Tuesday, February 20, 2024 12:01 AM

= COVARIANCE =

$$Cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$ → population covariance

$$Cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$ → sample covariance

Covariance Matrix

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY}^2 & \sigma_{XZ}^2 \\ \sigma_{XY}^2 & \sigma_Y^2 & \sigma_{YZ}^2 \\ \sigma_{XZ}^2 & \sigma_{YZ}^2 & \sigma_Z^2 \end{bmatrix}$$

multivariate normal

$$\phi(x) = \frac{1}{\sqrt{(2\pi)^2 |\hat{\Sigma}|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \hat{\Sigma}^{-1} (\vec{x}-\vec{\mu})}$$

Another formula:

$$Cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right] = E(XY) - E(X)E(Y)$$

Significance:

- Covariance +ve ⟹ X, Y increase/decrease together
- Covariance −ve ⟹ As X increases, Y decreases and vice versa
- Covariance = 0 ⟹ X and Y are uncorrelated

Units:

(X units)(Y units)

EXAMPLE:

	$X=-1$	$X=0$	$X=1$	$P(Y=t)$
$Y=-1$	1/15	2/15	2/15	$\frac{1}{3}$
$Y=0$	2/15	1/15	2/15	$\frac{1}{3}$
$Y=1$	2/15	2/15	1/15	$\frac{1}{3}$
$P(X=t)$	$\frac{5}{15} = \frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

$$\sum E\left[(x - E(x))(y - E(y))\right]$$

$$E(x) = E(y) = 0$$

$$\sum_{i=1}^{m} X_i P(X = x_i)$$

$$Cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$$

$$= (-1)(-1)\left(\frac{1}{15}\right) + (-1)(1)\left(\frac{2}{15}\right) + (1)(-1)\left(\frac{2}{15}\right) + (1)(1)\left(\frac{1}{15}\right)$$

$$= \frac{1}{15} - \frac{2}{15} - \frac{2}{15} + \frac{1}{15}$$

$$= \frac{-2}{15}$$

Properties

① $Cov(X,X) = Var(X)$

② $Cov(X,Y) = E(XY) - E(X)E(Y)$

③ Covariance is symmetric

$Cov(X,Y) = Cov(Y,X)$

④ Linear qty.

$Cov(X, aY + bZ) = a\,Cov(X,Y) + b\,Cov(X,Z)$

$Cov(aX + bY, Z) = a\,Cov(X,Z) + b\,Cov(Y,Z)$

= COVARIANCE & INDEPENDENCE =

- X and Y are independent $\Longrightarrow$ X and Y are uncorrelated ($Cov(X,Y) = 0$)
- Converse need not be true.
  Uncorrelated X, Y may still be dependent.
- Being independent is a much stronger notion of being unrelated compared to being uncorrelated.

= CORRELATION COEFFICIENT =

Correlation coefficient $\longrightarrow$ normalised version of covariance $\longrightarrow \rho / r$

$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{SD(X)\,SD(Y)}}$$

Sir Francis Galton, 1888

Properties

- $-SD(X)SD(Y) \leq Cov(X,Y) \leq SD(X)SD(Y)$
- $E\left[\left(\frac{X - E(X)}{SD(X)} + \frac{Y - E(Y)}{SD(Y)}\right)^2\right] \geq 0$ $\longrightarrow$ lower bound

  $E\left[\left(\frac{X - E(X)}{SD(X)} - \frac{Y - E(Y)}{SD(Y)}\right)^2\right] \geq 0$ $\longrightarrow$ upper bound
- $-1 \leq \rho(X,Y) \leq 1$
- Summarises trend b/w two random variables
- Dimensionless qty

Results:

① $\rho(X,Y)$ close to 0

$\Longrightarrow$ X, Y close to being uncorrelated; no clear trend

② $\rho(X,Y) = 1$ / $\rho(X,Y) = -1$

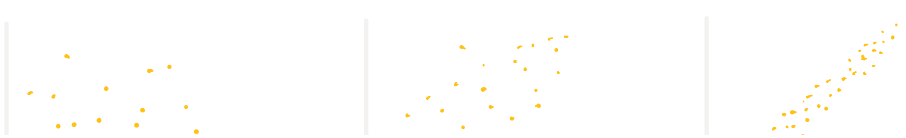
$\Longrightarrow$ There exists $a \neq 0$, b such that $Y = aX + b$ with probability 1.

Y is a linear function of X.

③ $|\rho(X,Y)| \longrightarrow 1$

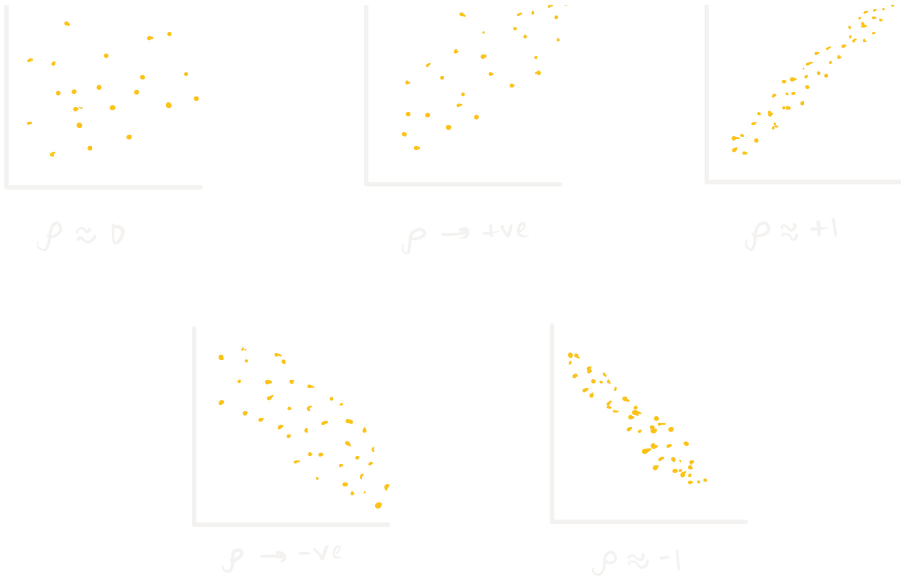
$\Longrightarrow$ X, Y have strong correlation; increase in X likely to match up with increase in Y.

= SCATTER PLOTS =



NOTE:

Correlation coefficient remains unaltered even if:



Correlation coefficient remains unaltered even if:

- X interchanged w/ Y
- Adding some constant to each value of a variable
- Multiplying each value by a positive constant

= CONFOUNDING VARIABLE =

- Hidden variable that disrupts your analysis by influencing both the independent variable and the dependent variable
- Increases variance and/or introduces bias