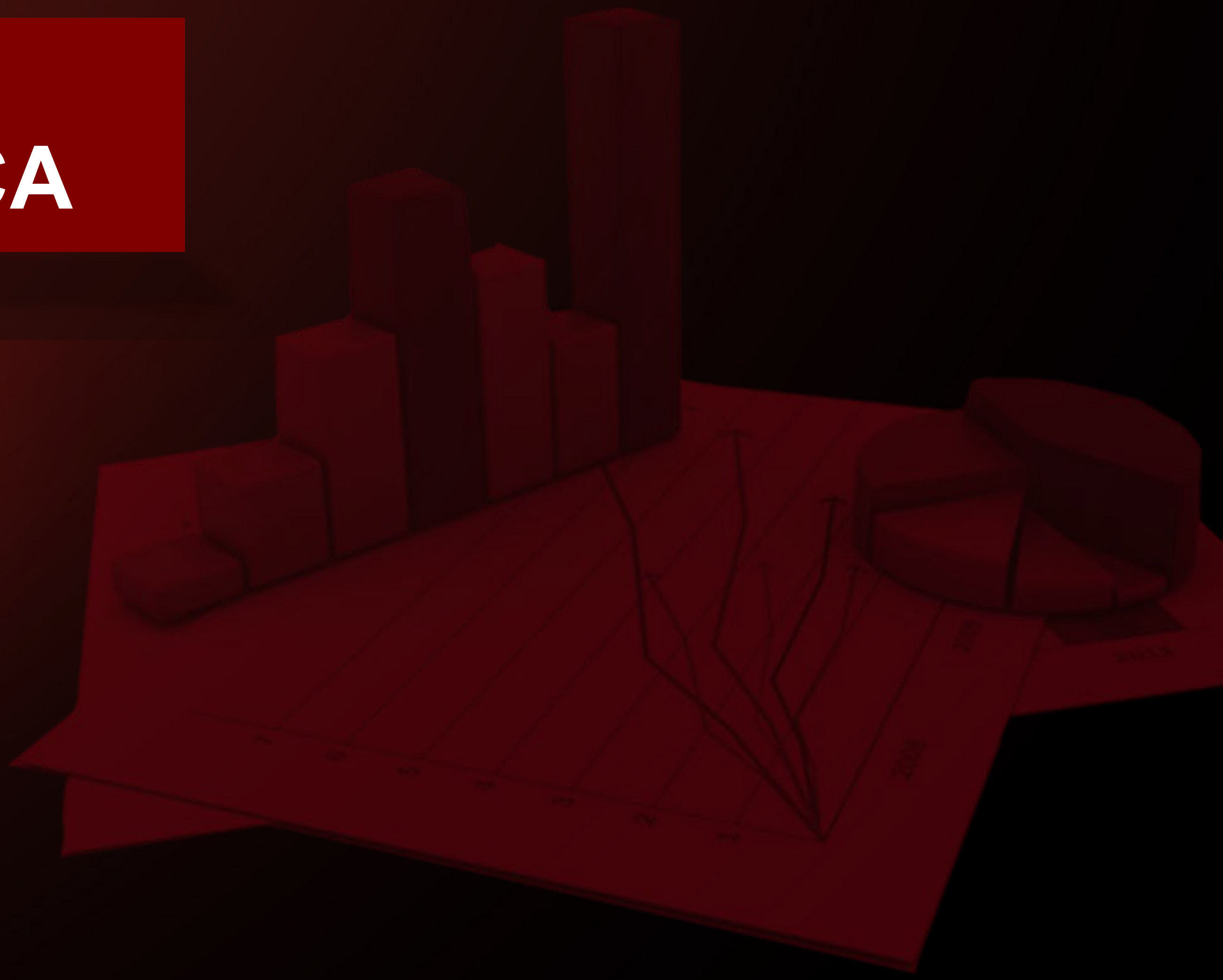
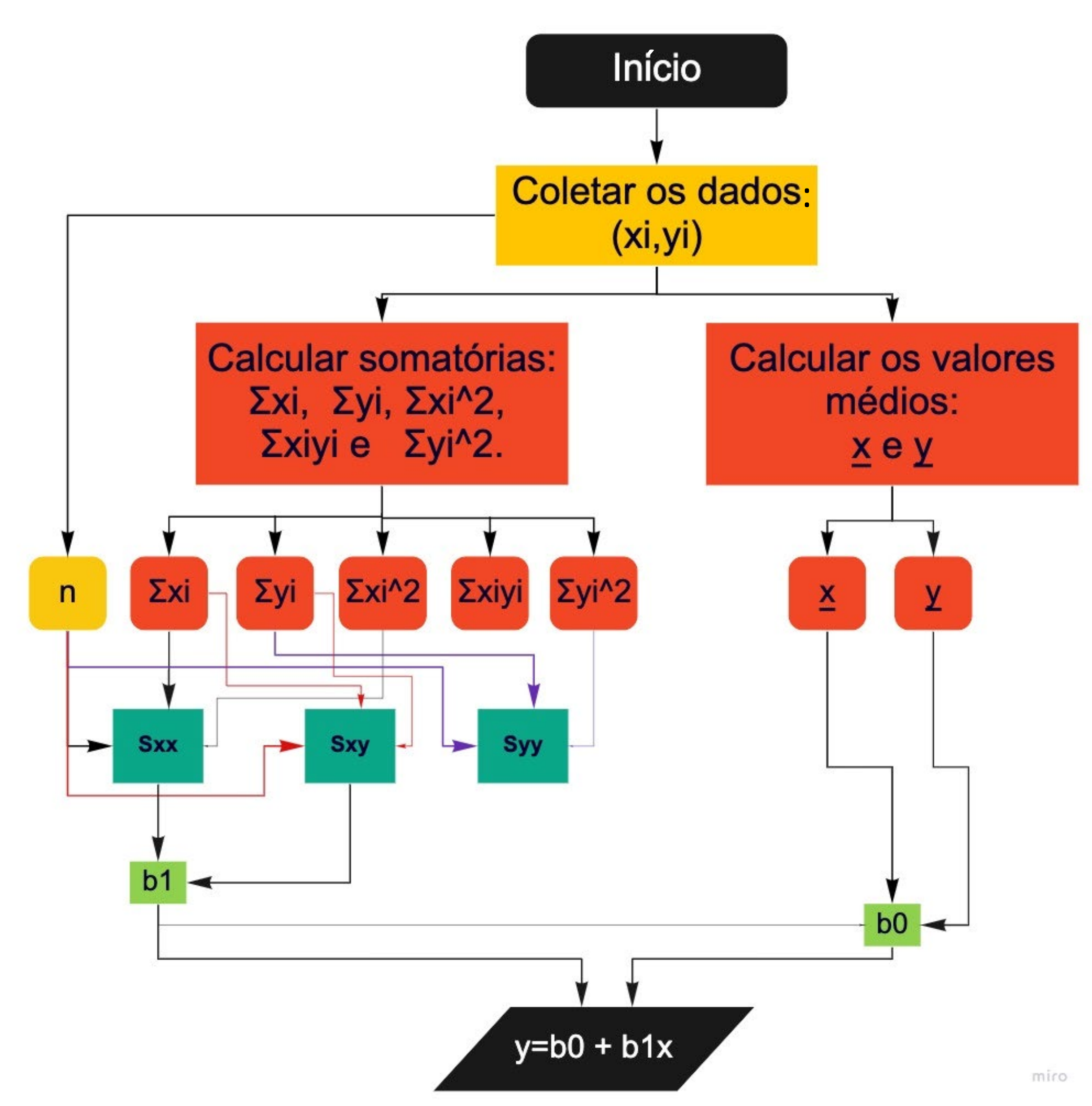
MODELAGEM E INFERÊNCIA ESTATÍSTICA

Análise de ajustes: Exemplo prático



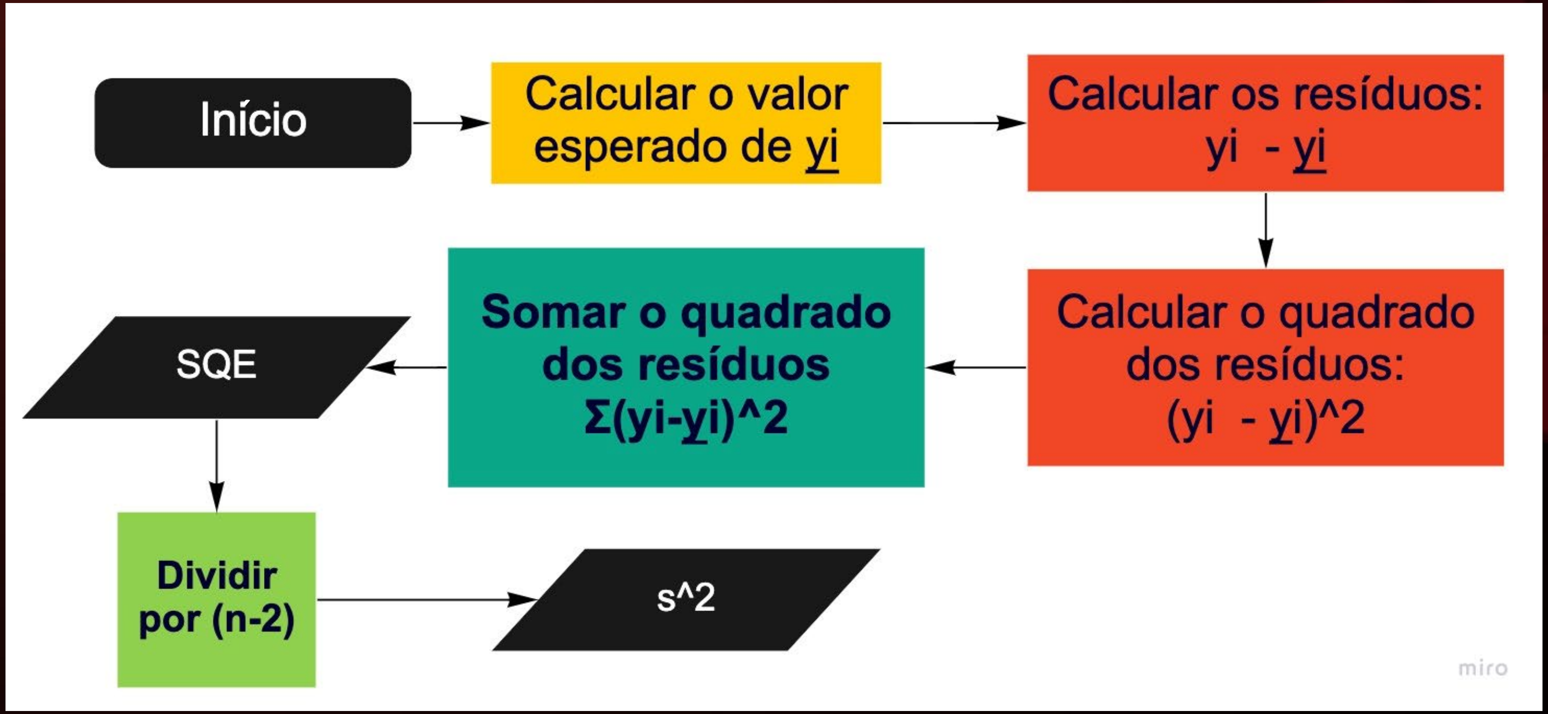
ESTIMAR OS PARÂMETROS



Resumo do processo

ESTIMAR σ^2 e σ

Procedimento



COEFICIENTE DE DETERMINAÇÃO r^2

SQE (Soma de Quadrados dos Erros) → medida da quantidade de variação em y não explicada.

$$\mathrm{SQE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

SQT (Soma de Quadrados Total)

$$\mathrm{SQT} = S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

SQR (Soma de Quadrados da Regressão)

$$SQR = SQT - SQE$$

COEFICIENTE DE DETERMINAÇÃO r^2

Coeficiente de determinação, proporção da variação do *y* observado que pode ser explicada pelo modelo de regressão linear simples.

$$r^2 = 1 - \frac{SQE}{SQT}$$

Valores altos de r^2 → o modelo de regressão linear simples pode ser usado para explicar a variação de *y*.

Valores baixos de r^2 → procurar outro modelo.

EXEMPLO

- Observe na tabela a seguir, os dados de 28 carros disponíveis na loja GT Auto, onde a capacidade volumétrica (cc) pode ser considerada como variável preditora e o consumo (km/l) como variável resposta.

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

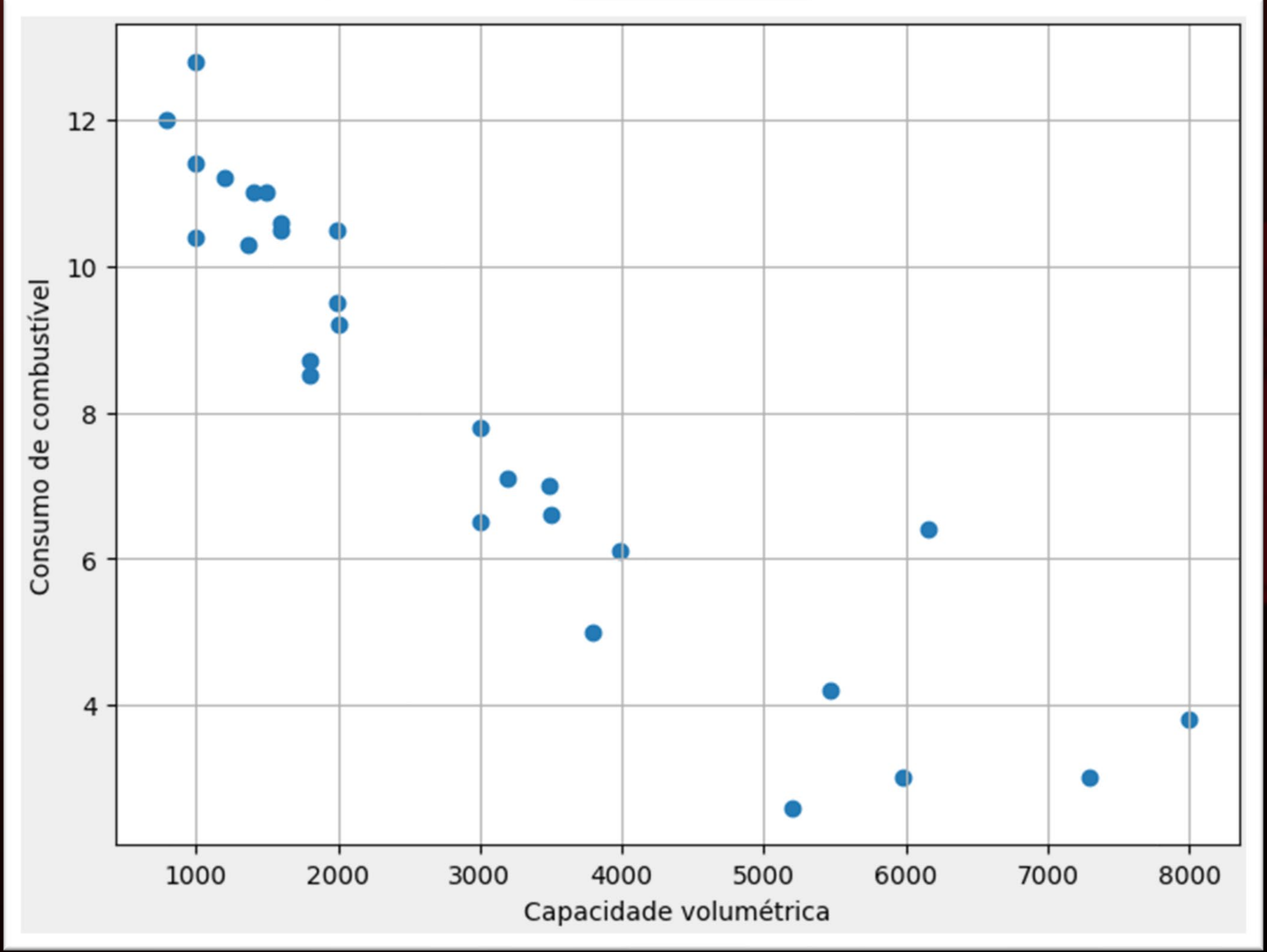
Se os valores Σx_i=84084, Σy_i= 226,7, Σx_i^2=365239312, $\Sigma x_i y_i$=528830,6 e Σy_i^2=2081,25 e os valores médios são $\bar{x} = 840840$ e $\bar{y} = 226{,}7$.

a) Obtenha a reta que descreve a relação entre as variáveis.
b) Qual o valor esperado de Y para x=3000.
c) Calcular SQE e a variância estimada s^2 e o desvio padrão estimado s.
d) Calcular SQT e o coeficiente de determinação r^2.
e) Obter o erro padrão estimado e o IC de β_1 para o nível de confiança de 95%.

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

a) Diagrama de dispersão



EXEMPLO

a) Obter a reta de regressão

Σx_i=84084, Σy_i= 226,7, Σx_i^2=365239312, $\Sigma x_i y_i$=528830,6 e Σy_i^2=2081,25 e os valores médios são $\bar{x} = 3003$ e $\bar{y} = 8{,}0964$.

Calcular as relações Sxx, Sxy, lembrando que n=28:

$$S_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 365239312 - \frac{84084^2}{28}$$

$$S_{xx} = 112735060$$

$$S_{xy} = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} = 528830{,}6 - \frac{(84084)(226{,}7)}{28}$$

$$S_{xy} = -151950$$

EXEMPLO

a) Obter a reta de regressão

Com os valores de Sxx e Sxy, calcular os parâmetros β_0 e β_1.

$$S_{xx} = 112735060 \text{ e } S_{xy} = -151950$$

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{-151950}{112735060}$$

$$\beta_1 = -0{,}00135$$

Com $\bar{x} = 3003$ e $\bar{y} = 8{,}0964$

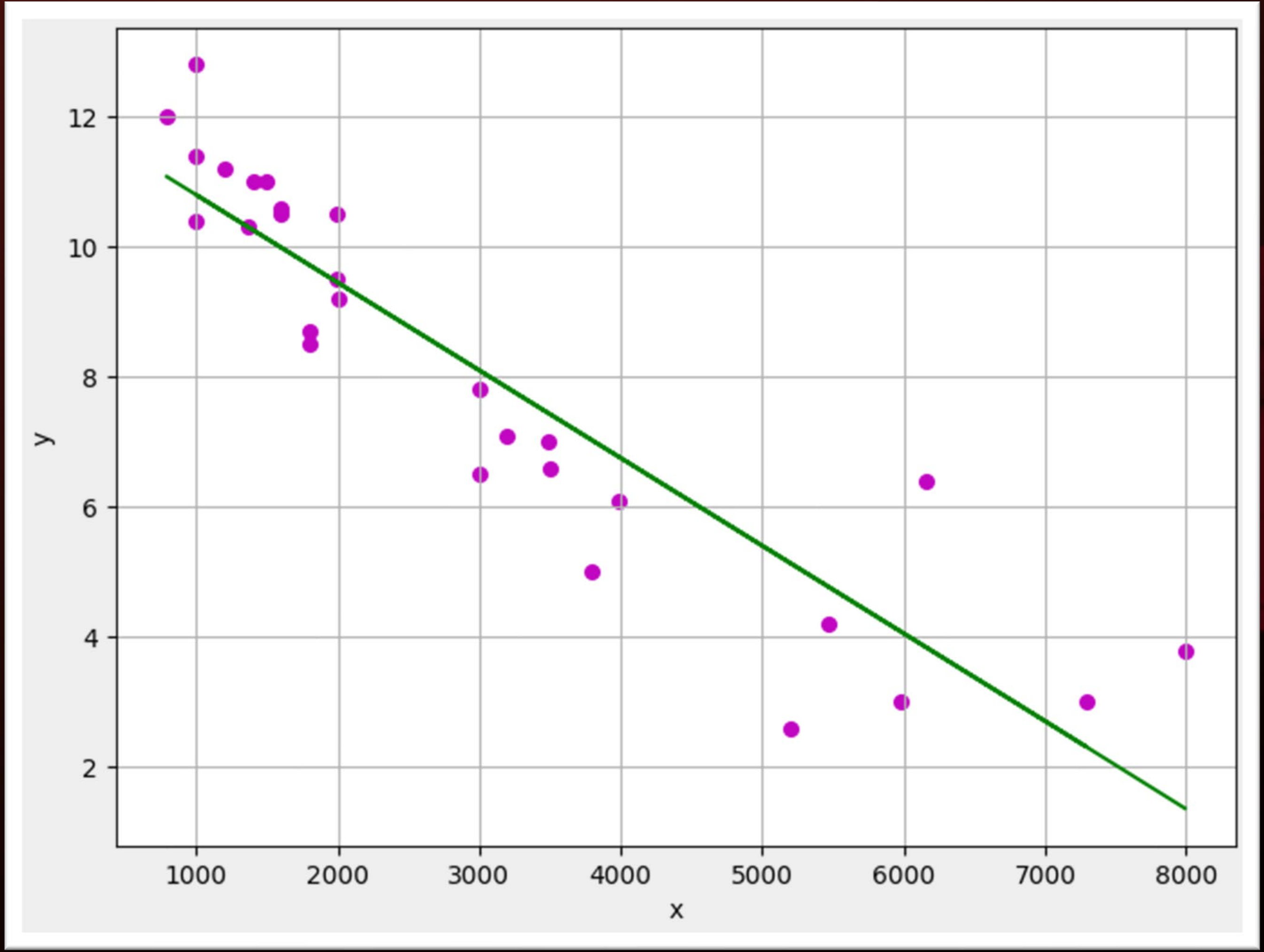
$\beta_0 = \bar{y} - \beta_1 \bar{x} = 8{,}0964 - (-0{,}00135)(3003)$

$$\beta_0 = 12{,}14401$$

Portanto:

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

EXEMPLO: RETA DOS MÍNIMOS QUADRADOS



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

Consumo de combustível

Capacidade volumétrica

EXEMPLO

RETA DE REGRESSÃO

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

Consumo de combustível Capacidade volumétrica

b) Valor esperado de y para x= 3000

$$\hat{\mu}_{y:3000} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

$$\hat{\mu}_{y:3000} = 12{,}14401 - (0{,}00135)(3000)$$

$$\hat{\mu}_{y:3000} = 8{,}094$$

Observe que o valor esperado ($\hat{\mu}_{y:x*}$ ou y) nem sempre será igual ao valor observado ou medido (y_i)

EXEMPLO

c) Calcular a SQE

O valor médio de y=8,0964,

1. Calcular o valor esperado de y → $\hat{y}_i$

$$\hat{y}_i = 12{,}14401 - 0{,}00135x_i$$

Ex: da tabela de dados para i=7 x=1997, portanto:

Consumo	12	10,4	12,8	10,3	10,5	8,5	9,5
Cap vol.	792	994	1000	1368	1598	1796	1997

$$\hat{y}_7 = 12{,}14401 - 0{,}00135x_7$$

$$\hat{y}_7 = 12{,}14401 - (0{,}00135)(1997)$$

$$\hat{y}_7 = 9{,}44806$$

EXEMPLO

c) Calcular a SQE

Valor esperado de y → $\hat{y}_i$ para todos os dados

$$\hat{y}_i = 12{,}14401 - 0{,}00135x_i$$

Aplicando a todos os dados, obtém-se a seguinte tabela:

y_i	12	10,4	12,8	10,3	10,5	8,5	9,5
$\hat{y}_i$	11,07	10,80	10,79	10,30	9,99	9,72	9,45
y_i	9,2	6,5	7,1	6,6	4,2	6,4	3
$\hat{y}_i$	9,45	8,10	7,83	7,42	4,77	3,83	2,30
y_i	11,4	11,2	11	11	10,6	8,7	10,5
$\hat{y}_i$	10,80	10,53	10,26	10,12	9,99	9,72	9,45
y_i	7,8	7	5	6,1	2,6	3	3,8
$\hat{y}_i$	8,10	7,43	7,02	6,77	5,12	4,07	1,35

EXEMPLO

c) Calcular a SQE

2. Calcular os resíduos →$y_i - \hat{y}_i$

Ex: da tabela de dados para i=7 x=1997:

y_i	12	10,4	12,8	10,3	10,5	8,5	9,5
$\hat{y}_i$	11,07	10,80	10,79	10,30	9,99	9,72	9,45

O valor esperado é $\hat{y}_7 = 9{,}44806 \cong 9{,}45$

O valor observado é $y_7 = 9{,}5$

Resíduo: $y_7 - \hat{y}_7 = 9{,}5 - 9{,}45$

$y_7 - \hat{y}_7 = 0{,}05$

EXEMPLO

c) Calcular a SQE

Resíduos → $(y_i - \hat{y}_i)$ → (+/-)
Aplicando a todos os dados, obtém-se a seguinte tabela:

y_i	12	10,4	12,8	10,3	10,5	8,5	9,5
y_i	11,07	10,80	10,79	10,30	9,99	9,72	9,45
$y_i - \hat{y}_i$	0,93	-0,40	2,01	0,00	0,51	-1,22	0,05
y_i	9,2	6,5	7,1	6,6	4,2	6,4	3
y_i	9,45	8,10	7,83	7,42	4,77	3,83	2,30
$y_i - \hat{y}_i$	-0,25	-1,60	-0,73	-0,82	-0,57	2,57	0,70
y_i	11,4	11,2	11	11	10,6	8,7	10,5
y_i	10,80	10,53	10,26	10,12	9,99	9,72	9,45
$y_i - \hat{y}_i$	0,60	0,67	0,74	0,88	0,61	-1,02	1,05
y_i	7,8	7	5	6,1	2,6	3	3,8
y_i	8,10	7,43	7,02	6,77	5,12	4,07	1,35
$y_i - \hat{y}_i$	-0,30	-0,43	-2,02	-0,67	-2,52	-1,07	2,45

EXEMPLO

c) Calcular a SQE

3. Calcular os quadrados dos resíduos → $(y_i - \hat{y}_i)^2$.
Aplicando a todos os dados, obtém-se a seguinte tabela:

y_i	12	10,4	12,8	10,3	10,5	8,5	9,5
y_i	11,07	10,80	10,79	10,30	9,99	9,72	9,45
$y_i - \hat{y}_i$	0,93	-0,40	2,01	0,00	0,51	-1,22	0,05
$(y_i - \hat{y}_i)^2$	0,86	0,16	4,02	0,00	0,26	1,49	0,00
y_i	9,2	6,5	7,1	6,6	4,2	6,4	3
y_i	9,45	8,10	7,83	7,42	4,77	3,83	2,30
$y_i - \hat{y}_i$	-0,25	-1,60	-0,73	-0,82	-0,57	2,57	0,70
$(y_i - \hat{y}_i)^2$	0,06	2,56	0,53	0,68	0,33	6,63	0,49
y_i	11,4	11,2	11	11	10,6	8,7	10,5
y_i	10,80	10,53	10,26	10,12	9,99	9,72	9,45
$y_i - \hat{y}_i$	0,60	0,67	0,74	0,88	0,61	-1,02	1,05
$(y_i - \hat{y}_i)^2$	0,37	0,46	0,55	0,77	0,38	1,03	1,11
y_i	7,8	7	5	6,1	2,6	3	3,8
y_i	8,10	7,43	7,02	6,77	5,12	4,07	1,35
$y_i - \hat{y}_i$	-0,30	-0,43	-2,02	-0,67	-2,52	-1,07	2,45
$(y_i - \hat{y}_i)^2$	0,09	0,18	4,06	0,45	6,34	1,15	5,99

EXEMPLO

c) Calcular a SQE

SQE calculando da somatória da tabela (soma na calculadora ou no Excel)

$$\text{SQE} = \sum (y_i - \hat{y}_i)^2 = 40{,}9868$$

SQE de acordo com as fórmulas:

$$\text{SQE} = \sum {y_i}^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

$$\text{SQE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

EXEMPLO

c) Calcular a SQE

SQE de acordo com as fórmulas:

$$\mathrm{SQE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

Lembrado que, do exemplo 1, $S_{xy} = -151950$,

$\beta_1 = -0{,}00135$ e S_{yy} deve ser calculado

$$S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 2081{,}25 - \frac{226{,}7^2}{28}$$

$$S_{yy} = 245{,}7896$$

$$\mathrm{SQE} = S_{yy} - \hat{\beta}_1 S_{xy} = 245{,}7896 - (-0{,}00135)(-151950)$$

$$\mathrm{SQE} = 40{,}6571$$

EXEMPLO

c) Calcular s^2 e s.

Resultados finais SQE, s^2 e s:

$$\mathrm{SQE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

$$\mathrm{SQE} = 40{,}6571$$

Lembrando que n=28

$$s^2 = \frac{SQE}{n-2} = \frac{40{,}6571}{28-2} = 1{,}563$$

Finalmente, o desvio padrão estimado:

$$s = \sqrt{s^2} = \sqrt{1{,}563} = 1{,}250$$

EXEMPLO

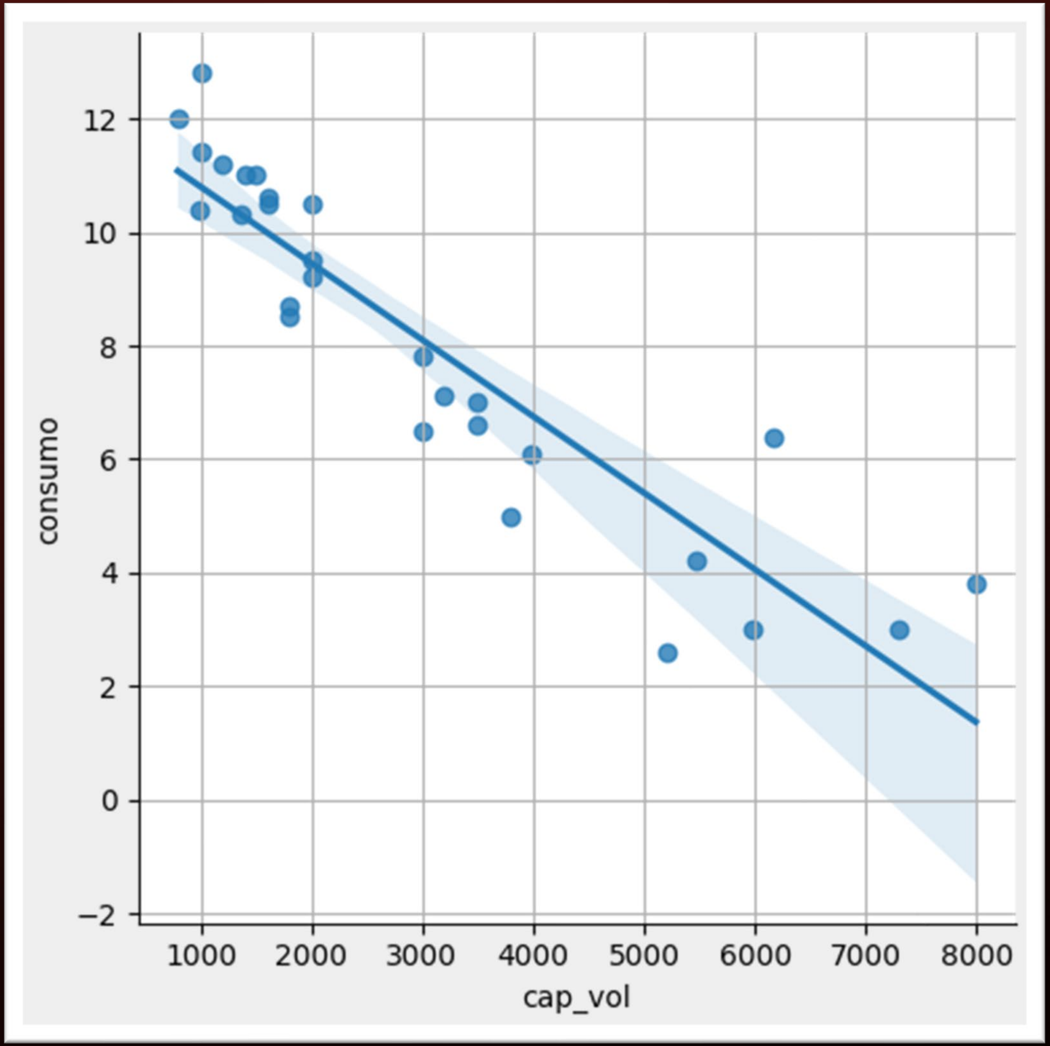
d) Calcular SQT e r²

$$\mathrm{SQT} = S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 2081{,}25 - \frac{226{,}7^2}{28}$$

$$SQT = 245{,}7896$$

$$r^2 = 1 - \frac{SQE}{SQT} = 1 - \frac{40{,}6571}{245{,}7896} = 0{,}8346 \; ou \; 83{,}46\ \%$$

EXEMPLO IC DE β



EXEMPLO

e) Obter o erro padrão estimado

Se $s = 1{,}250$ e $S_{xx} = 112735060.$

calcular o erro padrão de $\boldsymbol{\beta_1}$ $\quad s_{\hat{\beta}_1} = \dfrac{s}{\sqrt{S_{xx}}}$

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} = \frac{1{,}250}{\sqrt{112735060}}$$

$$s_{\hat{\beta}_1} = 1{,}177 \times 10^{-4}$$

EXEMPLO

e) IC de β_1 para o nível de confiança de 95%

Se $s_{\hat{\beta}_1} = 1{,}177 \times 10^{-4}$

Um nível de confiança de 95% indica que 100(1-α)=95%

Assim α=0,05 e α/2=0,025

$$t_{0{,}025{,}26} = 2{,}056$$

Para um nível de confiança de 95%

$\beta_1 = -0{,}00135$

$\hat{\beta}_1 = \beta_1 \pm \boldsymbol{t_{\alpha/2,n-2}}\ s_{\hat{\beta}_1}$

$\hat{\beta}_1 = -0{,}00135 \pm (2{,}056)(1{,}177 \times 10^{-4})$

$\hat{\beta}_1 = -0{,}00135 \pm 0{,}000295$

$\hat{\beta}_1$(-0,001645, -0,001055)

MODELAGEM E INFERÊNCIA ESTATÍSTICA

Análise de ajustes: Exemplo prático