

# **ESTATÍSTICA PARA SAÚDE COLETIVA**

## **Aula 14**

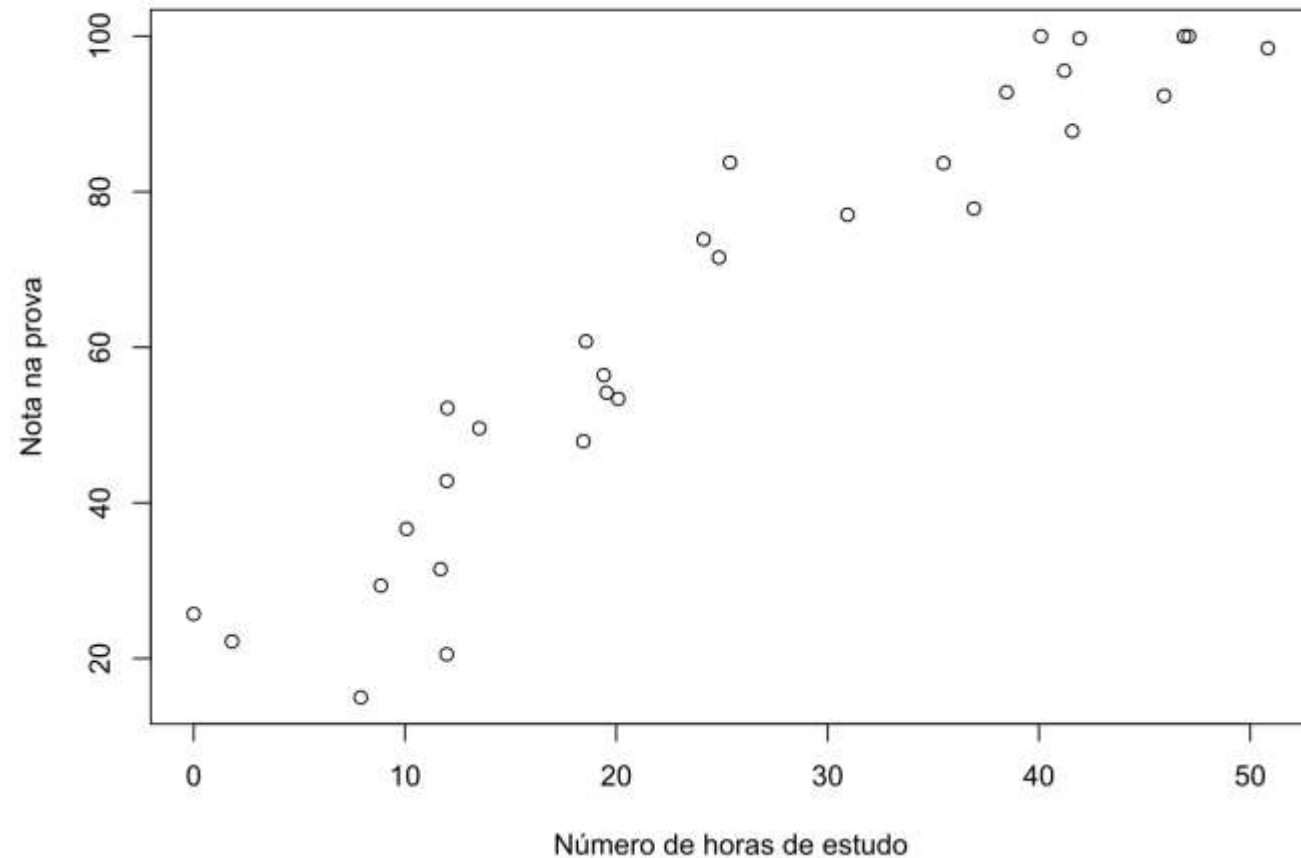
# Seminários de hoje

Nomes (Aluno ou Dupla)	Data seminário	Nome do artigo
ANDERSON SILVA	15/10/2020	Características da epidemia de dengue em Pinhalzinho, Santa Catarina.
ROSA MARIA GARCIA + ANA LEITE	15/10/2020	Consumo de alimentos fora do lar no Brasil segundo locais de aquisição
TAINARA PRADELLA + ADRIANA BARROS	15/10/2020	Doenças Crônicas Não Transmissíveis e fatores de risco e proteção em adultos com ou sem plano de saúde
PAMELA MONTE CRUZ	15/10/2020	Prevalência de transtornos mentais comuns e fatores associados em moradores da área urbana de São Paulo, Brasil

# Revisão: Correlações

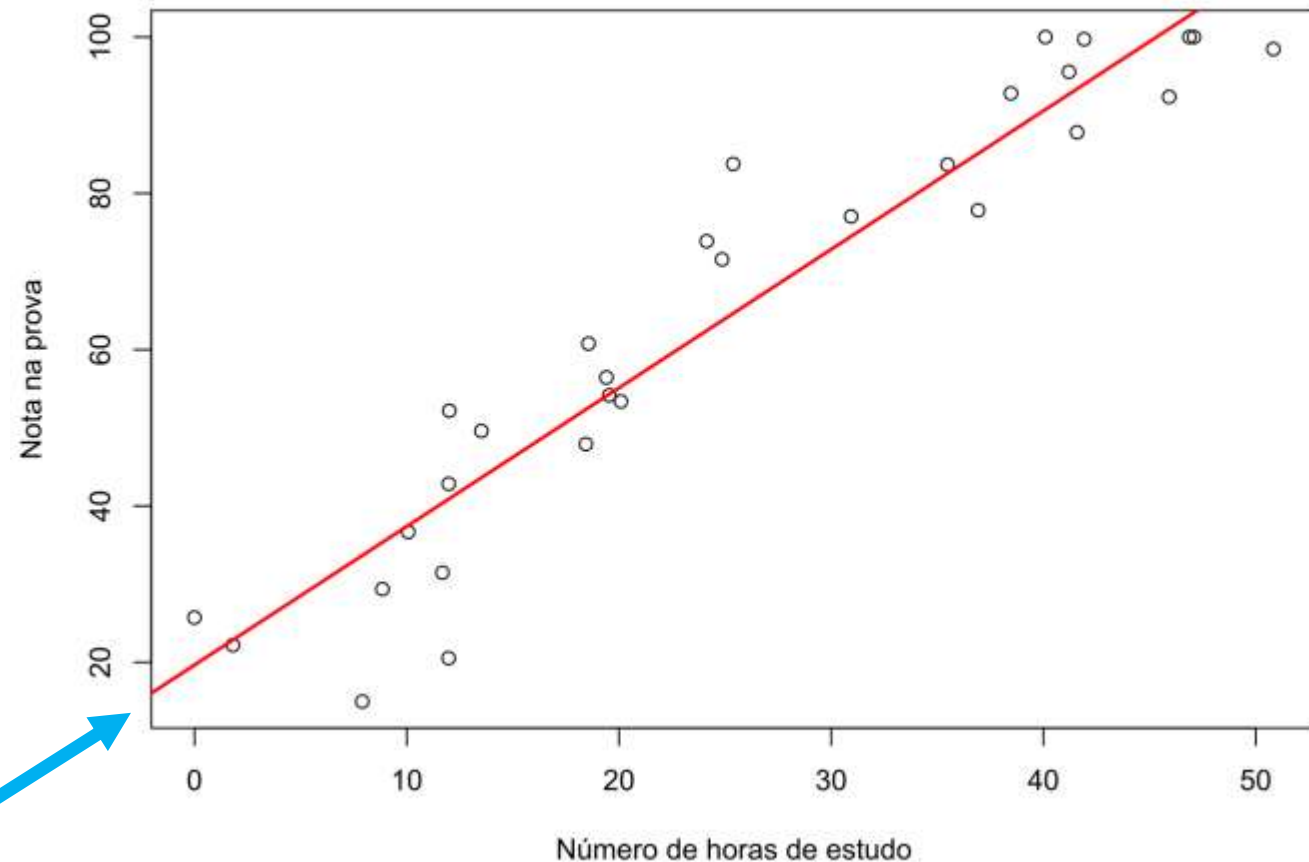
# Revisão correlação

- Compare o numero de horas de estudo vs. nota em um teste
- Os pontos indicam a nota e numero de horas de estudo para cada aluno



# Revisão correlação

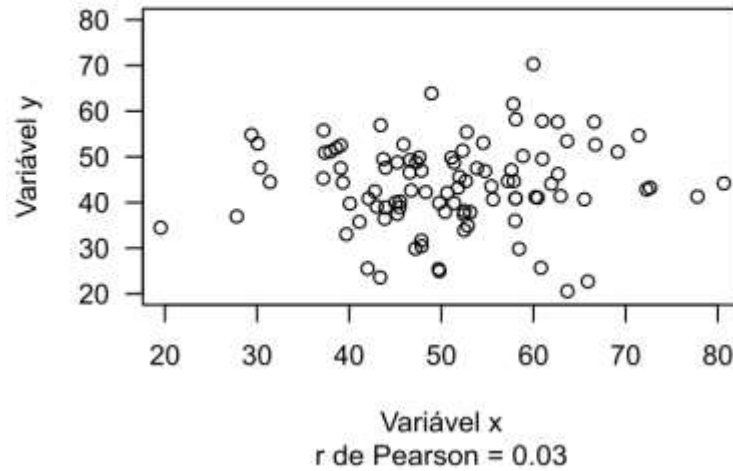
- Compare o numero de horas de estudo vs. nota em um teste
- Os pontos indicam a nota e numero de horas de estudo para cada aluno



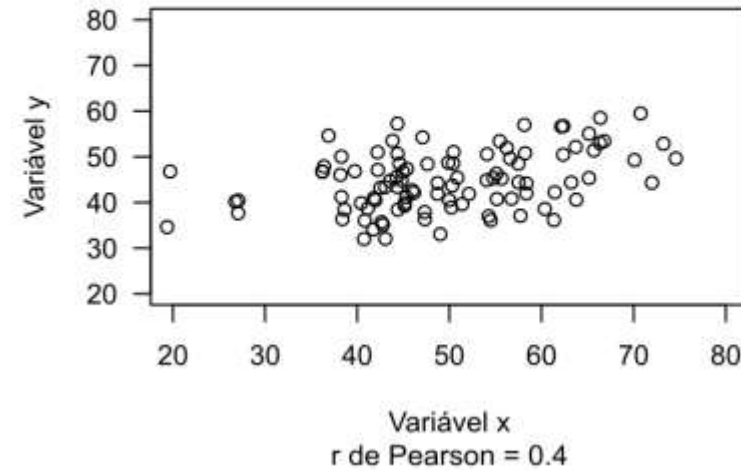
Quanto mais próximo os dados estiverem  
dessa reta imaginaria, maior será a correlação

# Revisão correlação

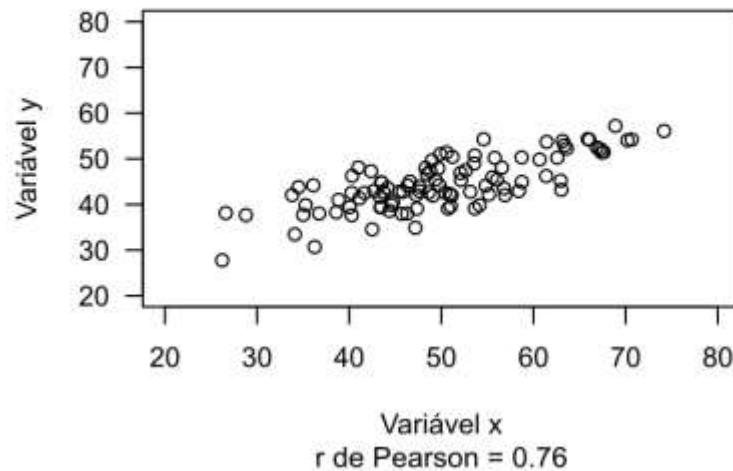
**Correlação pequena**



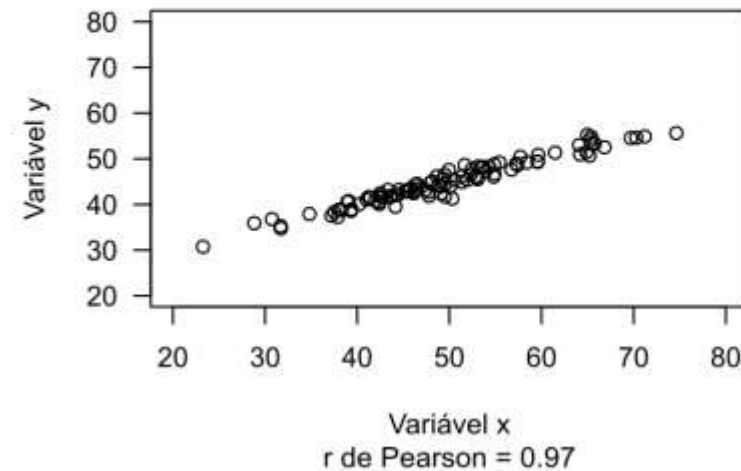
**Correlação fraca**



**Correlação moderada**

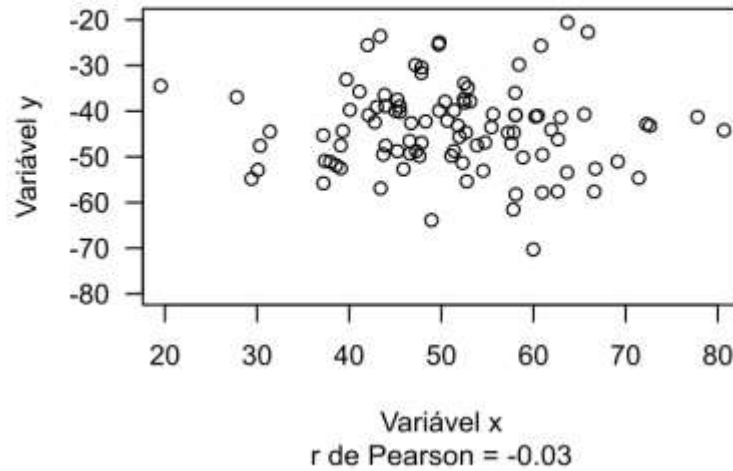


**Correlação forte**

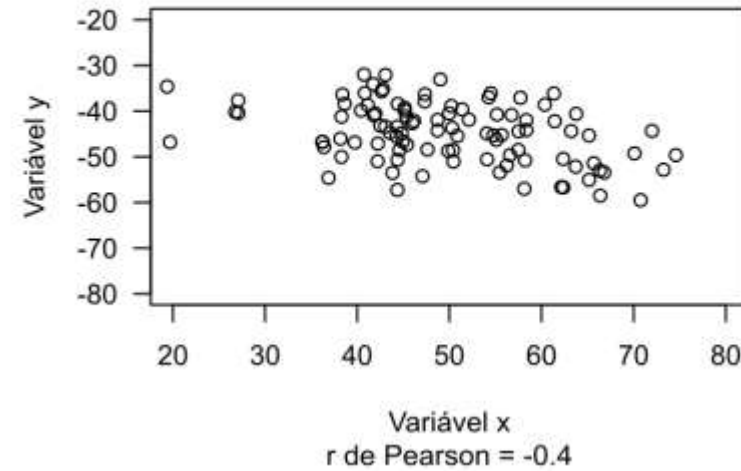


# Revisão correlação

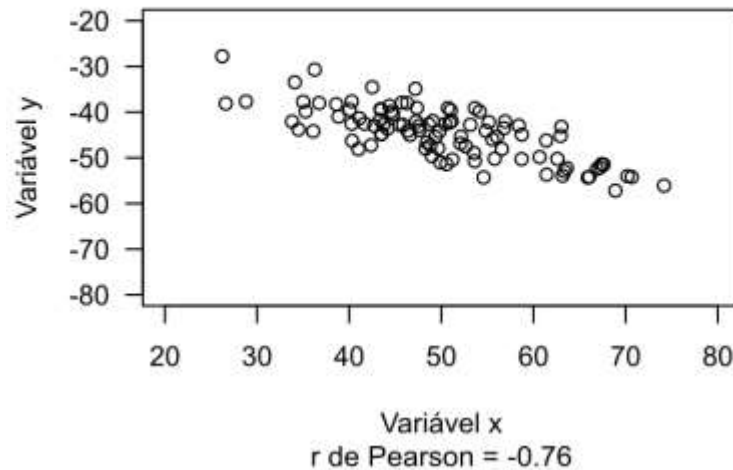
**Correlação pequena**



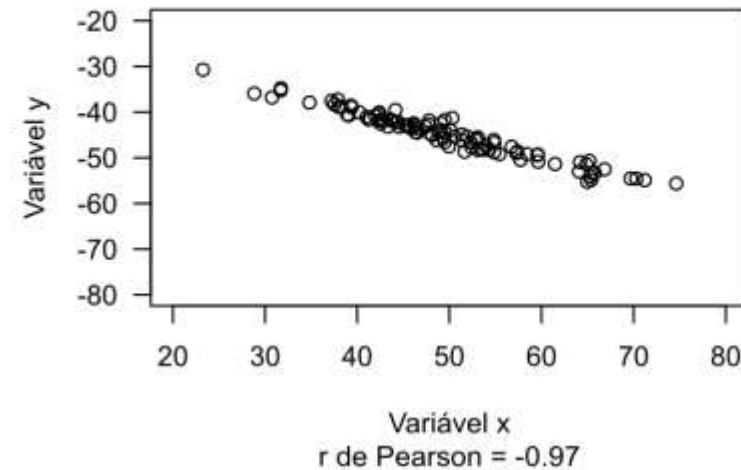
**Correlação fraca**



**Correlação moderada**



**Correlação forte**



Aula de hoje



Regressão linear simples

# Regressão linear simples

Variável resposta

Variável preditora

1 variável  
quantitativa

1 variável  
quantitativa



# Regressões



- Pense nos modelos matemáticos como manequins, onde você ajusta seus dados a uma estrutura previamente estabelecida

$$y_i = \beta_0 + \beta_1 x_{i1}$$

Equação que define valores esperados de uma dada variável resposta ( $y$ ), dado a observação de uma variável preditora ( $x$ )

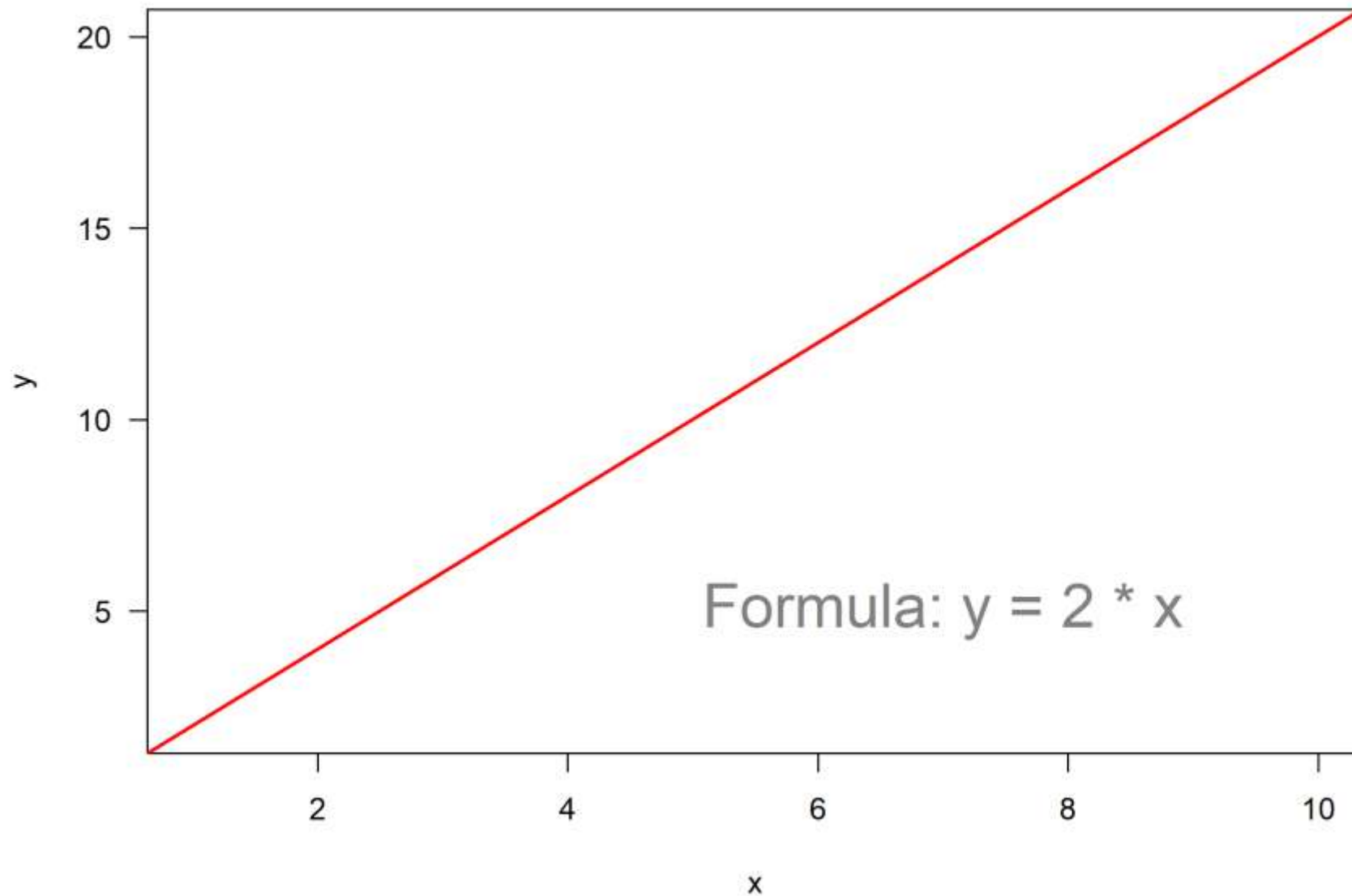
Para que servem modelos preditivos?



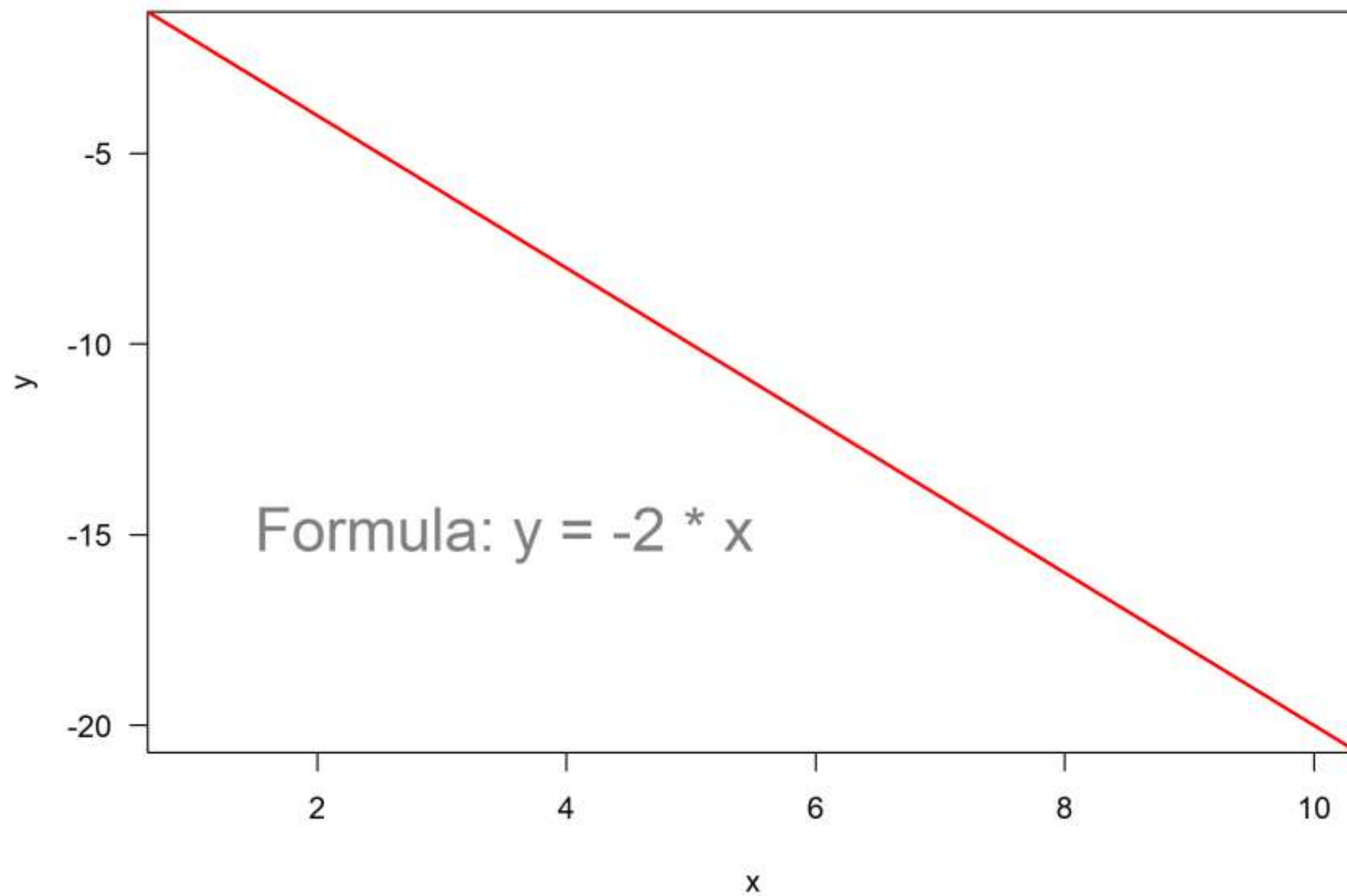
$$\beta_1$$

$$y_i = \beta_0 + \beta_1 x_{i1}$$

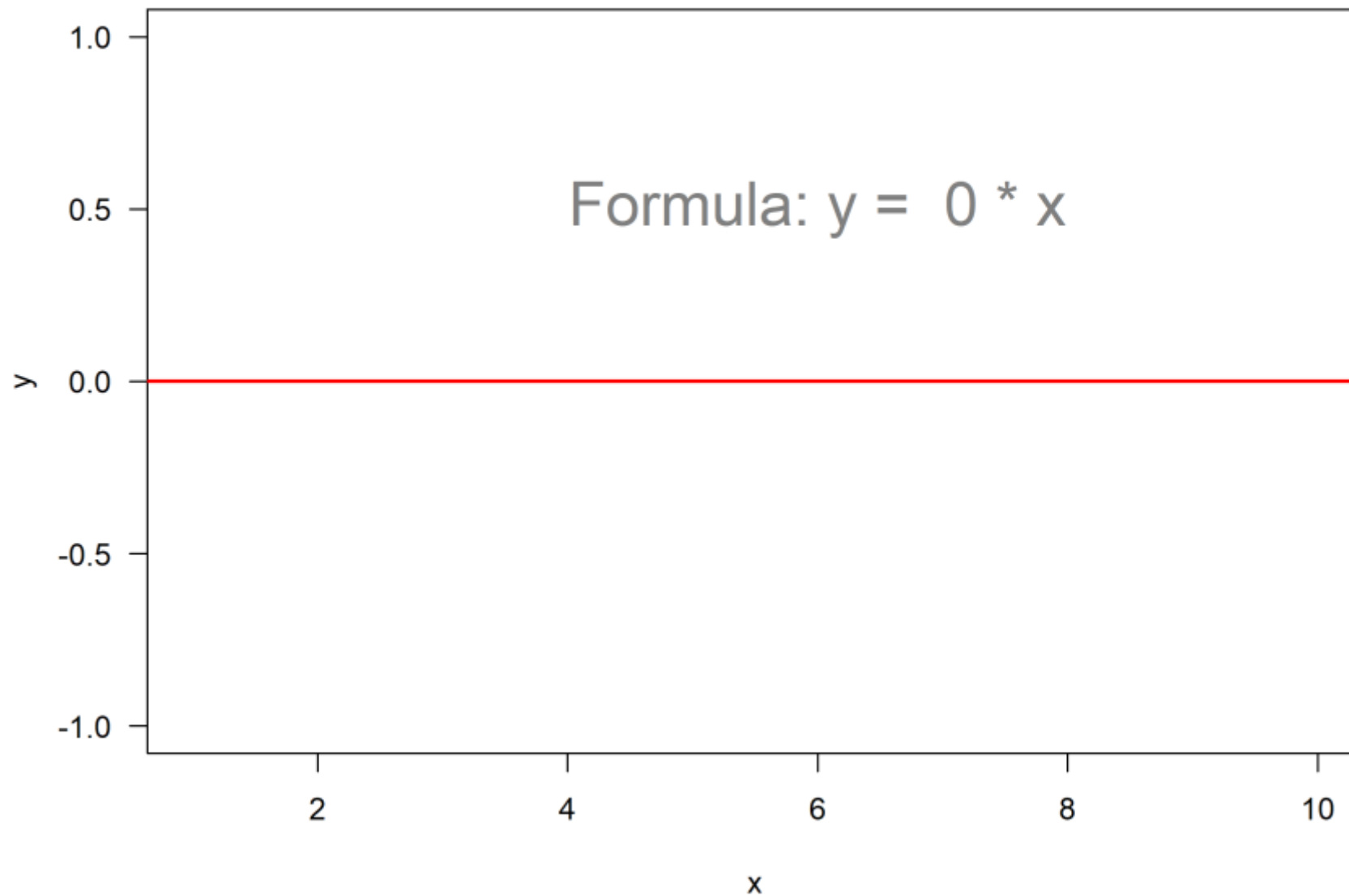
# Regressão linear $\beta_1$



# Regressão linear $\beta_1$



# Regressão linear $\beta_1$





# Regressão linear $\beta_1$

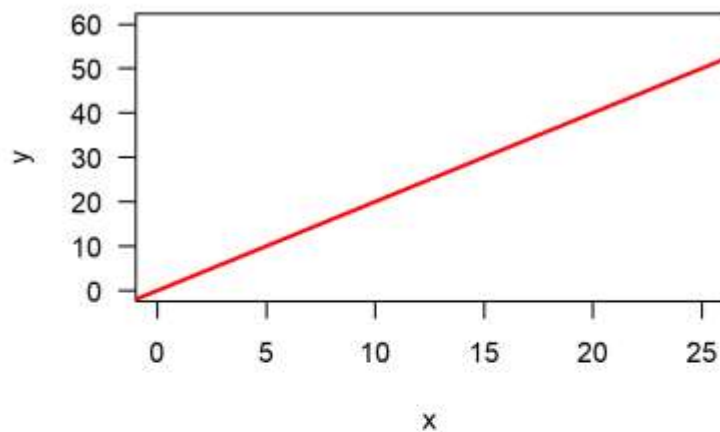
- Em regressão linear, chamamos a constante que multiplica o valor de  $x$  por  $\beta_1$
- Nos exemplos
  - Em  $Y = 2 * x$  o valor de  $\beta_1 = 2$
  - Em  $Y = -2 * x$  o valor de  $\beta_1 = -2$
  - Em  $Y = 0 * x$  o valor de  $\beta_1 = 0$

$$\beta_0$$

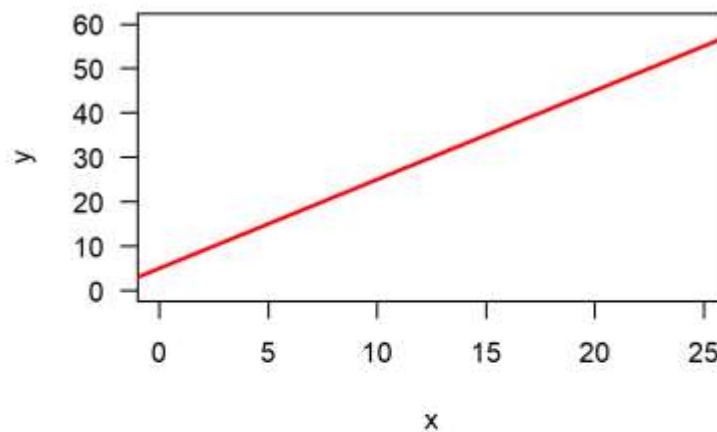
$$y_i = \beta_0 + \beta_1 x_{i1}$$

# Regressão linear $\beta_0$

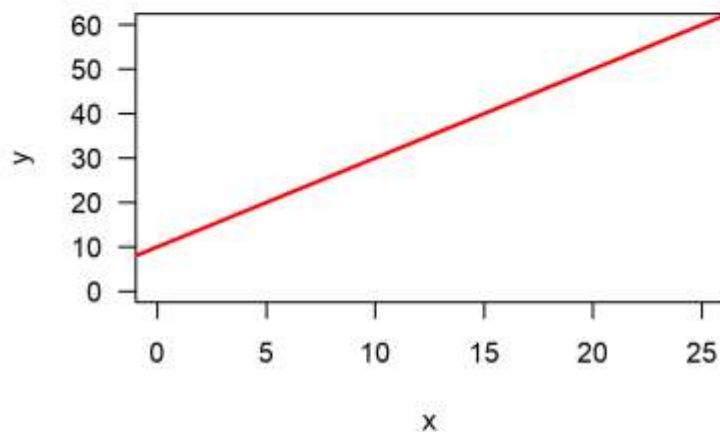
Formula:  $y = 0 + 2 * x$



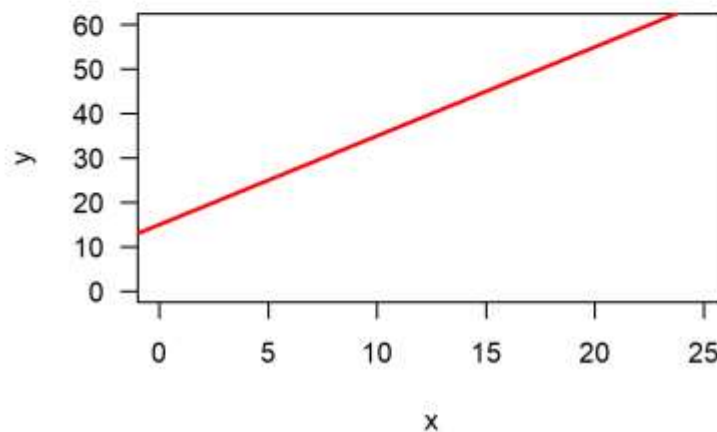
Formula:  $y = 5 + 2 * x$



Formula:  $y = 10 + 2 * x$



Formula:  $y = 15 + 2 * x$

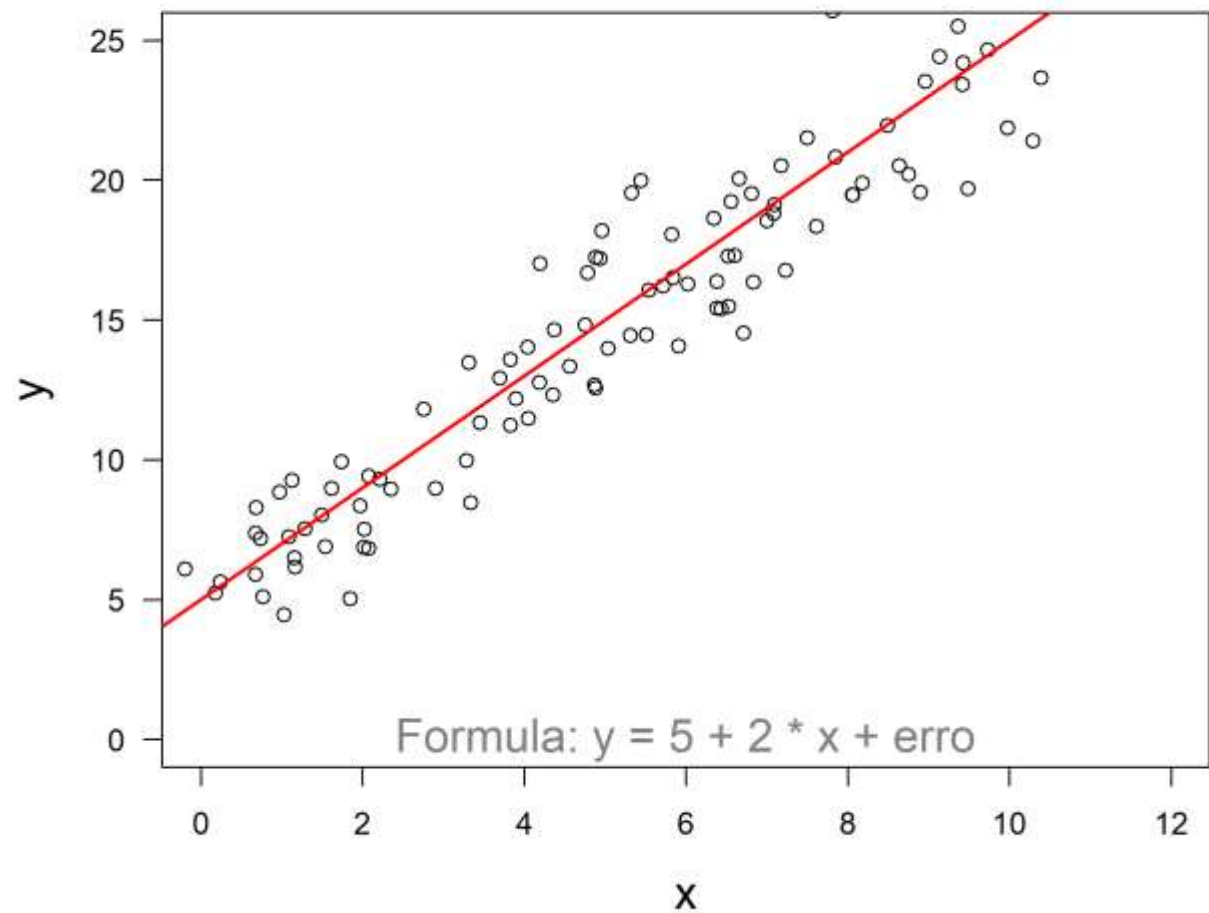
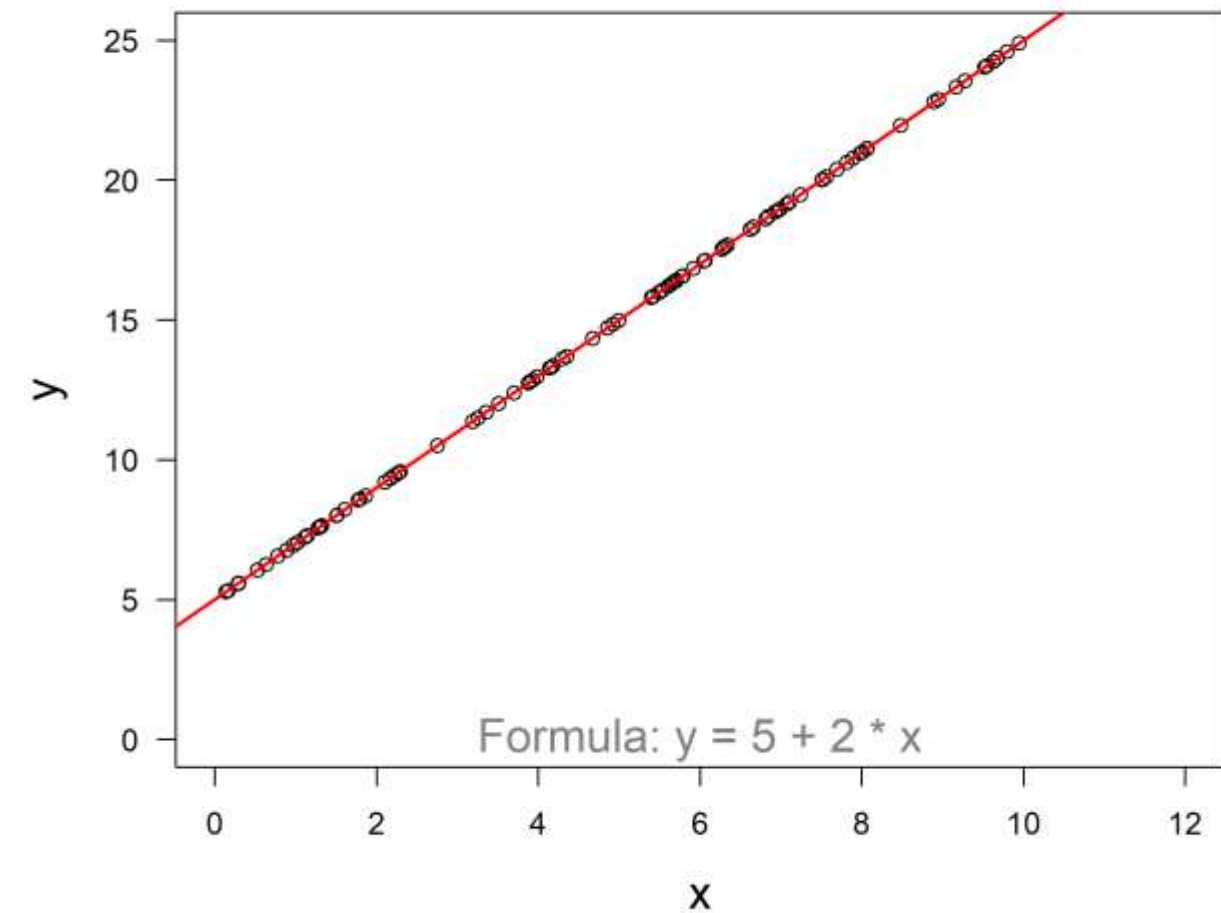


# Regressão linear $\beta_0$

- Em regressão linear, chamamos o valor que corta o eixo y de  $\beta_0$
- Nos exemplos
  - Em  $Y = 0 + 2 * x$  o valor de  $\beta_0 = 0$
  - Em  $Y = 5 + 2 * x$  o valor de  $\beta_0 = 5$
  - Em  $Y = 10 + 2 * x$  o valor de  $\beta_0 = 10$
  - Em  $Y = 15 + 2 * x$  o valor de  $\beta_0 = 15$

Erro ( $\varepsilon$ )

# Erro ( $\varepsilon$ )



Coeficiente de determinação ( $R^2$ )

# Coeficiente de determinação ( $R^2$ )

- $R^2$  varia entre 0 e 1
- Coeficiente de determinação é a proporção da variação de  $y$  explicada pela variação de  $x$
- Quanto maior o valor de  $R^2$ , mais os pontos estarão espalhados ao redor da reta

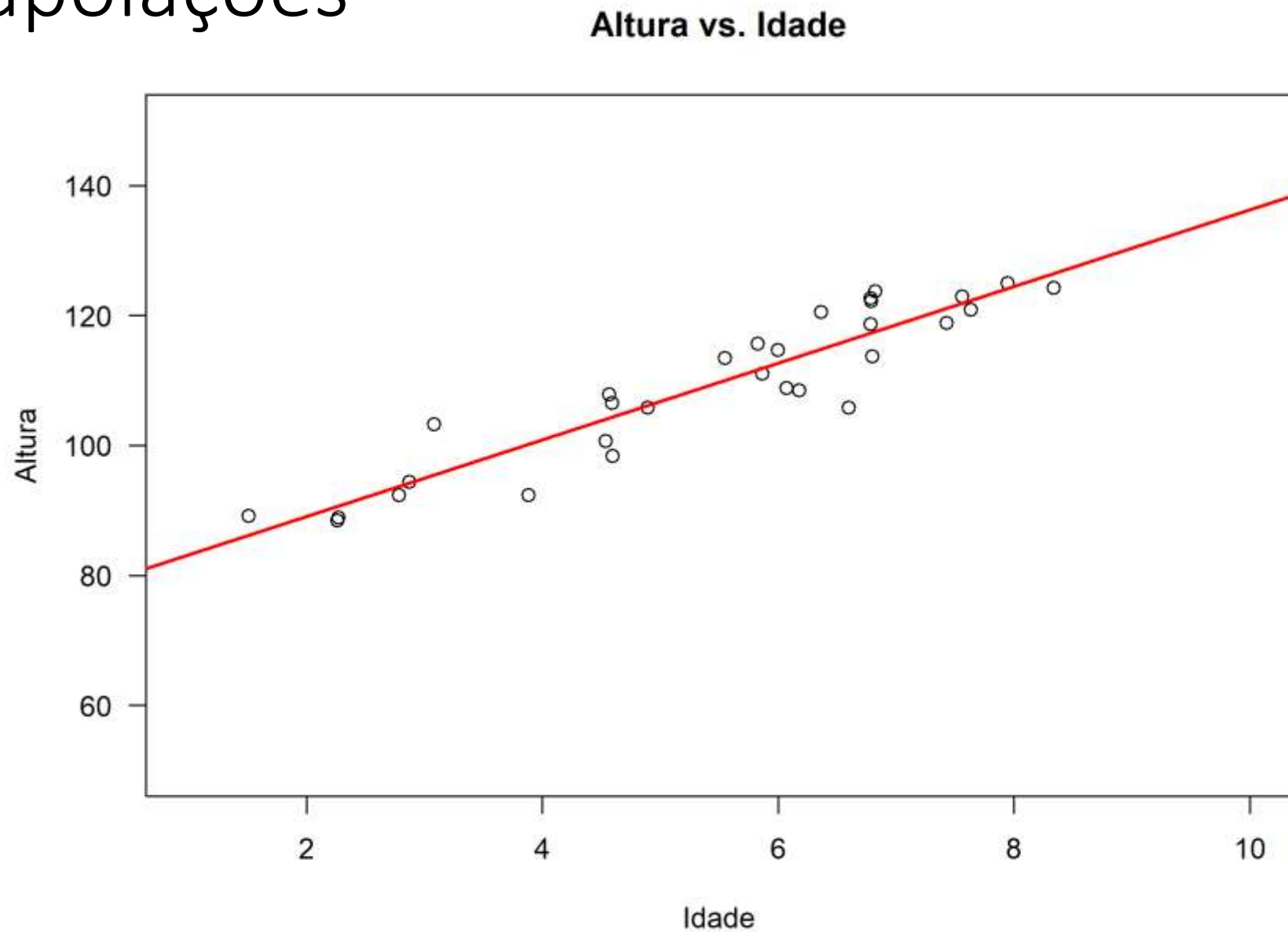


Extrapolações

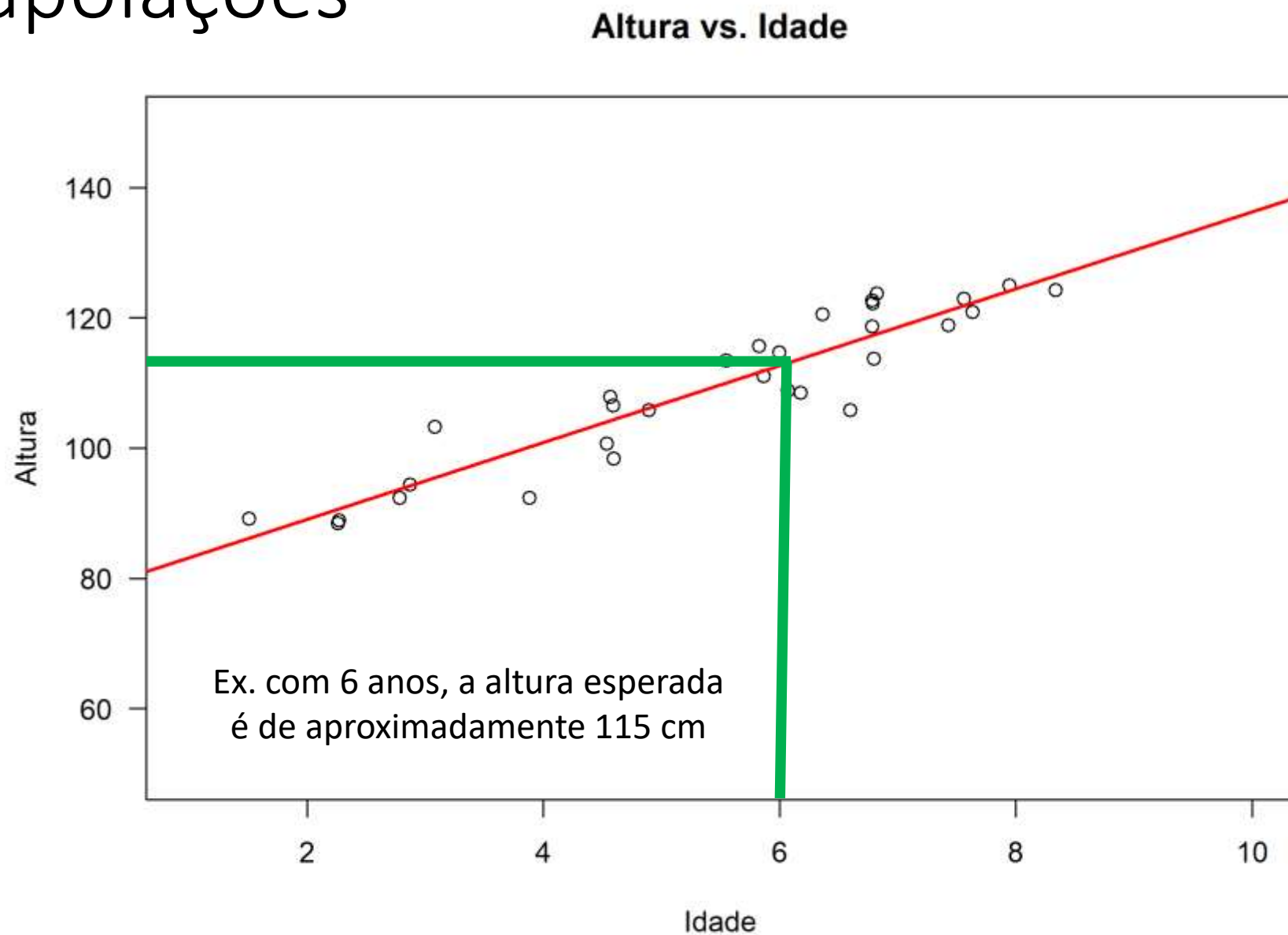
# Extrapolações

- A reta de regressão linear serve portanto para você prever valores de  $y$ , dado uma observação de  $x$ .
- Porém, você só pode fazer extrapolações dentro dos limites que você estudou.
- Por exemplo, em uma relação de idade vs altura, se você estudou crianças de 2 à 8 anos, você não pode usar o mesmo modelo para prever a altura de alguém de 80 anos.

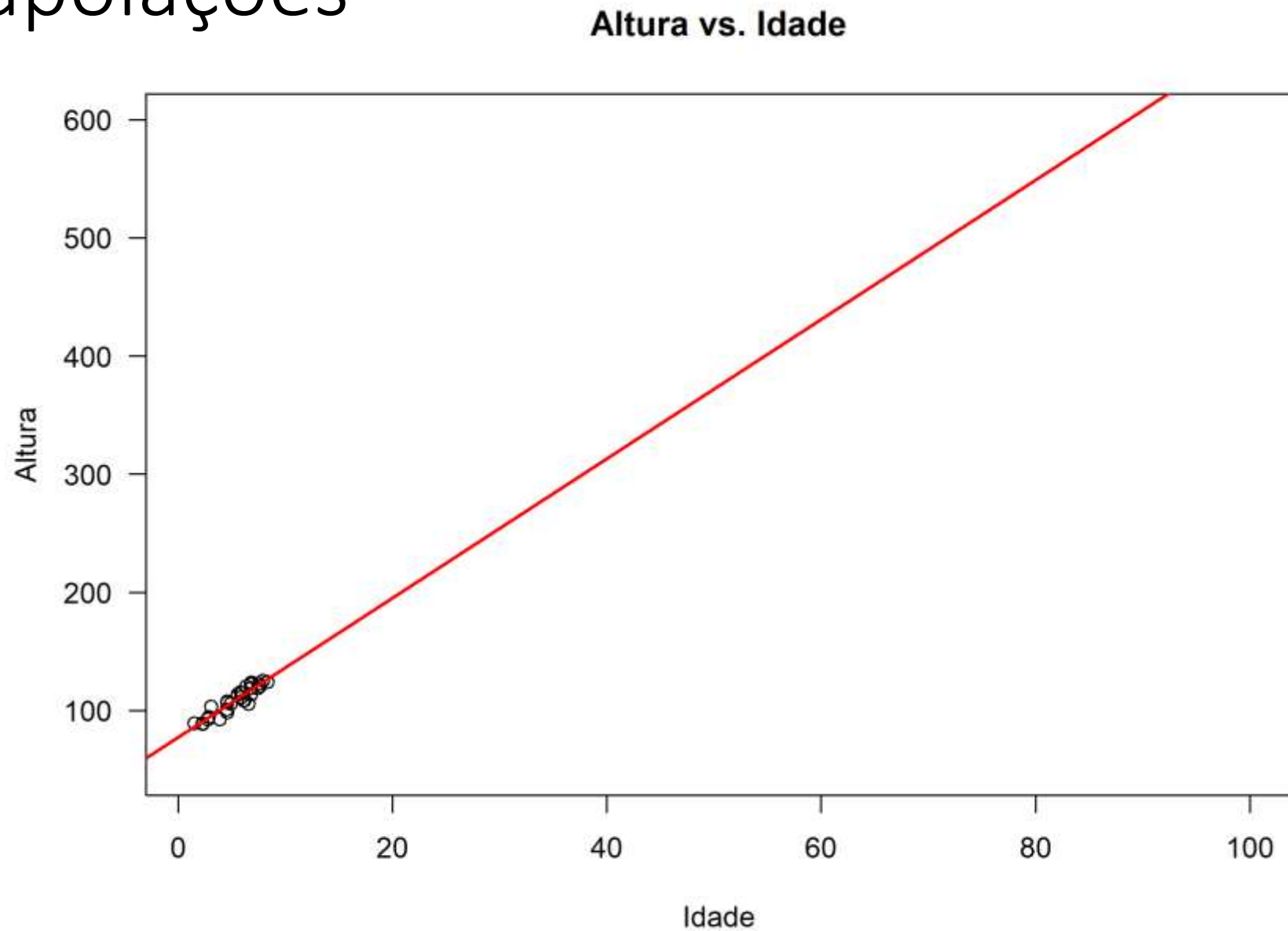
# Extrapolações



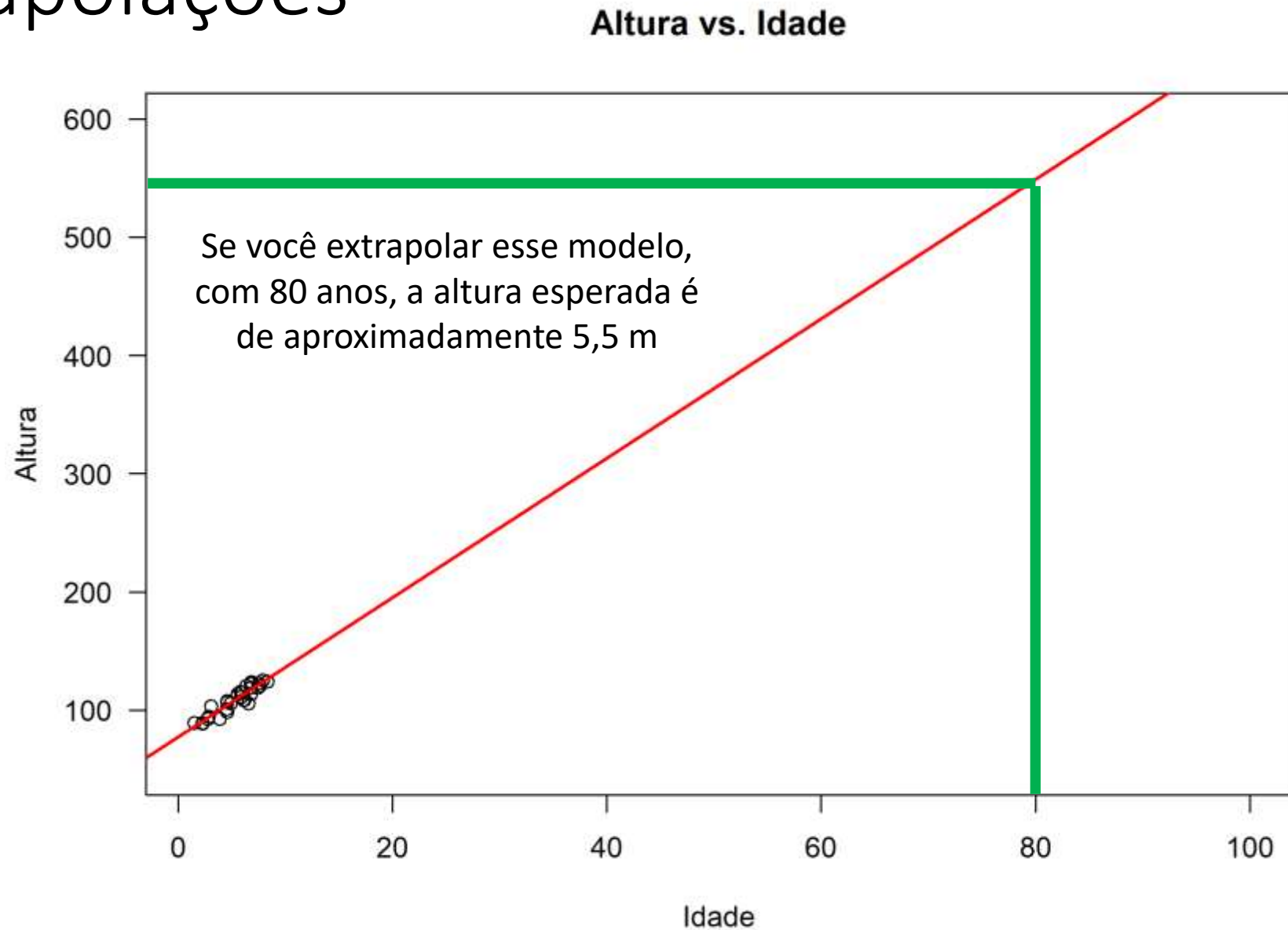
# Extrapolações



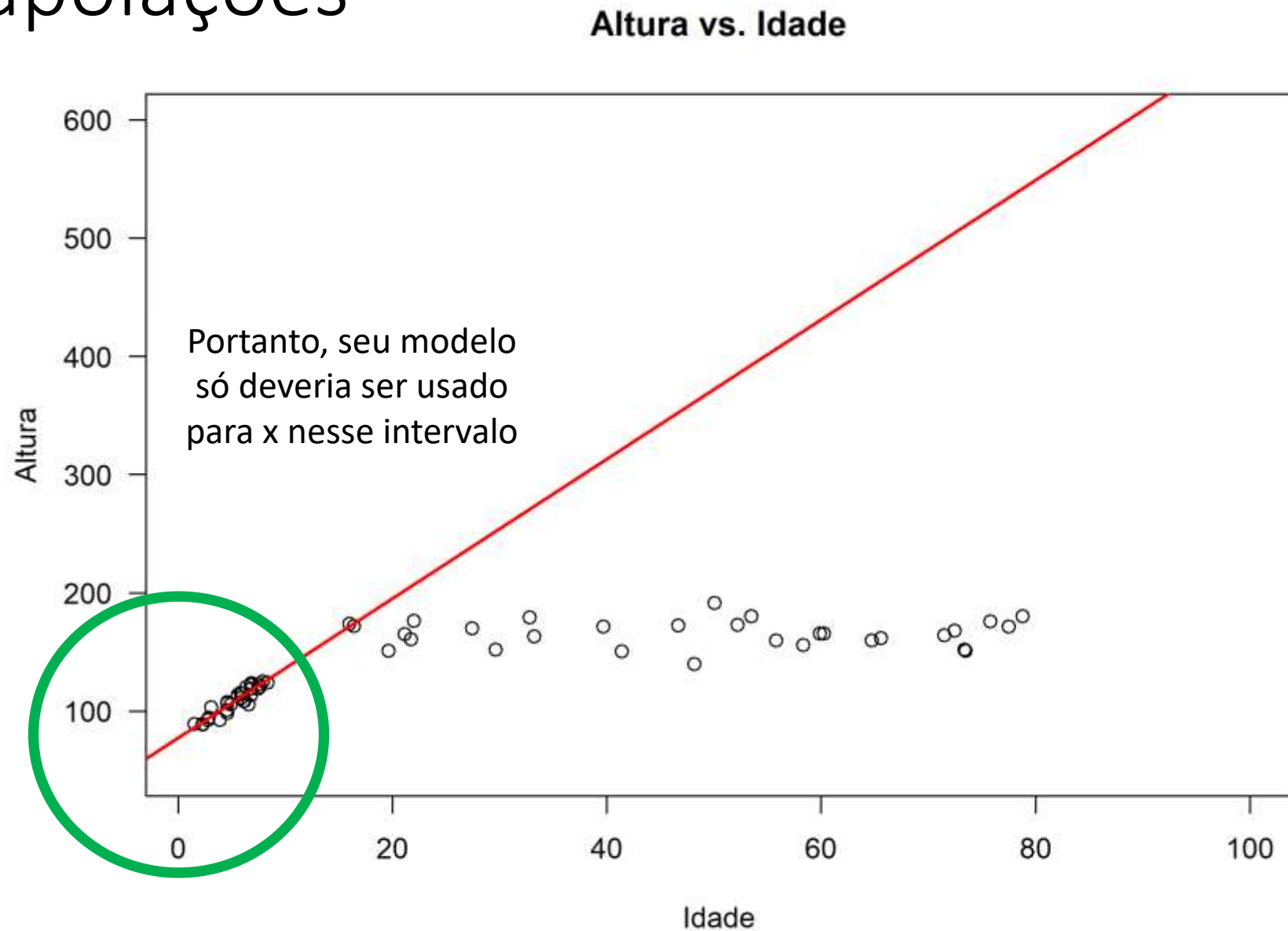
# Extrapolações



# Extrapolações



# Extrapolações



Aplicação de uma regressão linear



# Interpretar resultado

```
> summary(modelo)
```

```
call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-10.3226	-2.1808	-0.0269	3.4070	7.8274

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	77.2792	2.5127	30.76	< 2e-16	***
x	5.8954	0.4378	13.47	9.37e-14	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.411 on 28 degrees of freedom
```

```
Multiple R-squared:  0.8663,    Adjusted R-squared:  0.8615
```

```
F-statistic: 181.3 on 1 and 28 DF,  p-value: 9.365e-14
```

# Interpretar resultado

```
> summary(modelo)
```

```
call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-10.3226	-2.1808	-0.0269	3.4070	7.8274

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	77.2792	2.5127	30.76	< 2e-16	***
x	5.8954	0.4378	13.47	9.37e-14	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.411 on 28 degrees of freedom
```

```
Multiple R-squared:  0.8663,    Adjusted R-squared:  0.8615
```

```
F-statistic: 181.3 on 1 and 28 DF,  p-value: 9.365e-14
```

$\beta_0$

# Interpretar resultado

```
> summary(modelo)
```

```
call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-10.3226	-2.1808	-0.0269	3.4070	7.8274

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	77.2792	2.5127	30.76	< 2e-16	***
x	5.8954	0.4378	13.47	9.37e-14	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.411 on 28 degrees of freedom
```

```
Multiple R-squared:  0.8663,    Adjusted R-squared:  0.8615
```

```
F-statistic: 181.3 on 1 and 28 DF,  p-value: 9.365e-14
```

$\beta_1$

# Hipótese testada

- $H_0$ : O coeficiente não é diferente de zero
- $H_1$ : O coeficiente é diferente de zero

# Interpretar resultado

Resultado: P menor que 0.05

Conclusão:  $H_1$  é verdadeira

## Hipóteses testadas

$H_0$ :  $\beta_1$  não é diferente de zero. Portanto, não há associação entre as variáveis.

$H_1$ :  $\beta_1$  é diferente de zero. Portanto, há (sim) associação entre as variáveis.

## Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	77.2792	2.5127	30.76	< 2e-16	***
x	5.8954	0.4378	13.47	9.37e-14	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.411 on 28 degrees of freedom

Multiple R-squared: 0.8663, Adjusted R-squared: 0.8615

F-statistic: 181.3 on 1 and 28 DF, p-value: 9.365e-14

# Coeficiente de determinação ( $R^2$ )

> summary(modelo)

Coeficiente de determinação é a proporção da variação de y explicada pela variação de x

OBS: esse valor varia entre 0 – 1

Residual standard error: 4.411 on 28 degrees of freedom  
Multiple R-squared: 0.8663, Adjusted R-squared: 0.8615  
F-statistic: 181.3 on 1 and 28 DF, p-value: 9.365e-14

Validação de modelos

# Pré-requisitos

- Normalidade
- Verificar se não existem outliers
- O modelo deve passar por uma validação (análise de resíduos)



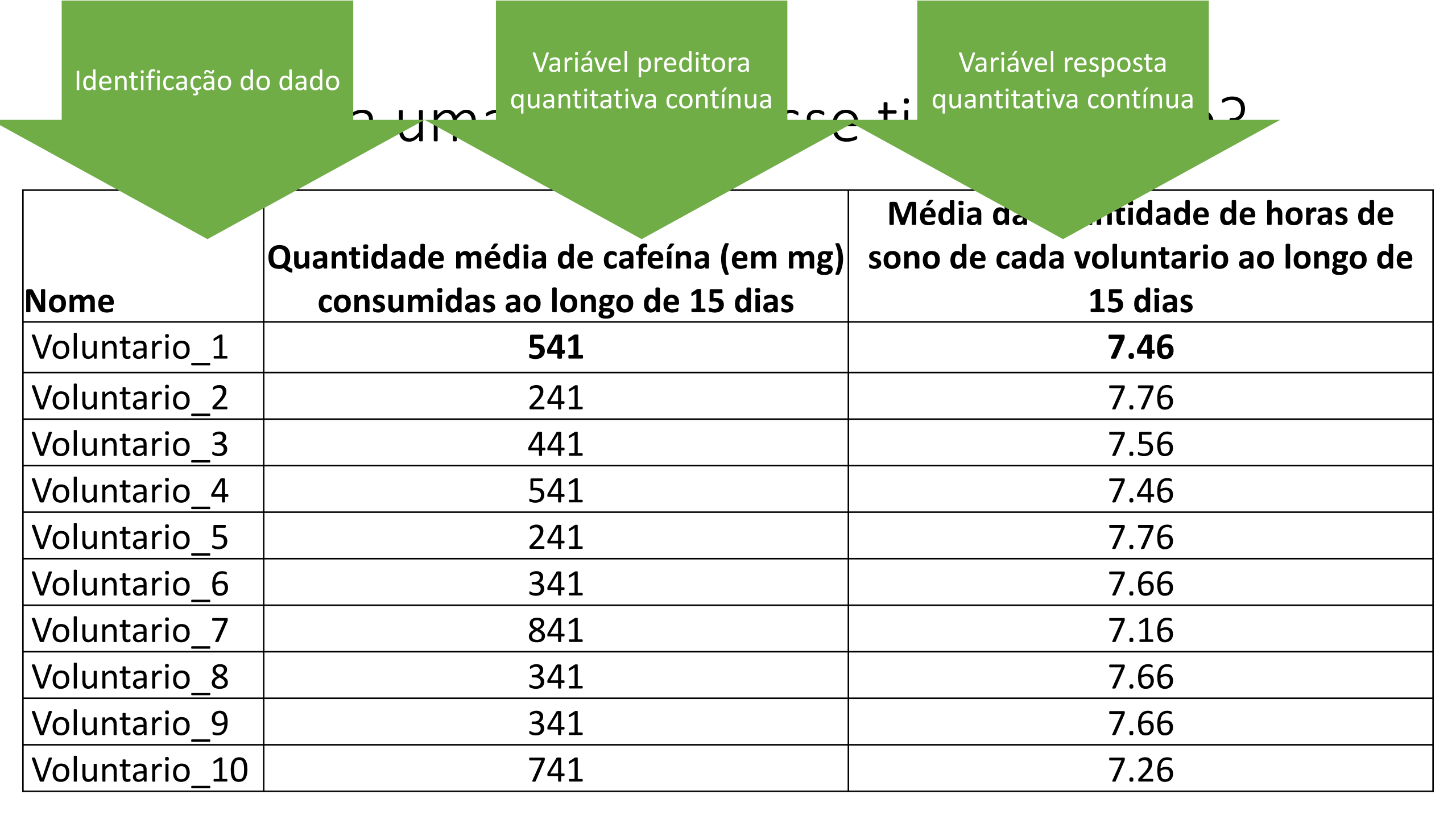
# Validação de modelos / Analise de resíduos

- O que são resíduos? R. Diferença entre valores esperados (de acordo com a formula do modelo) e valores observados
  - OBS: Como isso vai aparecer no R
    - Fitted values = valores previstos
    - Residuals = valores previstos – valores observados
- Como analisar os resíduos? Inspeção gráfica (Cap 2, Zuur et al 2009)
  - Fazer um gráfico de resíduos vs valores previstos para avaliar se
    1. A relação é linear
    2. Existe homocedasticidade (pontos estão igualmente distribuídos igualmente em toda a área do gráfico)
  - Avaliar por histograma/boxplot se os resíduos tem distribuição normal
  - Resíduos vs. variável explanatória (para testar independência)

Estrutura do banco de dados

Como fica uma tabela desse tipo de dado?

<b>Nome</b>	<b>Quantidade média de cafeína (em mg) consumidas ao longo de 15 dias</b>	<b>Média da quantidade de horas de sono de cada voluntario ao longo de 15 dias</b>
Voluntario_1	541	7.46
Voluntario_2	241	7.76
Voluntario_3	441	7.56
Voluntario_4	541	7.46
Voluntario_5	241	7.76
Voluntario_6	341	7.66
Voluntario_7	841	7.16
Voluntario_8	341	7.66
Voluntario_9	341	7.66
Voluntario_10	741	7.26

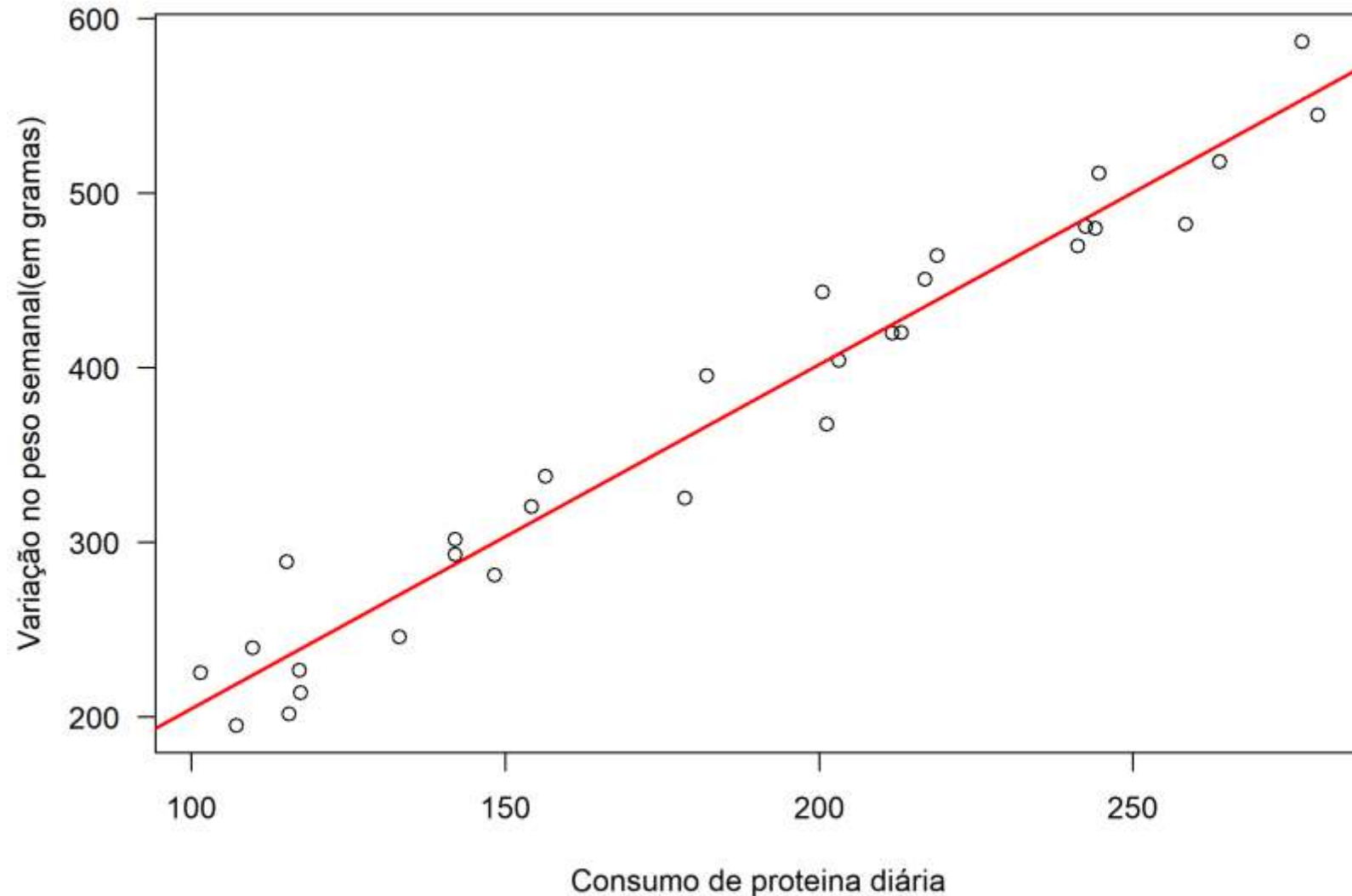


Exemplos dos alunos

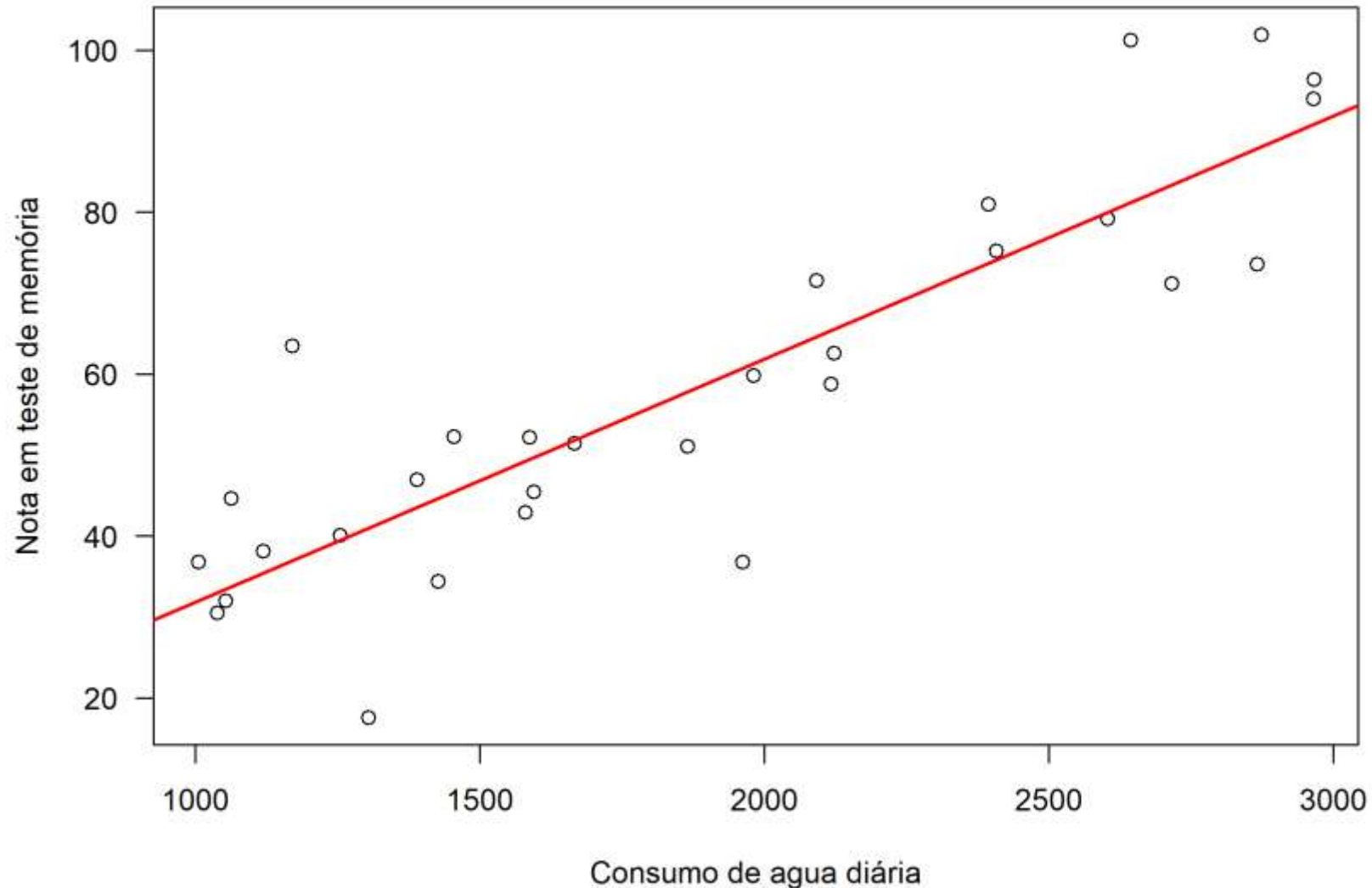
# Exemplos de adaptações de perguntas da Lista um que podem ser respondidas com essa metodologia

- 1) Realizar dieta hiperproteica ajuda no emagrecimento? (Alex)
- 2) Beber pouca água pode causar perda de memória? (Leticia)
- 3) O sono está relacionado ao crescimento da criança? (Pamela)
- 4) A taxa de infecção por dengue das cidades, está associada a proporção de casas que recebem saneamento básico? (Rafaela Reimberg)
- 5) Existe correlação entre taxa de mortalidade infantil e nível de pobreza (estimado pelo PIB) nos diferentes estados brasileiros? (Rosa)

# Realizar dieta hiperproteica ajuda no emagrecimento?

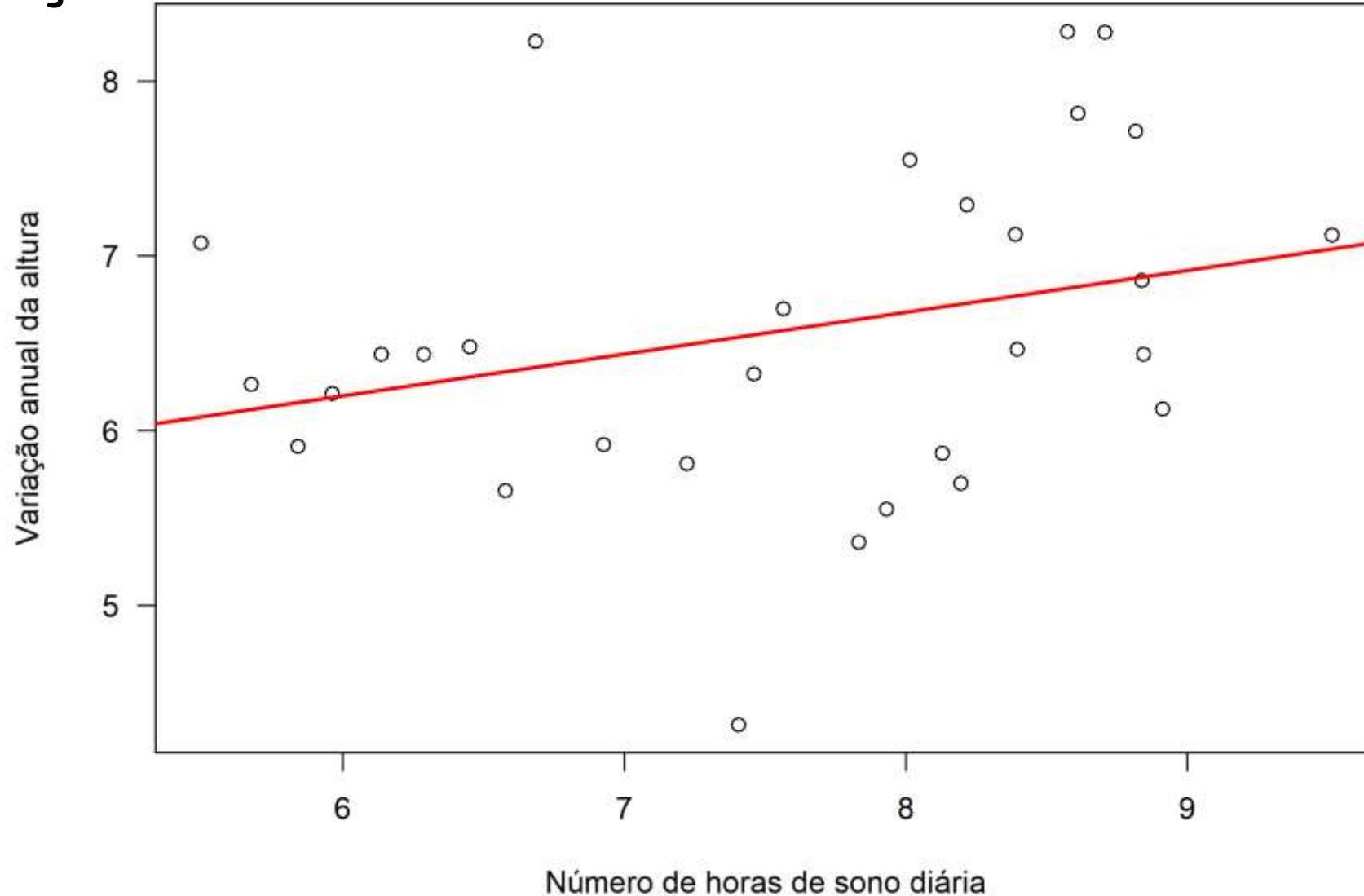


# Beber pouca água pode causar perda de memória?

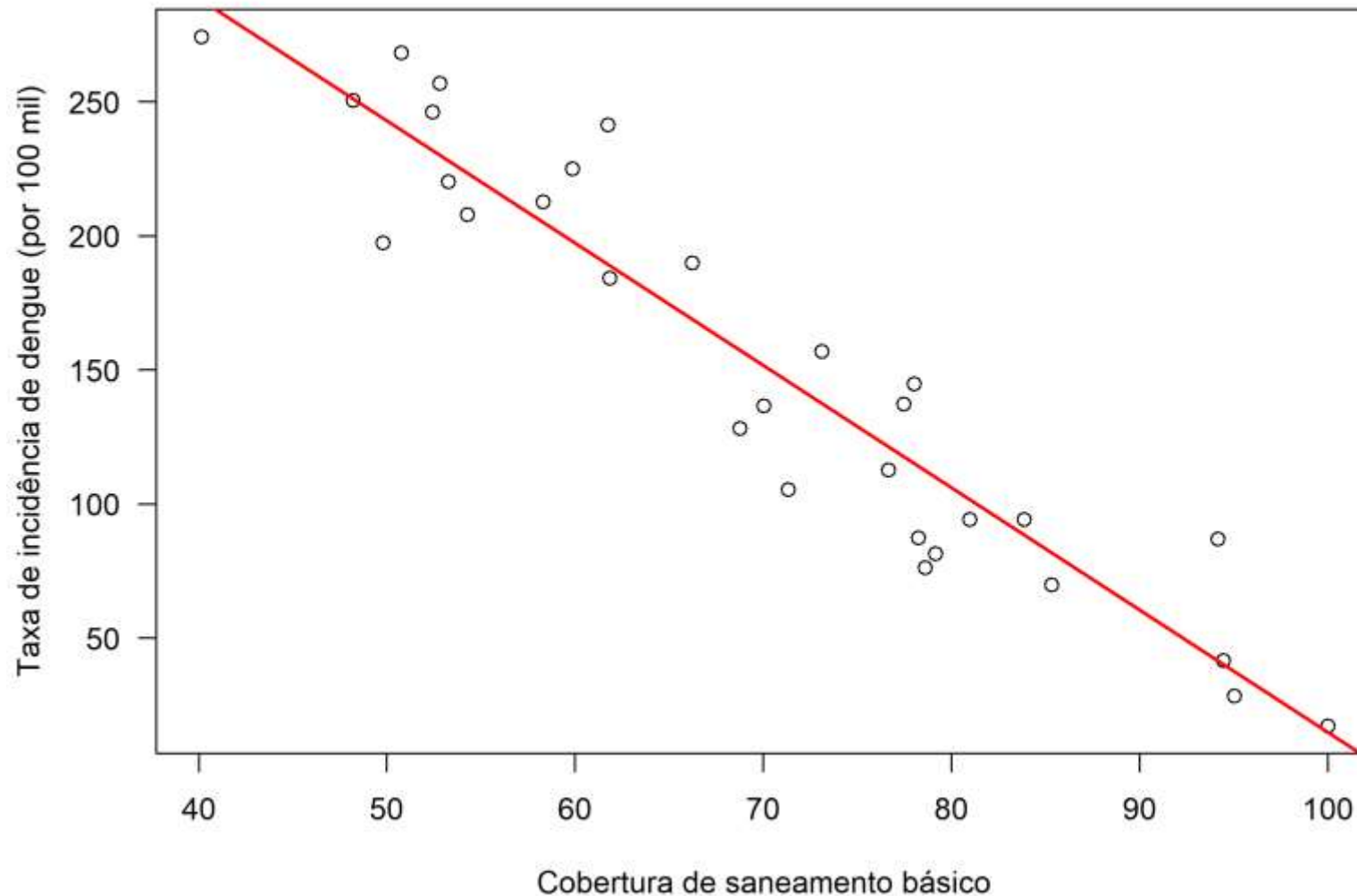




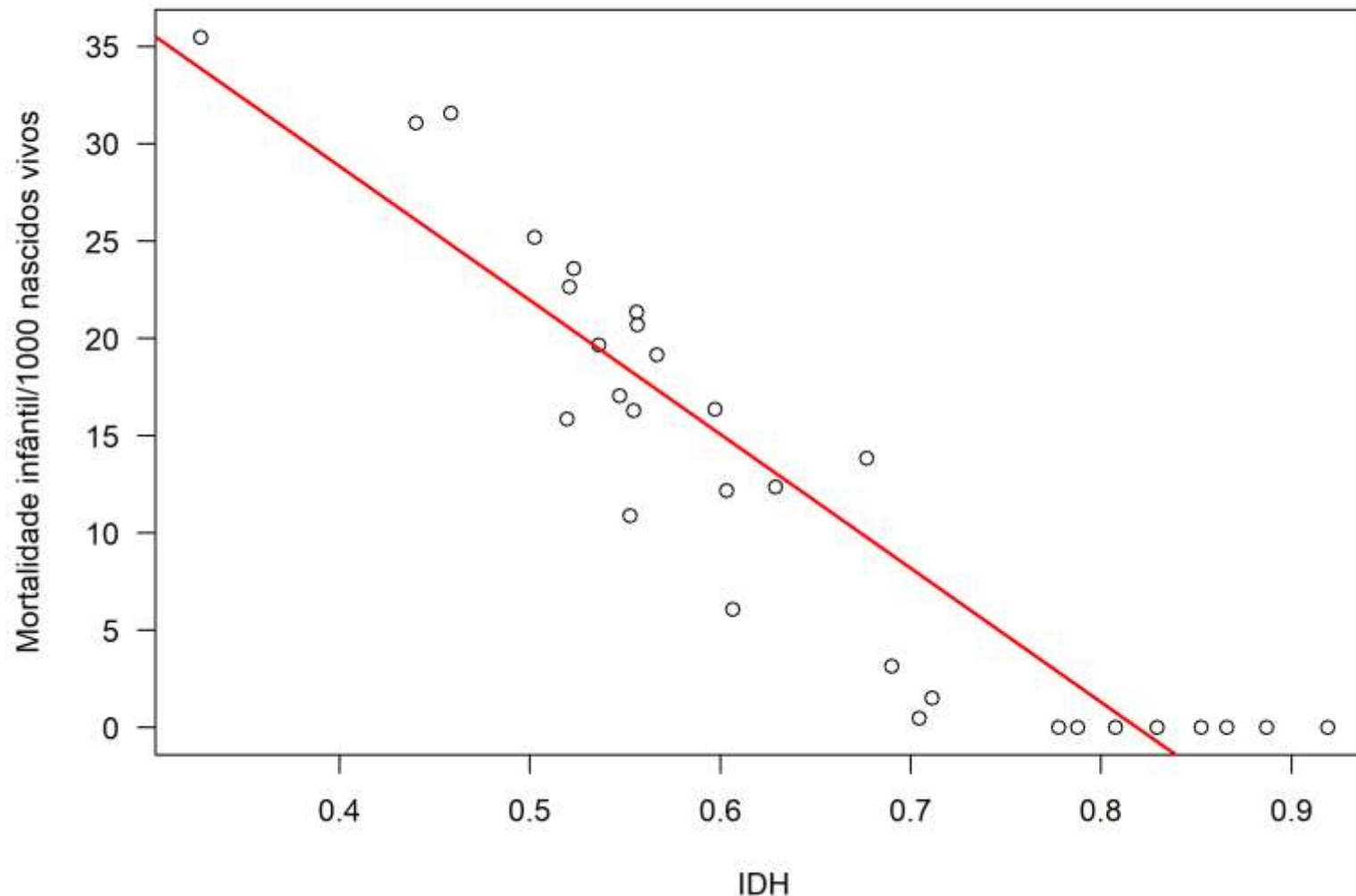
# O sono está relacionado ao crescimento da criança?



A taxa de infecção por dengue das cidades, está associada a proporção de casas que recebem saneamento básico?



Existe correlação entre taxa de mortalidade infantil e nível de pobreza (estimado pelo PIB) nos diferentes estados brasileiros?



Prática

Prática – 1 A pressão sanguínea diastólica pode ser prevista pelo tempo em repouso, após a prática de atividade física?

- Tabela “DadosAula13.xlsx”
- Colunas da tabela:
  - Pressao: Variável quantitativa contínua
  - Repouso: Variável quantitativa contínua

# Interpretação dos resultados

```
> summary(modelo)
```

```
call:
```

```
lm(formula = tabela1$Pressao ~ tabela1$Repouso)
```

```
Residuals:
```

Min	1Q	Median	3Q
-4.2191	-1.4858	-0.1014	1.3140

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.72152	1.50690	6.451	4.23e-09 ***
tabela1\$Repouso	-0.00201	0.14924	-0.013	0.989

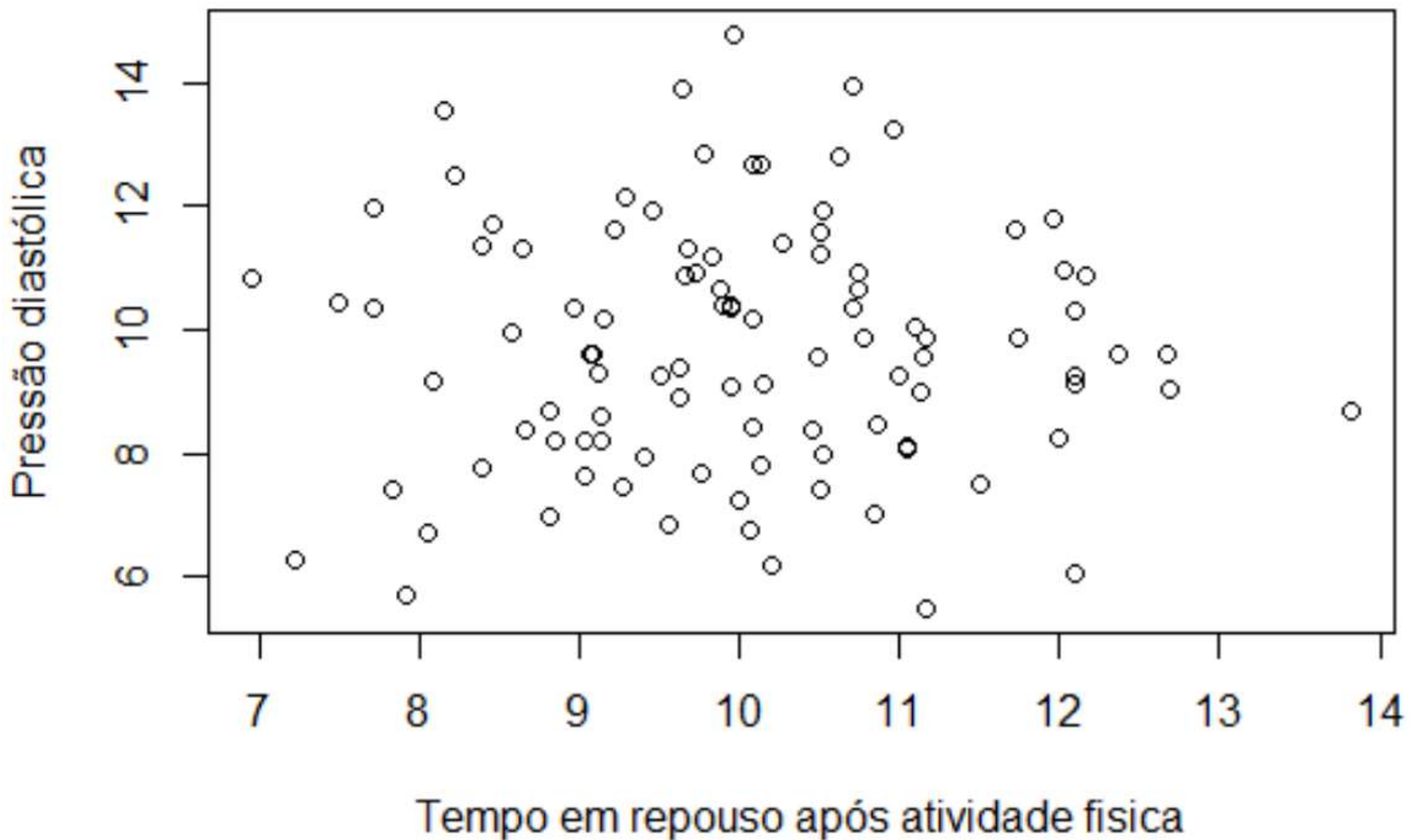
Resultado: P maior que 0.05  
Conclusão:  $H_0$  é verdadeira

Hipóteses testadas

$H_0$ :  $\beta_1$  não é diferente de zero. Portanto, não há associação entre as variáveis.

$H_1$ :  $\beta_1$  é diferente de zero. Portanto, há (sim) associação entre as variáveis.

# Gráfico ou diagrama de dispersão



Prática – 2 A pressão sanguínea diastólica pode ser prevista quantidade de sódio consumida na ultima hora?

- Tabela “DadosAula12.xlsx”
- Colunas da tabela:
  - Pressao: Variável quantitativa contínua
  - Sodio: Variável quantitativa contínua



# Interpretação dos resultados

```
> summary(modelo)
```

Call:

```
lm(formula = tabela1$Pressao ~ tabela1$Sodio)
```

Residuals:

Min	1Q	Median	3Q
-1.3668	-0.4557	0.0677	0.3699

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.5936	0.1444	38.73	<2e-16 ***
tabela1\$Sodio	8.6317	0.2758	31.29	<2e-16 ***

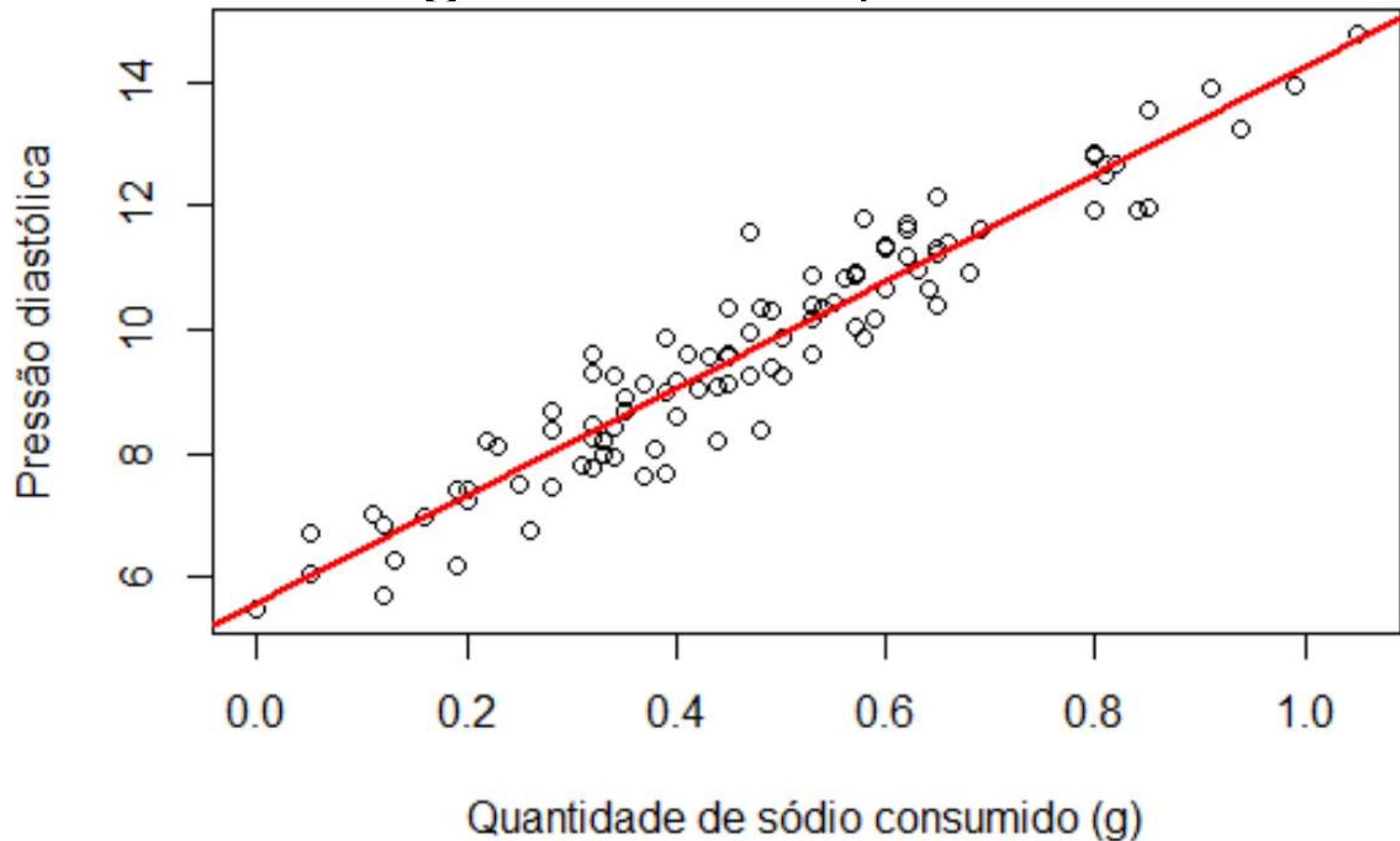
Resultado: P menor que 0.05  
Conclusão:  $H_1$  é verdadeira

Hipóteses testadas

$H_0$ :  $\beta_1$  não é diferente de zero. Portanto, não há associação entre as variáveis.

$H_1$ :  $\beta_1$  é diferente de zero. Portanto, há (sim) associação entre as variáveis.

# Gráfico ou diagrama de dispersão



# Validação do modelo

