Covariance

Eg

X	Y
Age	weight
14	40
15	45
18	61
20	68
25	74

$$\left.\begin{array}{ll} \text{If } X\uparrow & Y\uparrow \\ X\downarrow & Y\downarrow \end{array}\right\} \text{+ve covariance}$$

$$\left.\begin{array}{ll} \text{If } X\uparrow & Y\downarrow \\ X\downarrow & Y\uparrow \end{array}\right\} \text{-ve covariance}$$

Covariane

population

$$\mathrm{Cov}(x,y) = \sum_{i=1}^{N} \frac{(x_i-\bar{x})(y_i-\bar{y})}{N}$$

$$\text{Sample } \operatorname{Cov}(x,y) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})(y_i-\bar{y})}{(n-1)}$$

Eg:- Economic growth (%) | Nifty (50) index

Economic growth (%)	Nifty (50) index
2.1	8
2.5	12
3.6	10
4.0	14

calculation

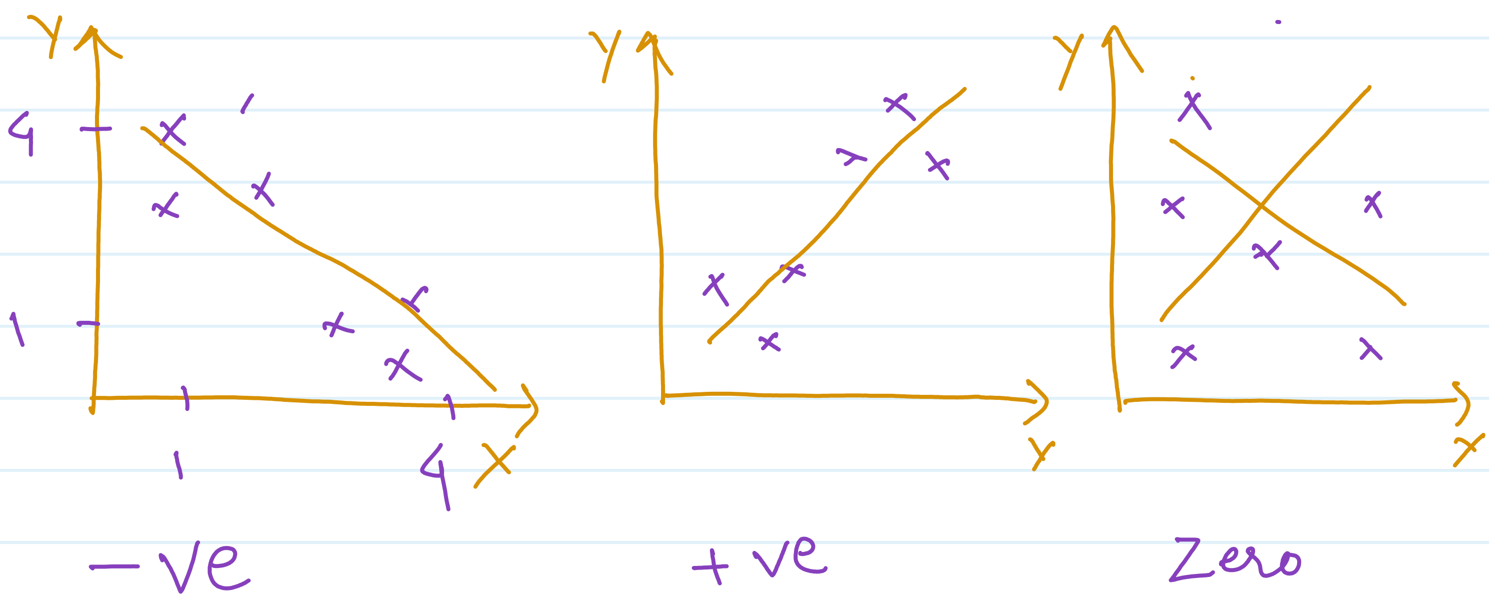
x	y	$\bar{x}$	$\bar{y}$	$(x-\bar{x})$	$(y-\bar{y})$
2.1	8			-1	-3
2.5	12	3.05	11	-0.6	1
3.6	10			0.5	-1
4.0	14			0.9	3

$$\operatorname{Cov}(x,y) = \frac{(-1)(-3) + (-0.6)(1) + (0.5)(-1) + (0.9)(3)}{4-1}$$

$$= \frac{4.6}{3} = \underline{\underline{1.533}}$$

Covariance -1000 to $+1000$

Zero means no covariance b/w x and y



Diff. b/w covariance and co-relation

in covariance we can find relation b/w x and y but can't say about it's strength

In co-relation we can find relation b/w x and y as well its strength which is calculate by -1 to $+1$

☆ Pearson corelation coffiecient

Eg:

X growth	Y Nifty50
2.1	8
2.5	12
3.6	10
4.0	14

$$\Rightarrow \quad \rho_{(X,Y)} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y}$$

standard devi.

$$\sigma_X = \sqrt{\frac{(-1)^2 + (-0.6)^2 + (0.9)^2 + (0.5)^2}{4-1}}$$

$$\sigma_X = \sqrt{0.8060} = 0.89$$

$$\sigma_Y = \sqrt{\frac{(-3)^2 + (1)^2 + (3)^2 + (-1)^2}{4-1}}$$

$$\sigma_Y = \sqrt{6.66} = \underline{2.58}$$

$$\rho_{(x,y)} = \frac{1.533}{(0.89)(2.58)}$$

$$\Rightarrow 0.66$$

$$= \underline{\underline{66\%}}$$

$\left.\begin{array}{ll} X\uparrow & Y\uparrow \\ X\downarrow & Y\downarrow \end{array}\right\}$ +ve

$\left.\begin{array}{ll} X\uparrow & Y\downarrow \\ X\downarrow & Y\uparrow \end{array}\right\}$ −ve

BMI

weight	height	BMI
X_1	X_2	Y

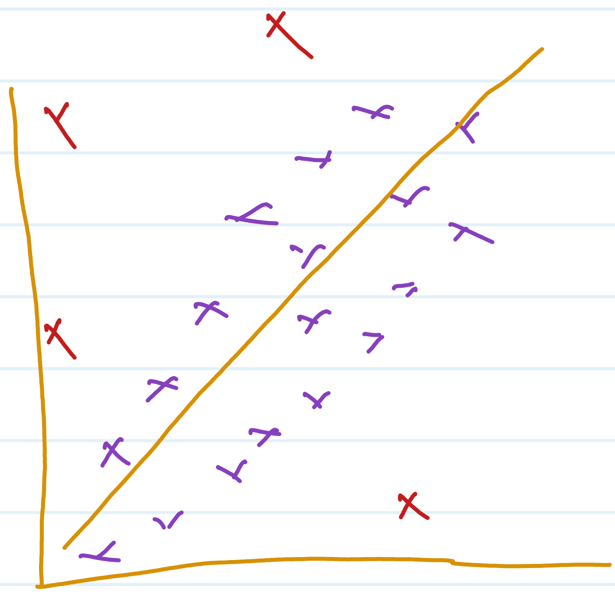
$\underbrace{X_1 \quad X_2}_{\text{feature column}}$ $\underbrace{Y}_{\text{Target column}}$

feature column → Independent ~~column~~ variable

Target column → Dependent ~~column~~ variable

Disadvantages of P. C. C. is,

It is able to capture linear property but is not able to capture non-linear property.



Spearman Rank correlation

Power law Dist.

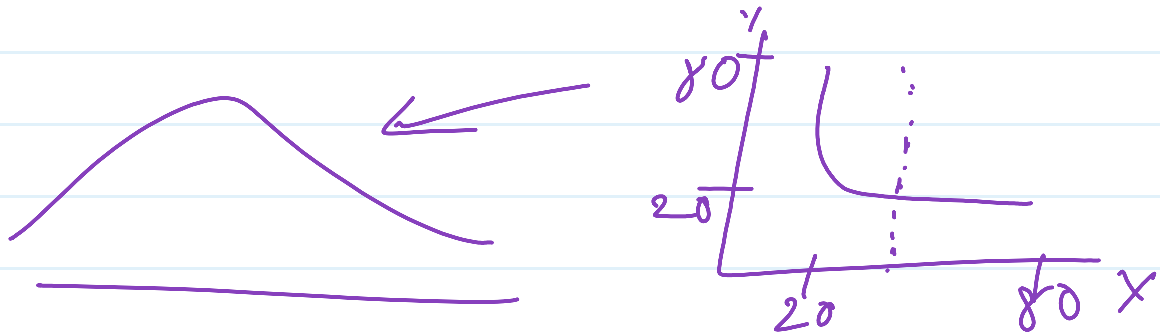
80 - 20

pareto Dist.

→ 80% Run — 20% Batters

→ 80% work - 20% employee.

→ 80% family income - 20% parent



To convert pareto dist. into Gaussian Dist. by "box-cox transformation"

$$y_i = \frac{x_i^{\alpha} - 1}{\alpha}$$

α is weight of Dist

if $\alpha \neq 0$, Box-cox transformation

if $\alpha = 0$, Log transformation