

Introdução a Econometria 6ª Edição - (Wooldridge)

Análise da decisão de políticas com agrupamentos de cortes transversais - Exemplo 13.3

Fábio Rocha | rochaviannafj@gmail.com

24 de janeiro, 2023

Efeito da localização de um incinerador de lixo sobre os preços dos imóveis.

A ideia geral que o exercício traz é reproduzir uma análise econométrica empreendida por Kiel e McClain no sentido de verificar a influencia da construção de um incinerador numa determinada região da cidade sobre os preços dos imóveis. Wooldridge assume que a análise produzida pelos autores foi um tanto mais sofisticada e complexa mas ele traz de forma mais simples e didática.

O que se passa é que na cidade de North Andover, no ano de 1978, havia rumores de que um novo incinerador de lixo seria construído por lá. Entretanto, só foi iniciar a construção em 1981. O incinerador iniciou a operação em 1985.

O exercício então, consiste em avaliar duas amostras de dados, uma de 1978 e outra de 1981, contendo preços e localizações dos imóveis. A hipótese geral é de que imóveis próximos (4,8km) ao incinerador teriam preços mais baixos em relação aos mais distantes.

Começamos então carregando os dados para reproduzir o exercício.¹

Carregamento de pacotes - pacotes necessário para o exercício

```
library(tidyverse)
library(wooldridge)
library(stargazer)
```

Carregar o dado do exercício a partir do pacote wooldridge

```
data("kielmc")
```

Um analista ingenuo, diz Wooldridge, usaria apenas a amostra para os anos 1981 e estimaria uma Regressão Simples conforme mostramos abaixo.

Filtrando os dados apenas para 1981

```
kielmc_1981 <- kielmc %>%
  filter(year == 1981)
```

Estimando o modelo de Regressão Linear Simples - 1981

De forma geral o modelo teórico é este:

¹Há duas formas de carregar os dados: a primeira é carregando a partir do pacote wooldridge, a segunda é indo até o site da Cengage e baixando todos por lá na seção onde se encontra o livro Introdução a Econometria 6ª Edição.

$$rprice = \beta_0 + \beta_1 nearinc + \varepsilon \quad (1)$$

Em que : *rprice* é o preço dos imóveis, *beta zero* o preço médio dos imóveis que não estão próximos ao incinerador, *beta um* o coeficiente angular da função e *nearinc* indica a proximidade do imóvel ou não ao incinerador. **nearinc** é uma variável binária e assume 1 se for próximo ao incinerador e 0 se distante.

```
modelo.kielmc_1981 <- lm(rprice ~ nearinc, kielmc_1981)

stargazer::stargazer(modelo.kielmc_1981,
  font.size = 'tiny',
  title='Regressão Linear Simples',
  header=FALSE)
```

Table 1: Regressão Linear Simples

<i>Dependent variable:</i>	
	<i>rprice</i>
<i>nearinc</i>	-30,688.270*** (5,827.709)
Constant	101,307.500*** (3,093.027)
Observations	142
R ²	0.165
Adjusted R ²	0.159
Residual Std. Error	31,238.040 (df = 140)
F Statistic	27.730*** (df = 1; 140)

Note: * p<0.1; ** p<0.05; *** p<0.01

O modelo estimado na tabela, por conter apenas uma variável dependente e binária, mostra o seguinte; O intercepto é preço médio dos imóveis afastados do incinerador, ou seja, o preço médio desses imóveis é de 101.307,5. Já o coeficiente *nearinc* indica que o preço dos imóveis próximos ao incinerador é 30.688,27 a menos do que os que estão distantes do incinerador: 101.307,5 - 30.688,27 = 70.619,23

A estatística t é significativa, o que implica dizer que podemos rejeitar com certeza a hipótese de que os preços médios de venda dos imóveis situados próximos do incinerador e daqueles distantes deles sejam os mesmos, diz Wooldridge, e o R quadrado de 0.165 (16,5%) - o que implica dizer que a variabilidade de *nearinc* explica 16.5% da variabilidade de *rprice*.

Infelizmente este modelo simples não implica que a causa dos preços médio de venda dos imoveis próxima do incinerador seja de fato a sua localização. Basta reproduzirmos esse mesmo exercício, mas com o conjunto de dados para 1978, antes da construção do incinerador, para vermos isso, conforme orienta o exercício de Wooldridge.

Filtrando os dados apenas para 1978

```
kielmc_1978 <- kielmc %>%
  filter(year == 1978)
```

Estimando o modelo de Regressão Linear Simples - 1978

```
modelo.kielmc_1978 <- lm(rprice ~ nearinc, kielmc_1978)

stargazer::stargazer(modelo.kielmc_1978,
  font.size = 'tiny',
  title='Regressão Linear Simples',
  header=FALSE)
```

Table 2: Regressão Linear Simples

		Dependent variable:
		rprice
nearinc		-18,824.370*** (4,744.594)
Constant		82,517.230*** (2,653.790)
Observations		179
R ²		0.082
Adjusted R ²		0.076
Residual Std. Error		29,431.960 (df = 177)
F Statistic		15.741*** (df = 1; 177)
Note:		*p<0.1; **p<0.05; ***p<0.01

Como se vê, mesmo antes de qualquer comentário sobre a construção do incinerador, o preço médio de venda de um imóvel próximo do local já era \$18.824,37 menor que um imóvel distante do incinerador, conforme mostra a tabela 2 para o coeficiente *nearinc*.

Identificando o efeito sobre o preço

A forma pela qual se verifica como a construção do incinerador reduziu o preço médio de venda dos imóveis próximo, é observando como o coeficiente *nearinc* mudou entre 1978 e 1981. Isto é, a diferença entre as médias $30.688,27 - 18.824,37 = 11.863,9$.

Essa é a estimativa do efeito da construção do incinerador sobre o preço dos imóveis próximos a ele. Esse valor é chamado em economia empírica de **estimador de diferença em diferenças**

Entretanto, precisamos identificar esse valor encontrado é estatisticamente significativo (diferente de zero) precisamos estimar uma nova regressão. Podendo ser assim estimado:

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + \varepsilon \quad (2)$$

Em que :

- *rprice* é o preço dos imóveis,
- *beta_0*, intercepto, preço médio de um imóvel distante do incinerador em 1978.
- *delta_0* indica as alterações em todos os valores dos imóveis de 1978 a 1981. Variável binária igual a um se 1981 e zero se 1978.
- *beta_1* mede o efeito da localização, que não é em razão da presença do incinerador. Variável binária igual a um se for próximo e zero se distante.
- Finalmente o parâmetro de interesse é o termo de interação entre as variáveis **y81 multiplicado por nearinc**. *delta_1* mede o declínio dos valores dos imóveis em razão do novo incinerador, supondo que tanto os imóveis distantes e próximos não tenham sido valorizados por outras razões a taxas diferentes.

Estimando o modelo com ambos períodos 1978-1981

```
modelo.kielmc <- lm(rprice ~ y81 + nearinc + y81nrinc, kielmc)

stargazer::stargazer(modelo.kielmc,
                      font.size = 'tiny',
                      title='Regressão Linear Multipla',
                      header=FALSE)
```

O modelo estimado retronou o *intercepto/constante* como significativo; *y81*, que representa a variação média dos dos preços de vendas dos imóveis entre 1978 e 1981, também significativo, e *nearinc*, que indica se o imóvel está próximo ou não do incinerador, também significativo. Entretando o *delta*, que é a interação entre *y81* e *nearinc* não foi significativo a 10% - talvez na margem, como diz Wooldridge, seria.

Table 3: Regressão Linear Multipla

	Dependent variable:
	rprice
y81	18,790.290*** (4,050.065)
nearinc	-18,824.370*** (4,875.322)
y81nrinc	-11,863.900 (7,456.646)
Constant	82,517.230*** (2,726.910)
Observations	321
R ²	0.174
Adjusted R ²	0.166
Residual Std. Error	30,242.900 (df = 317)
F Statistic	22.251*** (df = 3; 317)
Note: *p<0.1; **p<0.05; ***p<0.01	

Inserção de variáveis de controle

Kiel e McClain, diz Wooldridge, inseriram mais variáveis (características) dos imóveis em suas análises, por duas razões. A primeira é que os imóveis vendidos em 1978 podem ter características distintas dos de 1981. Desta forma, inserir mais variáveis que contam com essas características podem ser relevantes para **controlar** essas diferenças. Por outro lado, embora tais características não apresentem grande relevância, isto é, não ser tao diferente entre os anos, a inserção dessas características pode reduzir a variancia do erro e por sua vez o erro padrão (usado para o calculo da estatística t)

Vejamos como fica o nosso quarto modelo com variáveis de controle. Nesse quarto modelo, controlaremos apenas com a variável idade do imóvel (mais essa variável ao quadrado já que ela consegue capturar o efeito de valorização e desvalorização do imovel ao longo do tempo).

Conforme vemos na tabela 4, a simples inserção da variável *age* (idade do imóvel) já muda significativamente o R-quadrado, além do erro padrão e a intensidade do coeficiente *y81.nrinc*, que por sua vez passa ser significativo.

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + \beta_2 age + \beta_3 age^2 + \varepsilon \quad (3)$$

```
modelo.kielmc.age <- lm(rprice ~y81 + nearinc + y81nrinc + age + I(age^2), kielmc)
stargazer::stargazer(modelo.kielmc.age,
  font.size = 'tiny',
  title='Regressão Linear Multipla',
  header=FALSE)
```

Table 4: Regressão Linear Multipla

	Dependent variable:
	rprice
y81	21,321.040*** (3,443.631)
nearinc	9,397.936* (4,812.222)
y81nrinc	-21,920.270*** (6,359.745)
age	-1,494.424*** (131.860)
I(age^2)	8.691*** (0.848)
Constant	89,116.540*** (2,406.051)
Observations	321
R ²	0.414
Adjusted R ²	0.405
Residual Std. Error	25,543.290 (df = 315)
F Statistic	44.591*** (df = 5; 315)

Note: *p<0.1; **p<0.05; ***p<0.01

Por fim o quinto modelo inserimos mais variáveis para controlar, conforme consta no exemplo do livro. São inseridos *intst*, distancia até a rodovia estadual, área do terreno *land*, área construída *area*, o número de quartos *rooms* e o número de banheiros *baths*. Essa estimativa contendo essas variáveis de controle produz uma estimativa mais próxima daquela inicial sem nenhum controle.

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + \beta_2 intst + \beta_3 land + \beta_4 area + \beta_5 rooms + \beta_6 baths + \beta_7 age + \beta_8 age^2 + \varepsilon \quad (4)$$

```
modelo.kielmc.controlado <- lm(rprice ~ y81 + nearinc +
                               y81nrinc + intst + land + area +
                               rooms + baths + age + I(age^2), kielmc)

stargazer(modelo.kielmc.controlado,
           font.size = 'tiny',
           title='Regressão Linear Multipla',
           header=FALSE)
```

As estimativas do modelo 5 são preferidas já que possuem maior controle dos fatores que afetam os preços e possuem menores erro padrão, o que leva a estatística t e significancia estatística dos parametros estimados ser significativa.

Se bem observarmos veremos que o coeficiente *nearinc* diminuiu em relação a modelo anterior e passou a não ser significativo. Isso mostra que as características incluídas no ultimo modelo indicam em grande parte as características dos imóveis mais importantes para determinação dos preços dos imóveis.

Table 5: Regressão Linear Multipla

	<i>Dependent variable:</i>
	rprice
y81	13,928.480*** (2,798.747)
nearinc	3,780.337 (4,453.415)
y81nrinc	-14,177.930*** (4,987.267)
intst	-0.539*** (0.196)
land	0.141*** (0.031)
area	18.086*** (2.306)
rooms	3,304.227** (1,661.248)
baths	6,977.317*** (2,581.321)
age	-739.451*** (131.127)
I(age^2)	3.453*** (0.813)
Constant	13,807.670 (11,166.590)
Observations	321
R ²	0.660
Adjusted R ²	0.649
Residual Std. Error	19,619.020 (df = 310)
F Statistic	60.189*** (df = 10; 310)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Vejamos os três modelos juntos numa unica tabela:

```
stargazer(modelo.kielmc, modelo.kielmc.age, modelo.kielmc.controlado,
          font.size = 'tiny',
          title='Regressão Linear Multipla',
          header=FALSE)
```

De fato é possível perceber que depois de controlado, *y81nearinc* teve seu coeficiente aumentado (negativamente) e passou ser significativo, enquanto que naquele terceiro modelo isso não acontecia embora tinha valor exatamente ao calculado pelas diferenças de médias dos modelos 1 e 2. Isso significa que de fato a localização próxima ao incinerador teve algum efeito sobre os preços de venda dos imóveis após a construção. E ao que tudo indica o ultimo modelo estimado o número de quartos e banheiros representam um grande peso na formação dos preços de venda.

Table 6: Regressão Linear Multipla

	Dependent variable:		
	(1)	rprice (2)	(3)
y81	18,790.290*** (4,050.065)	21,321.040*** (3,443.631)	13,928.480*** (2,798.747)
nearinc	-18,824.370*** (4,875.322)	9,397.936* (4,812.222)	3,780.337 (4,453.415)
y81nrinc	-11,863.900 (7,456.646)	-21,920.270*** (6,359.745)	-14,177.930*** (4,987.267)
intst			-0.539*** (0.196)
land			0.141*** (0.031)
area			18.086*** (2.306)
rooms			3,304.227** (1,661.248)
baths			6,977.317*** (2,581.321)
age		-1,494.424*** (131.860)	-739.451*** (131.127)
I(age^2)		8.691*** (0.848)	3.453*** (0.813)
Constant	82,517.230*** (2,726.910)	89,116.540*** (2,406.051)	13,807.670 (11,166.590)
Observations	321	321	321
R ²	0.174	0.414	0.660
Adjusted R ²	0.166	0.405	0.649
Residual Std. Error	30,242.900 (df = 317)	25,543.290 (df = 315)	19,619.020 (df = 310)
F Statistic	22.251*** (df = 3; 317)	44.591*** (df = 5; 315)	60.189*** (df = 10; 310)

Note: *p<0.1; **p<0.05; ***p<0.01

Verificando esse efeito em termos percentuais

```
modelo.kielmc <- lm(log(rprice) ~ y81 + nearinc + y81nrinc, kielmc)

stargazer::stargazer(modelo.kielmc,
  font.size = 'tiny',
  title='Regressão Linear Multipla',
  header=FALSE)
```

Table 7: Regressão Linear Multipla

	Dependent variable:
	log(rprice)
y81	0.193*** (0.045)
nearinc	-0.340*** (0.055)
y81nrinc	-0.063 (0.083)
Constant	11.285*** (0.031)
Observations	321
R ²	0.246
Adjusted R ²	0.239
Residual Std. Error	0.338 (df = 317)
F Statistic	34.470*** (df = 3; 317)

Note: *p<0.1; **p<0.05; ***p<0.01

O termo de interação do modelo 3, *y81nearinc* indica a redução percentual aproximada dos imóveis próximos do incinerador. Isso implica dizer que, os imóveis próximos ao incinerador desvalorizaram na média aproximadamente 6,3%. Porém já vimos que esse coeficiente assim estimado não é estatisticamente significativo. A

forma pela qual contornamos esse problema é controlar por outras variáveis (características) do modelo cinco²

```
modelo.kielmc.controlado <- lm(log(price) ~ y81 + nearinc +
                               y81nrinc + log(intst) + log(land) + log(area) +
                               rooms + baths + age + I(age^2), kielmc)

stargazer(modelo.kielmc.controlado,
           font.size = 'tiny',
           title='Regressão Linear Multipla',
           header=FALSE)
```

Table 8: Regressão Linear Multipla

	Dependent variable:
	log(price)
y81	0.426*** (0.028)
nearinc	0.032 (0.047)
y81nrinc	-0.132** (0.052)
log(intst)	-0.061* (0.032)
log(land)	0.100*** (0.024)
log(area)	0.351*** (0.051)
rooms	0.047*** (0.017)
baths	0.094*** (0.028)
age	-0.008*** (0.001)
I(age^2)	0.00004*** (0.00001)
Constant	7.652*** (0.416)
Observations	321
R ²	0.790
Adjusted R ²	0.784
Residual Std. Error	0.204 (df = 310)
F Statistic	116.909*** (df = 10; 310)
Note: *p<0.1; **p<0.05; ***p<0.01	

Assim, é possível constatar que a desvalorização dos imóveis próximo ao incinerador foi de 13,2% depois de sua construção.

²Conforme orientação do professor Wooldridge, é necessário colocar as variáveis *intst*, *land* e *area* em log.