Regressão linear simples e correlação

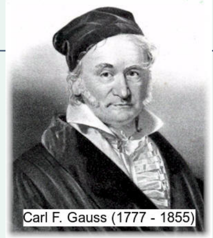
Fabio Cop

Instituto de Ciências do Mar - UNIFESP

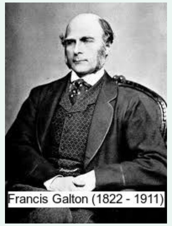
05 junho, 2021

Um pouco de história

- A primeira solução para o problema da regressão (relacionar uma variável resposta Y a uma variável preditora X) foi o **Método dos Mínimos Quadrados (MMQ)**, publicado por por Gauss (1777 – 1855) em 1809, embora haja relatos históricos de que Gauss pensou e resolveu o problema quando tinha apenas 11 anos. Gauss aplicou o método obter predições sobre as órbitas dos corpos ao redor do Sol a partir de observações astronômicas.


Carl F. Gauss (1777 - 1855)

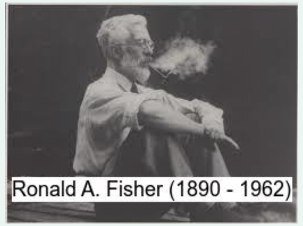
- O termo **regressão** foi empregado por Francis Galton em 1866, um dos pais da Biometria e primo de *Charles Darwin*, no séc. XIX, para descrever o fenômeno biológico onde a altura dos descendentes de pais altos tende a regressar em direção à média. Desta forma, pais muito altos tenderiam a ter descendentes mais baixos que eles próprios e vice versa. A altura dos descendentes tenderia portanto a **regressar** à média da população.


Francis Galton (1822 - 1911)

- Para Galton a regressão tinha apenas um significado biológico, mas suas idéias foram estendidas por Udny Yule e Karl Pearson para um contexto estatístico mais geral. Na formulação de Yule e Pearson, assume-se que a distribuição conjunta da variável resposta e da variável preditora $f(Y, X)$ é Gaussiana (Normal), o que confunde os conceitos de regressão e correlação.


Karl Pearson (1857 - 1936)

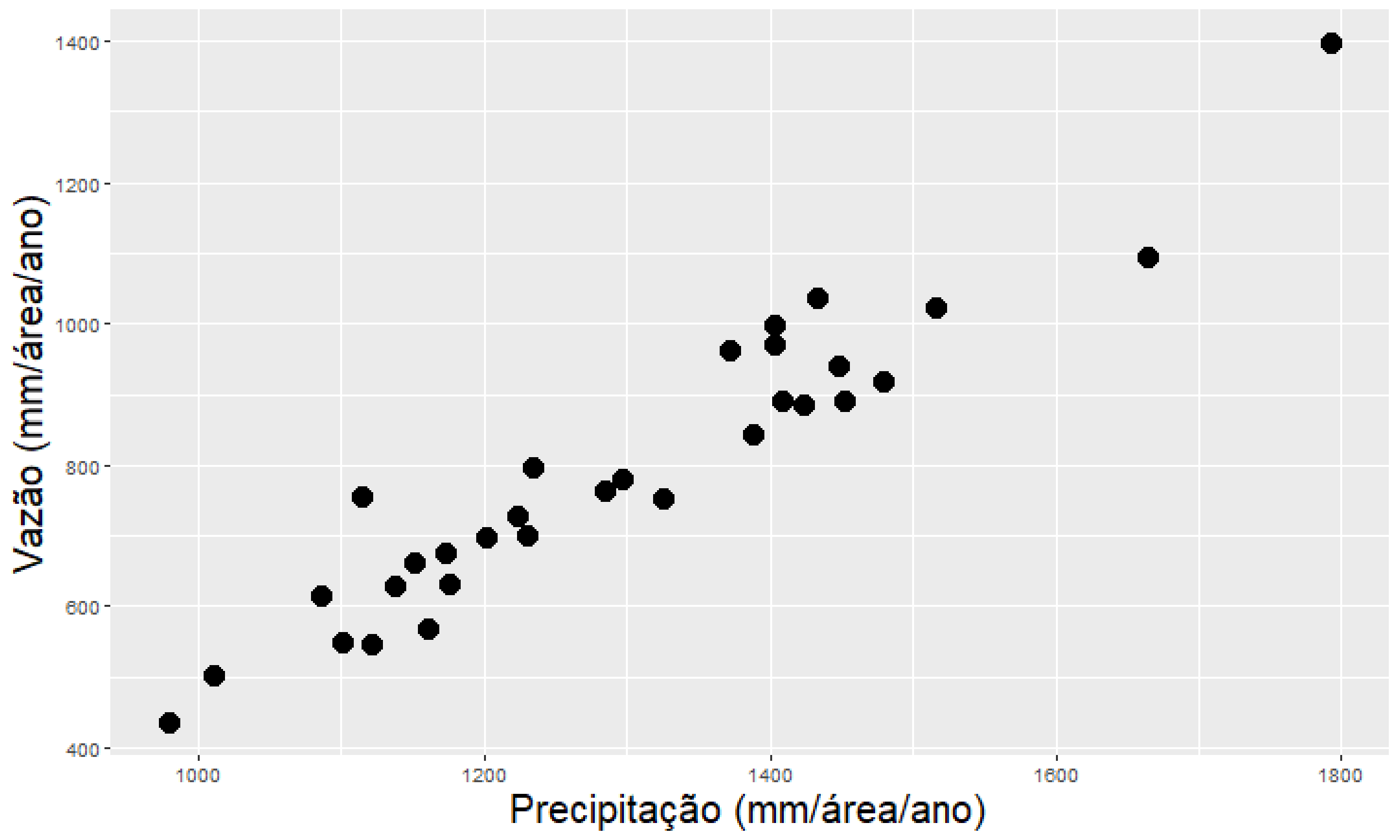
- Esta suposição foi modificada por R. A. Fisher em 1922 e 1925. Fisher assumiu que a distribuição condicional da variável resposta $f(Y|X)$ é Gaussiana, mas a conjunta não precisa ser. Esta solução é mais próxima daquela formulada por Gauss. Fisher desenvolveu também o método da **Máxima Verossimilhança (MV)**. Para uma variável em que $f(Y|X)$ é Gaussiana, a solução pelo **MMQ** e pela **MV** convergem.


Ronald A. Fisher (1890 - 1962)

Conteúdo da Aula

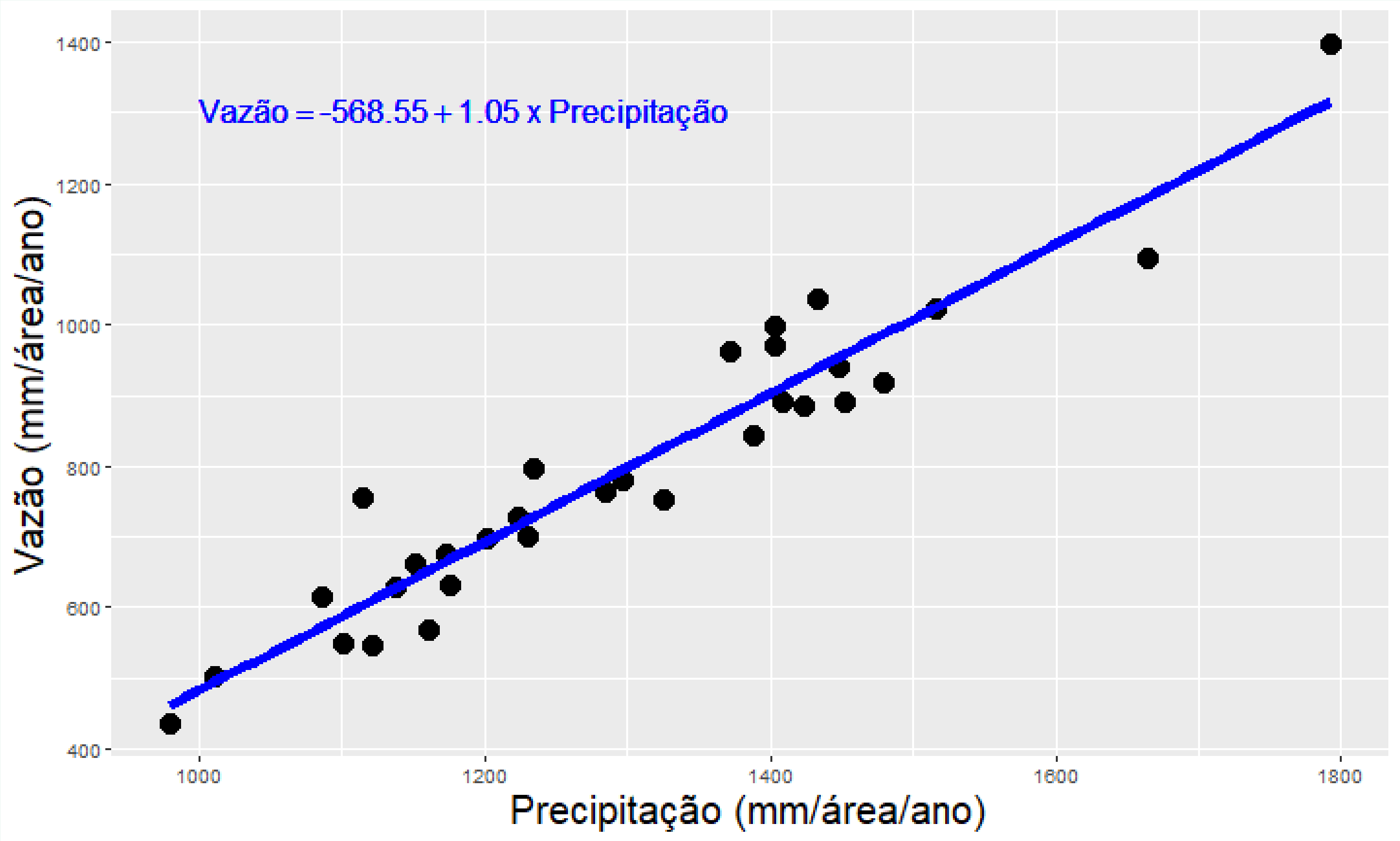
1. Descrevendo relações funcionais

O Serviço Florestal americano estabeleceu a Floresta Experimental de Hubbard Brook (HBEF) em 1955 como um centro de pesquisa hidrológica. Um serviço ecossistêmico óbvio das bacias hidrográficas nesta região é o fornecimento hídrico. Podemos supor que o volume de água anual que uma bacia pode fornecer tem relação com o volume de chuva.



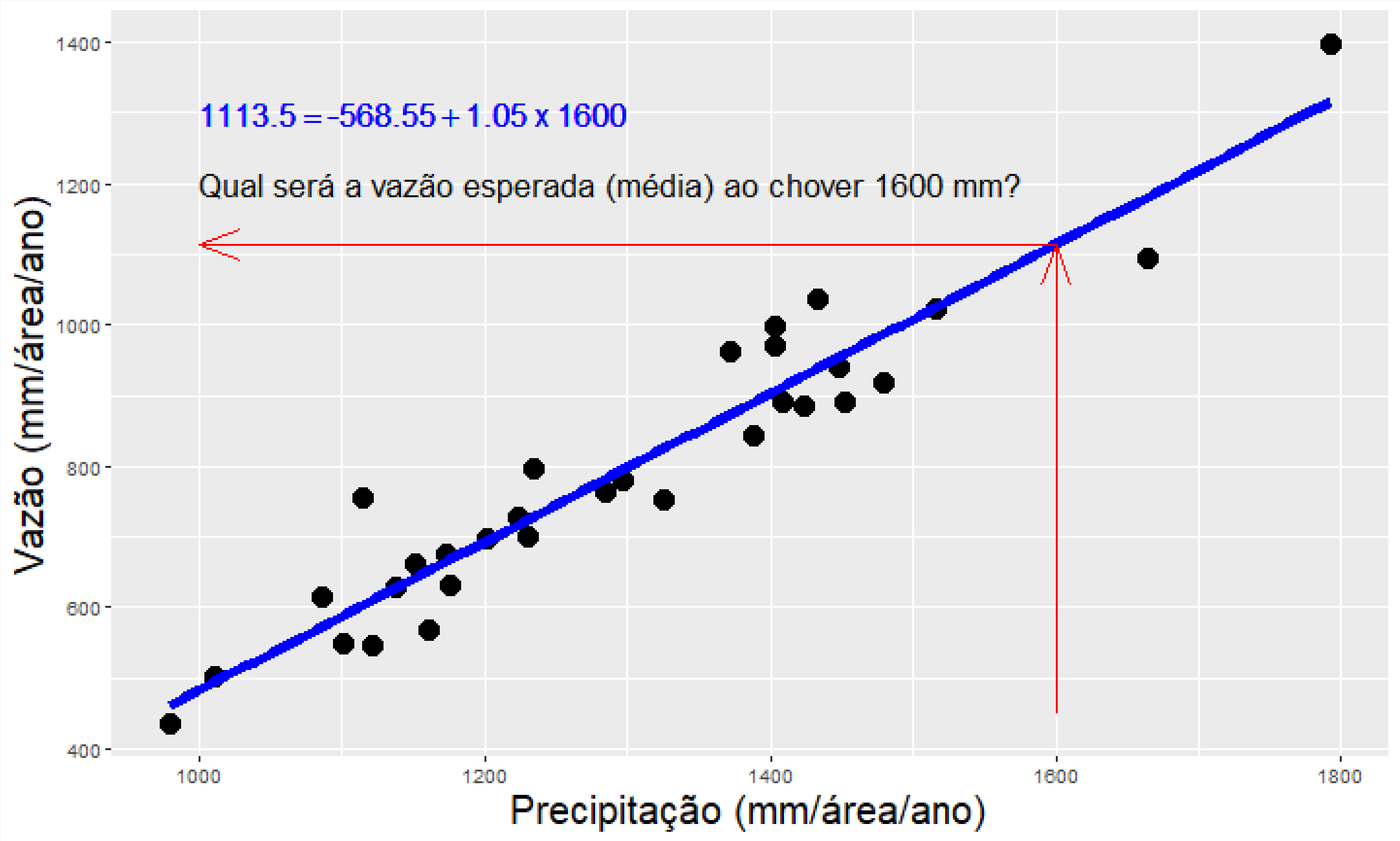
1. Descrevendo relações funcionais

O Serviço Florestal americano estabeleceu a Floresta Experimental de **Hubbard Brook (HBEF)** em 1955 como um centro de pesquisa hidrológica. Um serviço ecossistêmico óbvio das bacias hidrográficas nesta região é o fornecimento hídrico. Podemos supor que o volume de água anual que uma bacia pode fornecer tem relação com o volume de chuva.



1. Predições sobre fenômenos ambientais

O Serviço Florestal americano estabeleceu a Floresta Experimental de Hubbard Brook (HBEF) em 1955 como um centro de pesquisa hidrológica. Um serviço ecossistêmico óbvio das bacias hidrográficas nesta região é o fornecimento hídrico. Podemos supor que o volume de água anual que uma bacia pode fornecer tem relação com o volume de chuva.



1. Predições sobre fenômenos ambientais

Taxa de fotossíntese em folhas do mangue-vermelho (*Rhizophora mangle*)

$$Y = \frac{k \times X}{D + X}$$

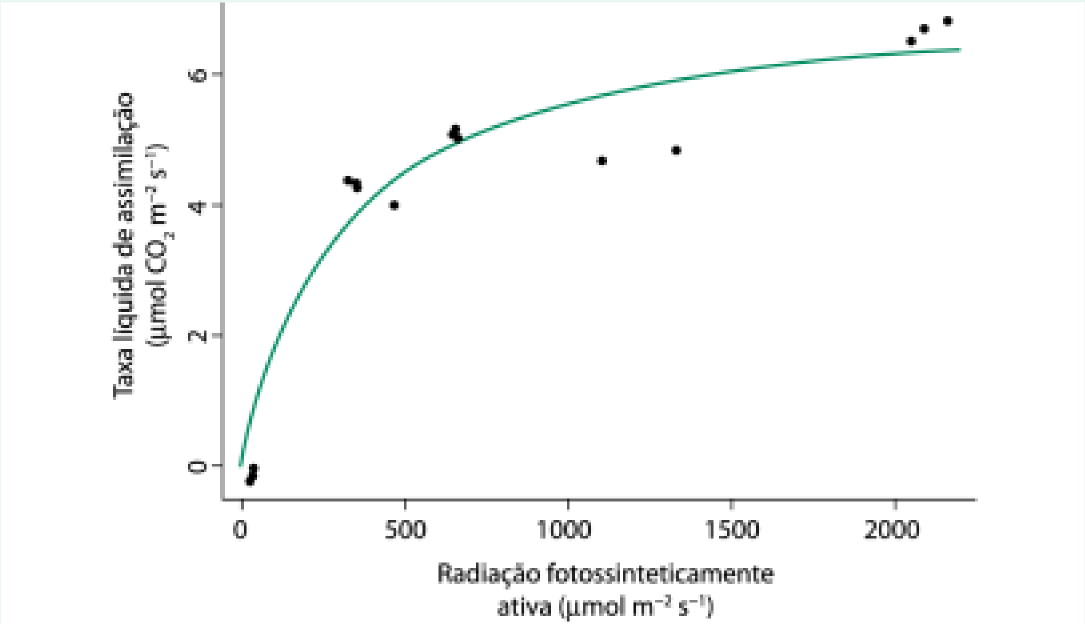
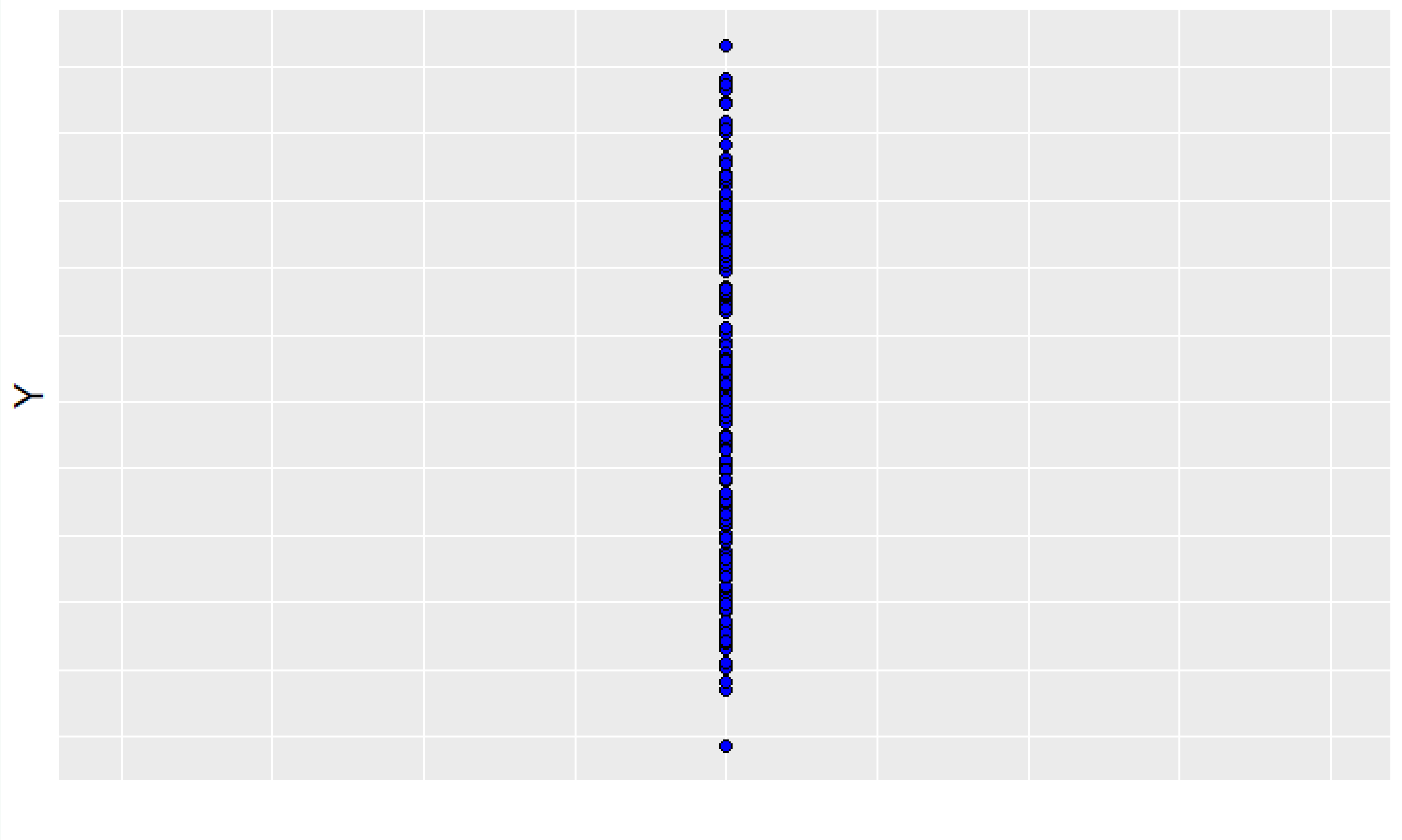


Figura 4.3 Relação entre a intensidade luminosa e a taxa fotossintética. Os dados são medidas da taxa de assimilação líquida e da radiação fotossinteticamente ativa para $n = 15$ folhas jovens da planta do mangue-vermelho *Rhizophora mangle* em Belize (Farnsworth e Ellison, 1996b). A equação de Michaelis-Menten com a forma $Y = kX/(D + X)$ foi ajustada aos dados. As estimativas dos parâmetros ±1 desvio-padrão são $k = 7{,}3 \pm 0{,}59$ e $D = 313 \pm 86{,}6$.

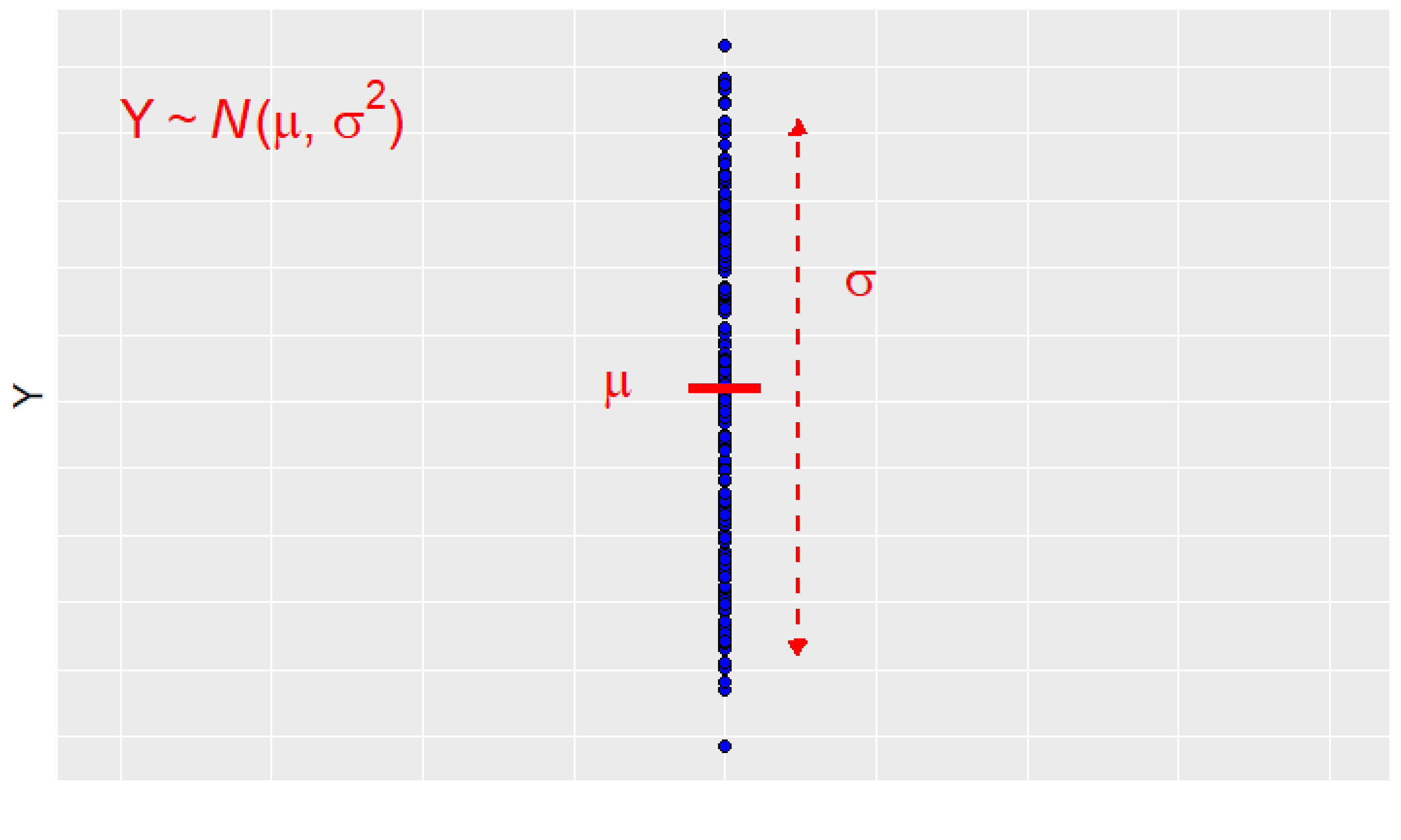
1. Estrutura geral do modelo de regressão

Seja uma variável aleatória Y com distribuição normal proveniente de um *experimento aleatório*.



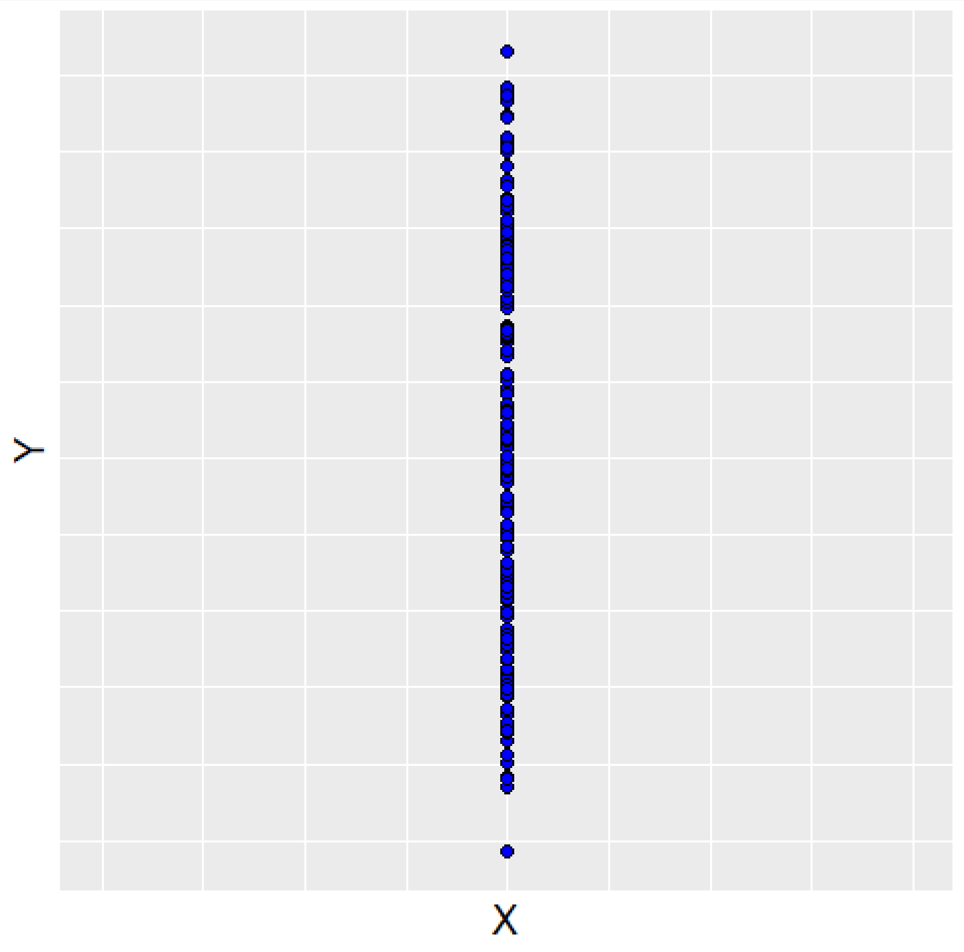
1. Estrutura geral do modelo de regressão

Seja uma variável aleatória Y com distribuição normal proveniente de um *experimento aleatório*.



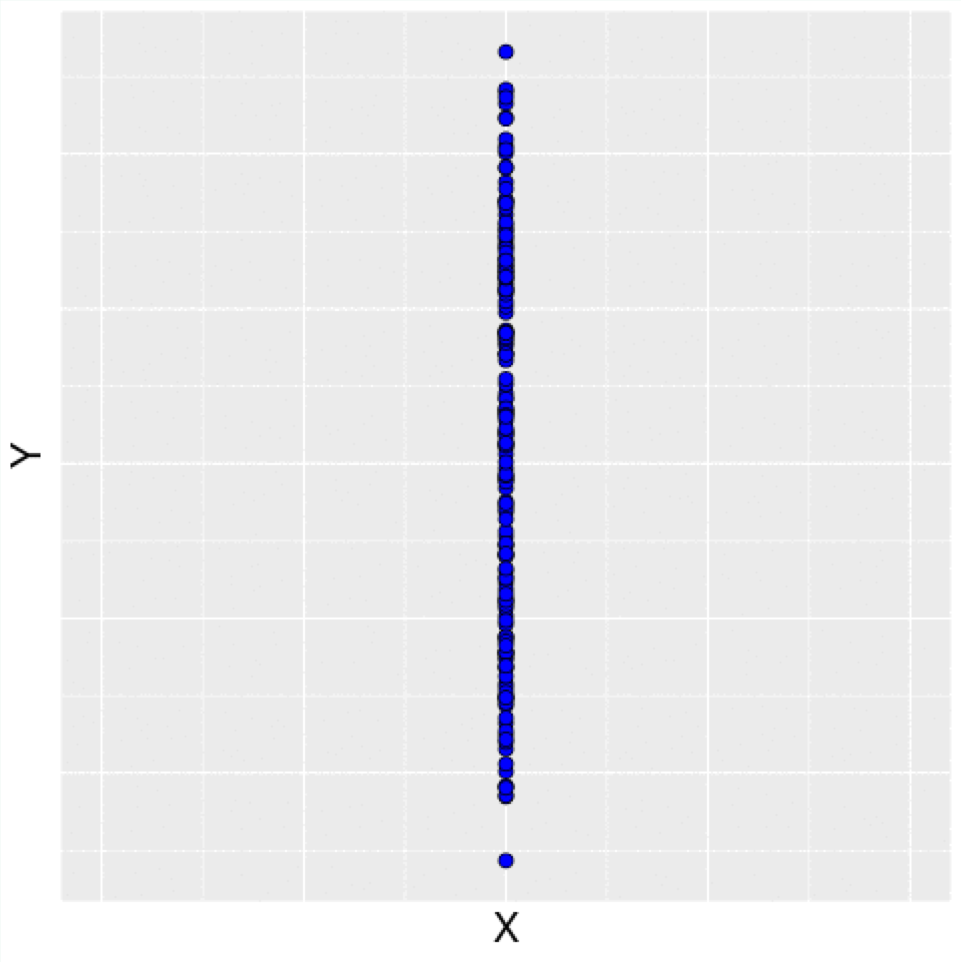
1. Estrutura geral do modelo de regressão

Para cada observação y_i é conhecida também uma informação sobre x_i.



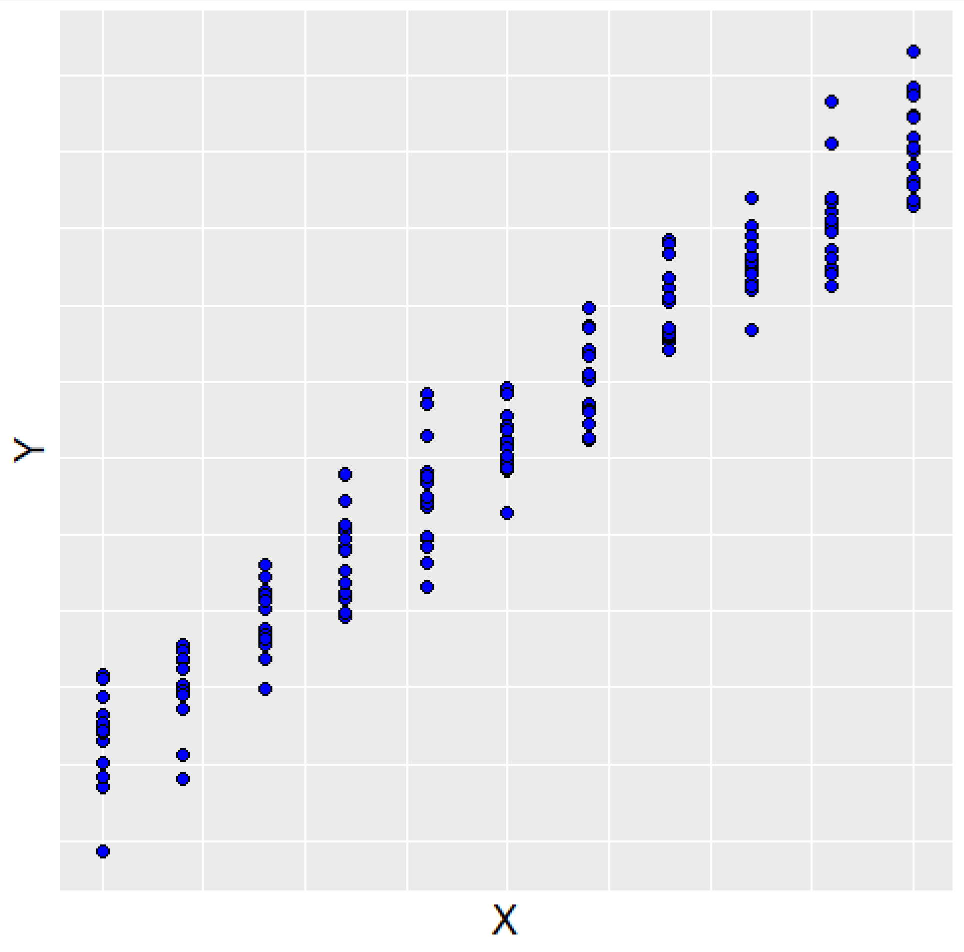
1. Estrutura geral do modelo de regressão

Para cada observação y_i é conhecida também uma informação sobre x_i.



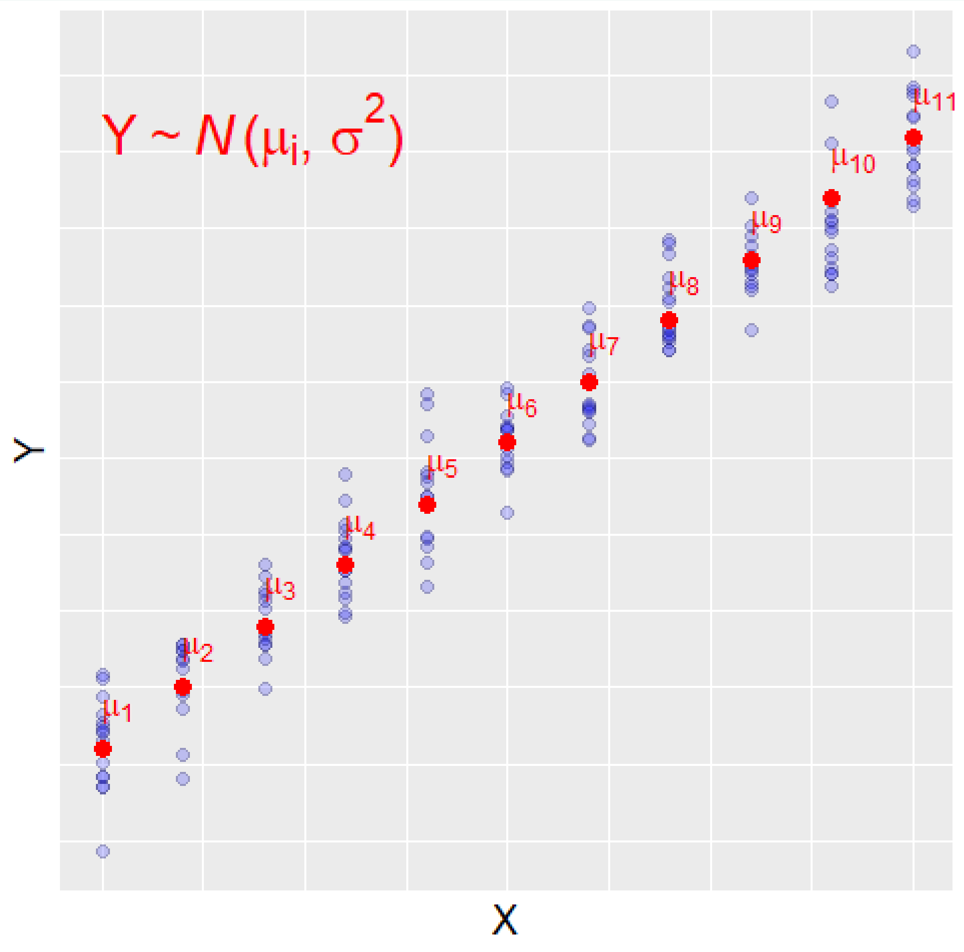
1. Estrutura geral do modelo de regressão

Para cada observação y_i é conhecida também uma informação sobre x_i.



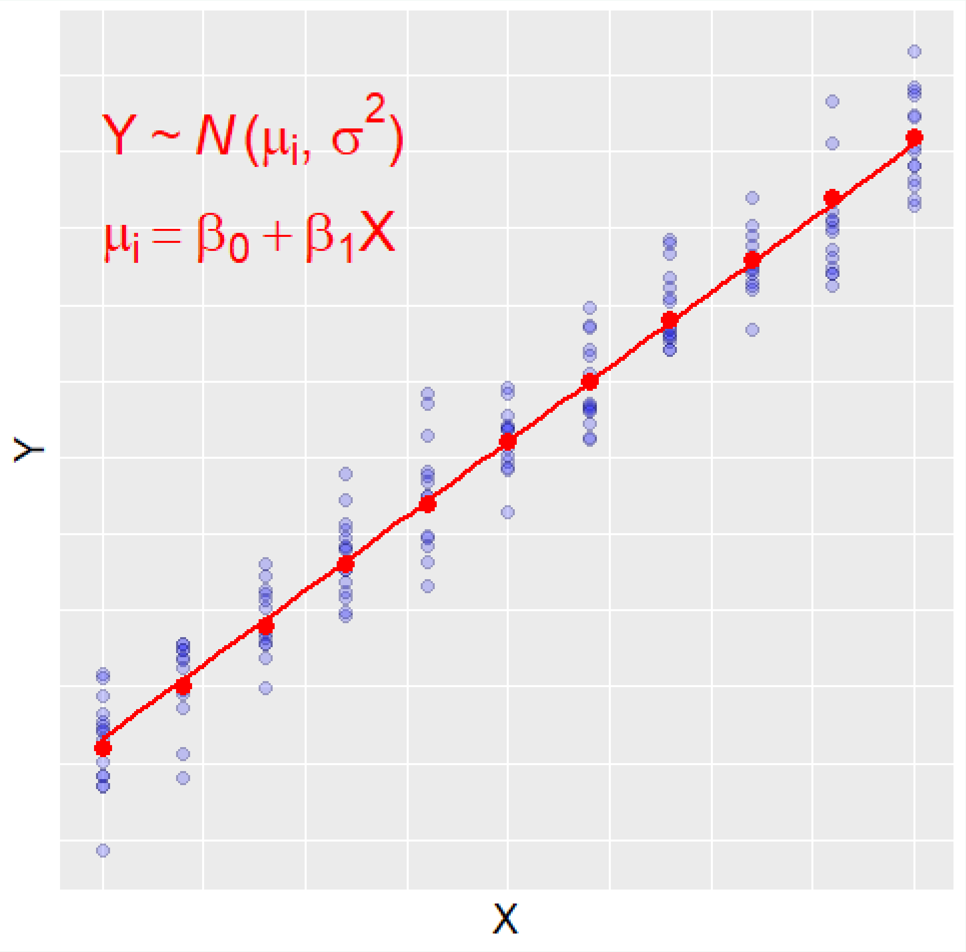
1. Estrutura geral do modelo de regressão

Para cada observação y_i é conhecida também uma informação sobre x_i.



1. Estrutura geral do modelo de regressão

Para cada observação y_i é conhecida também uma informação sobre x_i.



1. Estrutura geral do modelo de regressão

1 - As observações em Y e X compõem um par (y_i, x_i) de modo que:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

2 - X é determinada **experimentalmente** e **sem erros**.

3 - Y é uma variável aleatória normalmente distribuída, com μ_i variância σ^2.

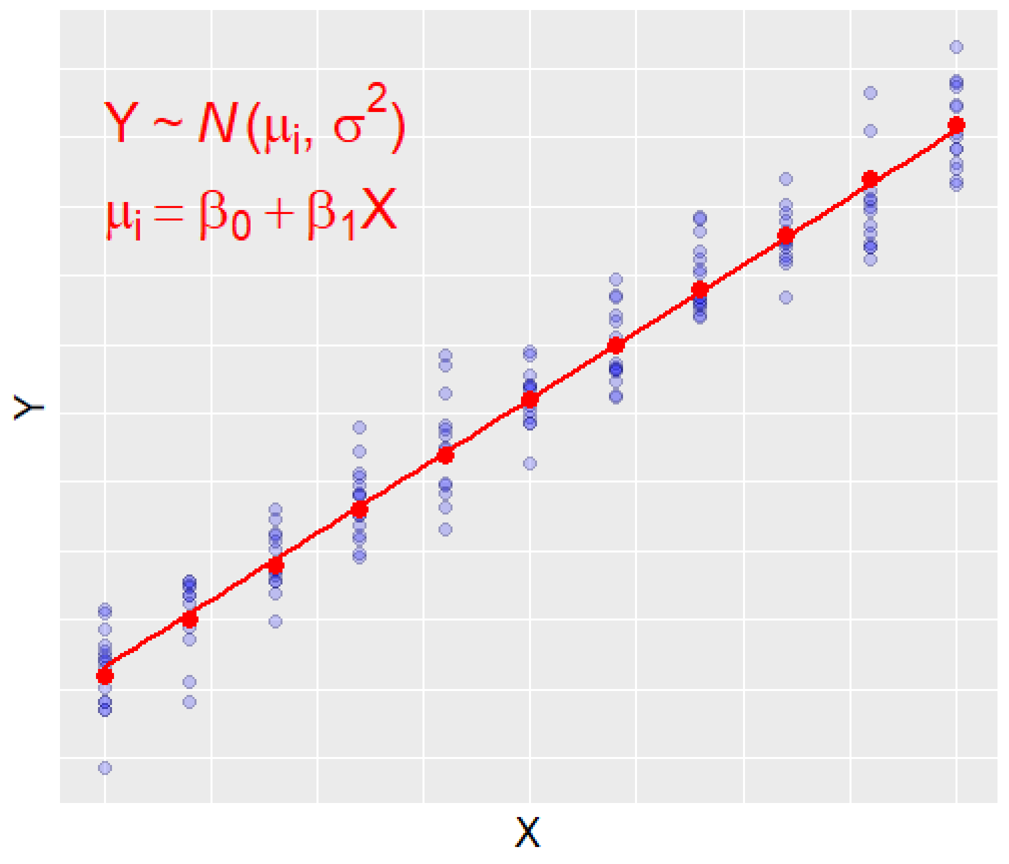
$$Y \sim \mathcal{N}(\mu_i, \sigma^2)$$

4 - μ_i é representado por um **modelo linear** que expressa o valor esperado de y_i para um dado valor de x_i. Compõe a **parcela determinística** do modelo.

$$E(Y|x_i) = \mu_i = \beta_0 + \beta_1 x_i$$

5 - β_0 e β_1 são as contantes a serem estimadas, representando o **intercepto** e o **coeficience de inclinação da reta**, repectivamente.

6 - σ^2 é a **variância** de Y e ser estimada. σ^2 é **constante** para todos os valores em X.



2. O modelo matemático

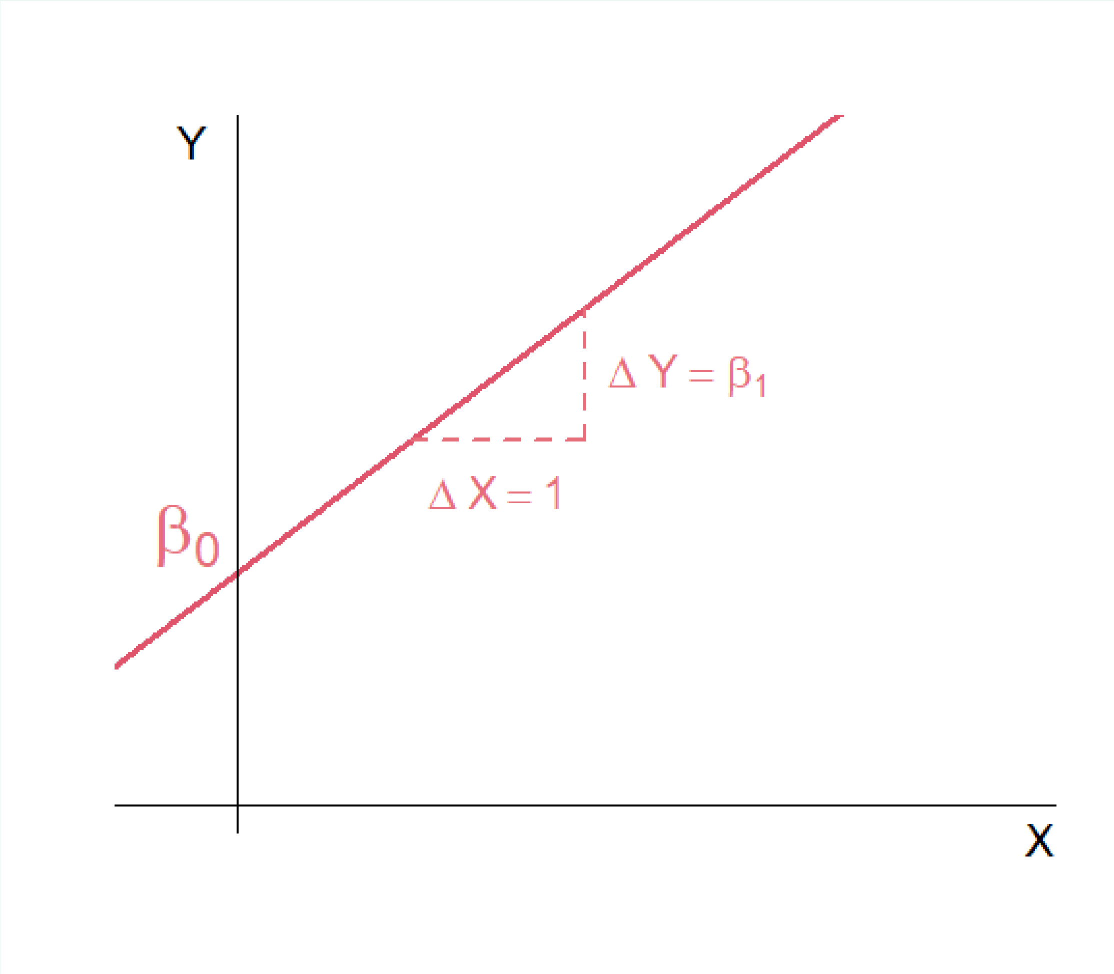
1. Y: variável resposta (dependente);
2. X: variável preditora (**in**dependente);

$$E(Y|x_i) = \beta_0 + \beta_1 x_i$$

1. Parâmetros do mdelo

β_0: Intercepto;

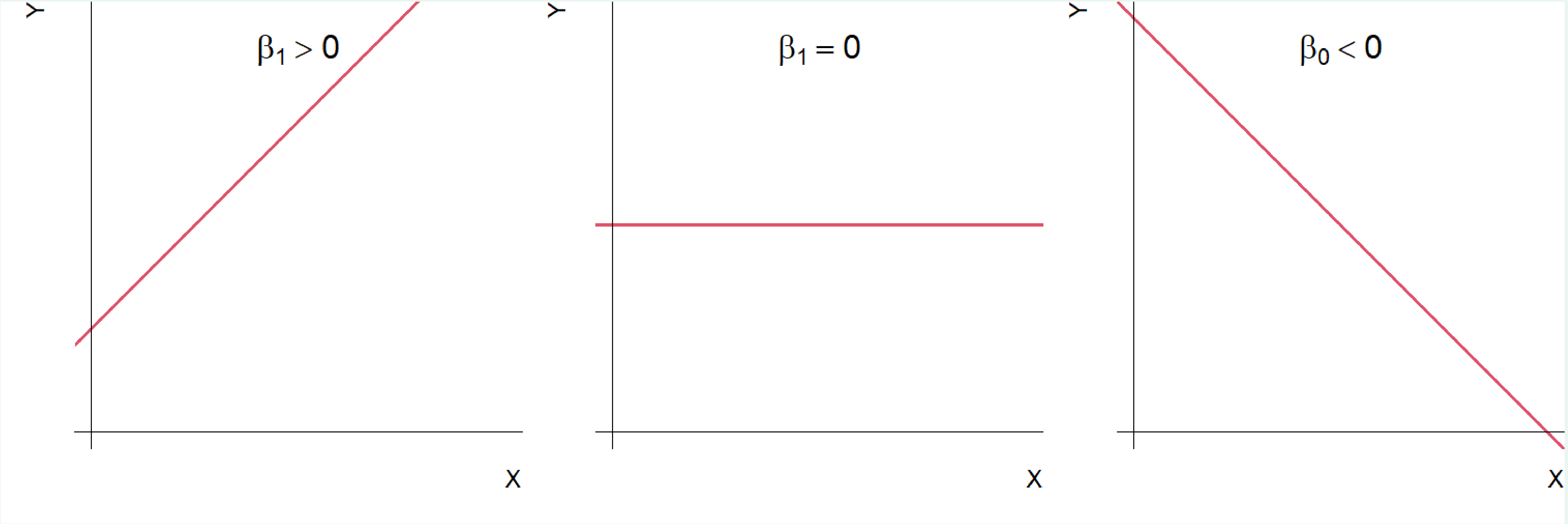
β_1: coeficiente de inclinação da reta (coeficiente de regressão);



2. O modelo matemático

Se o intercepto β_0 e a inclinação β_1 são conhecidos, podemos PREDIZER qualquer valor y_i para um dado valor em x_i.

$$E(Y|x_i) = \beta_0 + \beta_1 x_i$$



2. A tabela e o gráfico de dispersão

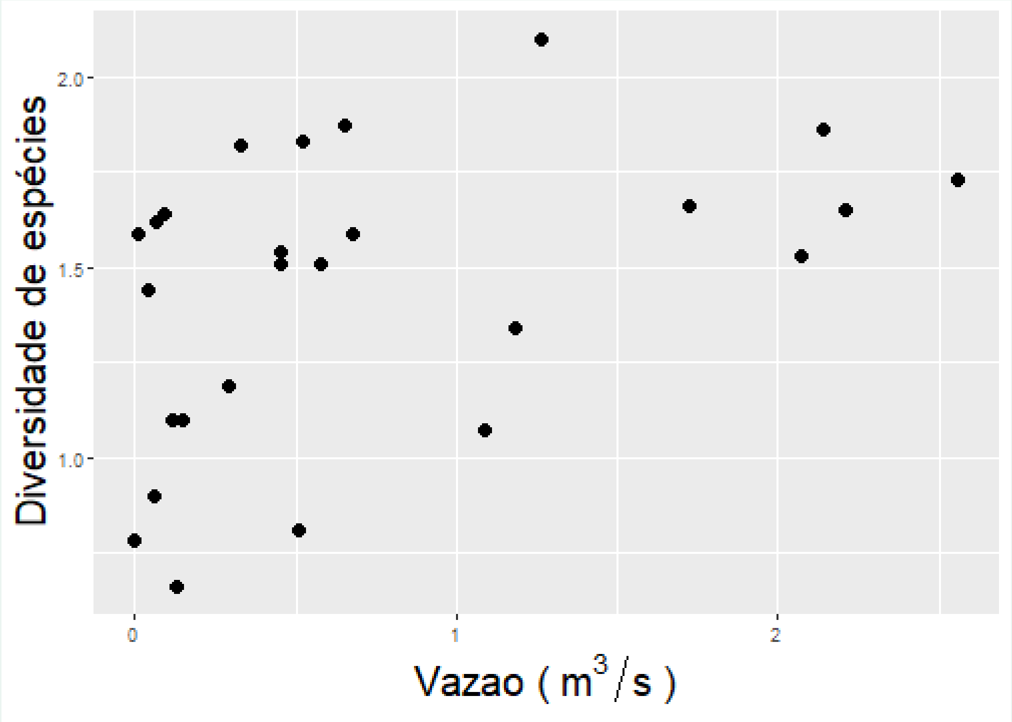
$$y_i|x_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Show 5 entries Search:

	Diversidade	Vazao
1	0.78	0
2	1.59	0.01
3	1.44	0.04
4	0.9	0.06
5	1.62	0.07

Showing 1 to 5 of 26 entries

Previous 1 2 3 4 5 6 Next



2. O modelo estatístico: reta de regressão e resíduos (ε_i)

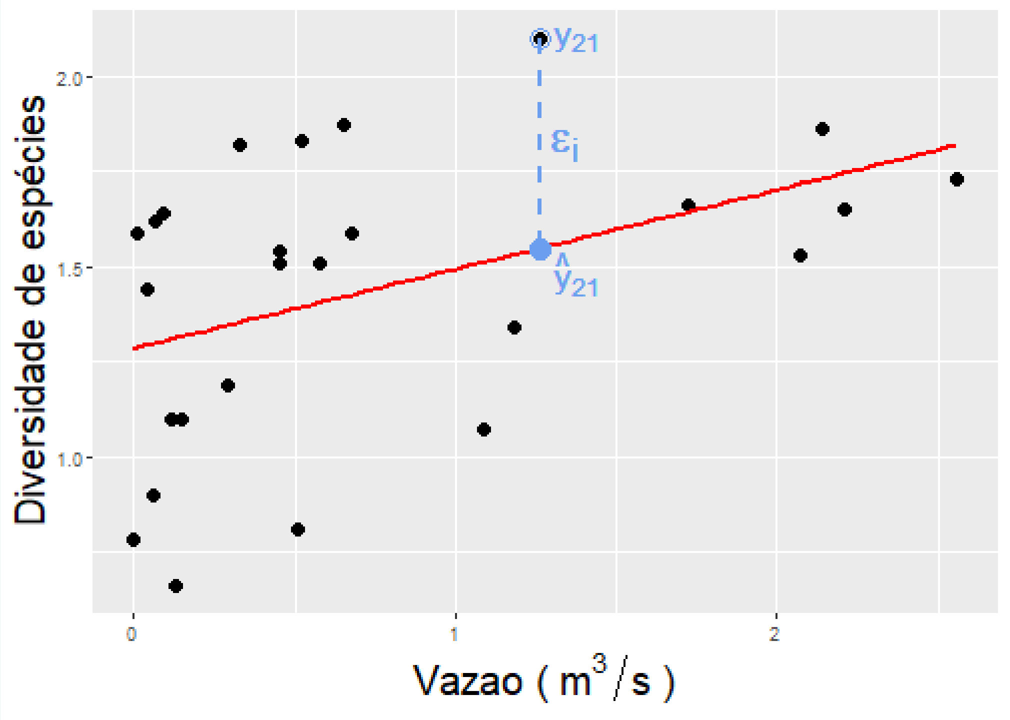
$$y_i|x_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Show 5 entries Search:

	y_i	x_i	$\hat{y}_i$	ε_i
1	0.78	0	1.28	-0.5
2	1.59	0.01	1.29	0.3
3	1.44	0.04	1.29	0.15
4	0.9	0.06	1.3	-0.4
5	1.62	0.07	1.3	0.32

Showing 1 to 5 of 26 entries

Previous 1 2 3 4 5 6 Next



2. O modelo estatístico: reta de regressão e resíduos (ε_i)

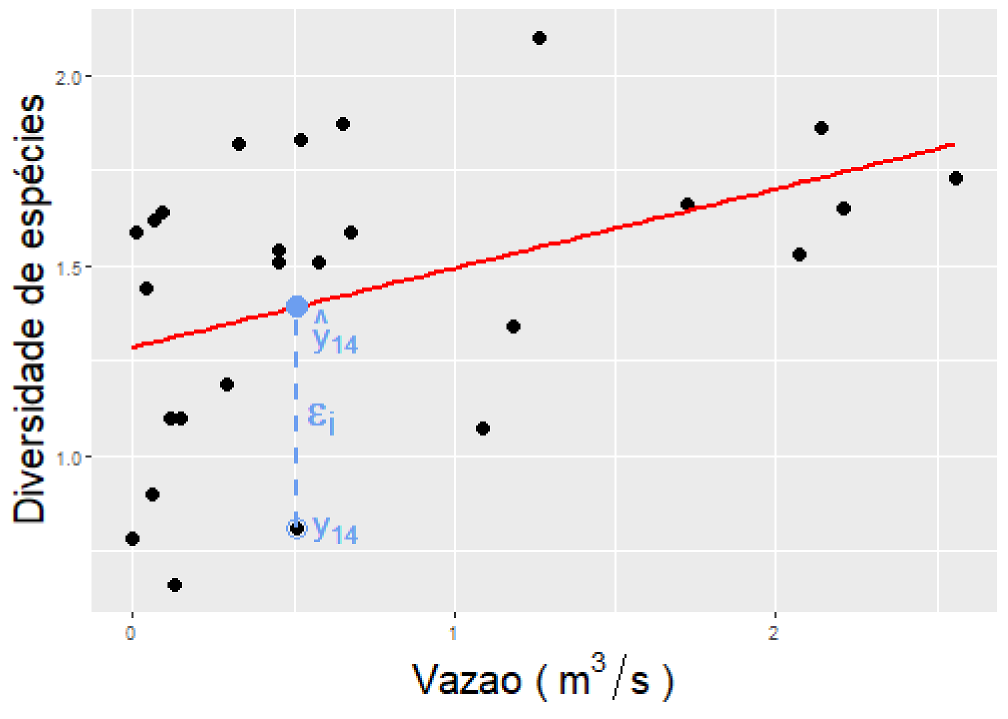
$$y_i|x_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Show 5 entries Search:

	y_i	x_i	$\hat{y}_i$	ε_i
1	0.78	0	1.28	-0.5
2	1.59	0.01	1.29	0.3
3	1.44	0.04	1.29	0.15
4	0.9	0.06	1.3	-0.4
5	1.62	0.07	1.3	0.32

Showing 1 to 5 of 26 entries

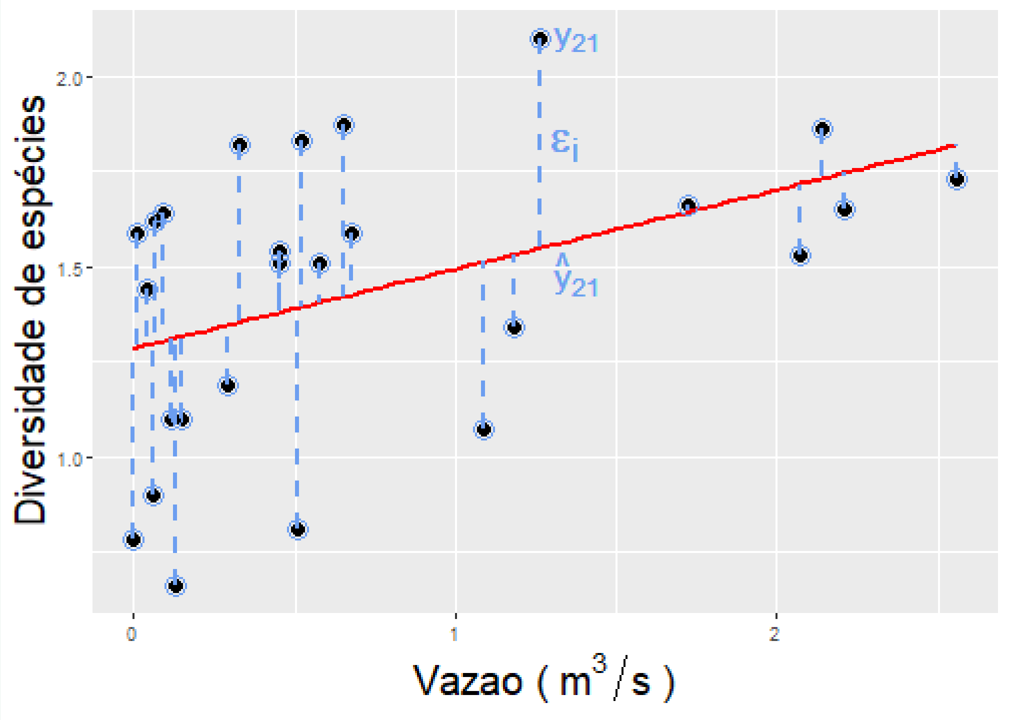
Previous 1 2 3 4 5 6 Next



2. O modelo estatístico: reta de regressão e resíduos (ε_i)

$$y_i|x_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

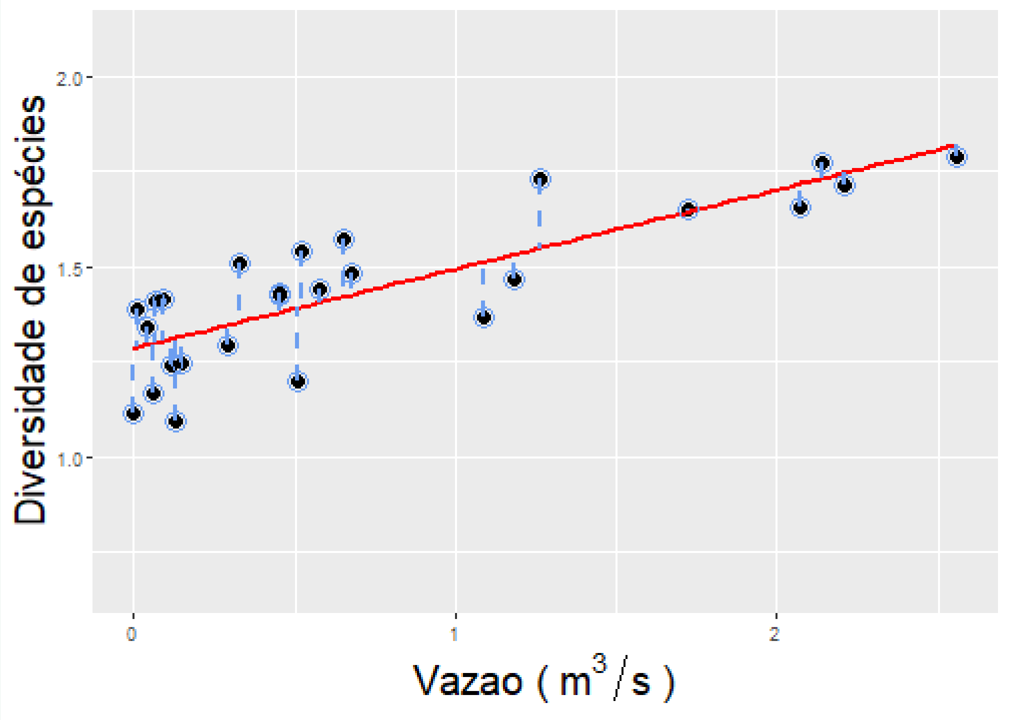
- y_i: variável resposta - i: $1 \cdots n$;
- x_i: variável preditora - i: $1 \cdots n$;
- n: tamanho da amostra;
- β_0: intercepto;
- β_1: coeficiente inclinação da reta;
- ε_i: resíduo - responsável pela variação de y_i em torno do valor **predito** $(\hat{y}_i)$ pela reta de regressão.



2. O modelo estatístico: reta de regressão e resíduos (ε_i)

$$y_i|x_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- y_i: variável resposta - i: $1 \cdots n$;
- x_i: variável preditora - i: $1 \cdots n$;
- n: tamanho da amostra;
- β_0: intercepto;
- β_1: coeficiente inclinação da reta;
- ε_i: resíduo - responsável pela variação de y_i em torno do valor **predito** $(\hat{y}_i)$ pela reta de regressão.



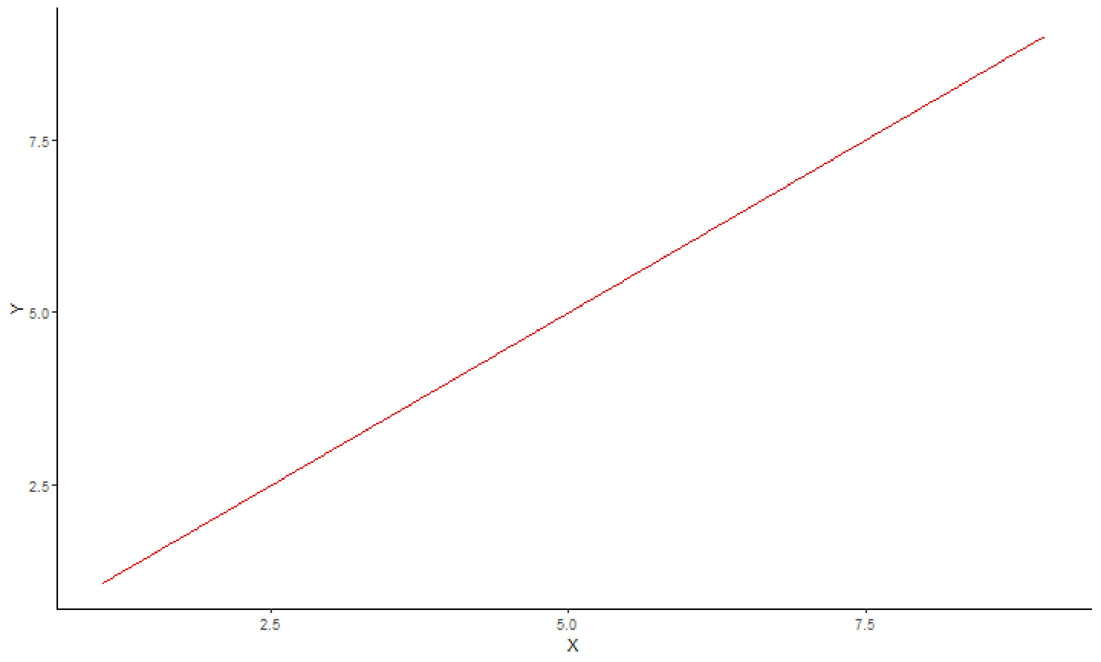
O resíduo associado a cada observação diminui ou aumenta à medida que o pontos está mais próximo ou distante da reta de regressão.

2. Estimativa dos parâmetros

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

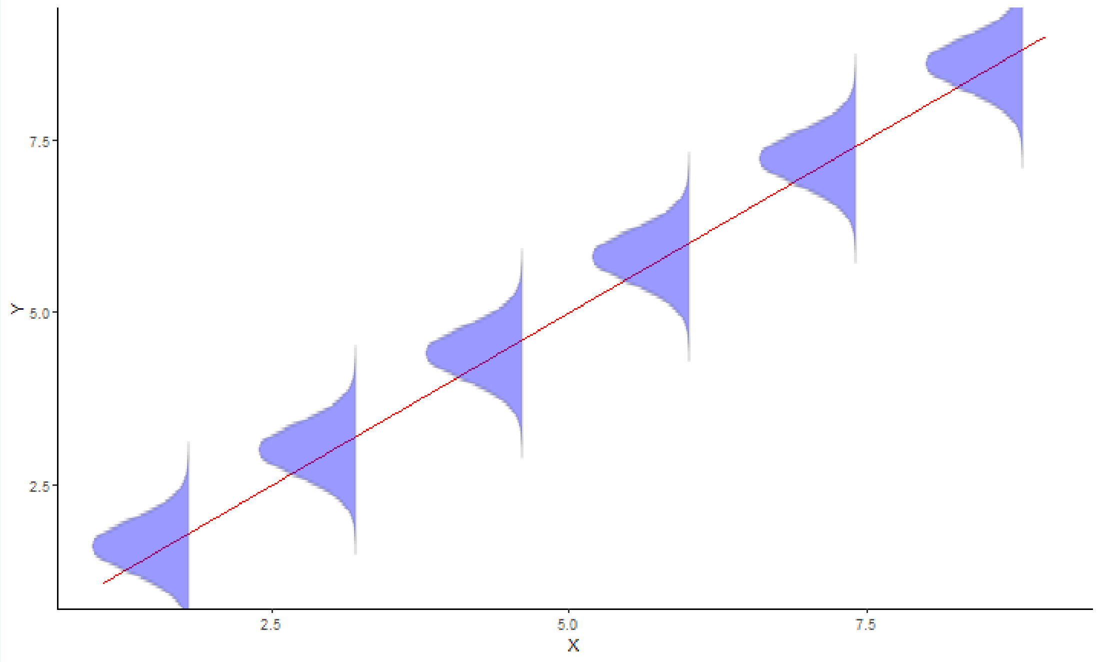
Parte determinística: β_0 e β_1

$$\beta_0 + \beta_1 x_i$$



Parte estocástica: σ^2

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$



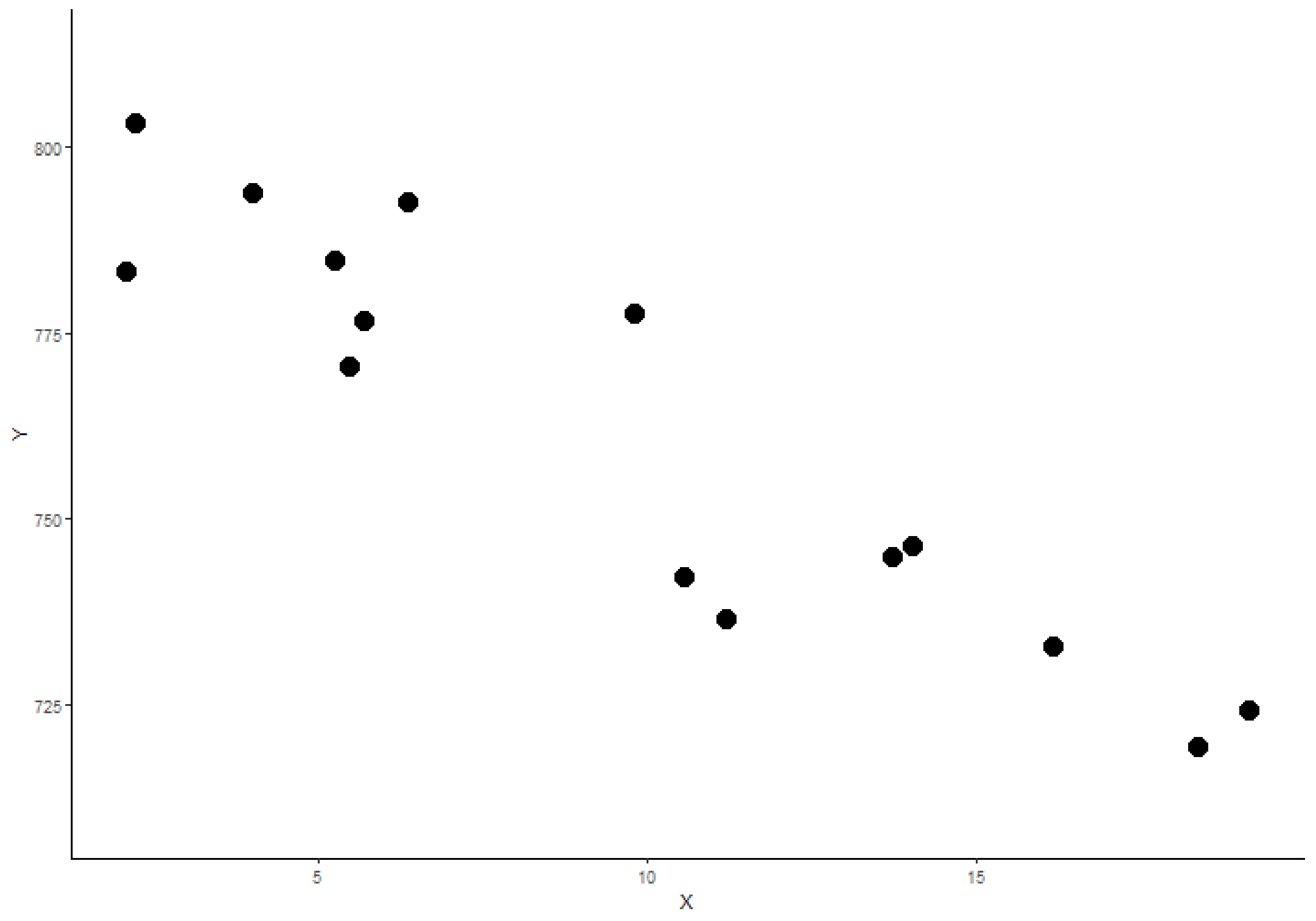
2. Estimativa dos parâmetros

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

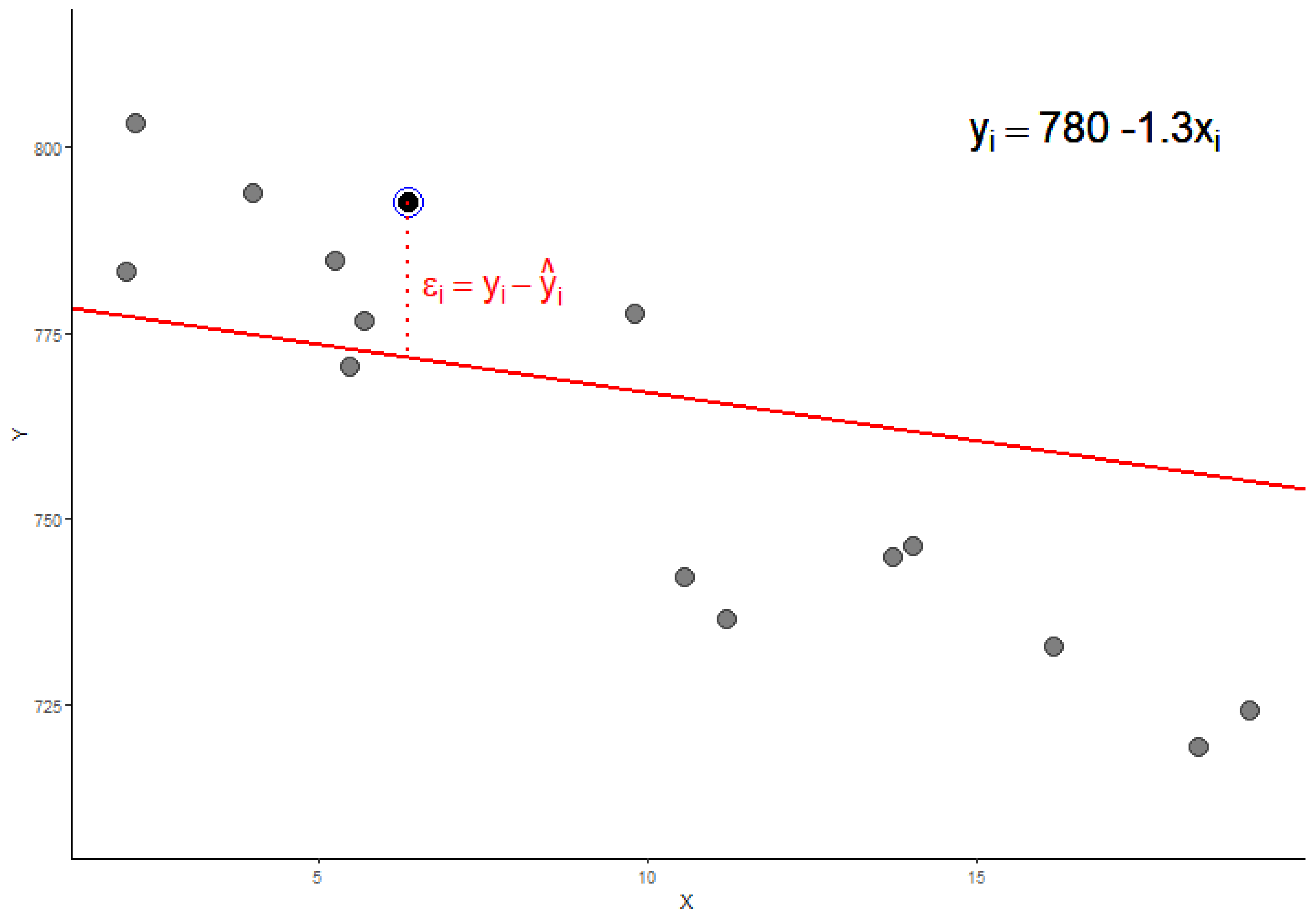
Como estimar os parâmetros de um modelo de regressão?

1. Método dos Mínimos Quadrados (MMQ)
2. Estimador de Máxima Verossimilhança (EMV)

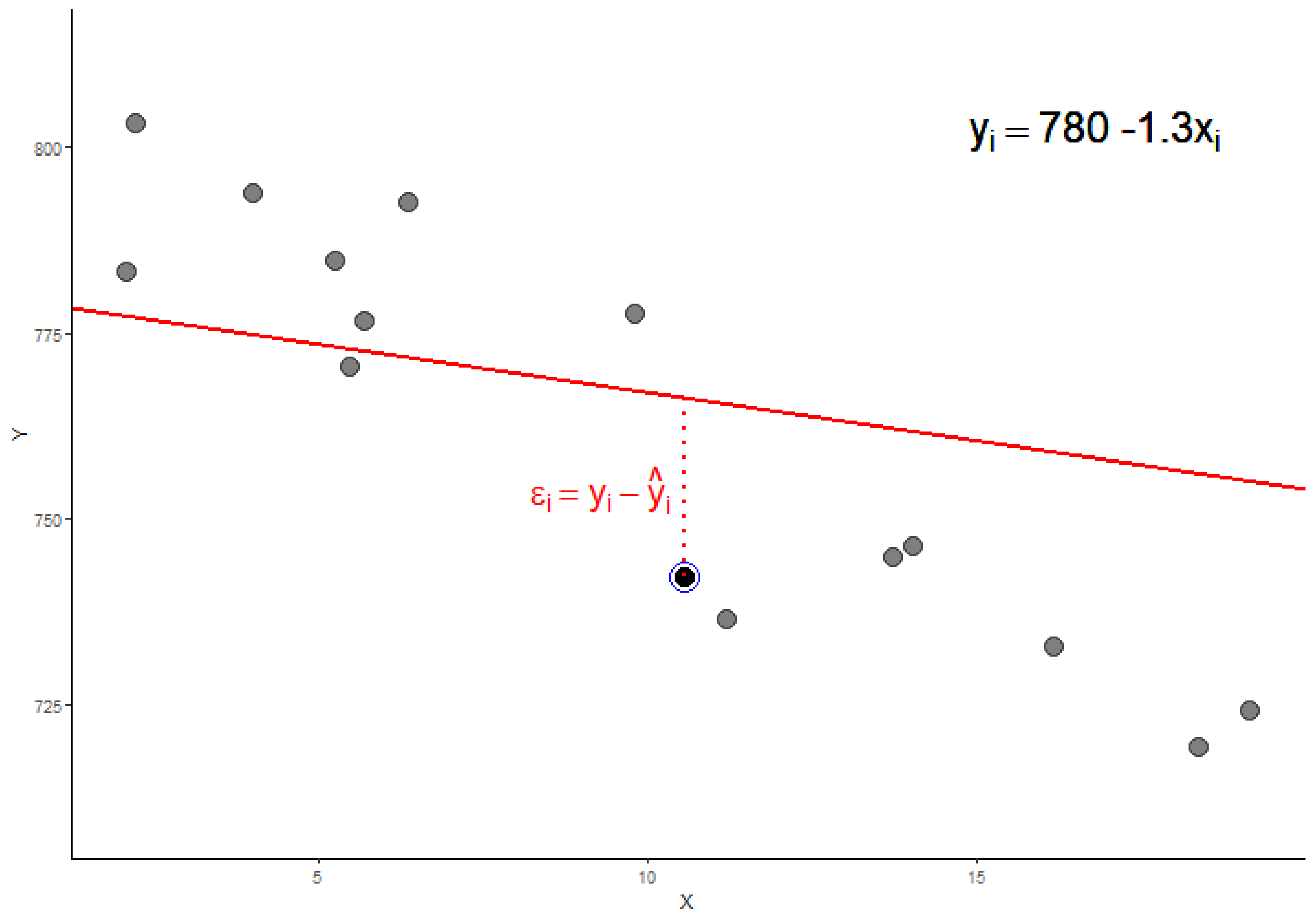
2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*



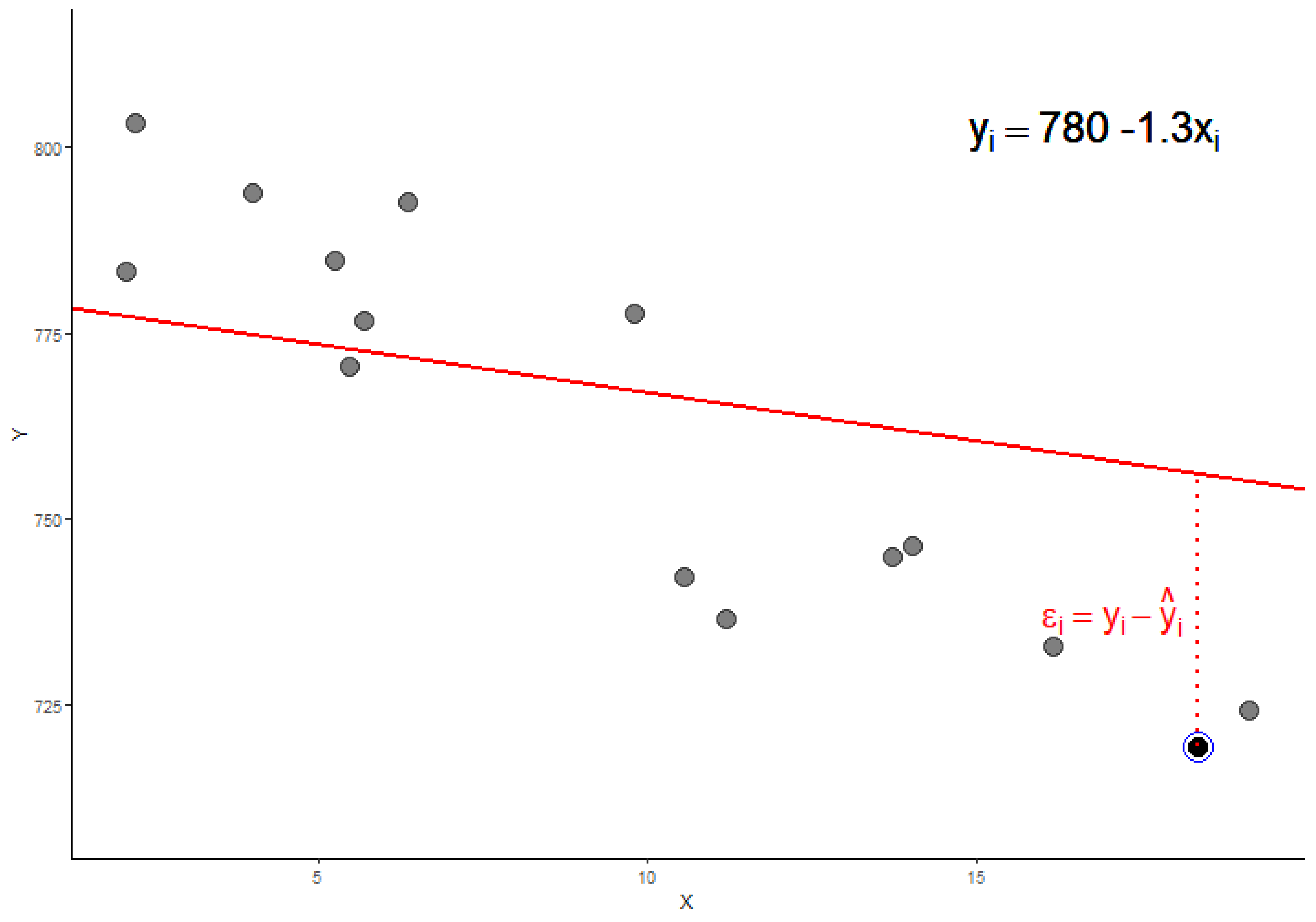
2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*



2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

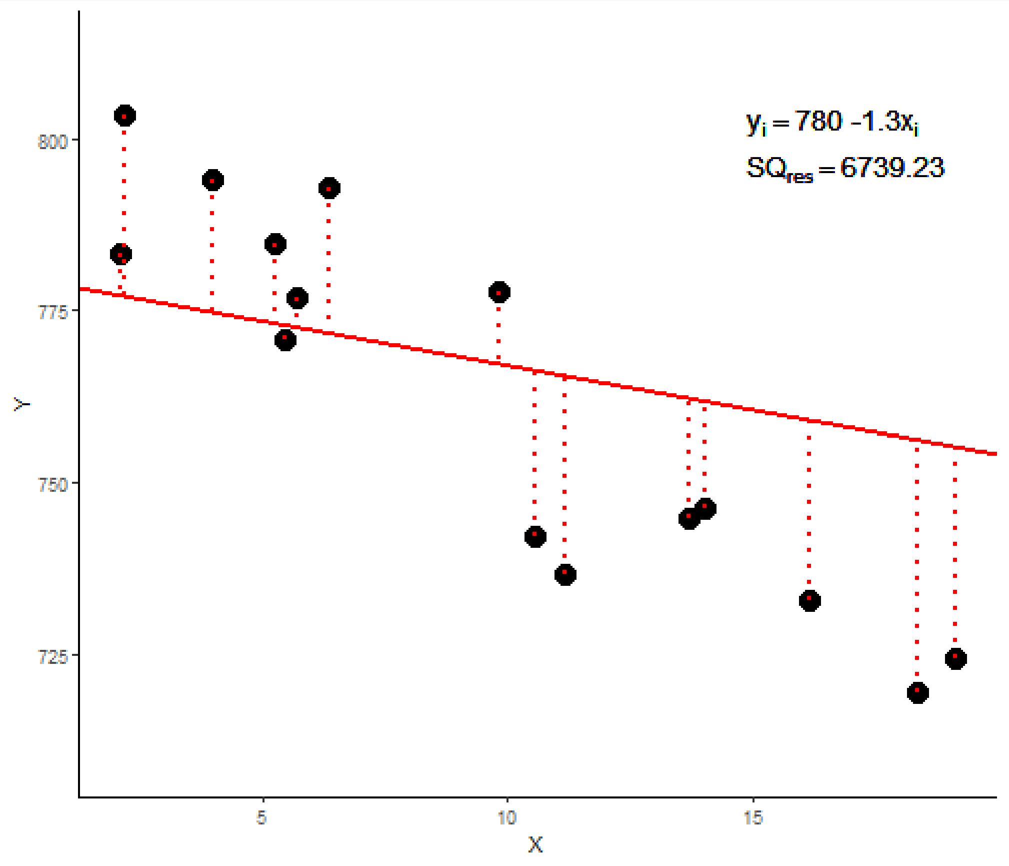


2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*



2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

Soma dos quadrados dos resíduos (SQ_{Res})

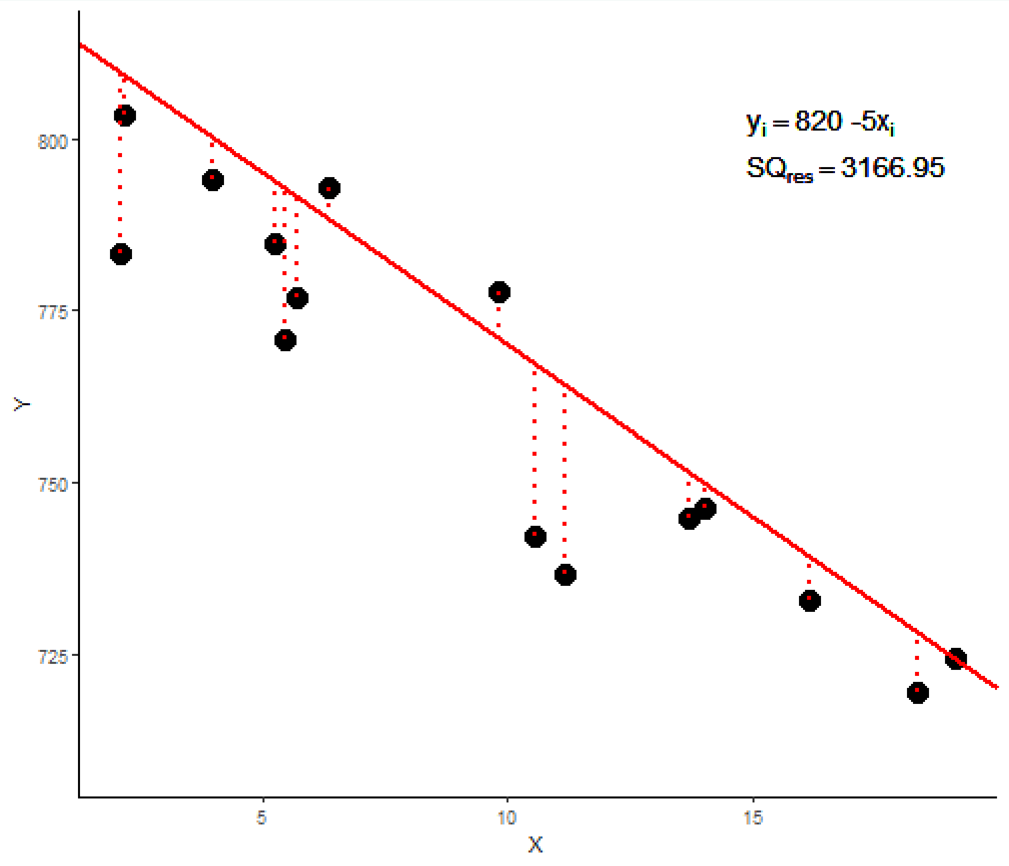


$$SQ_{Res} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

O método dos mínimos quadrados consiste em encontrar a reta que MINIMIZA o somatório dos quadrados dos resíduos.

2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

Soma dos quadrados dos resíduos (SQ_{Res})

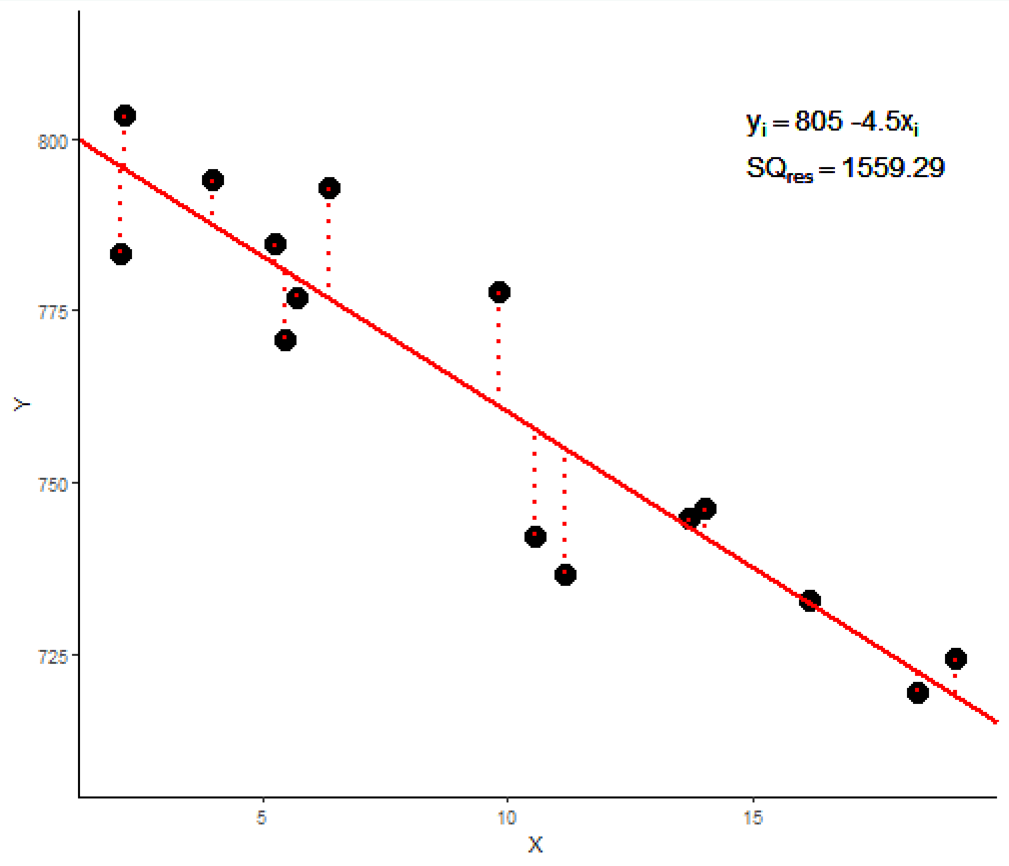


$$SQ_{Res} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

O método dos mínimos quadrados consiste em encontrar a reta que MINIMIZA o somatório dos quadrados dos resíduos.

2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

Soma dos quadrados dos resíduos (SQ_{Res})



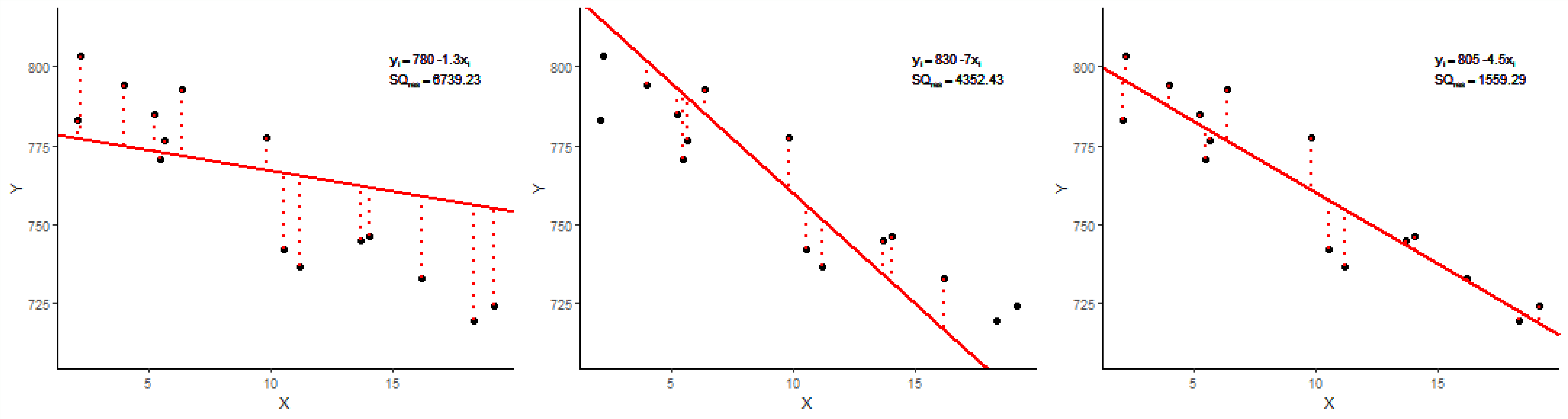
$$SQ_{Res} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

O método dos mínimos quadrados consiste em encontrar a reta que MINIMIZA o somatório dos quadrados dos resíduos.

2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

O método dos mínimos quadrados consiste em encontrar a reta que MINIMIZA o somatório dos quadrados dos resíduos.

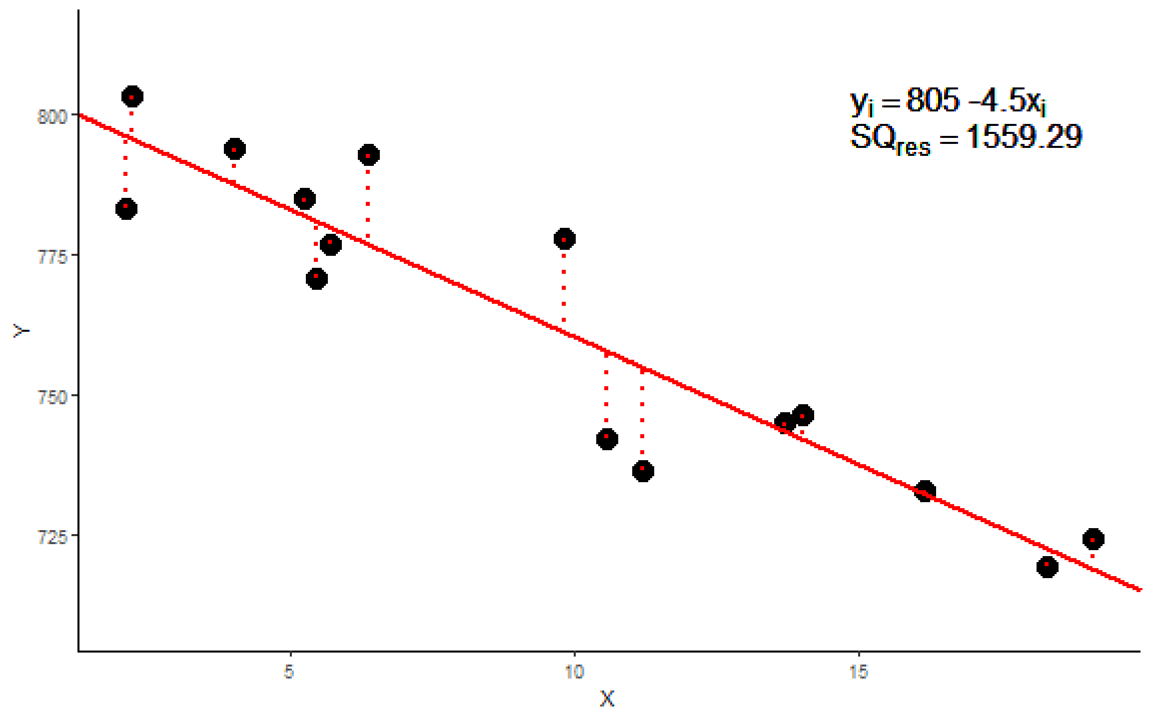
$$SQ_{Res} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$



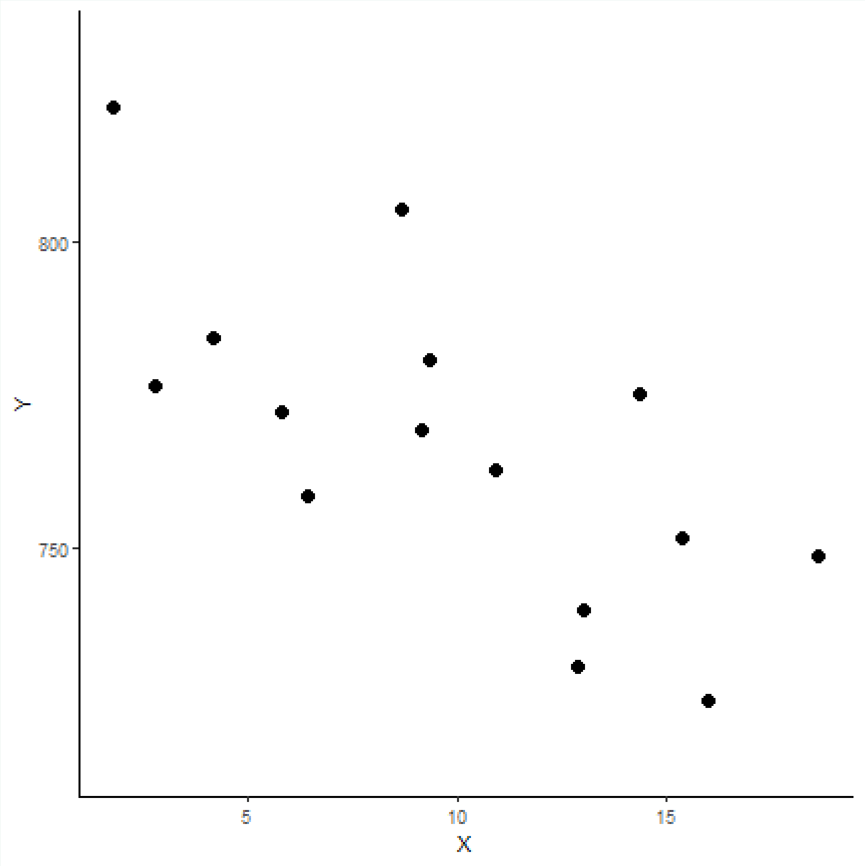
2. Estimativa dos parâmetros: *O Método dos Mínimos Quadrados (MMQ)*

---> *Estime* $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimize a quantia:

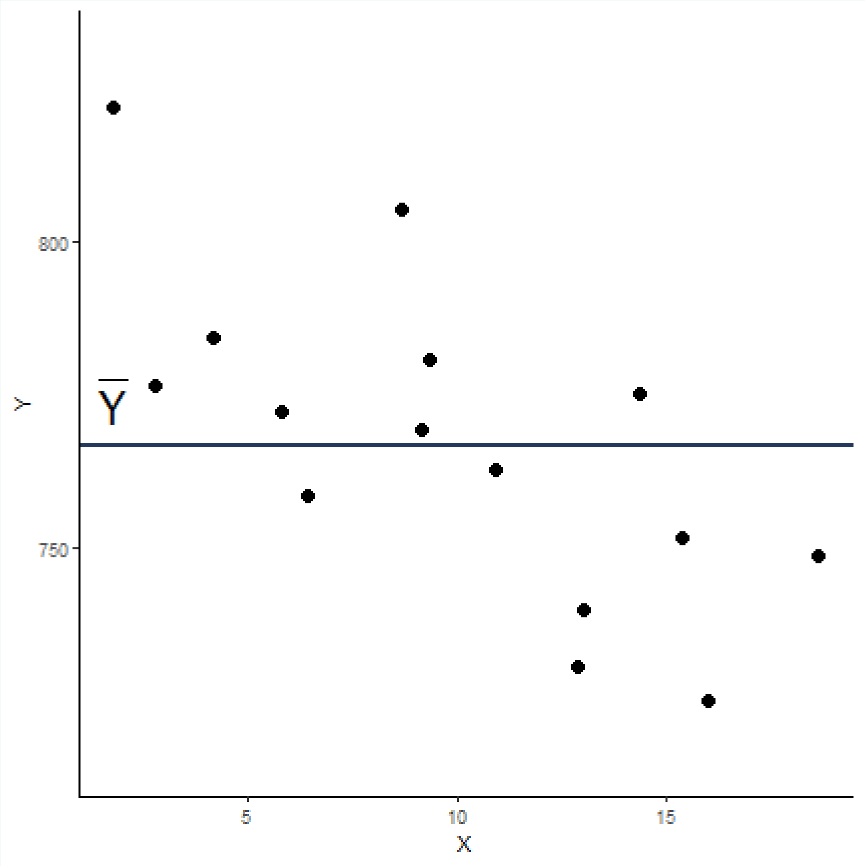
$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$



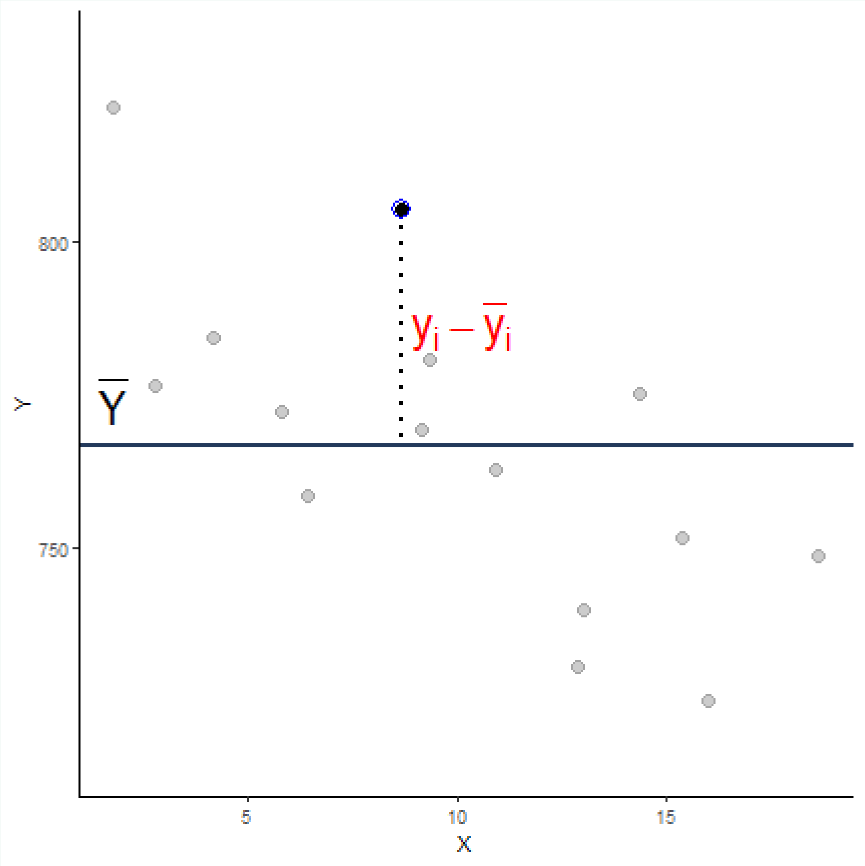
2. Variâncias e Covariâncias



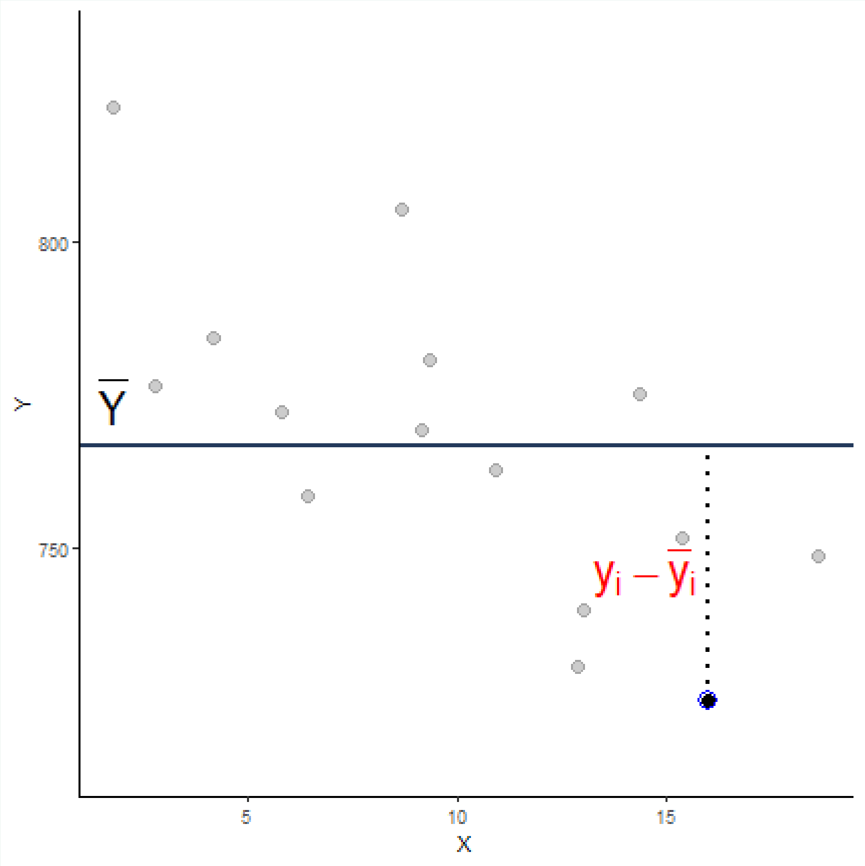
2. Variâncias e Covariâncias



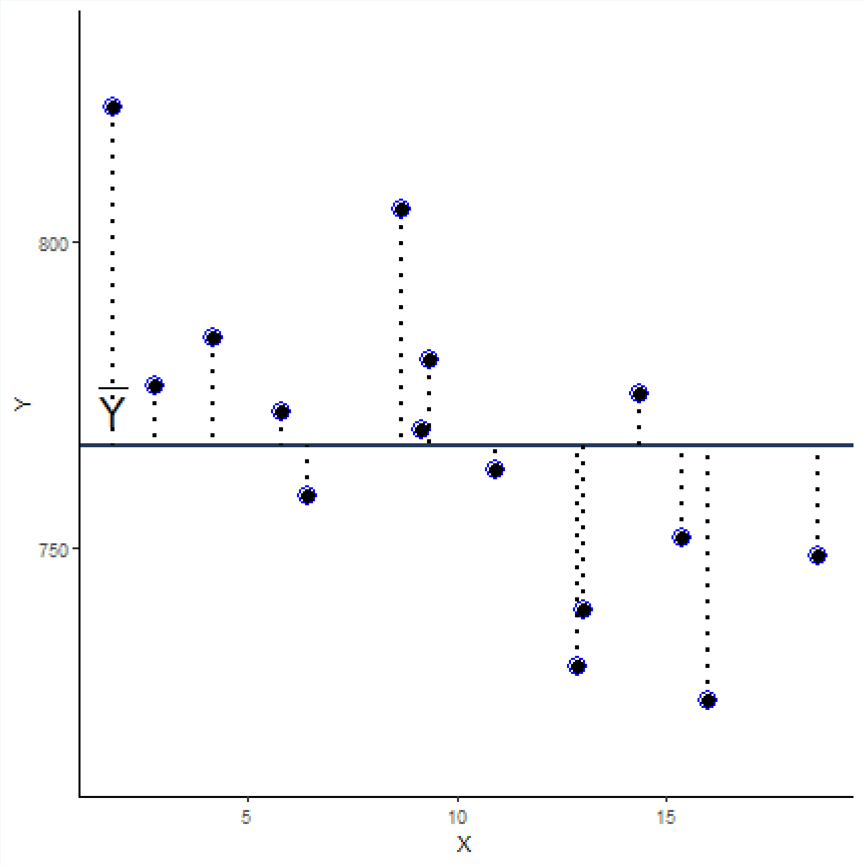
2. Variâncias e Covariâncias



2. Variâncias e Covariâncias



2. Variâncias e Covariâncias



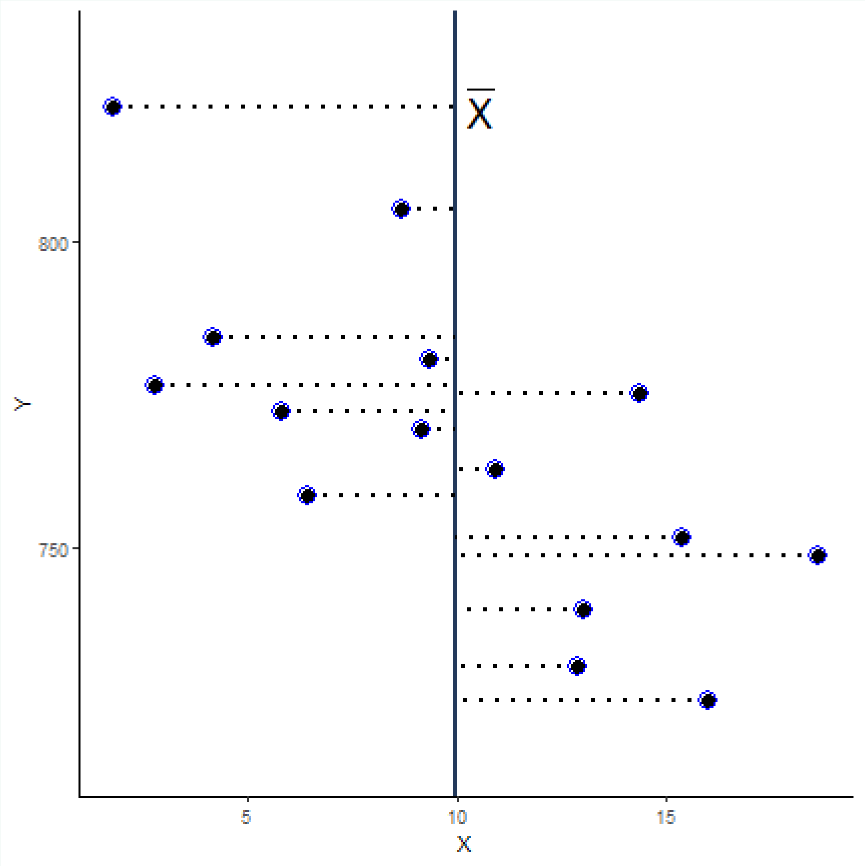
Soma dos Quadrados de Y

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

Variância amostral de Y

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

2. Variâncias e Covariâncias



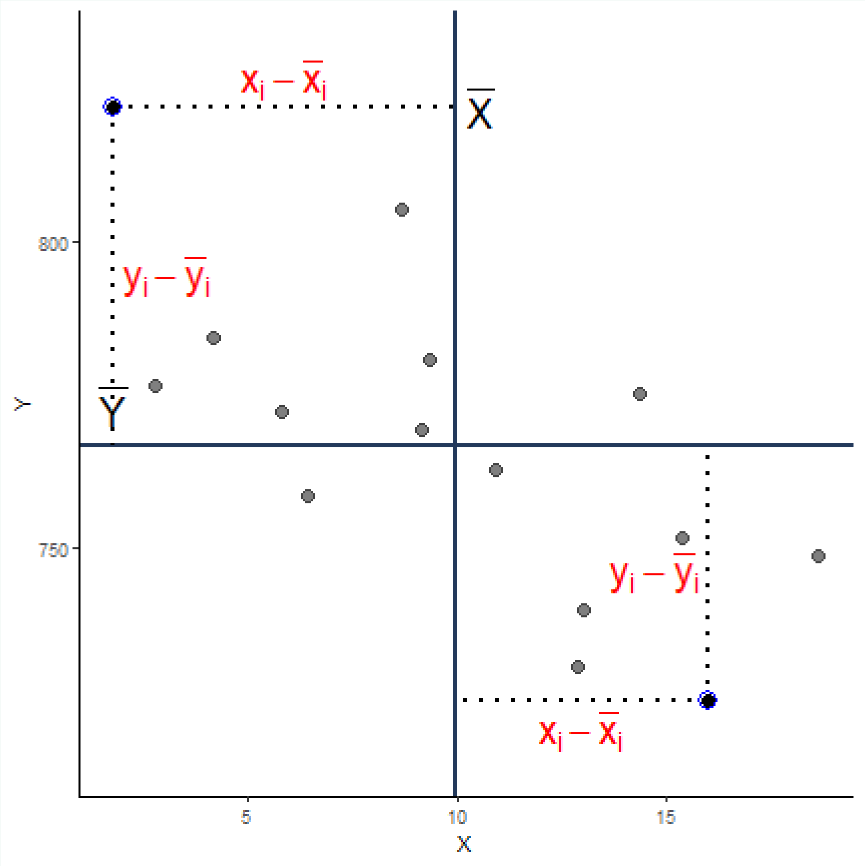
Soma dos Quadrados de X

$$SQ_X = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})$$

Variância amostral de X

$$s_X^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

2. Variâncias e Covariâncias



Soma dos produtos cruzados de Y e X

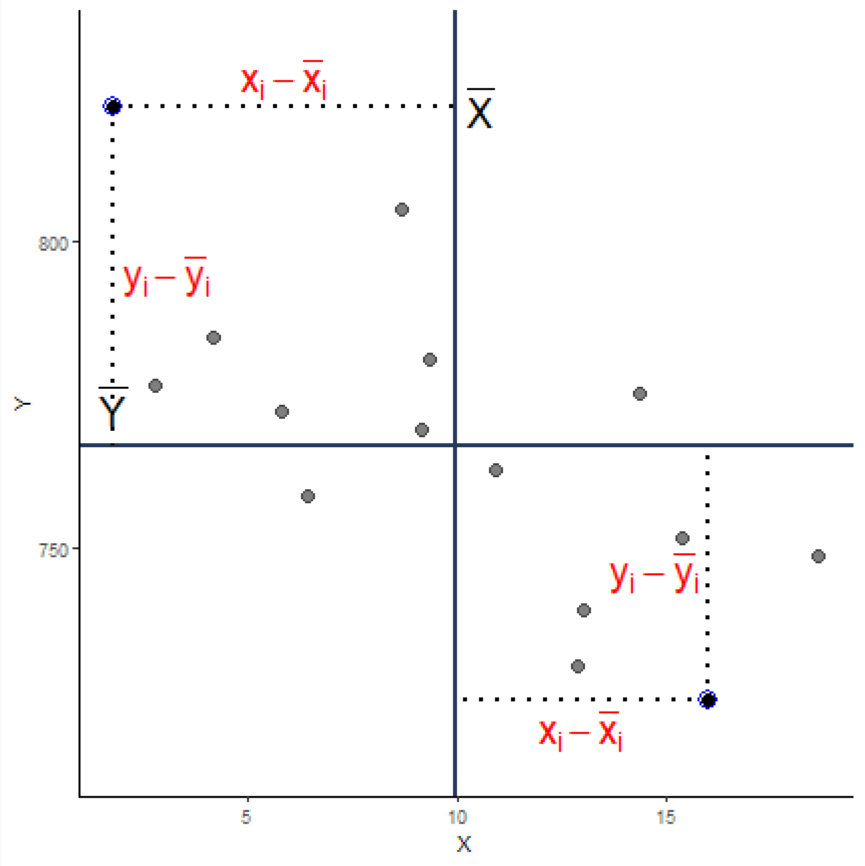
$$SQ_{YX} = \sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})$$

Covariância amostral entre Y e X

$$s_{YX} = \frac{\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})}{n-1}$$

2. Variâncias e Covariâncias

A covariância pode ser NEGATIVA



Quando:

$(y_i - \bar{y}) > 0$; $(x_i - \bar{x}) < 0$

ou

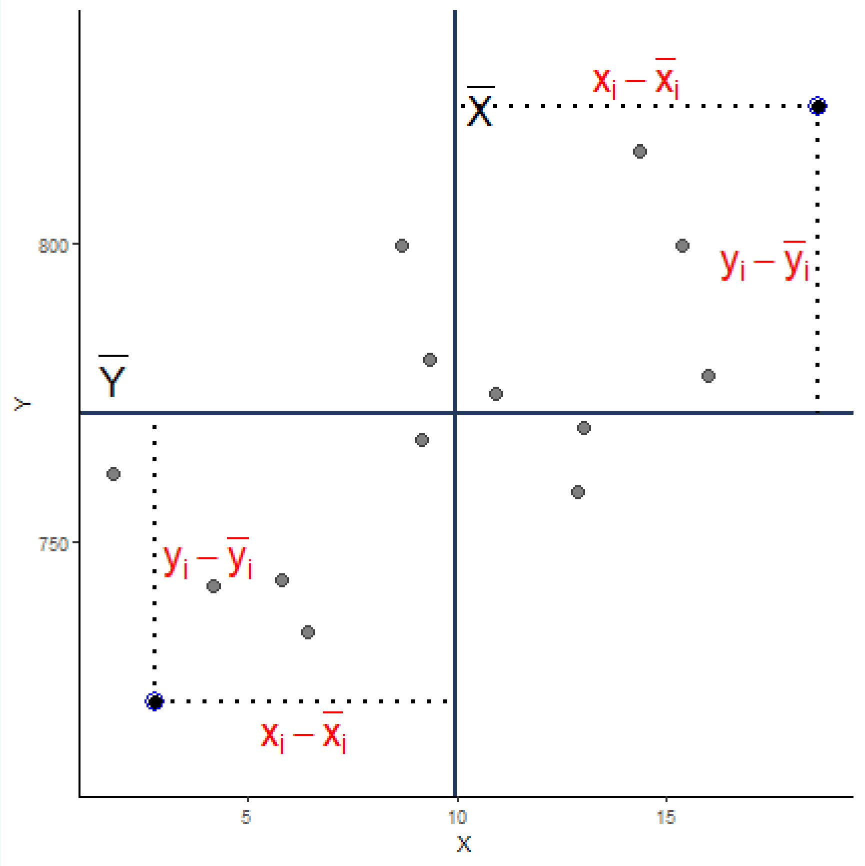
$(y_i - \bar{y}) < 0$; $(x_i - \bar{x}) > 0$

de modo que:

$$s_{YX} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{n-1} < 0$$

2. Variâncias e Covariâncias

A covariância pode ser POSITIVA



Quando:

$(y_i - \overline{y}) > 0$; $(x_i - \overline{x}) > 0$

ou

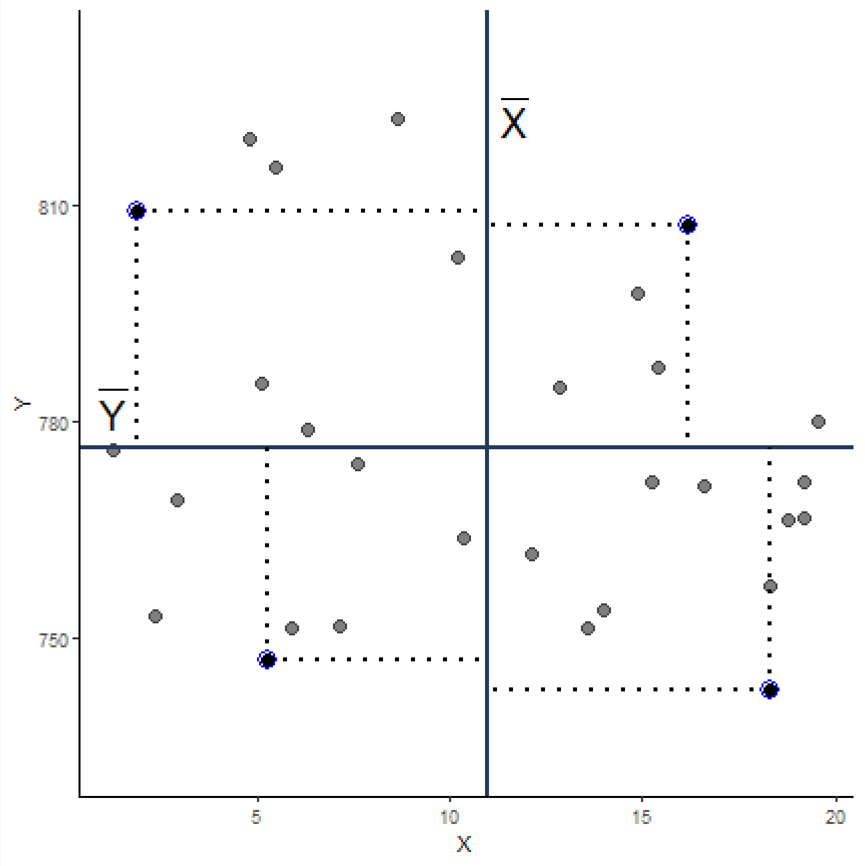
$(y_i - \overline{y}) < 0$; $(x_i - \overline{x}) < 0$

de modo que:

$$s_{YX} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{n-1} > 0$$

2. Variâncias e Covariâncias

A covariância pode ser NULA



Quando:

$$(y_i - \bar{y}) \approx 0; (x_i - \bar{x}) \approx 0$$

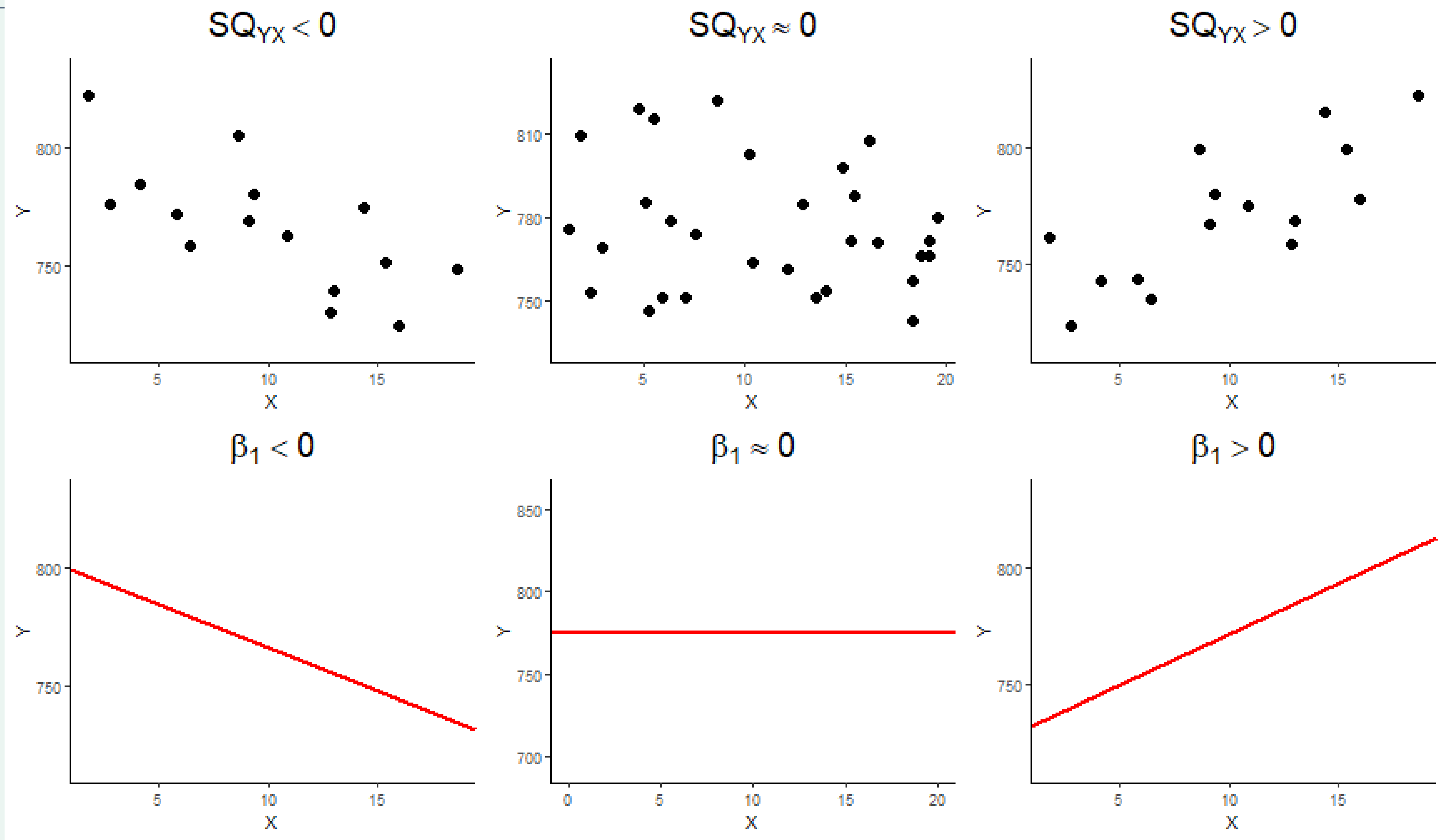
ou

$$(y_i - \bar{y}) \approx 0; (x_i - \bar{x}) \approx 0$$

de modo que:

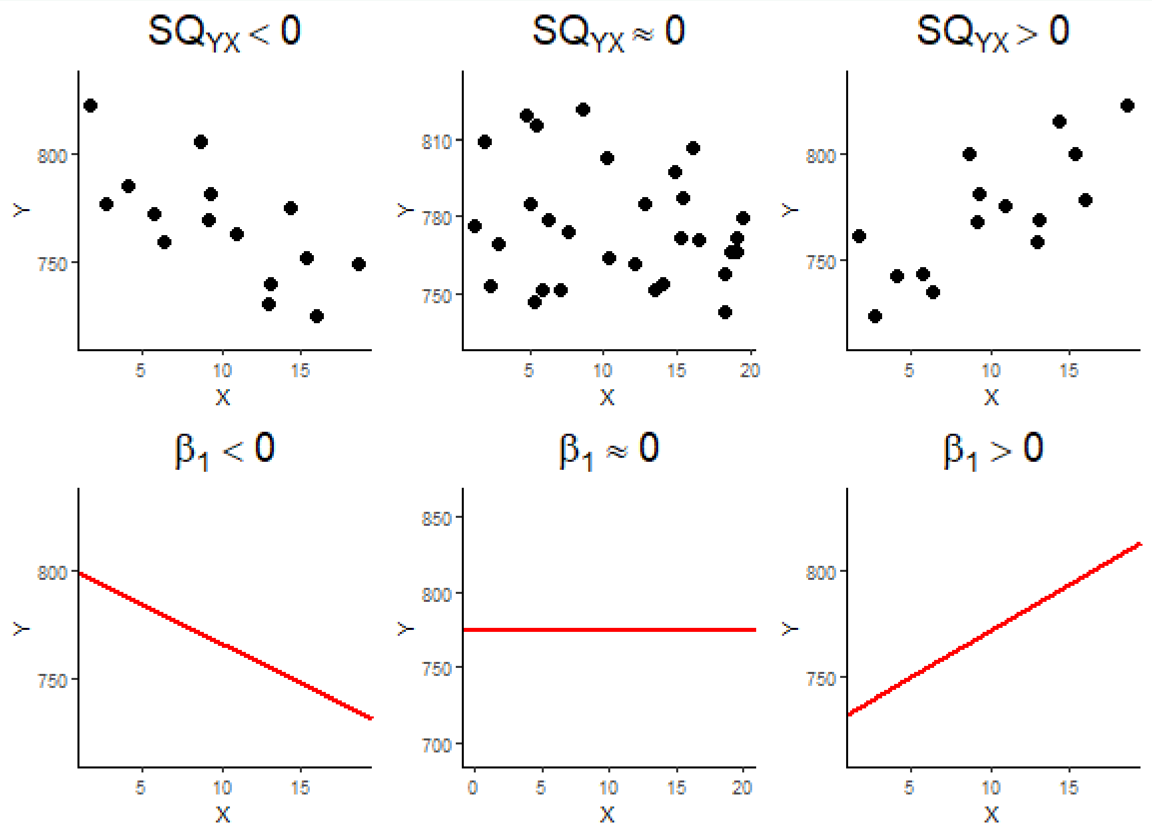
$$s_{YX} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{n-1} \approx 0$$

2. Variâncias e Covariâncias



2. Variâncias e Covariâncias: *estimando* β_1

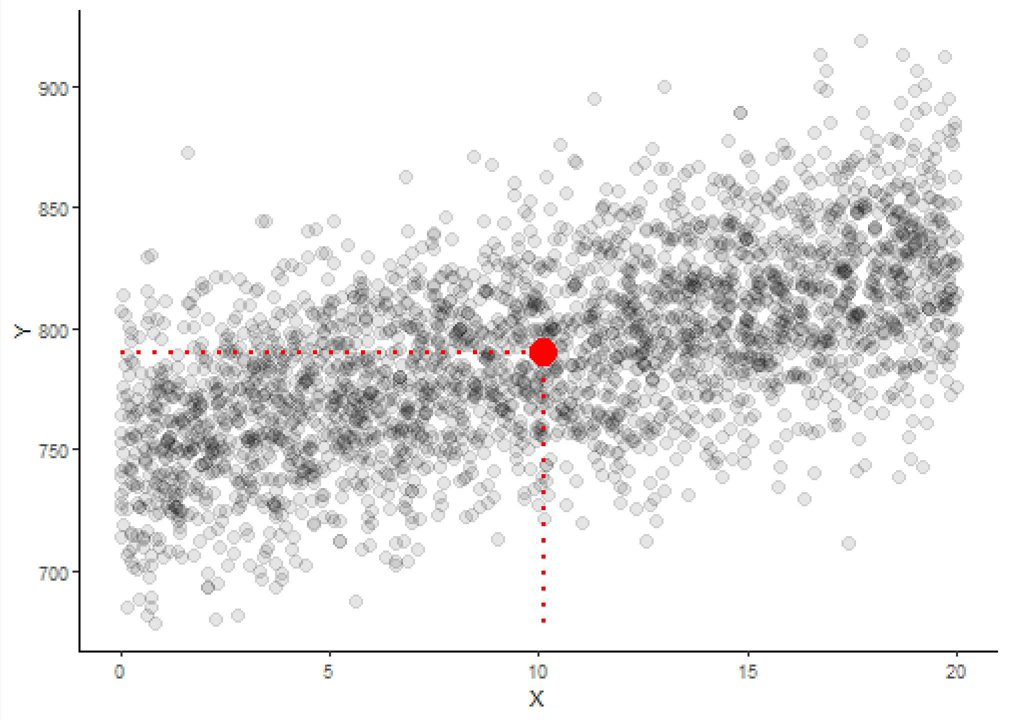
$$\hat{\beta}_1 = \frac{SQ_{YX}}{SQ_X} = \frac{s_{XY}}{s^2_X}$$



2. Variâncias e Covariâncias: *estimando* β_0

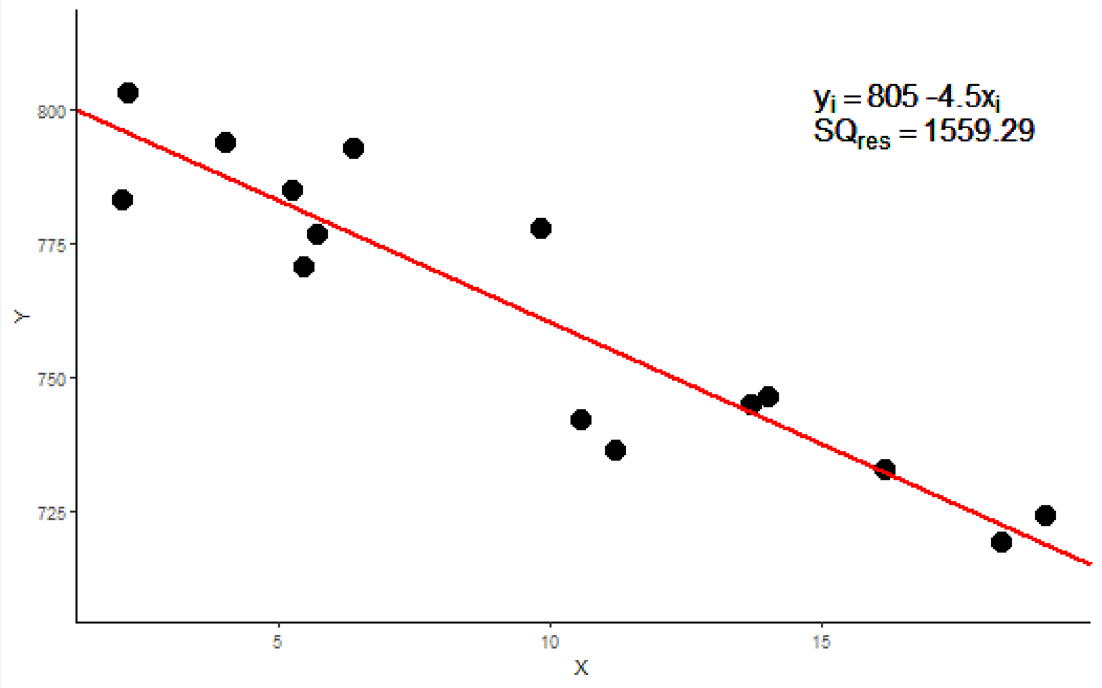
$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$



2. Variâncias e Covariâncias: *estimando* σ^2

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$$



O Quadrado Médio do Resíduo (QM_{Res})

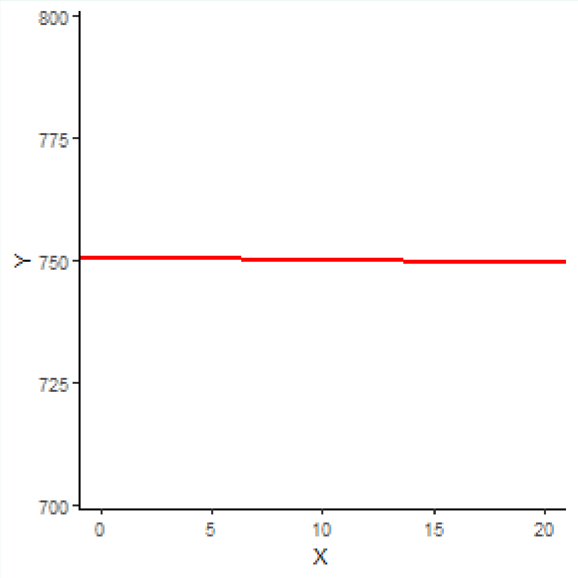
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\hat{\sigma}^2 = QM_{Res} = \frac{SQ_{Res}}{n-2}$$

3. Teste de hipóteses para β_1

Hipótese nula

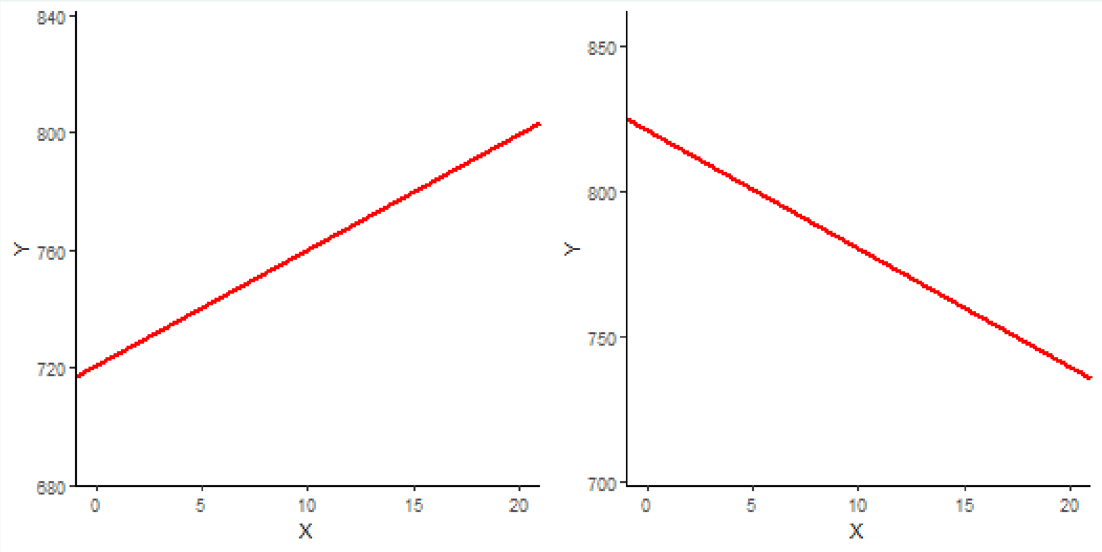
$$H_0 : \beta_1 = 0$$



$$y_i = \beta_0 + \varepsilon_i$$

Hipótese alternativa

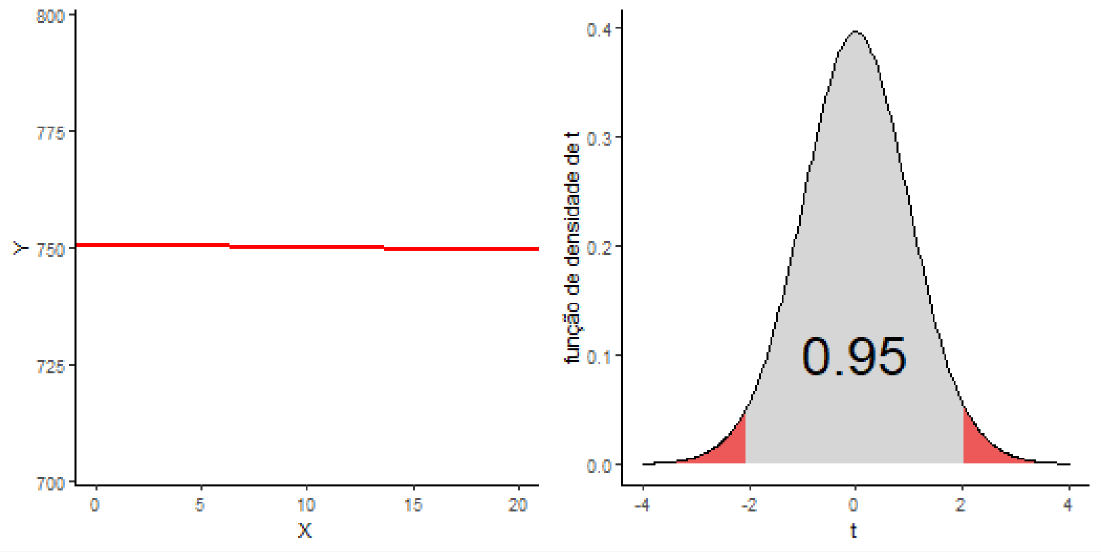
$$H_1 : \beta_1 \neq 0$$



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

3. Teste de hipóteses para β_1

H_0 pode ser testada por meio do teste t para o estimador $\hat{\beta}_1$



$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$

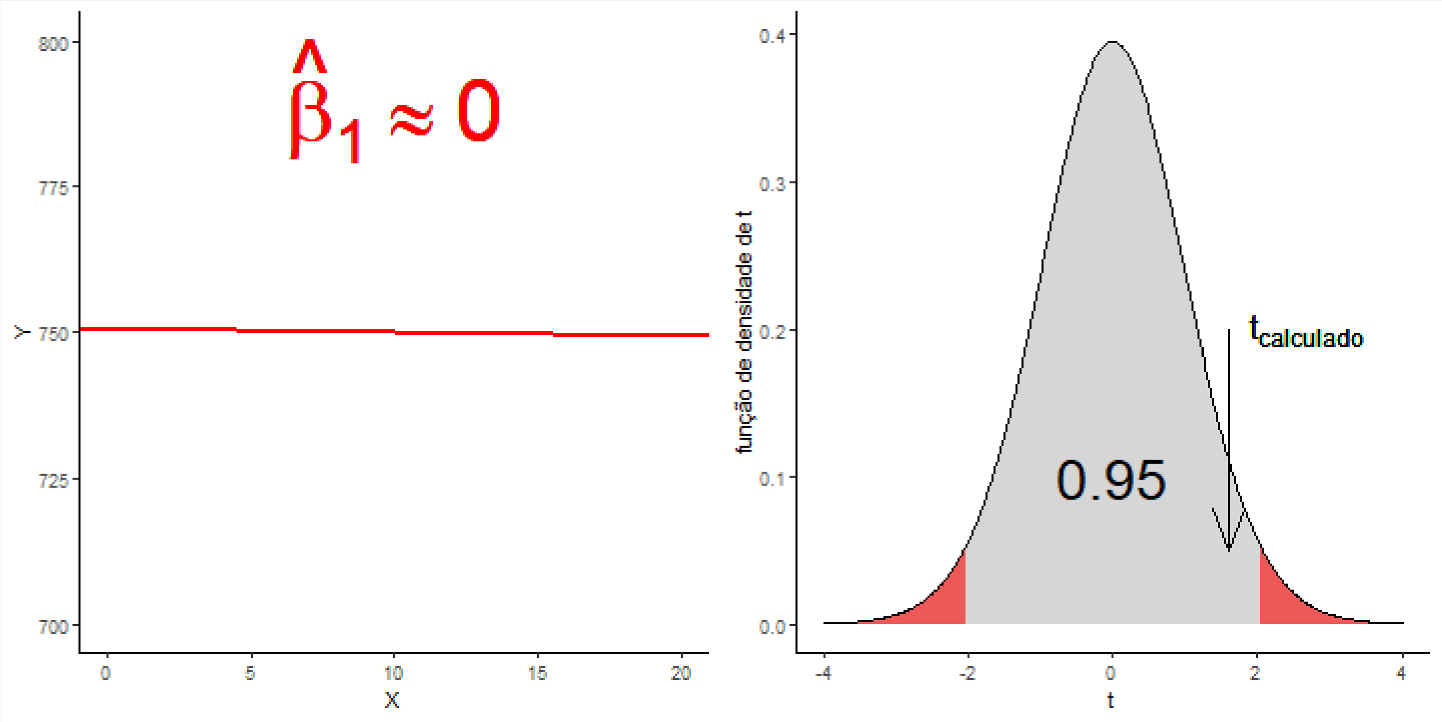
Erro padrão de $\hat{\beta}_1$

$$s_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{SQ_X}}$$

3. Teste de hipóteses para β_1

$t_{calculado}$ depende da **magnitude de $\hat{\beta}_1$**

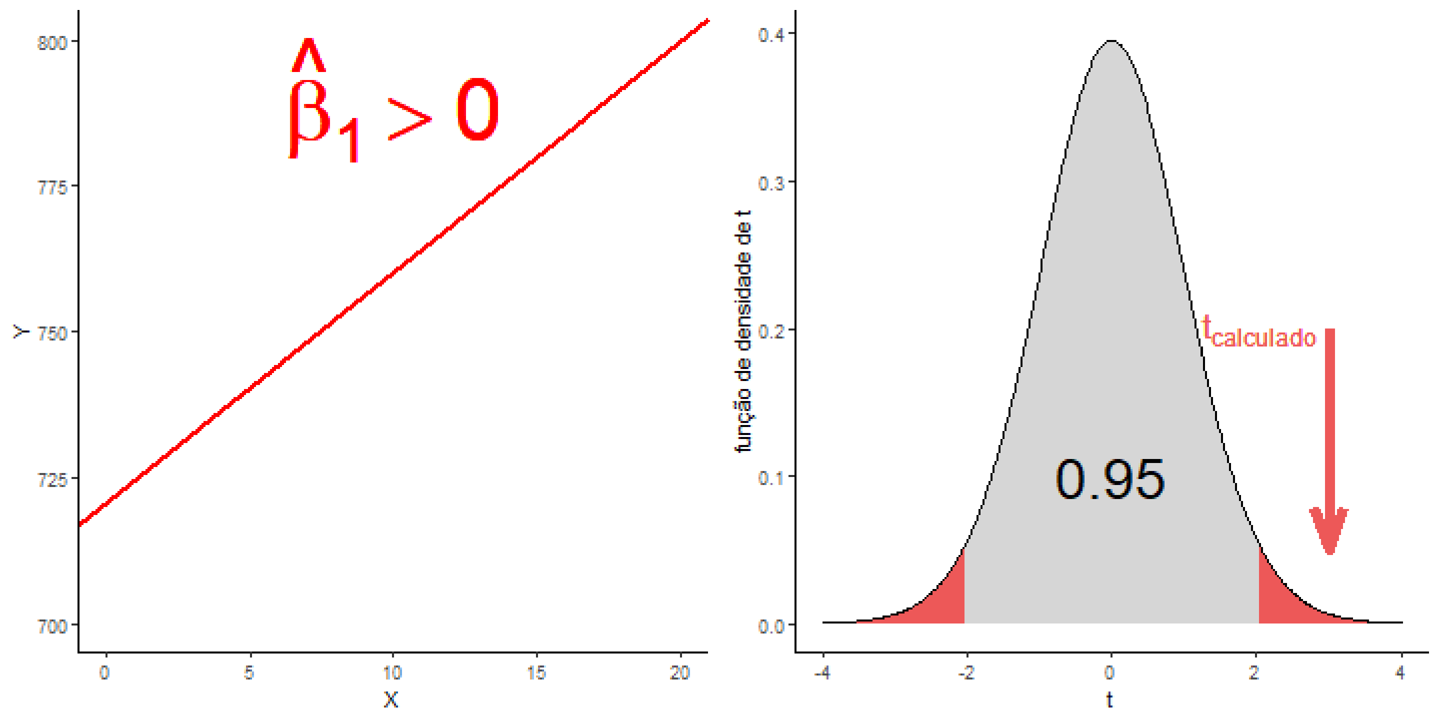
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende da **magnitude de** $\hat{\beta}_1$

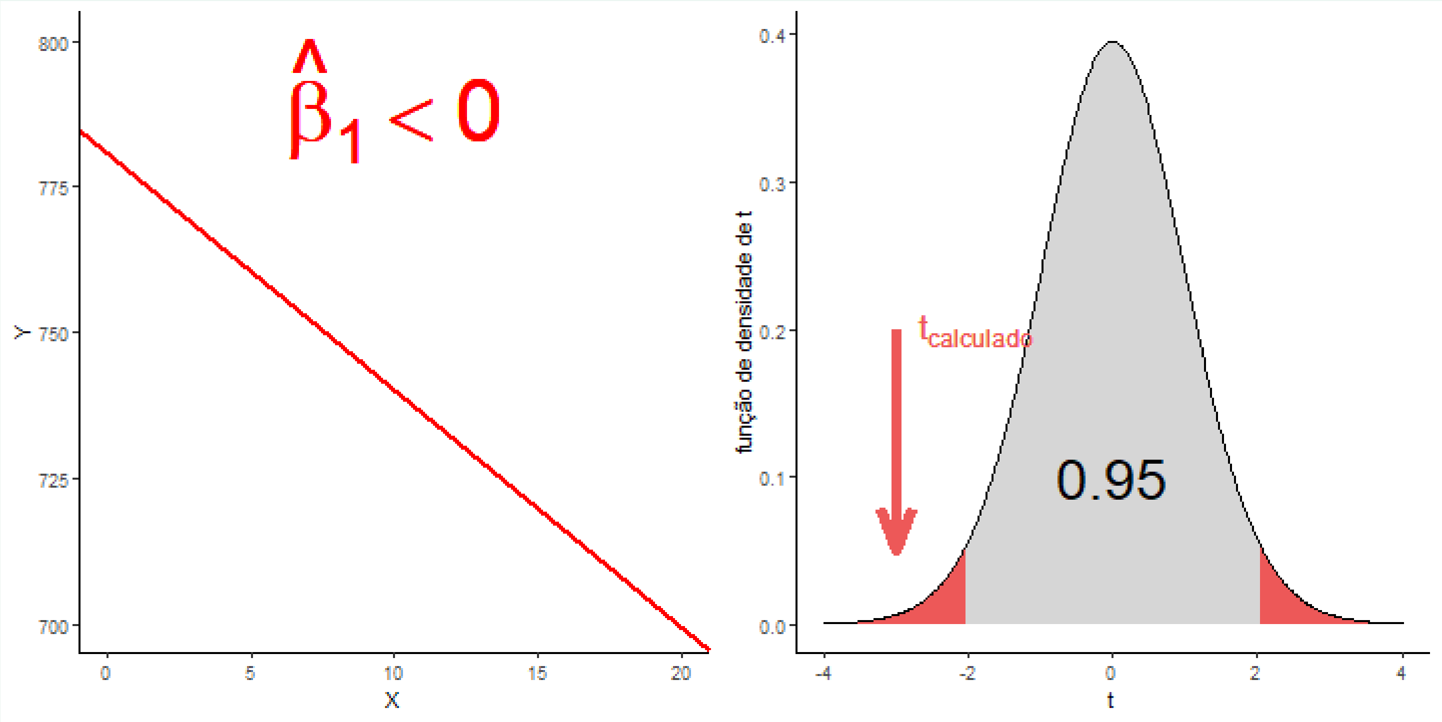
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende da **magnitude de** $\hat{\beta}_1$

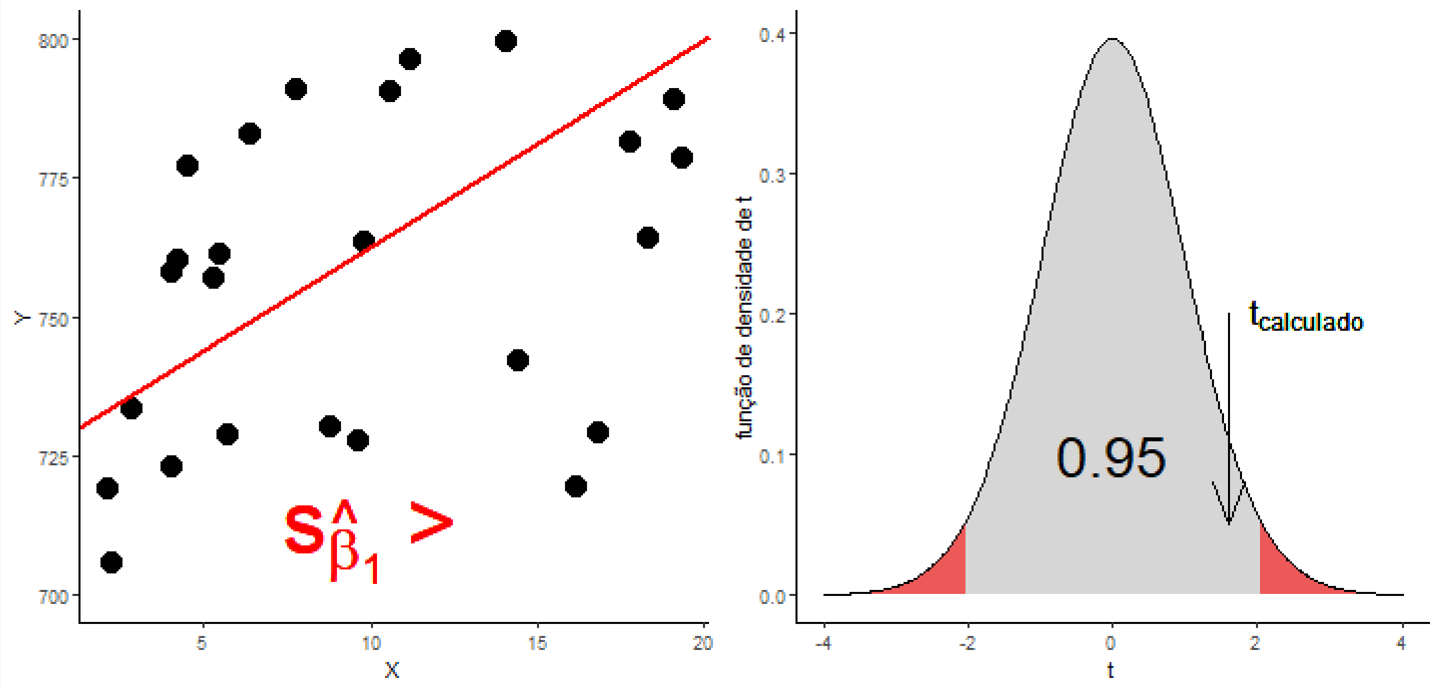
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende da variância residual - $s_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{SQ_X}}$

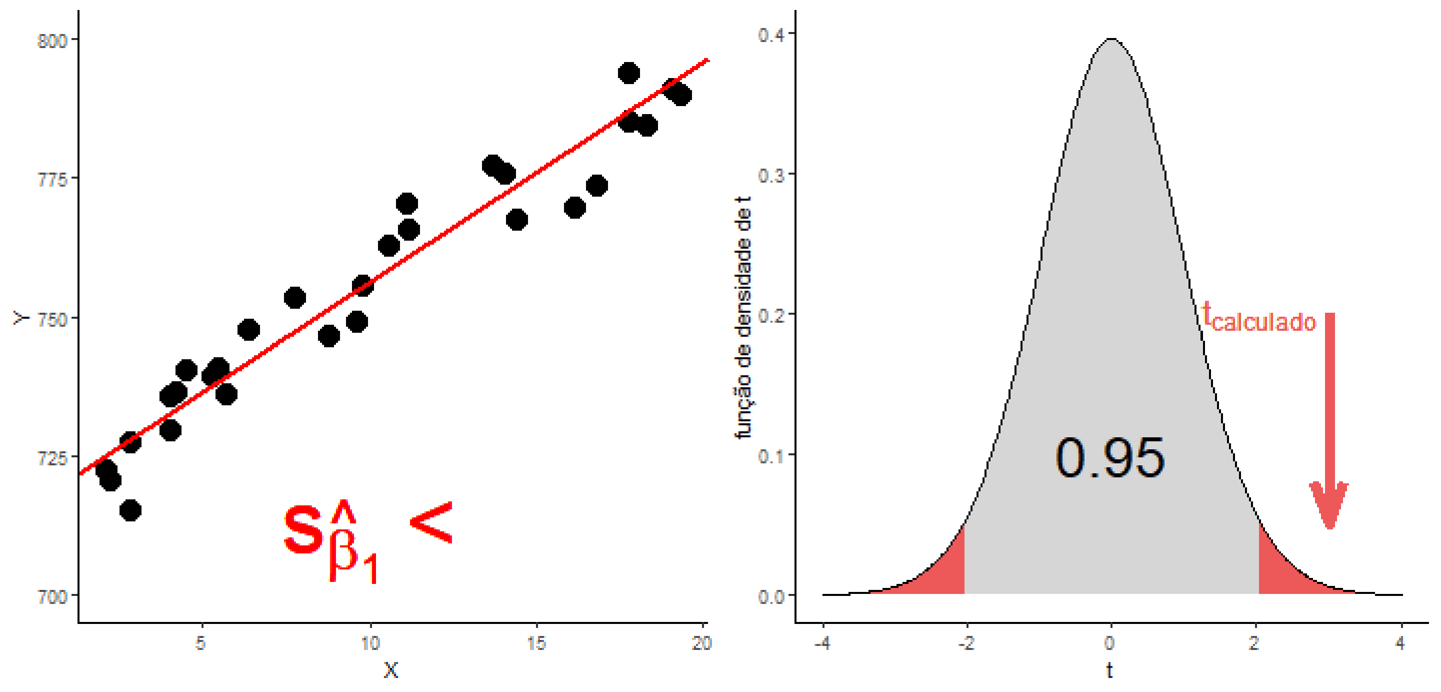
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende da **variância residual** - $s_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{SQ_X}}$

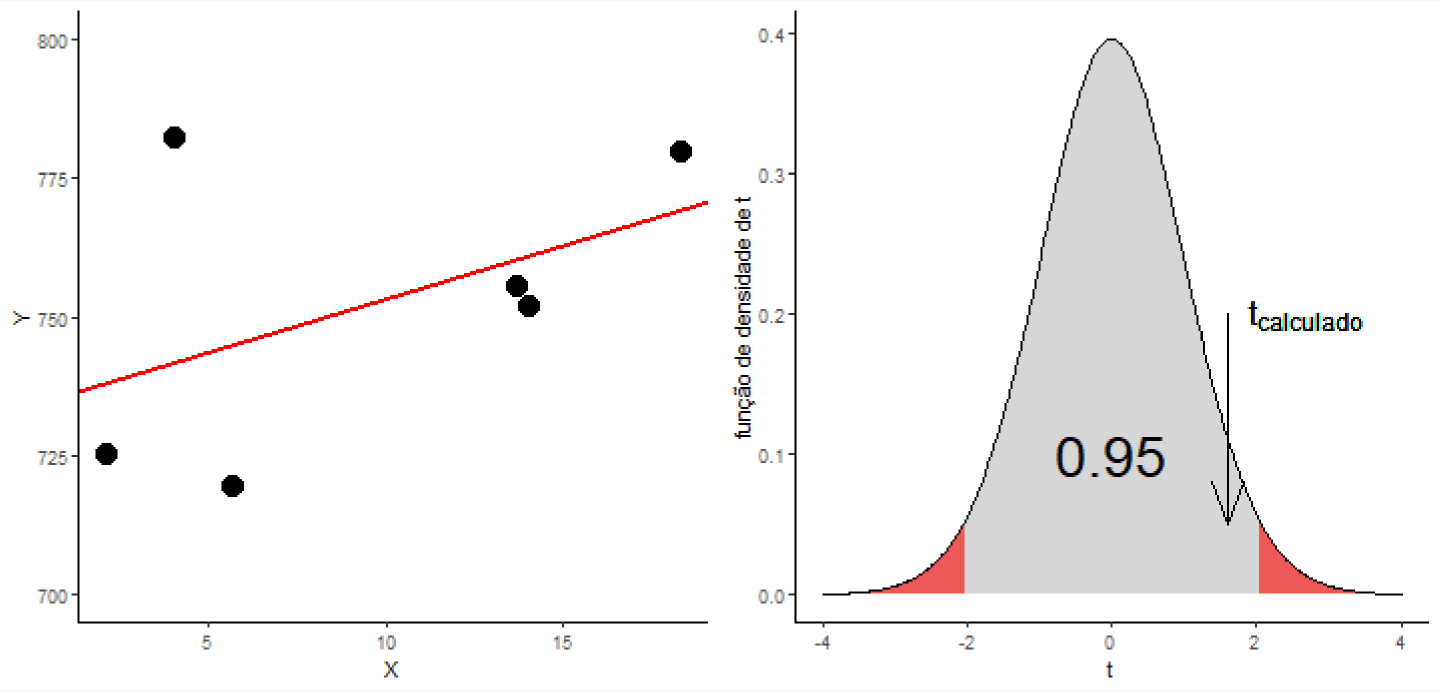
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende do tamanho da amostra - n

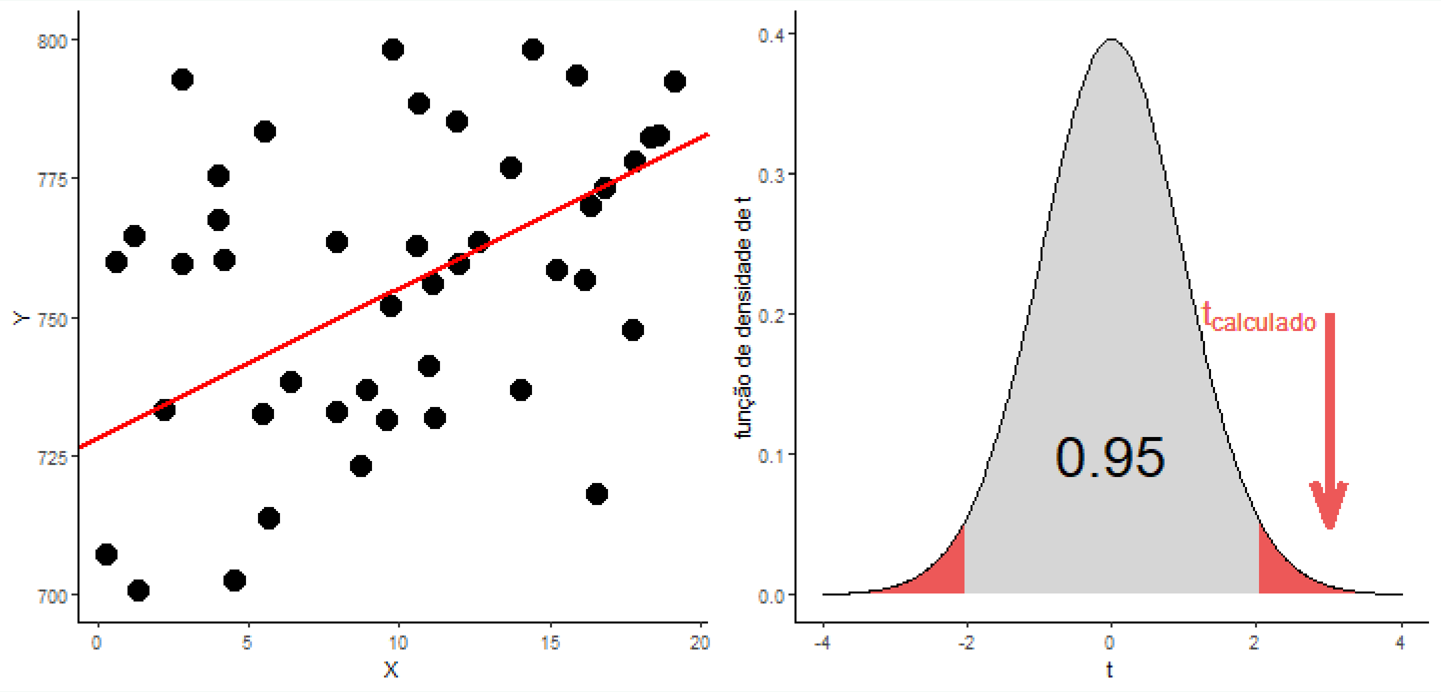
$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses para β_1

$t_{calculado}$ depende do tamanho da amostra - n

$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$



3. Teste de hipóteses: diversidade de espécies e vazão

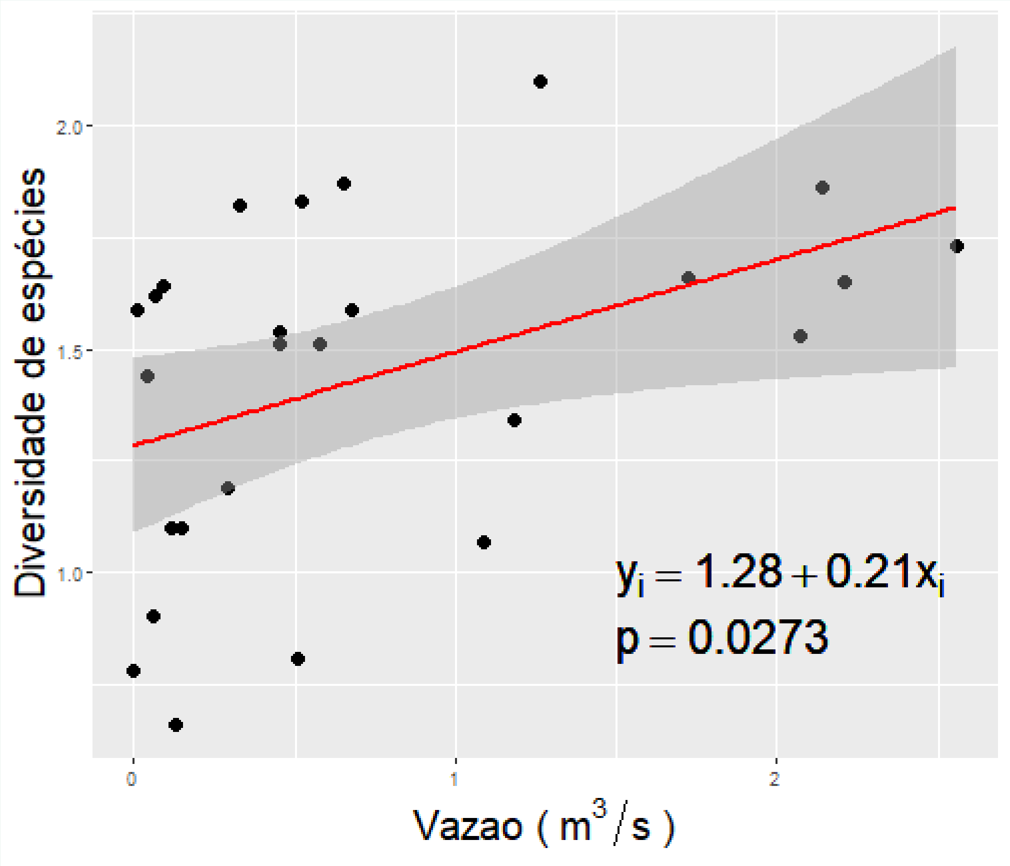
Na figura ao lado, os coeficientes de regressão foram estimados pelo MMQ em $\hat{\beta}_0 = 1.28$ e $\hat{\beta}_1 = 0.21$.

O valor de t foi:

$$t_{calculado} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} = \frac{0.21 - 0}{0.089} = 2.351$$

Embora exista uma alta variabilidade ao redor da reta de regressão o valor de p associado a este resultado foi $p = 0.0273$, o que se interpretado ao nível de significância $\alpha = 0,05$ nos leva a rejeitar H_0.

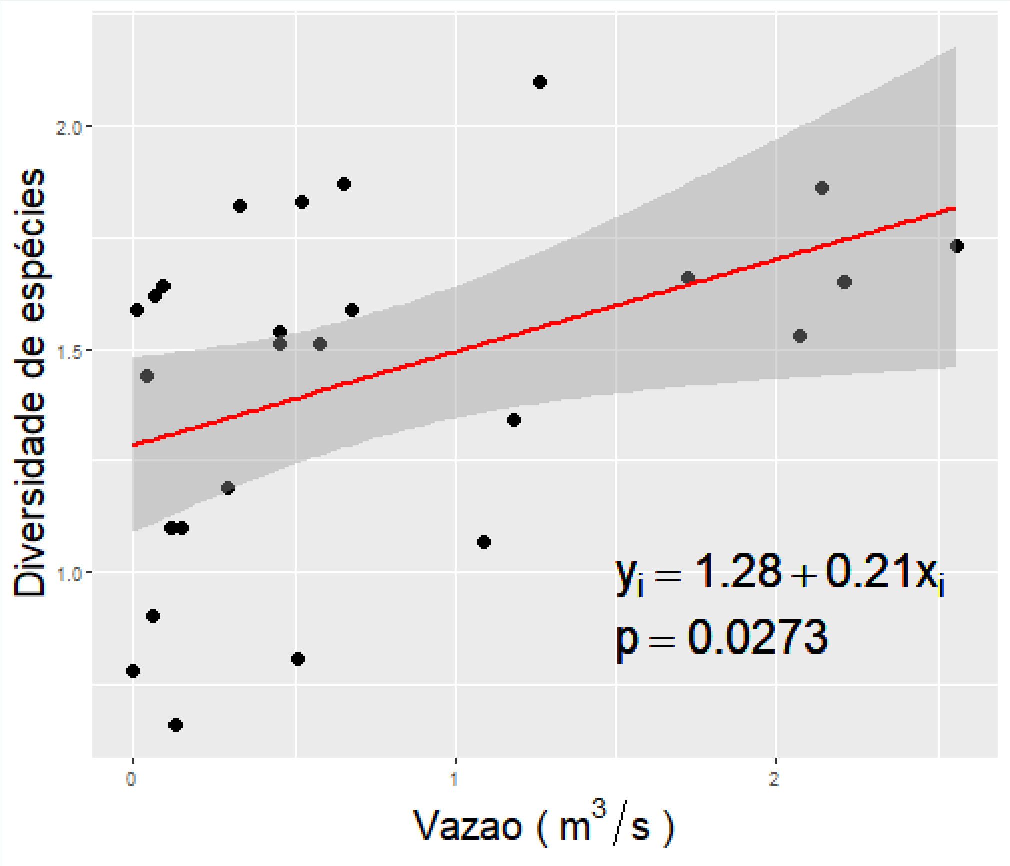
Nossa conclusão é de que existe uma relação crescente entre a Diversidade de espécies e a vazão dos riachos.



3. Teste de hipóteses: diversidade de espécies e vazão

```
rdiv <- lm(Diversidade ~ Vazao, data = peixes)
summary(rdiv)
```

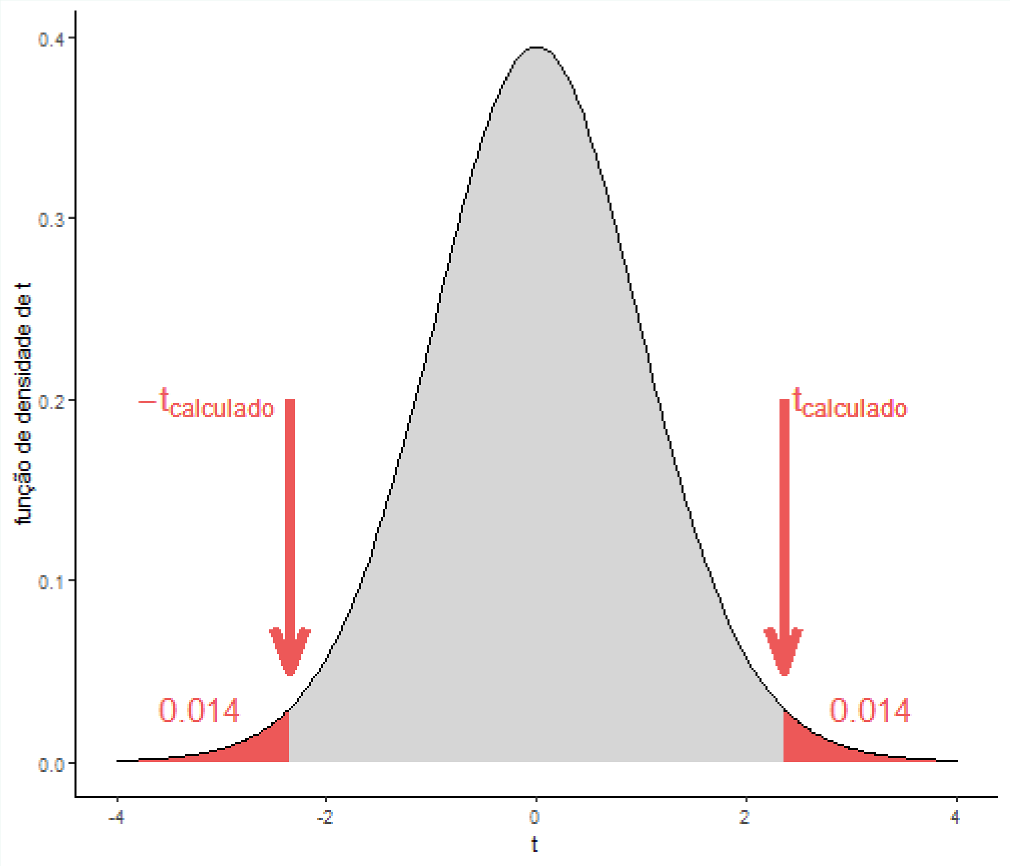
```
##
## Call:
## lm(formula = Diversidade ~ Vazao, data = peixes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65179 -0.20498  0.06043  0.26830  0.55249
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.28467    0.09508  13.512 1.03e-12 ***
## Vazao        0.20861    0.08873   2.351   0.0273 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3486 on 24 degrees of freedom
## Multiple R-squared:  0.1872,    Adjusted R-squared:  0.1533
## F-statistic: 5.527 on 1 and 24 DF,  p-value: 0.02728
```



3. Teste de hipóteses: diversidade de espécies e vazão

```
rdiv <- lm(Diversidade ~ Vazao, data = peixes)
summary(rdiv)
```

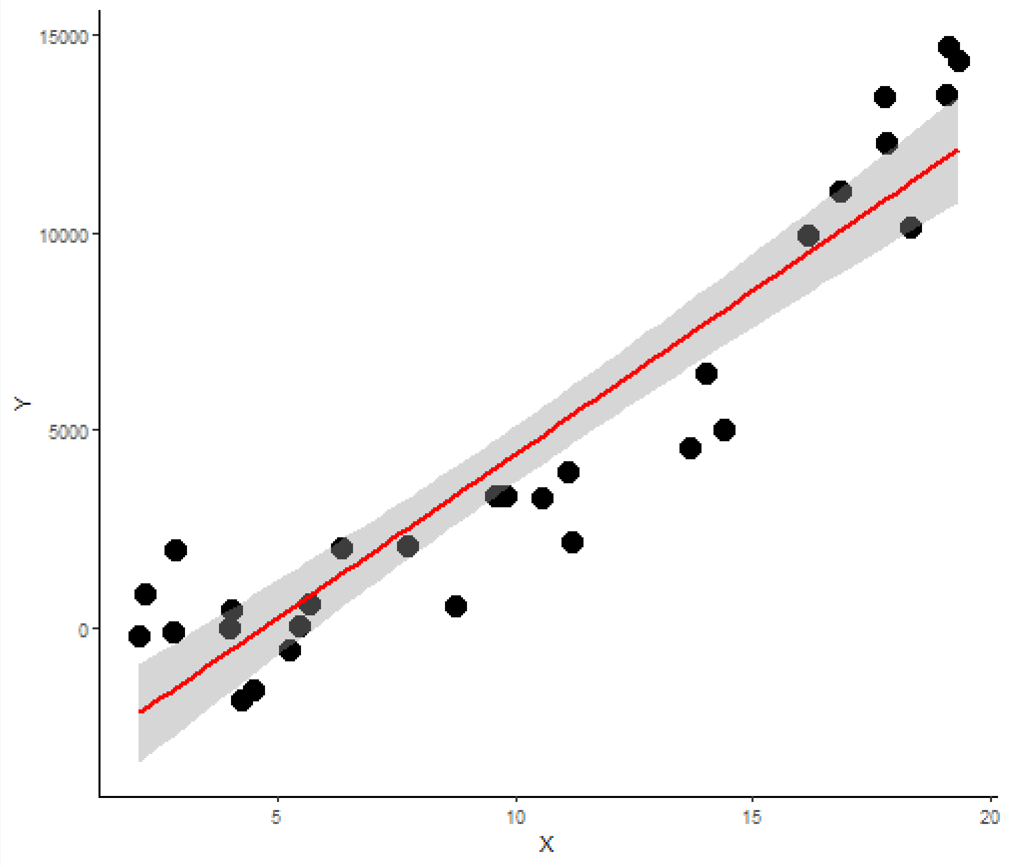
```
##
## Call:
## lm(formula = Diversidade ~ Vazao, data = peixes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65179 -0.20498  0.06043  0.26830  0.55249
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.28467    0.09508  13.512 1.03e-12 ***
## Vazao        0.20861    0.08873   2.351   0.0273 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3486 on 24 degrees of freedom
## Multiple R-squared:  0.1872,    Adjusted R-squared:  0.1533
## F-statistic: 5.527 on 1 and 24 DF,  p-value: 0.02728
```



4. Pressupostos da regressão linear

Ao realizar uma regressão linear simples, devemos assumir como verdadeiros alguns pressupostos.

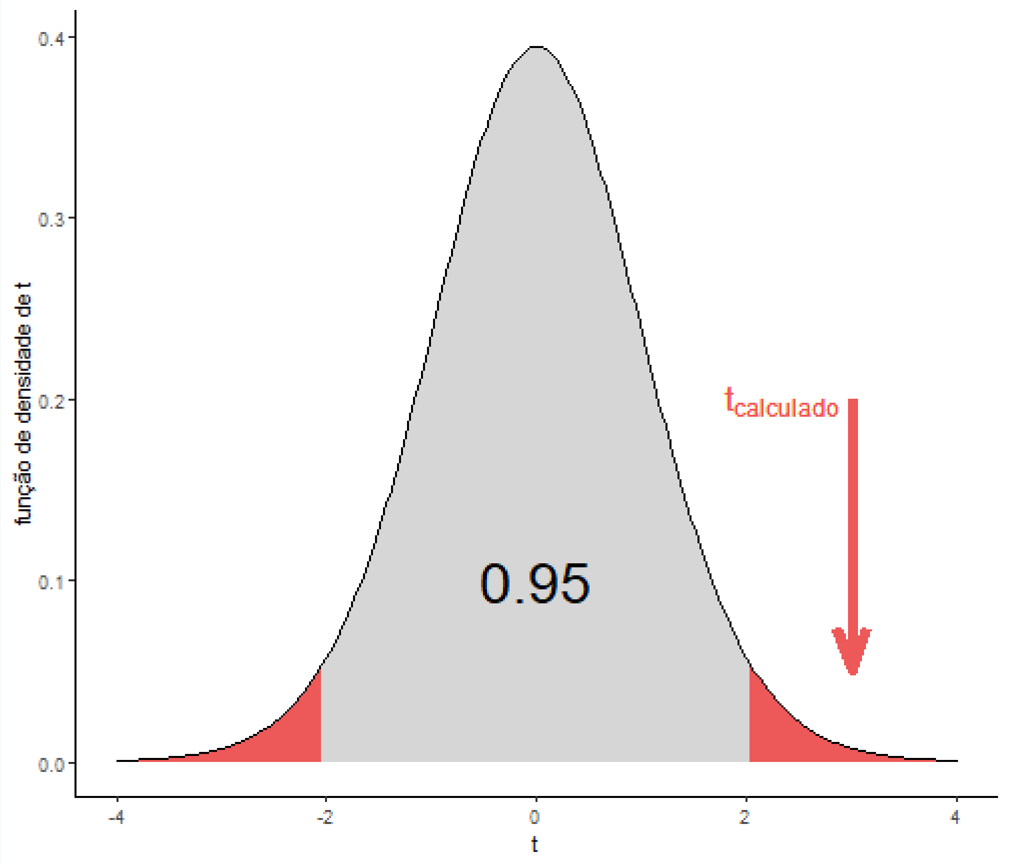
1. O modelo linear descreve adequadamente a relação funcional entre Y e X;
2. Cada par de observação (y_i, x_i) é independente dos demais;
3. A variável X é medida sem erros;
4. Os resíduos têm distribuição normal;
5. A variância residual σ^2 é constante ao longo de X.



4. Pressupostos da regressão linear

Ao realizar uma regressão linear simples, devemos assumir como verdadeiros alguns pressupostos.

1. O modelo linear descreve adequadamente a relação funcional entre Y e X;
2. Cada par de observação (y_i, x_i) é independente dos demais;
3. A variável X é medida sem erros;
4. Os resíduos têm distribuição normal;
5. A variância residual σ^2 é constante ao longo de X.



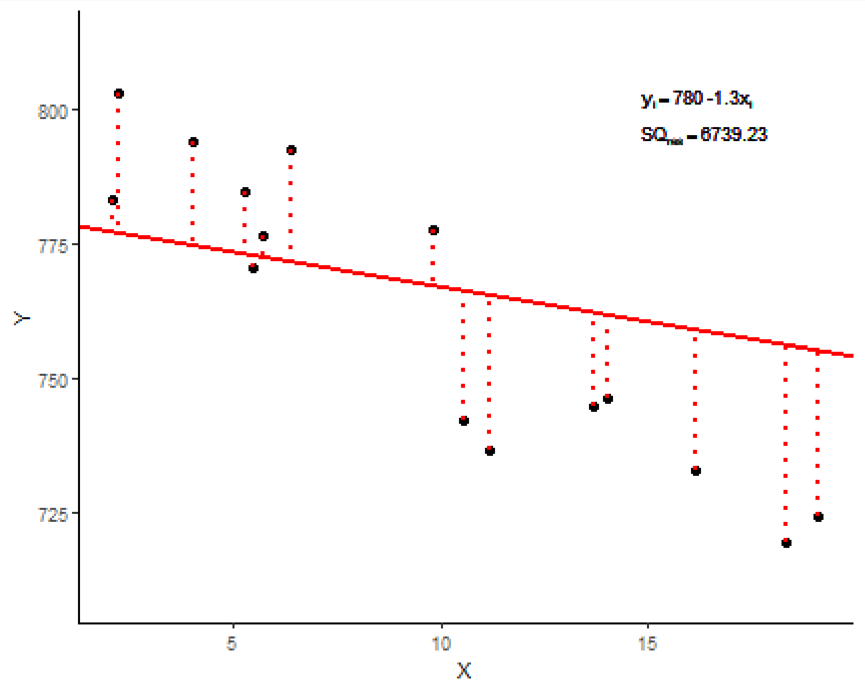
4. Pressupostos da regressão linear

Ao realizar uma regressão linear simples, devemos assumir como verdadeiros alguns pressupostos.

1. O modelo linear descreve adequadamente a relação funcional entre Y e X;
2. Cada par de observação (y_i, x_i) é independente dos demais;
3. A variável X é medida sem erros;
4. Os resíduos têm distribuição normal;
5. A variância residual σ^2 é constante ao longo de X.

$$Y \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$E(Y|x_i) = \mu_i = \beta_0 + \beta_1 x_i$$

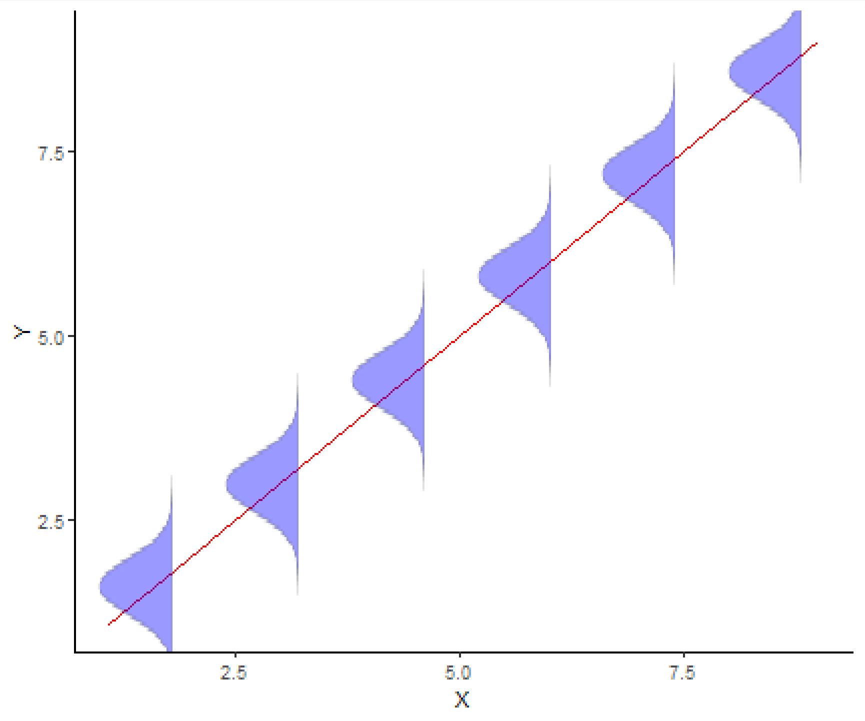


4. Pressupostos da regressão linear

Ao realizar uma regressão linear simples, devemos assumir como verdadeiros alguns pressupostos.

1. O modelo linear descreve adequadamente a relação funcional entre Y e X;
2. Cada par de observação (y_i, x_i) é independente dos demais;
3. A variável X é medida sem erros;
4. Os resíduos têm distribuição normal;
5. A variância residual σ^2 é constante ao longo de X.

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$



4. Pressupostos da regressão linear

Ao realizar uma regressão linear simples, devemos assumir como verdadeiros alguns pressupostos.

1. O modelo linear descreve adequadamente a relação funcional entre Y e X;
2. Cada par de observação (y_i, x_i) é independente dos demais;
3. A variável X é medida sem erros;
4. Os resíduos têm distribuição normal;
5. A variância residual σ^2 é constante ao longo de X.

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$