# Notes

Monday, August 15, 2022 8:56 AM

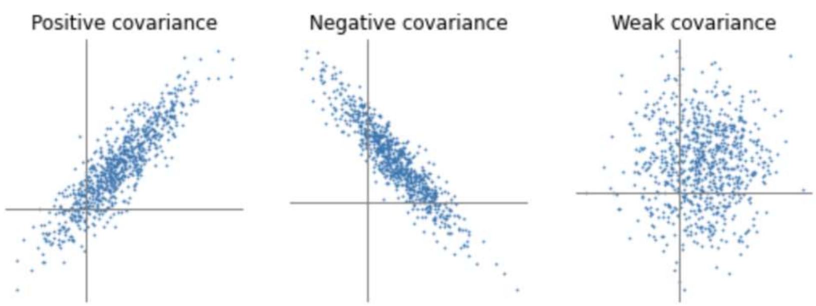
*Myroslava Sánchez Andrade A01730712 | 14/08/2022*

## 3 Measures of linear relationship

The linear (proportional change) relationships measure whether there is a pattern of *movement* for a random variable when another variable moves up or down. The main two measures of linear relationship between 2 random variables are:

### - Covariance

Is a measure of the joint probability of two random variables. If the greater the values of one variable mainly correspond with the greater values of the other variable => the covariance is positive [similar behavior]; if the greater values of one variable mainly correspond to the lesser values of other => the covariance is negative [opposite behavior].
The **sign** of the variance shows the tendency in the linear relationship between the variables. We cannot understand the magnitude of covariance; only its sign (+, -, 0).




The covariance is defined as the expected value (mean) of the product of their deviations from the mean of each variable:

$$\operatorname{cov}(X, Y) = \sum_{i=1}^{N} \frac{(x_i - \overline{x})(y_i - \overline{y})}{N}.$$

Example:

Suppose that $X$ and $Y$ have the following joint probability mass function,[6] in which the six central cells give the discrete joint probabilities $f(x, y)$ of the six hypothetical realizations $(x, y) \in S = \{(5,8),(6,8),(7,8),(5,9),(6,9),(7,9)\}$:

| $f(x,y)$ | | x | | | $f_Y(y)$ |
|---|---|---|---|---|---|
| | | **5** | **6** | **7** | |
| **y** | **8** | 0 | 0.4 | 0.1 | 0.5 |
| | **9** | 0.3 | 0 | 0.2 | 0.5 |
| $f_X(x)$ | | 0.3 | 0.4 | 0.3 | 1 |

$X$ can take on three values (5, 6 and 7) while $Y$ can take on two (8 and 9). Their means are $\mu_X = 5(0.3) + 6(0.4) + 7(0.1 + 0.2) = 6$ and $\mu_Y = 8(0.4 + 0.1) + 9(0.3 + 0.2) = 8.5$. Then,

$$\begin{aligned}\operatorname{cov}(X,Y) = \sigma_{XY} &= \sum_{(x,y)\in S} f(x,y)\,(x-\mu_X)\,(y-\mu_Y) \\ &= (0)(5-6)(8-8.5) + (0.4)(6-6)(8-8.5) + (0.1)(7-6)(8-8.5) + \\ &\quad (0.3)(5-6)(9-8.5) + (0)(6-6)(9-8.5) + (0.2)(7-6)(9-8.5) \\ &= -0.1\,.\end{aligned}$$

## - Correlation

Statistical relationship between two random variables. It refers to the degree to which a pair of variables are linearly related, thus, the interpretation of the Covariance in percentage.

$$Corr(x,y) = \frac{Cov(x,y)}{SD(x) \cdot SD(y)}$$

Then:

$$P_{xy} = \frac{\sigma_{xy}}{\sigma_x \ \sigma_y} = \frac{\sum (xi - \bar{x})\ (yi - \bar{y})}{\sqrt{\left(\sum (xi - \bar{x})^2\right)\left(\sum (yi - \bar{y})^2\right)}}$$

Where, $\sigma_x, \sigma_y \rightarrow$ Population Standard Deviation

$\sigma_{xy} \rightarrow$ Population Covariance

$\bar{X}, \bar{Y} \rightarrow$ Population Mean

If Corr(x, y) = 0.30, then about 30% of the cases, when x goes up, y goes up; and when x goes down, y goes dow.
If Corr(x, y) = -0.30, then about 30% of the cases, when x goes up, y goes down, and when x goes down, y goes up.