

# MAE0350 - Modelos de Regressão I

## Aula 03

Silvia Nagib Elian

Universidade de São Paulo - Instituto de Matemática e Estatística



**IME**

2023

# SUMÁRIO



Coeficiente de determinação



Distribuição normal bivariada



Referências

## O COEFICIENTE DE CORRELAÇÃO POPULACIONAL

👑 Seja dada uma amostra de vetores aleatórios bivariados denotados por  $\{(X_i, Y_i)\}_{i \in \{1, 2, \dots, n\}}$ , o coeficiente de correlação populacional entre  $X$  e  $Y$  é dado por  $\rho$ . Supondo então que o vetor  $(X, Y)$  tem distribuição **normal** bivariada, é proposto o seguinte teste

$$\begin{cases} \mathcal{H}_0 : \rho = 0, \\ \mathcal{H}_a : \rho \neq 0. \end{cases}$$

👑 Rejeita-se  $\mathcal{H}_0$  a nível de significância  $\alpha \in (0, 1)$  se  $|t_{\text{obs}}| \geq t_c$ , onde  $t_c$  é definido por  $\mathbb{P}(t_{n-2} \geq t_c) = \alpha/2$  e

$$t_{\text{obs}} = \frac{r(x, y) \sqrt{n-2}}{\sqrt{1 - [r(x, y)]^2}}.$$

EQUIVALÊNCIA ENTRE  $t_{\text{OBS}}$  E  $F_{\text{OBS}}$ 

👑 Observa-se o seguinte

$$\begin{aligned} t_{\text{obs}}^2 &= \frac{[r(x, y)]^2(n-2)}{1 - [r(x, y)]^2} = \frac{\frac{\text{SSR}}{\text{SST}}(n-2)}{1 - \frac{\text{SSR}}{\text{SST}}}, \\ &= \frac{\frac{\text{SSR}}{\text{SST}}(n-2)}{\frac{\text{SSE}}{\text{SST}}} = \frac{\text{SSR}}{\frac{\text{SSE}}{n-2}} = \frac{\text{MSR}}{\text{MSE}} = F_{\text{obs}}. \end{aligned}$$

👑 Uma vez que  $t_{\text{obs}}^2 = F_{\text{obs}}$ , o teste  $\mathcal{H}_0 : \rho = 0$  contra  $\mathcal{H}_a : \rho \neq 0$  é equivalente ao teste  $\mathcal{H}_0 : \beta_1 = 0$  contra  $\mathcal{H}_a : \beta_1 \neq 0$ , no sentido de rejeitarmos  $\mathcal{H}_0 : \rho = 0$  se e somente se rejeitamos  $\mathcal{H}_0 : \beta_1 = 0$ .

👑 Vale ressaltar que a interpretação da conclusão de ambos testes é diferente, assim como as situações em que são aplicáveis.

## VISUALIZAÇÃO

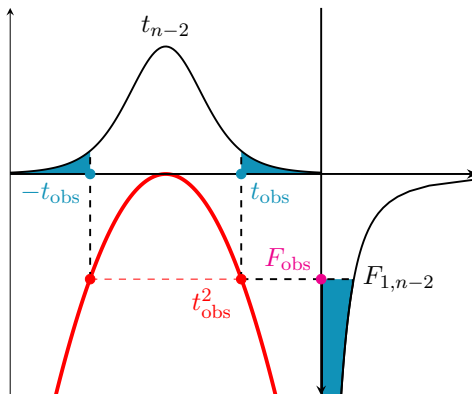


Figura 1: Visualização da equivalência entre a distribuição das estatísticas  $t_{obs}$  e  $F_{obs}$ . Nota-se que as regiões preenchidas em azul tem mesma área.

## FUNÇÃO DE DENSIDADE

👑 Dizemos que o vetor aleatório  $(X, Y)$  tem distribuição normal bivariada se sua densidade é escrita como

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right]\right\},$$

com  $\mu_X \in \mathbb{R}$ ,  $\mu_Y \in \mathbb{R}$ ,  $\sigma_X > 0$ ,  $\sigma_Y > 0$  e  $\rho \in (-1, 1)$ .

👑 Onde  $\rho = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X\sigma_Y}$ , e temos que se  $\rho = 0$ ,  $X$  e  $Y$  são independentes.

## VISUALIZAÇÃO

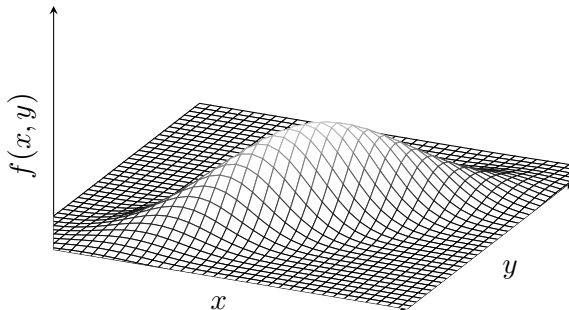


Figura 2: Visualização da função de densidade de probabilidade de um vetor normal bivariado.

## DISTRIBUIÇÃO CONDICIONAL

👑 Se  $(X, Y)$  tem distribuição normal bivariada, a distribuição condicional de  $Y$  dado  $X = x$  é normal com média  $\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$  e variância  $(1 - \rho^2)\sigma_Y^2$ .

👑 Com isto,  $\mathbb{E}[Y|X = x] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}x - \rho \frac{\sigma_Y}{\sigma_X}\mu_X$  e, na distribuição normal bivariada a regressão de  $Y$  em  $X$  é linear da forma  $\beta_0 + \beta_1 x$ , onde

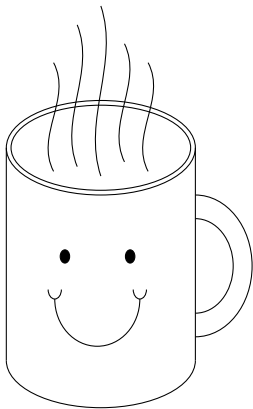
$$\mathbb{E}[Y|X = x] = \underbrace{\mu_Y - \rho \frac{\sigma_Y}{\sigma_X}\mu_X}_{\beta_0} + \underbrace{\rho \frac{\sigma_Y}{\sigma_X}}_{\beta_1} x.$$

👑 Podemos notar que  $\beta_1 = 0 \Leftrightarrow \rho = 0$ . Neste caso,  $\mathbb{E}[Y|X = x] = \mu_Y = \beta_0$ . Observa-se então que é bastante natural que, uma vez aceita a hipótese de  $\mathcal{H}_0 : \beta_1 = 0$ , seja adotado o modelo  $\hat{y} = \bar{y}$ .



## REFERÊNCIAS

 Elian, Silvia Nagib (1988). *Análise de regressão*.



*Bom dia!*