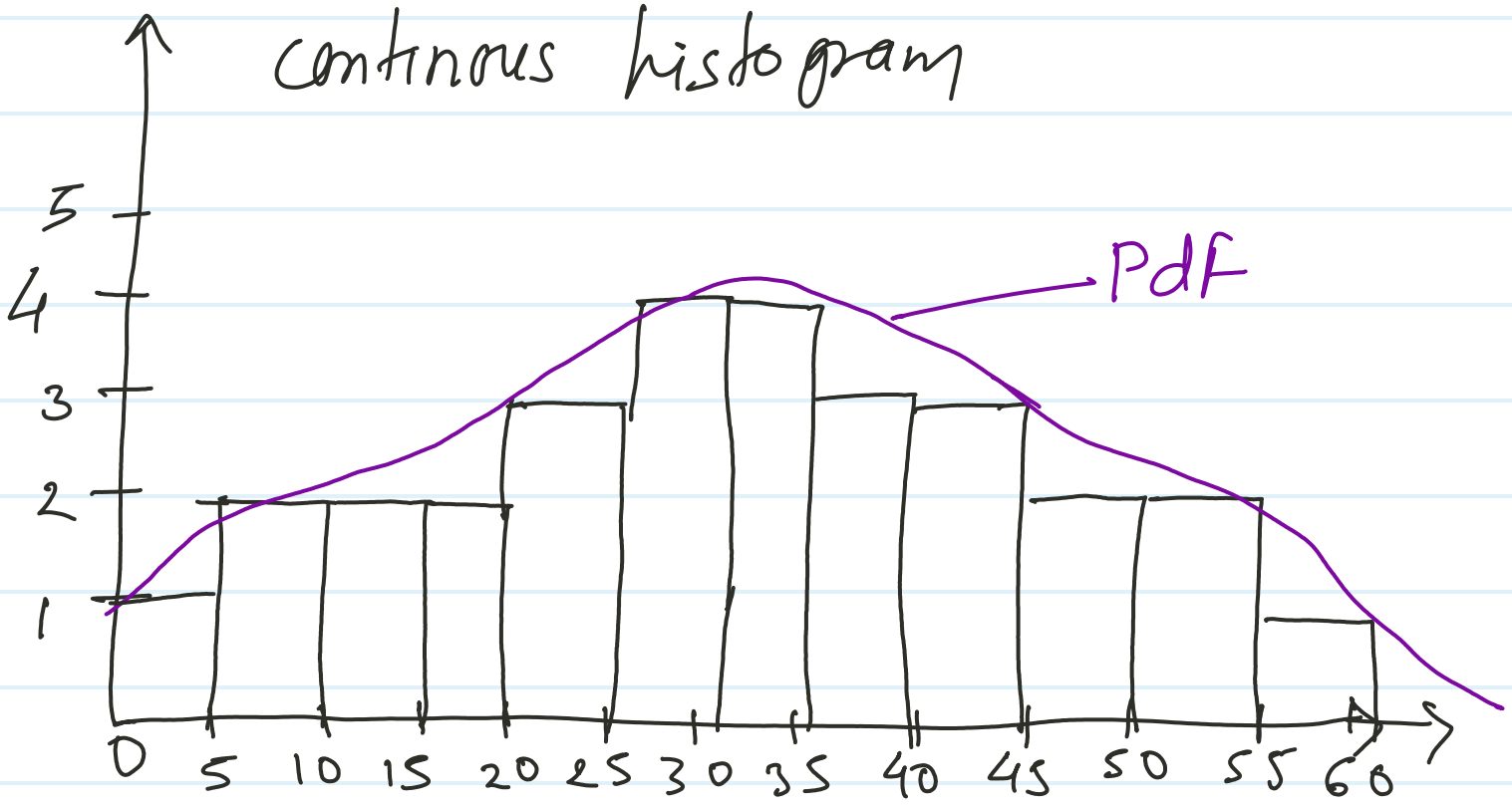
Histogram

[2, 3, 7, 8, 11, 13, 16, 20, 21, 22, 23, 26, 27, 28, 30, 31, 33, 34, 35, 37, 38, 40, 41, 44, 47, 49, 51, 52, 56]

Range = 56 - 2 = 54

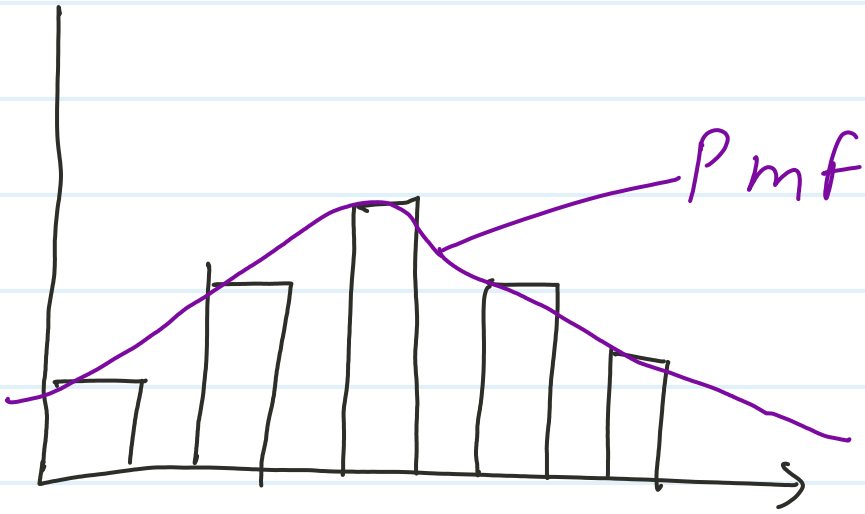
Assume bin size = 5

$$\text{no of bin} = \frac{\text{Range}}{\text{bin size}} = \frac{54}{5} = 10.8$$



Pdf = Probability density function

Descrete histogram



Pmf = Probability mass function.

Visualization

1. Histogram
2. Box & whisker plot
3. Bar plot

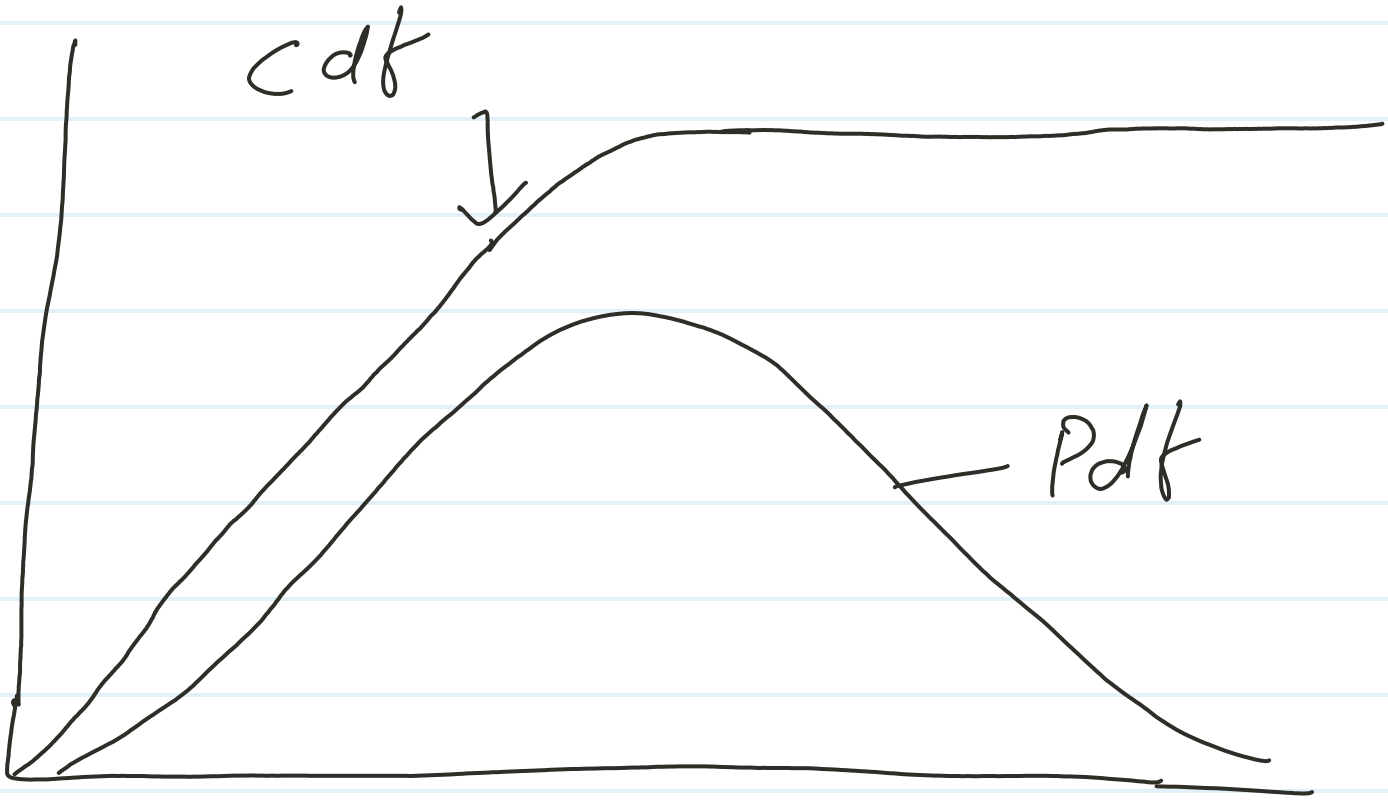
④ column plot

⑤ Pie chart

⑥ Scatter plot

x	frequency
2	3
3	2
4	1
6	4
8	2
9	4
	16

X		Cummulative
2	3	3
3	2	2+3 = 5
4	1	5+1 = 6
6	2	6+2 = 8
8	4	8+4 = 12
9	3	12+3 = 15
	2	15+2 = 17



cdf = cummulative Density function.

Covariance and correlation.

covariance

X	Y
age	height
2	2
4	2.5
6	3
10	3.5
15	4.5
20	5.5

$X \uparrow \quad Y \uparrow$

$X \downarrow \quad Y \downarrow$

} positive covariance

$X \uparrow \quad Y \downarrow$

$X \downarrow \quad Y \uparrow$

} negative covariance

X ↑ Y

X ↓ Y

X ↑ Y ↑

X ↑ Y ↓

} zero covariance

Age	weight	height	BMI
1	2	3	4

Independent

Target

1 = 4
2 = 4
3 = 4
} +/-

1 = 2
2 = 3
1 = 3
} zero covariance

$$Cov(X,Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

limit $(+\infty, -\infty)$

Correlation

① Pearson's correlation coefficient
② Spearman Rank corr coeff.

① Pearson's correlation coefficient

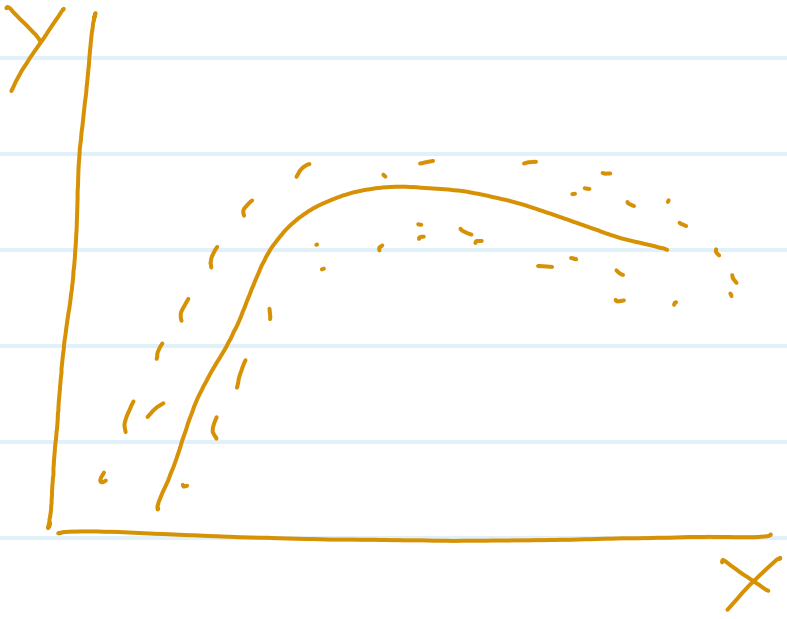
$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

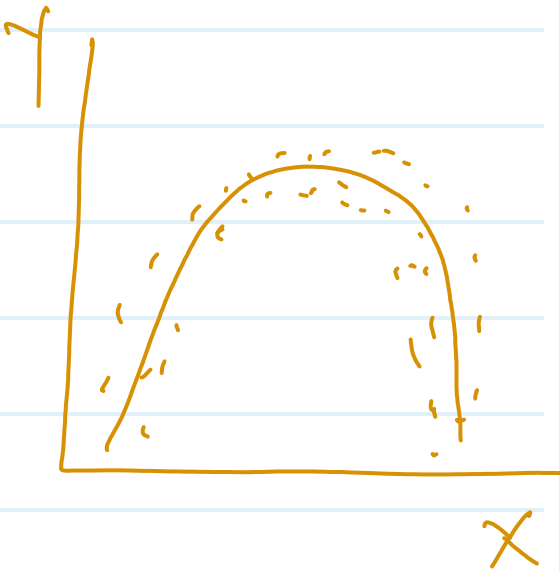
limit $(-1, 1)$

Disavantage -

Can't identified non-linear relation

+/−





(II) Spearman Rank corr coff.

$$corr_R(X, Y) = \frac{cov(X, Y)}{\sigma_{RX} \cdot \sigma_{RY}}$$

Eg

X	Y
2	20
4	28
6	34
3	30

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

$\bar{X} = 3.75$

$\bar{Y} = 28$

$x_i - \bar{x}$ $\quad$ $y_i - \bar{y} =$

$2 - 3.75 = -1.75$ $\quad$ $20 - 28 = -8$

$4 - 3.75 = 0.25$ $\quad$ $28 - 28 = 0$

$6 - 3.75 = 2.25$ $\quad$ $34 - 28 = 6$

$3 - 3.75 = -0.75$ $\quad$ $30 - 28 = 2$

$$\Rightarrow \frac{14 + 0 + 13.5 - 1.5}{3}$$

$$\Rightarrow \frac{26}{3} \Rightarrow 8.66$$

$= 0$

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$\sqrt{\frac{1}{n-1}\sum (X_i - \bar{X})^2}$$

$$\Rightarrow -1.75 = 3.0625$$

$$\Rightarrow 0.25 = 0.0625$$

$$\Rightarrow 2.25 = 5.0625$$

$$\Rightarrow -0.75 = 0.5625$$

$$\sqrt{\frac{8.75}{3}}$$

$$\Rightarrow \sqrt{2.9} \Rightarrow 1.7$$