Covariance and correlation

Covariance -

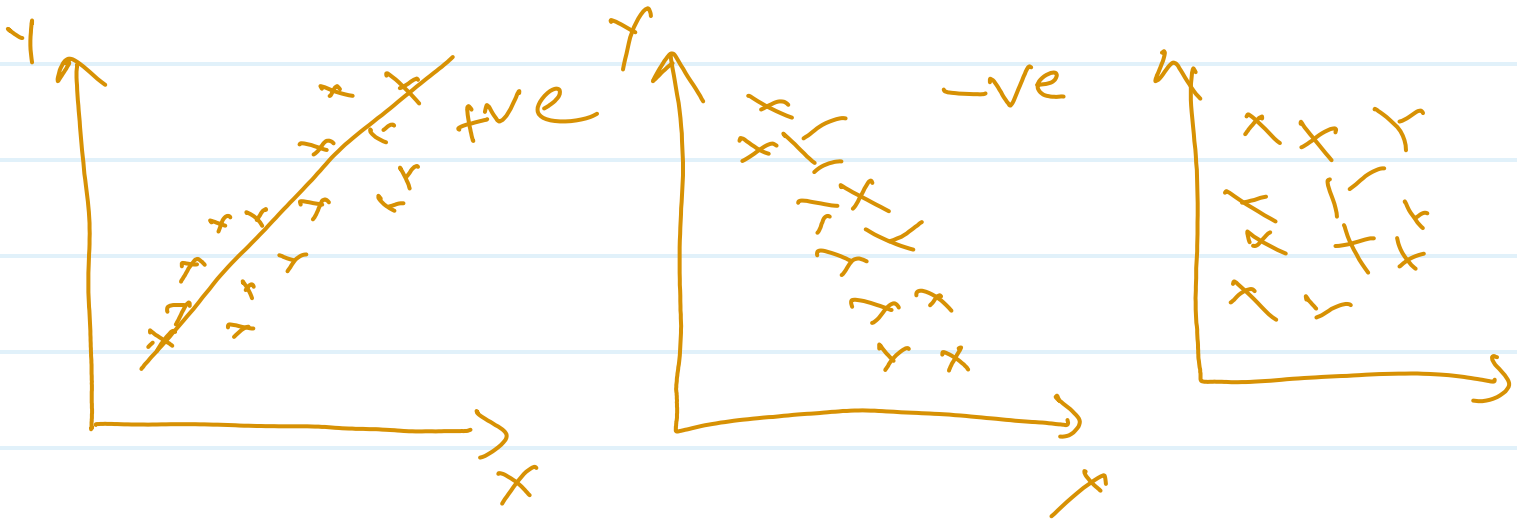
X	Y
age	hight
5	80cm
10	100cm
15	120cm
20	150cm

X_1 X_2 X_3 X_4 X_5 X_6 | Y

$X\uparrow$ $Y\uparrow$ = Positive cova.

$X\downarrow$ $Y\downarrow$ = negative cova.

$X\uparrow$ $Y\uparrow\downarrow$ = Zero Cova.



limit $(\infty, -\infty)$

$$Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Co-relation — Relation as well as strength.

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

Positive corr = +1

zero corr = 0

Negative corr = -1

Type correlation

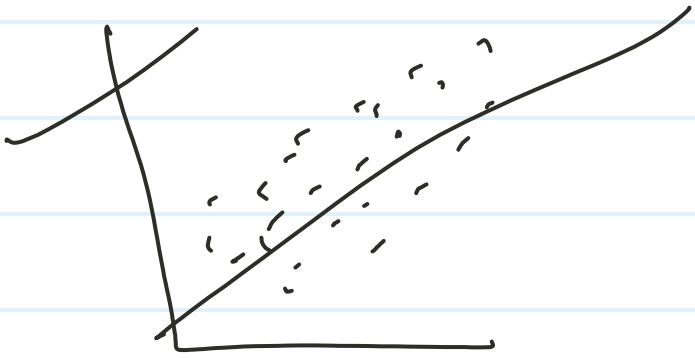
1. Pearson's correlation coefficient
2. Spearman's Rank correlation coefficient

1. Pearson's –

It describe the linear relation b/w quantitative/Numby variable.

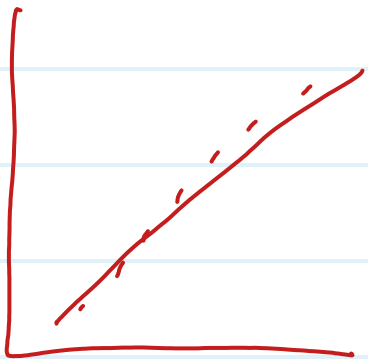
$$P(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

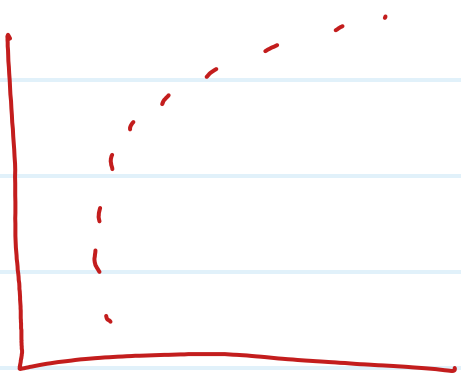
Disadvantage — It doesn't work on non-linear relation variable

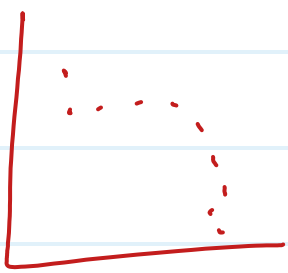


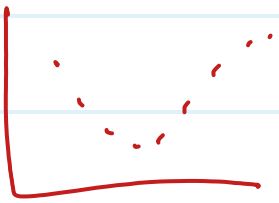
② Spearman's Rank corr. coefficient

Formula $r_{(X,Y)} = \frac{cov(X,Y)}{R_{ox} R_{oy}}$









X_1	X_2	X_3	X_4	Y
0	10	100	1000	—
$			5000	—
10	100	1000	10000	=

$$\frac{0.1}{10000} = 0.00001$$

$$\frac{5}{1000} = 0.0005$$

$$\frac{5000}{10000} \Rightarrow 0.5$$

$$\frac{500}{10000} = 0.0500$$

$$\frac{50}{10000} = 0.005$$

$$\frac{10000}{10000} = 1$$

0 - 1

Center limit theorem

No of sample = 40

{ Sample size >= 30 }

$(n_1, n_2, n_3 \text{ - - - - } n_{40})$

$(\bar{x}_1, \bar{x}_1, \bar{x}_3 \text{ - - - - - - } \bar{x}_{40})$

Dataset $(x_1, x_2, x_3 \text{ - - - } x_{40})$

max no. of sample will lead to normalize the data curve.