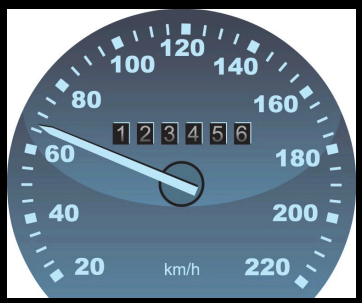
Descriptive Stats



Inferential Stats
"conclude"

What is this?
How is this relevant to today's class?

"central tendency"
"variability"
"percentile"

1) You are driving at 60 Km/hr
2) You will reach in one hour → infer
   ↳ don't call → infer traffic jam / accident

1) Vote share of candidate is 70% — actual poll
2) Exit poll → subset of data
   ↳ candidate may have 70% vote share
   ↳ conclusions

Central Tendency

Salary 30L, 30L, 35L, 40L, 40L, 3Cr

mean ("average") $\frac{30+30+35+40+40}{5} = 35$ Median = 35

mean $\frac{30+30+35+40+40+300}{6} = 79$ Median = 37.5

Median is more robust to outlies than mean

Computer (Python) should tell me that 3 Cr is an outlies

not common, atypical, rare "outlier"

There are 4 people whose average age is 24.
We know the age of three people: 20, 22, and 28.
What is the median age of these 4 people?

fourth person → `x'

$$\frac{20+22+28+x}{4} = 24 \longrightarrow x = 26$$

20, 22, 26, 28

median is 24

Weighted average

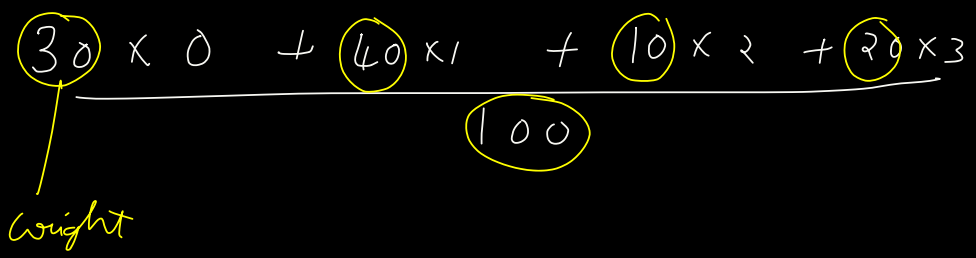
	Credits	GRADE	
Math	3	5	3×5 +
History	4	4	4×4 +
Chem	3	5	3×5 +
English	2	3	2×3

$$\frac{3 \times 5 + 4 \times 4 + 3 \times 5 + 2 \times 3}{(3 + 4 + 3 + 2)}$$

GPA

grade point average

A survey of number of pets in a town saw that -
30% people had 0 pets, 40% had 1 pet, 10% had 2 pets, 20% had 3 pets.
What is the average number of pets?



The mean weight of 2 children in a family is 40 Kgs.
If the weight of the mother is included, the mean becomes 45.
What is the weight of the mother?

$$\frac{x+y}{2} = 40 \qquad \frac{x+y+m}{3} = 45$$

$$x+y = 80 \qquad \frac{80+m}{3} = 45$$

$$m = 55$$

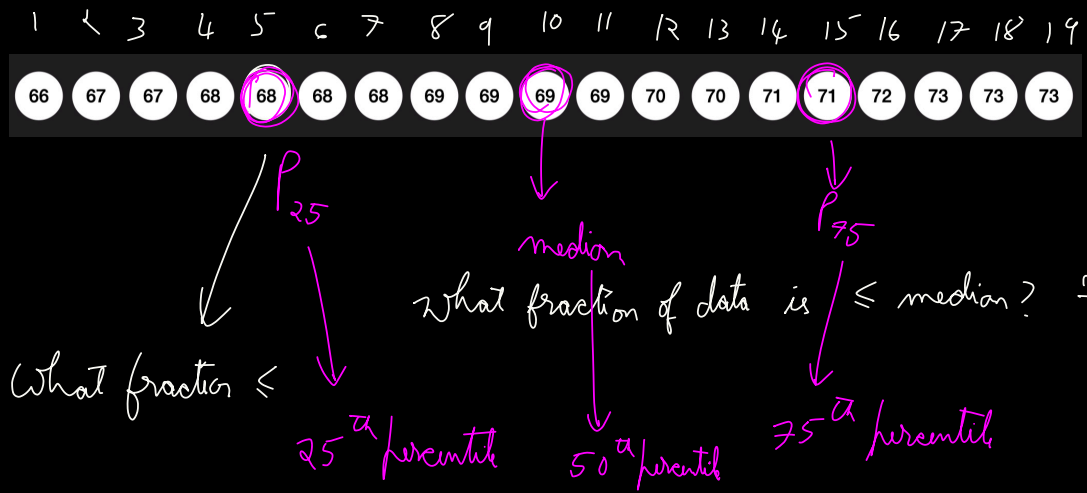
Range

Cricketer 20, 25, 60, 100

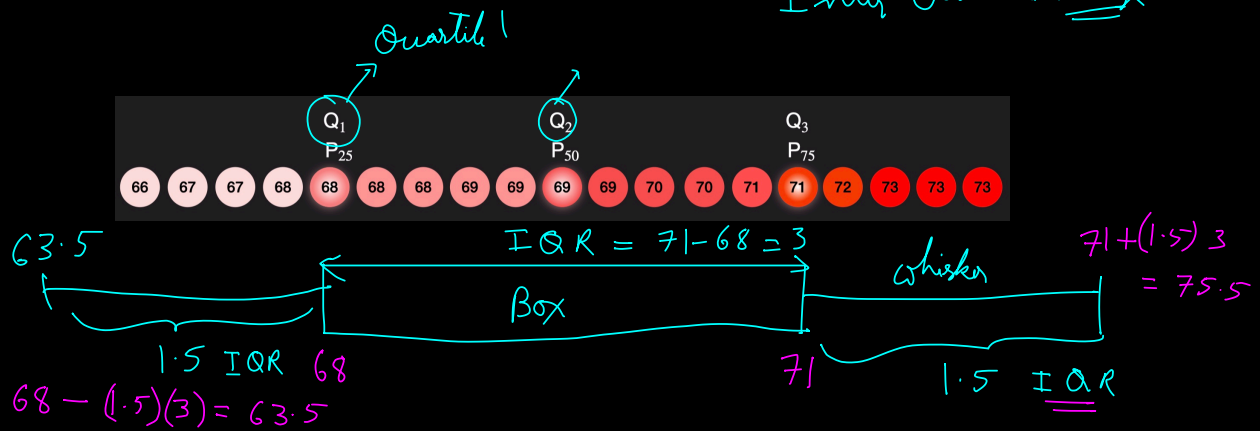
range = 100 − 20 = 80

Salary
① 30, 30, 35, 40, 40 range = 40 − 30 = 10
② 30, 30, 35, 40, 40, 300 range = 270

Range is NOT robust (stable) to outliers

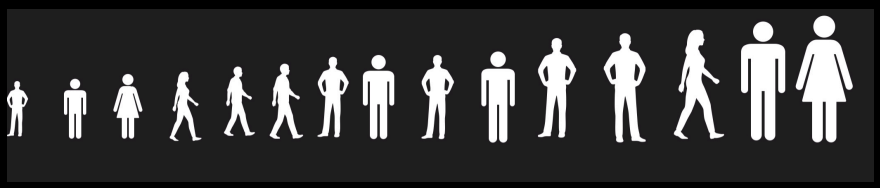

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
66 67 67 68 68 68 68 69 69 69 69 70 70 71 71 72 73 73 73
P_{25}
median
P_{75}
What fraction of data is ≤ median? ≃ 50%
What fraction ≤
25th percentile
50th percentile
75th percentile

Inter Quartile Range

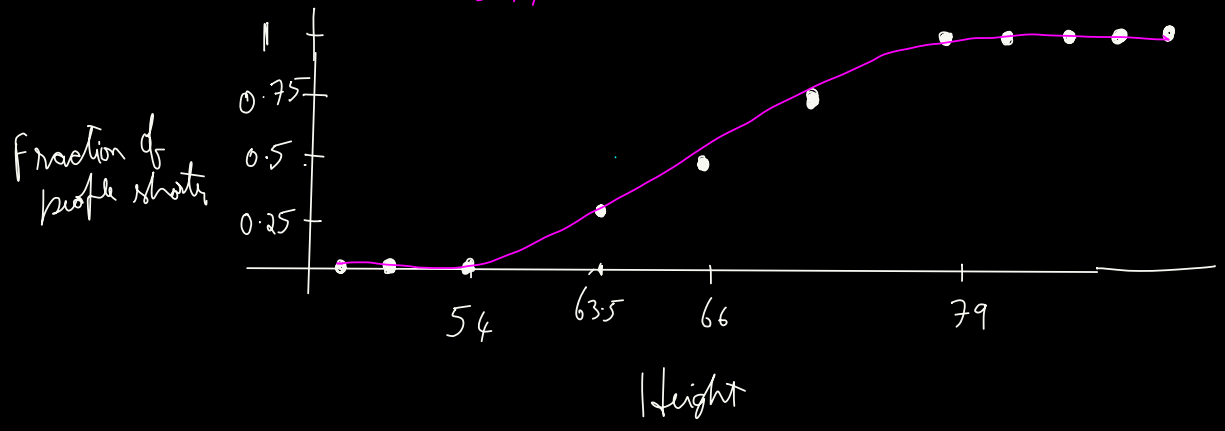


[63.5 to 75.5]

anything outside, we call outlier



54, 54.2, 54.2 79



Game : Betting

Droid → Bet how many runs

Rules you bet 34

If actual score is 44 → you lose 10 Rs

actual score is 24 → you lose 10 Rs

35 → 1 Rs

34 → win

(bet)

μ → mean (34)

x → actual score

$|x - \mu|$

Variance (Std dev) $x_1, x_2, x_3 \ldots x_{300}$ → all the score

$$Var = \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_{300}-\mu)^2}{300}$$

Std dev: $\sqrt{Var}$

captures how much you deviate from the mean