- **Última aula:** modelo RLS

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Amostra: $\{(x_i, y_i),\ i = 1, \dots, n\}$

Estimadores dos mínimos quadrados:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x} \quad ; \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i Y_i - n \overline{x}\,\overline{Y}}{\sum_{i=1}^{n} x_i^2 - n \overline{x}^2}$$

$\hat{Y}_i = \widehat{E(Y|x_i)} = \hat{\beta}_0 + \hat{\beta}_1 x_i$ = estimador de mínimos quadrados da reta de regressão.

Coeficiente de determinação R^2

$$\underbrace{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}_{SST} = \underbrace{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}_{SSE} + \underbrace{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2}_{SSR}$$

$$R^2 = \frac{SSR}{SST} = \frac{\left(\sum_{i=1}^{n} x_i Y_i - n \overline{x}\,\overline{Y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n \overline{x}^2\right)\left(\sum_{i=1}^{n} Y_i^2 - n \overline{Y}^2\right)}$$

- **Estimação de** $Var(\varepsilon_i) = \sigma^2$

$$\boxed{\hat{\sigma}^2} = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$\boxed{= \frac{1}{n-2} \left[\left(\sum_{i=1}^{n} Y_i^2 - n \overline{Y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^{n} x_i^2 - n \overline{x}^2 \right) \right]}$$

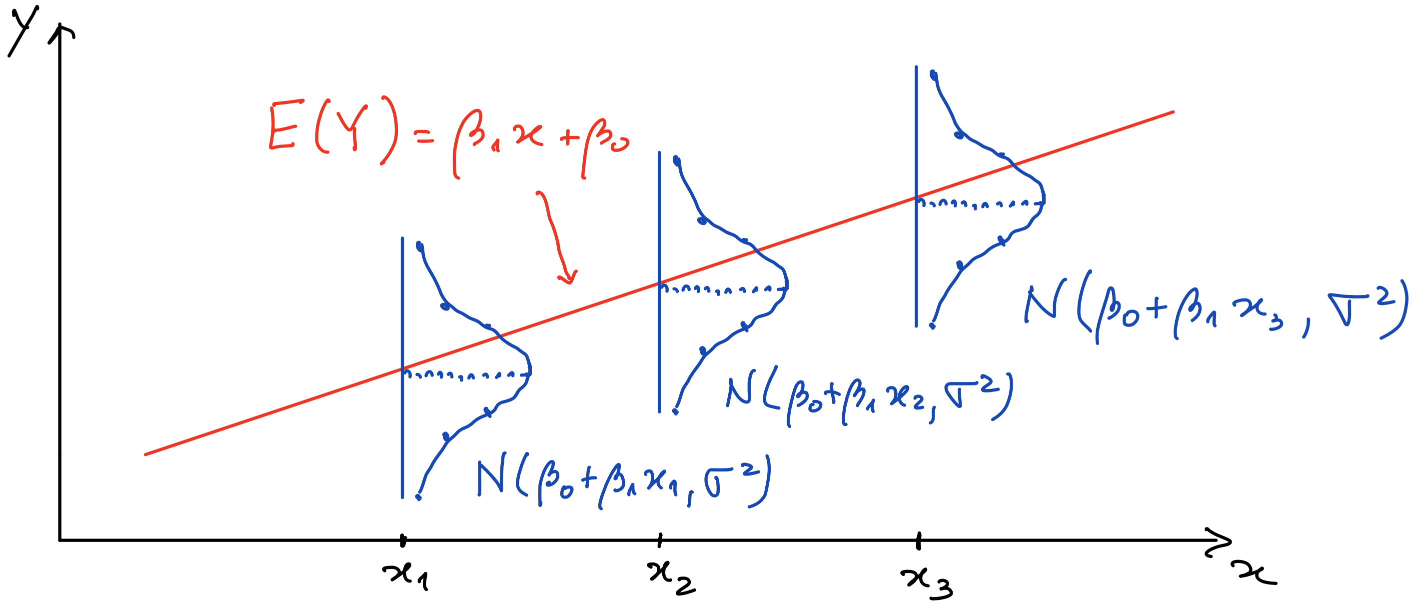
Notas:

- $\hat{\sigma}^2$ é também um estimador da variância de Y_i.
- $\hat{\sigma}$ = erro padrão dos resíduos.

9.2 Inferência sobre os parâmetros RLS

→ intervalos de confiança, testes de hipóteses

Para isso vamos assumir que os erros têm distribuição normal, i.e. $\varepsilon_i \underset{i.i.d.}{\sim} N(0, \sigma^2)$.



É então possível mostrar que:

$$\boxed{\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\right)\hat{\sigma}^2}} \sim t_{(n-2)}}$$

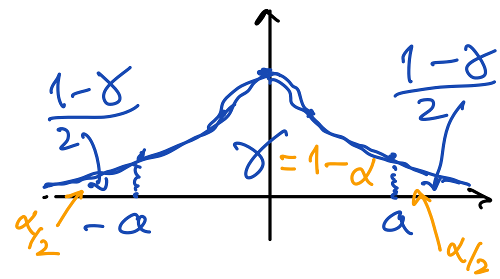
$$\boxed{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}} \sim t_{(n-2)}}$$

Isto permite calcular intervalos de confiança e realizar testes de hipóteses relativos a β_0 e β_1.

Notação no contexto desta aula: $q_x := \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

• IC a $(\gamma = 1-\alpha)\times 100\%$ para β_0

V.a. fulcral: $$T = \frac{\hat\beta_0 - \beta_0}{\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{q_x}\right)\hat\sigma^2}} \sim t_{(n-2)}$$



$$P(-a \le T \le a) = \gamma$$

$$\Leftrightarrow \quad a = F^{-1}_{t_{(n-2)}}\left(\gamma + \frac{1-\gamma}{2}\right) = F^{-1}_{t_{(n-2)}}\left(\frac{1+\gamma}{2}\right)$$
$$= F^{-1}_{t_{(n-2)}}\left(1 - \alpha/2\right) \quad \text{e}$$

$$-a \le T \le a \Leftrightarrow \beta_0 \in \left[\hat\beta_0 - a\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{q_x}\right)\hat\sigma^2},\ \hat\beta_0 + a\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{q_x}\right)\hat\sigma^2}\right]$$
$$= IC_{(1-\alpha)\times 100\%}(\beta_0)$$

• Teste de hipóteses relativo a β_0

$$H_0: \beta_0 = \beta_{0,0} \quad \text{vs} \quad H_1: \beta_0 >, <, \neq \beta_{0,0}$$

Estatística de teste: $$T_{H_0} = \frac{\hat\beta_0 - \beta_{0,0}}{\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{q_x}\right)\hat\sigma^2}} \underset{\text{sob } H_0}{\sim} t_{(n-2)}$$

Região crítica ao nível de significância α e valor-p (= menor nível de significância p/ rejeitar H_0)

- $H_1: \beta_0 \neq \beta_{0,0}$, $RC_\alpha: |T_{H_0}| > c$ com $c = F^{-1}_{t_{(n-2)}}(1-\alpha/2)$

  Valor-p $= 2\min\{P(T_{H_0} < t_{obs}),\ P(T_{H_0} > t_{obs})\}$

[$\beta_{0,0} \notin IC_{1-\alpha}(\beta_0) \Rightarrow$ ~~não~~ rejeitar H_0]

- $H_1: \beta_0 > \beta_{0,0}$, $RC_\alpha: T_{H_0} > c$ com $c = F^{-1}_{t_{(n-2)}}(1-\alpha)$

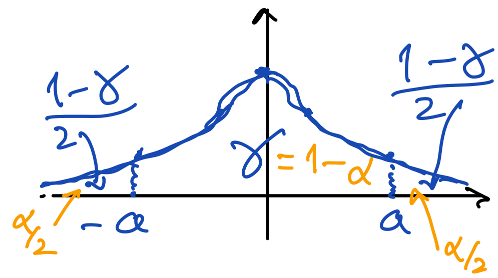
  valor-p $= P(T_{H_0} > t_{obs})$

- $H_1: \beta_0 < \beta_{0,0}$, $RC_\alpha: T_{H_0} < c$ com $c = F^{-1}_{t_{(n-2)}}(\alpha)$

  valor-p $= P(T_{H_0} < t_{obs})$

- IC a $(\gamma = 1-\alpha)\times 100\%$ para β_1

V.a. fulcral: $T = \dfrac{\hat\beta_1 - \beta_1}{\sqrt{\dfrac{\hat\sigma^2}{q_x}}} \sim t_{(n-2)}$



$P(-a \leq T \leq a) = \gamma$

$\Leftrightarrow \; a = F^{-1}_{t_{(n-2)}}\left(\gamma + \frac{1-\gamma}{2}\right) = F^{-1}_{t_{(n-2)}}\left(\frac{1+\gamma}{2}\right)$

$= F^{-1}_{t_{(n-2)}}(1-\alpha/2)$ e

$-a \leq T \leq a \Leftrightarrow \beta_1 \in \left[\hat\beta_1 - a\sqrt{\dfrac{\hat\sigma^2}{q_x}}, \; \hat\beta_1 + a\sqrt{\dfrac{\hat\sigma^2}{q_x}}\right]$

$= IC_{(1-\alpha)\times 100\%}(\beta_1)$

- Teste de hipóteses relativo a β_1

$H_0: \beta_1 = \beta_{1,0}$ vs $H_1: \beta_1 >, <, \neq \beta_{1,0}$

Estatística de teste: $T_{H_0} = \dfrac{\hat\beta_1 - \beta_{1,0}}{\sqrt{\dfrac{\hat\sigma^2}{q_x}}} \underset{\text{sob } H_0}{\sim} t_{(n-2)}$

Região crítica ao nível de significância α e valor-p (= menor nível de significância p/ rejeitar H_0)

- $H_1: \beta_1 \neq \beta_{1,0}$, $RC_\alpha: |T_{H_0}| > c$ com $c = F^{-1}_{t_{(n-2)}}(1-\alpha/2)$

  $\text{Valor-p} = 2 \min \{ P(T_{H_0} < t_{obs}), P(T_{H_0} > t_{obs}) \}$

[$\beta_{1,0} \notin IC_{1-\alpha}(\beta_1) \Rightarrow$ ~~não~~ rejeitar H_0]

- $H_1: \beta_1 > \beta_{1,0}$, $RC_\alpha: T_{H_0} > c$ com $c = F^{-1}_{t_{(n-2)}}(1-\alpha)$

  $\text{valor-p} = P(T_{H_0} > t_{obs})$

- $H_1: \beta_1 < \beta_{1,0}$, $RC_\alpha: T_{H_0} < c$ com $c = F^{-1}_{t_{(n-2)}}(\alpha)$

  $\text{valor-p} = P(T_{H_0} < t_{obs})$

Nota: um caso importante é o chamado teste à significância da regressão:

$$H_0: \beta_1 = 0 \quad \text{vs} \quad H_1: \beta_1 \neq 0$$

Não rejeitar H_0 significa haver evidência para a não existência de uma associação linear entre x e Y.

- Inferências relativas a $E[Y|x_0] = \beta_0 + \beta_1 x_0$

Feitas de forma análoga às anteriores, com base no seguinte resultado

$$\boxed{\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\right)\hat{\sigma}^2}} \sim t_{(n-2)}}$$

- <u>Exercício 9.3</u> $\quad n=10$

$$\sum_{i=1}^{10} x_i = 19.4, \quad \sum_{i=1}^{10} x_i^2 = 38.06, \quad \sum_{i=1}^{10} y_i = 14.8,$$

$$\sum_{i=1}^{10} y_i^2 = 22.76, \quad \sum_{i=1}^{10} x_i y_i = 28.12$$

$$IC_{95\%}(\beta_1) = ?$$

Assumindo como hipótese de trabalho que

$$\varepsilon_i \underset{i.i.d.}{\sim} N(0,\sigma^2), \quad \forall i=1,\dots,10$$

vimos que

$$IC_{95\%}(\beta_1) = \left[\hat{\beta}_1 - a\sqrt{\frac{\hat{\sigma}^2}{q_x}}, \hat{\beta}_1 + a\sqrt{\frac{\hat{\sigma}^2}{q_x}}\right]$$

com $a = F_{t_8}^{-1}\left(1-\frac{0.05}{2}\right) = F_{t_8}^{-1}(0.975) \underset{\text{tabela}}{=} 2.306$ //

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{10} x_i y_i - 10\bar{x}\bar{y}}{\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2} = \frac{28.12 - 10\left(\frac{19.4}{10}\right)\left(\frac{14.8}{10}\right)}{38.06 - 10\left(\frac{19.4}{10}\right)^2} =$$

$$= \frac{-0.592}{0.424} = -1,396$$

$q_x =$ 0.424

$$\hat{\sigma}^2 = \frac{1}{10-2}\left[\left(\sum_{i=1}^{10} y_i^2 - 10\bar{y}^2\right) - \hat{\beta}_1^2 \underbrace{\left(\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2\right)}_{q_x}\right]$$

$$= \frac{1}{8}\left[\left(22.76 - 10\left(\frac{14.8}{10}\right)^2\right) - (1.396)^2 \times 0.424\right]$$

$$= \frac{1}{8}\left[0.856 - 0.826\right] = 0.00375$$

$$\Rightarrow IC_{95\%}(\beta_1) = \left[-1.396 - 2.306\sqrt{\frac{0.00375}{0.424}}\ ,\ -1.396 + 2.306\sqrt{\frac{0.00375}{0.424}}\right]$$

$$= [-1.396 - 0.217\ ,\ -1.396 + 0.217]$$

$$= [-1.613, -1.179] \ //$$

Nota: $R^2 = \dfrac{(-0.592)^2}{0.424 \times 0.856} \simeq 0.966$ (muito alto)

$\Rightarrow$ modelo RLS é bastante ajustado à situação.