

Lista – Variáveis Instrumentais

Curso de Microeconometria - IE-UFRJ, 2021.2

Prof. Romero Rocha

Regras gerais: este trabalho deverá ser entregue pelo classroom. O produto final a ser entregue constará de (i) um arquivo de texto (Word ou PDF) com as tabelas e as respostas escritas; e (ii) um arquivo de códigos no Stata ou no R.

O objetivo deste trabalho consiste em capacitá-los(as) no uso do Stata ou do R em aplicações sobre dados brasileiros, e fazê-los(as) pensar nos conceitos estudados no curso. A base de dados a ser disponibilizada será a do Censo 2000 para o estado do Rio de Janeiro.

O primeiro passo deste trabalho será a montagem de uma base de microdados (indivíduos + domicílios) para o município do Rio de Janeiro. Você poderá usar a documentação do Censo (dicionário, também anexado aos arquivos) para identificar e entender as variáveis da base.

A pergunta que guiará este trabalho é: o número de filhos é um determinante importante para a participação feminina e salário das mulheres no mercado de trabalho? Neste trabalho, vamos investigar mais a fundo o efeito de um terceiro filho na participação, no número de horas trabalhadas e na renda das mulheres. Aqui estão os passos a serem seguidos, e as questões a serem respondidas:

- 1) Eu já montei a base completa em dta para o município do Rio de Janeiro, fazendo a junção das bases do Censo para indivíduos e domicílio. Antes de começarmos, você deverá limpar a base, através da criação ou renomeação de variáveis, de forma a obter as seguintes variáveis (utilize a documentação do censo como referência):

d_mulher (dummy de sexo feminino); idade (idade em anos, variável contínua); idade_2 (idade em anos ao quadrado); educa (anos de estudo); branco (dummy de raça branca); trab (dummy de trabalho remunerado na semana de referência); salario (rendimento total no trabalho principal em reais, variável contínua); horas (número de horas trabalhadas no trabalho principal); sal_hr (rendimento por hora trabalhada); ln_sal_hr (log do rendimento por hora trabalhada); n_fil (número de filhos nascidos vivos); n_fil_o (número de filhos homens nascidos vivos); n_fil_a (número de filhas mulheres nascidas vivas); d_fil (dummy para número de filhos maior ou igual a um).

Por fim, selecione a amostra para mulheres com idade de 18 a 60 anos.

Responda: (i) quantas observações existem nessa amostra? (ii) summarize (calcule média, desvio-padrão, máximo e mínimo) as variáveis mencionadas acima (d_mulher até n_fil_a, com exceção de idade_2); (iii) por fim, calcule a média das variáveis trab, salario e horas para mulheres com e sem filhos; calcule a diferença entre essas médias e não se esqueça de calcular a significância estatística desta diferença. Apresente as tabulações, descreva o que você encontrou e interprete os resultados destas tabulações.

- 2) Rode as regressões abaixo e analise o coeficiente encontrado para d_fil.

Regressão da variável trab na variável d_fil (1)

Regressão da variável trab na variável d_fil, com erro padrão robusto (2)

Responda: (i) interprete o coeficiente d_fil da regressão (1) e o compare com a resposta encontrada na questão anterior; (ii) com base no coeficiente encontrado, você pode afirmar que ter filhos afeta a participação de mulheres no mercado de trabalho? Você consegue identificar um efeito causal? Justifique sua resposta; (iii) qual das duas regressões é preferível e por que?

- 3) Rode agora outras três regressões, todas com erro padrão robusto (a partir de agora coloque erro padrão robusto em todas as regressões até o final da lista).

Regressão de trab em d_fil, idade e idade_2 (3)

Regressão de trab em d_fil, idade, idade_2 e educa (4)

Regressão de trab em d_fil, idade, idade_2, educa e branco (5)

Responda: (i) compare os coeficientes da variável d_fil encontrados nas regressões (2) a (5), e interprete o que aconteceu em cada etapa: o coeficiente aumentou ou diminuiu (em termos absolutos)? Explique o que aconteceu em cada etapa (de (2) para (3), de (3) para (4), etc); (iii) qual regressão é preferível? Esta regressão nos fornece uma estimativa de efeito causal de ter filhos na participação de mulheres no mercado de trabalho?

- 4) Agora selecione a amostra para mulheres com apenas 2 ou 3 filhos.

Responda: repita o item (iii) da primeira questão para esta nova amostra selecionada, de forma a comparar a média das variáveis trab, salario e horas para mulheres com dois e três filhos e sua significância.

- 5) Em seguida, crie uma dummy chamada d_tresfil para número de filhos=3 e rode as regressões abaixo.

Regressão de trab em d_tresfil, idade, idade_2, educa e branco (6)

Regressão de horas em d_tresfil, idade, idade_2, educa e branco para os casos onde trab=1 (7)

Regressão de ln_sal_hr em d_tresfil, idade, idade_2, educa e branco para os casos onde trab=1 (8)

Responda: (i) interprete os coeficientes da variável d_tresfil nas regressões acima; (ii) exemplifique situações para as quais os efeitos estimados acima estariam viesados.

- 6) Iremos utilizar a metodologia de variáveis instrumentais de forma a nos aproximar do efeito causal desejado. A literatura costuma utilizar a combinação de sexo dos dois primeiros filhos como instrumento para decisão de se ter um terceiro filho. A ideia por trás disso é que casais que tiveram os dois primeiros filhos de sexos diferentes teriam menor probabilidade de tentar um terceiro filho.

Responda: (i) você considera a combinação de sexos dos dois primeiros filhos pode ser um bom instrumento? Por que? Explique de forma a contextualizar os requisitos abordados na teoria para este caso prático; (ii) em sociedades machistas, qual poderia ser um bom instrumento para decisão de um filho a mais? Explique.

- 7) Precisamos criar uma nova variável (d_filcas) para denotar nosso instrumento: dummy de um filho homem & uma filha mulher entre os dois primeiros filhos da mulher.

Orientações de como criar esta variável: Primeiramente, crie uma dummy chamada ult_fil_o para último filho homem e outra chamada ult_fil_a para última filha mulher a partir da variável v0464 disponível na base. Em seguida, crie a variável d_filcas (instrumento) com valor zero. Depois, substitua os valores dessa variável para 1 em caso de mulher com dois filhos, onde o número total de filhos homens (n_fil_o) é 1 e o número total de filhos mulheres (n_fil_a) também é 1. Finalmente, substitua o valor dessa variável novamente para 1 em caso de mulheres com três filhos onde o número total de filhos homens até o segundo filho é 1 e o número total de filhas mulheres até o segundo filho também é 1 (chega-se até essas últimas pela subtração entre as variáveis n_fil_o e ult_fil_o e subtração entre as variáveis n_fil_a e ult_fil_a).

Após ter criado a variável d_filcas, rode o 1º estágio abaixo:

Regressão de d_tresfil em d_filcas, idade, idade_2, educa e branco (9)

Responda: (i) qual a proporção de mulheres com dois e três filhos nesta amostra?; (ii) a relação entre a composição de sexos dos dois primeiros filhos com a decisão pelo terceiro filho se mostrou conforme esperada?; (iii) por que precisamos rodar o primeiro estágio? Você diria que o instrumento tem cara de ser um instrumento fraco ou um instrumento forte? Por quê?

- 8) Crie uma nova variável chamada d_tresfil_prev com o valor previsto para a probabilidade de se ter um terceiro filho a partir da especificação do 1º estágio e rode o segundo estágio, conforme abaixo:

Regressão de `trab` em `d_tresfil_prev`, `idade`, `idade_2`, `educa` e `branco` (10)

Regressão de horas em `d_tresfil_prev`, `idade`, `idade_2`, `educa` e `branco` para os casos onde `trab=1` (11)

Regressão de `ln_sal_hr` em `d_tresfil_prev`, `idade`, `idade_2`, `educa` e `branco` para os casos onde `trab=1` (12)

Responda: (i) podemos dizer que existe efeito de se ter um terceiro filho na participação feminina no mercado de trabalho? E no número de horas trabalhadas (dado que ela trabalha)? E no salário das mulheres no mercado de trabalho (dado que ela trabalha)? Interprete os resultados. Como você os explicaria? (Existe uma boa explicação para todos).

(ii) Vamos tentar criticar o instrumento. Qual seria uma razão pela qual esse instrumento não é bom nessas regressões. A composição dos dois primeiros filhos pode afetar a escolha de trabalho das mulheres por outros caminhos que não sejam um terceiro filho? Exemplifique

(iii) Eu não sei a resposta do item (ii) para a regressão de participação, mas vejo um problema (de seleção) grande nas duas outras regressões (horas e salários). Você conseguiria me explicar qual é esse problema e como ele pode estar afetando os coeficientes? Por que você acha que o instrumento pode não estar resolvendo esse problema?

- 9) **Responda:** Verifique que os resultados são os mesmos se você rodar estas mesmas regressões dos quesitos 7 e 8, mas usando diretamente os comandos de variáveis instrumentais do R ou do Stata em vez de rodar o primeiro estágio, calcular o valor previsto e rodar o segundo estágio. Aproveite e descubra algum teste de instrumentos fracos nesses comandos, apresente a estatística de teste e o p-valor do resultado do teste e interprete.