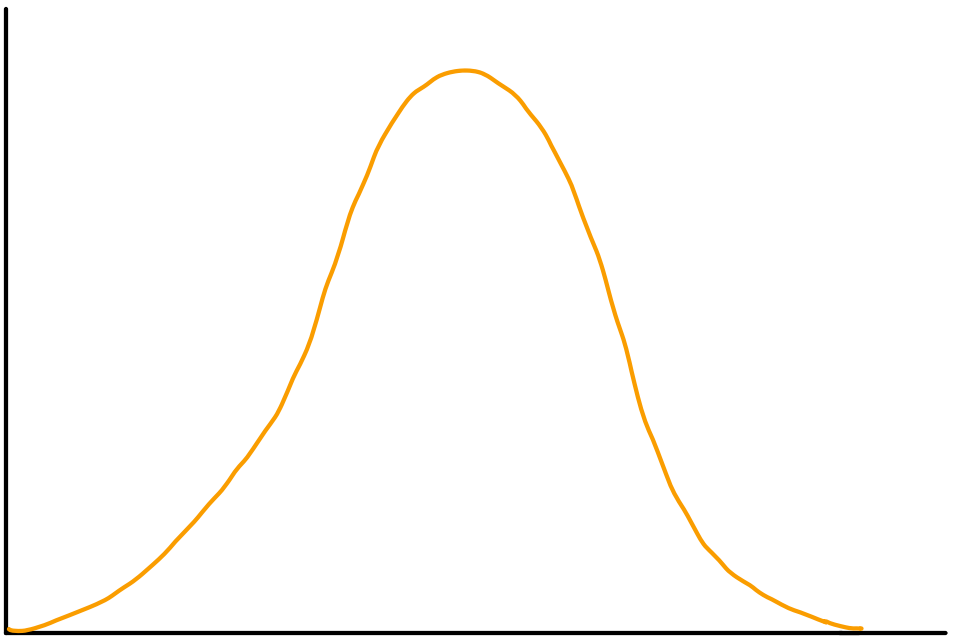
Agenda

- Descriptive Statistics
  - * Measure of Central Tendency
  - * Measure of Variability
- Inferential Statistics
- Weighted Mean / Average
- IQR : Box-plot , outliers
- Random Variables



Descriptive Statistics

Statistics

Descriptive

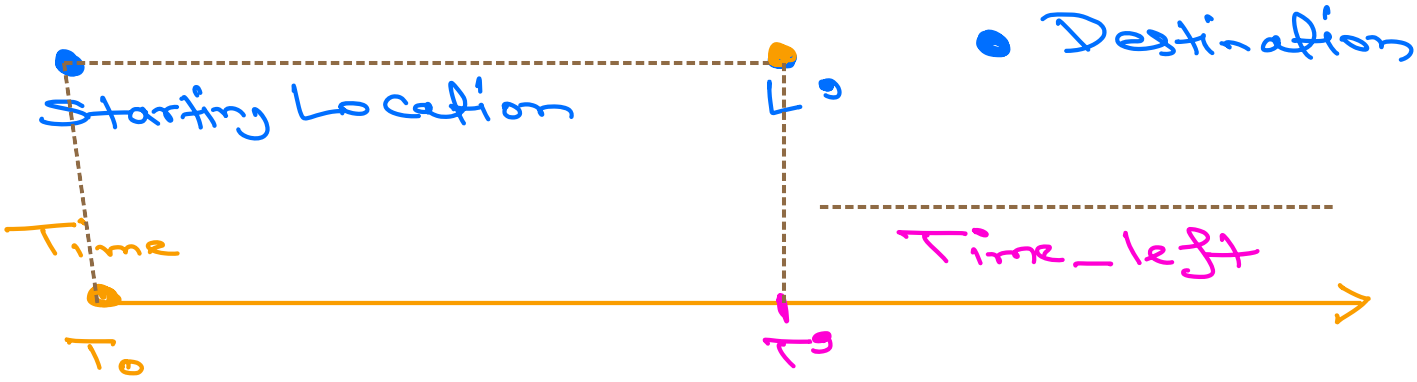
Describing the Existing Historical Data

Avg Speed

Inferential

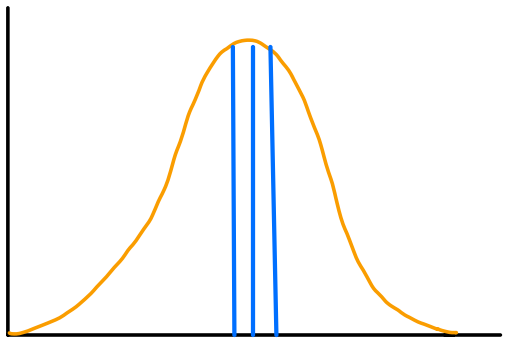
Inferring predicting Future from Historical Data

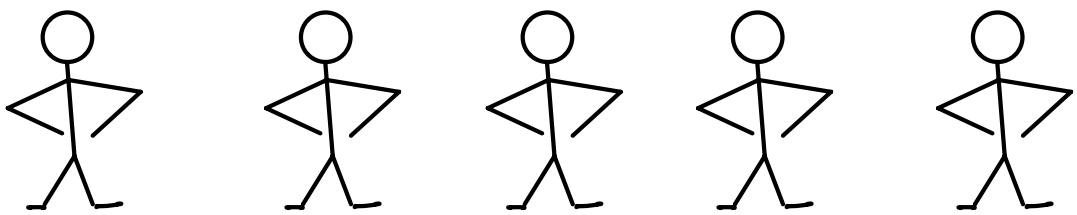
Estimate Time taken to Reach Destination



Measure of Central Tendency

1. Mean
2. Median
3. Mode





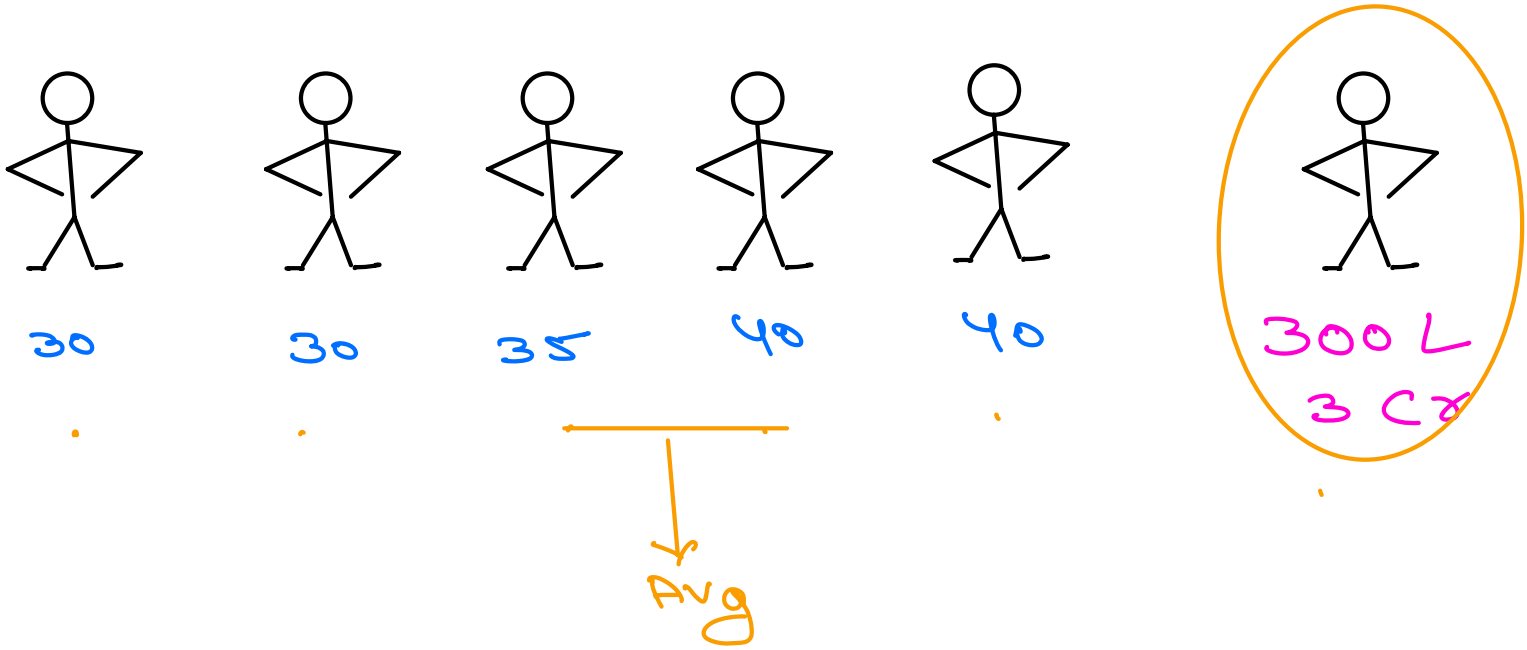
30 30 35 40 40

* mean = $\frac{30+30+35+40+40}{5}$

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n}$$

② Median: Midpoint of Sorted Data

outlier



if n odd : $\left[n//2 \right]$ Floor Div of Length

else : $\dfrac{\left[n//2 \right] \quad \left[n+1//2 \right]}{2}$

Discrete

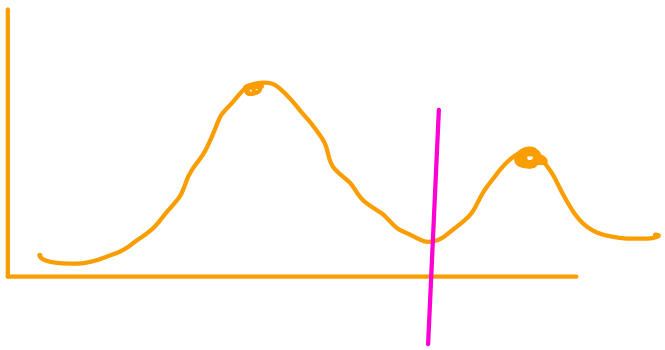
③ Mode: Most frequent element

(a) 30 30 35 40 40 40

⇒ 40 Uni Modal

(b) 30 30 35 40 40

⇒ 30, 40 Bi Modal Data



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?

mean

$P_4 = ?$

median = ?

mean ⇒ 24

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

$$x = 26$$

20 22 26 28

Avg 24 Median

Weighted Average / Mean

Loss function for Imbalance Dataset

SUBJECT	CREDIT	GRADE
Math	3	5
History	4	4
Chemistry	3	5
English	2	3

Standard Avg = $\frac{5+4+5+3}{4}$

= 17/4

Weighted Avg = $\frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$

= $\frac{3*5 + 4*4 + 3*5 + 2*3}{3+4+3+2}$

= $\frac{15+16+15+6}{12}$ = 4.33

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?

%	Num_pets
30	0
40	1
10	2
20	3
100	

$$\Rightarrow \frac{30 * 0 + 40 * 1 + 10 * 2 + 20 * 3}{100}$$

$$\Rightarrow 40 + 20 + 60 \Rightarrow \frac{120}{100} \Rightarrow 1.2$$

Expected Value

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?

$C_1 \rightarrow x$

$C_2 \rightarrow y$

Mo m

$$\frac{(x+y)}{2} = 40$$

$$x+y = 80 \quad (1)$$

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

$x+y+m = 45 \times 3$ (2)

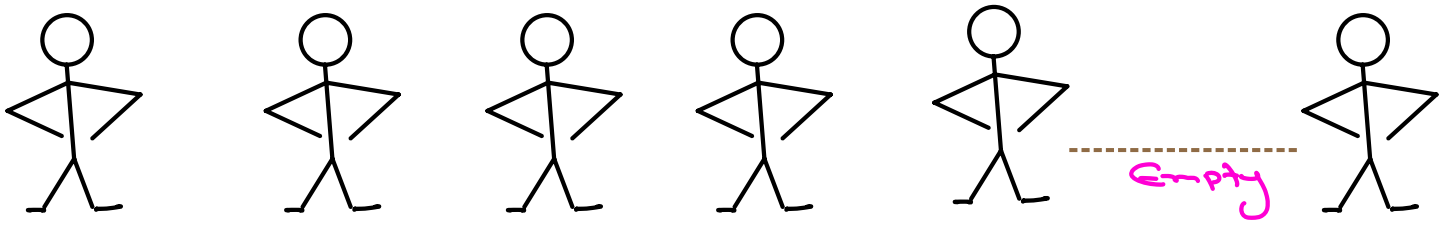
Solving equation (1) and (2)

$m = 45 \times 3 - 80$

$m = 135 - 80 = 55$

Measure of Variability

1. Range
2. Variance
3. Standard Deviation



30 30 35 40 40 300 L

Range = max - min

min → 30

max → 300

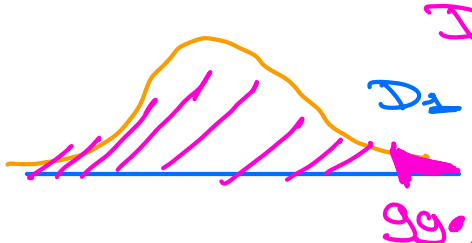
300 - 30 = 270

* Range is sensitive to Outli

* Inter Quartile Range

IQR = Q3 − Q1

* Percentile

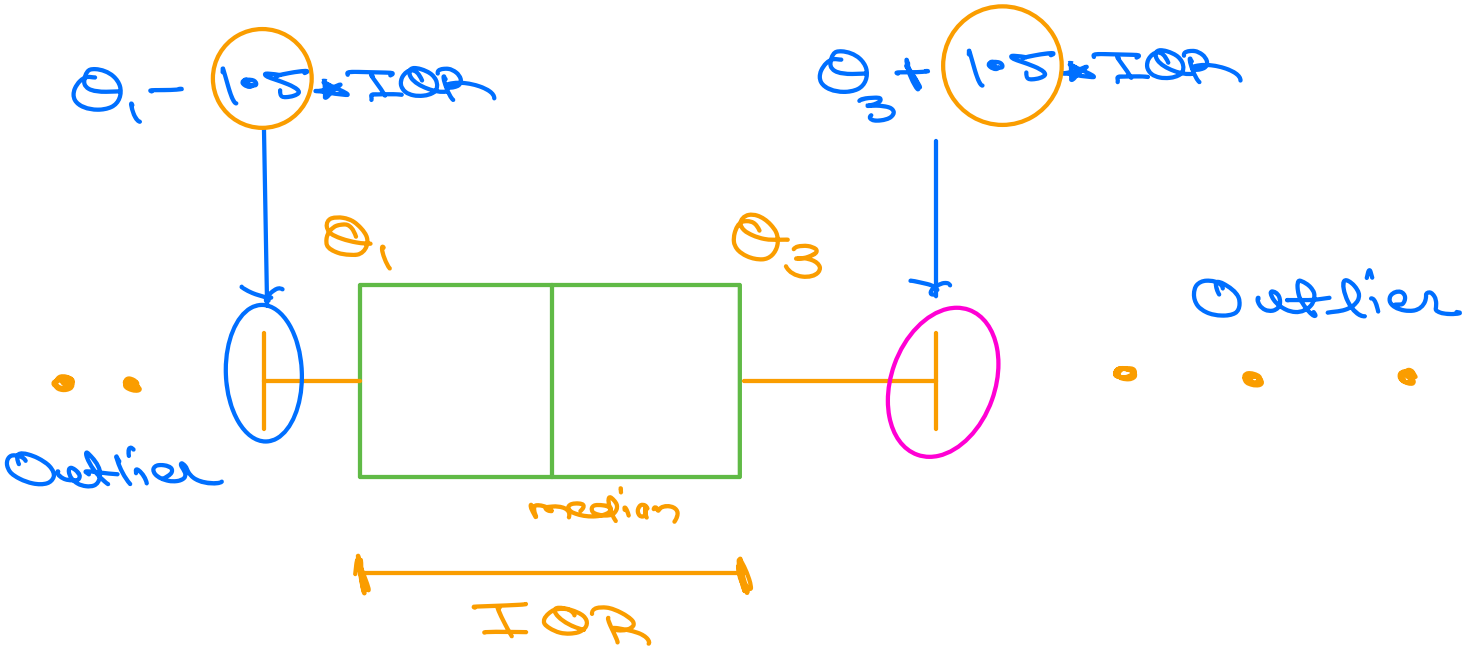


99% of Datapoint are less than D1

Q1 25th Percentile

Q2 50 Percentile

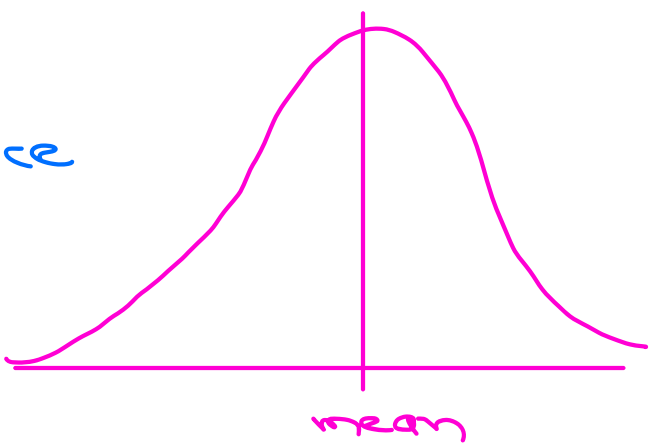
Q3 75 Percentile



Range

Variance

Avg Square of Distance from mean



$$\text{Variance} = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \sigma^2$$

* Higher the Variance more spread out the data

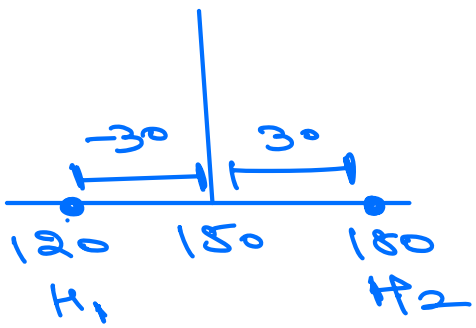
$H_i - \mu \Rightarrow e_i$

$H_1 - \mu \Rightarrow e_1$ (error/mistake) the mean model made

Total ⇒ Sum of Errors/n

⇒ −30 + 30

Total ⇒ 0 ✗



Error ⇒ Sum of Square e_i/n

⇒ 900 + 900 ⇒ $\frac{1800}{2}$ ✓

Total Error ⇒ 900 cm^2

Not interpretable

Standard Deviation

SD ⇒ $\sqrt{\text{Variance}}$ ⇒ 6

Random Variables

Q = S Non Random

Q can be defined as a Random Variable when we are not certain about the outcom

a ⇒ Outcome of Dice Roll

X ⇒ Rain Tmrw

* Discre Random Variable

C ⊙ {H, T}

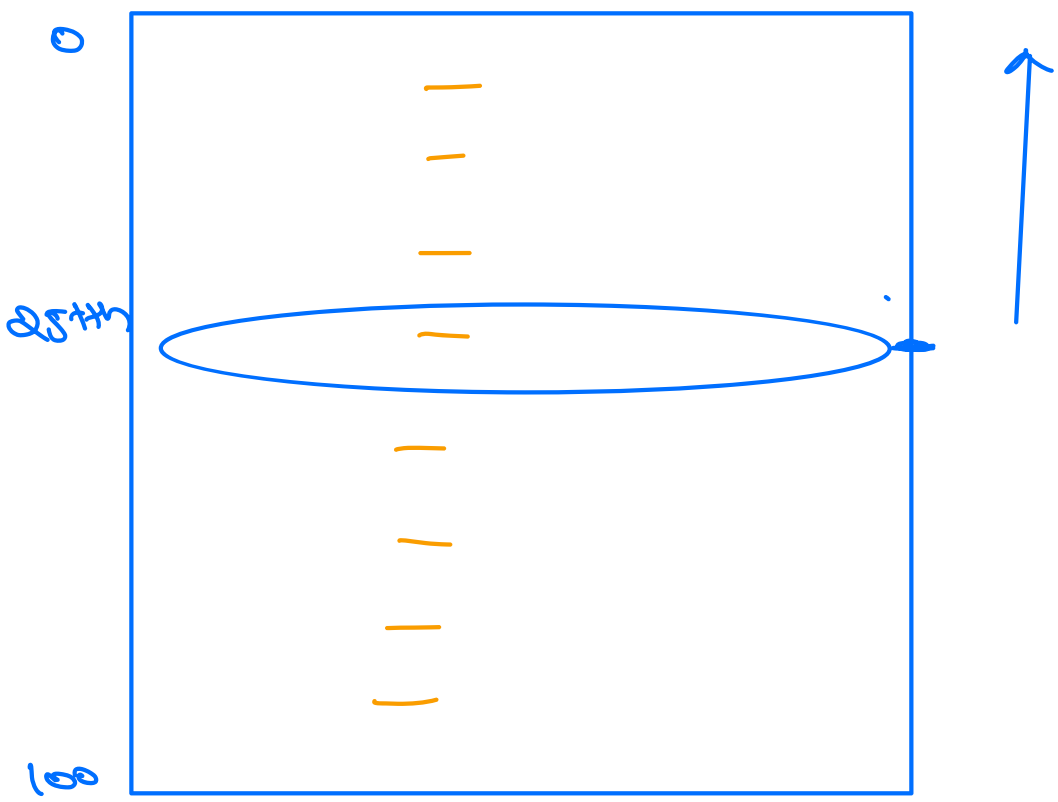
D ⊙ {1 2 3 4 5 6}

* Continous R V

Hieght of People ⊙ RE(10cm, 250cm)

Rain Forecast ⊙ (0mm, 100mm) 175.5552...

Sort all Data points



25% of your total Population is present

100

Q1 ⇒ 25%

Q2 ⇒ 50%
Median

Q3 ⇒ 75%

Probability = $\frac{\#\text{Possible Outcomes}}{\text{Total Outcomes}}$

20 Cars → 20 Spaces

Total $\frac{20}{1} \cdot \frac{19}{1} \cdots \frac{1}{1} = 20!$

Way 3 Arranging Cars in Alternate

10	10	9	9	---11
H	F	H	F	
F	H	F	H	

$2 * 10! * 10!$

$2 \times 10! * 10!$ / $20!$

$\frac{2}{{}^{20}C_{10}}$

${}^{n}C_r = \frac{n!}{(n-r)!r!}$

$\frac{20}{(20-10)! * 10!}$

$\frac{20!}{10!\,10!}$

$P(\text{Affected}) = 0.001 = \frac{1}{1000}$

$P(T/A) = 0.9$

$P(T/A') = 0.5$

$P(T) = P(T/A) * P(A) + P(T/A') * P(A')$ ①

Bayes Theorem

$P(A/T) = \frac{P(T/A) * P(A)}{P(T)}$ ②