MODELAGEM E INFERÊNCIA ESTATÍSTICA

Teste de hipótese e Previsão da resposta

O QUE VOU ESTUDAR HOJE?

Lembrar o conceito e aplicação do teste e hipótese

- **Aplicar no teste de utilidade o modelo β_0 e β_1.**

Inferências sobre $\mu_{y,\, x^*}$

Previsão de valores da variável Y.

TESTE HIPÓTESE

Teste de utilidade do modelo, isto é, confirmar ou rejeitar sua utilidade.

Hipótese nula: H_0: $\beta_1 = \beta_{10}$

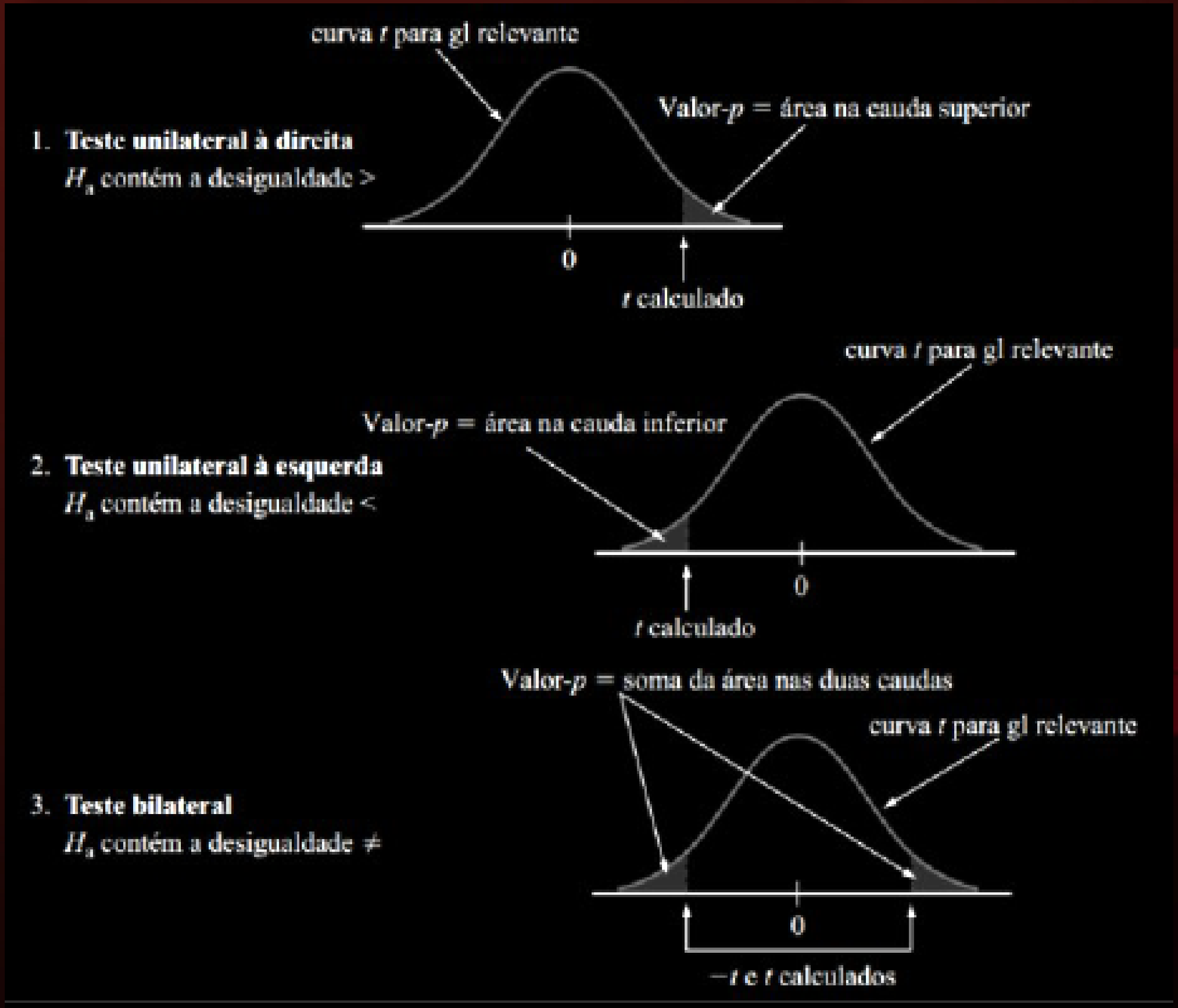
Valor da estatística do teste: $t = \dfrac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$

Hipótese alternativa	Determinação do valor-*p*
H_a: $\beta_1 > \beta_{10}$	Área sob a curva t_{n-2} à direita de t
H_a: $\beta_1 < \beta_{10}$	Área sob a curva t_{n-2} à esquerda de t
H_a: $\beta_1 \neq \beta_{10}$	$2 \cdot$ (Área sob a curva t_{n-2} à direita de $\|t\|$)

O **teste de utilidade do modelo** é o teste de H_0: $\beta_1 = 0$ *versus* H_a: $\beta_1 \neq 0$, sendo o valor da estatística de teste a **razão** $t = \hat{\beta}_1 / s_{\hat{\beta}_1}$.

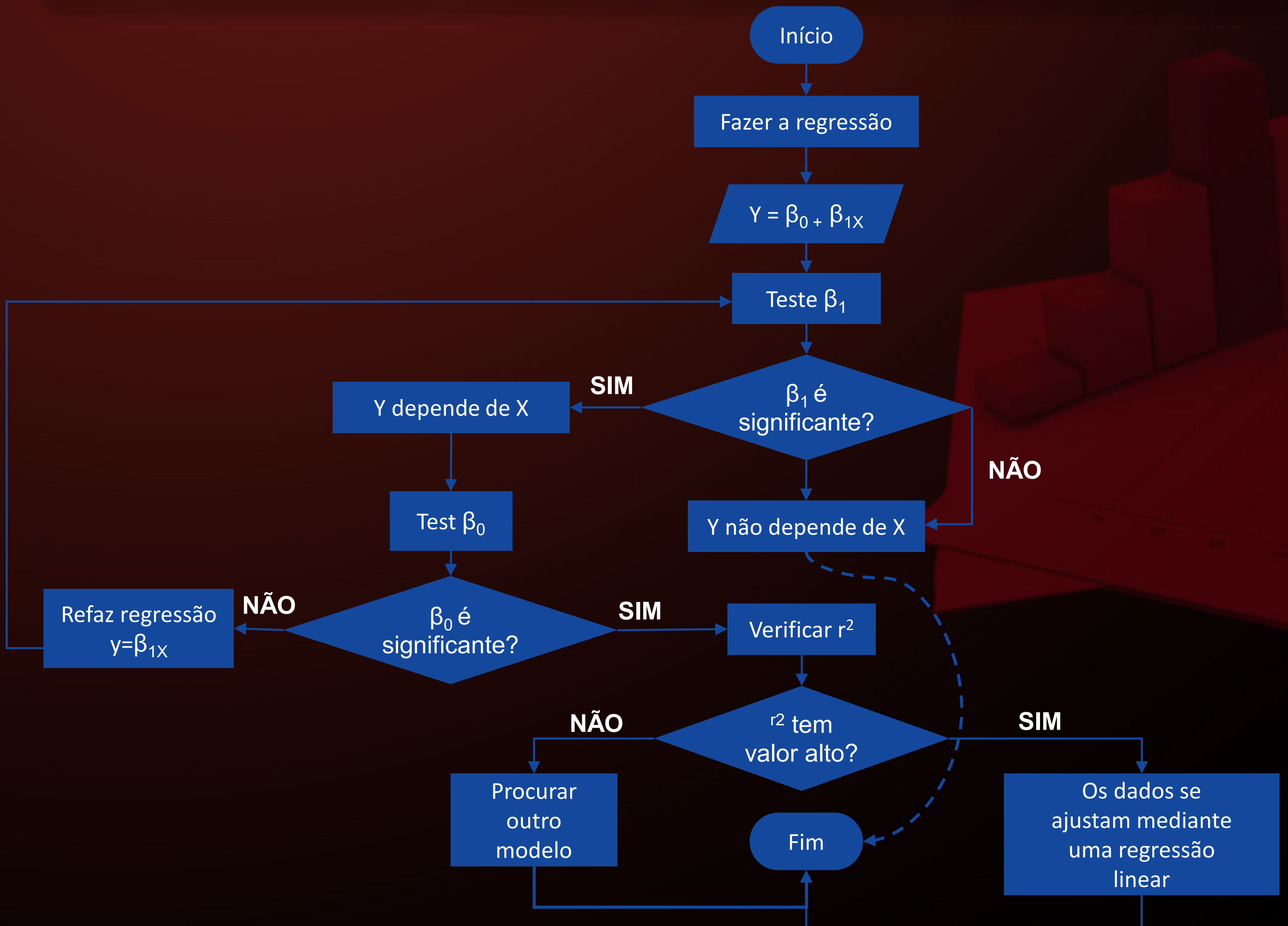
Fonte: (DEVORE, 2018, p. 482)

TESTE HIPÓTESE



Fonte: (DEVORE, 2018, p. 315)

POR QUE REALIZAR O TESTE DE HIPÓTESE?



PROCEDIMENTOS PARA O TESTE DE HIPÓTESE

1. Obter a reta de regressão $y = \beta_o + \beta_1 x$
2. Definir o valor de n (número de amostras).
3. Definir o valor de k (número de variáveis).
4. Calcular os graus de liberdade gl = n-k.
5. Se a intenção é testar a utilidade do modelo iniciar testando $H_0\colon \beta_1 = 0$
   frente a $H_a\colon \beta_1 \neq 0$
   depois testar $H_0\colon \beta_0 = 0$ frente a $H_a\colon \beta_0 \neq 0$

PROCEDIMENTOS PARA O TESTE DE HIPÓTESE β_1

1. Definir a hipótese nula $H_0: \beta_{10} = 0$ frente a $H_a: \beta_{10} \neq 0$.
2. Definir o intervalo de confiança α se conhecido.
3. Definir o intervalo crítico (t_{crit}) na tabela t-student para um determinado α e gl.
4. Definir a estatística de teste $t = \frac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$.

   Como $H_0: \beta_{10} = 0 \rightarrow t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$.
5. Se $|t| \geq t_{crit}$ rejeitar H_0 em favor de $H_a: \beta_{10} \neq 0$.
6. Caso contrário comparar se p-valor < α.

PROCEDIMENTOS PARA O TESTE DE HIPÓTESE β_0

1. Definir a hipótese nula $H_0: \beta_{00} = 0$ frente a $H_a: \beta_{00} \neq 0$.
2. Dos dados do problema 1- α= 0,95 → α=0,05 e α/2=0,025
3. Definir o intervalo crítico (t_{crit}) na tabela t-student para um determinado α e gl.
4. Definir a estatística de teste $t = \frac{\hat{\beta}_0 - \beta_{00}}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}}$.

   Como $H_0: \beta_{00} = 0$→ $t = \frac{\hat{\beta}_0}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}}$.
5. Se $|t| \geq t_{crit}$ rejeitar H_0 em favor de $H_a: \beta_{00} \neq 0$.
6. Caso contrário comparar se 2(p-valor) < α se sim, rejeitar H_0 em favor de $H_a: \beta_{00} \neq 0$

INFERÊNCIAS SOBRE μ_{y,x^*}

Seja $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$, em que x^* é algum valor fixo de x. Então

1. O valor médio de $\hat{Y}$ é

$$E(\hat{Y}) = E(\hat{\beta}_0 + \hat{\beta}_1 x^*) = \mu_{\hat{\beta}_0 + \hat{\beta}_1 x^*} = \beta_0 + \beta_1 x^*$$

Assim, $\hat{\beta}_0 + \hat{\beta}_1 x^*$ é um estimador não viciado para $\beta_0 + \beta_1 x^*$ (isto é, para $\mu_{Y \cdot x^*}$).

2. A variância de $\hat{Y}$ é

$$V(\hat{Y}) = \sigma^2_{\hat{Y}} = \sigma^2\left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum x_i^2 - (\sum x_i)^2/n}\right] = \sigma^2\left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right]$$

e o desvio padrão $\sigma_{\hat{Y}}$ é a raiz quadrada dessa expressão. O desvio padrão estimado de $\hat{\beta}_0 + \hat{\beta}_1 x^*$, representado por $s_{\hat{Y}}$ ou $s_{\hat{\beta}_0 + \hat{\beta}_1 x^*}$, resulta da substituição de σ por sua estimativa s:

$$s_{\hat{Y}} = s_{\hat{\beta}_0 + \hat{\beta}_1 x^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

3. $\hat{Y}$ tem uma distribuição normal.

Fonte: (DEVORE, 2018, p. 488)

INFERÊNCIAS SOBRE μ_{y,x^*}

Definir uma variável T para obter a estatística de teste para reta

$$T = \frac{\hat{\beta}_0 + \hat{\beta}_1 x^* - \beta_0 + \beta_1 x^*}{s_{\hat{\beta}_0 + \hat{\beta}_1 x^*}}$$

$$T = \frac{\hat{Y} - \beta_0 + \beta_1 x^*}{s_{\hat{\beta}_0 + \hat{\beta}_1 x^*}}$$

Considerando gl=n-2

Da mesma forma que foi calculado o intervalo de confiança para β0 e β1 produzir um IC de 100(1-α)% para μ_{y,x^*} ,

o valor esperado de Y quando x=x*

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\frac{\alpha}{2},n-2}\, s_{\hat{\beta}_0 + \hat{\beta}_1 x^*} = \hat{y} \pm t_{\frac{\alpha}{2},n-2}\, s_{\hat{y}}$$

INTERVALO DE PREVISÃO PARA UM VALOR FUTURO DE Y

- Intervalo de confiança (IC).
  - ✓ Referido a um parâmetro ou população com valor fixo.
  - ✓ Valor desconhecido para o pesquisador.

- Intervalo de previsão (IP).
  - ✓ Valor futuro de Y.
  - ✓ Variável aleatória.

INTERVALO DE PREVISÃO PARA UM VALOR FUTURO DE Y

$$\begin{aligned} V[Y-(\hat{\beta}_0+\hat{\beta}_1 x^*)] &= \text{variância do erro de previsão} \\ &= V(Y)+V(\hat{\beta}_0+\hat{\beta}_1 x^*) \\ &= \sigma^2+\sigma^2\left[\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2\left[1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right] \end{aligned}$$

$$T=\frac{Y-(\hat{\beta}_0+\hat{\beta}_1 x^*)}{S\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}}$$

Um IP de $100(1-\alpha)\%$ para uma observação Y futura a ser feita quando $x = x^*$ é

$$\begin{aligned} &\hat{\beta}_0+\hat{\beta}_1 x^* \pm t_{\alpha/2,n-2}\cdot s\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}} \\ &= \hat{\beta}_0+\hat{\beta}_1 x^* \pm t_{\alpha/2,n-2}\cdot\sqrt{s^2+s^2_{\hat{\beta}_0+\hat{\beta}_1 x^*}} \qquad (12.7) \\ &= \hat{y} \pm t_{\alpha/2,n-2}\cdot\sqrt{s^2+s^2_{\hat{Y}}} \end{aligned}$$

Fonte: (DEVORE, 2018, p. 491-492)

EXEMPLO 1: TESTE DE HIPÓTESE β_1

Observe na seguinte tabela os dados de 28 carros disponíveis na loja GT Auto, a capacidade volumétrica (cc) como variável preditora e o consumo (km/l) como variável resposta. Calcule o valor da estatística do teste, ao nível de confiança de 95%, e realize o teste de utilidade do modelo, se a reta é:

Consumo de combustível $= 12,14401 - 0,00135$ Capacidade volumétrica

$$y = 12,14401 - 0,00135x$$

E o erro padrão é $s_{\hat{\beta}_1} = \frac{s}{\sqrt{s_{xx}}} = 0,0001177$

Consumo	12	10,4	12,8	10,3	10,5	8,5	9,5
Cap vol.	792	994	1000	1368	1598	1796	1997
Consumo	9,2	6,5	7,1	6,6	4,2	6,4	3
Cap vol.	1999	2996	3197	3498	5461	6162	7291
Consumo	11,4	11,2	11	11	10,6	8,7	10,5
Cap vol.	999	1199	1399	1498	1598	1798	1998
Consumo	7,8	7	5	6,1	2,6	3	3,8
Cap vol.	2995	3493	3799	3982	5204	5980	7993

EXEMPLO 1: TESTE DE HIPÓTESE β_1

O teste de utilidade do modelo é verificar $H_0: \beta_{10} = 0$ frente a $H_a: \beta_{10} \neq 0.$
Seguir o passo a passo para n=28, k=2, gl= 28-2=26

1. **Verificar $H_0: \beta_{10} = 0$ frente a $H_a: \beta_{10} \neq 0$ em $y = 12,14401 - 0,00135x$**
2. **Dos dados do problema 1- α= 0,95 → α=0,05 e α/2=0,025**
3. **Da tabela t-student obter $t_{crit} = t_{(\alpha/2,gl)} = t_{(0,025,26)}$ → t_{crit} =2,0555**

Distribuição t de *Student*

gl	Teste Unilateral								
	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%
	Teste Bilateral								
	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%
1	1,9626	3,0777	6,3137	12,7062	15,8945	31,8210	63,6559	318,2888	636,5776
2	1,3862	1,8856	2,9200	4,3027	4,8487	6,9645	9,9250	22,3285	31,5998
3	1,2498	1,6377	2,3534	3,1824	3,4819	4,5407	5,8408	10,2143	12,9244
24	1,0593	1,3178	1,7109	2,0639	2,1715	2,4922	2,7970	3,4668	3,7454
25	1,0584	1,3163	1,7081	2,0595	2,1666	2,4851	2,7874	3,4502	3,7251
26	1,0575	1,3150	1,7056	2,0555	2,1620	2,4786	2,7787	3,4350	3,7067
27	1,0567	1,3137	1,7033	2,0518	2,1578	2,4727	2,7707	3,4210	3,6895

EXEMPLO 1: TESTE DE HIPÓTESE β_1

O teste de utilidade do modelo é verificar $H_0: \beta_{10} = 0$ frente a $H_a: \beta_{10} \neq 0.$
Seguir o passo a passo para n=28, k=2, gl= 28-2=26

1. Verificar $H_0: \beta_{10} = 0$ frente a $H_a: \beta_{10} \neq 0$ em $y = 12,14401 - 0,00135x$
2. Dos dados do problema 1- α= 0,95 → α=0,05 e α/2=0,025
3. Da tabela t-student $t_{crit} = t_{(\alpha/2,gl)} = t_{(0,025,26)}$→$t_{crit} = 2,0555$.
4. Definir a estatística de teste. Como $H_0: \beta_{10} = 0$→ $t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$.

   E o erro padrão é $s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} = 0,0001177$ e $\hat{\beta}_1 = -0,00135$

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{-0,00135}{0,0001177} = -11,469$$

5. Se $|t| \geq t_{crit}$ rejeitar H_0 em favor de $H_a: \beta_{10} \neq 0.$

$$|-11,469\,| \geq 2,0555$$

Não há evidência suficiente de que $H_0: \beta_{10} = 0.$
No nível de significância do 5% β_1 é significantemente $\neq 0.$

EXEMPLO 1 VERIFICANDO RESULTADO NO PYTHON COM O RESULTADO MANUAL

```
modelo = sm.OLS(y, x).fit()
print(modelo.summary())
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         12.1440      0.427     28.435      0.000      11.266      13.022
cap_vol       -0.0013      0.000    -11.398      0.000      -0.002      -0.001
==============================================================================
```

Valor-p

Valor da estatística do teste $t = \frac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$

Com a hipótese nula $H_0: \beta_1 = 0$

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{-0,00135}{0,0001177} = -11,469$$

EXEMPLO 1 VERIFICANDO RESULTADO NO PYTHON COM O RESULTADO MANUAL

O Python mediante o sm. OLS confirma a rejeição apresentando o valor-p = 0,000

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                consumo   R-squared:                       0.833
Model:                            OLS   Adj. R-squared:                  0.827
Method:                 Least Squares   F-statistic:                     129.9
Date:                Thu, 13 Jan 2022   Prob (F-statistic):           1.30e-11
Time:                        04:46:18   Log-Likelihood:                -45.064
No. Observations:                  28   AIC:                             94.13
Df Residuals:                      26   BIC:                             96.79
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         12.1440      0.427     28.435      0.000      11.266      13.022
cap_vol       -0.0013      0.000    -11.398      0.000      -0.002      -0.001
==============================================================================
Omnibus:                        0.268   Durbin-Watson:                   1.555
Prob(Omnibus):                  0.875   Jarque-Bera (JB):                0.200
Skew:                           0.181   Prob(JB):                        0.905
Kurtosis:                       2.798   Cond. No.                     6.50e+03
==============================================================================
```

EXEMPLO 2: TESTE DE HIPÓTESE β_0

Realize o teste de funcionalidade do modelo para intercepto, lembrando que a reta de regressão é $y = 12,14401 - 0,00135x$, s=1,25, $\bar{x} = 3003$ e $S_{XX} = 112735060$.

Após calcule o intervalo de confiança do intercepto da reta.

Intercepto → β_0

Teste de funcionalidade do modelo

$$H_0: \beta_{00} = 0$$
$$H_a: \beta_{00} \neq 0.$$

EXEMPLO 2: TESTE DE HIPÓTESE β_0

Dados n=28, k=2, gl= 28-2,, s=1,25, $\bar{x} = 3003$ e $S_{XX} = 112735060$

1. Verificar $H_0 : \beta_{00} = 0$ frente a $H_a : \beta_{00} \neq 0$ em $y = 12,144010 - 0,00135x$
2. Definir o intervalo de confiança α se conhecido.
3. Do exemplo 1 $t_{crit} = t_{(\alpha/2,gl)} = t_{(0,025,26)}$ → t_{crit} =2,0555
4. Definir a estatística de teste $t = \dfrac{\hat{\beta}_0}{s\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{s_{XX}}}}$.

$$t = \frac{12,14401}{(1,25)\sqrt{\frac{1}{26}+\frac{\overline{(3003)}^2}{112735060}}}$$

$$t = 28,56089$$

EXEMPLO 2: TESTE DE HIPÓTESE β_0

Dados n=28, k=2, gl= 28-2,, s=1,25, $\bar{x} = 3003$ e $S_{XX} = 112735060$

1. Verificar $H_0: \beta_{00} = 0$ frente a $H_a: \beta_{00} \neq 0$ em $y = 12,14401 - 0,00135x$
2. Definir o intervalo de confiança α se conhecido.
3. Do exemplo 1 $t_{crit} = t_{(\alpha/2,gl)} = t_{(0,025,26)}$ → t_{crit} =2,0555
4. Definir a estatística de teste $t = \dfrac{\hat{\beta}_0}{s\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{s_{XX}}}} \rightarrow t = 28,56089$
5. Se $|t| \geq t_{crit}$ rejeitar H_0 em favor de $H_a: \beta_{00} \neq 0.$

$$|28,56089| \geq 2,055$$

$H_0: \beta_{00} = 0$ é rejeitada validando $H_a: \beta_{00} \neq 0$

EXEMPLO 2: TESTE DE HIPÓTESE β_0

Dados n=28, k=2, gl= 28-2,, s=1,25, $\bar{x} = 3003$, $S_{XX} = 112735060$ e

$$y = 12,14401 - 0,00135x$$

Da etapa anterior $s_{\hat{\beta}_0} = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}} = (1,25)\sqrt{\frac{1}{28} + \frac{\overline{(3003)}^2}{112735060}} =$

$s_{\hat{\beta}_0} =$ 0,43022

$$\hat{\beta}_0 = \beta_0 \pm t_{\alpha/2,n-2} s_{\hat{\beta}_0}$$

$$t_{0,025,26} = 2,0555$$

$$\hat{\beta}_0 = 12,14401 \pm (2,055)(0,4252)$$

IC para o nível do 95%

$\hat{\beta}_0$ =(11,27022,13,0178)

EXEMPLO 2 VERIFICANDO RESULTADO NO PYTHON COM O RESULTADO MANUAL

O Python mediante o `sm.OLS` confirma a rejeição de H_0 apresentando o valor-p = 0,000

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                consumo   R-squared:                       0.833
Model:                            OLS   Adj. R-squared:                  0.827
Method:                 Least Squares   F-statistic:                     129.9
Date:                Thu, 13 Jan 2022   Prob (F-statistic):           1.30e-11
Time:                        04:46:18   Log-Likelihood:                -45.064
No. Observations:                  2                                     94.13
Df Residuals:                      2                                     96.79
Df Model:                           
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         12.1440      0.427     28.435      0.000      11.266      13.022
cap_vol       -0.0013      0.000    -11.398      0.000      -0.002      -0.001
==============================================================================
Omnibus:                        0.268   Durbin-Watson:                   1.555
Prob(Omnibus):                  0.875   Jarque-Bera (JB):                0.200
Skew:                           0.181   Prob(JB):                        0.905
Kurtosis:                       2.798   Cond. No.                     6.50e+03
==============================================================================
```

$s_{\hat{\beta}_0} = 0{,}43022$

$$t = \frac{\hat{\beta}_0}{s_{\hat{\beta}_0}} \rightarrow t = 28{,}2278$$

EXEMPLO 3: IC DE UMA AMOSTRA

Considerando os dados da GT Auto, cuja reta de regressão é

$$y = 12,14401 - 0,00135x$$

Calcule o intervalo de confiança para o nível de confiança de 95% para uma capacidade volumétrica média de 3500.

Seguem alguns valores calculados nos exemplos anteriores:

$r^2 = 0,8346$, Σx_i=84084, Σy_i= 226,7, Σx_i^2=365239312, $\Sigma x_i y_i$=528830,6 Σy_i^2=2081,25, $\bar{x} = 3003$, $\bar{y} = 226,7$, $S_{xx} = 112735060$, $S_{xy} = -151950$, $S_{yy} = 245{,}7896$, $s^2 = 1,563$, $s = 1,250$, SQE=40,6571, $SQT = 245{,}7896$ e $t_{(0{,}025{,}26)} = 2{,}0555$.

EXEMPLO 3: IC DE UMA AMOSTRA

Dados necessários $y = 12,14401 - 0,00135$, $x^*=3500$, $n=28$, $\bar{x} = 3003$, $S_{xx} = 112735060$, $t_{(0,025,26)} = 2,0555$ e $s = 1,250$

O intervalo está centralizado em:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^* = \hat{\beta}_0 + \hat{\beta}_1(3500)$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^* = 12,14401 - 0,00135(3500)$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^* = 7,419$$

O desvio padrão estimado é:

$$\boxed{s_{\hat{Y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}}$$

$$s_{\hat{Y}} = (1,250)\sqrt{\frac{1}{28} + \frac{(3500 - 3003)^2}{1127350600}} = 0,24337$$

EXEMPLO 3: IC DE UMA AMOSTRA

Assim $\hat{y} \pm t_{\frac{\alpha}{2},n-2} s_{\hat{y}}$ pode ser construída mediante:

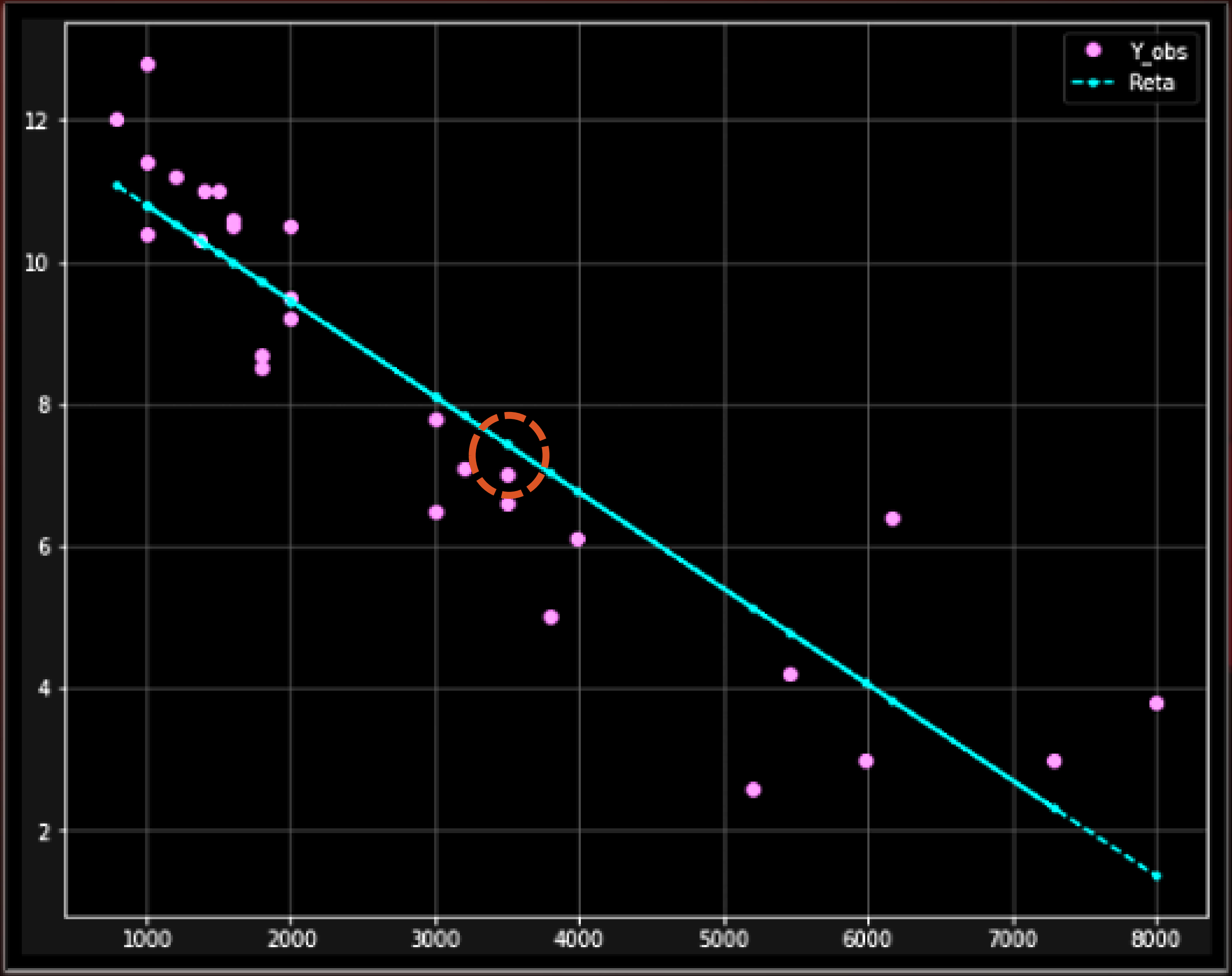
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$ **=7,419**
$s_{\hat{Y}} = 0,24337$ **e**

$t_{(0,025,26)} =$ **2,0555**

O intervalo de confiança será:

$\hat{\mu}_{Y,3500}$**=**$7,419 \pm$ **(2,0555)(0,24337)**
$\hat{\mu}_{Y,3500}$**=(6,9188, 7,9192)**

EXEMPLO 3: IC DE UMA AMOSTRA



$\hat{\mu}_{Y,3500}$=(6,9188, 7,9192)

6,9188<$\hat{\mu}_{Y,3500}$ < 7,9192

EXEMPLO 4: IP

Finalmente calcular o intervalo de previsão para o nível de confiança de 95% para uma capacidade volumétrica média de 3500.
Considerando os dados da GT Auto, cuja reta de regressão é

$$y = 12,14401 - 0,00135x$$

$$\bar{y} = 226,7, t_{(0,025,26)} = 2,0555, s = 1,250 \text{ e } s_{\hat{Y}} = 0,24337$$

A equação do IP

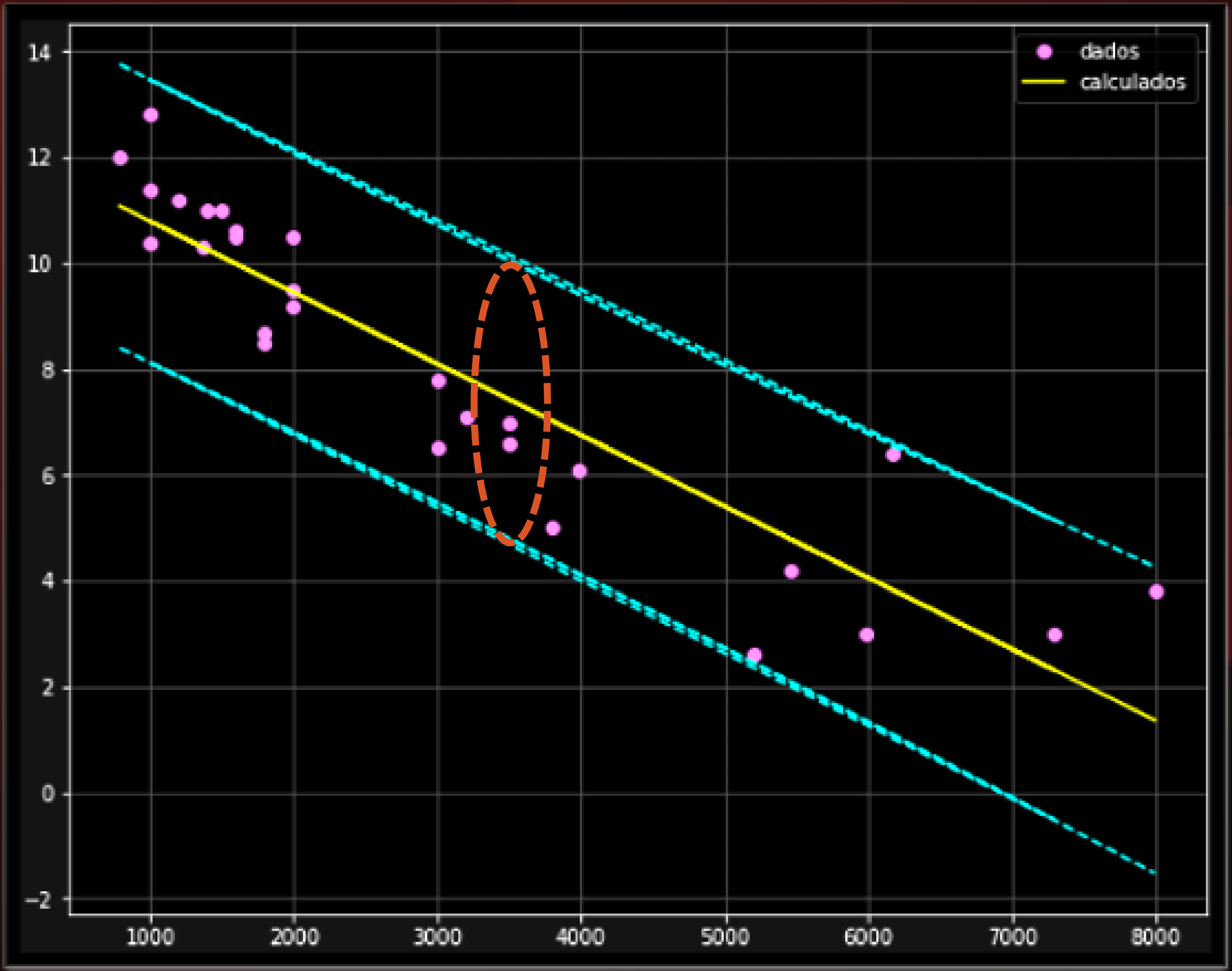
$$\boxed{\hat{y} \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}}$$

$$\hat{y} \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2} = 7,419 \pm (2,0555)\sqrt{1,250^2 - 0,24337^2}$$

IP: $\hat{Y}$=(4,8988, 9,9392)

4,8988 <$\hat{Y}$< 9,9392

EXEMPLO 3



IC

$\hat{\mu}_{Y,3500}$=(6,9188, 7,9192)

6,9188<$\hat{\mu}_{Y,3500}$ < 7,9192

IP

$\hat{Y}$=(4,8988, 9,9392)

4,8988 <$\hat{Y}$< 9,9392