REVISTA DE CIÊNCIA ELEMENTAR

Covariância amostral

Maria Eugénia Graça Martins
Universidade de Lisboa
memartins@fc.ul.pt

CITAÇÃO
Martins, E.G.M. (2018)
Covariância amostral,
Rev. Ciência Elem., V6(01):022.
doi.org/10.24927/rce2018.022

EDITOR
José Ferreira Gomes,
Universidade do Porto

EDITOR CONVIDADO
Luís Vítor Duarte,
Universidade de Coimbra

RECEBIDO EM
03 de fevereiro de 2012

ACEITE EM
28 de janeiro de 2018

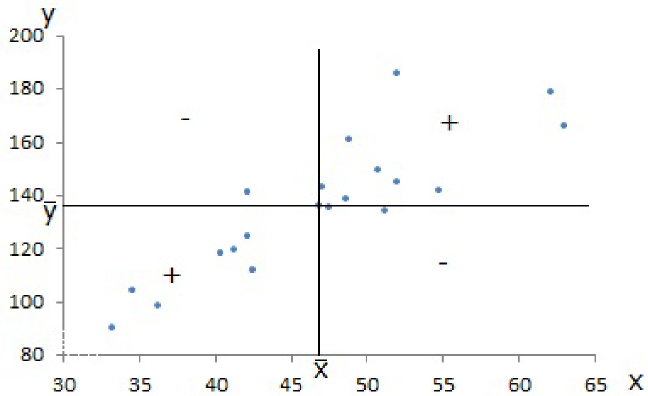
PUBLICADO EM
14 de março de 2018

A Covariância amostral entre duas variáveis, de tipo quantitativo, descreve a direção e o grau com que as variáveis se associam linearmente.

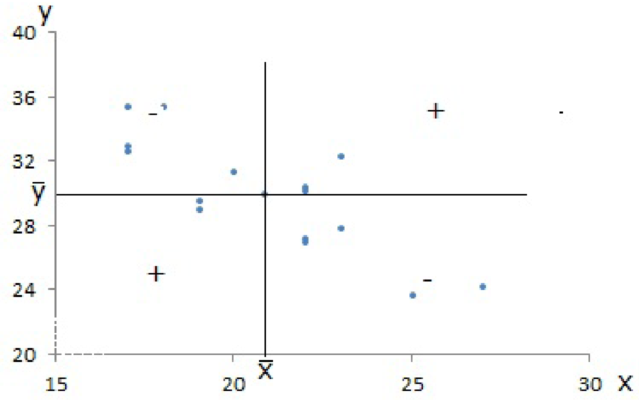
Se representarmos por **(x,y)**={ (x_i,y_i) }, com i = 1,...,n, uma amostra de dados bivariados, a **covariância amostral** entre as variáveis **x** e **y** é dada pela seguinte expressão:

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

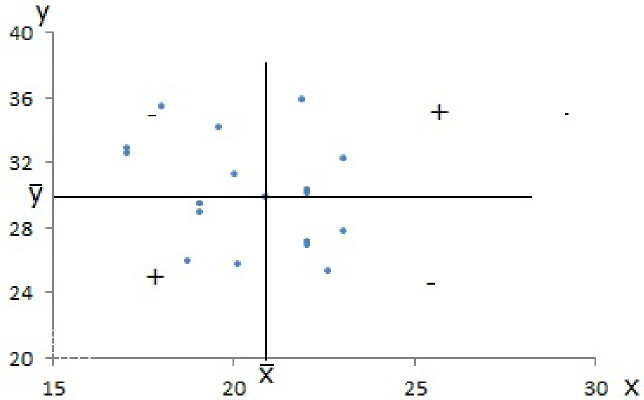
Uma associação linear entre os x's e os y's, do mesmo sentido, isto é, quando a valores grandes (pequenos) de x correspondem, de um modo geral, valores grandes (pequenos) de y, faz com que predominem as parcelas positivas na expressão da covariância, pois quando $(x_i-\bar{x}) > 0 \ (< 0)$, tende a ser $(y_i-\bar{y}) > 0 \ (< 0)$. Então a covariância vem positiva. Geometricamente, tem-se:



Uma associação linear entre os x's e os y's, de sentido contrário, isto é, quando a valores grandes (pequenos) de x correspondem, de um modo geral, valores pequenos (grandes) de y, faz com que predominem as parcelas negativas na expressão da covariância, pois quando $(x_i-\bar{x}) > 0 \ (< 0)$, tende a ser $(y_i-\bar{y}) > 0 \ (< 0)$. Então a covariância vem negativa. Geometricamente, tem-se:



Se não se verificar uma associação linear entre as variáveis, então nem predominam as parcelas positivas, nem as negativas, obtendo-se para a covariância um valor próximo de 0. Geometricamente tem-se:



A covariância é uma medida que tem o inconveniente de depender das unidades com que se apresentam os elementos da amostra, pelo que não é normalmente usada. Em sua substituição utiliza-se o coeficiente de correlação amostral, que não depende das unidades utilizadas.