

Fecundidade em São Paulo

Maria Cruz e Guilherme Davi

1 Introdução

O objetivo deste trabalho é usar conhecimentos da disciplina de Econometria 1 para analisar a fecundidade da cidade de São Paulo. Fecundidade representa o potencial reprodutivo, normalmente abordado para determinada região. A taxa de fecundidade é o número médio de filhos por mulher em idade reprodutiva (15 a 49 anos).

2 Dados

A base possui 8146 observações e 13 variáveis que abordam indicadores pessoais como idade das mulheres, raça, aspectos de inserção ao mercado de trabalho, como atuação no trabalho doméstico, carga horária, salário e educação e por fim, variáveis de quantidade de filhos por mulher.

A faixa etária da base vai de 7 a 93 anos de idade, como esse grupo não é considerado nas análises de fecundidade de um determinado local pois não está inserido na faixa de idade reprodutiva, consideramos dois modelos por recorte de faixa etária: a base completa, e a de idade reprodutiva.

A ideia é que a população de mulheres mais novas que 15 anos e mais velhas que 49 anos não são representativas do potencial de reprodução do local, afinal nessas idades questões comportamentais e intrínsecas a sua saúde reprodutiva afetam a possibilidade de conceber um filho.

Assim, para a base com amostra de mulheres em idade reprodutiva temos 6465 observações.

3 Literatura

Sabe-se que a fecundidade depende de variáveis de escolha e condição das mulheres. A Pesquisa Nacional de Demografia e Saúde (PNDS) comparou níveis de fecundidade entre 1996 e 2006 para o Brasil e como evidência é possível observar que os níveis de fecundidade diminuíram em média para o país. Esse fenômeno esteve presente para mulheres entre 20 e 24 anos, entretanto para mulheres menores que 20 anos, tais níveis não caíram. Além disso, a pesquisa apresenta que houve uma diminuição na idade do primeiro filho para a região nordeste, que possuía maiores níveis de fecundidade para mulheres com menos de 20 anos.

Este fenômeno de queda da fecundidade é denominado transição da fecundidade. As principais mudanças que acarretam esse processo são inserção feminina no mercado de trabalho, maior acesso a educação, maior taxa de urbanização, maior acesso a contraceptivos.

Os primeiros estudos são de Becker (1991), em que se propõe a ideia de maximizar o número de crianças dada restrição econômica,

- Black (2008): crianças como bens normais: mais renda aumenta fecundidade;
- Ponczek and Souza (2012): qualidade vs. número de crianças no Brasil;

4 Questão 1

4.1 Questão A

4.2 Questão B

4.3 Questão C

4.4 Questão D

4.5 Questão E

4.6 Questão F

5 Questão 2

5.1 Questão A

Para avaliar fecundidade na cidade de São Paulo, podemos usar um modelo que permite formas funcionais, portanto, pode-se fazer uso de variáveis binárias, função log entre outros.

No nosso estudo a variável explicada se refere aos filhos, logo, não parece necessário usar logaritmo natural nessa variável.

As variáveis dummy disponibilizadas são: raça (1 para mulheres negras e 0 para brancas, amarelas, pardas); trabalha (1 para mulheres que trabalham e 0 para aquelas que não trabalham); domestico (1 para aquelas que fazem trabalho doméstico e 0 para aquelas que não fazem trabalho doméstico); educ1 (1 para aquelas que possuem entre 1 a 3 anos de educação e 0 para mais que 3 anos); educ2 (1 para aquelas que possuem entre 4 a 7 anos de educação e 0 para mais que 7 anos); educ3 (1 para aquelas que possuem entre 8 a 11 anos de educação e 0 para mais que 11 anos); educ4 (1 para aquelas que possuem de 12 ou mais anos de educação e 0 para menos que 12 anos); filho (1 se possui pelo menos um filho e 0 não possui nenhum filho).

A variável de comparação para as dummies de educação seria a de mulheres que não possuem nenhum ano de escolaridade, mas como era poucas observações, criou-se a variável categórica para Fundamental 1 incompleto, Fundamental 2 incompleto, Ensino

médio incompleto, Ensino médio completo e Ensino superior.

Estimou-se os três seguintes modelos para cada um de dois grupos: “Até ensino básico” (Fundamental 1 e 2 Incompleto, Ensino médio Incompleto e Completo) e “Mais que ensino básico” (Ensino Superior).

5.1.1 Modelo 1 - Relação de quantidade de filhos e educação

$$\text{Quantidade de filhos} = \beta_0 + \sum_{i=1}^{n=4} \beta_i \text{educ}_i + u$$

5.1.2 Modelo 2 - Relação de quantidade de filhos e salário por hora

$$\text{Quantidade de filhos} = \beta_0 + \beta_1 \log(\text{Salário por hora}) + u$$

5.1.3 Modelo 3 - Relação de quantidade de filhos, salário por hora linear e ao quadrado

$$\text{Quantidade de filhos} = \beta_0 + \beta_1 \log(\text{Salário por hora}) + \beta_2 \log(\text{Salário por hora})^2 + u$$

5.2 Questão B

As variáveis criadas foram de salário por hora (somente dividindo salário mensal por total de horas de trabalho mensal), uma variável categórica de educação -Fundamental 1 Incompleto; Fundamental 2 Incompleto; Ensino Médio Incompleto; Ensino Médio Completo e Ensino Superior-, e alteração de variáveis com erros.

Em relação a inclusão, usou-se com base na literatura a motivação de entender os seguintes mecanismos: Maior escolaridade implica que mulheres terão menos filhos - seja por condições biológicas ou seja por escolha? Maior renda tem a mesma relação que escolaridade?

Entendendo as relações destas variáveis que tangem ao tema de inserção das mulheres no mercado de trabalho e renda, podemos observar as heterogeneidades para raça, idade, desigualdade econômica e potencial dedicação ao trabalho.

5.3 Questão C

Para o modelo 1, temos que o coeficiente de impacto de grau de escolaridade diminui em $\hat{\beta}_i \forall i \in (1, 4)$ unidades de filhos, quando comparadas a uma mulher com maior grau

de escolaridade como fundamental 1 incompleto, as mulheres com mais anos completos de ensino têm menos filhos. Uma surpresa é a diferença do estimador de ensino médio completo para o estimador de ensino superior. Há um aumento desses graus de ensino, que não sabemos se é estatisticamente significativo entre si, afinal comparamos com o estimador de fundamental incompleto 1 e vemos significância de 1%.

Impacto da escolaridade na quantidade de filhos	
	Quantidade de Filhos
Intercepto	2.742*** (0.1053)
Fundamental 2 incompleto	-0.9257*** (0.1150)
Ensino médio incompleto	-1.701*** (0.1074)
Ensino médio completo	-2.048*** (0.1263)
Ensino superior	-1.959*** (0.1094)
N	6,314
R-quadrado	0.12497
R-quadrado ajustado	0.12441
Significância: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

Quando estimamos em regressões separadas para cada regressão, temos no segundo modelo os seguintes resultados:

Impacto linear do salário por hora na quantidade de filhos		
	Quantidade de Filhos	
	<Ens. Médio	≥ Ens. Médio
Intercepto	1.340*** (0.0866)	0.4822*** (0.1049)
Log Salário por hora	-0.0076 (0.0470)	0.1110** (0.0404)
N	4,434	1,459
R-quadrado	9.24e-6	0.00646
R-quadrado ajustado	-0.00022	0.00578
Significância: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		

5.4 Questão D

Temos diversos problemas para identificar uma causalidade completa na análise que requer melhor estrutura de dados para obter-se um desenho d modelo. Entretanto, dadas as ferramentas presentes, temos a possibilidade de separar e entender o efeito de salário e escolaridade separadamente.

5.5 Questão E

Como dito anteriormente, gostaríamos de estimar o efeito de rendimento, mas não temos dados de recebimentos para além do salário, e sabe-se que para pessoas nos quantis mais altos de distribuição de renda, parte relativamente importante é composta por ganhos de capital e não de trabalho. Provavelmente existe um viés quando observamos renda (mesmo que usando salário como um indicador possível).

Se temos um modelo original dado por:

$$\text{Quantidade de filhos} = \beta_0 + \beta_1 \log(\text{Salário por hora}) + \beta_2 \log(\text{Outros rendimentos}) + u$$

Se temos o modelo estimado como:

$$\text{Quantidade de filhos} = \beta_0 + \beta_1 \log(\text{Salário por hora}) + u_2$$

O estimador de $\hat{\beta}_1$ é dado por:

$$\hat{\beta}_1 =$$

5.6 Questão F

5.7 Questão G

Temos que as hipóteses clássicas são: - Lineariedade nos parâmetros; - Amostragem aleatória; - Inexistência de multicolinearidade perfeita (ou existência de variabilidade nas variáveis dependentes); - Exogeneidade estrita; - Homocedasticidade e inexistência de autocorrelação; - Normalidade dos erros.

Para as hipóteses 1 a 3 não temos problemas aparentes, afinal definimos os modelos como lineares nos parâmetros, temos uma amostra aleatória e não fazemos uso de variáveis que possam gerar multicolinearidade (como variáveis dummy com a mesma informação).

Para a hipótese 4, sobre exogeneidade precisamos garantir que tudo que está nos erros não é correlacionado com as variáveis x, portanto fazemos algumas análises do que poderia interferir no modelo.

Para a hipótese 5, fazemos a correção para heterocedasticidade e autocorrelação dos

erros no próprio estimador do software R, utilizamos o parâmetro “vcov” na função “feols” do pacote “fixest”.

Para a hipótese 6,

Questão H

6 Questão 3

6.1 Questão A

6.2 Questão B

6.3 Questão C