ACA228 - Modelos de Regressão e Previsão

Regressão Linear Multipla

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

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

Aula 7

Introdução

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$
- Contudo, pensar que uma única variável X pode explicar Y é bastante ingênuo

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$
- Contudo, pensar que uma única variável X pode explicar Y é bastante ingênuo
- $\mathbb{E}(u|X) = 0$ implica $Cov(u, X) = 0$

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$
- Contudo, pensar que uma única variável X pode explicar Y é bastante ingênuo
- $\mathbb{E}(u|X) = 0$ implica $Cov(u, X) = 0$
- No contexto de RLS, dizer que todos os outros fatores que afetam Y (e incorporados em u) são não correlacionados com X é bastante irrealista.

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$
- Contudo, pensar que uma única variável X pode explicar Y é bastante ingênuo
- $\mathbb{E}(u|X) = 0$ implica $Cov(u, X) = 0$
- No contexto de RLS, dizer que todos os outros fatores que afetam Y (e incorporados em u) são não correlacionados com X é bastante irrealista.
- Um modelo que inclua mais do que uma variavél explicativa parece ser bem mais razoável.

Introdução

- Nas aulas 3 e 5 focamos em modelos do tipo $Y = \beta_0 + \beta_1 X + u$
- Contudo, pensar que uma única variável X pode explicar Y é bastante ingênuo
- $\mathbb{E}(u|X) = 0$ implica $Cov(u, X) = 0$
- No contexto de RLS, dizer que todos os outros fatores que afetam Y (e incorporados em u) são não correlacionados com X é bastante irrealista.
- Um modelo que inclua mais do que uma variavél explicativa parece ser bem mais razoável.
- Se incluirmos no nosso modelo mais variáveis que sejam úteis para explicar Y, esperamos que uma maior parte da variabilidade de Y seja explicada.

Regressão Linear Multipla

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente
- $X_1, \cdots, X_k$ são as k variáveis explicativas

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente
- $X_1, \cdots, X_k$ são as k variáveis explicativas
- $\beta = [\beta_0 \quad, \beta_1 \quad \ldots \quad \beta_k]'$ é o vetor de parâmetros

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente
- $X_1, \cdots, X_k$ são as k variáveis explicativas
- $\beta = [\beta_0 \quad , \beta_1 \quad \ldots \quad \beta_k]'$ é o vetor de parâmetros
- u é o termo de erro.

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente
- $X_1, \cdots, X_k$ são as k variáveis explicativas
- $\beta = [\beta_0 \quad , \beta_1 \quad \ldots \quad \beta_k]'$ é o vetor de parâmetros
- u é o termo de erro.

Regressão Linear Multipla

MRLM

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

- Y é a variavél dependente
- $X_1, \cdots, X_k$ são as k variáveis explicativas
- $\beta = [\beta_0 \quad , \beta_1 \quad \ldots \quad \beta_k]'$ é o vetor de parâmetros
- u é o termo de erro.

Na prática, nunca conhecemos os β's, então precisamos estimá-los utilizando os dados

Método MQO

Estimação

Sejam $(y_1, x_{1,1}, \ldots, x_{1,k})$, $\ldots$ $(y_n, x_{n,1}, \ldots, x_{n,k})$ uma a.a. de tamanho n.

Estimação

Sejam $(y_1, x_{1,1}, \dots, x_{1,k})$, ... $(y_n, x_{n,1}, \dots, x_{n,k})$ uma a.a. de tamanho n.

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_k x_{1,k} + u_1 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_k x_{n,k} + u_n
\end{aligned} \tag{1}
$$

Estimação

Sejam $(y_1, x_{1,1}, \ldots, x_{1,k})$, $\ldots$ $(y_n, x_{n,1}, \ldots, x_{n,k})$ uma a.a. de tamanho n.

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_k x_{1,k} + u_1 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_k x_{n,k} + u_n
\end{aligned} \tag{1}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

Estimação

Sejam $(y_1, x_{1,1}, \dots, x_{1,k})$, $\dots$ $(y_n, x_{n,1}, \dots, x_{n,k})$ uma a.a. de tamanho n.

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_k x_{1,k} + u_1 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_k x_{n,k} + u_n \end{aligned} \tag{1}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

Em forma matricial

$$Y = X\beta + u$$

Estimação

Queremos os $b_0, b_1, \cdots, b_k$ que minimizem

$$SQR = \sum_{i=1}^{n} \hat{u}_i^2$$

em que $\hat{u}_i = y_i - \hat{y}_i$ com $\hat{y}_i = b_0 + b_1 x_{i,1} + \cdots, + b_k x_{,k}$.

Estimação

Queremos os $b_0, b_1, \cdots, b_k$ que minimizem

$$SQR = \sum_{i=1}^{n} \hat{u}_i^2$$

em que $\hat{u}_i = y_i - \hat{y}_i$ com $\hat{y}_i = b_0 + b_1 x_{i,1} + \cdots, +b_k x_{,k}$.

$$\begin{aligned}
\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k]' &= \underset{b}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{u}_i^2 \\
&= \underset{b}{\operatorname{argmin}} \sum_{i=1}^{n} (\underbrace{y_i - b_0 - b_1 x_{i,1} - \cdots - b_k x_{i,k}}_{\hat{u}_i})^2 \qquad (2)
\end{aligned}$$

Estimação

Equivalentemente, a equação (2) em forma matricial é dada por:

$$\hat{\beta} = \underset{b}{\operatorname{argmin}}\, \hat{u}'\hat{u} = \underset{b}{\operatorname{argmin}}(Y - \underbrace{Xb}_{\hat{Y}})'(Y - \underbrace{Xb}_{\hat{Y}})$$

Estimação

Equivalentemente, a equação (2) em forma matricial é dada por:

$$\hat{\beta} = \underset{b}{\text{argmin}}\ \hat{u}'\hat{u} = \underset{b}{\text{argmin}}(Y - \underbrace{Xb}_{\hat{Y}})'(Y - \underbrace{Xb}_{\hat{Y}})$$

Para minimizar SQR, igualamos a primeira derivada a zero e depois verificamos se é ponto de mínimo (segunda derivada).

Estimação

Equivalentemente, a equação (2) em forma matricial é dada por:

$$\hat{\beta} = \underset{b}{\operatorname{argmin}}\, \hat{u}'\hat{u} = \underset{b}{\operatorname{argmin}} (Y - \underbrace{Xb}_{\hat{Y}})'(Y - \underbrace{Xb}_{\hat{Y}})$$

Para minimizar SQR, igualamos a primeira derivada a zero e depois verificamos se é ponto de mínimo (segunda derivada).

$$\begin{aligned}
\text{SQR} &= \hat{u}'\hat{u} && (3)\\
&= (Y - Xb)'(Y - Xb) && (4)\\
&= (Y' - b'X')(Y - Xb) && (5)\\
&= Y'Y - b'X'Y - Y'Xb + b'X'Xb && (6)
\end{aligned}$$

Estimação

Regras de derivação para matrices/vetores:

Sejam $a_{n\times 1}$, $b_{n\times 1}$ e $A_{n\times n}$

$$\frac{\partial b'Ab}{\partial b} = (A+A')b; \quad \frac{\partial Ab}{\partial b'} = A; \quad \frac{\partial a'b}{\partial b} = \frac{\partial b'a}{\partial b} = a$$

Estimação

Regras de derivação para matrices/vetores:

Sejam $a_{n\times 1}$, $b_{n\times 1}$ e $A_{n\times n}$

$$\frac{\partial b'Ab}{\partial b} = (A + A')b; \quad \frac{\partial Ab}{\partial b'} = A; \quad \frac{\partial a'b}{\partial b} = \frac{\partial b'a}{\partial b} = a$$

$$\begin{aligned}
\frac{\partial SQR}{\partial b} &= \frac{\partial(Y'Y - b'X'Y - Y'Xb + b'X'Xb)}{\partial b} && (7)\\
&= \frac{\partial Y'Y}{\partial b} - \frac{\partial b'X'Y}{\partial b} - \frac{\partial Y'Xb}{\partial b} + \frac{\partial b'X'Xb}{\partial b} && (8)\\
&= 0 - X'Y - (Y'X)' + (X'X + X'X)b && (9)\\
&= -2X'Y + 2X'Xb && (10)
\end{aligned}$$

Estimação

$$\frac{\partial SQR}{\partial b} = -2X'Y + 2X'Xb$$

Estimação

$$\frac{\partial SQR}{\partial b} = -2X'Y + 2X'Xb$$

Igualando a zero e cancelando o -2 , temos

$$X'Y = X'X\hat{\beta} \quad \text{ou equivalentemente} \quad \underbrace{(X'X)^{-1}X'X}_{I}\hat{\beta} = (X'X)^{-1}X'Y$$

Estimação

$$\frac{\partial SQR}{\partial b} = -2X'Y + 2X'Xb$$

Igualando a zero e cancelando o -2 , temos

$$X'Y = X'X\hat{\beta} \quad \text{ou equivalentemente} \quad \underbrace{(X'X)^{-1}X'X}_{I}\hat{\beta} = (X'X)^{-1}X'Y$$

Será que é ponto de mínimo? (Segundas derivadas...)

Estimação

$$\frac{\partial SQR}{\partial b} = -2X'Y + 2X'Xb$$

Igualando a zero e cancelando o -2 , temos

$$X'Y = X'X\hat{\beta} \quad \text{ou equivalentemente} \quad \underbrace{(X'X)^{-1}X'X}_{I}\hat{\beta} = (X'X)^{-1}X'Y$$

Será que é ponto de mínimo? (Segundas derivadas. . .)

$$\frac{\partial}{\partial b'}\left(\frac{\partial SQR}{\partial b}\right) = \frac{\partial^2 SQR}{\partial b \partial b'} = \frac{\partial(-2X'Y + 2X'Xb)}{\partial b'} = 2X'X \geq 0,$$

então $\hat{\beta}$ é ponto de mínimo.

Estimação

Estimador MQO

$$\hat{\beta} = [X'X]^{-1}X'Y$$

Estimação

Estimador MQO

$$\hat{\beta} = [X'X]^{-1}X'Y$$

Note que,

$$\begin{aligned}
\hat{\beta} &= [X'X]^{-1}X'Y = \\
&= [X'X]^{-1}X'(\underbrace{X\beta + u}_{Y}) \\
&= \underbrace{[X'X]^{-1}X'X}_{I}\beta + [X'X]^{-1}X'u \\
&= \beta + [X'X]^{-1}X'u
\end{aligned} \tag{11}$$

Estimação

```r
library(wooldridge)
modelo <- lm(log(wage) ~ educ + exper + tenure, data = wage1)
coef(modelo)
```

```
## (Intercept)        educ       exper      tenure
## 0.284359555 0.092028987 0.004121109 0.022067217
```

Estimação

```r
library(wooldridge)
modelo <- lm(log(wage) ~ educ + exper + tenure, data = wage1)
coef(modelo)
```

```
## (Intercept)        educ      exper      tenure
## 0.284359555 0.092028987 0.004121109 0.022067217
```

- Mantendo os fatores *exper* e *tenure* fixos, quando *educ* (anos de educção formal) aumenta em 1, espera-se que o salário aumente em $9.2\%(100 \times 0.092028987)$

Qualidade de ajuste

RLM: Qualidade de ajuste - R2

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

R2

$$R^2 = 1 - SQR/SQT = SQE/SQT$$

RLM: Qualidade de ajuste - R2

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

R2

$$R^2 = 1 - SQR/SQT = SQE/SQT$$

- R^2 já foi introduzido na RLS.

RLM: Qualidade de ajuste - R2

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

R2

$$R^2 = 1 - SQR/SQT = SQE/SQT$$

- R^2 já foi introduzido na RLS.
- R^2 : proporção da variabilidade de y que é explicada pelo modelo.

RLM: Qualidade de ajuste - R2

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

R2

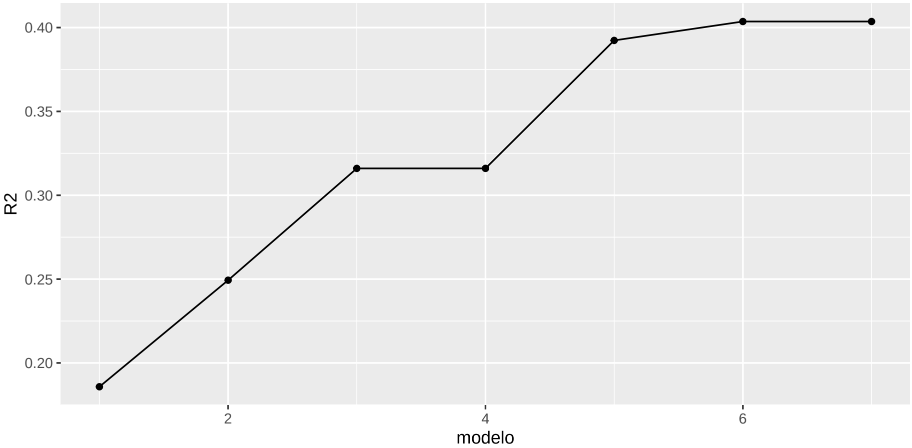
$$R^2 = 1 - SQR/SQT = SQE/SQT$$

- R^2 já foi introduzido na RLS.
- R^2 : proporção da variabilidade de y que é explicada pelo modelo.
- $100 \times R^2$: porcentagem da variabilidade de y que é explicada pelo modelo.

RLM: Qualidade de ajuste - R2

```
modelo1 = lm(log(wage)~educ,data = wage1)
modelo2 = lm(log(wage)~educ+exper,data = wage1)
modelo3 = lm(log(wage)~educ+exper+tenure,data = wage1)
modelo4 = lm(log(wage)~educ+exper+tenure+nonwhite,data = wage1)
modelo5 = lm(log(wage)~educ+exper+tenure+nonwhite
             +female,data = wage1)
modelo6 = lm(log(wage)~educ+exper+tenure+nonwhite
             +female + married,data = wage1)
modelo7 = lm(log(wage)~educ+exper+tenure+nonwhite+female
             +married+numdep,data = wage1)
```

RLM: Qualidade de ajuste - R2



RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)

RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)
- Uma alternativa é usar uma nova medida de qualidade de ajuste.

RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)
- Uma alternativa é usar uma nova medida de qualidade de ajuste.

RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)
- Uma alternativa é usar uma nova medida de qualidade de ajuste.

R2-Ajustado

$$R_A^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)
- Uma alternativa é usar uma nova medida de qualidade de ajuste.

R2-Ajustado

$$R_A^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

R^2-Ajustado penaliza o número de variáveis incluidas no modelo.

RLM: Qualidade de ajuste -R2 ajustado

- R^2 tem a desvantagem que nunca diminui quando incluimos uma nova variavel no modelo (mesmo se a variável não for importante)
- Uma alternativa é usar uma nova medida de qualidade de ajuste.

R2-Ajustado

$$R_A^2 = 1 - \frac{n-1}{n-(k+1)}(1 - R^2)$$

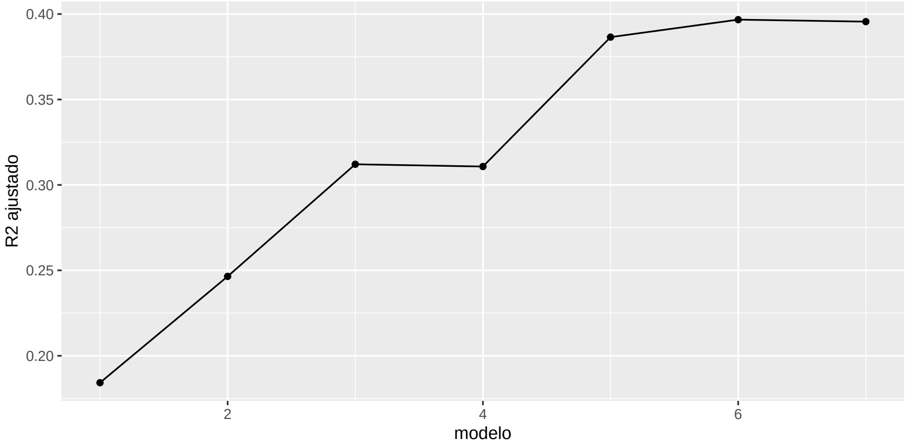
R^2-Ajustado penaliza o número de variáveis incluidas no modelo.

No R:

```
summary(modelo1)$adj.r.squared
```

```
## [1] 0.1842527
```

RLM: Qualidade de ajuste -R2 ajustado



RLM: Qualidade de ajuste - R2 ajustado

Comparando R^2 com R^2-ajustado

```
#R2
round(R2,4)
```

```
## [1] 0.1858 0.2493 0.3160 0.3160 0.3923 0.4036 0.4036
```

```
#R2-Ajustado
round(R2adj,4)
```

```
## [1] 0.1843 0.2465 0.3121 0.3108 0.3865 0.3967 0.3956
```

RLM: Qualidade de ajuste - R2 ajustado

Comparando R^2 com R^2-ajustado

```
#R2
round(R2,4)
```

```
## [1] 0.1858 0.2493 0.3160 0.3160 0.3923 0.4036 0.4036
```

```
#R2-Ajustado
round(R2adj,4)
```

```
## [1] 0.1843 0.2465 0.3121 0.3108 0.3865 0.3967 0.3956
```

R^2 ou R^2-ajustado? Prefere-se o R^2-ajustado.

Propriedades do estimador MQO

Propriedades

HRLM1: Linear nos parâmetros

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u \tag{12}$$

Propriedades

HRLM1: Linear nos parâmetros

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u \tag{12}$$

HRLM2: Amostragem aleatória

$(y_1, x_{1,1}, \ldots, x_{1,k}), \cdots, (y_n, x_{n,1}, \ldots, x_{n,k})$ constituem uma a.a. de tamanho n do modelo populacional (12)

Propriedades

HRLM1: Linear nos parâmetros

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u \qquad (12)$$

HRLM2: Amostragem aleatória

$(y_1, x_{1,1}, \ldots, x_{1,k}), \cdots, (y_n, x_{n,1}, \ldots, x_{n,k})$ constituem uma a.a. de tamanho n do modelo populacional (12)

HRLM3: Colinearidade não perfeita

Não há relações linerares exatas entre as variáveis independêntes e nenhuma das variáveis independentes é constante.

RLM: Propriedades

HRLM4: Média condicional zero

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

RLM: Propriedades

HRLM4: Média condicional zero

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

Teorema: Inexistência do viés MQO

Sob HRLM1–HRLM4,

$$\mathbb{E}(\hat{\beta}) = \beta$$

RLM: Propriedades

Prova

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'u$$

RLM: Propriedades

Prova

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'u$$

$$\begin{aligned}
\mathbb{E}(\hat{\beta}|X) &= \mathbb{E}(\beta + (X'X)^{-1}X'u|X) && (13)\\
&= \underbrace{\mathbb{E}(\beta|X)}_{\beta} + \underbrace{\mathbb{E}((X'X)^{-1}X'u|X)}_{(X'X)^{-1}X'\underbrace{\mathbb{E}(u|X)}_{0}} && (14)\\
&= \beta + 0 = \beta && (15)
\end{aligned}$$

RLM: Propriedades

Prova

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'u$$

$$\mathbb{E}(\hat{\beta}|X) = \mathbb{E}(\beta + (X'X)^{-1}X'u|X) \tag{13}$$

$$= \underbrace{\mathbb{E}(\beta|X)}_{\beta} + \underbrace{\mathbb{E}((X'X)^{-1}X'u|X)}_{(X'X)^{-1}X'\underbrace{\mathbb{E}(u|X)}_{0}} \tag{14}$$

$$= \beta + 0 = \beta \tag{15}$$

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

$$\underbrace{\mathbb{E}[\mathbb{E}(\hat{\beta}|X)]}_{\mathbb{E}(\hat{\beta})} = \mathbb{E}[\beta] = \beta$$

RLM: Propriedades

HRLM5: Variância constante (Homocedasticidade)

$$\mathbb{V}(u|X) = \mathbb{E}[uu'|X] = \sigma^2 I$$

RLM: Propriedades

HRLM5: Variância constante (Homocedasticidade)

$$\mathbb{V}(u|X) = \mathbb{E}[uu'|X] = \sigma^2 I$$

Variância dos EMQO

Sob HRLM1–HRLM5,

$$\mathbb{V}(\hat{\beta}|X) = \sigma^2 (X'X)^{-1} \tag{16}$$

RLM: Propriedades

Prova

Sabemos que $\hat{\beta} = \beta + [X'X]^{-1}X'u \quad \longrightarrow \quad \hat{\beta} - \beta = [X'X]^{-1}X'u$

RLM: Propriedades

Prova

Sabemos que $\hat{\beta} = \beta + [X'X]^{-1}X'u \quad \longrightarrow \quad \hat{\beta} - \beta = [X'X]^{-1}X'u$

$$
\begin{aligned}
\mathbb{V}(\hat{\beta}|X) &= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] \\
&= \mathbb{E}\Big[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'|X\Big] \\
&= \mathbb{E}[(X'X)^{-1}X'uu'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X'\underbrace{\mathbb{E}[uu'|X]}_{\sigma^2 I}X(X'X)^{-1} \\
&= \sigma^2\underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
\tag{17}
$$

RLM: Propriedades

- Na prática, σ^2 não é conhecido, então precisamos estima-lo

RLM: Propriedades

- Na prática, σ^2 não é conhecido, então precisamos estima-lo
- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} \hat{u}_i^2}{n - (k+1)}$

RLM: Propriedades

- Na prática, σ^2 não é conhecido, então precisamos estima-lo
- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} \hat{u}_i^2}{n - (k+1)}$
- Pode-se mostrar que, sob HRLM1–HRLM5[1], $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$

[1] As hipóteses HRLM1–HRLM5 são conhecidas como Hipóteses de Gauss-Markov

Teorema de Gauss-Markov

Teorema de Gauss-Markov

Sob as hipóteses HRLM1–HRLM5, $\hat{\beta}$ é o melhor estimador linear não viesado (Best Linear Unbiased Estimator –BLUE) de β.

Teorema de Gauss-Markov

Teorema de Gauss-Markov

Sob as hipóteses HRLM1–HRLM5, $\hat{\beta}$ é o melhor estimador linear não viesado (Best Linear Unbiased Estimator –BLUE) de β.

- $\tilde{\beta}$ é **linear** se $\tilde{\beta} = A'Y$, onde $A_{n\times(k+1)}$ função de X

Teorema de Gauss-Markov

Teorema de Gauss-Markov

Sob as hipóteses HRLM1–HRLM5, $\hat{\beta}$ é o melhor estimador linear não viesado (Best Linear Unbiased Estimator –BLUE) de β.

- $\tilde{\beta}$ é **linear** se $\tilde{\beta} = A'Y$, onde $A_{n\times(k+1)}$ função de X
- **Melhor:** menor variância. $V(\hat{\beta}|X) \leq V(\tilde{\beta}|X)$, para qualquer estimador linear não viesado $\tilde{\beta}$

MRLM: Desafíos na prática

Incluir variáveis irrelevantes

Incluir mais variáveis das que deveríamos.

Incluir variáveis irrelevantes

Incluir mais variáveis das que deveríamos.

Suponha que o modelo populacional seja

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

mas estimamos

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

Qual é o efeito de incluir a variavel irrelevante x_3?

- Em termos da inexistencia do viés não há ifeito nenhum $E(\hat{\beta}) = [\beta_0, \beta_1, \beta_2, 0]'$

Incluir variáveis irrelevantes

Incluir mais variáveis das que deveríamos.

Suponha que o modelo populacional seja

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

mas estimamos

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

Qual é o efeito de incluir a variavel irrelevante x_3?

- Em termos da inexistencia do viés não há ifeito nenhum $E(\hat{\beta}) = [\beta_0, \beta_1, \beta_2, 0]'$
- Contudo, incluir variáveis irrelevantes pode ter efeitos indesejáveis nas variâncias dos estimadores MQO.

Incluir variáveis irrelevantes

$$\underbrace{Y = 2 + 1.2X_1 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2}_{\text{Modelo Estimado}}$$

Incluir variáveis irrelevantes

$$\underbrace{Y = 2 + 1.2X_1 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2}_{\text{Modelo Estimado}}$$

```r
dados_sim <-  simular_dados_ir(n = 1000,rho = 0,
                              betas = c(2,1.2))
summary(lm(y~x1, data = dados_sim))$coefficients
```

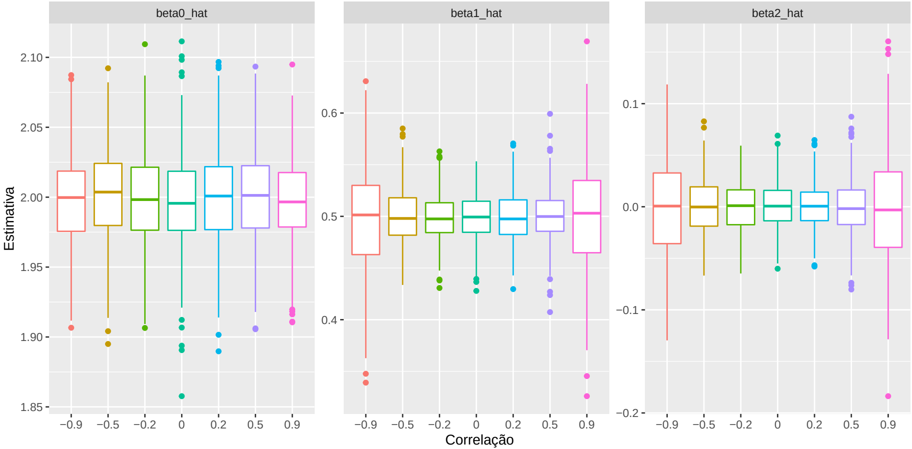
```
##             Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 1.995151 0.02203071 90.56229        0
## x1          1.202188 0.02218742 54.18333        0
```

```r
summary(lm(y~x1+x2, data = dados_sim))$coefficients
```

```
##                Estimate Std. Error   t value  Pr(>|t|)
## (Intercept)  1.99557544 0.02205083 90.498900 0.0000000
## x1           1.20206971 0.02219281 54.164836 0.0000000
## x2          -0.01091434 0.02161174 -0.505019 0.6136012
```

Incluir variáveis irrelevantes

$$\underbrace{Y = 2 + 1.2X_1 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2}_{\text{Modelo Estimado}}$$



Omissão de variáveis

Considerar menos variáveis do que deveríamos.

Omissão de variáveis

Considerar menos variáveis do que deveríamos.

Suponha que o modelo populacional seja

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

mas estimamos

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

Qual é o efeito de omitir a variavel X_3?

Omissão de variáveis

Considerar menos variáveis do que deveríamos.

Suponha que o modelo populacional seja

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

mas estimamos

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

Qual é o efeito de omitir a variavel X_3?

- Em geral, produz estimadores viesados

Omissão de variáveis

$$\underbrace{Y = 2 + 0.5X_1 + 1.2X_2 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1}_{\text{Modelo Estimado}}$$

Omissão de variáveis

$$\underbrace{Y = 2 + 0.5X_1 + 1.2X_2 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1}_{\text{Modelo Estimado}}$$

```
dados_sim <- simular_dados_om(n = 1000, rho = 0.7,
                              betas = c(2,0.5,1.2))
coef(lm(y ~ x1 + x2, data = dados_sim))
```

```
## (Intercept)          x1          x2
##   2.0342043   0.4791449   1.1564827
```

```
coef(lm(y ~ x1, data = dados_sim))
```

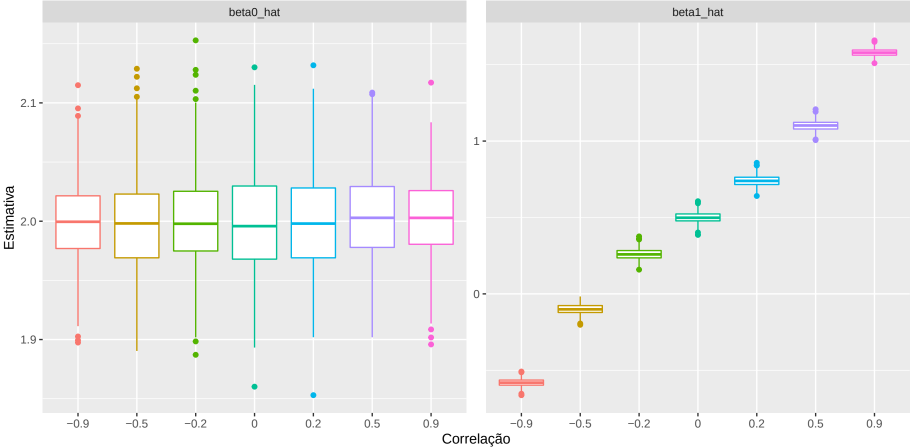
```
## (Intercept)          x1
##    2.032879    1.312008
```

Omissão de variáveis

$$\underbrace{Y = 2 + 0.5X_1 + 1.2X_2 + u}_{\text{Modelo Verdadeiro}} \quad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1}_{\text{Modelo Estimado}}$$

Omissão de variáveis

$$\underbrace{Y = 2 + 0.5X_1 + 1.2X_2 + u}_{\text{Modelo Verdadeiro}} \qquad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1}_{\text{Modelo Estimado}}$$



Multicolinearidade

- A HRLM3 nos diz que não existe colinearidade perfeita entra as variáveis independentes

Multicolinearidade

- A HRLM3 nos diz que não existe colinearidade perfeita entra as variáveis independentes
- Contudo, na prática podemos ter variaveis explicativas fortemente correlacionadas (mas $\neq \pm 1$)

Multicolinearidade

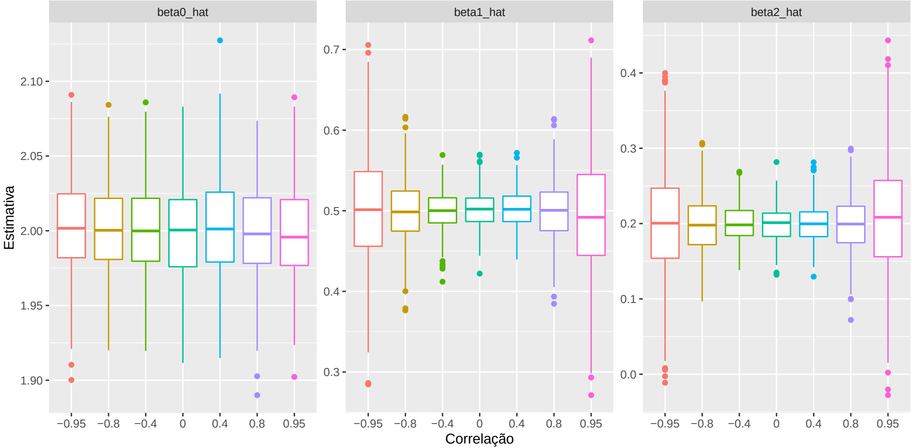
- A HRLM3 nos diz que não existe colinearidade perfeita entra as variáveis independentes
- Contudo, na prática podemos ter variaveis explicativas fortemente correlacionadas (mas $\neq \pm 1$)
- Este fenômeno é conhecido na literatura como **multicolinearidade**

Multicolinearidade

- A HRLM3 nos diz que não existe colinearidade perfeita entra as variáveis independentes
- Contudo, na prática podemos ter variaveis explicativas fortemente correlacionadas (mas $\neq \pm 1$)
- Este fenômeno é conhecido na literatura como **multicolinearidade**
- Multicolinearidade tem consequencias tanto na estimação dos parâmetros quanto na estimação das suas respectivas variâncias.

Multicolinearidade

$$\underbrace{Y = 2 + 0.5X_1 + 0.2X_2 + u}_{\text{Modelo Verdadeiro}} \qquad \underbrace{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2}_{\text{Modelo Estimado}}$$



Multicolinearidade: VIF

O Fator de Inflação de Variância (**V**ariance **I**nflation **F**actor –VIF) é um critério para identificar se estamos com problema de multicolineariedade.

Multicolinearidade: VIF

O Fator de Inflação de Variância (**V**ariance **I**nflation **F**actor –VIF) é um critério para identificar se estamos com problema de multicolineariedade.

VIF

$$VIF_j = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o R^2 da regressão da j-ésima variavel (x_j) sobre as outras variáveis regressoras.

Multicolinearidade: VIF

O Fator de Inflação de Variância (**V**ariance **I**nflation **F**actor –VIF) é um critério para identificar se estamos com problema de multicolineariedade.

VIF

$$VIF_j = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o R^2 da regressão da j-ésima variavel (x_j) sobre as outras variáveis regressoras.

- Valores grandes de VIF_j podem indicar multicolinearidade

Multicolinearidade: VIF

O Fator de Inflação de Variância (**V**ariance **I**nflation **F**actor –VIF) é um critério para identificar se estamos com problema de multicolineariedade.

VIF

$$VIF_j = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o R^2 da regressão da j-ésima variavel (x_j) sobre as outras variáveis regressoras.

- Valores grandes de VIF_j podem indicar multicolinearidade
- **Quanto é grande?** Algumas vezes 10 é considerado grande, mas nao existe uma regra (10 representaria um $R_j^2 = 0.9$)

Multicolinearidade: VIF

```r
library(car)
dados_sim <-simular_dados_mul(n = 1000,
                              rho = 0.995,
                              betas = c(2,0.5,0.2))
modelo <-  lm(y~x1+x2, data = dados)
vif(modelo)
```

```
##       x1       x2
## 9.857007 9.857007
```

Multicolinearidade: VIF

```r
modelo = lm(log(wage) ~ educ + exper + tenure, data = wage1)
vif(modelo)
```

```
##     educ    exper   tenure
## 1.112771 1.477618 1.349296
```

O que fazer?

- Se tivermos evidencia de multicolineariedade, devemos excluir as variaveis que causan a multicolineariedade, ou formar novas variáveis que sejam uma combinação das variáveis originais (por exemplo: utilizar *análise de componentes principais*)

O que fazer?

- Se tivermos evidencia de multicolineariedade, devemos excluir as variaveis que causan a multicolineariedade, ou formar novas variáveis que sejam uma combinação das variáveis originais (por exemplo: utilizar *análise de componentes principais*)
- Variáveis irrelevantes causam um problema maior se as variaveis irrelevantes incluidas forem altamente correlacionada. Podemos calcular a matriz de correlação, verificar se existem variáveis altamente correlacionadas e excluí-las da análise. Podemos também utilizar o VIF para verificar se a inclussão de alguma variavel causa problemas de multicolineariedade

O que fazer?

- Se tivermos evidencia de multicolineariedade, devemos excluir as variaveis que causan a multicolineariedade, ou formar novas variáveis que sejam uma combinação das variáveis originais (por exemplo: utilizar *análise de componentes principais*)
- Variáveis irrelevantes causam um problema maior se as variaveis irrelevantes incluidas forem altamente correlacionada. Podemos calcular a matriz de correlação, verificar se existem variáveis altamente correlacionadas e excluí-las da análise. Podemos também utilizar o VIF para verificar se a inclussão de alguma variavel causa problemas de multicolineariedade
- O problema de variáveis omitidas é mais complicado de resolver, pois muitas vezes está além do nosso alcance. Podemos tentar incluir novas variáveis no modelo (se tivermos).

O que fazer?

- Se tivermos evidencia de multicolineariedade, devemos excluir as variaveis que causan a multicolineariedade, ou formar novas variáveis que sejam uma combinação das variáveis originais (por exemplo: utilizar *análise de componentes principais*)
- Variáveis irrelevantes causam um problema maior se as variaveis irrelevantes incluidas forem altamente correlacionada. Podemos calcular a matriz de correlação, verificar se existem variáveis altamente correlacionadas e excluí-las da análise. Podemos também utilizar o VIF para verificar se a inclussão de alguma variavel causa problemas de multicolineariedade
- O problema de variáveis omitidas é mais complicado de resolver, pois muitas vezes está além do nosso alcance. Podemos tentar incluir novas variáveis no modelo (se tivermos).
- Para evitar o problema de variáveis omitidas é importante pensar (antes mesmo de obter os dados) nas variáveis que podem ser importantes (por isso é importante conhecermos o problema com que

Leituras recomendadas

Leituras recomendadas

- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 3**