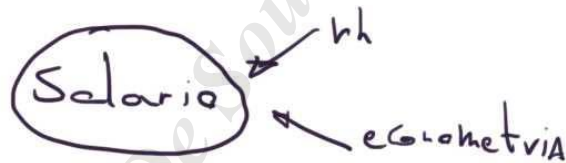PRINTS REALIZADOS DURANTE A AULA DE 13/08/2024:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

Multicolinearidade:

$$Corr[X_{p_i}, X_{q_i}] \neq 0$$

$$\boxed{Salario} \rightarrow rh$$
$$\leftarrow econometria$$

Cenário 1:

$$\hat{Salario}_i = \alpha + \beta_1 \cdot rh_i + \beta_2 \cdot econometria1_i$$

$$Corr[rh_i, econometria_i] \rightarrow \underline{BAIXA}$$

Cenário 2:

$$\hat{Salario}_i = \alpha + \beta_1 \cdot rh_{2i} + \beta_2 \cdot econometria2_i$$

$$Corr[rh_{2i}, econometria2_i] \rightarrow \underline{Muito\ Alta!}$$

Diagnóstico de Multicolinearidade:

$$\hat{Y}_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

$$X_1 = f(X_2, X_3, \ldots, X_k)$$

$$Z_0 \, R_1^2$$

$$\text{Tolerance} = 1 - R_1^2 \qquad (1 \; a \; 0)$$

$$\text{VIF} = 1/\text{Tolerance} \qquad (1 \; a \; \infty)$$

(Variance Inflation Factor)

VIF ↓ : ausência multicolin.

VIF ↑ : existência multicolin.

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \mu_i$$

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}
=
\begin{bmatrix}
1 & X_{11} & X_{21} & \cdots & X_{K1} \\
1 & X_{12} & X_{22} & \cdots & X_{K2} \\
\vdots & \vdots & \vdots & & \vdots \\
1 & X_{1n} & X_{2n} & & X_{Kn}
\end{bmatrix}
\cdot
\begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}
+
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}
$$

$n \times 1$ → Y $n \times (k+1)$ → X β $(k+1) \times 1$ $n \times 1$ → U

$$Y = X \cdot \beta + U$$

OLS: $\boxed{\beta = (x'x)^{-1} \cdot (x'y)}$

1º CASO: Correlação Baixa:

Exemplo:

$X = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$

$x' = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$

$x'x = \begin{bmatrix} 5 & 10 \\ 10 & 25 \end{bmatrix}$

$\det(x'x) = 25$

$(x'x)^{-1} = \begin{bmatrix} 1 & -0{,}4 \\ -0{,}4 & 0{,}2 \end{bmatrix}$

2º CASO: Correlação Muito Alta:
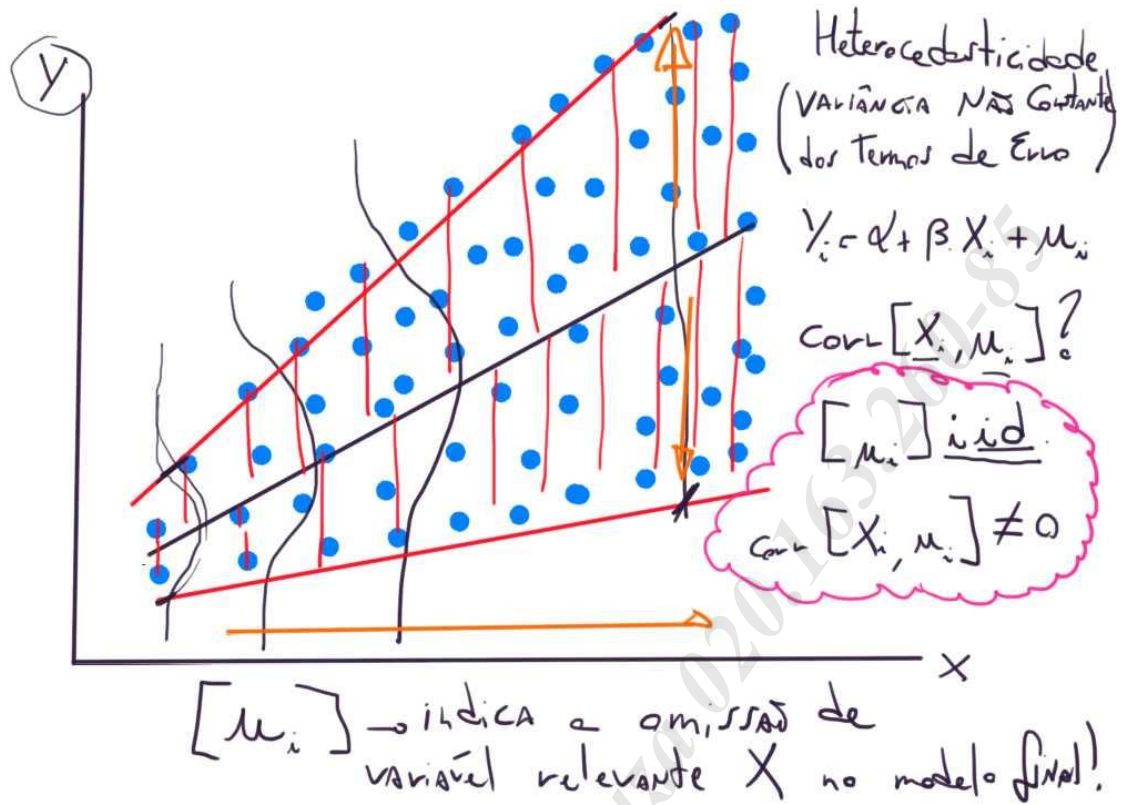
Exemplo:

$X = \begin{bmatrix} 1 & 4 \\ 2 & 7{,}9 \end{bmatrix}$

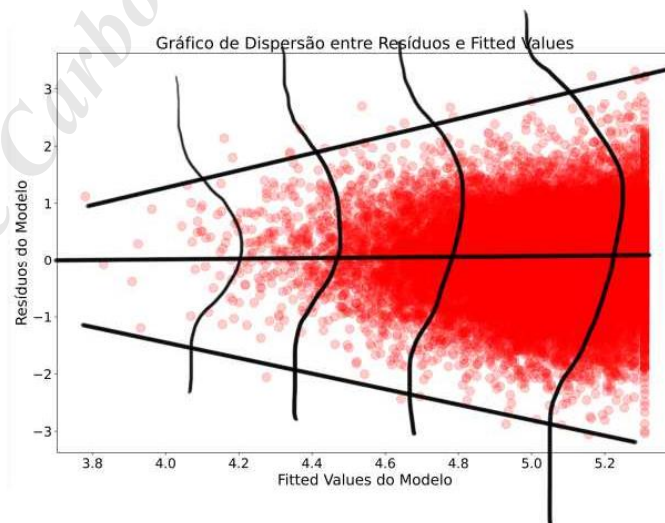$x'x = \begin{bmatrix} 5 & 19{,}8 \\ 19{,}8 & 78{,}41 \end{bmatrix}$

$\det(x'x) = 0{,}01$

$(x'x)^{-1} = \begin{bmatrix} 7841 & -1980 \\ -1980 & 500 \end{bmatrix}$

denominador estatística t

Heterocedasticidade
(VARIÂNCIA NÃO Constante dos Termos de Erro)

$$Y_i = \alpha + \beta X_i + \mu_i$$

$Corr[X_i, \mu_i]$?

$[\mu_i]$ i.i.d.

$Corr[X_i, \mu_i] \neq 0$

$[\mu_i] \to$ indica a omissão de variável relevante X no modelo final!

Saeb $\to$ rendimento

Gráfico de Dispersão entre Resíduos e Fitted Values

Diagnóstico de Heterocedasticidade:

Breusch - Pagan (1980)
The Review of Economic Studies.

$$\mu_p = \frac{\mu^2}{\left(\sum\limits_{i=1}^{n} u^2\right)/n} \Rightarrow \mu_{pi} = \alpha + \beta \cdot \hat{Y}_i + \varepsilon_i$$

$$\Big\} \text{ANOVA}$$

$$\frac{SQReg}{2} \sim \chi^2 \ 1df$$

$$\begin{cases} p\text{-value} \geqslant 5\% \ (H_0) \rightarrow \text{Homocedasticidade} \\ \rightarrow \ p\text{-value} < 5\% \ (H_1) \rightarrow \text{Heterocedasticidade} \end{cases}$$

idade

despmed ← d creh

plano ↗ ↑ relda
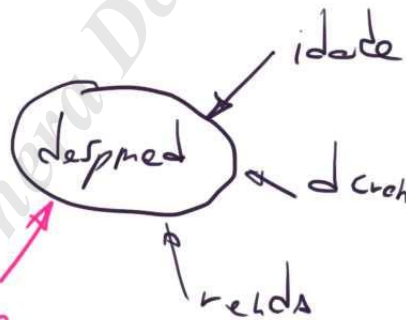
```
================================================================
               coef    std err      t      P>|t|   [0.025   0.975]
----------------------------------------------------------------
Intercept    α 3.5983    0.017   211.642   0.000    3.565    3.632
Q('dcron')   β₁ 0.0271   0.004     7.582   0.000    0.020    0.034
Q('plano_esmeralda') β₂ -0.0903  0.011  -8.137  0.000  -0.112  -0.068
Q('plano_ouro') β₃ -0.1910  0.014  -13.437  0.000  -0.219  -0.163
```
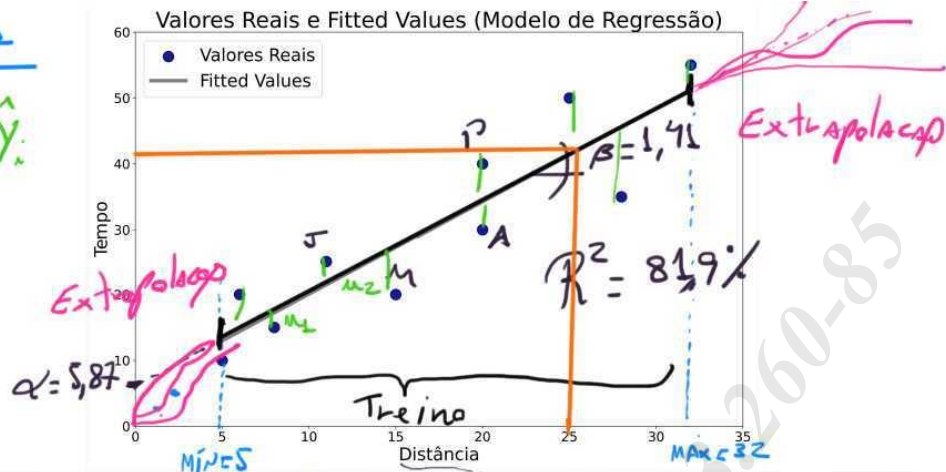
$$\underset{-0,14}{\widehat{despmed}_i} = \underset{\alpha}{3,598} + \underset{\beta_1}{0,027} \cdot dcron_i$$
$$- \underset{\beta_2}{0,090} \cdot Esmeralda_i - \underset{\beta_3}{0,191} \cdot Ouro_i$$

$$\xrightarrow{-0,14} -1$$

Resumo 1

$$\mu_i = Y_i - \hat{Y}_i$$

Valores Reais e Fitted Values (Modelo de Regressão)

$\alpha = 5,87$ $\beta = 1,4$ $R^2 = 81,9\%$ Extrapolação   Treino   MÍN = 5   MAX = 32

① CORRELAÇÃO NÃO IMPLICA CAUSALIDADE.

② PREDIÇÃO NA INTERPOLAÇÃO

③ PARÂMETROS α, β

④ R^2

⑤ OLS ou MQO:
$$\sum_{i=1}^{n} \mu_i = 0$$
$$\sum_{i=1}^{n} \mu_i^2 = \text{mín}$$

⑥ teste F

⑦ teste t

⑧ α sempre fica!

⑨ IC's.

⑩ $(n-1)$ dummies ONE HOT encoding

CONTINUA

Resumo 2:

⑪ teste de Shapiro-Francia $\left(\begin{array}{c}\text{Pressuposto da} \\ \text{Normalidade dos} \\ \text{Resíduos}\end{array}\right)$

⑫ transformação de Box-Cox (Normalização)
$$y^* = \frac{y^\lambda - 1}{\lambda}$$

⑬ Procedimento Stepwise $\left\{\begin{array}{l}\text{- Não ser significante p/ explicar } y. \\ \text{- Multicolinearidade.} \\ \text{- Forma Funcional não linear.}\end{array}\right.$

⑭ Diagnóstico de Multicolinearidade. (VIF e Tolerância).

⑮ Diagnóstico de Heterocedasticidade (teste de Breush-Pagan)