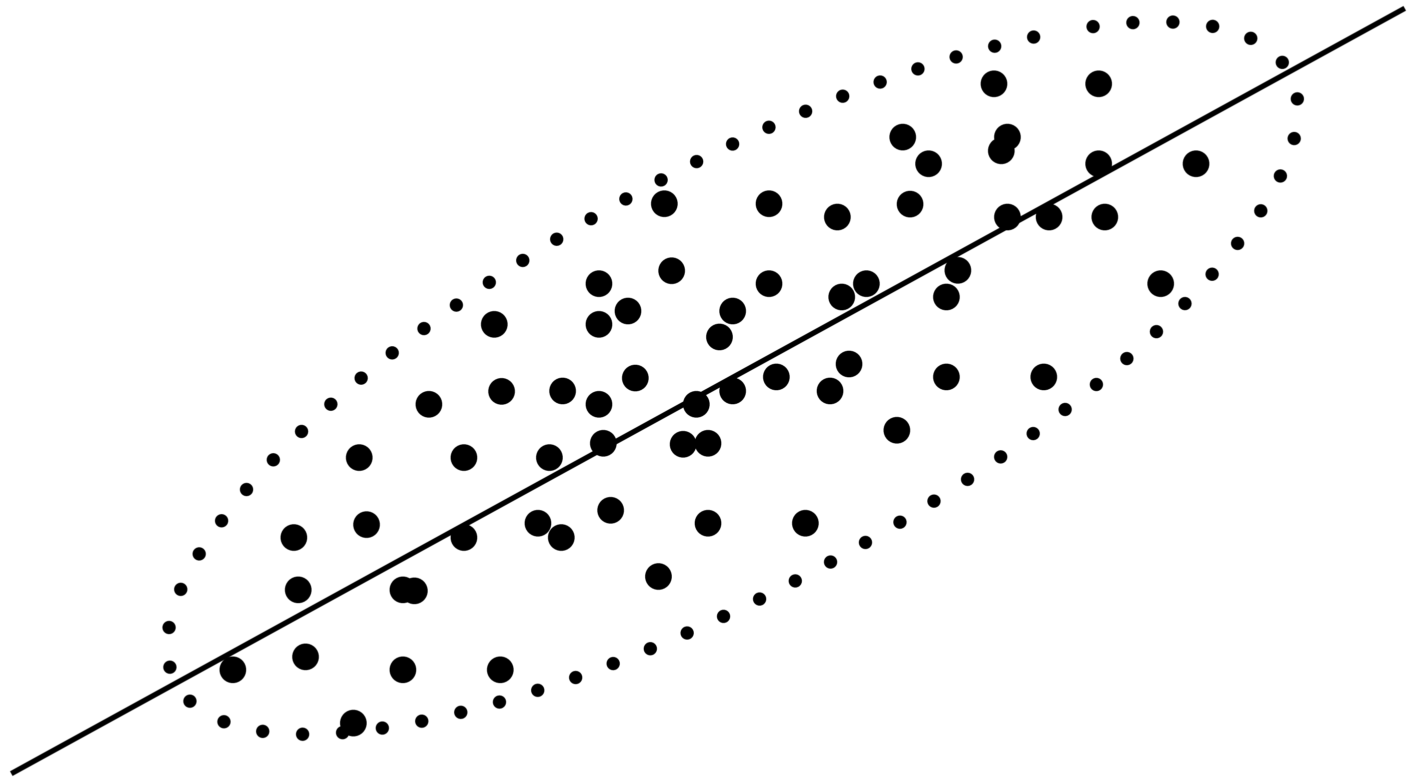
Esempio del podcast



Le 3 coppie di punti: $(2; 10), (4; 14), (6; 15)$.

La media di X è $\bar{x} = \frac{1}{3}(2 + 4 + 6) = 4$. Quella di Y è $\bar{y} = 13$.

La covarianza tra X e Y è

$$\text{cov}(X, Y) = \frac{1}{3}\Big((2-4)(10-13) + (4-4)(14-13) + (6-4)(15-13)\Big) = \frac{10}{3} = 3.3\bar{3}.$$

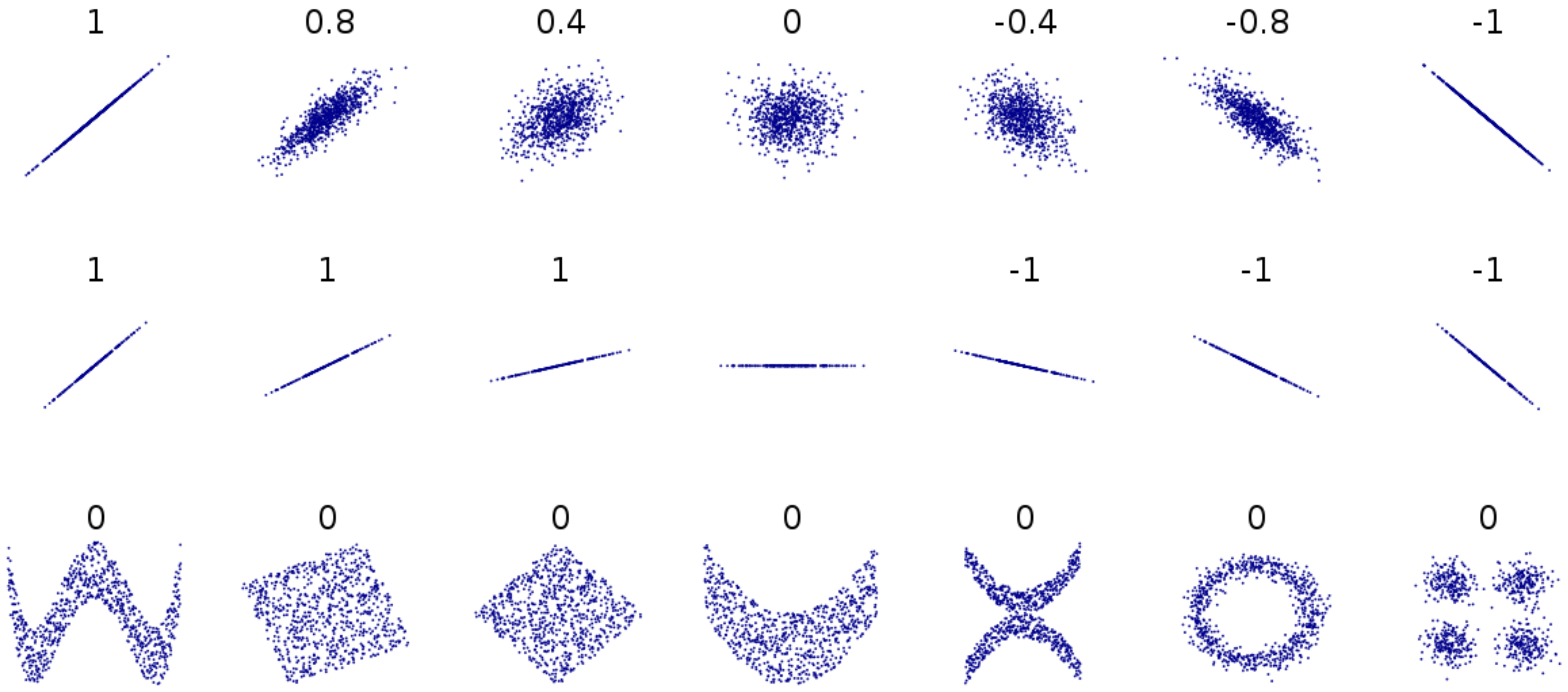
Se si divide per $n - 1 = 2$ (più corretto visto il campione piccolo*) si ottiene $\text{cov}(X, Y) = 5$.

* Ricordate la differenza tra valore teorico e valore empirico per varianza e deviazione standard? È la stessa motivazione.

Nuvole di punti (X,Y) e correlazione di Pearson

Se la correlazione ha senso, valori più alti (in senso assoluto) indicano una migliore approssimazione della retta che dovrebbe legare X e Y. Valori minori indicano che i punti sono più dispersi nello spazio.

Qui la correlazione di Pearson non ha senso teoricamente, e nella pratica sarà misurata vicino a 0.



Fonte: Wikipedia CC0

Le formule

$$\text{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Per campioni piccoli, si usa $n - 1$ nel calcolo.

$$\rho(X, Y) = \frac{\overbrace{\text{cov}(X, Y)}^{\text{covarianza}}}{\underbrace{\text{sd}(X)\text{sd}(Y)}_{\text{Prodotto delle dev. standard}}} \in [-1, 1]$$

Correlazione di Pearson

@DrCirillo