Random variable

$X = 10$

$Y = \text{Pass}$

age	weight	hieght	
X_1	X_2	X_3	Y
-	-		
-	-		
-	-		
-	-		

$X \approx Y$

Random variable

$X = \text{coin (Toss)}$

$X = T/H$

$X = \text{Dice}$

$x = 1/2/3/4/5/6$

Type of Random variable

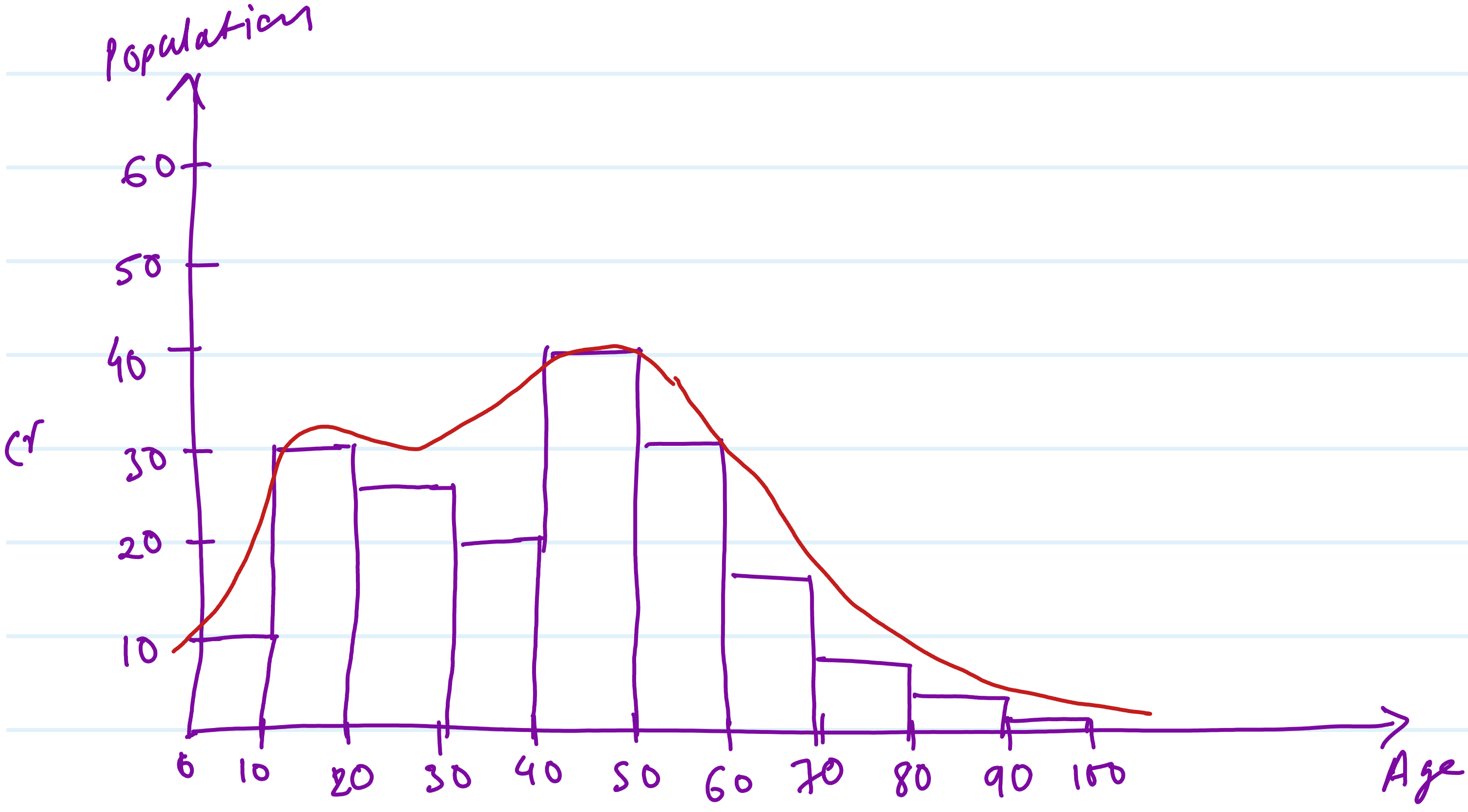
Descrite random variable

Continous Random variable

$X_1 =$ Gender
$= M | F | T$

$X_2 =$ Result
(P / F)

$X_1 =$ Rain
(1mm, 6.5mm)

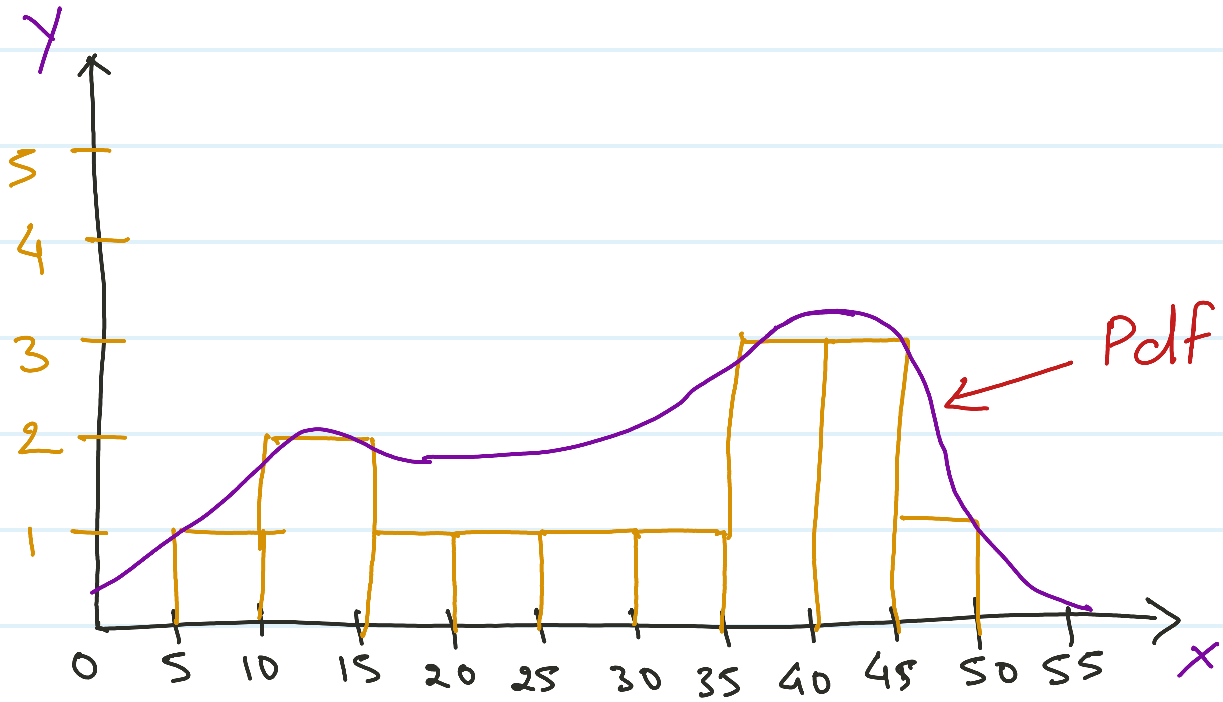


✱ Histogram

Dataset = [10, 12, 14, 18, 24, 30, 35, 36, 37, 40, 41, 42, 43, 50, 51]

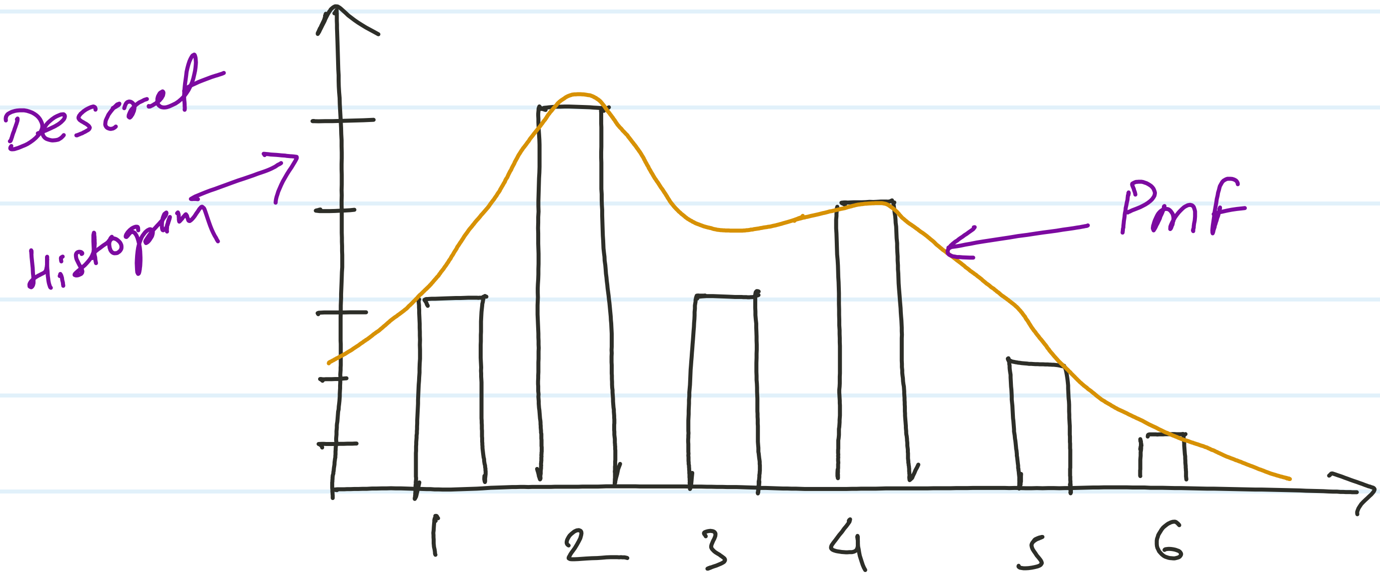
Bin | Bin size   Assume bin size = 5

$\Rightarrow$ No of Bin = $\frac{50}{5}$ = 10



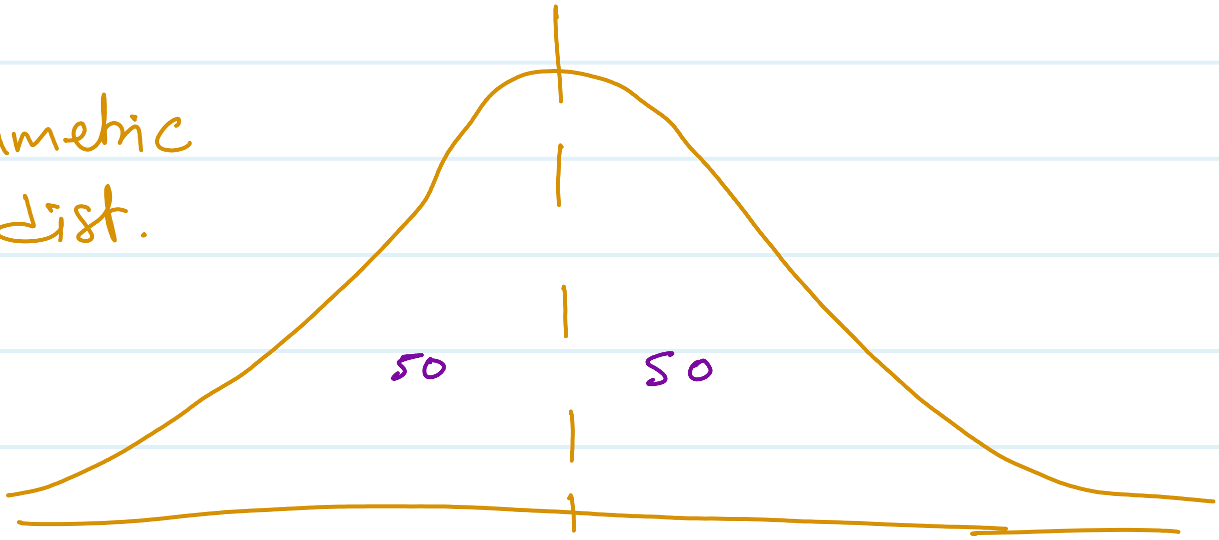
Pdf = Probability density function.

continuous histogram



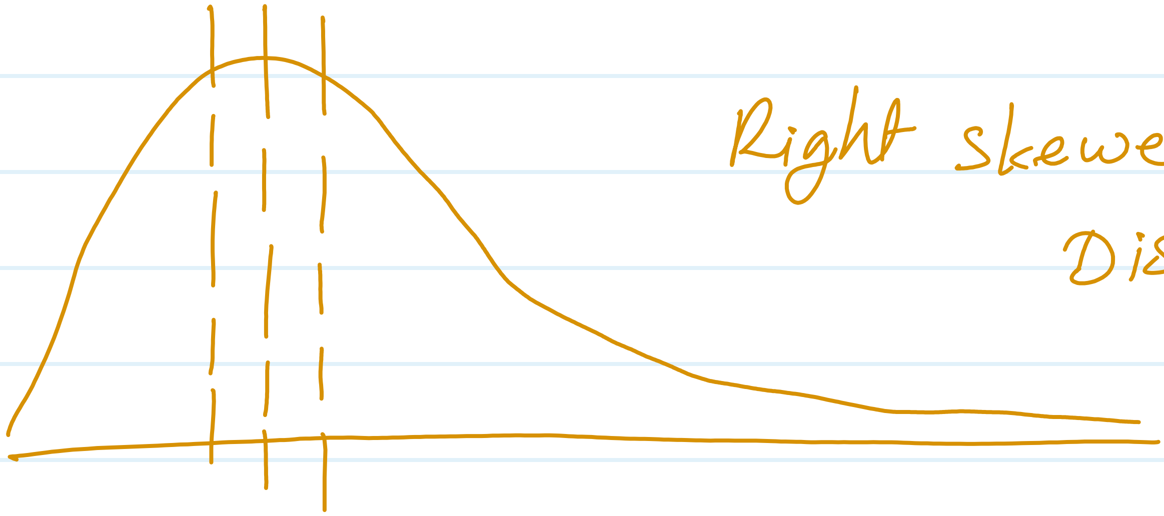
Pmf = Probability mass function

① Symmetric dist.



mean = median = mode

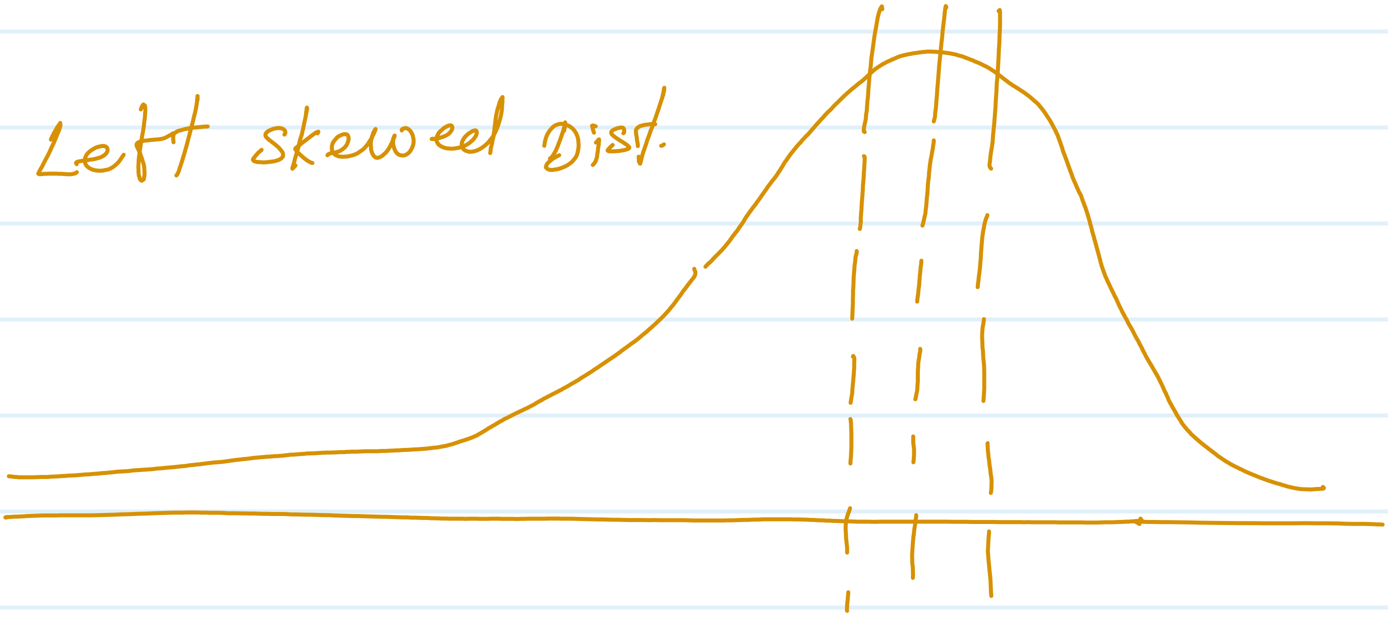
②



Right skewed Dist.

mean > median > mode

③ Left skewed Dist.



mean < median < mode

Sampling methods

↓

Probability sampling

↓

1. Simple Random Samp
2. clustered samp.
3. Systematic samp.
4. stratified Random samp

Non-probability sampling

1. Convenience Sample.
2. consicutive samp.
3. Quato sample.
4. Purposive | Judgement Sample.
5. Snowball sampling.

<u>Quarter</u>

Percentile and Quartile

100%.

25%. percentile = Q_1

50% percentile = Q_2 / median

75% —||— = Q_3

100 %. —||— = Q_4

Score	Rank
30	1
33	2
43	3
53	4
56	5
67	6
68	7
72	8

find out where is the 25th percentile is in the above list.

Rank at 25th percentile

$\text{Rank} = \frac{\text{Percentile}}{100} \times (\overset{\checkmark}{n}+1)$

$= \frac{25}{100} \times 9$

$= 0.25 \times 9$

$= 2.25$

when rounding up/down the closest value will be Rank.

so Rank is $= 2$

<u>Rank 75%</u>

$$= \frac{75}{100} \times (n+1)$$

$= 075 \times 9$

$= 6.75$

Rank $= 7$

25% = Q_1

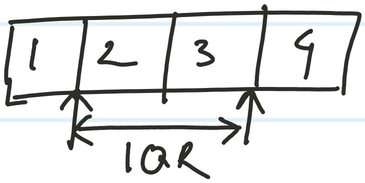
50% = Q_2 / median

75% = Q_3

100% = Q_4

IQR (Inter Quartile Range)

$$IQR = Q_3 - Q_1$$



5 - Number Summery

1. min
2. Q_1
3. median
4. Q_3
5. max

IQR

= [1, 2, 3, 4, 6, 8, 11, 14, 18, 19, 5, 21, 82, 95, 7]

1, 2, 3, 4, 5, 6, 8, 11, 14, 18, 19, 21, 82, 95, 140

To find outlies we use 5 numbes summery to display values in Box-whiskes plot

Formula

uppes limit = $Q_3 + 1.5(IQR)$

lowes limit = $Q_1 - 1.5(IQR)$

$$Q_1 = \frac{Q_1}{100} \times (n+1)$$

$$= \frac{25}{100} \times (15+1)$$

$$= \frac{25}{100} \times 16 \quad =$$

$$Q_1 = 4$$

$$Q_3 = \frac{\overset{3}{\cancel{75}}}{\underset{25}{\cancel{100}}} \times \overset{4}{\cancel{16}}$$

$$Q_3 = 12$$

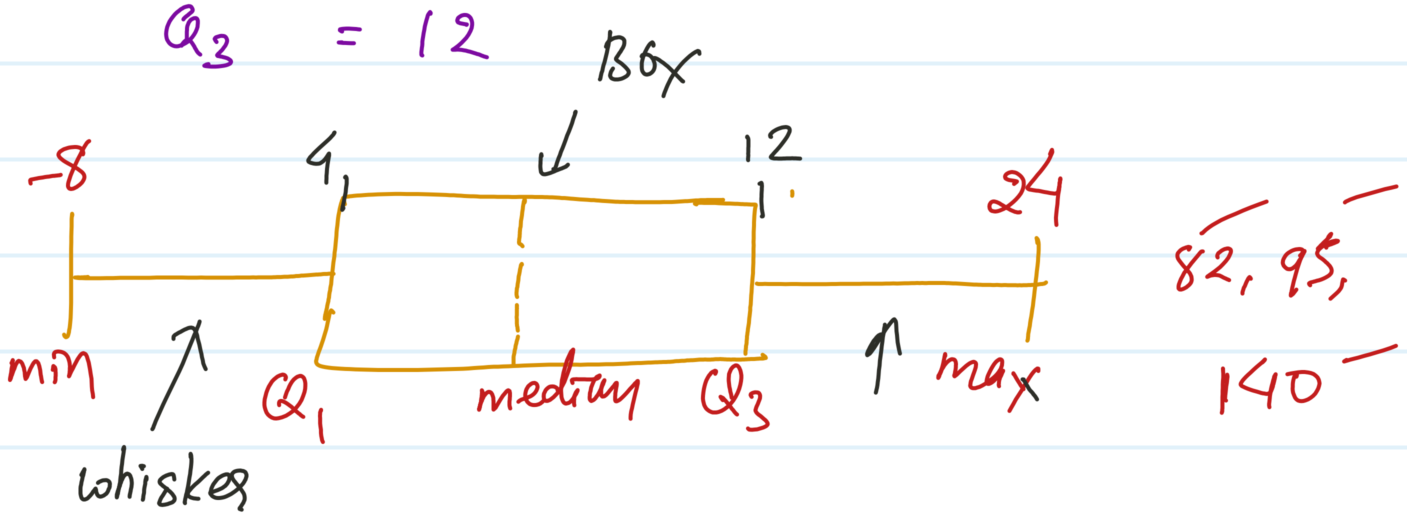
$$IQR = Q_3 - Q_1$$
$$= 12 - 4$$
$$= 8$$

lower limit $= 4 - 1.5 \times 8$
$= -8$

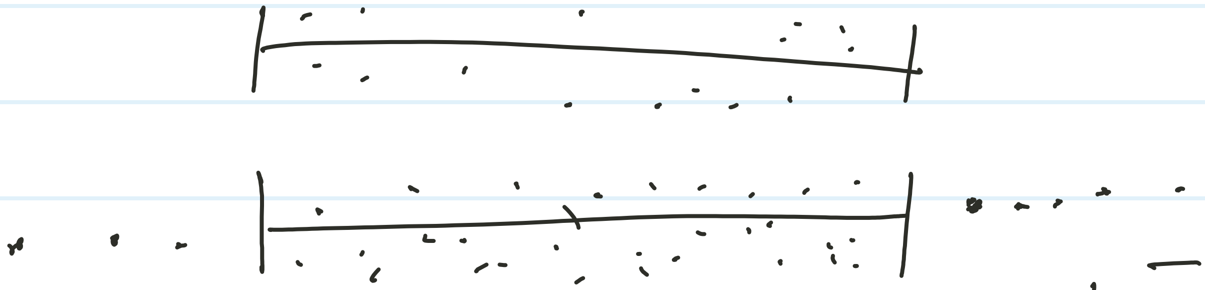
upper limit $= 12 + 1.5 \times 8$
$= 24$

min $= -8$
max $= 24$
$Q_1 = 4$
$Q_3 = 12$

Box and whisker plot



To treat outlier, we can use median



① mean

② median

③ mode

1
2
3
4
5
6 → 6.5
7
8
9

outlier $\frac{45+75+80+89}{12} = \frac{289}{12} \Rightarrow 24$

24	75	6.5
24	80	6.5
24	89	6.5

7
1
2
7.5 NAN 15.2
5 →
6
9
7.5 NAN 15.2
11
7.5 NAN 15.2
12

1
2
5
6 → 7.5
9
11
12
77

15.2

$\frac{122}{8}$

$\frac{54}{8} \Rightarrow 8.8$

missing → mean, median, mode

mean, median → numeric data

P
F
P
F
NA - P
F
NA - P
P
NA - P
P

mode