

Random variable

$$X = 10 \sim$$

$$Y = \text{Pass} \sim$$

age	weight	height	
X_1	X_2	X_3	Y
1	—		
1	—		
1	—		
1	—		

$$X \approx Y$$

Random variable

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

$$X = T/H$$

$$X = \text{Dice}$$

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

Type of Random variable

1

↓
Discrete random variable

↓
Continuous Random variable

$X_1 = \text{Gender}$
 $= M/F/T$

$X_1 = \text{Rain}$
(1mm, 6.5mm)

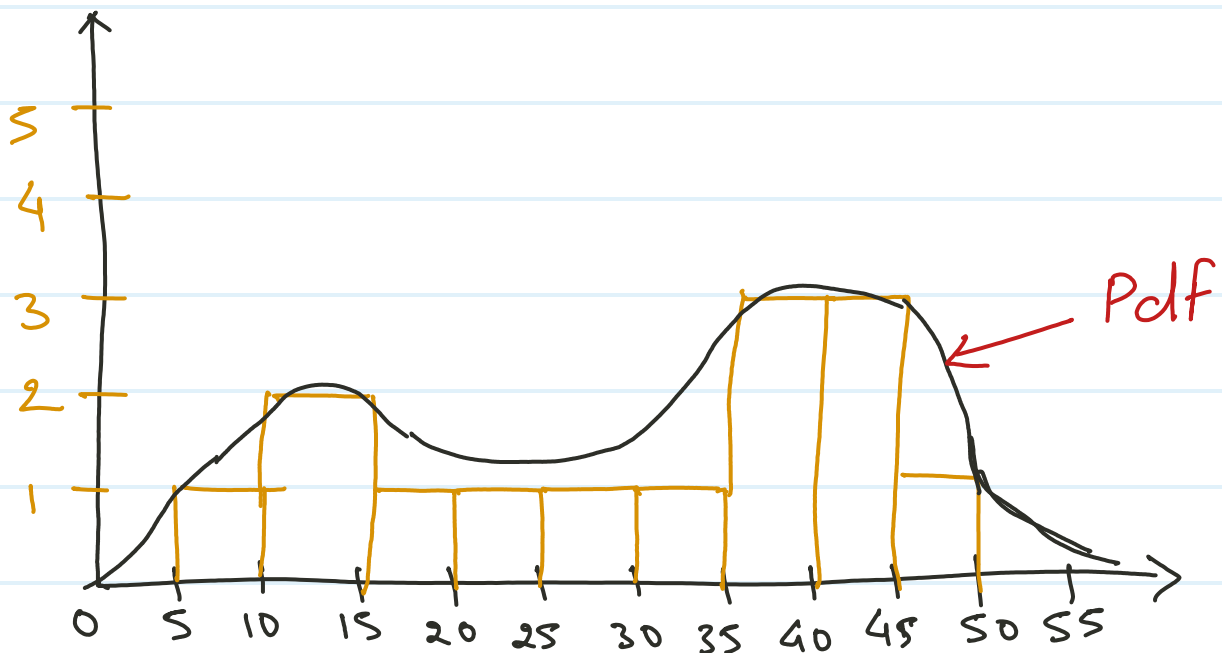
$X_2 = \text{Result}$
(P/F)

* Histogram

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

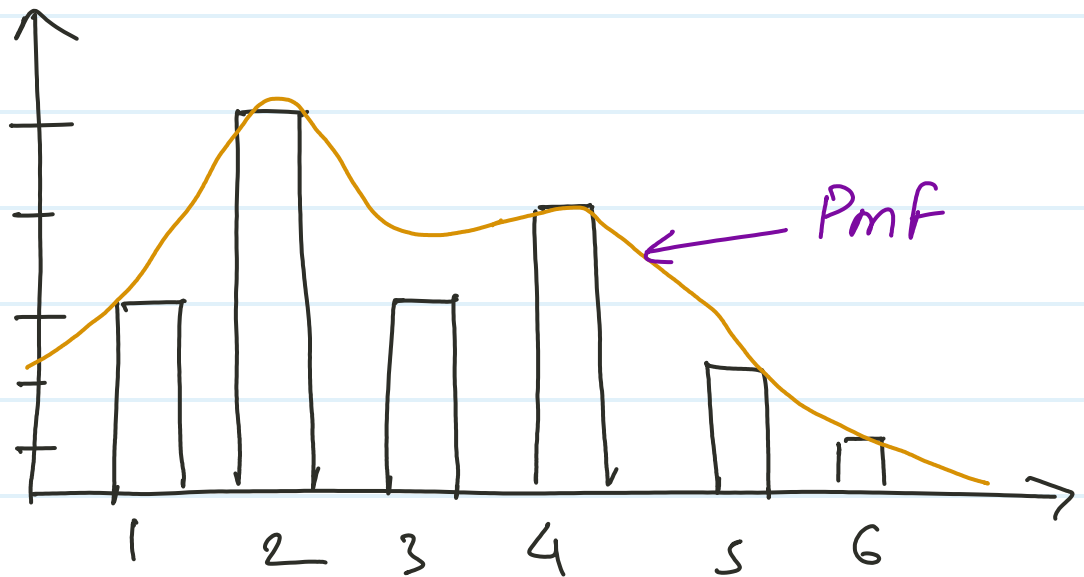
Bin / Bin size Assume bin size = 5

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



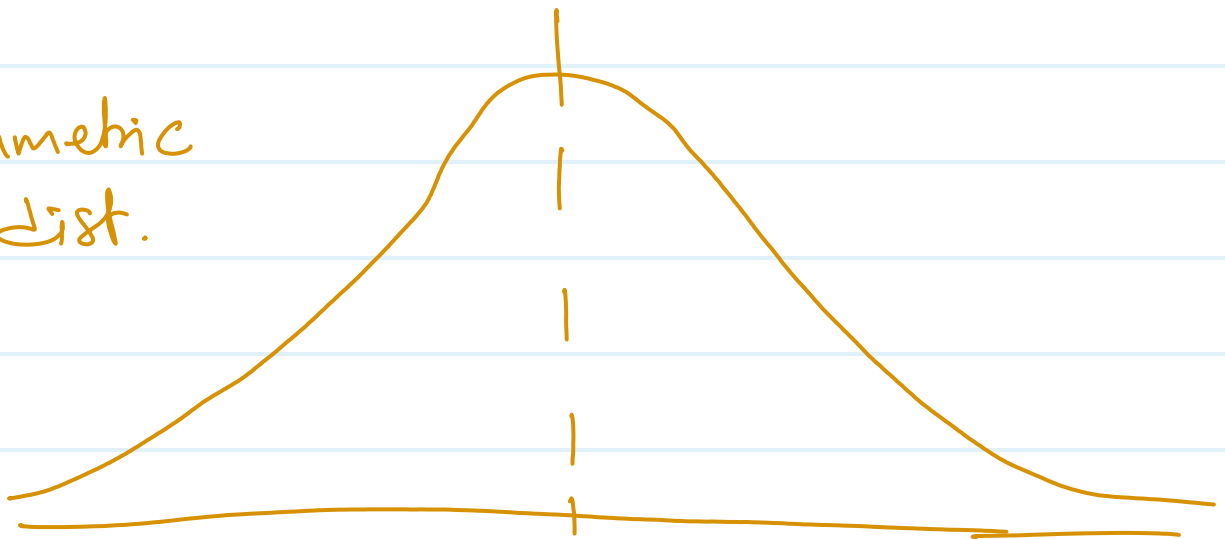
Pdf = Probability density function.

continuous histogram



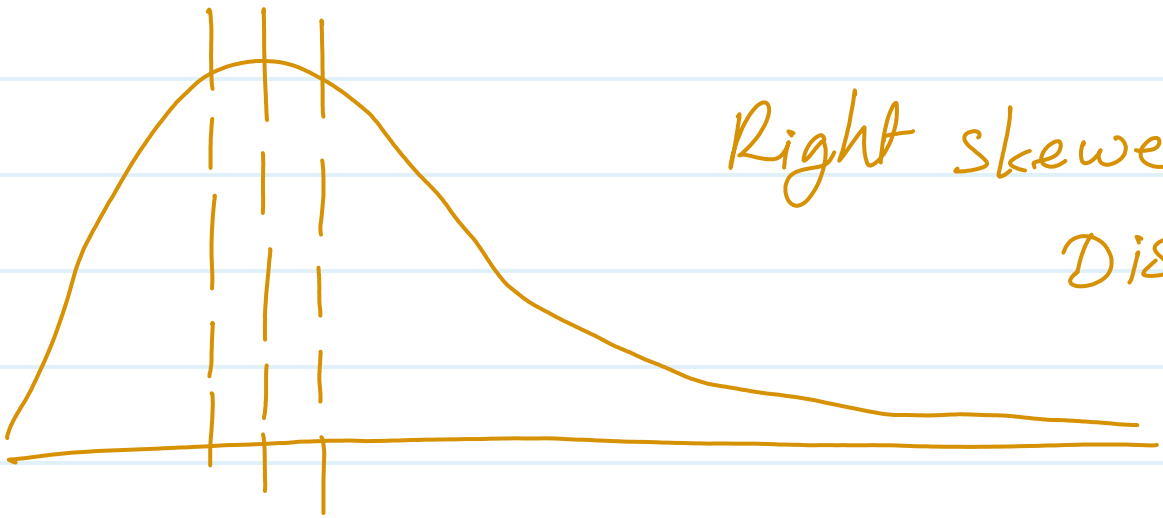
Pmf = Probability mass function

① Symmetric dist.



mean = median = mode

(9)



Right skewed
Dist.

$$\text{mean} > \text{median} > \text{mode}$$

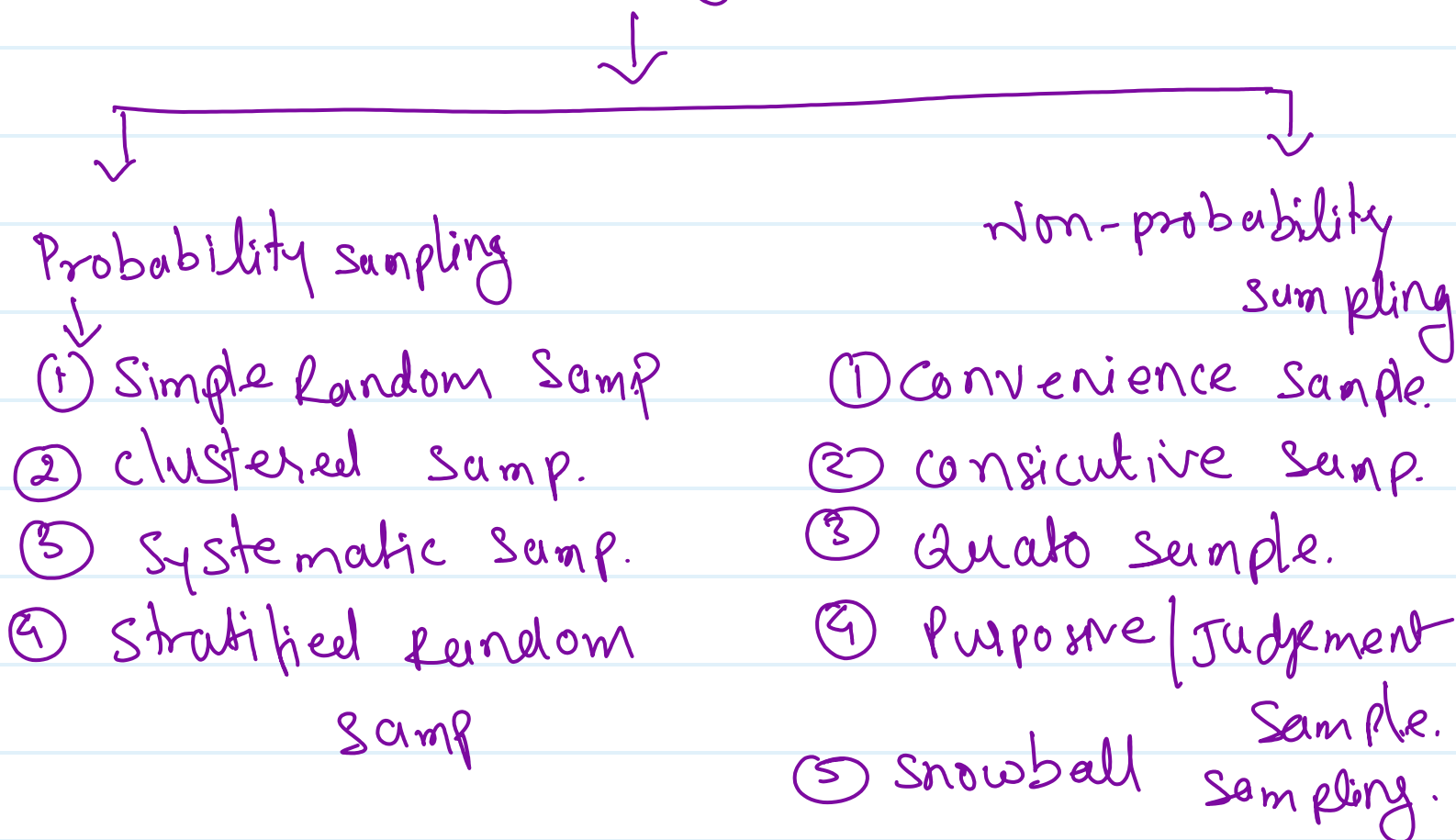
(5)

Left skewed Dist.



$$\text{mean} < \text{median} < \text{mode}$$

Sampling methods



* 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

$$\begin{aligned}\text{Rank} &= \frac{\text{Percentile}}{100} \times (n+1) \\ &= \frac{25}{100} \times 9\end{aligned}$$

$$= 0.25 \times 9$$

$$= 2.25$$

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

$$\text{So Rank is } = 2$$

$$\underline{\text{Rank } 75\%}$$

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

$$= 0.75 \times 9$$

$$= 6.75$$

$$\text{Rank } = 7$$

$$\begin{aligned}
 \star \quad 25\% &= Q_1 \\
 50\% &= Q_2 / \text{median} \\
 75\% &= Q_3 \\
 100\% &= Q_4
 \end{aligned}$$

IQR (Inter Quatile Range)

$$IQR = Q_3 - Q_1$$

★ 5-Number Summary

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

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

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

↑
11
11
11

To find outliers we use 5 number summary to display values in Box-whisker plot

Formula upper limit = $Q_3 + 1.5(IQR)$
 lower limit = $Q_1 - 1.5(IQR)$

$$\begin{aligned} Q_1 &= \frac{Q_1}{100} \times (n+1) \\ &= \frac{25}{100} \times (15+1) \\ &= \frac{25}{100} \times 16 = \end{aligned}$$

$$Q_1 = 4$$

$$Q_3 = \frac{\overset{3}{\cancel{78}}}{\underset{\cancel{28}}{+100}} \times 16$$

$$Q_3 = 12$$

$$IQR = Q_3 - Q_1$$

$$= 12 - 4$$

$$= 8$$

$$\begin{aligned} \text{lower limit} &= 4 - 1.5 \times 8 \\ &= -8 \end{aligned}$$

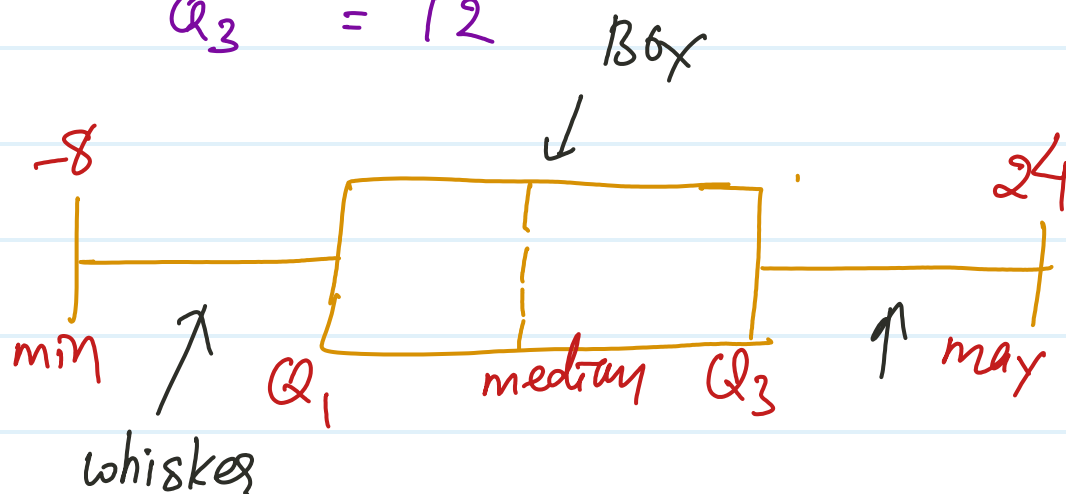
$$\begin{aligned} \text{upper limit} &= 12 + 1.5 \times 8 \\ &= 24 \end{aligned}$$

$$\text{min} = -8$$

$$\text{max} = 24$$

$$Q_1 = 4$$

$$Q_3 = 12$$



82, 95,
140

To treat outliers, we can use median