Planejamento e Análise

modelos lineares simples II

Alexandre Adalardo de Oliveira

PlAnEco 2025

Modelos Lineares II

Modelos Lineares II

Unificação metodológica:

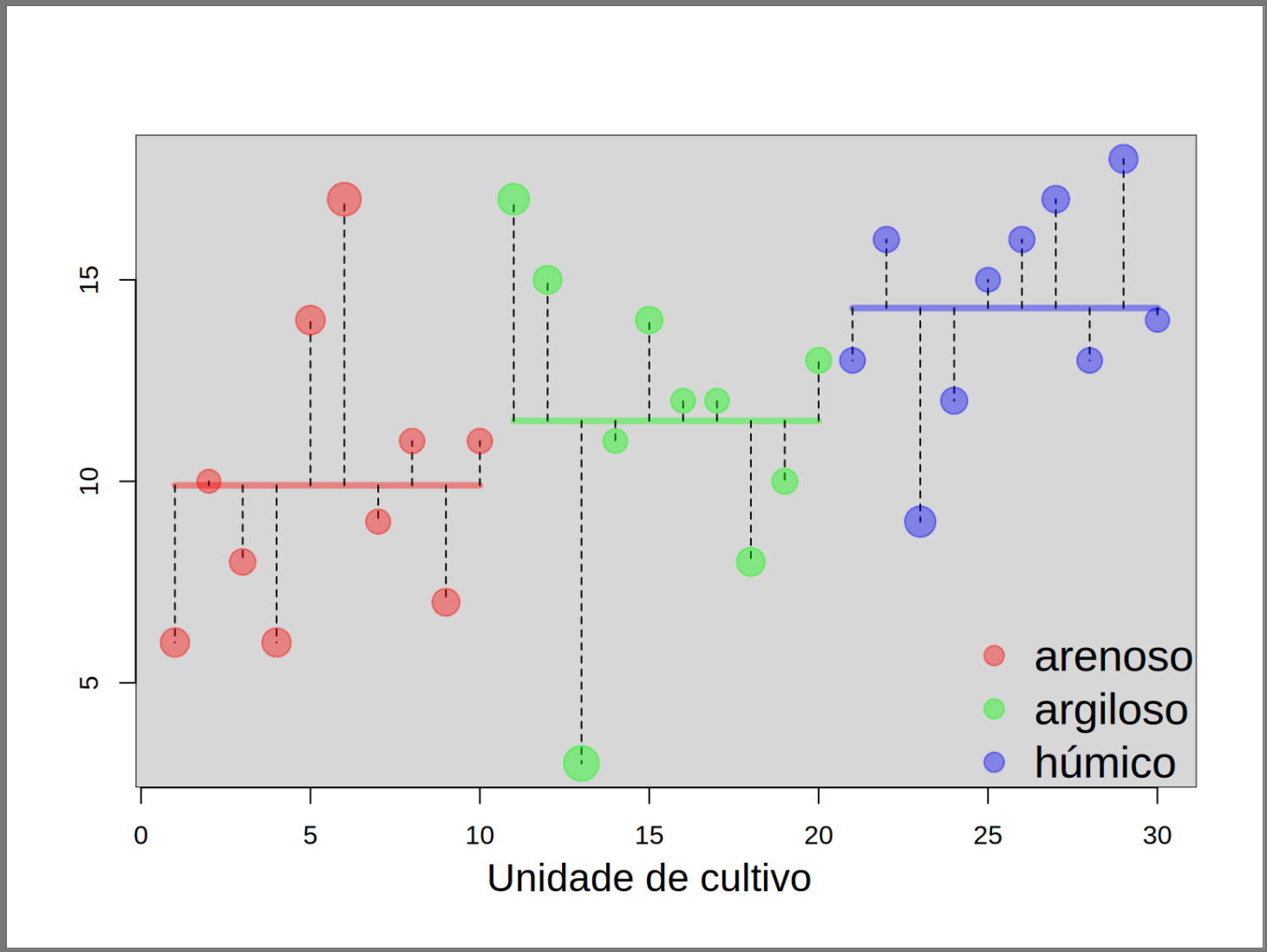
- variáveis indicadoras (Dummies)
- partição da variação dos dados
- coeficiente de determinação
- comparação com o modelo nulo

Unificação Metodológica

Resposta	Preditoras	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência
Contínua	Categórica(2)	Teste-t	$\mu_1 = \mu_2$
Contínua	Categórica (>2)	Anova	$\mu_1 = \mu_2 = \mu_3$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0; \beta_n = 0$
Contínua	Cont + Cat	Ancova	$\beta_1 = \beta_2; \alpha_1 = \alpha_2$
Proporção	Contínua	Reg. Logística	$logit(\beta_1) = 1$

Anova como Modelo Linear

Modelo linear?!



Rendimento Colheita: dados

	solo	colhe
1	are	6
2	are	10
3	are	8
4	are	6
5	are	14
11	arg	17
12	arg	15
13	arg	3
14	arg	11
15	arg	14
21	hum	13
22	hum	16
23	hum	9
24	hum	12

	solo	colhe
25	hum	15

Variável Indicadora

Indicadoras $(n_{levels} - 1)$

	colhe	solo	arg	hum
1	6	are	0	0
2	10	are	0	0
3	8	are	0	0
4	6	are	0	0
5	14	are	0	0
11	17	arg	1	0
12	15	arg	1	0
13	3	arg	1	0
14	11	arg	1	0
15	14	arg	1	0
21	13	hum	0	1
22	16	hum	0	1
23	9	hum	0	1

	colhe	solo	arg	hum
24	12	hum	0	1
25	15	hum	0	1

Modelo linear: dummy

Modelo

$$y = \alpha_{d_1} + \beta_2 x_{d_2} + \beta_3 x_{d_3}$$

Intercepto:

$$\alpha_{d_1} = \bar{x}_1$$

Coeficientes:

$$\beta_2 = \bar{x}_2 - \bar{x}_1$$

$$\beta_3 = \bar{x}_3 - \bar{x}_1$$

Residuos:

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

	colhe	arg	hum
1	6	0	0
2	10	0	0
3	8	0	0
4	6	0	0
5	14	0	0
11	17	1	0
12	15	1	0
13	3	1	0
14	11	1	0
15	14	1	0
21	13	0	1
22	16	0	1
23	9	0	1
24	12	0	1

Variável Dummy

	colhe	arg	hum
25	15	0	1

```
lmDum <- lm(colhe ~ arg + hum, data =  croplin)
summary(lmDum)
```

```

Call:
lm(formula = colhe ~ arg + hum, data = croplin)

Residuals:
   Min     1Q Median     3Q    Max
  -8.5   -1.8    0.3    1.7    7.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.900      1.081   9.158 9.04e-10 ***
arg            1.600      1.529   1.047  0.30456
hum            4.400      1.529   2.878  0.00773 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.418 on 27 degrees of freedom
Multiple R-squared:  0.2392,    Adjusted R-squared:  0.1829
F-statistic: 4.245 on 2 and 27 DF,  p-value: 0.02495
```

Variável Categórica

```r
lmCrop <- lm(colhe~solo, data = solos)
summary(lmCrop)
```

```
Call:
lm(formula = colhe ~ solo, data = solos)

Residuals:
   Min     1Q Median     3Q    Max
  -8.5   -1.8    0.3    1.7    7.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.900      1.081   9.158 9.04e-10 ***
soloarg        1.600      1.529   1.047  0.30456
solohum        4.400      1.529   2.878  0.00773 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.418 on 27 degrees of freedom
Multiple R-squared:  0.2392,    Adjusted R-squared:  0.1829
F-statistic: 4.245 on 2 and 27 DF,  p-value: 0.02495
```

Indicadora x Categórica

```
Call:
lm(formula = colhe ~ arg + hum, data = croplin)

Residuals:
   Min     1Q Median     3Q    Max
  -8.5   -1.8    0.3    1.7    7.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.900      1.081   9.158 9.04e-10 ***
arg            1.600      1.529   1.047  0.30456
hum            4.400      1.529   2.878  0.00773 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.418 on 27 degrees of freedom
Multiple R-squared:  0.2392,    Adjusted R-squared:  0.1829
F-statistic: 4.245 on 2 and 27 DF,  p-value: 0.02495
```

```
Call:
lm(formula = colhe ~ solo, data = solos)

Residuals:
   Min     1Q Median     3Q    Max
  -8.5   -1.8    0.3    1.7    7.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.900      1.081   9.158 9.04e-10 ***
soloarg        1.600      1.529   1.047  0.30456
solohum        4.400      1.529   2.878  0.00773 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.418 on 27 degrees of freedom
Multiple R-squared:  0.2392,    Adjusted R-squared:  0.1829
F-statistic: 4.245 on 2 and 27 DF,  p-value: 0.02495
```

Coeficientes do modelo

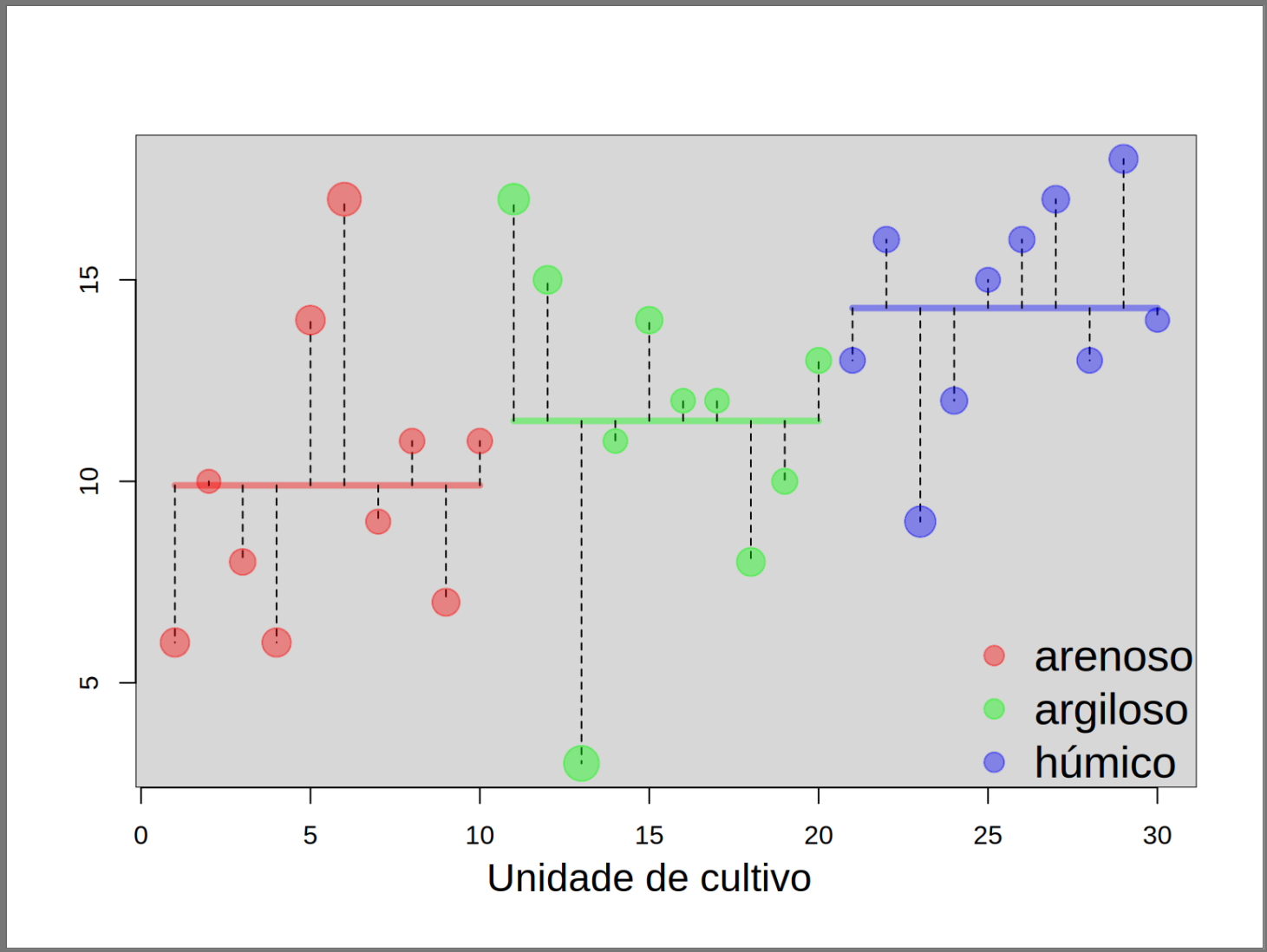
```
coef(lmDum)
```

```
(Intercept)         arg         hum
        9.9         1.6         4.4
```

```
tapply(solos$colhe, solos$solo, mean)
```

```
 are  arg  hum
 9.9 11.5 14.3
```

Modelo linear



Modelo Dummy

$$y = \hat{\alpha}_{d_1} + \hat{\beta}_2 x_{d_2} + \hat{\beta}_3 x_{d_3}$$

	colhe	solo	arg	hum	pred
1	6	are	0	0	9.9
2	10	are	0	0	9.9
3	8	are	0	0	9.9
11	17	arg	1	0	11.5
12	15	arg	1	0	11.5
13	3	arg	1	0	11.5
21	13	hum	0	1	14.3
22	16	hum	0	1	14.3
23	9	hum	0	1	14.3

```
(Intercept)         arg         hum
        9.9         1.6         4.4
```

Modelo

$y = \alpha_{d_1} + \beta_2 x_{d_2} + \beta_3 x_{d_3}$

Intercepto:

$\alpha_{d_1} = \bar{x}_1$

Coeficientes:

$\beta_2 = \bar{x}_2 - \bar{x}_1$

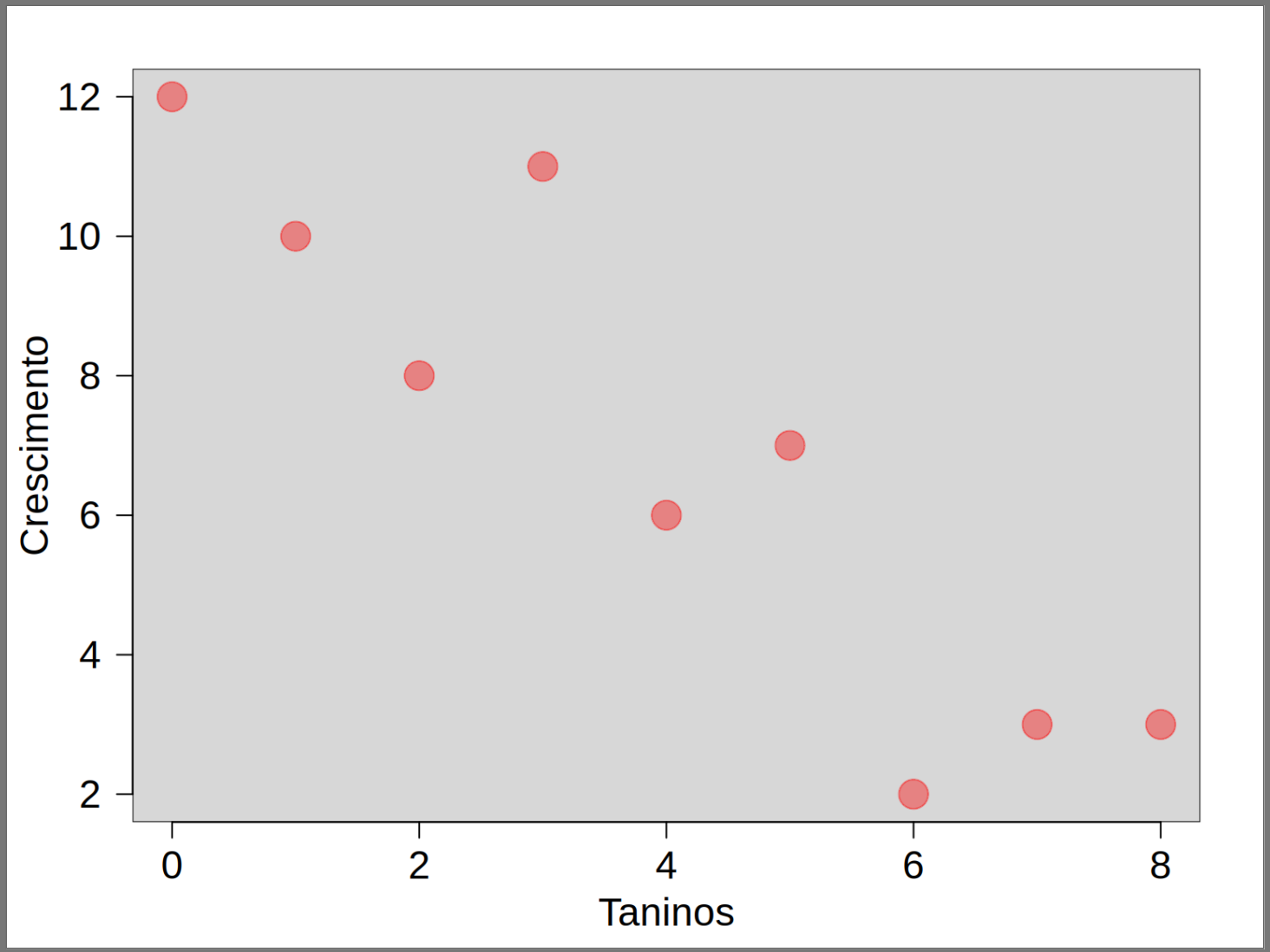
$\beta_3 = \bar{x}_3 - \bar{x}_1$

Residuos:

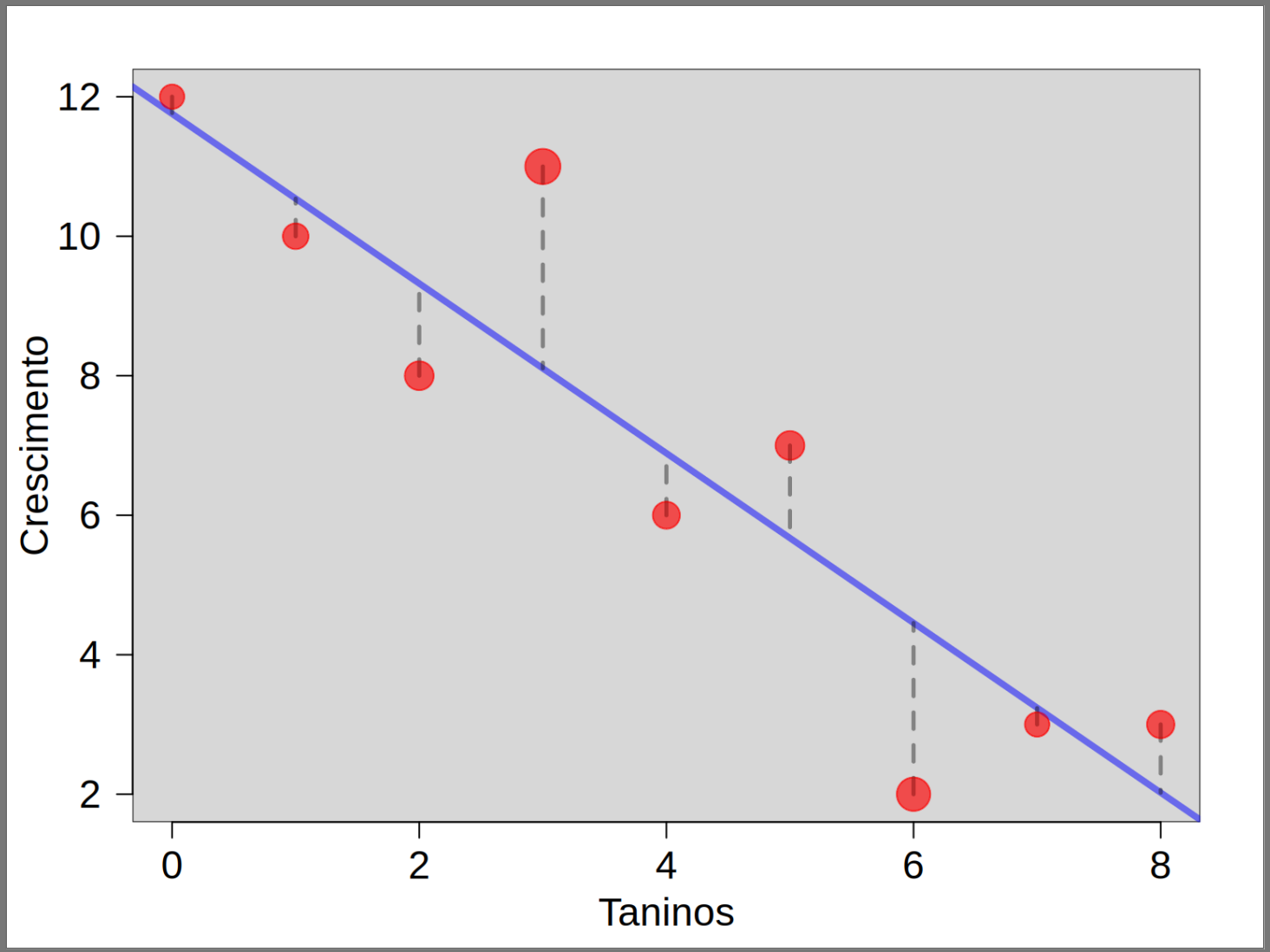
$\epsilon = N(0, \sigma)$

Dieta de Lagarta

Exemplo: dieta de lagarta



Modelo Linear: lagartas

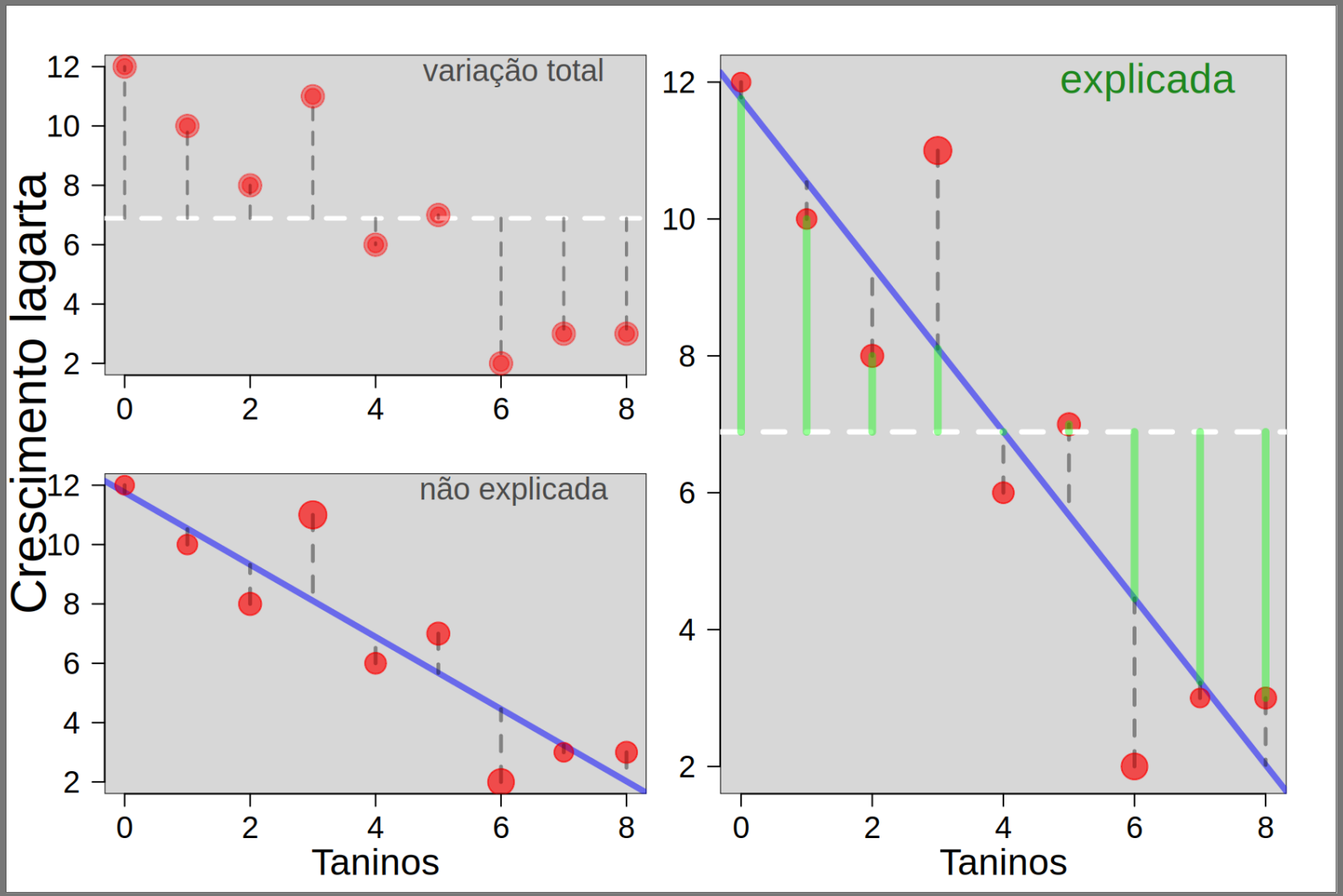


lm(growth~tannin)

```
summary(lmlag)
```

```
Call:
lm(formula = growth ~ tannin, data = lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4556 -0.8889 -0.2389  0.9778  2.8944 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  11.7556     1.0408  11.295 9.54e-06 ***
tannin       -1.2167     0.2186  -5.565 0.000846 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.693 on 7 degrees of freedom
Multiple R-squared:  0.8157,    Adjusted R-squared:  0.7893 
F-statistic: 30.97 on 1 and 7 DF,  p-value: 0.0008461
```

Partição da Variância



Partição da Variância

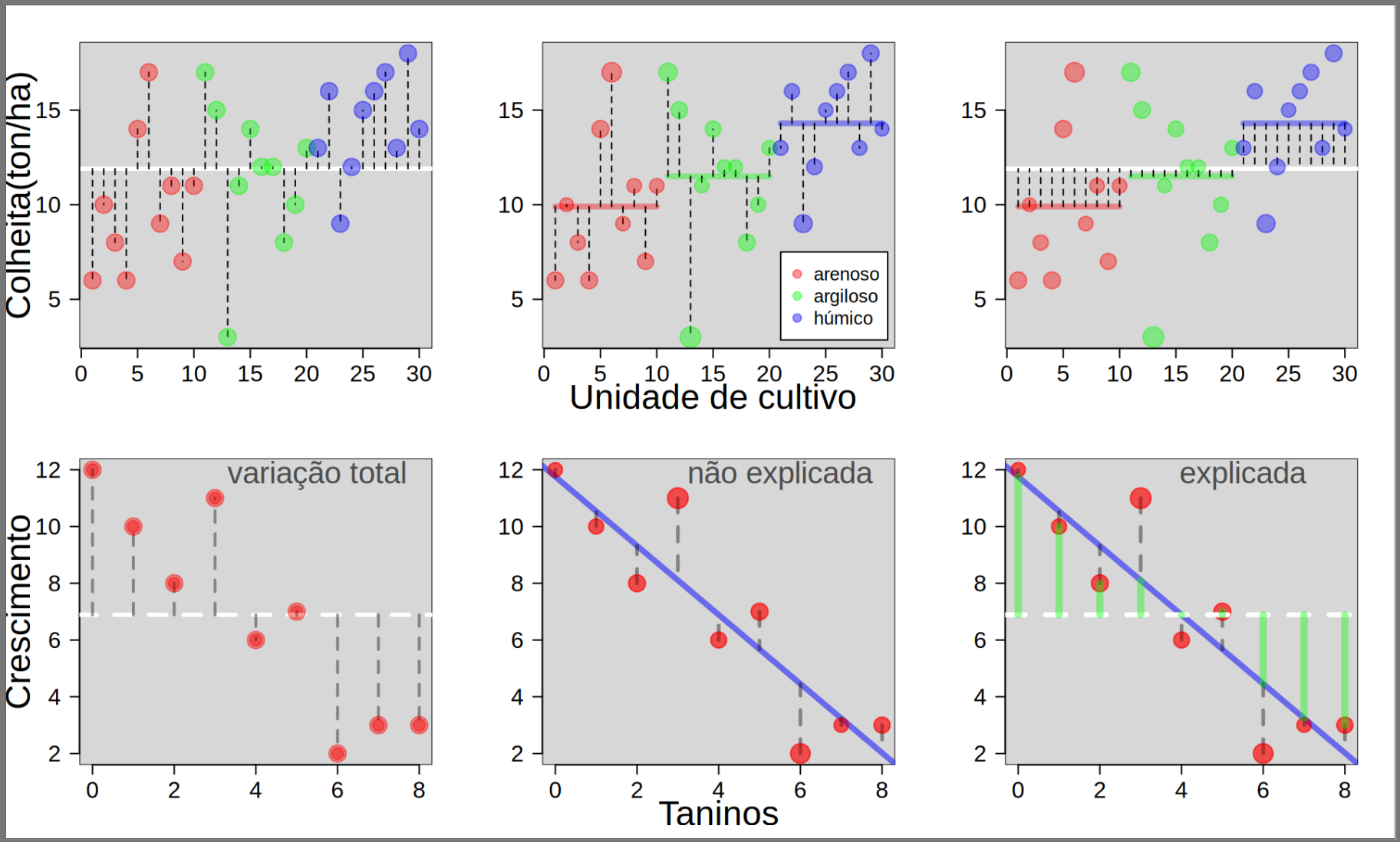
Anova

$$SQ_{total} = SQ_{entre} + SQ_{intra}$$

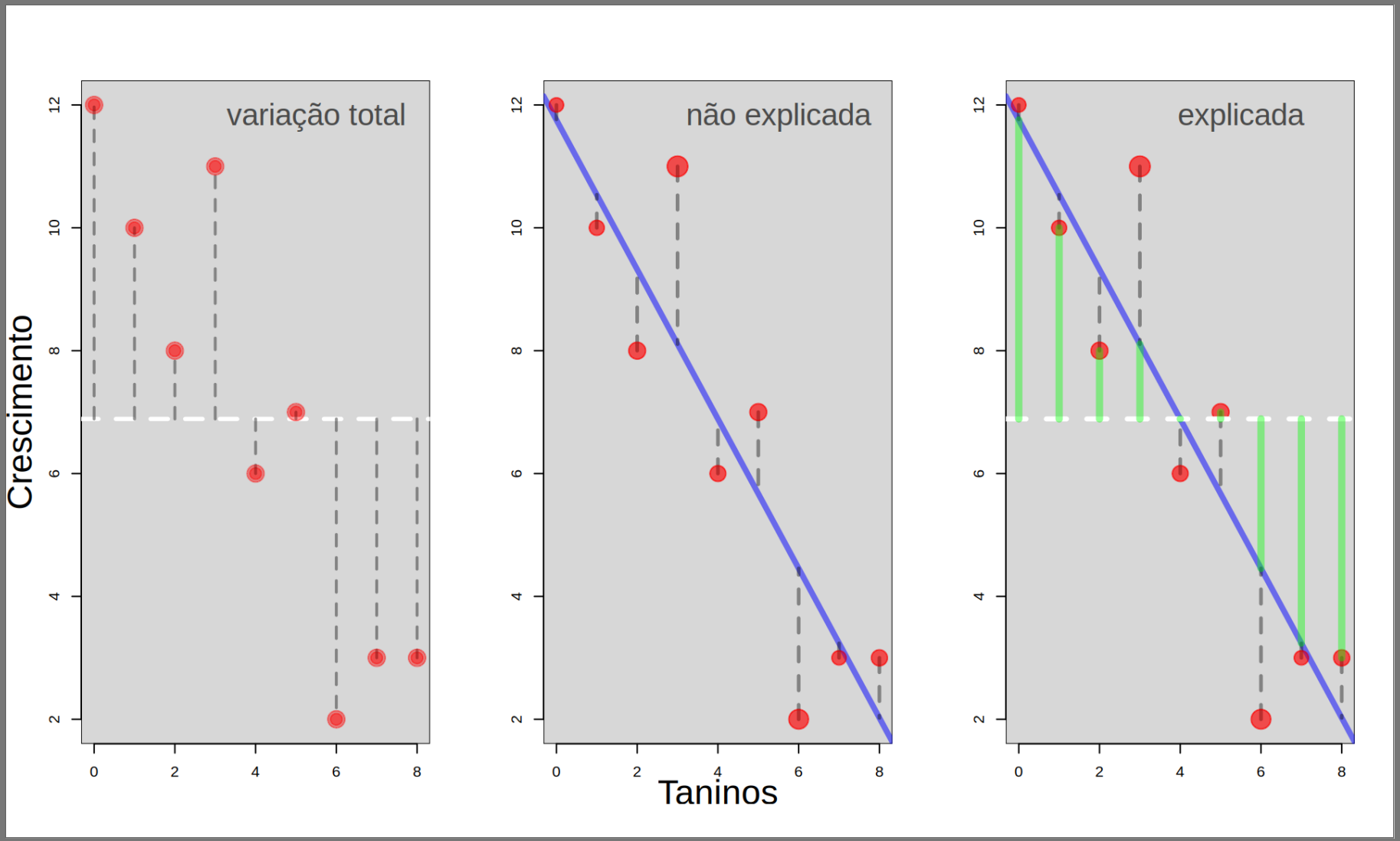
Modelo Linear

$$SQ_{total} = SQ_{mod} + SQ_{res}$$

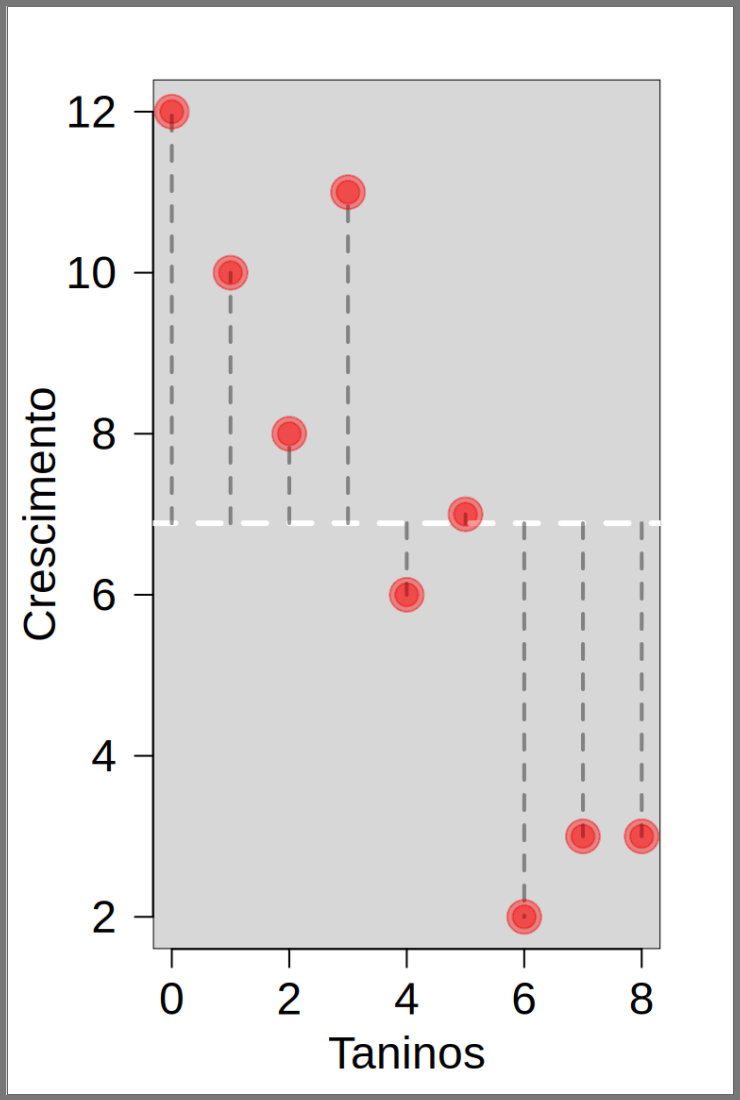
Anova x Modelo Linear



Partição da Variância: lm



Desvios Quadráticos: total

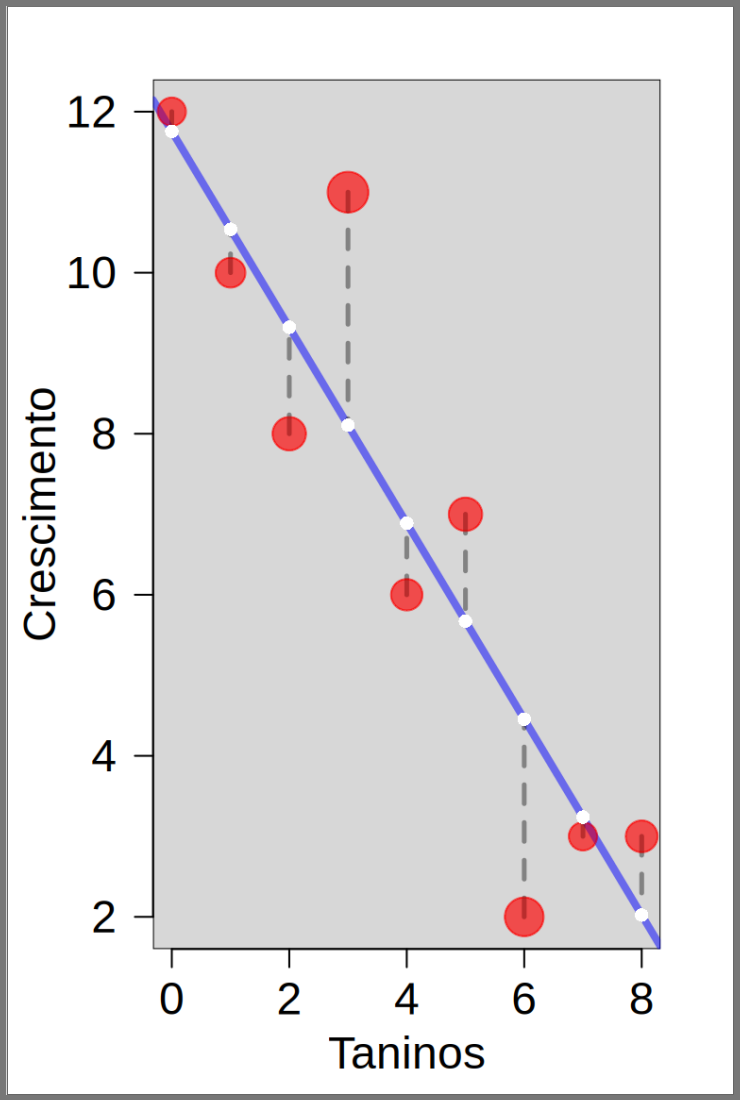


$$SQ_{total} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

tannin	growth	mediag	devTotal	dqT
0	12	6.89	5.11	26.12
1	10	6.89	3.11	9.68
2	8	6.89	1.11	1.23
3	11	6.89	4.11	16.90
4	6	6.89	−0.89	0.79
5	7	6.89	0.11	0.01
6	2	6.89	−4.89	23.90
7	3	6.89	−3.89	15.12
8	3	6.89	−3.89	15.12

```
[1] 108.89
```

Desvios Quadráticos: resíduo

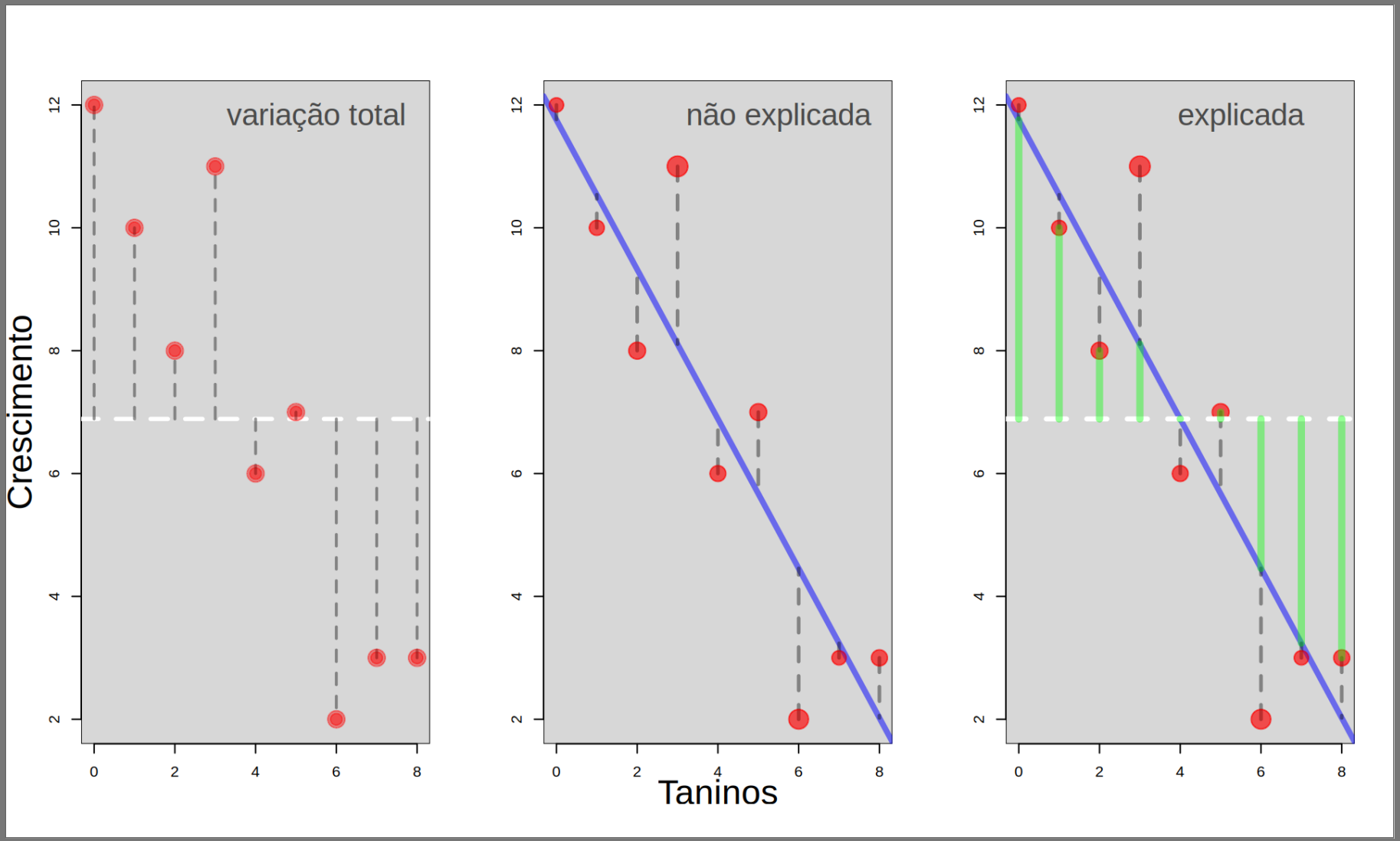


$$SQ_{error} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

tannin	growth	preditos	residuos	dq
0	12	11.76	0.24	0.06
1	10	10.54	−0.54	0.29
2	8	9.32	−1.32	1.75
3	11	8.11	2.89	8.38
4	6	6.89	−0.89	0.79
5	7	5.67	1.33	1.76
6	2	4.46	−2.46	6.03
7	3	3.24	−0.24	0.06
8	3	2.02	0.98	0.96

```
[1] 20.07
```

Desvios Quadráticos: modelo



Lógica do Modelo Linear

$$SQ_{total} = SQ_{mod} + SQ_{res}$$
$$SQ_{mod} = SQ_{total} - SQ_{res}$$

Desvios Quadraticos do Modelo

```
[1] 88.81667
```

Tabela de Anova

Fonte	SQ	GL	MQ
Modelo	88.82	1	88.82
Residuo	20.07	7	2.87
Total	108.89	8	

Modelo Linear

Teste de hipótese: F

Fonte	SQ	GL	MQ	F	pvalor
Modelo	88.82	1	88.82	30.97	0.00085
Residuo	20.07	7	2.87		
Total	108.89	8			

Modelo Linear: lagarta

Tabela de Anova no R

```
anova(lmlag)
```

```
Analysis of Variance Table

Response: growth
          Df Sum Sq Mean Sq F value    Pr(>F)
tannin     1 88.817  88.817  30.974 0.0008461 ***
Residuals  7 20.072   2.867
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fonte	SQ	GL	MQ	F	pvalor
Modelo	88.82	1	88.82	30.97	0.00085
Residuo	20.07	7	2.87		
Total	108.89	8			

Coeficiente de Determinação

$$R^2 = \frac{SQ_{mod}}{SQ_{total}}$$

Fonte	SQ	GL	MQ	F	pvalor
Modelo	88.82	1	88.82	30.97	0.00085
Residuo	20.07	7	2.87		
Total	108.89	8			

R^2

```
[1] 0.816
```

R^2 Ajustado

$$R^2_{adj} = 1 - (1 - R^2) * \frac{n - 1}{n - 2}$$

R^2_{adj}

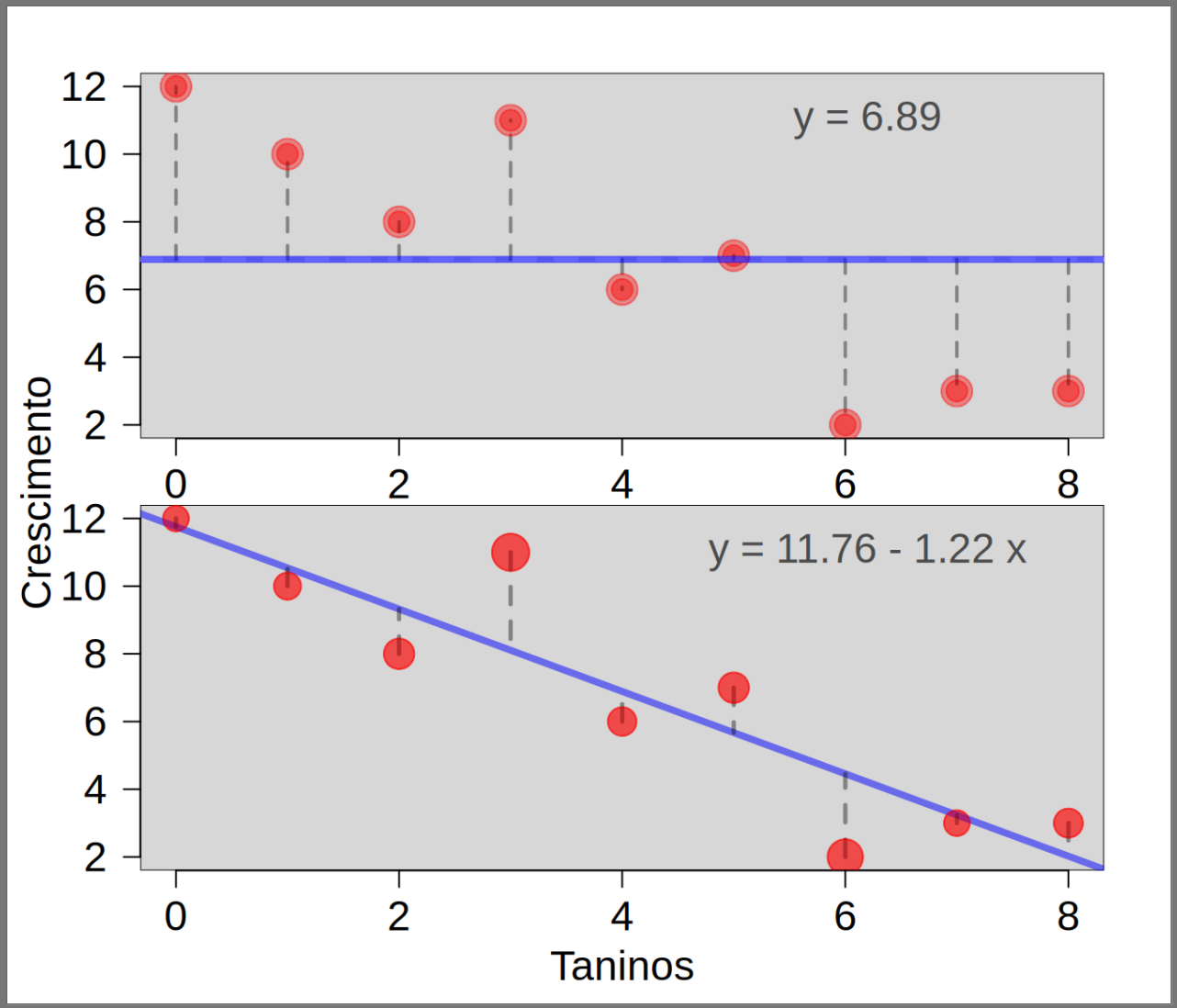
```
[1] 0.7897143
```

Resumo do Modelo

```
summary(lmlag)
```

```
Call:
lm(formula = growth ~ tannin, data = lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4556 -0.8889 -0.2389  0.9778  2.8944 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  11.7556     1.0408  11.295 9.54e-06 ***
tannin       -1.2167     0.2186  -5.565 0.000846 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.693 on 7 degrees of freedom
Multiple R-squared:  0.8157,    Adjusted R-squared:  0.7893 
F-statistic: 30.97 on 1 and 7 DF,  p-value: 0.0008461
```

Comparando Modelos

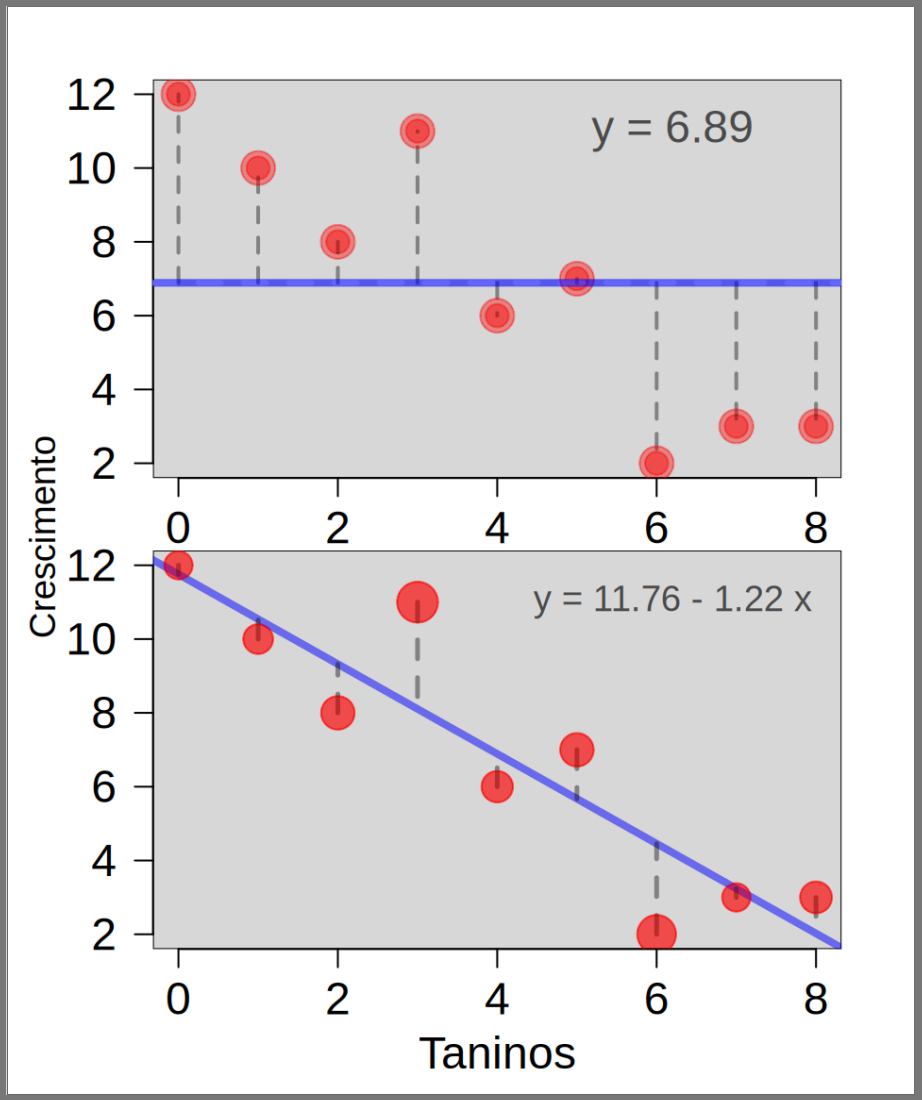


Modelo Mínimo (nulo)

```r
nullag <- lm(growth ~ 1, data = lag)
summary(nullag)
```

```
Call:
lm(formula = growth ~ 1, data = lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.8889 -3.8889  0.1111  3.1111  5.1111 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    6.889      1.230   5.602 0.000509 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.689 on 8 degrees of freedom
```

Comparando Modelos




```
anova(nullag, lmlag)
```

```
Analysis of Variance Table

Model 1: growth ~ 1
Model 2: growth ~ tannin
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1      8 108.889
2      7  20.072  1    88.817 30.974 0.0008461 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
```

```
anova(lmlag)
```

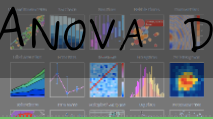
```
Analysis of Variance Table

Response: growth
          Df Sum Sq Mean Sq F value    Pr(>F)
tannin     1 88.817  88.817  30.974 0.0008461 ***
Residuals  7 20.072   2.867
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
```

Comparando Modelos

ANOVA DO MODELO: ANOVA(LMLAG)

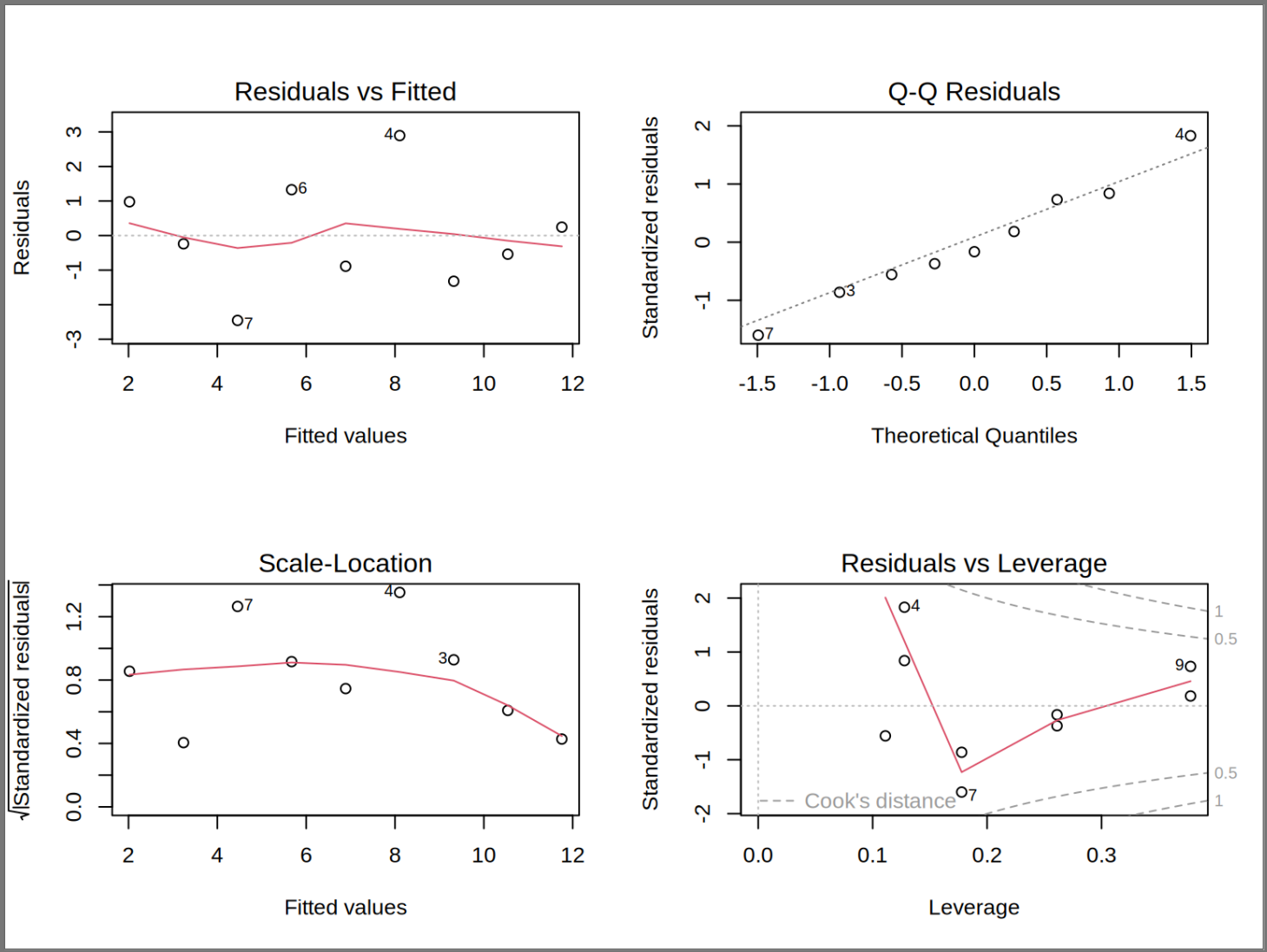
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tannin	1	88.81667	88.81667	30.97398	0.0008461
Residuals	7	20.07222	2.86746		

ANOVA DA COMPARAÇÃO DE MODELOS: ANOVA(NULLAG, LMLAG)



Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
8	108.88889				
7	20.07222	1	88.81667	30.97398	0.0008461

Diagnóstico do Modelos



Atividade