

Introdução a Econometria - Wooldridge

Agrupamento independente de cortes transversais ao longo do tempo

Fábio Rocha - rochaviannafj@gmail.com*

14 de abril, 2024

```
library(dplyr)
library(wooldridge)
library(tidyverse)

data("fertil1")
attach(fertil1)
```

Fertilidade feminina ao longo do tempo

Exercício 13.1

Utilizaremos os dados FERTIL1 para estimar um modelo que explique o número total de nascimentos por mulheres (kids) conforme especificado no modelo abaixo. A questão especificamente delimitada é: **após controlarmos por fatores observáveis, o que aconteceu com a natalidade ao longo do tempo?** Os fatores são: anos de educação idade, raça, região do país que as mulheres viviam quando tinham 16 anos e ambiente que viviam quanto tinham essa mesma idade.

$$\begin{aligned} \text{kids} = & \alpha + \beta_1 \text{educ} + \beta_2 \text{age} + \beta_3 (\text{age}^2) + \beta_4 \text{black} + \beta_5 \text{east} \\ & + \beta_6 \text{northcen} + \beta_7 \text{west} + \beta_8 \text{farm} + \beta_9 \text{othrural} + \beta_{10} \text{town} \\ & + \beta_{11} \text{smcity} + \delta_1 \text{y74} + \delta_2 \text{y76} + \delta_3 \text{y78} + \delta_4 \text{y80} \\ & + \delta_5 \text{y82} + \delta_6 \text{y84} + u \end{aligned} \tag{1}$$

Em que:

- kids = número de filhos
- educ = anos de educação
- age = idade
- black = variável dummy (1 se preto/pardo/idígena, 0 caso contrário).
- east = variável dummy (1 se vivia no leste aos 16 anos de idade, 0 caso contrário).
- northcen = variável dummy 1 se morava no centro-norte aos 16 anos, 0 caso contrário).
- west = variável dummy (1 se morava no oeste aos 16 anos, 0 caso contrário).
- farm = variável dummy (1 se estava em uma fazenda aos 16 anos, 0 caso contrário).
- othrural = variável dummy (1 se em outra área rural aos 16 anos, 0 caso contrário).
- town = variável dummy (1 se morava em uma cidade aos 16 anos, 0 caso contrário).
- smcity = variável dummy (1 se em uma pequena cidade aos 16 anos, 0 caso contrário).
- y74 = variável dummy (1 se o ano for 74, 0 caso contrário).
- u = termo de erro aleatório (variáveis que afetam kids mas não estão especificadas no modelo)

*Filosofo e economista pela Universidade Federal de Ouro Preto. Mestrando em economia pela Universidade Federal do ABC e analista de pesquisa socioeducacional no Instituto Unibanco

O ano base é 1972. Os coeficientes das variáveis dummy anuais evidenciam uma queda da fertilidade no início dos anos 1980. y_{82} , por exemplo, mostra que mantendo educação, age e outros fatores constantes, uma mulher teve em média 0,52 menos filhos que em 1982 do que em 1972 (10 anos atrás); De forma mais intuitiva, 100 mulheres teriam 52 filhos a menos se comparadas com 100 mulheres de 1972.

Como estamos controlando pelas variáveis educação, essa queda é separada do declínio da fertilidade em razão do aumento dos níveis de educação. (a média de anos de educação em 1972 é de 12,2 e 1984, 13,3). Os coeficientes y_{82} e y_{84} representam queda na fertilidade por razões que não são captadas nas variáveis explicativas.

Analisando B1, nota-se que mulheres com mais anos de escolaridade tem menor número de filhos. Um outro exercício que consta no Wooldridge é simular o que acontece com número de filhos de mulheres com apenas ensino médio, comparado com mulheres com ensino superior. Isto é, como os anos médios de estudo nesse conjunto de dados varia entre 12,2 e 13,3 - o que provavelmente é o suficiente para indicar escolaridade ensino médio. O que ele faz é adicionar 4 anos ao coeficiente negativo estimado.

$-0,128 \times (4) = 0,512$ que corresponde ao período médio de uma formação superior. Assim, sugere que mantendo os demais fatores constantes, 100 mulheres com ensino superior terão 51 filhos a menos que 100 mulheres apenas com ensino médio.

```
M1 <- lm(kids ~ 1 + educ + age + I(age^2) + black + east +
         northcen + west + farm + othrural + town + smcity +
         y74 + y76 + y78 + y80 + y82 + y84)

stargazer::stargazer(M1,
                     font.size = 'tiny',
                     header=FALSE)
```

Interação de variável dummy anual com variáveis explicativas.

Se quisermos identificar mudanças dos efeitos parciais das variáveis explicativas mudou ao longo do tempo, a interação dessas variáveis pode ser uma relevante estratégia.

Mudanças no retorno da educação e a diferença salarial por gênero.

Considere o seguinte modelo:

$$\log(\text{wage}) = \beta_0 + \delta_0 y_{85} + \beta_1 \text{educ} + \delta_1 \text{educ} * y_{85} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{union} + \beta_5 \text{female} + \delta_5 \text{female} * y_{85} + u \quad (2)$$

Em que:

- $\log(\text{wage})$ = salário por hora
- educ = anos de educação
- exper = experiência
- union = variável dummy (1 se pertence a um sindicato, 0 se o contrário).
- y_{85} = variável dummy (1 se o ano for o ano de 1985, 0 caso contrário).
- u = termo de erro aleatório (variáveis que afetam o salário mas não estão especificadas no modelo)

```
data("cps78_85")

M2 <- lm(lwage ~ y85 + educ + I(educ*y85) + exper + I(exper^2) +
         union + female + female*y85, cps78_85)

stargazer::stargazer(M2,
```

Table 1:

	<i>Dependent variable:</i>
	kids
educ	-0.128*** (0.018)
age	0.532*** (0.138)
I(age^2)	-0.006*** (0.002)
black	1.076*** (0.174)
east	0.217 (0.133)
northcen	0.363*** (0.121)
west	0.198 (0.167)
farm	-0.053 (0.147)
othrural	-0.163 (0.175)
town	0.084 (0.125)
smcity	0.212 (0.160)
y74	0.268 (0.173)
y76	-0.097 (0.179)
y78	-0.069 (0.182)
y80	-0.071 (0.183)
y82	-0.522*** (0.172)
y84	-0.545*** (0.175)
Constant	-7.742** (3.052)
Observations	1,129
R ²	0.130
Adjusted R ²	0.116
Residual Std. Error	1.555 (df = 1111)
F Statistic	9.723*** (df = 17; 1111)
<i>Note:</i> * p<0.1; ** p<0.05; *** p<0.01	

```
font.size = 'tiny',
header=FALSE)
```

Table 2:

	Dependent variable:
	lwage
y85	0.118 (0.124)
educ	0.075*** (0.007)
I(educ *y85)	0.018** (0.009)
exper	0.030*** (0.004)
I(exper^2)	-0.0004*** (0.0001)
union	0.202*** (0.030)
female	-0.317*** (0.037)
y85:female	0.085* (0.051)
Constant	0.459*** (0.093)
Observations	1,084
R ²	0.426
Adjusted R ²	0.422
Residual Std. Error	0.413 (df = 1075)
F Statistic	99.804*** (df = 8; 1075)
Note: * p<0.1; ** p<0.05; *** p<0.01	

O intercepto de 1978 é β_0 , e o intercepto de 1985 é $\beta_0 + \delta_0$. O retorno da educação em 1978 é β_1 e o retorno da educação em 1985 é $\beta_1 + \delta_1$. Portanto, δ_1 mede como o retorno de mais um ano de estudo mudou ao longo do período de sete anos.

No que se refere à diferença no logaritmo do salário por hora ($\log(\text{wage})$) entre homens e mulheres em 1978, isso é representado por β_5 , enquanto que a diferença em 1985 é representada por $\beta_5 + \delta_5$.

O retorno da educação em 1978 é de aproximadamente 7,5%; enquanto que o retorno da educação em 1985 é de 1,8 mais alto, ou cerca de 9,3%. Já a diferença salarial por gênero, em 1978, mantendo os demais fatores constantes, as mulheres recebiam 31,7% menos que um homem (27,2% é uma estimativa mais precisa), enquanto que em 1985 essa diferença passou para 0,31 - 0,08 = 0,23, ou 23,2%. Isto é, a diferença salarial entre 1978 e 1985 caiu cerca de 8,5%

O que acontece se fizermos interação de variável dummy anual com todas as explicativas?

Isso é possível e equivale ao mesmo que estimar varios modelos com periodos de tempo diferentes. No nosso exercicio estimaos para os periodos 1978 e 1985, então se interassessmos todas as variáveis explicativas com y85 o resultado seria o mesmo verificado em dois modelos, um para 1978 e outro para 1985.

Exercícios complementares.

$$\begin{aligned} \log(\text{wage}) = & \beta_0 + \delta_0 y85 + \beta_1 \text{educ} + \delta_1 \text{educ} * y85 * \text{female} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 \\ & + \beta_4 \text{union} + \beta_5 \text{female} + \delta_5 \text{female} * y85 + u \end{aligned} \quad (3)$$

A interação entre *y85*, *female* e *educ* estimada indica que um ano a mais de educação para mulheres em 1985 teve efeito, mantendo os demais fatores constantes, de diminuir 1% na diferença salarial.

```
M3 <- lm(lwage ~ y85 + educ + I(educ*y85*female) + exper +
        I(exper^2) + union + female, cps78_85)

stargazer::stargazer(M3,
                      font.size = 'tiny',
                      header=FALSE)
```

Table 3:

	Dependent variable:
	lwage
y85	0.338*** (0.032)
educ	0.082*** (0.005)
I(educ *y85 *female)	0.010** (0.004)
exper	0.030*** (0.004)
I(exper^2)	-0.0004*** (0.0001)
union	0.204*** (0.030)
female	-0.338*** (0.036)
Constant	0.382*** (0.078)
Observations	1,084
R ²	0.426
Adjusted R ²	0.422
Residual Std. Error	0.413 (df = 1076)
F Statistic	114.114*** (df = 7; 1076)
Note:	*p<0.1; **p<0.05; ***p<0.01