Planejamento e Análise

modelos lineares

Alexandre Adalardo de Oliveira

PlAnEco 2024

Modelos Lineares

Modelos Lineares I

- estrutura do modelo linear
- mínimos quadrados
- resíduos do modelo
- erro padrão das estimativas
- interpretação do summary

Modelo Linear Simples

$y = \alpha + \beta x + \epsilon$

$\epsilon = N(0, \sigma)$

Simulando dados

$y = \alpha + \beta x + \epsilon$

$y = 12 + 2.5x$

$\epsilon_i = N(0, 5)$

$y_i = 12 + 2.5x_i + \epsilon_i$

Simulando dados (x)

$y_i = 12 + 2.5x_i + N(0, 5)$

x1	y0	res	y1
0.0			
0.5			
1.0			
1.5			
2.0			
2.5			
3.0			
3.5			
4.0			
4.5			
5.0			

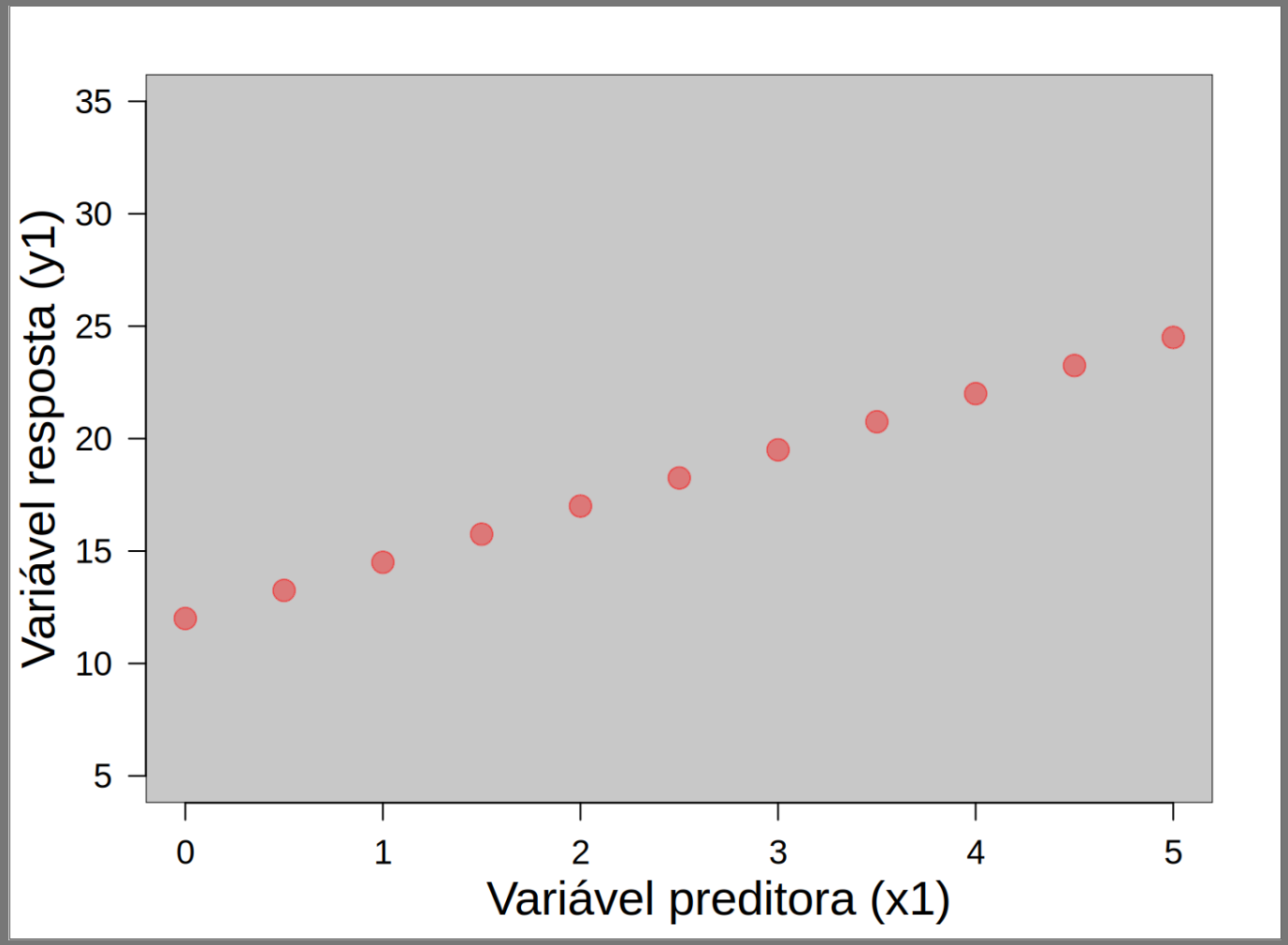
Estrutura deterministica

$y_i = 12 + 2.5x_i$

x1	y0	res	y1
0.0	12.00		
0.5	13.25		
1.0	14.50		
1.5	15.75		
2.0	17.00		
2.5	18.25		
3.0	19.50		
3.5	20.75		
4.0	22.00		
4.5	23.25		
5.0	24.50		

Estrutura determinística

$$y_0 = 12 + 2.5x_1$$



Estrutura estocástica

$\epsilon = N(0, 5)$

x1	y0	res	y1
0.0	12.00	6.85	
0.5	13.25	−2.82	
1.0	14.50	1.82	
1.5	15.75	3.16	
2.0	17.00	2.02	
2.5	18.25	−0.53	
3.0	19.50	7.56	
3.5	20.75	−0.47	
4.0	22.00	10.09	
4.5	23.25	−0.31	
5.0	24.50	6.52	

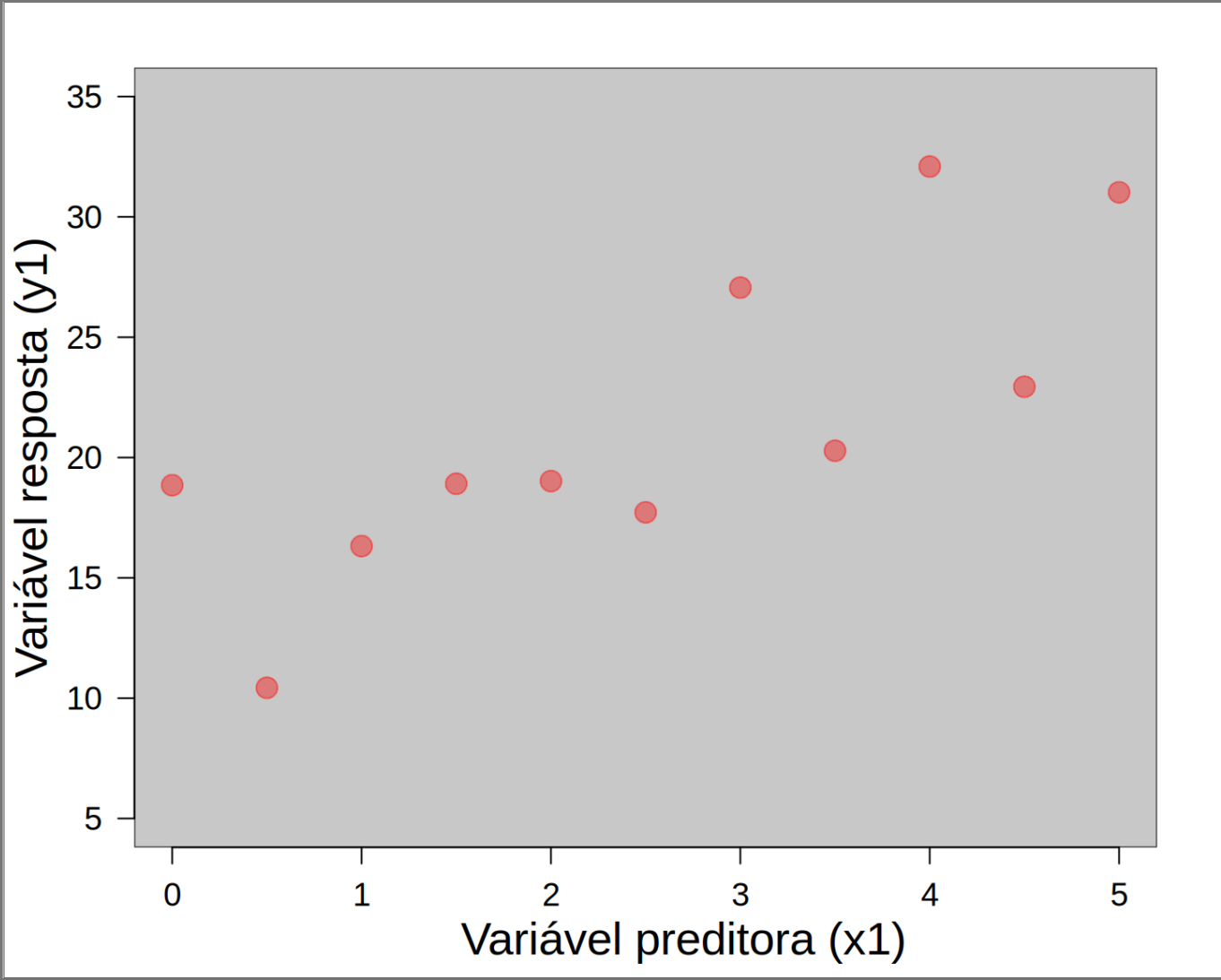
Dados simulados

$y_i = 12 + 2.5x_i + N(0, 5)$

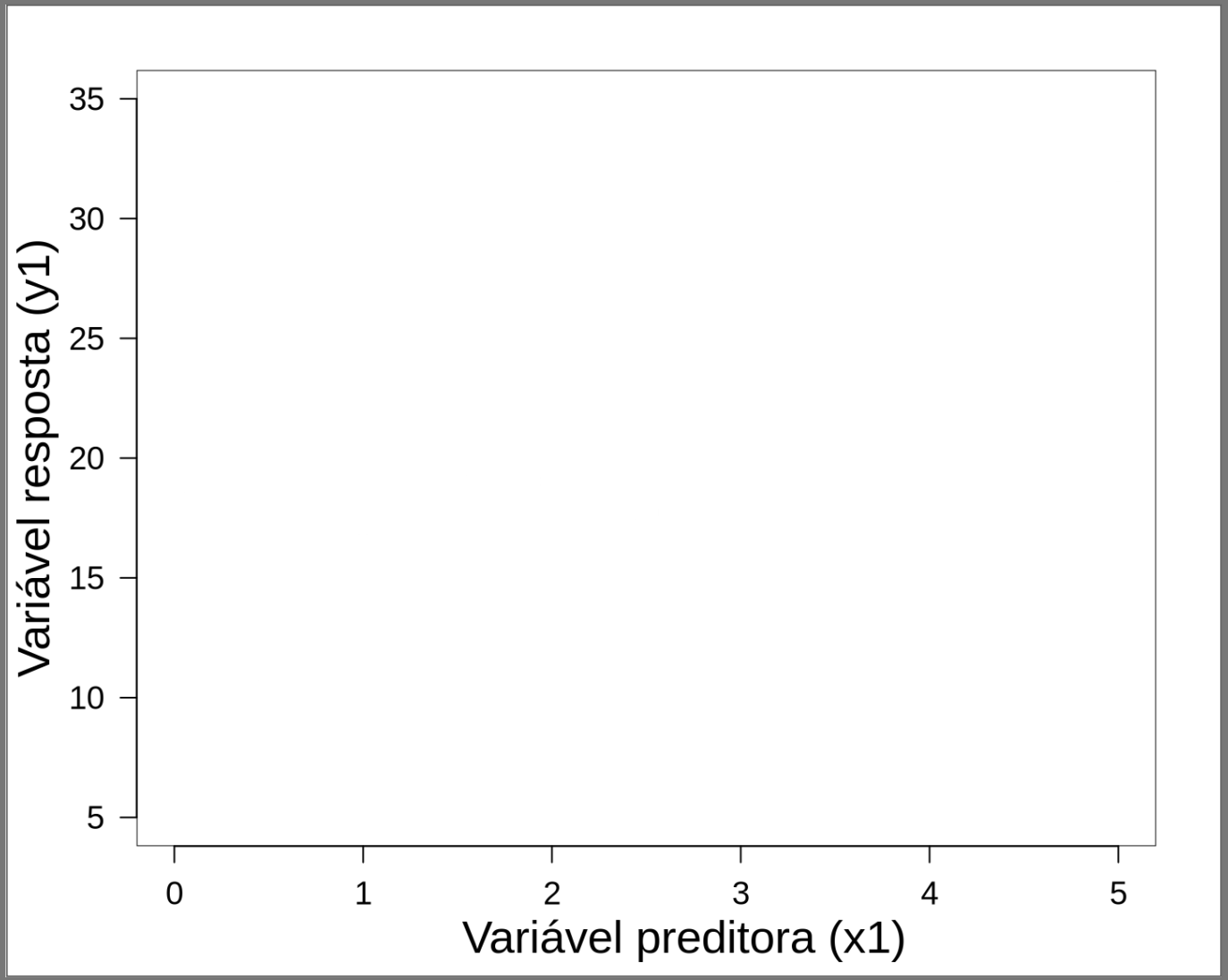
x1	y0	res	y1
0.0	12.00	6.85	18.85
0.5	13.25	−2.82	10.43
1.0	14.50	1.82	16.32
1.5	15.75	3.16	18.91
2.0	17.00	2.02	19.02
2.5	18.25	−0.53	17.72
3.0	19.50	7.56	27.06
3.5	20.75	−0.47	20.28
4.0	22.00	10.09	32.09
4.5	23.25	−0.31	22.94
5.0	24.50	6.52	31.02

Dados simulados

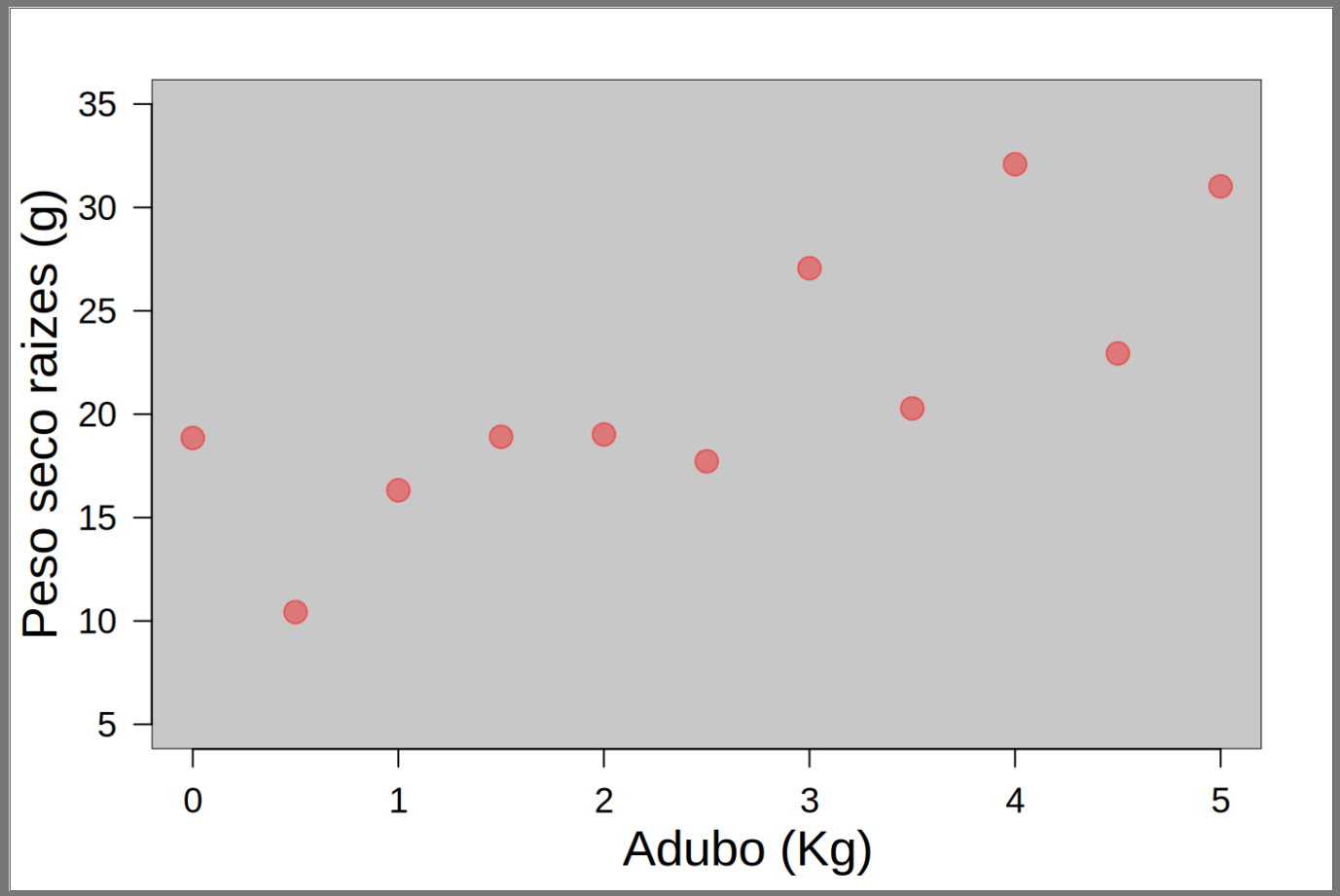
$$y_i = 12 + 2.5x_i + N(0, 5)$$



Dados REAIS

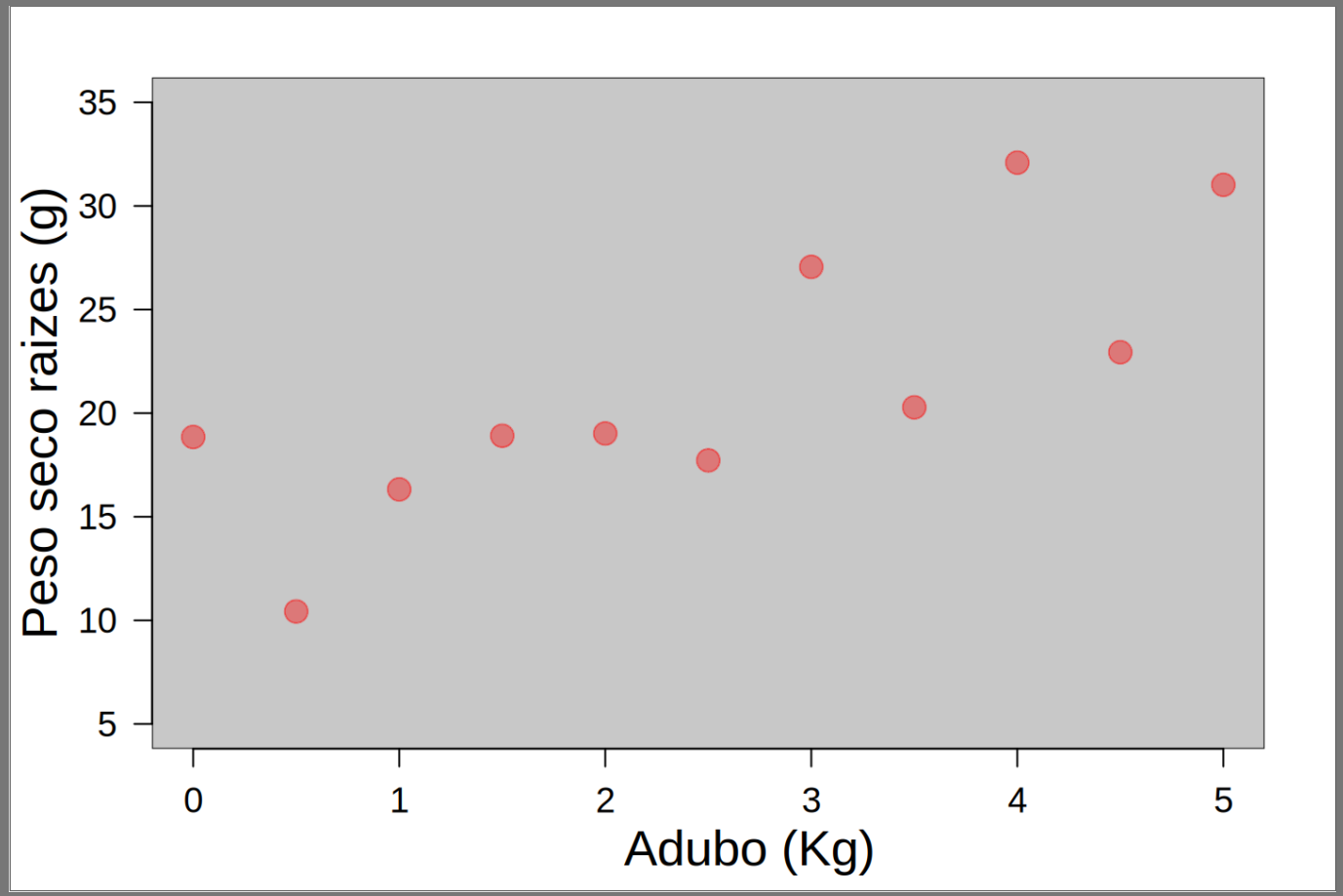

35
30
25
20
15
10
5
Variável resposta (y1)
0
1
2
3
4
5
Variável preditora (x1)

Dados REAIS

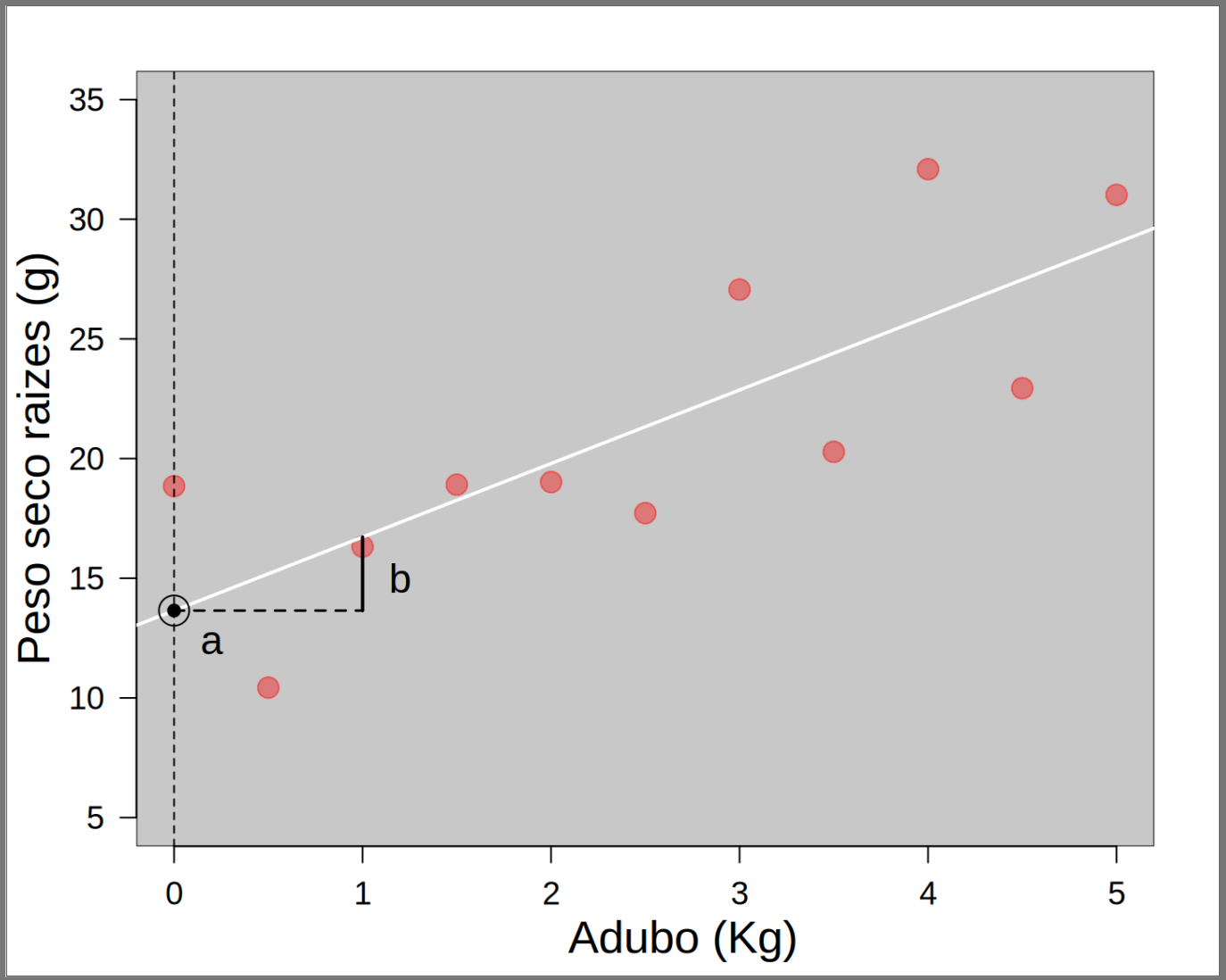


Estimar os parâmetros

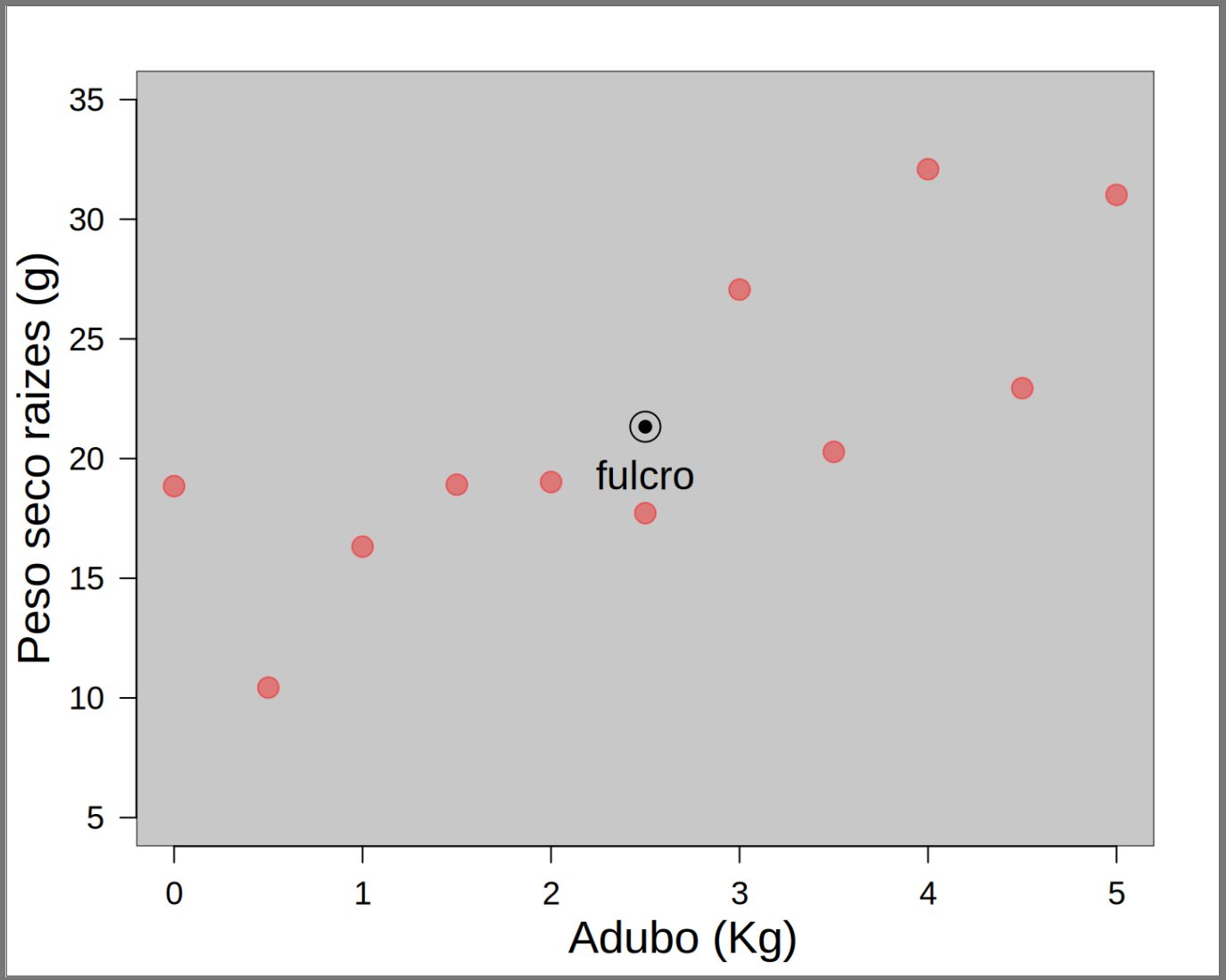
$$y = a + bx$$



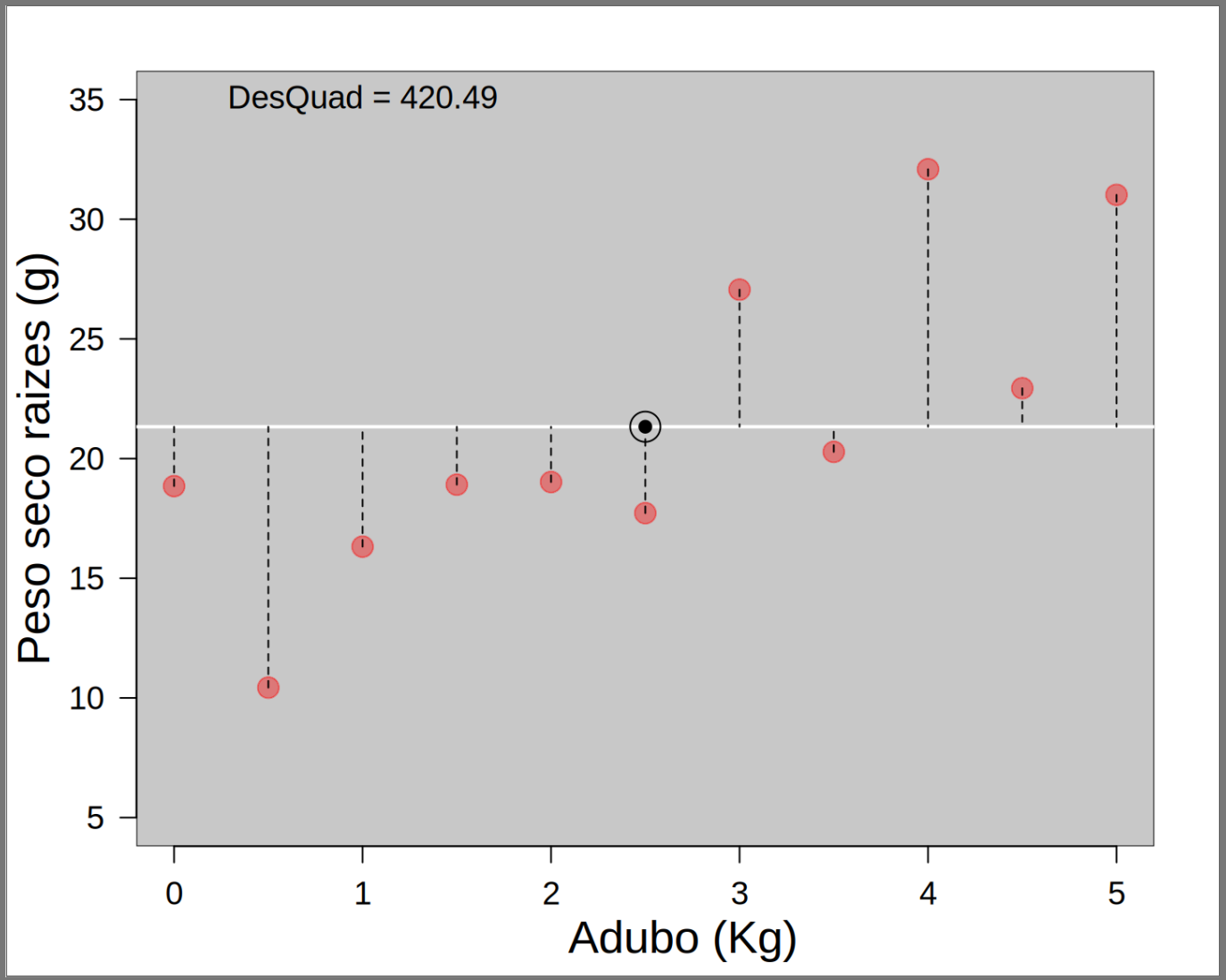
Modelo Linear



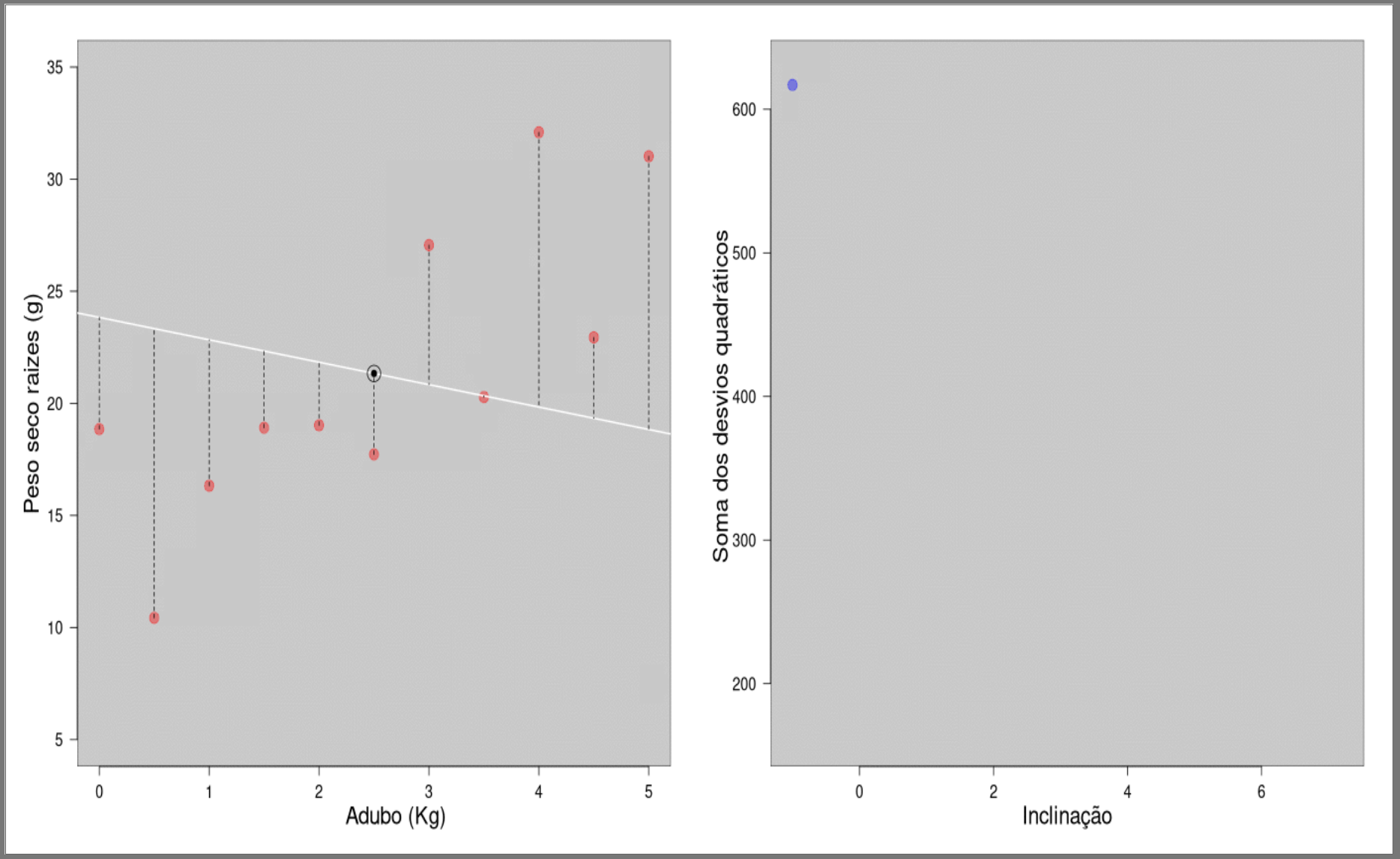
Minimos Quadrados



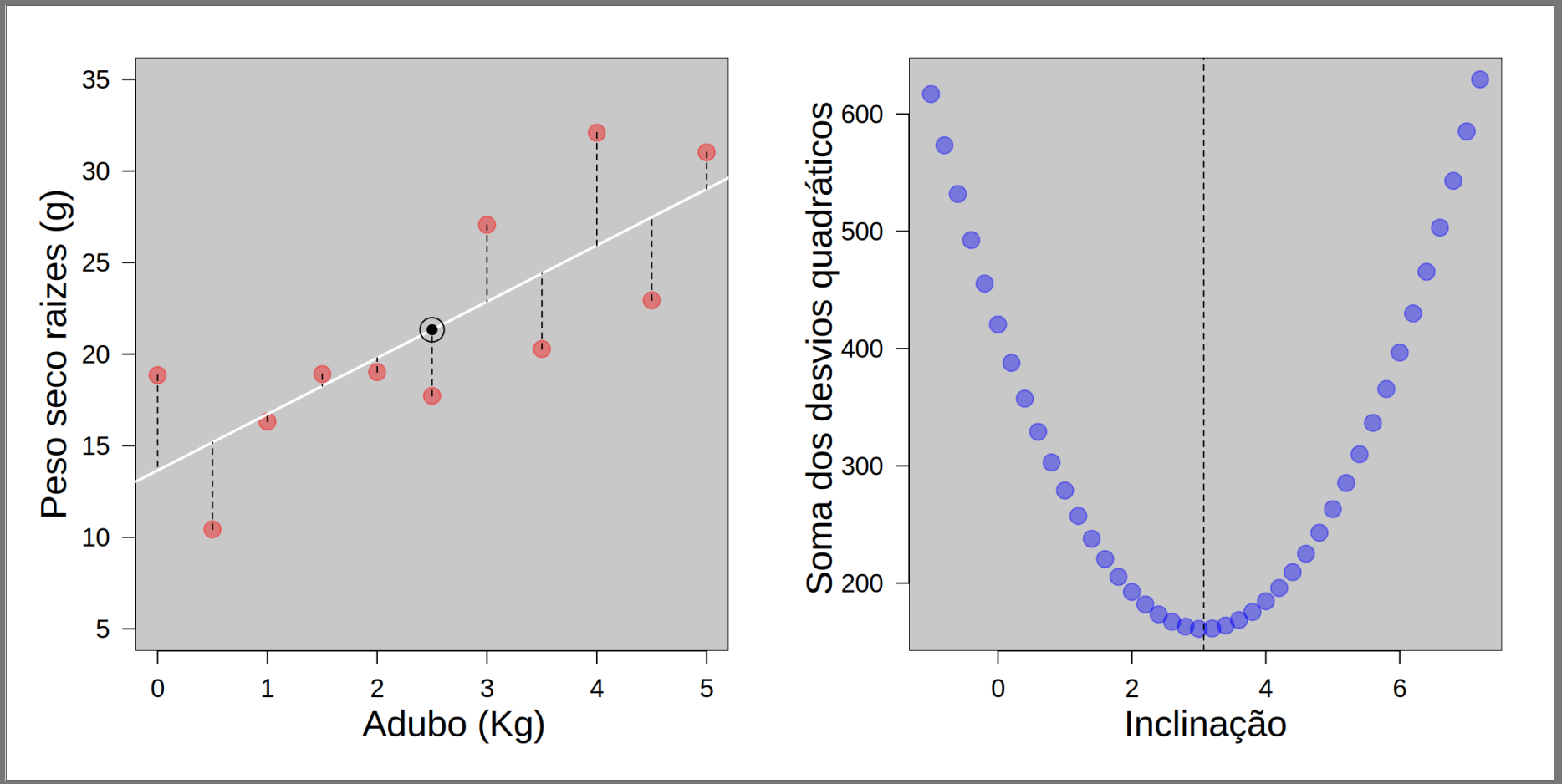
Um modelo possível



Mínimos Quadrados

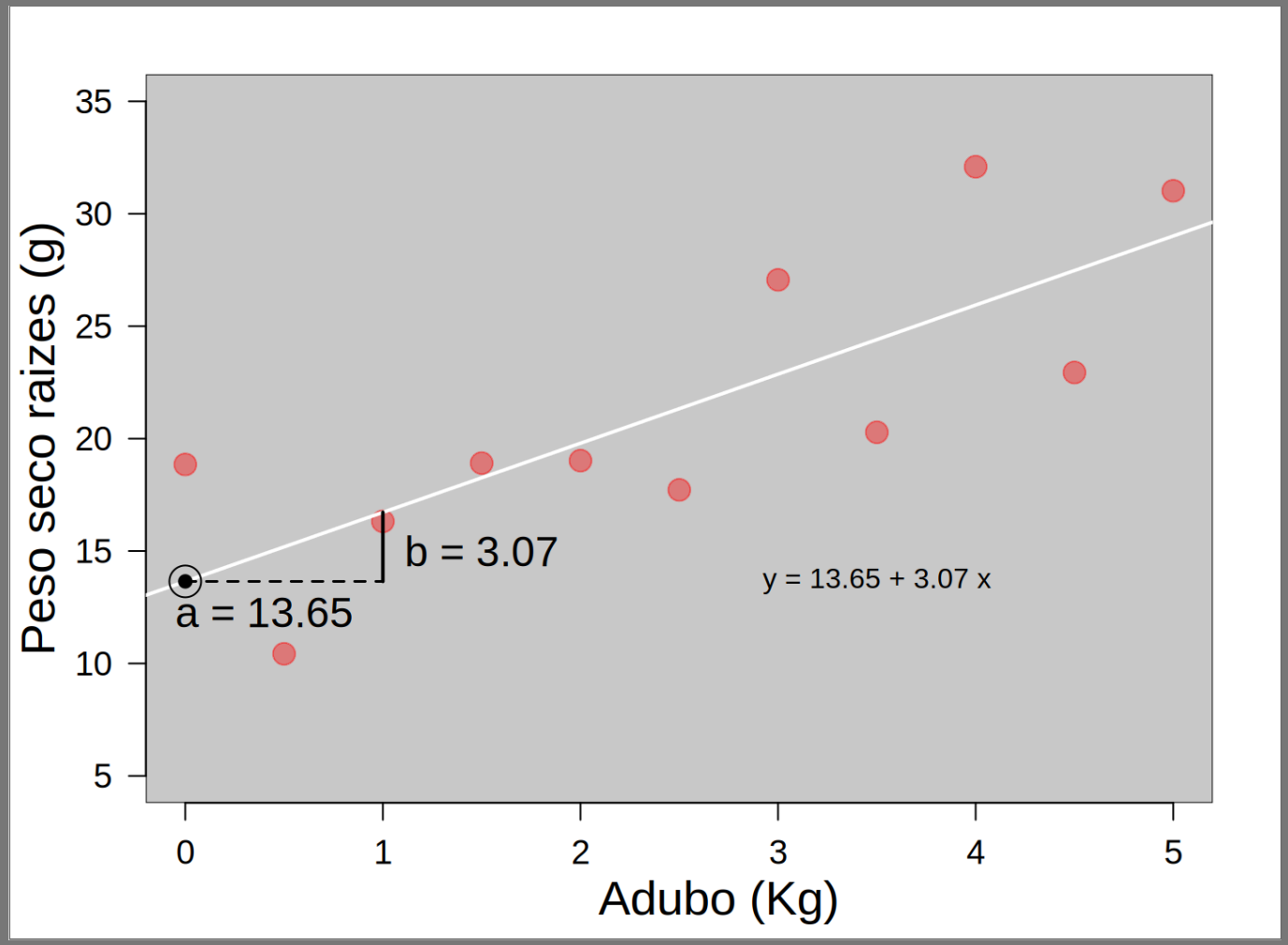


Mínimos Quadrados



Parâmetros estimados

$$y_i = 12 + 2.5x_i + N(0, 5)$$

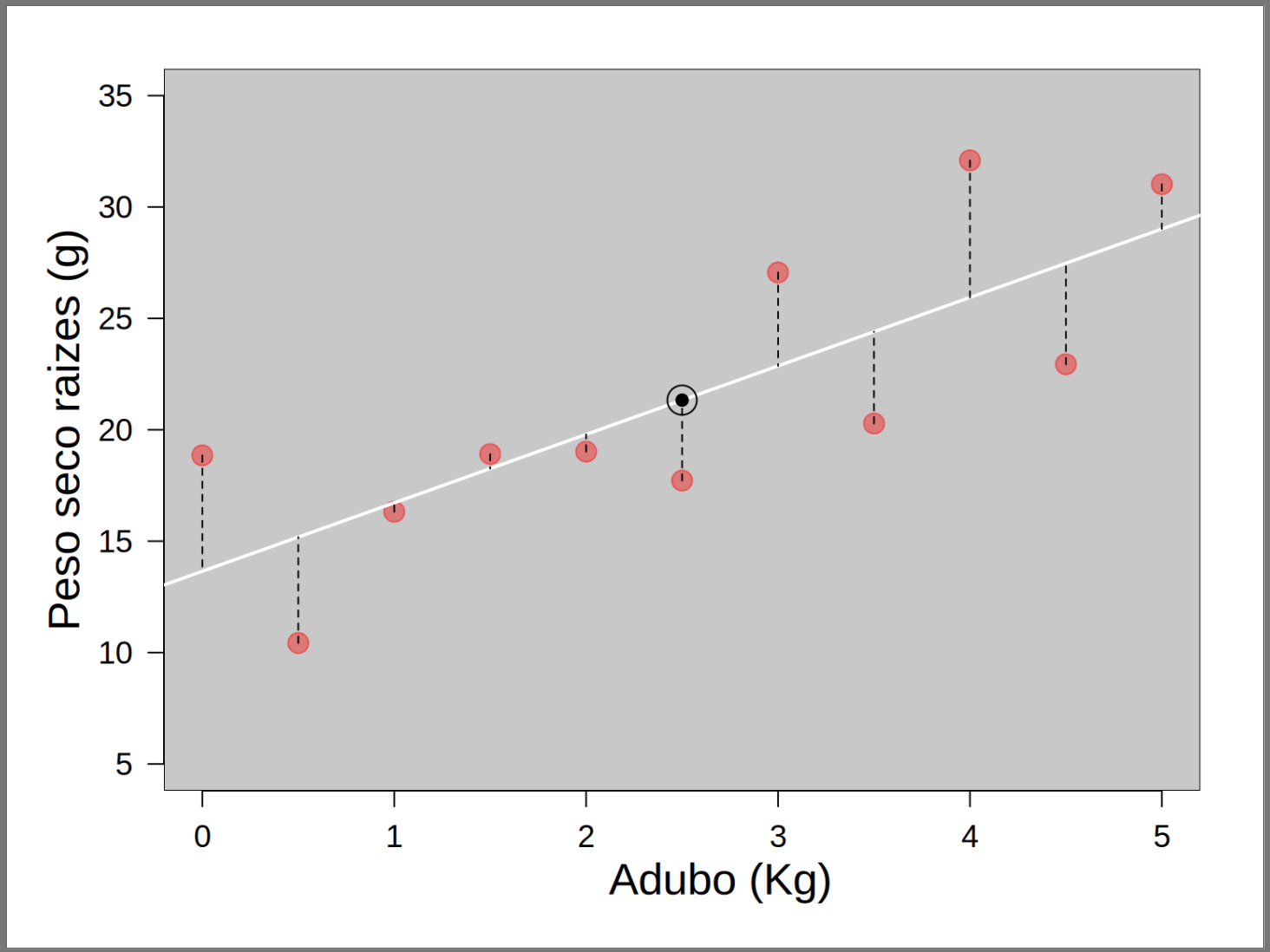


Modelo Linear

```r
lm01 <- lm(y1 ~ x1, data = xy)
summary(lm01)
```

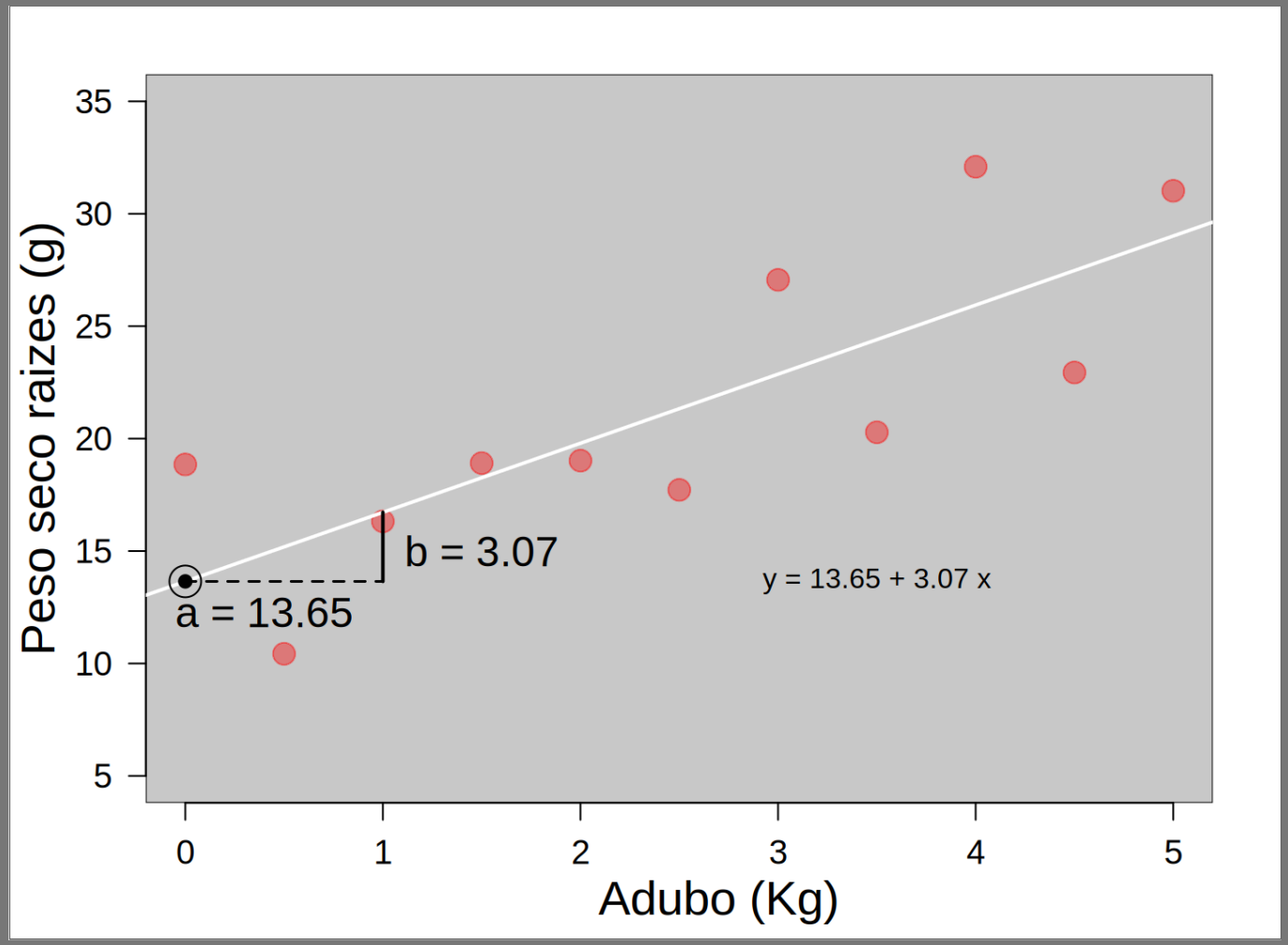
```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

Resíduos



Parâmetros estimados

$$y_i = 12 + 2.5x_i + N(0, 5)$$



Estimativas de incerteza

```r
lm01 <- lm(x1 ~ y1, data = xy)
summary(lm01)
```

Erro padrão da inclinação

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

Erro Padrão da inclinação

A incerteza da estimativa da inclinação

$$se_\beta = \sqrt{\frac{1}{n-2}\frac{\sum (y_i - \hat{y}_i)^2}{\sum (x_i - \bar{x}_i)^2}}$$

```
sqrt(1/(length(x1)-2) *
     sum((y1 - predict(lm01))^2)/sum((x1- mean(x1))^2))
```

```
[1] 0.8062941
```

Erro padrão da inclinação

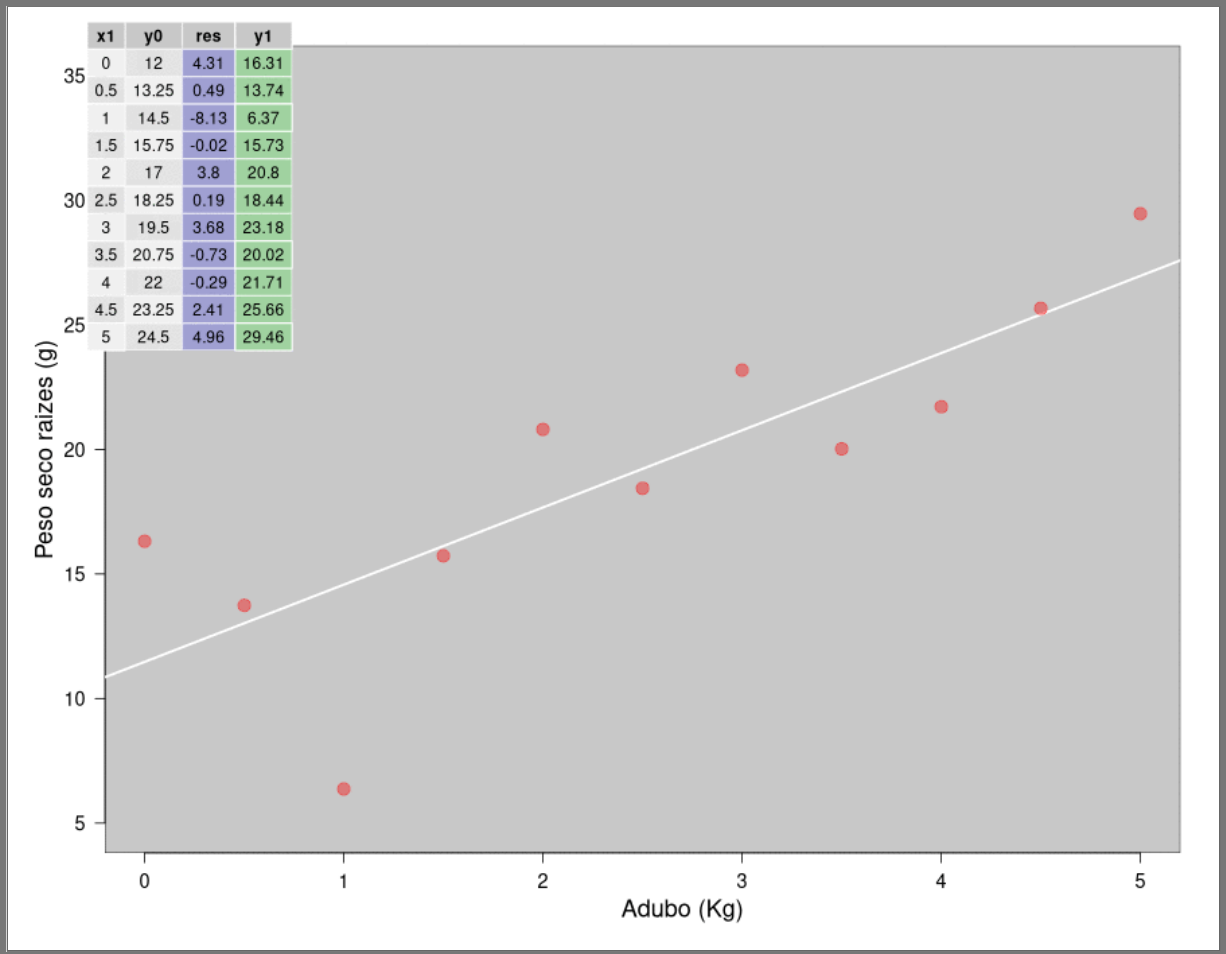
Medida de incerteza da estimativa

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

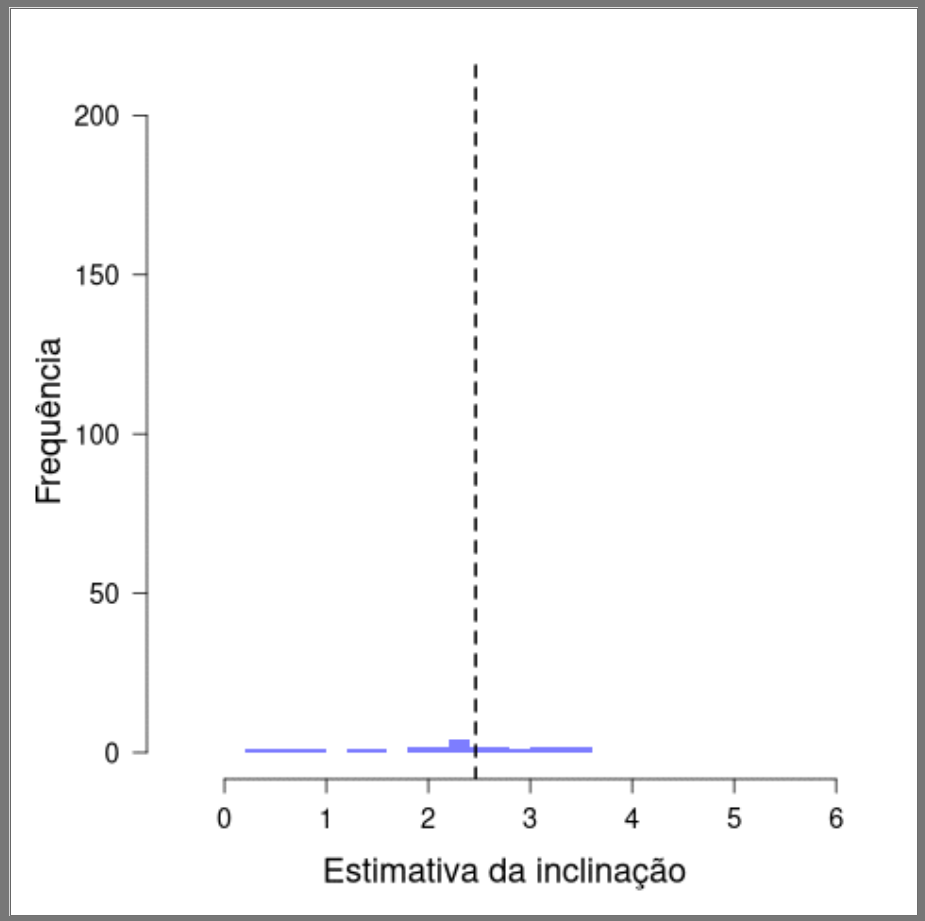
Repetindo a simulação

$$y_i = 12 + 2.5x_i + N(0, 5)$$

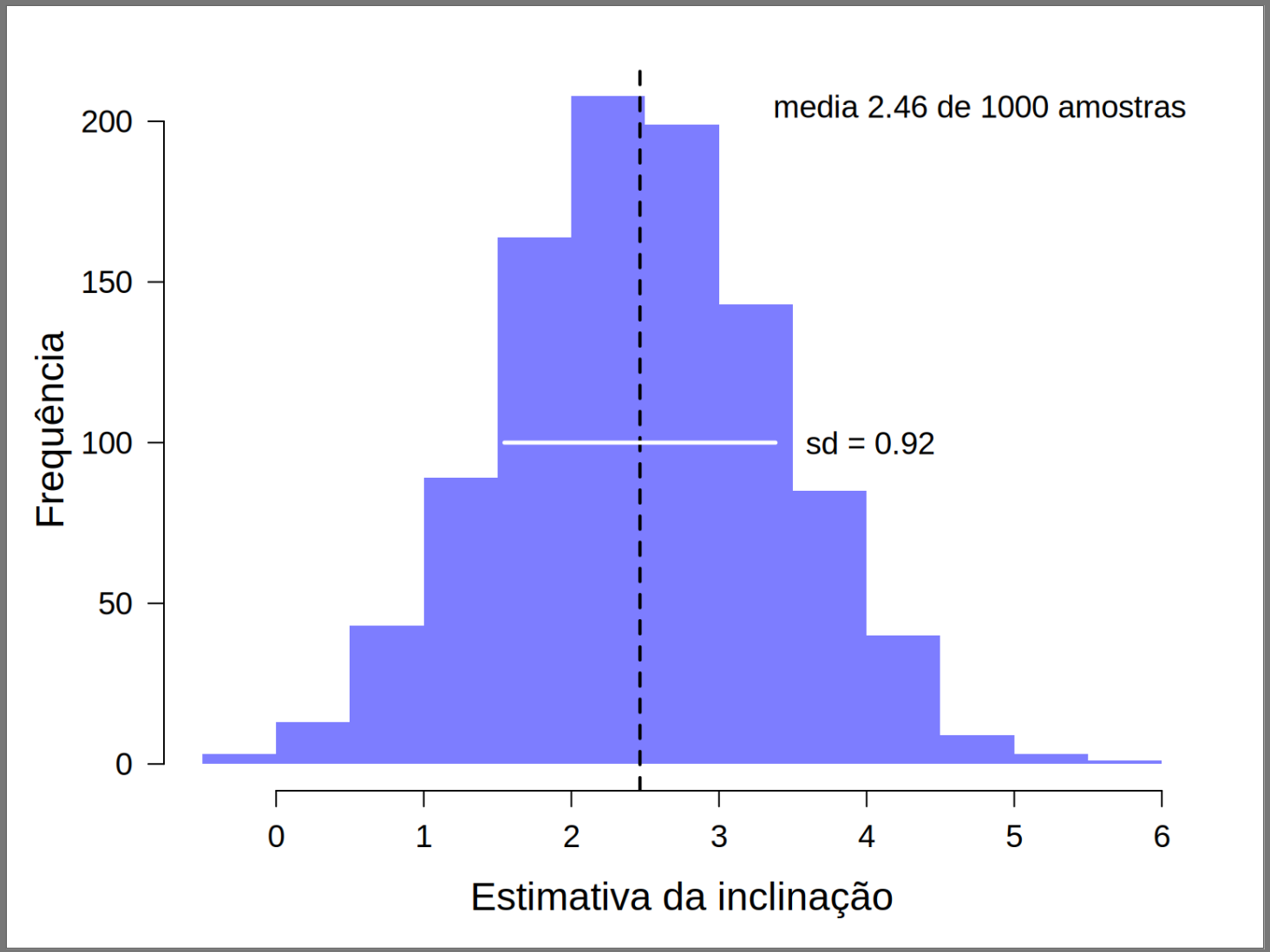
x1	y0	res	y1
0	12	4.31	16.31
0.5	13.25	0.49	13.74
1	14.5	-8.13	6.37
1.5	15.75	-0.02	15.73
2	17	3.8	20.8
2.5	18.25	0.19	18.44
3	19.5	3.68	23.18
3.5	20.75	-0.73	20.02
4	22	-0.29	21.71
4.5	23.25	2.41	25.66
5	24.5	4.96	29.46



Mil estimativas da inclinação



Distribuição de mil amostras

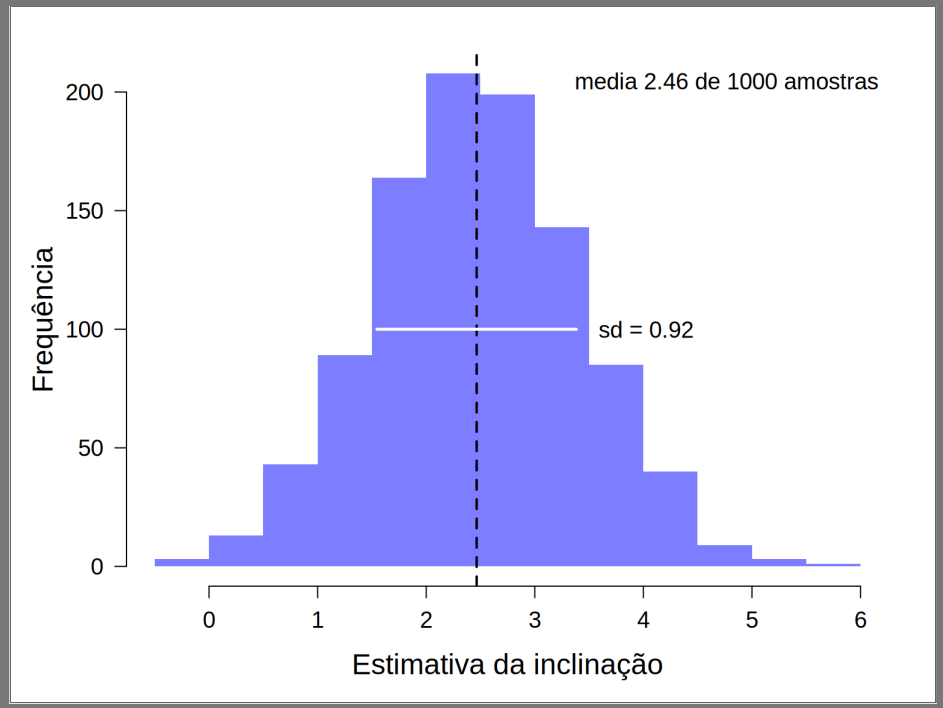


Erro padrão da inclinação

Estima o desvio padrão de inclinações de múltiplas realizações de experimentos

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

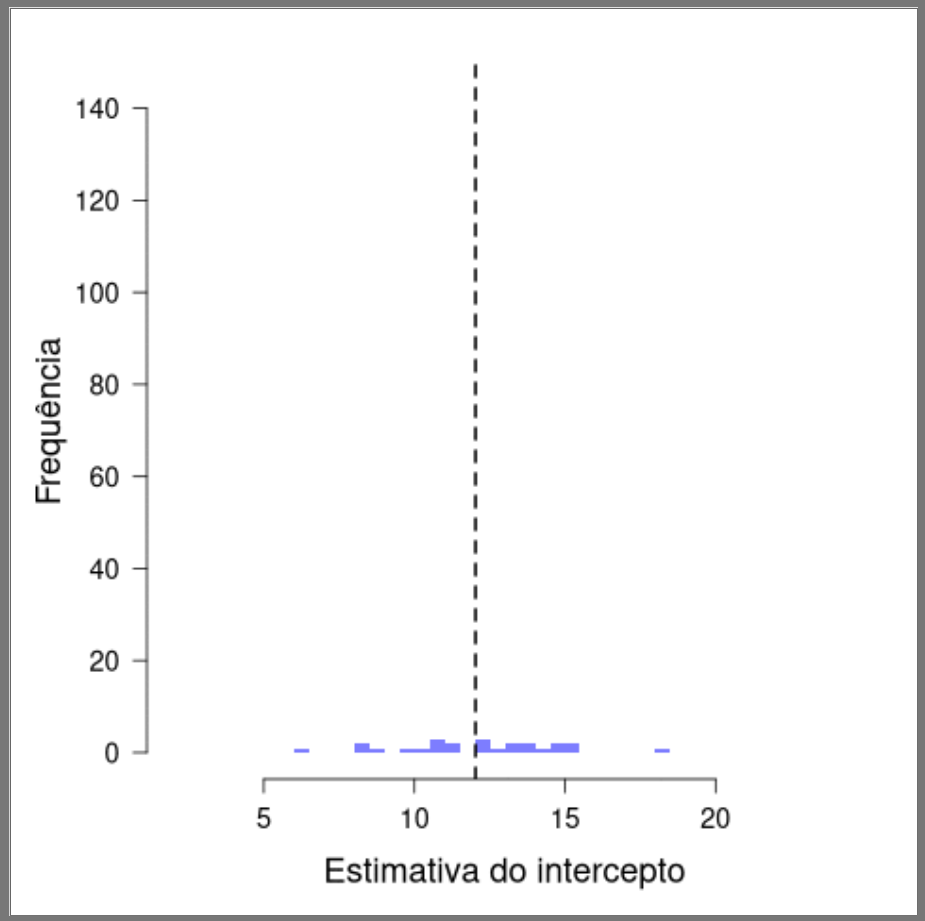
Estimativa de uma amostra



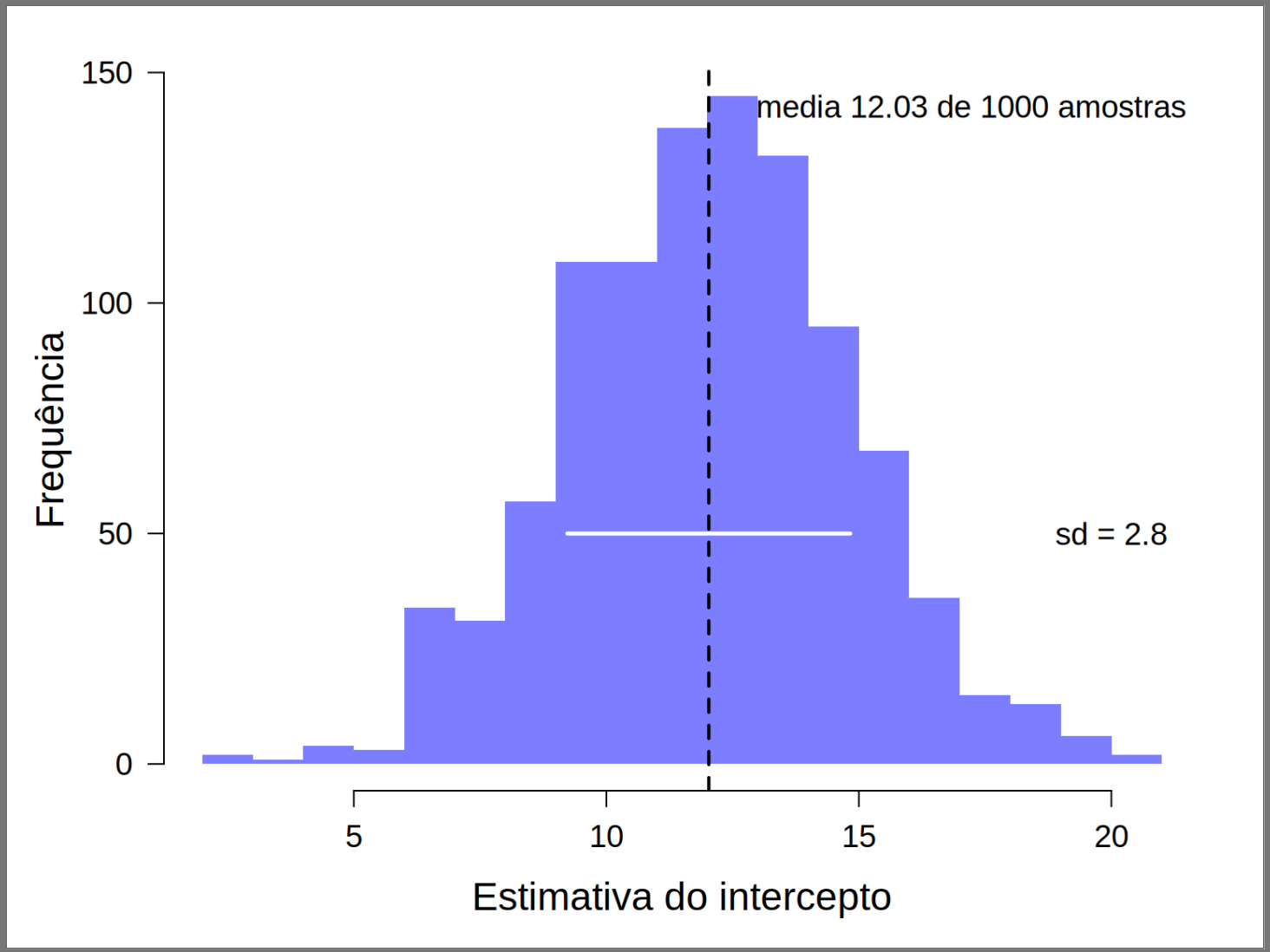

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286
***
x1            3.0724     0.8063   3.810 0.004150
**
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of
freedom
Multiple R-squared:  0.6173,    Adjusted R-
squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value:
0.00415
```

$$y_i = 12 + 2.5x_i + N(0, 5)$$

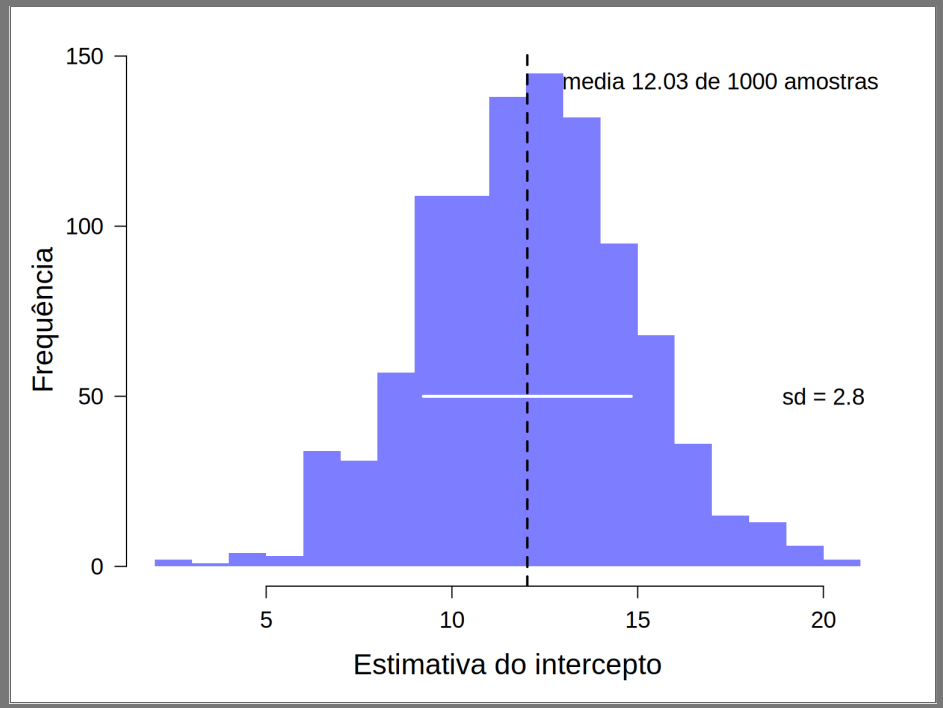
Imprecisão do intercepto



Mil estimativas do intercepto



Estimativa de uma amostra




```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286
***
x1            3.0724     0.8063   3.810 0.004150
**
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of
freedom
Multiple R-squared:  0.6173,    Adjusted R-
squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value:
0.00415
```

$$y_i = 12 + 2.5x_i + N(0, 5)$$

Erro padrão dos resíduos

```
summary(lm01)
```

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7562 -3.8671 -0.4024  3.1005  6.1505 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748 
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

Erro padrão dos resíduos

$$RSE = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2}}$$

```
sqrt(sum((y1 - predict(lm01))^2)/(length(x1)-2))
```

```
[1] 4.228242
```

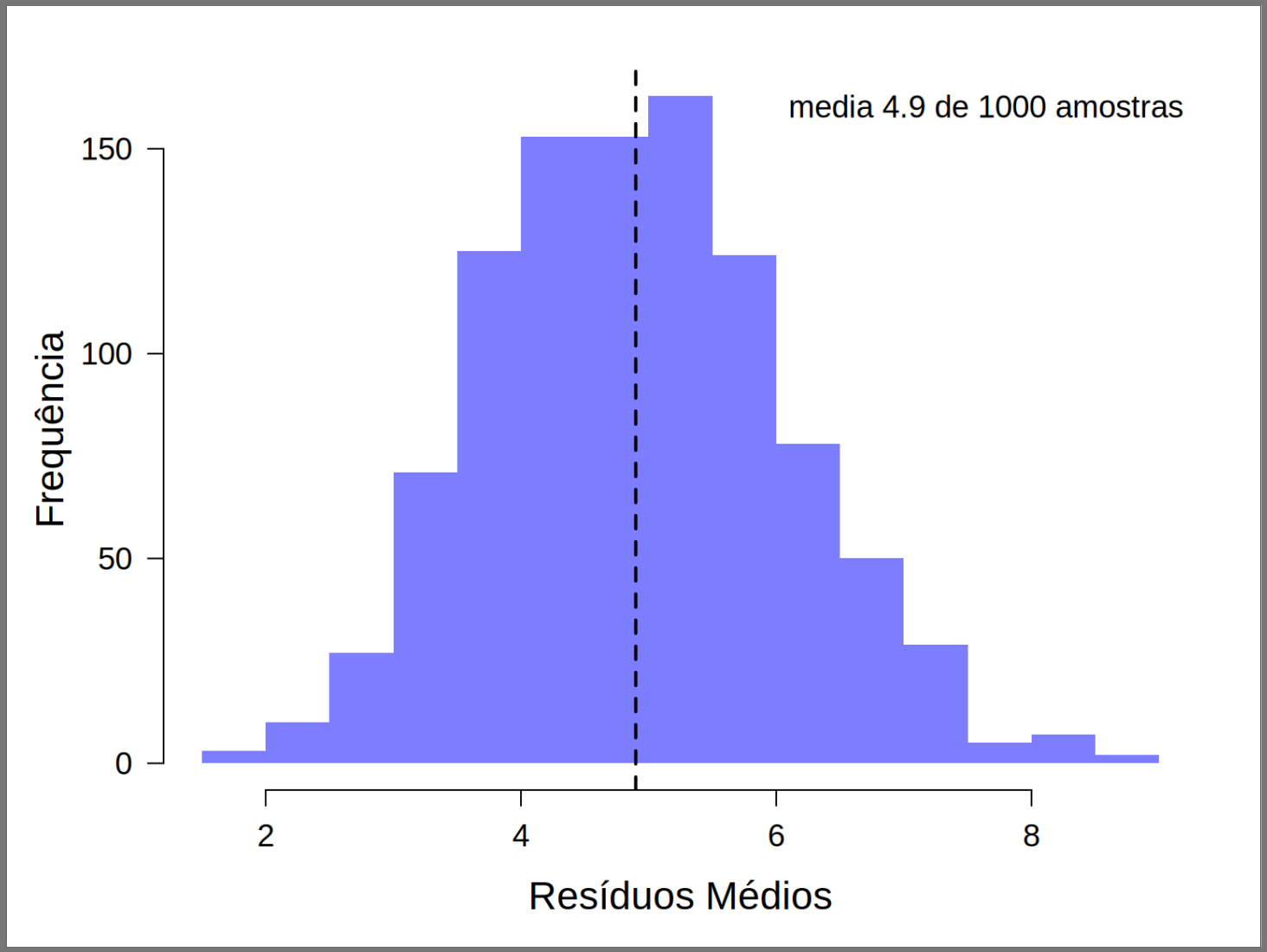
Erro padrão dos resíduos

```
summary(lm01)
```

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7562 -3.8671 -0.4024  3.1005  6.1505 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748 
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

Erro padrão dos resíduos

$$y_i = 12 + 2.5x_i + N(0, 5)$$



Resumo do Summary

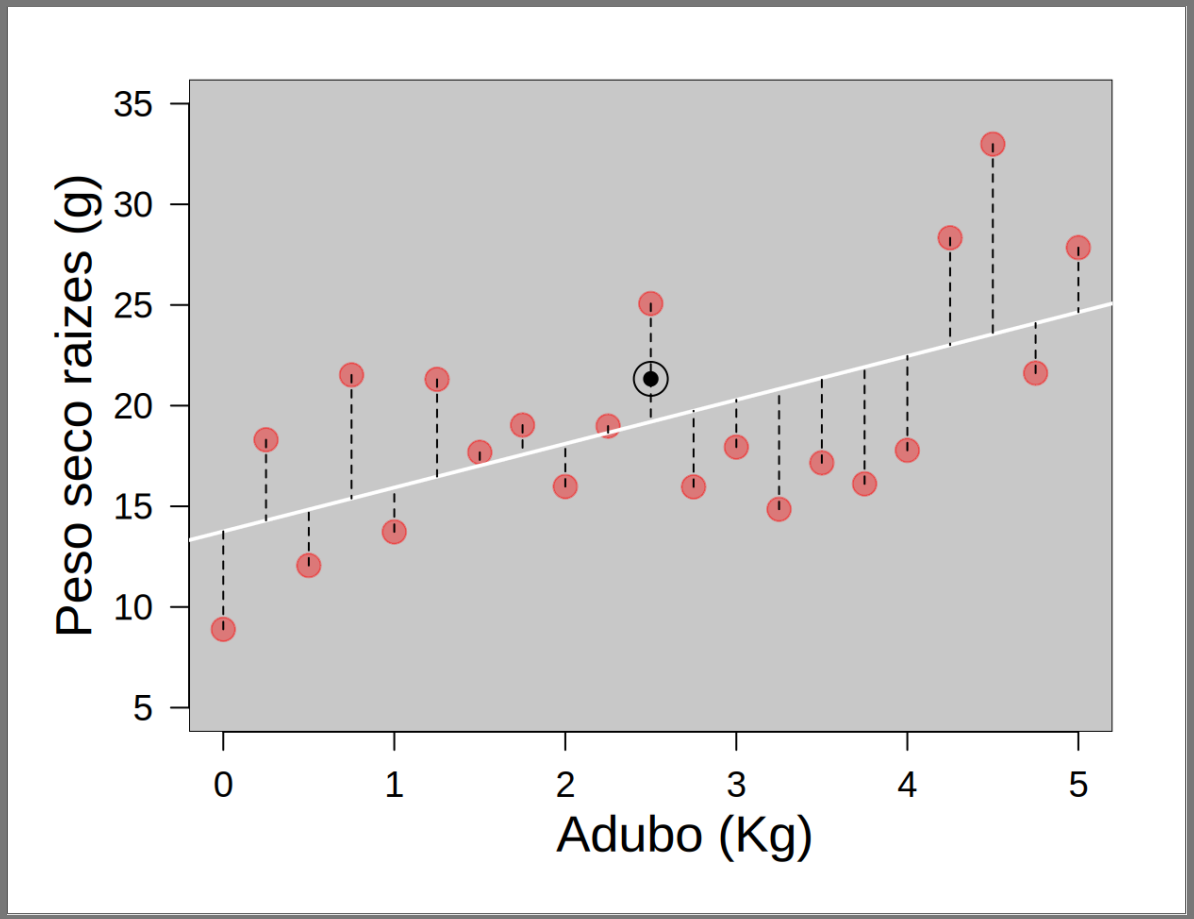
$y_i = 12 + 2.5x_i + N(0, 5)$

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

Fontes de Variabilidade

Esforço amostral

$$y_i = 12 + 2.5x_i + N(0, 5)$$



Comparando os resultados

$$y_i = 12 + 2.5x_i + N(0, 5)$$

$$n = 11$$

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

$$n = 21$$

```
Call:
lm(formula = y2 ~ x2)

Residuals:
   Min     1Q Median     3Q    Max
-5.976 -3.778 -2.130  4.005  9.437

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.755      1.976   6.963 1.24e-06 ***
x2             2.177      0.676   3.221   0.0045 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.689 on 19 degrees of freedom
Multiple R-squared:  0.3532,    Adjusted R-squared:  0.3192
F-statistic: 10.38 on 1 and 19 DF,  p-value: 0.004497
```

Variabilidade do sistema

$n = 11$

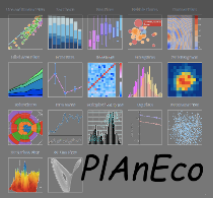
$y_i = 12 + 2.5x_i + N(0, 5)$

```
Call:
lm(formula = y1 ~ x1, data = xy)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7562 -3.8671 -0.4024  3.1005  6.1505

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6500     2.3851   5.723 0.000286 ***
x1            3.0724     0.8063   3.810 0.004150 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.228 on 9 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.5748
F-statistic: 14.52 on 1 and 9 DF,  p-value: 0.00415
```

$y_i = 12 + 2.5x_i + N(0, 10)$

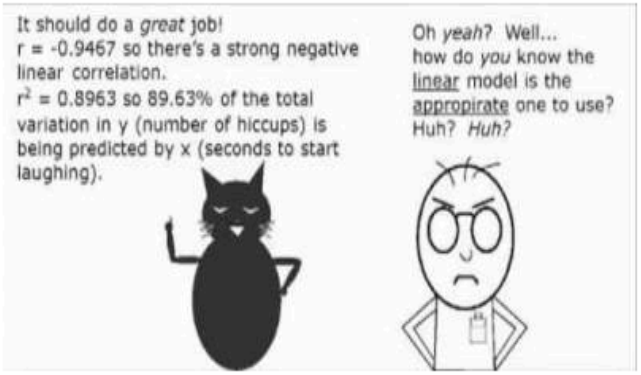
```
Call:
lm(formula = y3 ~ x1)

Residuals:
   Min     1Q Median     3Q    Max
-8.653 -5.792 -1.371  7.074 11.458

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.433      4.329   3.334  0.00874 **
x1             3.433      1.463   2.346  0.04361 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.674 on 9 degrees of freedom
Multiple R-squared:  0.3794,    Adjusted R-squared:  0.3105
F-statistic: 5.503 on 1 and 9 DF,  p-value: 0.04361
```

Atividade



cursos/planeco/roteiro/08-lm_rcmdr.txt ·
Última modificação: 2020/03/01 12:39

Modelos Lineares Simples I



Os modelos lineares são uma generalização dos testes de hipótese clássicos mais simples. Uma regressão linear, por exemplo, só pode ser aplicada para dados em que as variáveis são contínuas. Enquanto uma análise de variância é utilizada quando a variável preditora é categórica. Os modelos lineares não tem essa limitação, podemos usar variáveis contínuas ou categóricas indistintamente.

No nosso quadro de testes clássicos frequentistas, definimos os testes, baseados na natureza das variáveis respostas e preditoras.

Tipo de Variável		Estatística Clássica	
Resposta	Preditora	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência