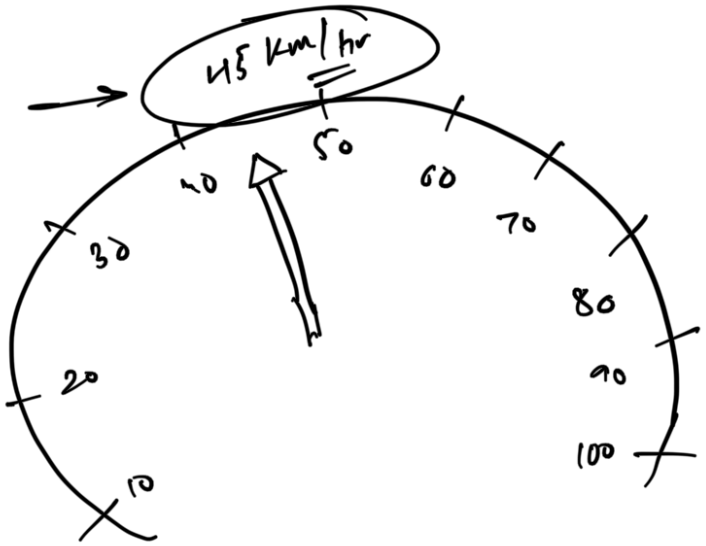
Descriptive Statistics

Descriptive → Describe

↑

statistics that simply describes features from a collection of information



← Speedometer

"Infer" something

↑

Inferential Statistics ← Draws conclusion

↑

Descriptive statistics → Summary statistics

↑ ↓ ↑

Summarises

Q: DS job at FAANG.

Sample salaries : 30L, 30L, 35L, 40L, 40L

Median → 35L

Salary Expectation : 35L ?

Avg

Outlier

New Med ⇒ 37.5

30L, 30L, 35L, 40L, 40L, 300L

→ New Mean → 179L

Median

old Median 35L

New Median 37.5L

① Sorted

No Mode

10, 20, 30, 40, 50, 60, 70, 80

② (7) → odd → $\left(\frac{n+1}{2}\right)^{th}$ term ⇒ $\frac{7+1}{2}$ = 4th term

(8) → Even → $\left(\frac{n}{2}\right)^{th}$ term + $\left(\frac{n}{2}+1\right)^{th}$ term

$$\frac{(\quad)^{th} + (\quad)^{th}}{2}$$

$$\rightarrow \frac{\left(\frac{8}{2}\right)^{th} \text{term} + \left(\frac{8}{2}+1\right)^{th} \text{term}}{2}$$

$$\Rightarrow \frac{4^{th} \text{term} + 5^{th} \text{term}}{2}$$

$$\Rightarrow \frac{40 + 50}{2} = \frac{90}{2} = 45$$ ✓

Mode

90, 90, 90, 80, 80, 45, 45, 20 → 90 (Mode)

(3)

2, 2, 5, 2, 3, 3, 3, 4 → 2 & 3

Mode

(Bi-modal)

Avg. Age → 24

$$\rightarrow \frac{20 + 22 + 28 + x}{4} = 24$$

$$\frac{70 + x}{4} = 24$$

$$70 + x = 96$$

$$x = 96 - 70 = 26$$

x ↓

Age of the 3rd person

Q: Median Age ⇒ 20, 22, 26, 28

$$\frac{22 + 26}{2} = 24$$

100

↳ 10 → 3

Q: Avg. weight → 40kgs.

(1) $\frac{x+y}{} = 40 \Rightarrow (x+y) = 80$

② $\frac{x+y+m}{3} = 45 \Rightarrow \boxed{(x+y)+m = 135}$

② $x+y + m = 135$

① $x+y = 80$

$m = 55$

Q:

30% — 0

40% — 1

10% — 2

20% — 3

→ x

$\left(\frac{30}{100} \times\right) \times (0) + \left(\frac{40}{100} \times\right) \times (1) + \left(\frac{10}{100}\right) \times 2 + \left(\frac{20}{100} \times\right) \times 3$

$\frac{100}{100} \times (x)$

$0 + 40 + 20 + 60$

100

$\frac{120}{100} = 1.2$

Weighted Average

GPA ?

Sub	Credit	Grade
EM	4	8
Phy	5	7
Thermo...	5	6
E Draw	4	5

$$\text{GPA} \Rightarrow \frac{4\times 8 + 5\times 7 + 5\times 6 + 4\times 5}{\text{Sum of weights } (4+5+5+4)}$$

$$\Rightarrow \frac{32 + 35 + 30 + 20}{18}$$

$$\Rightarrow 6.5$$

Mean, Median, Mode, Weighted Avg.

Range — Describing the overall spread of the data.

→ 30, 30, 35, 40, 40

Range → highest val — lowest value

→ 40 — 30 → 10

30, 30, 35, 40, 40, 300

New Range → 300 − 30 = 270

Inter Quartile Range

Percentile → what fraction of values are lesser than a specific no.

33rd percentile → 33% of employees have salary less than or equal to 30L

↓

20 Lacs

$$IQR = Q3 - Q1$$

$Q3$ → 75th percentile

$Q1$ → 25th percentile

1. Upper limit → $Q3 + 1.5 * IQR$
2. Lower limit → $Q1$ till $Q1 - 1.5 * IQR$

Cumulative Distribution function (CDF)

Die → 6 faces

$$P(\varepsilon = 3) = 1/6 \quad \times$$

$$P(\varepsilon \leq 3) = \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right)$$

$$= \frac{3}{x} = \frac{1}{2}$$

CDF $P(E \leq 1) = \frac{1}{6}$

CDF $P(E \leq 6) = 1$

$P(E \leq 0) = 0$

$P(E \leq 7) = 1$

$P(E \leq x_1) = y\_value[0]$

$= \quad " \quad 1$

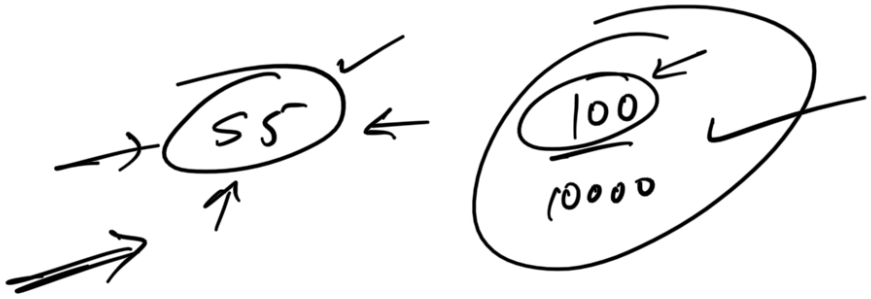
$(\quad x_2) \quad 2$

$(\quad x_3)$

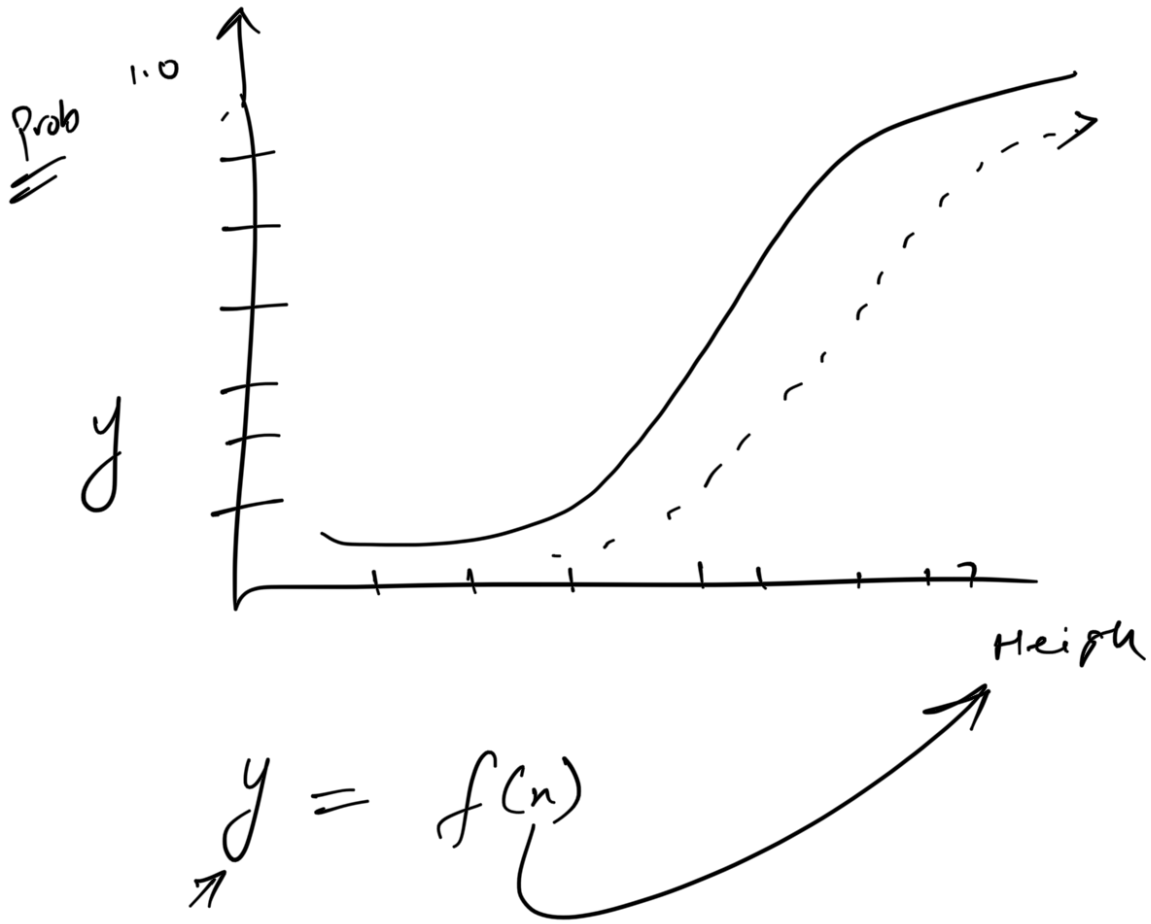
$(\quad x_4)$

$(\quad \vdots)$

$(\quad x_{100}) \quad 100$



CDF



$y = f(x)$

Inverse of percentile

Percentile → 25 → 63.5

CDF → 63.5 → 25