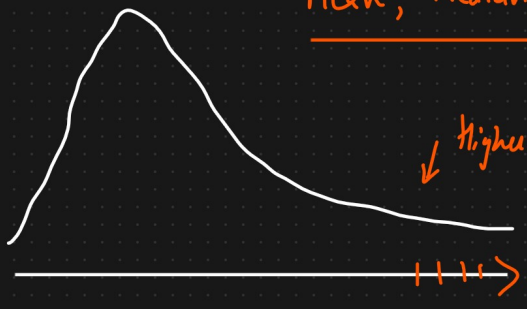


$\Rightarrow$ Gaussian Distri

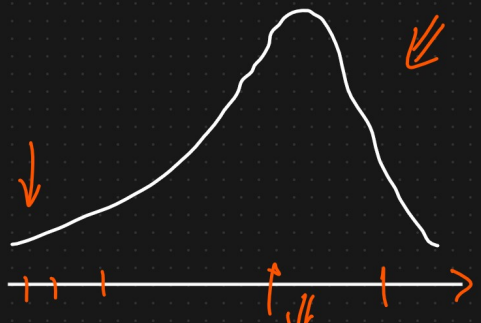
$x = \{ \quad \}$ $\leftarrow e^{(y)} \leftarrow$

Assignment

Mean, Median, Mode



Higher



IT people

Life span