ACA228 - Modelos de Regressão e Previsão

RLS: Interpretação e Propriedades

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 5

Revisão da aula anterior

Qualidade de ajuste

Interpretação dos parâmetros: unidades de medida e não-lineariedades

Propriedades dos Estimadores

Variância do erro

Revisão da aula anterior

Na aula anterior...

- Conhecimos o modelo de regressão linar simples (RLS), que é um modelo da forma
$$Y = \beta_0 + \beta_1 X + u$$

Na aula anterior. . .

- Conhecimos o modelo de regressão linar simples (RLS), que é um modelo da forma
$$Y = \beta_0 + \beta_1 X + u$$
- Vimos que na prática $\beta = [\beta_0 \quad \beta_1]'$ nunca é conhecido e precisamos estimá-lo utilizando os dados.

Na aula anterior...

- Conhecimos o modelo de regressão linar simples (RLS), que é um modelo da forma
$$Y = \beta_0 + \beta_1 X + u$$
- Vimos que na prática $\beta = [\beta_0 \quad \beta_1]'$ nunca é conhecido e precisamos estimá-lo utilizando os dados.
- Discutimos o estimador MQO (que é um método para estimar os β's) e vimos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Na aula anterior...

- Conhecimos o modelo de regressão linar simples (RLS), que é um modelo da forma
$$Y = \beta_0 + \beta_1 X + u$$
- Vimos que na prática $\beta = [\beta_0 \quad \beta_1]'$ nunca é conhecido e precisamos estimá-lo utilizando os dados.
- Discutimos o estimador MQO (que é um método para estimar os β's) e vimos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Na aula anterior. . .

- Conhecimos o modelo de regressão linar simples (RLS), que é um modelo da forma
$$Y = \beta_0 + \beta_1 X + u$$
- Vimos que na prática $\beta = [\beta_0 \quad \beta_1]'$ nunca é conhecido e precisamos estimá-lo utilizando os dados.
- Discutimos o estimador MQO (que é um método para estimar os β's) e vimos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Mas..será que esses $\hat{\beta}'s$ são *"bons"*? e quanto será que essa reta se ajusta aos nossos dados?

Qualidade de ajuste

Qualidade de ajuste

A qualidade de ajuste nos ajudará a saber até que ponto nosso modelo se ajusta aos dados.

Qualidade de ajuste

A qualidade de ajuste nos ajudará a saber até que ponto nosso modelo se ajusta aos dados.

Um dos critérios mais utilizados é conhecido como R^2 (R-quadrado), mas para conhece-lo e entende-lo, precisamos introduzir os seguintes conceitos:

Qualidade de ajuste

A qualidade de ajuste nos ajudará a saber até que ponto nosso modelo se ajusta aos dados.

Um dos critérios mais utilizados é conhecido como R^2 (R-quadrado), mas para conhece-lo e entende-lo, precisamos introduzir os seguintes conceitos:

- SQT: Soma de Quadrados Totais
- SQE. Soma de Quadrados Explicados
- SQR: Soma de Quadrados dos Resíduos

SQT, SQE, SQR

Tinhamos visto que

$$\hat{u}_i = y_i - \hat{y}_i \quad \text{ou equivalentemente} \quad y_i = \hat{y}_i + \hat{u}_i$$

SQT, SQE, SQR

Tinhamos visto que

$$\hat{u}_i = y_i - \hat{y}_i \quad \text{ou equivalentemente} \quad y_i = \hat{y}_i + \hat{u}_i$$

Vamos decompor $\sum_{i=1}^{n}(y_i - \bar{y})^2$ em 2 partes: **variabilidade explicada pelo modelo** e **variabilidade dos resíduos**.

SQT, SQE, SQR

Tinhamos visto que

$$\hat{u}_i = y_i - \hat{y}_i \quad \text{ou equivalentemente} \quad y_i = \hat{y}_i + \hat{u}_i$$

Vamos decompor $\sum_{i=1}^{n}(y_i - \bar{y})^2$ em 2 partes: **variabilidade explicada pelo modelo** e **variabilidade dos resíduos**.

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\underbrace{y_i - \hat{y}_i}_{\hat{u}_i} + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{u}_i + \hat{y}_i - \bar{y})^2$$

SQT, SQE, SQR

Tinhamos visto que

$$\hat{u}_i = y_i - \hat{y}_i \quad \text{ou equivalentemente} \quad y_i = \hat{y}_i + \hat{u}_i$$

Vamos decompor $\sum_{i=1}^{n}(y_i - \bar{y})^2$ em 2 partes: **variabilidade explicada pelo modelo** e **variabilidade dos resíduos**.

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\underbrace{y_i - \hat{y}_i}_{\hat{u}_i} + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{u}_i + \hat{y}_i - \bar{y})^2$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{SQT} = \underbrace{\sum_{i=1}^{n}\hat{u}_i^2}_{SQR} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{SQE} + 2\underbrace{\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i - \bar{y})}_{0}$$

SQT, SQE, SQR

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = 0?$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i \hat{y}_i - \sum_{i=1}^{n} \hat{u}_i \bar{y}$$

SQT, SQE, SQR

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = 0?$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i\hat{y}_i - \sum_{i=1}^{n} \hat{u}_i\bar{y}$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y}\underbrace{\sum_{i=1}^{n} \hat{u}_i}_{0}$$

SQT, SQE, SQR

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = 0?$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i \hat{y}_i - \sum_{i=1}^{n} \hat{u}_i \bar{y}$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y} \underbrace{\sum_{i=1}^{n} \hat{u}_i}_{0}$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i \hat{\beta}_0 + \sum_{i=1}^{n} \hat{u}_i \hat{\beta}_1 x_i = \hat{\beta}_0 \underbrace{\sum_{i=1}^{n} \hat{u}_i}_{0} + \hat{\beta}_1 \underbrace{\sum_{i=1}^{n} \hat{u}_i x_i}_{0}$$

SQT, SQE, SQR

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = 0?$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i\hat{y}_i - \sum_{i=1}^{n} \hat{u}_i\bar{y}$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y}\underbrace{\sum_{i=1}^{n} \hat{u}_i}_{0}$$

$$\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} \hat{u}_i\hat{\beta}_0 + \sum_{i=1}^{n} \hat{u}_i\hat{\beta}_1 x_i = \hat{\beta}_0\underbrace{\sum_{i=1}^{n} \hat{u}_i}_{0} + \hat{\beta}_1\underbrace{\sum_{i=1}^{n} \hat{u}_i x_i}_{0}$$

Ou seja,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y})$$

Qualidade de ajuste

$$SQT = SQR + SQE$$

Qualidade de ajuste

$$SQT = SQR + SQE$$

$$1 = \frac{SQR}{SQT} + \frac{SQE}{SQT}$$

Qualidade de ajuste

$$SQT = SQR + SQE$$

$$1 = \frac{SQR}{SQT} + \frac{SQE}{SQT}$$

$$1 - \frac{SQR}{SQT} = \frac{SQE}{SQT}$$

Qualidade de ajuste

$$SQT = SQR + SQE$$

$$1 = \frac{SQR}{SQT} + \frac{SQE}{SQT}$$

$$1 - \frac{SQR}{SQT} = \frac{SQE}{SQT}$$

R-quadrado ou coeficiente de determinação

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

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

- $0 \leq R^2 \leq 1$

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

- $0 \leq R^2 \leq 1$
- razão entre a variação explicada e a variação total.

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

- $0 \leq R^2 \leq 1$
- razão entre a variação explicada e a variação total.
- $100 \times R^2$ é a porcentagem de variação de y explicada pelo modelo.

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

- $0 \leq R^2 \leq 1$
- razão entre a variação explicada e a variação total.
- $100 \times R^2$ é a porcentagem de variação de y explicada pelo modelo.
- R^2 próximo de **0** indica um ajuste **ruim**, R^2 próximo de **1** indica um **bom** ajuste.

Qualidade de ajuste

R-quadrado ou coeficiente de determinação

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

- $0 \leq R^2 \leq 1$
- razão entre a variação explicada e a variação total.
- $100 \times R^2$ é a porcentagem de variação de y explicada pelo modelo.
- R^2 próximo de **0** indica um ajuste **ruim**, R^2 próximo de **1** indica um **bom** ajuste.
- R^2 é o quadrado do coeficiente de correlação entre y e $\hat{y}$.

Qualidade de ajuste

Como é o ajuste em cada um dos modelos vistos na aula anterior?

Qualidade de ajuste

Como é o ajuste em cada um dos modelos vistos na aula anterior?

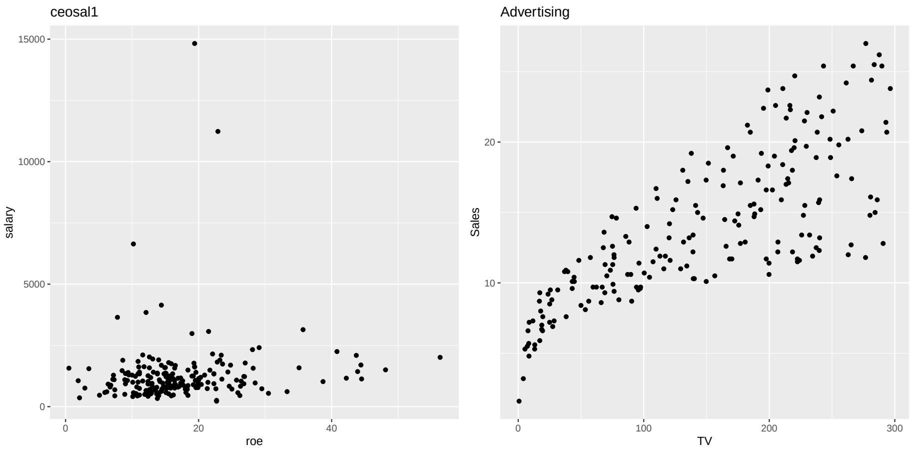
```r
modelo1 = lm(salary~roe, data = ceosal1)
summary(modelo1)$r.squared
```

```
## [1] 0.01318862
```

```r
modelo2 = lm(Sales~TV, data = Advertising)
summary(modelo2)$r.squared
```

```
## [1] 0.6118751
```

Qualidade de ajuste



Interpretação dos parâmetros: unidades de medida e não-lineariedades

Unidades de medida

- Quando trabalhamos com dados, é comum que cada *analista de dados* tenha as variáveis da forma que para ele/ela seja mais facil de entender e interpretar.

Unidades de medida

- Quando trabalhamos com dados, é comum que cada *analista de dados* tenha as variáveis da forma que para ele/ela seja mais facil de entender e interpretar.
- Isto significa que para um mesmo problema, João pode trabalhar com a variável Salário (em USD), Maria pode trabalhar com a variável Salário (em milhares de USD) e Lucas pode trabalhar com a variável Salário (em Reais).

Unidades de medida

- Quando trabalhamos com dados, é comum que cada *analista de dados* tenha as variáveis da forma que para ele/ela seja mais facil de entender e interpretar.
- Isto significa que para um mesmo problema, João pode trabalhar com a variável Salário (em USD), Maria pode trabalhar com a variável Salário (em milhares de USD) e Lucas pode trabalhar com a variável Salário (em Reais).
- Será que isso afeta nosso modelo e/ou a qualidade de ajuste dele?

Unidades de medida

Sejam $salarydol = salary \times 1000$ e $roedec = roe/100$

Unidades de medida

Sejam $salarydol = salary \times 1000$ e $roedec = roe/100$

```
coef(lm(salary~roe, data = ceosal1))
```

```
## (Intercept)         roe
##   963.19134    18.50119
```

```
coef(lm(salarydol~roe, data = ceosal1))
```

```
## (Intercept)         roe
##  0.96319134  0.01850119
```

```
coef(lm(salary~roedec, data = ceosal1))
```

```
## (Intercept)      roedec
##    963.1913   1850.1186
```

Unidades de medida

- Se a variável dependente for cy, então os $\hat{\beta}$'s da regressão cy sobre x serão $c\hat{\beta}_0$ e $c\hat{\beta}_1$ (onde $\hat{\beta}_0, \hat{\beta}_1$ são os betas da regressão de y sobre x).

Unidades de medida

- Se a variável dependente for cy, então os $\hat{\beta}$'s da regressão cy sobre x serão $c\hat{\beta}_0$ e $c\hat{\beta}_1$ (onde $\hat{\beta}_0, \hat{\beta}_1$ são os betas da regressão de y sobre x).
- Se a variavel independente for cx, então o novo coeficiente da regressão y sobre cx será $\hat{\beta}_1/c$ ($\hat{\beta}_0$ permanece igual).

Unidades de medida

- Se a variável dependente for cy, então os $\hat{\beta}$'s da regressão cy sobre x serão $c\hat{\beta}_0$ e $c\hat{\beta}_1$ (onde $\hat{\beta}_0, \hat{\beta}_1$ são os betas da regressão de y sobre x).
- Se a variavel independente for cx, então o novo coeficiente da regressão y sobre cx será $\hat{\beta}_1/c$ ($\hat{\beta}_0$ permanece igual).

Unidades de medida

- Se a variável dependente for cy, então os $\hat{\beta}$'s da regressão cy sobre x serão $c\hat{\beta}_0$ e $c\hat{\beta}_1$ (onde $\hat{\beta}_0, \hat{\beta}_1$ são os betas da regressão de y sobre x).
- Se a variavel independente for cx, então o novo coeficiente da regressão y sobre cx será $\hat{\beta}_1/c$ ($\hat{\beta}_0$ permanece igual).

Importante

Para uma correta interpretação, devemos conhecer as unidades de medida utilizadas nas variaveis.

Unidades de medida

O que acontece com o R^2 quando mudamos a unidade de medida de algumas das variáveis?

Unidades de medida

O que acontece com o R^2 quando mudamos a unidade de medida de algumas das variáveis?

```
summary(lm(salary~roe, data = ceosal1))$r.squared
```

```
## [1] 0.01318862
```

```
summary(lm(salarydol~roe, data = ceosal1))$r.squared
```

```
## [1] 0.01318862
```

```
summary(lm(salary~roedec, data = ceosal1))$r.squared
```

```
## [1] 0.01318862
```

Mudanças na unidade de medida não afetam o R^2.

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável
- Gera um modelo com (aproximadamente) um efeito percentual constante $\log(y) = \beta_0 + \beta_1 x + u$

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável
- Gera um modelo com (aproximadamente) um efeito percentual constante $\log(y) = \beta_0 + \beta_1 x + u$
- $\Delta \log(y) = \log(y_1) - \log(y_0) \approx (y_1 - y_0)/y_0 = \Delta y/y_0$, então
$$100 \times \Delta \log(y) \approx \% \Delta y$$

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável
- Gera um modelo com (aproximadamente) um efeito percentual constante $\log(y) = \beta_0 + \beta_1 x + u$
- $\Delta \log(y) = \log(y_1) - \log(y_0) \approx (y_1 - y_0)/y_0 = \Delta y/y_0$, então
$$100 \times \Delta \log(y) \approx \%\Delta y$$
- Logo, se $\Delta u = 0$, temos que
$$\%\Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável
- Gera um modelo com (aproximadamente) um efeito percentual constante $\log(y) = \beta_0 + \beta_1 x + u$
- $\Delta \log(y) = \log(y_1) - \log(y_0) \approx (y_1 - y_0)/y_0 = \Delta y/y_0$, então
$$100 \times \Delta \log(y) \approx \% \Delta y$$
- Logo, se $\Delta u = 0$, temos que
$$\% \Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

Não linearidade

Até aqui, temos enfatizado as relações lineares no nosso MLRS, porém relações lineares não são sempre fáceis de encontrar.

Transformação logaritmica

- A transformação pode linearizar a variável
- Gera um modelo com (aproximadamente) um efeito percentual constante $\log(y) = \beta_0 + \beta_1 x + u$
- $\Delta \log(y) = \log(y_1) - \log(y_0) \approx (y_1 - y_0)/y_0 = \Delta y / y_0$, então
$$100 \times \Delta \log(y) \approx \% \Delta y$$
- Logo, se $\Delta u = 0$, temos que
$$\% \Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

(para uma revisao de $\log(\cdot)$ ver Apêndice **A.4b**)

Não linearidade

Exemplo

O dataset `wage1` do pacote `wooldridge` contém informação de 526 pessoas. A variável `wage` é o salário médio por hora e a variável `educ` são os anos de educação.

Não linearidade

Exemplo

O dataset `wage1` do pacote `wooldridge` contém informação de 526 pessoas. A variável `wage` é o salário médio por hora e a variável `educ` são os anos de educação.

Ajustando um modelo RLS $\log(wage) = \beta_0 + \beta_1 educ + u$, obtemos:

```
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Não linearidade

Exemplo

O dataset `wage1` do pacote `wooldridge` contém informação de 526 pessoas. A variável `wage` é o salário médio por hora e a variável `educ` são os anos de educação.

Ajustando um modelo RLS $\log(wage) = \beta_0 + \beta_1 educ + u$, obtemos:

```
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Qual interpretação é correta?

- A cada ano adicional de educação, o salário médio por hora aumenta em ≈ 0.0827 USD
- A cada ano adicional de educação, o salário médio por hora aumenta em $\approx 8.27\%$

Não linearidade

```r
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Não linearidade

```
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Tinhamos visto que

$$\%\Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

Não linearidade

```r
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Tinhamos visto que

$$\%\Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

No nosso caso:

$$\%\Delta wage \approx (100\beta_1)\Delta educ$$

Não linearidade

```
coef(lm(log(wage)~educ, data = wage1))
```

```
## (Intercept)        educ
##  0.58377267  0.08274437
```

Tinhamos visto que

$$\%\Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x$$

No nosso caso:

$$\%\Delta wage \approx (100\beta_1)\Delta educ$$

A cada ano adicional de educação, o salario aumenta $(100 \times 0.08274 = 8,274\%)$

Não linearidade

Seja a regressão

$$\log(y) = \beta_0 + \beta_1 \log(x) + u, \tag{1}$$

Não linearidade

Seja a regressão

$$\log(y) = \beta_0 + \beta_1 \log(x) + u, \tag{1}$$

Se utilizarmos o fato que $\%\Delta y \approx 100 \times \Delta \log(y)$, temos que

$$\frac{\%\Delta y}{\%\Delta x} = \frac{100\Delta \log(y)}{100\Delta \log(x)} = \beta_1$$

Quando x aumenta 1%, y aumenta em β_1%

Não linearidade

Seja a regressão

$$\log(y) = \beta_0 + \beta_1 \log(x) + u, \tag{1}$$

Se utilizarmos o fato que $\%\Delta y \approx 100 \times \Delta \log(y)$, temos que

$$\frac{\%\Delta y}{\%\Delta x} = \frac{100\Delta \log(y)}{100\Delta \log(x)} = \beta_1$$

Quando x aumenta 1%, y aumenta em β_1%

Nota:

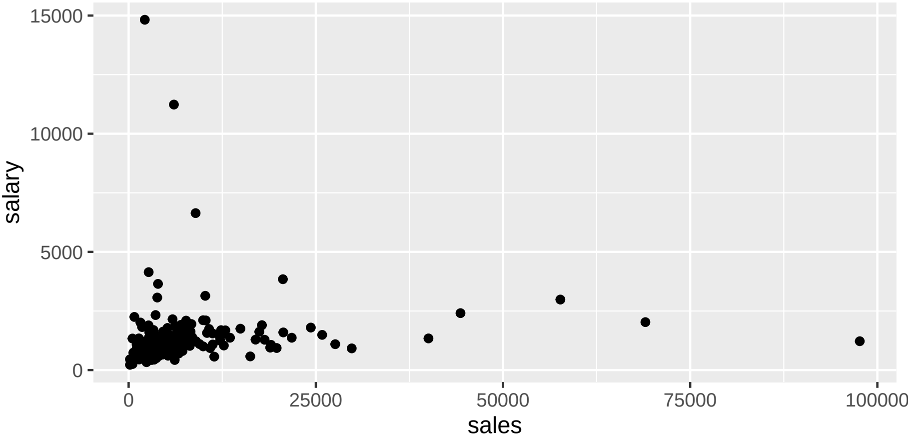
A *Elasticidade* de y em relação a x é definida como variação percentual de y quando x aumenta 1%. Então na regressão (1), β_1 é a elasticidade de y em relação a x.

Não linearidade

Como saberei se devo aplicar alguma transformação?

Não linearidade

Como saberei se devo aplicar alguma transformação?

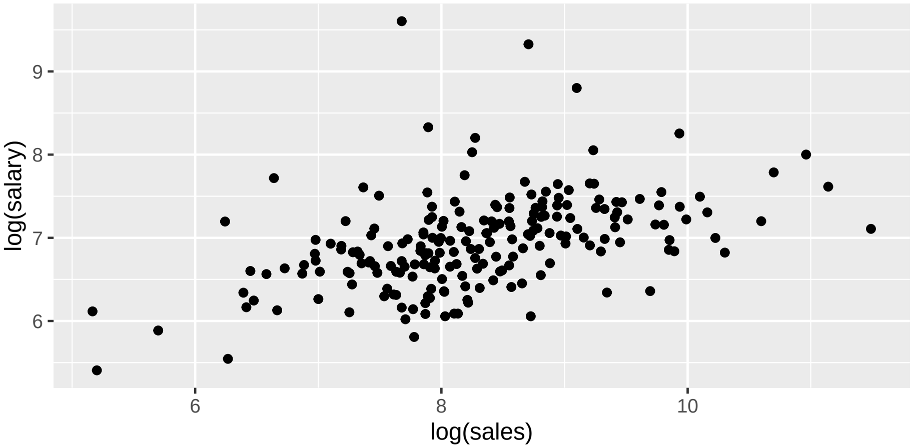


Não linearidade

A transformação logaritmica costuma funcionar bastante bem

Não linearidade

A transformação logaritmica costuma funcionar bastante bem



Não linearidade

```
coef(lm(log(salary)~log(sales),data = ceosal1))
```

```
## (Intercept)  log(sales)
##   4.8219965   0.2566717
```

Não linearidade

```r
coef(lm(log(salary)~log(sales),data = ceosal1))
```

```
## (Intercept)  log(sales)
##   4.8219965   0.2566717
```

$$\frac{\%\Delta Salario}{\%\Delta Vendas} = \frac{\Delta \log(Salario)}{\Delta \log(Vendas)} = \beta_1$$

A Elasticidade do salário em relação às vendas é 0.257. Isto implica que o aumento de 1% nas vendas, aumenta o salário dos CEOs em 0.257%

Não linearidade

Modelo	V. Dependente	V. Independente	Interpretação β_1
Nível-Nível	y	x	$\Delta y = \beta_1 \Delta x$
Nível-Log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-Nível	$\log(y)$	x	$\%\Delta y = 100\beta_1 \Delta x$
Log-Log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Importante

O termo linear no modelo de regressão linear, refere-se à linearidade nos parametros β_0, β_1

Propriedades dos Estimadores

Propriedades dos Estimadores: Hipóteses

Hipóteste RLS1: Lineariedade nos parâmetros

No modelo populacional: $y = \beta_0 + \beta_1 x + u$

Propriedades dos Estimadores: Hipóteses

Hipóteste RLS1: Lineariedade nos parâmetros

No modelo populacional: $y = \beta_0 + \beta_1 x + u$

Hipóteste RLS2: Amostragem aleatória

$(x_1, y_1), \ldots (x_n, y_n)$ constituem uma amostra aleatoria de tamanho n do modelo populacional do HRLS1.

Propriedades dos Estimadores: Hipóteses

Hipóteste RLS1: Lineariedade nos parâmetros

No modelo populacional: $y = \beta_0 + \beta_1 x + u$

Hipóteste RLS2: Amostragem aleatória

$(x_1, y_1), \ldots (x_n, y_n)$ constituem uma amostra aleatoria de tamanho n do modelo populacional do HRLS1.

Hipóteste RLS3: Variância amostral da variavel independênte

$x_1, \ldots, x_n$ não são todos iguais.

Propriedades dos Estimadores: Hipóteses

Hipóteste RLS1: Lineariedade nos parâmetros

No modelo populacional: $y = \beta_0 + \beta_1 x + u$

Hipóteste RLS2: Amostragem aleatória

$(x_1, y_1), \ldots (x_n, y_n)$ constituem uma amostra aleatoria de tamanho n do modelo populacional do HRLS1.

Hipóteste RLS3: Variância amostral da variavel independênte

$x_1, \ldots, x_n$ não são todos iguais.

Hipóteste RLS4: Média condicional zero

$$\mathbb{E}(u|x) = 0$$

Propriedades dos Estimadores: EMQO são não-viesados

Estimadores MQO são não-viesados

Sob HRLS1 – HRLS4,

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \text{e} \quad \mathbb{E}(\hat{\beta}_1) = \beta_1$$

ou seja, $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores não-viesados. \end{block}

Propriedades dos Estimadores: EMQO são não-viesados

Estimadores MQO são não-viesados

Sob HRLS1 – HRLS4,

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \text{e} \quad \mathbb{E}(\hat{\beta}_1) = \beta_1$$

ou seja, $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores não-viesados. \end{block}

Definição: Estimadores não-viesados

Um estimador $\hat{\theta}$ de θ é não-viesado se

$$\mathbb{E}(\hat{\theta} - \theta) = 0 \quad \text{ou equivalentemente} \quad \mathbb{E}(\hat{\theta}) = \theta$$

Propriedades dos Estimadores: EMQO são não-viesados

Prova

Vamos utilizar os seguintes resultados:

- $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \sum(x_i - \bar{x})\bar{y} = \sum(x_i - \bar{x})y_i - \bar{y}\underbrace{\sum(x_i - \bar{x})}_{0} = \sum(x_i - \bar{x})y_i$

Propriedades dos Estimadores: EMQO são não-viesados

Prova

Vamos utilizar os seguintes resultados:

- $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \sum(x_i - \bar{x})\bar{y} = \sum(x_i - \bar{x})y_i - \bar{y}\underbrace{\sum(x_i - \bar{x})}_{0} = \sum(x_i - \bar{x})y_i$
- $\sum(x_i - \bar{x})x_i = \sum(x_i - \bar{x})(x_i - \bar{x} + \bar{x}) = \sum(x_i - \bar{x})^2 + \bar{x}\underbrace{\sum(x_i - \bar{x})}_{0} = \sum(x_i - \bar{x})^2$

Propriedades dos Estimadores: EMQO são não-viesados

Prova

Vamos utilizar os seguintes resultados:

- $\sum(x_i-\bar{x})(y_i-\bar{y}) = \sum(x_i-\bar{x})y_i - \sum(x_i-\bar{x})\bar{y} = \sum(x_i-\bar{x})y_i - \bar{y}\underbrace{\sum(x_i-\bar{x})}_{0} = \sum(x_i-\bar{x})y_i$
- $\sum(x_i-\bar{x})x_i = \sum(x_i-\bar{x})(x_i - \bar{x} + \bar{x}) = \sum(x_i-\bar{x})^2 + \bar{x}\underbrace{\sum(x_i-\bar{x})}_{0} = \sum(x_i-\bar{x})^2$

Propriedades dos Estimadores: EMQO são não-viesados

Prova

Vamos utilizar os seguintes resultados:

- $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \sum(x_i - \bar{x})\bar{y} = \sum(x_i - \bar{x})y_i - \bar{y}\underbrace{\sum(x_i - \bar{x})}_{0} = \sum(x_i - \bar{x})y_i$
- $\sum(x_i - \bar{x})x_i = \sum(x_i - \bar{x})(x_i - \bar{x} + \bar{x}) = \sum(x_i - \bar{x})^2 + \bar{x}\underbrace{\sum(x_i - \bar{x})}_{0} = \sum(x_i - \bar{x})^2$

Então podemos reescrever $\hat{\beta}_1$ da forma

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\overbrace{(\beta_0 + \beta_1 x_i + u_i)}^{y_i}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Propriedades dos Estimadores: EMQO são não-viesados

$$\hat{\beta}_1 = \frac{\beta_0 \sum_{i=1}^{n}(x_i - \bar{x}) + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})x_i + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Propriedades dos Estimadores: EMQO são não-viesados

$$\hat{\beta}_1 = \frac{\beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i + \sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Mas, $\sum_{i=1}^n (x_i - \bar{x}) = 0$ e $\sum_{i=1}^n (x_i - \bar{x})x_i = \sum_{i=1}^n (x_i - \bar{x})^2$, então.

Propriedades dos Estimadores: EMQO são não-viesados

$$\hat{\beta}_1 = \frac{\beta_0 \sum_{i=1}^{n}(x_i - \bar{x}) + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})x_i + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Mas, $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ e $\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2$, então.

$$\hat{\beta}_1 = \frac{\beta_0 \times 0 + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Propriedades dos Estimadores: EMQO são não-viesados

$$\hat{\beta}_1 = \frac{\beta_0 \sum_{i=1}^{n}(x_i - \bar{x}) + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})x_i + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Mas, $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ e $\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2$, então.

$$\hat{\beta}_1 = \frac{\beta_0 \times 0 + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Propriedades dos Estimadores: EMQO são não-viesados

Aplicando $\mathbb{E}(\cdot|x)$ em ambos os lados.

Propriedades dos Estimadores: EMQO são não-viesados

Aplicando $\mathbb{E}(\cdot|x)$ em ambos os lados.

$$\mathbb{E}(\hat{\beta}_1|x) = \mathbb{E}\Big(\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}|x\Big) = \beta_1 + \underbrace{\mathbb{E}\Big(\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n \left(x_i - \bar{x}\right)^2}|x\Big)}_{\sum_{i=1}^n \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \underbrace{\mathbb{E}(u_i|x)}_{0}} = \beta_1$$

Propriedades dos Estimadores: EMQO são não-viesados

Aplicando $\mathbb{E}(\cdot|x)$ em ambos os lados.

$$\mathbb{E}(\hat{\beta}_1|x) = \mathbb{E}\Big(\beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}|x\Big) = \beta_1 + \underbrace{\mathbb{E}\Big(\frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}|x\Big)}_{\sum_{i=1}^{n}\frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\underbrace{\mathbb{E}(u_i|x)}_{0}} = \beta_1$$

$$\underbrace{\mathbb{E}[\mathbb{E}(\hat{\beta}_1|x)] = \mathbb{E}(\beta_1)}_{\text{propriedade: } \mathbb{E}(\mathbb{E}(\cdot|x)) = \mathbb{E}(\cdot)} = \beta_1$$

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = (\underbrace{\beta_0 + \beta_1 \bar{x} + \bar{u}}_{\bar{y}} - \hat{\beta}_1 \bar{x})$$

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = (\underbrace{\beta_0 + \beta_1\bar{x} + \bar{u}}_{\bar{y}} - \hat{\beta}_1\bar{x})$$

Aplicando $\mathbb{E}(\cdot|x)$ em ambos os lados.

$$\mathbb{E}(\hat{\beta}_0|x) = \mathbb{E}(\beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}|x) = \beta_0 + \beta_1\underbrace{\mathbb{E}(\bar{x}|x)}_{\bar{x}} + \underbrace{\mathbb{E}(\bar{u}|x)}_{0} - \underbrace{\mathbb{E}(\hat{\beta}_1\bar{x}|x)}_{\bar{x}\mathbb{E}(\hat{\beta}_1|x) = \bar{x}\beta_1}$$

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = (\underbrace{\beta_0 + \beta_1\bar{x} + \bar{u}}_{\bar{y}} - \hat{\beta}_1\bar{x})$$

Aplicando $\mathbb{E}(\cdot|x)$ em ambos os lados.

$$\mathbb{E}(\hat{\beta}_0|x) = \mathbb{E}(\beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}|x) = \beta_0 + \beta_1\underbrace{\mathbb{E}(\bar{x}|x)}_{\bar{x}} + \underbrace{\mathbb{E}(\bar{u}|x)}_{0} - \underbrace{\mathbb{E}(\hat{\beta}_1\bar{x}|x)}_{\bar{x}\mathbb{E}(\hat{\beta}_1|x)=\bar{x}\beta_1}$$

$$\mathbb{E}[\mathbb{E}(\hat{\beta}_0|x)] = \mathbb{E}(\beta_0) = \beta_0$$

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $\mathbb{E}(\hat{\beta}_0) = \beta_0$ e $\mathbb{E}(\hat{\beta}_1) = \beta_1$, *i.e.* a distribuição amostral de $\hat{\beta}_1$ ($\hat{\beta}_0$) está centrada em torno de β_1 (β_0).

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $\mathbb{E}(\hat{\beta}_0) = \beta_0$ e $\mathbb{E}(\hat{\beta}_1) = \beta_1$, *i.e.* a distribuição amostral de $\hat{\beta}_1$ $(\hat{\beta}_0)$ está centrada em torno de β_1 (β_0).
- Agora estamos interessados em saber o quão distante podemos esperar que $\hat{\beta}_1$ esteja de β_1 e $\hat{\beta}_0$ esteja de β_0

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $\mathbb{E}(\hat{\beta}_0) = \beta_0$ e $\mathbb{E}(\hat{\beta}_1) = \beta_1$, *i.e.* a distribuição amostral de $\hat{\beta}_1$ ($\hat{\beta}_0$) está centrada em torno de β_1 (β_0).
- Agora estamos interessados em saber o quão distante podemos esperar que $\hat{\beta}_1$ esteja de β_1 e $\hat{\beta}_0$ esteja de β_0
- Em particular, queremos conhecer $\mathbb{V}(\hat{\beta}_0|x)$, $\mathbb{V}(\hat{\beta}_1|x)$

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $\mathbb{E}(\hat{\beta}_0) = \beta_0$ e $\mathbb{E}(\hat{\beta}_1) = \beta_1$, *i.e.* a distribuição amostral de $\hat{\beta}_1$ ($\hat{\beta}_0$) está centrada em torno de β_1 (β_0).
- Agora estamos interessados em saber o quão distante podemos esperar que $\hat{\beta}_1$ esteja de β_1 e $\hat{\beta}_0$ esteja de β_0
- Em particular, queremos conhecer $\mathbb{V}(\hat{\beta}_0|x)$, $\mathbb{V}(\hat{\beta}_1|x)$

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $\mathbb{E}(\hat{\beta}_0) = \beta_0$ e $\mathbb{E}(\hat{\beta}_1) = \beta_1$, *i.e.* a distribuição amostral de $\hat{\beta}_1$ ($\hat{\beta}_0$) está centrada em torno de β_1 (β_0).
- Agora estamos interessados em saber o quão distante podemos esperar que $\hat{\beta}_1$ esteja de β_1 e $\hat{\beta}_0$ esteja de β_0
- Em particular, queremos conhecer $\mathbb{V}(\hat{\beta}_0|x)$, $\mathbb{V}(\hat{\beta}_1|x)$

Hipóteste RLS5: Homocedasticidade

$$\mathbb{V}(u|x) = \mathbb{E}(u^2|x) - (\mathbb{E}(u|x))^2 = \mathbb{E}(u^2|x) = \sigma^2$$

Propriedades dos Estimadores: variância dos EMQO

Variância dos EMQO

Sob HRLS1 – HRLS5,

$$\mathbb{V}(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{e} \quad \mathbb{V}(\hat{\beta}_0|x) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

onde σ^2 é a variância do erro ($\mathbb{V}(u) = \sigma^2$)

Propriedades dos Estimadores: variância dos EMQO

Variância dos EMQO

Sob HRLS1 – HRLS5,

$$\mathbb{V}(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{e} \quad \mathbb{V}(\hat{\beta}_0|x) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

onde σ^2 é a variância do erro ($\mathbb{V}(u) = \sigma^2$)

Qual o problema com essas formulas?

Propriedades dos Estimadores: variância dos EMQO

Variância dos EMQO

Sob HRLS1 – HRLS5,

$$\mathbb{V}(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{e} \quad \mathbb{V}(\hat{\beta}_0|x) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

onde σ^2 é a variância do erro ($\mathbb{V}(u) = \sigma^2$)

Qual o problema com essas formulas?

- Na prática, não conhecemos σ^2

Variância do erro

Estimação da variância do erro

- Na prática, não conhecemos σ^2
- Utilizaremos os dados para estimar σ^2, esse valor é denotado por $\hat{\sigma}^2$.

Estimação não-viesada de σ^2

Sob HRLS1–HRLS5,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}$$

é um estimador não-viessado para σ^2 ($\mathbb{E}(\hat{\sigma}^2) = \sigma^2$).

Estimação da variância do erro

Muitas vezes existe uma confussão entre `erro` e `resíduo`.

Estimação da variância do erro

Muitas vezes existe uma confussão entre `erro` e `resíduo`.

Erros vs. Resíduo

- u no modelo populacional é o erro, $\hat{u}$ são os resíduos e aparecem na equação estimada
- Os erros (u) são não-observáveis, já os resíduos ($\hat{u}$) são calculados a partir dos dados
- $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i$

Leituras recomendadas

Leituras recomendadas

- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 2.4-2.6**