2100122 - Oceanografia Integrativa I
Aula Correlação e Tendência

Eduarda Valério de Jesus

May 2024

1 Covariância

A covariância é uma medida estatística onde é possível comparar duas variáveis, permitindo entender como elas se relacionam entre si.

$$C_{xy} = \langle x'y' \rangle \tag{1}$$

Sabendo que $\langle x \rangle$ é dado por:

$$\langle x \rangle = \frac{1}{N}\sum_{i=1}^{N}(x) \tag{2}$$

A equação (2) é a chamada média amostral.
Ainda sim, temos que $x' = x - \langle x \rangle$, igualmente para y', $y' = x - \langle y \rangle$.

A covariância entre a mesma amostra resulta na **variância**.

$$C_{xx} = \langle x'^2 \rangle = \langle (x_i - \langle x \rangle)^2 \rangle = \frac{1}{N}\sum_{i=1}^{N}(x_i - \langle x \rangle)^2 \tag{3}$$

2 Coeficiênte de correlação

O coeficiente de correlação varia entre 1 e -1 e mede o grau de relação entre duas séries temporais, é representado por:

$$\rho_{xy} = \frac{\langle x'y' \rangle}{\sqrt{\langle x'^2 \rangle \langle y'^2 \rangle}} \tag{4}$$

$$\rho_{yx} = \frac{\langle x'y' \rangle}{\sqrt{\langle y'^2 \rangle \langle x'^2 \rangle}} \tag{5}$$

ρ = 1: Correlação perfeita
ρ = -1: Anticorrelação, as séries são espelhadas

3 Relação linear

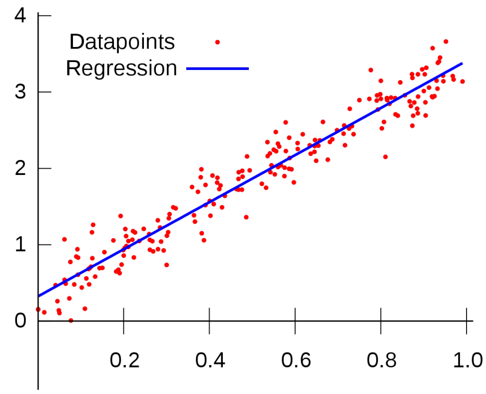

Figure 1: representação gráfica de relação linear

A reta de regressão representa o modelo, onde a correlação é perfeita ($\rho = 1$). A equação do modelo é dada por, $\hat{y}' = \alpha x'$.

A partir do **ERRO MÉDIO QUADRÁTICO** podemos minimizar a diferença entre o modelo e a correlação com os valores oferecidos. O erro médio quadrático é calculado por (6).

$$E = \langle (y' - \hat{y}')^2 \rangle \tag{6}$$

Manipulando a equação (6) chegamos em:

$$E = \langle (y')^2 \rangle - 2\alpha \langle y'x' \rangle + \alpha^2 \langle (x')^2 \rangle \tag{7}$$

A partir de (7) devemos encontrar o valor de alfa que minimiza o erro. Para isso igualamos a derivada do erro médio quadratico a 0.

$$\frac{\partial E}{\partial \alpha} = 0 \tag{8}$$

$$\frac{\partial}{\partial \alpha} \langle y'^2 \rangle - \frac{\partial}{\partial \alpha} (2\alpha \langle x'y' \rangle) + \frac{\partial}{\partial \alpha} (\alpha^2 \langle x'^2 \rangle) = 0 \tag{9}$$

$$\alpha = \frac{\langle x'y' \rangle}{\langle x'^2 \rangle} \tag{10}$$

4 Relação entre α e ρ

A variância explicada é a fração de variância que o modelo consegue explicar, sendo uma medida da eficiência do modelo.
Essa fração também é chamada de **skill** e representa a relação entre o modelo e a variância dos dados.

$$Skill = \frac{\langle \hat{y}'^2 \rangle}{\langle y'^2 \rangle} \tag{11}$$

$$Skill = \frac{\langle (\alpha x')^2 \rangle}{\langle y'^2 \rangle} \tag{12}$$

Substituindo α em (12).

$$Skill = \frac{\langle x'y' \rangle^2}{\langle x'^2 \rangle \langle y'^2 \rangle} = \rho_{xy}^2 \tag{13}$$