Modelos de Regressão e Previsão

Modelo de Regressão Linear Simples

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

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

Aula 3

Interpretação dos coeficientes

Unidades de medida

```
CEOSAL1 = read.table("./DadosMRP/ceosal1.txt")[,c(1,4)]
colnames(CEOSAL1) = c("salario", "roe")
CEOSAL1$salariodol = CEOSAL1$salario*1000
CEOSAL1$roedec = CEOSAL1$roe/100
head(CEOSAL1)
```

```
##   salario  roe salariodol roedec
## 1    1095 14.1    1095000  0.141
## 2    1001 10.9    1001000  0.109
## 3    1122 23.5    1122000  0.235
## 4     578  5.9     578000  0.059
## 5    1368 13.8    1368000  0.138
## 6    1145 20.0    1145000  0.200
```

Unidades de medida

$salariodol = salario \times 1000$ e $roedec = roe/100$

```
coef(lm(salario~roe, data = CEOSAL1))
```

```
## (Intercept)         roe
##   963.19133    18.50119
```

```
coef(lm(salariodol~roe, data = CEOSAL1))
```

```
## (Intercept)         roe
##   963191.33    18501.19
```

```
coef(lm(salario~roedec, data = CEOSAL1))
```

```
## (Intercept)      roedec
##    963.1913   1850.1187
```

Unidades de medida

Importante

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

- 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

Importante

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

- 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

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(salario~roe, data = CEOSAL1))$r.squared
```

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

```
summary(lm(salariodol~roe, data = CEOSAL1))$r.squared
```

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

```
summary(lm(salario~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
- 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

$$\% \Delta y \approx 100 \times \Delta \log(y) = (100\beta_1)\Delta x \quad (1)$$

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

Não linearidade

```
WAGE1 = read.table("./DadosMRP/wage1.txt")[,c(1,2)]
colnames(WAGE1) = c("salario", "educacao")
coef(lm(log(salario)~educacao, data = WAGE1))
```

```
## (Intercept)    educacao
##  0.58377267  0.08274437
```

$$\%\Delta Salario \approx (100\beta_1)\Delta educacao$$

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

Não linearidade

Modelo de elasticidade constante

Elasticidade de y em relação a x: variação percentual de y quando x aumenta 1%.

$$\frac{\Delta y}{\Delta x}\frac{x_0}{y_0} = \frac{\%\Delta y}{\%\Delta x}$$

Não linearidade

Modelo de elasticidade constante

Elasticidade de y em relação a x: variação percentual de y quando x aumenta 1%.

$$\frac{\Delta y}{\Delta x}\frac{x_0}{y_0} = \frac{\%\Delta y}{\%\Delta x}$$

Se usarmos a aproximação (1), temos que:

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

Não linearidade

Modelo de elasticidade constante

Elasticidade de y em relação a x: variação percentual de y quando x aumenta 1%.

$$\frac{\Delta y}{\Delta x}\frac{x_0}{y_0} = \frac{\%\Delta y}{\%\Delta x}$$

Se usarmos a aproximação (1), temos que:

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

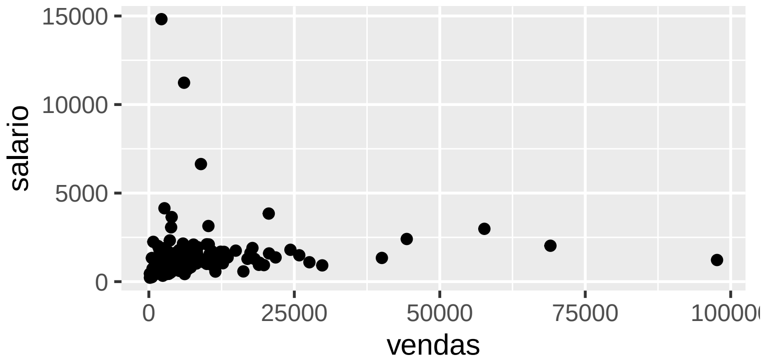
Então, na regressão

$$\log(y) = \beta_0 + \beta_1 \log(x) + u,$$

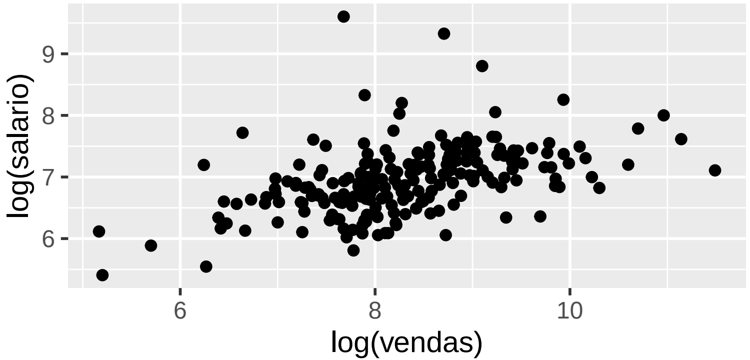
β_1 é a elasticidade de y em relação a x.

Não linearidade

```
CEOSAL1 = read.table("./DadosMRP/ceosal1.txt")[,c(1,3)]
colnames(CEOSAL1) = c("salario", "vendas")
```



Não linearidade



Não linearidade

```
CEOSAL1 = read.table("./DadosMRP/ceosal1.txt")[,c(1,3)]
colnames(CEOSAL1) = c("salario", "vendas")
coef(lm(log(salario)~log(vendas),data = CEOSAL1))
```

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

Não linearidade

```
CEOSAL1 = read.table("./DadosMRP/ceosal1.txt")[,c(1,3)]
colnames(CEOSAL1) = c("salario", "vendas")
coef(lm(log(salario)~log(vendas),data = CEOSAL1))
```

```
## (Intercept) log(vendas)
##   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. Ind	V. Dep	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

Hipótese RLS1: Lineariedade nos parâmetros

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

Hipótese RLS2: Amostragem aleatória

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

Hipótese RLS3: Variância amostral da variavel independênte

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

Propriedades dos Estimadores: Hipóteses

Hipótese RLS1: Lineariedade nos parâmetros

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

Hipótese RLS2: Amostragem aleatória

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

Hipótese RLS3: Variância amostral da variavel independênte

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

Hipótese RLS4: Média condicional zero

$$E(u|x) = 0$$

Propriedades dos Estimadores: EMQO são não-viesados

MQO são não-viessados

Sobre HRLS1 – HRLS4,

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

Propriedades dos Estimadores: EMQO são não-viesados

MQO são não-viessados

Sobre HRLS1 – HRLS4,

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

Prova

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

Propriedades dos Estimadores: EMQO são não-viesados

MQO são não-viessados

Sobre HRLS1 – HRLS4,

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

Prova

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

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

Propriedades dos Estimadores: EMQO são não-viesados

(. . . cont) Prova

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

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

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

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

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

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

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

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

Propriedades dos Estimadores: EMQO são não-viesados

(...cont) Prova

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

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

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

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

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 $E(\cdot|x)$ em ambos os lados.

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

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 $E(\cdot|x)$ em ambos os lados.

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

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

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$, *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 $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$, *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 $\hat{\beta}_1$ ($\hat{\beta}_0$) esteja de β_1 (β_0)

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$, *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 $\hat{\beta}_1$ ($\hat{\beta}_0$) esteja de β_1 (β_0)
- Em particular, queremos conhecer $V(\hat{\beta}_0)$, $V(\hat{\beta}_1)$

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$, *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 $\hat{\beta}_1$ ($\hat{\beta}_0$) esteja de β_1 (β_0)
- Em particular, queremos conhecer $V(\hat{\beta}_0)$, $V(\hat{\beta}_1)$

Propriedades dos Estimadores: variância dos EMQO

- Já sabemos que $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$, *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 $\hat{\beta}_1$ $(\hat{\beta}_0)$ esteja de β_1 (β_0)
- Em particular, queremos conhecer $V(\hat{\beta}_0)$, $V(\hat{\beta}_1)$

Hipótese RLS5: Homocedasticidade

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

Propriedades dos Estimadores: variância dos EMQO

Variância dos EMQO

Sobre HRLS1 – HRLS5,

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

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

Prova

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

Propriedades dos Estimadores: variância dos EMQO

Variância dos EMQO

Sobre HRLS1 – HRLS5,

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

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

Prova

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

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

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_1|x) = V\Big(\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}|x\Big) = \frac{V\big(\sum_{i=1}^{n}(x_i-\bar{x})u_i|x\big)}{\big[\sum_{i=1}^{n}(x_i-\bar{x})^2\big]^2}$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_1|x) = V\Big(\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}|x\Big) = \frac{V\big(\sum_{i=1}^n (x_i - \bar{x})u_i|x\big)}{\big[\sum_{i=1}^n (x_i - \bar{x})^2\big]^2}$$

Como os u_is são independentes (HRLS2)

$$V(\hat{\beta}_1|x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \overbrace{V(u_i|x)}^{\sigma^2}}{\big[\sum_{i=1}^n (x_i - \bar{x})^2\big]^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

Sabemos que,

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

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

Sabemos que,

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

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\bar{x} = \sum_{i=1}^{n}\Big(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big)y_i$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_0|x) = \sum_{i=1}^{n}\Big(\frac{1}{n} - \frac{(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big)^2 \underbrace{V(\overbrace{\beta_0+\beta_1 x_i+u_i}^{y_i}|x)}_{\sigma^2}$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_0|x) = \sum_{i=1}^{n} \Big(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big)^2 \underbrace{V(\overbrace{\beta_0 + \beta_1 x_i + u_i}^{y_i}|x)}_{\sigma^2}$$

$$= \sigma^2 \textstyle\sum_{i=1}^{n} \Big[\frac{1}{n^2} + \Big(\frac{(x_i - \bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big)^2 - 2\frac{(x_i - \bar{x})\bar{x}}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big]$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_0|x) = \sum_{i=1}^{n}\Big(\frac{1}{n} - \frac{(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big)^2 \underbrace{V(\overbrace{\beta_0+\beta_1 x_i + u_i}^{y_i}|x)}_{\sigma^2}$$

$$= \sigma^2 \textstyle\sum_{i=1}^{n}\Big[\frac{1}{n^2} + \Big(\frac{(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big)^2 - 2\frac{(x_i-\bar{x})\bar{x}}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big]$$

$$= \sigma^2\Big[\frac{n}{n^2} + \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2\bar{x}^2}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\Big] = \sigma^2\Big[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big]$$

Propriedades dos Estimadores: variância dos EMQO

(...continuação) Prova

$$V(\hat{\beta}_0|x) = \sigma^2 \Big[\frac{\overbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}^{\sum_{i=1}^n x_i^2 - n\bar{x}^2} + n\bar{x}^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} \Big] = \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

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 $\hat{\sigma}^2$.

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
- $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i$

Estimação da variância do erro

- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} u_i^2}{n}$?

Estimação da variância do erro

- $\hat{\sigma}^2 = \frac{\sum_{i=1}^n u_i^2}{n}$?
- Nunca conhecemos u

Estimação da variância do erro

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

Estimação da variância do erro

- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} u_i^2}{n}$?
- Nunca conhecemos u
- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n}$?
- É um estimador de σ^2, porém ele é viesado $(E(\hat{\sigma}^2) \neq \sigma^2)$

Estimação da variância do erro

- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} u_i^2}{n}$?
- Nunca conhecemos u
- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n}$?
- É um estimador de σ^2, porém ele é viesado ($E(\hat{\sigma}^2) \neq \sigma^2$)

Estimação da variância do erro

- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} u_i^2}{n}$?
- Nunca conhecemos u
- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n}$?
- É um estimador de σ^2, porém ele é viesado ($E(\hat{\sigma}^2) \neq \sigma^2$)

Estimação não-viesada de σ^2

Sobre HRLS1–HRLS5,

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

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

Estimação da variância do erro: Prova

Sabemos que

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i, \tag{2}$$

então, (somando os n elementos e dividindo po n)

Estimação da variância do erro: Prova

Sabemos que

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i, \tag{2}$$

então, (somando os n elementos e dividindo po n)

$$\underbrace{\bar{\hat{u}}}_{0} = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}, \tag{3}$$

Estimação da variância do erro: Prova

Sabemos que

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i, \tag{2}$$

então, (somando os n elementos e dividindo po n)

$$\underbrace{\bar{\hat{u}}}_{0} = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}, \tag{3}$$

Fazendo (2) - (3),

$$\hat{u}_i = u_i - \bar{u} - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x}), \tag{4}$$

Estimação da variância do erro: Prova

Sabemos que

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i, \tag{2}$$

então, (somando os n elementos e dividindo po n)

$$\underbrace{\bar{\hat{u}}}_{0} = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}, \tag{3}$$

Fazendo (2) - (3),

$$\hat{u}_i = u_i - \bar{u} - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x}), \tag{4}$$

Logo,

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

Estimação da variância do erro: Prova

Como $\sum_{i=1}^{n}(u_i - \bar{u})(x_i - \bar{x}) = \sum_{i=1}^{n} u_i(x_i - \bar{x})$

Estimação da variância do erro: Prova

Como $\sum_{i=1}^{n}(u_i - \bar{u})(x_i - \bar{x}) = \sum_{i=1}^{n} u_i(x_i - \bar{x})$

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

Estimação da variância do erro: Prova

Como $\sum_{i=1}^{n}(u_i - \bar{u})(x_i - \bar{x}) = \sum_{i=1}^{n} u_i(x_i - \bar{x})$

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

Pela lineariedade da Esperança, teremos três termos:

- **Primeiro termo:** $E(\sum_{i=1}^{n}(u_i - \bar{u})^2|x)$
- **Segundo termo:** $E((\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^{n}(x_i - \bar{x})^2|x)$
- **Terceiro termo:** $E(2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{n} u_i(x_i - \bar{x})|x)$

Estimação da variância do erro: Prova

Primeiro termo

$$\sum_{i=1}^{n}(u_i-\bar{u})^2 = \sum_{i=1}^{n}(u_i^2+\bar{u}^2-2u_i\bar{u}) = \sum_{i=1}^{n} u_i^2 + n\bar{u} - 2\bar{u}\underbrace{\sum_{i=1}^{n} u_i}_{n\bar{u}} = \sum_{i=1}^{n} u_i^2 - n\bar{u}^2$$

Estimação da variância do erro: Prova

Primeiro termo

$$\sum_{i=1}^{n}(u_i - \bar{u})^2 = \sum_{i=1}^{n}(u_i^2 + \bar{u}^2 - 2u_i\bar{u}) = \sum_{i=1}^{n} u_i^2 + n\bar{u} - 2\bar{u}\underbrace{\sum_{i=1}^{n} u_i}_{n\bar{u}} = \sum_{i=1}^{n} u_i^2 - n\bar{u}^2$$

Aplicando $E(\cdot|x)$ em ambos os lados

$$E\Big[\sum_{i=1}^{n}(u_i - \bar{u})^2|x\Big] = E\Big[\sum_{i=1}^{n} u_i^2|x\Big] - nE[\bar{u}^2] = \underbrace{\sum_{i=1}^{n} E(u_i^2|x)}_{n\sigma^2} - n\underbrace{E[\bar{u}^2]}_{\frac{\sigma^2}{n}} = (n-1)\sigma^2$$

Estimação da variância do erro: Prova

Segundo termo

$$E((\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 | x) = \sum_{i=1}^{n} (x_i - \bar{x})^2 \underbrace{E((\hat{\beta}_1 - \beta_1)^2 | x)}_{V(\hat{\beta}_1 | x) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

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

Estimação da variância do erro: Prova

Terceiro termo

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

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

Estimação da variância do erro: Prova

Terceiro termo

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

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

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

Estimação da variância do erro: Prova

Terceiro termo

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

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

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

Aplicando $E(\cdot|x)$ em ambos os lados,

$$E\left[2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^n u_i(x_i - \bar{x})|x\right] = \\ 2\sum_{i=1}^n (x_i - \bar{x})^2 \underbrace{E\left[(\hat{\beta}_1 - \beta_1)^2|x\right]}_{V(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = 2\sigma^2$$

Estimação da variância do erro: Prova

Juntanto os resultados do **primeiro**, **segundo** e **terceiro** termo,

$$E(\sum_{i=1}^{n} \hat{u}_i^2|x) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

Estimação da variância do erro: Prova

Juntanto os resultados do **primeiro**, **segundo** e **terceiro** termo,

$$E(\sum_{i=1}^{n} \hat{u}_i^2 | x) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

Como $E(E(\cdot|x)) = E(\cdot)$,

Estimação da variância do erro: Prova

Juntanto os resultados do **primeiro**, **segundo** e **terceiro** termo,

$$E(\sum_{i=1}^{n} \hat{u}_i^2|x) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

Como $E(E(\cdot|x)) = E(\cdot)$,

$$E\Big[E(\sum_{i=1}^{n} \hat{u}_i^2|x)\Big] = E\Big(\sum_{i=1}^{n} \hat{u}_i^2\Big) = (n-2)\sigma^2$$

Estimação da variância do erro: Prova

Juntanto os resultados do **primeiro**, **segundo** e **terceiro** termo,

$$E(\sum_{i=1}^{n} \hat{u}_i^2|x) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

Como $E(E(\cdot|x)) = E(\cdot)$,

$$E\Big[E(\sum_{i=1}^{n} \hat{u}_i^2|x)\Big] = E\Big(\sum_{i=1}^{n} \hat{u}_i^2\Big) = (n-2)\sigma^2$$

Logo, $E\Big(\frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}\Big) = \sigma^2$

Leituras recomendadas

Leituras recomendadas

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