Coeficiente de Correlação

Até o momento medimos a $E(X)$ e a $Var(X)$, ou seja, uma medida de posição e de variabilidade em relação a R_X. Entretanto, pode-se ter um vetor bidimensional, (X, Y) e outra medida surge, a qual tenta medir o "grau de associação" linear entre X e Y.

Definição:

Seja (X, Y) uma variável aleatória bidimensional, o coeficiente de correlação,$\rho_{X,Y}$, entre X e Y será:

$$\rho_{X,Y} = \frac{E[(X - E(X)(Y - E(Y)]}{\sqrt{Var(X)Var(Y)}}$$

Sendo a ***Covariância*** entre X e Y, $Cov(X, Y)$, dada por:

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

Algumas Propriedades da Correlação:

Teorema:

O coeficiente de correlação,$\rho_{X,Y}$, entre X e Y pode ser assim apresentado::

$$\rho_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}}$$

Demonstração:

$$E[(X - E(X))(Y - E(Y))] = E[XY - XE(Y) - E(X)Y + E(X)E(Y)]$$

$$= E(XY) - E(X)E(Y) - \cancel{E(X)E(Y)} + \cancel{E(X)E(Y)} = E(XY) - E(X)E(Y)$$

Teorema:

Se X e Y forem independentes então:

$$\rho_{X,Y} = 0$$

Demonstração:

Da propriedade da esperança sobre independência se X e Y forem independentes:

$$E(XY) = E(X)E(Y)$$

Considerando o primeiro teorema apresentado nessa seção tem-se que:

$$\rho_{X,Y} = 0$$

OBSERVE:

$Independ\hat{e}ncia \Rightarrow \rho_{X,Y} = 0$

mas

$\rho_{X,Y} = 0 \nRightarrow Independ\hat{e}ncia$

Teorema:

O Coeficiente de Correlação possui valores entre -1 e 1, ou seja:

$$-1 \leq \rho \leq 1$$

Demonstração:

Considere a seguinte desigualdade verdadeira:

$$\left(\frac{X - \mu_X}{\sigma_X} \mp \frac{X - \mu_Y}{\sigma_Y}\right)^2 \geq 0$$

A expressão continua verdadeira se aplicarmos o operador esperança:

$$E\left(\frac{X - \mu_X}{\sigma_X} \mp \frac{X - \mu_Y}{\sigma_Y}\right)^2 \geq 0$$

Desenvolvendo tem-se:

$$E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2+\left(\frac{X-\mu_Y}{\sigma_Y}\right)^2\mp 2\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{X-\mu_Y}{\sigma_Y}\right)\right]\geq 0$$

$$\left[\frac{1}{\sigma_X^2}E(X-\mu_X)^2+\frac{1}{\sigma_Y^2}E(X-\mu_Y)^2\mp 2\frac{1}{\sigma_X\sigma_Y}E(X-\mu_X)(X-\mu_Y)\right]\geq 0$$

$$\left[\frac{\sigma_X^2}{\sigma_X^2}+\frac{\sigma_Y^2}{\sigma_Y^2}\mp 2\rho\right]\geq 0$$

$$\mp 2\rho \geq -2$$

$$\therefore \begin{matrix}\rho \geq -1\\ \rho \leq 1\end{matrix}$$

Teorema:

Sendo $\rho^2 = 1$ tem-se que $Y = AX + B$, onde A e B são constantes. Assim, se $\rho = \pm 1$ então Y será uma função linear de X

Derivação:

Ainda não disponível para essas notas de aula.

Teorema:

Se X e Y forem duas variáveis aleatórias, onde $Y = AX + B$, onde A e B são constantes. Então $\rho^2 = 1$. Se $A > 0, \rho = 1$. Se $A < 0, \rho = -1$.

Demonstração:

$$Y = AX + B$$

$$E(Y) = AE(X) + B \text{ e } VAR(Y) = A^2VAR(X)$$

$$E(XY) = E(AX^2 + BX) \rightarrow AE(X^2) + BE(X)$$

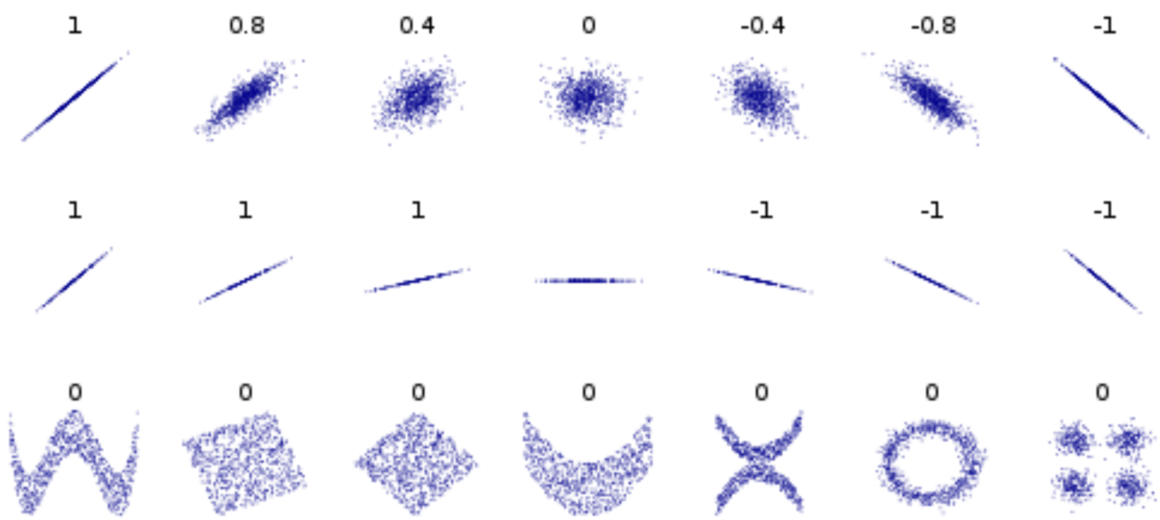
$$\rho^2 = \frac{[E(XY) - E(X)E(Y)]^2}{Var(X)Var(Y)} = \frac{[AE(X^2) + BE(X) - E(X)[AE(X) + B]]^2}{A^2Var(X)Var(X)}$$

$$= \frac{[AE(X^2) - A\big(E(X)\big)^2 + \cancel{BE(X)} - \cancel{BE(X)}]^2}{A^2(Var(X))^2} = \frac{[A(E(X^2) - E(X)^2)]^2}{A^2Var(X)}$$

$$\rho^2 = \frac{A^2Var(X)^2}{A^2Var(X)^2} = 1$$

Com base no exposto acima, tem-se que o coeficiente de correlação é uma medida do grau de linearidade entre X e Y. Dessa forma ρ próximo a 1 e -1 indicam alto grau de linearidade e ρ próximo a zero indica ausência de relação linear, mas não diz nada sobre relações não-lineares.

Vejamos alguns exemplos alguns scatterplot de dados e suas correlações:



Outras propriedades:

$$\rho(X + A;\ Y + b) \ = \ \rho(X, Y)$$

$$\rho(\mathrm{A}X; \mathrm{B}Y) = \frac{AB}{|AB|}\ \rho(X, Y)$$

Sendo

$\mathrm{AB}\ >\ 0 \rightarrow \rho(\mathrm{x}, \mathrm{y})$

$\mathrm{AB}\ <\ 0 \rightarrow -\rho(\mathrm{x}, \mathrm{y})$