

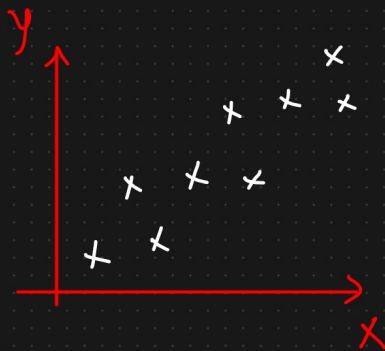
# Covariance And Correlation

[Relationship between X and Y]

X	Y
2	3
4	5
6	7
8	9

X ↑	Y ↑
X ↓	Y ↑
X ↓	Y ↓
X ↑	Y ↓

↓ ↑ Size of house → Price of house ↓ ↑



X ↑	Y ↑
X ↓	Y ↓



X ↑	Y ↓
X ↓	Y ↑

## Covariance

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

$x_i \rightarrow$  Datapoints of  $x$

$\bar{x} \rightarrow$  Sample mean of  $x$

$y_i \rightarrow$  Datapoints of  $y$

$\bar{y} \rightarrow$  Sample mean of  $y$

$$\begin{aligned} \text{Var}(x) &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{n-1} \end{aligned}$$

$$= \text{Cov}(x, x) \Rightarrow \text{Spread}$$

$Cov(x, y)$

$x \uparrow$	$y \uparrow$
$x \downarrow$	$y \downarrow$

 +ve Covariance

$x \uparrow$	$y \downarrow$
$x \downarrow$	$y \uparrow$

 -ve Covariance

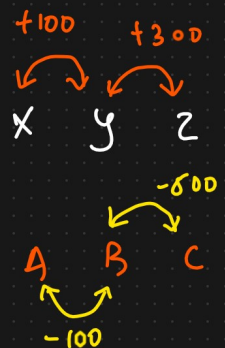
X	Y
→ 2	3
→ 4	5
→ 6	7
<u>          </u>	<u>          </u>
$\bar{x} = 4$	$\bar{y} = 5$

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

$$= \frac{[(2-4)(3-5) + (4-4)(5-5) + (6-4)(7-5)]}{n-1}$$

$$= \frac{-4 + 0 + 4}{2} = \frac{0}{2} = 0 \text{ tve value}$$

↓  
Positive  
Covariance



X & Y are having a positive Covariance

### Advantages

- ① Relationship between X and Y  
+ve or -ve value

### Disadvantages

- ① Covariance does not have a specific limit value

## ② Pearson Correlation Coefficient $[-1 \text{ to } 1]$

$$\rho_{x,y} = \frac{\text{Cov}(x,y)}{\sigma_x \cdot \sigma_y}$$

① The more the value towards  $+1$  the more +ve correlated it is  $(x,y)$

② The more the value towards  $-1$  the more -ve correlated it is  $(x,y)$

## ③ Spearman Rank Correlation $[-1 \text{ to } 1]$

$$\rho_s = \frac{\text{Cov}(R(x), R(y))}{\sigma(R(x)) * \sigma(R(y))}$$

$x$	$y$	$R(x)$	$R(y)$
1	2	5	5
3	4	4	4
5	6	3	3
7	8	2	1
0	7	6	2
8	1	1	6

## Feature Selection

+ve  
Size of  
house ↑

+ve  
No. of  
Rooms ↑

+ve  
Location ↑

~0  
~~No. of people  
staying~~

-ve  
Haunted

Price ↑