Covariance
&
Correlation

Covariance

- Variables may change in relation to each other
- *Covariance* measures how much the movement in one variable predicts the movement in a corresponding variable

Covariance of X and Y

$$\mathrm{Cov}(X,Y) =$$

$$= \left[E(X-\mu_X)(Y-\mu_Y) \right]$$

$$= \sum_x \sum_y (x-\mu_x)(y-\mu_y)\, p(x,y)$$

if discrete

$$= \iint (x-\mu_x)(y-\mu_y)\, f(x,y)\, dx\, dy$$

if continuous

$$\mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n} (x - \mu_x)(y - \bar{\mu}_y)}{n-1}$$

Covariance Formula

For Population

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

For Sample

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

Smoking v Lung Capacity Data

n	Cigarettes (x)	Lung Capacity (y)
1	0	45
2	5	42
3	10	33
4	15	31
5	20	29

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})\cdot(y-\bar{y})$
0	45	-10	9	-90
5	42	-5	6	-30
10	33	0	-3	0
15	31	5	-5	-25
20	29	10	-7	-70

$\Sigma = -215$

$$\bar{x} = \frac{\Sigma x}{n} = 10$$

$$\bar{y} = \frac{\Sigma y}{n} = 36$$

Co-variance

$$= \frac{\Sigma (x-\bar{x})(y-\bar{y})}{n-1}$$

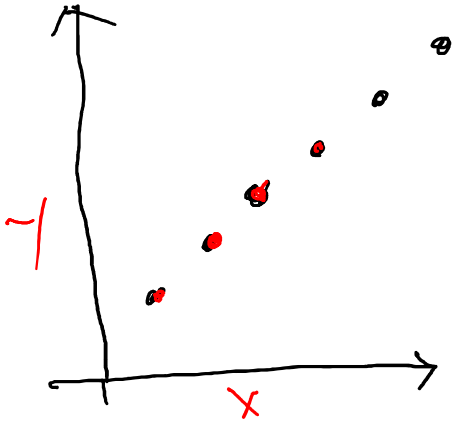
$$= \frac{-215}{5-1}$$

$$= -53.75$$

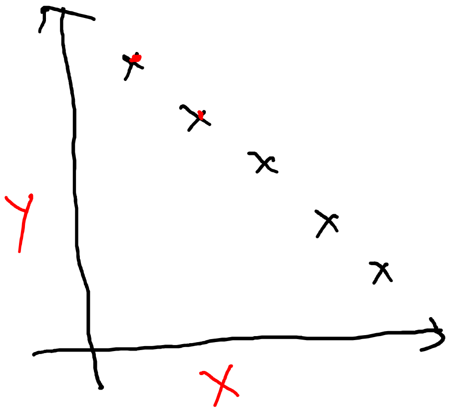
Correlation

- Finding the relationship between two quantitative variables without being able to infer causal relationships

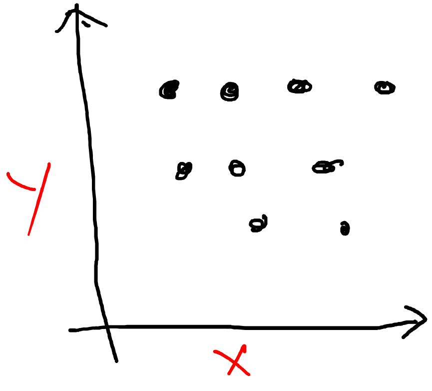
Correlation is a statistical technique used to determine the degree to which two variables are related



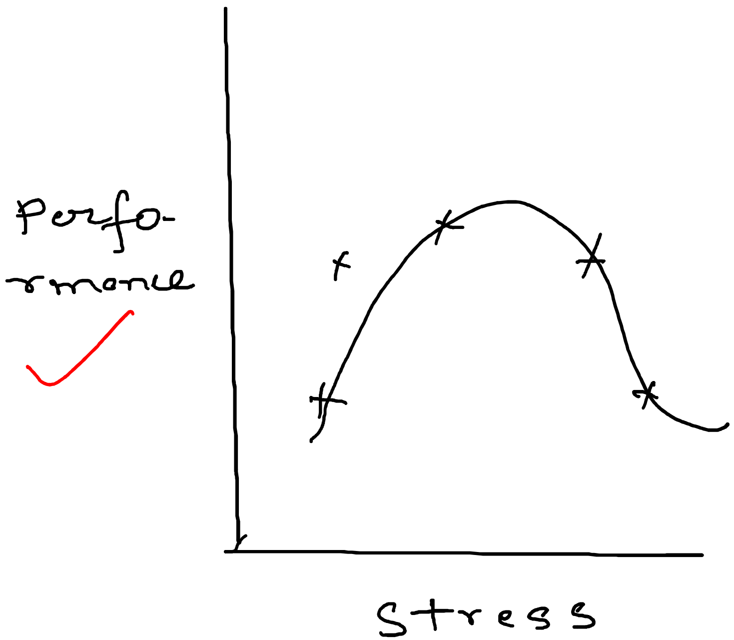
positive correlation

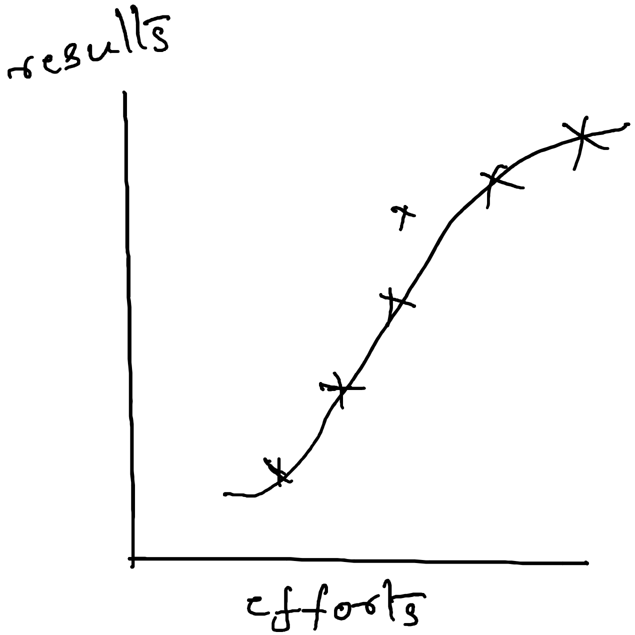


Negative correlation



No correlation


Performance
Stress


results
efforts

Coefficient of correlation:

$$r = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}}$$

where

$$X = x - \bar{x}$$

$$Y = y - \bar{y}$$

$$X^2 = (x - \bar{x})^2$$

$$Y^2 = (y - \bar{y})^2$$

Coefficient of Correlation

$r = 1 \Rightarrow$ Perfect and positive relation

$r = -1 \Rightarrow$ " " negative relation

$r = 0 \Rightarrow$ No relation

$0 < r < 1 \Rightarrow$ Partial positive relation

$-1 < r < 0 \Rightarrow$ " negative "

Example - 1

x	1	2	3	4	5	6	7	8	9
y	10	11	12	14	13	15	16	17	18

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{9} = 14$$

x	$X = x-5$	X^2	y	$Y = y-14$	Y^2	XY
1	-4	16	10	-4	16	16
2	-3	9	11	-3	9	9
3	-2	4	12	-2	4	4
4	-1	1	14	0	0	0
5	0	0	13	-1	1	0
6	1	1	15	1	1	1
7	2	4	16	2	4	4
8	3	9	17	3	9	9
9	4	16	18	4	16	16
					60	Σ 59

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \, \Sigma Y^2}}$$

$$= \frac{59}{\sqrt{60 \times 60}}$$

$$= 0.9833$$

x	$X = x-5$	X^2	y	$Y = y-14$	Y^2	XY
1	-4	16	10	-4	16	16
2	-3	9	11	-3	9	9
3	-2	4	12	-2	4	4
4	-1	1	14	0	0	0
5	0	0	13	-1	1	0
6	1	1	15	1	1	1
7	2	4	16	2	4	4
8	3	9	17	3	9	9
9	4	16	18	4	16	16
		60			60	Σ 59

$$\therefore \ \mathrm{cov}(x,y) = \frac{\Sigma XY}{n-1} = \frac{59}{8} = 7.375$$

Coefficient of Determination

r is coeff. of correlation

r^2 is coeff of determination

↓

Indicates the extent to which variation in one variable is explained by the variation in the other.

$r = 0.9 \Rightarrow r^2 = 0.81$

ie 81% of the variation in y due to variation in 'x'

remaining 19% is due to some other factors.

$r = 0.9833$

$\mathrm{cov}(x, y) = 7.375$

Interpretation

$r^2 = 0.81$

Thanks