

Modelo de Regressão

Material suplementar de <https://www.youtube.com/watch?v=4Z52eEKVorY>
Psicometristas Brasil (março de 2016)

Prof. Luis Anunciação

PUC-Rio

Psicometrista

Correlação (r)	Regressão (r ²)
Simétrico: $r(y,x) = r(x,y)$	Estimativas != real
Limitado: $ r(y,x) \leq 1$	Previsões (%)
Linear: $r(y,x) = \pm 1 \leftrightarrow Y = a + bX, (a,b) \in \mathbb{R}^2$	“Causalidade”
	-> Há uma associação
	-> A VI vem antes da VD (precedência temporal)
	-> A VD desaparece com a retirada da VI
	Causalidade de Granger: “o futuro não pode causar o presente!”
	9 critérios de Hill (Exposição e Doença) [Evidência experimental]

Noção de causalidade vs previsão



Mundo real

$$y = a + bX$$
$$y = f(x)$$
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots$$

Objetos

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + u$$

$U \sim N(\mu, DP)$ (Nor)
 $E(u|X) = 0$ (Média)
 $VAR(u) = (Homo)$
 $COV(u, x) = 0$ (Indep.)

Quando não houver, a previsão do modelo é em comparação com à média

Linear simples: preço do apartamento dado área + u

$$y = \beta_0 + \beta_1 x + u$$

(Ridge ou cumeeira: altera os coeficientes pela regularização de Tikhonov)

Múltipla: preço do apartamento dado área + vista + u

$$y = \beta_0 + \sum \beta_i x_{ij} + u$$

Logística (dicotômico, ordenada, multinomial): aprovação em estatística dado horas de estudo

$$y \in \{1, 0\}$$

Distribuição binomial: $y \sim \binom{n}{k} = \frac{n!}{k!(n-k)!}$

$$p(y = 1|x) = p(x) = \left(\frac{e^{\beta_0 + \sum \beta_i x_{ij}}}{1 + e^{\beta_0 + \sum \beta_i x_{ij}}} \right) = \left(\frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_{ij})}} \right)$$

Jackknife: estudar...

1. Variáveis, medida e banco de dados
2. Correlação e plots
3. Regressão
4. Interpretação

Thanks to!

Prof. Cristiano Fernandes (Elétrica – PUC-Rio)