= COVARIANCE =

$$\left.\begin{array}{l} Cov(X,Y) = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N} \end{array}\right] \rightarrow \text{population covariance}$$

$$\left.\begin{array}{l} Cov(X,Y) = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} \end{array}\right] \rightarrow \text{sample covariance}$$

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

Units:

(X units)(Y units)

EXAMPLE:

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

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

$$E(X) = E(Y) = 0$$

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

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

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

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

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

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.

$$\text{Cov}(X, aY + bZ) = a\,\text{Cov}(X,Y) + b\,\text{Cov}(X,Z)$$
$$\text{Cov}(aX + bY, Z) = a\,\text{Cov}(X,Z) + b\,\text{Cov}(Y,Z)$$

= COVARIANCE & INDEPENDENCE =

- X and Y are independent $\Longrightarrow$ X and Y are uncorrelated ($\text{Cov}(X,Y) = 0$)
- 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 $\longrightarrow$ ρ / r

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

Properties

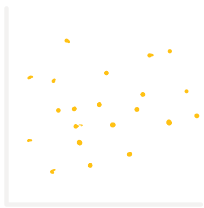
- $-SD(X)SD(Y) \le \text{Cov}(X,Y) \le SD(X)SD(Y)$
- $E\left[\left(\frac{X - E(X)}{SD(X)} + \frac{Y - E(Y)}{SD(Y)}\right)^2\right] \ge 0$ $\rightarrow$ lower bound

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

Results:

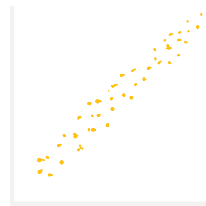
1. $\rho(X,Y)$ close to 0
   $\Rightarrow$ X, Y close to being uncorrelated; no clear trend
2. $\rho(X,Y) = 1$ / $\rho(X,Y) = -1$
   $\Rightarrow$ There exists $a \neq 0, b$ such that $Y = aX + b$ with probability 1.
   Y is a linear function of X.
3. $|\rho(X,Y)| \longrightarrow 1$
   $\Rightarrow$ X, Y have strong correlation; increase in X likely to match up with increase in Y.

= SCATTER PLOTS =

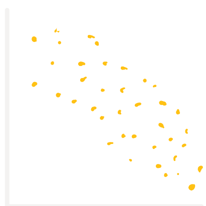


$\rho \approx 0$

$\rho \to +ve$



$\rho \approx +1$



$\rho \to -ve$

$\rho \approx -1$