Day 3 — Advance Statistics

① Distributions
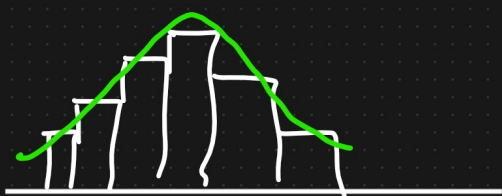- → Normal Distr
- → Standard Normal Distr
- → Z score
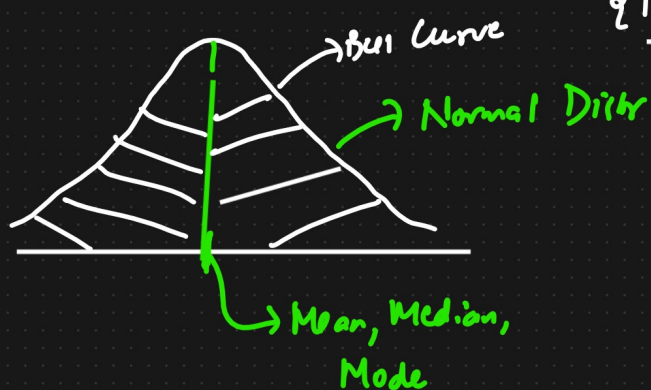- → Log Normal Distr
- → Bernoullis Distr
- → Binomial Distr

Practical
① Mean, Median, Mode
② Variance, Standard deviation
③ Histogram, pdf, Bar plot, Violin plot
④ IQR
⑤ Log Normal Distribution

① Distributions }

$Ages = \{ 24, 26, 27, 28, 30, 32 \ - \ - \ - \ . \ . \ . \ - \ - \ - \}$

⇓

① Gaussian/Normal Distribution

→ Bell Curve

→ Normal Distr

→ Mean, Median, Mode

Empirical Formula

68 - 95 - 99.7 % Rule

Dataset
{100 datapoints}

99.7%   68%
                95%

$\mu - 3\sigma$ $\mu - 2\sigma$ $\mu - \sigma$ μ 1std $\mu + \sigma$ $\mu + 2\sigma$ $\mu + 3\sigma$
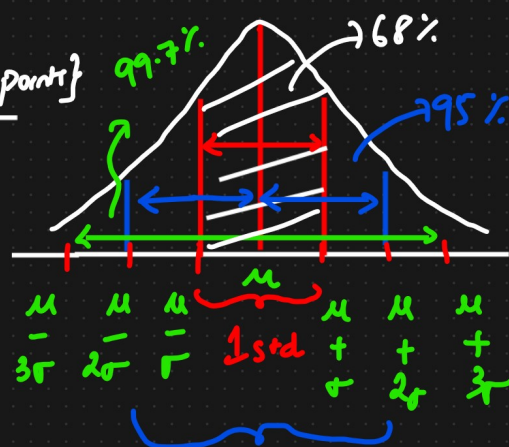
Eg : ① Height $\longrightarrow$ Normally Distributed

$\downarrow$

Domain Expert $\rightarrow \{ Doctor \}$

② Weight   ③ IRIS DATASET
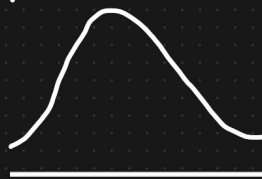
Eg : $\mu = 4$ $\sigma = 1$
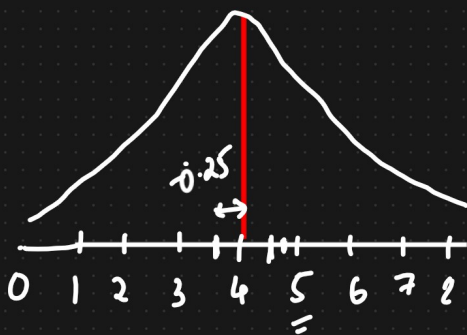
$\boxed{4.5} \color{red}{\longrightarrow}$ standard deviation

$+0.5$ sd

$\boxed{4.75}$??

0 1 2 3 4 5 6 7 8

$Z \text{ score} = \dfrac{x_i - \mu}{\sigma} \nearrow^{+ve}$

$= \dfrac{4.75 - 4}{1} = 0.75$ sd

3.75

$Z \text{ score} = \dfrac{3.75 - 4}{1} = -0.25$

$\mu = 4$
$\sigma = 1$
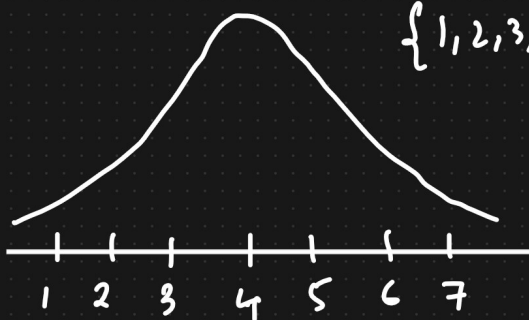
1 2 3 4 5 6 7

$\{1, 2, 3, 4, 5, 6, 7\}$

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

$\mu = 4$
$\sigma = 1$

$\{-3, -2, -1, 0, 1, 2, 3\}$

$Z(1) = \dfrac{1-4}{1} = -3$ $Z(3) = \dfrac{3-4}{1} = -1$ $Y \sim SND(\mu = 0, \sigma = 1)$

$Z(2) = \dfrac{2-4}{1} = -2$

$\{1, 2, 3, 4, 5, 6, 7\} \rightarrow$ Normal Distr

$\Downarrow$

$Z \text{ score}$

$\Downarrow$

Standard Normal Distr

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

$\{-3, -2, -1, 0, 1, 2, 3\}$ ← Satisfying this property

Practical Application

$\boxed{\mu = 0, \sigma = 1}$

$\Downarrow$

Standard Normal Distr

$\boxed{\text{Standardization}}$ ← Z score

DATASET

Age (years)	Salary (RS)	Weight (kg)
24	40K	70
25	80K	80
26	60K	55
27	70K	45

← Z score

Units

Normalization $\{\mu = 0, \sigma = 1\}$

$(-1 \text{ to } 1) \quad \hookrightarrow (0 \text{ to } 1)$

→ MinMax Scaler → $(0 \text{ to } 1)$

CNN → Image Classification → Divide by 255

Pixels   → 0-255

$\Downarrow\Downarrow$

MinMax Scaler

Normalization ↗ $\boxed{0-1}$

Practical Eg $\{$ India VS SA $\}$

① ODI Series   (CRICKET)

$\downarrow$ 2021

Series Average 2021 = 250

Standard Deviation = 10

Compared to both the series in which year Rishabh pant final

Series ← Team final score = 240   Score was better??

2020

Series Average 2020 = 260

Standard Deviation = 12

Team final score = 245

2021

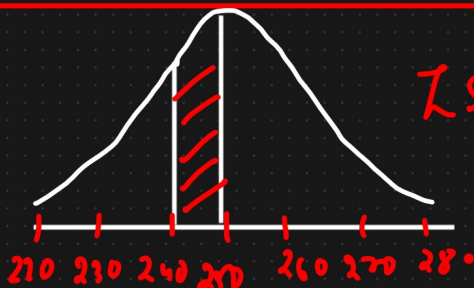$$Z \text{ Score} = \frac{x_i - \mu}{\sigma} = \frac{240 - 250}{10} = \frac{-10}{10} = -1 \Big\}$$

2020

$$Z \text{ Score} = \frac{x_i - \mu}{\sigma} = \frac{245 - 260}{12} = \frac{-15}{12} = -1.25 \Big\}$$

Final Match

In 2021 ✓

$$\mu = 250 \quad x_i = 240 \quad \sigma = 10$$

Z Score = −1

210 230 240 250 260 270 28o

In 2020 ✓

$$\mu = 260 \quad x_i = 245 \quad \sigma = 12$$

Z Score = −1.25

224 236 248 260 272 284 296

Assignment

Stats Interview Question

$X \rightarrow$ Score

Area of Body Curve

Body ↑   Jail ↗

0.5987

$\mu = 4$
$\sigma = 1$

$\frac{3}{7}$

Question? : What percentage of scores falls above 4.25?

1 − Left Area
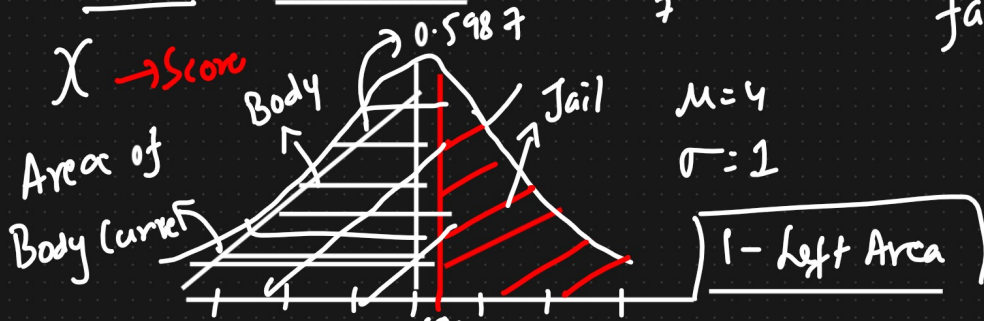
$\{ 1 \quad 2 \quad 3 \quad 4 \ 0.25 \quad 6 \quad 7 \}$ $\quad 1 - 0.5987$

$$\boxed{Z} = \frac{x_i - \mu}{\sigma} = \frac{4.25 - 4}{1} = 0.25 = 0.4013$$

$$\Downarrow 40\%$$

④ In India the average IQ is 100, with a standard deviation of 15. What percentage of the population would you expect to have an IQ lower than 85?

(Ans) $Z = \dfrac{85 - 100}{15} = \dfrac{-15}{15} = -1$

IQ 90   to   120

and

Body $1 - 0.84134$

55 70 85 100 115 130 145