

# Econometria 1

Pedro Henrique Rocha Mendes \*

## Lista 3

(Texto para as questões 1 a 3) Os governos dos estados brasileiros estão interessados em implementar ações com o intuito de melhorar a taxa de aprovação (ou de não-repetência) dos alunos de escolas estaduais. Resultados de estudos econométricos previamente realizados apontam que a qualidade da escola pública em que foi cursado o ensino fundamental é uma das principais causas do desempenho do aluno no ensino médio. Seja assumido que é possível obter, a partir dos dados disponíveis, uma medida razoavelmente precisa para a qualidade das escolas públicas, denotada por “quali\_pub”. Também será assumido que os governos estaduais são capazes de alterar a variável em questão. Com base nisso, seja, por simplicidade, considerada uma política pública estadual que procura aumentar a taxa de aprovação explorando, para isso, a via de aumento da qualidade das escolas públicas de ensino fundamental. Você foi contratado para estimar o impacto provável desta política como subsídio à decisão que definirá se ela vai ou não ser implementada. Os dados disponíveis compreendem três variáveis, (i) o desempenho, no ensino médio, de alunos que cursaram o ensino fundamental em escolas públicas, taxa\_aprov, (ii) quali\_pub, (iii) nível educacional dos pais dos alunos, educ\_pais. A partir destes dados é possível estimar duas Funções de Regressão Populacional (FRPs) distintas.

- **FRP 1:**  $\text{taxa\_aprov}_i = \alpha_0 + \alpha_1 \text{quali\_pub}_i + e_i$
- **FRP 2:**  $\text{taxa\_aprov}_i = \beta_0 + \beta_1 \text{quali\_pub}_i + \beta_2 \text{educ\_pais}_i + u_i$

Ao comparar as duas FRPs, deve-se levar em conta a existência de correlação entre quali\_pub e educ\_pais, justificada pela conjectura de que a capacidade de selecionar a melhor escola para os filhos é diretamente proporcional ao nível educacional. Além disso, há evidência de que o desempenho no ensino médio é positivamente correlacionado com a educação dos pais.

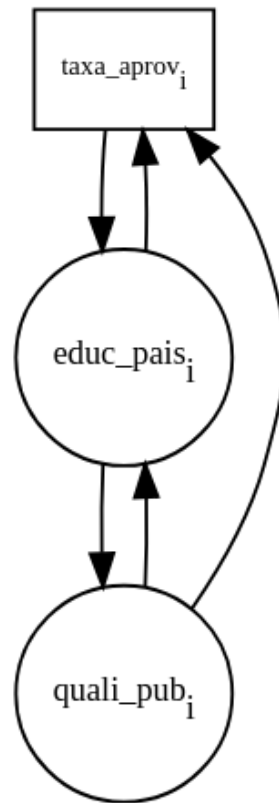
1)

*No enunciado acima, há três relações que conectam as três variáveis. Represente essas relações a partir de um diagrama de flechas tal como o que consta na nota de aula 4. Para isso, considere*

---

\*RA: 11201811516

duas convenções, quais sejam: (a) a causalidade deve ser indicada com uma flecha unidirecional,  $\rightarrow$ , em cujo início está a variável-causa e, em, cujo término, a variável-consequência e; (b) a correlação deve ser indicada com uma flecha bidirecional, “ $\leftrightarrow$ ”.



2)

Explique porque o coeficiente  $\beta_1$  é uma medida mais adequada do que o coeficiente  $\alpha_1$  para o impacto da política pública. Tome por base (i) o diagrama elaborado na questão anterior e (ii) a interpretação *ceteris paribus* da regressão múltipla.

Segundo Wooldridge<sup>1</sup>, sabendo que quali\_pub<sub>i</sub> é correlacionada com educ\_pais<sub>i</sub>, em uma mesma amostra, a relação entre os estimadores de  $\alpha_1$  e  $\beta_1$  (respectivamente,  $\hat{\alpha}_1$  e  $\hat{\beta}_1$ ) é

$$\hat{\alpha}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\delta}_1 \quad (1)$$

onde  $\hat{\delta}_1$  é o coeficiente de inclinação da regressão simples de educ\_pais<sub>i</sub> sobre quali\_pub<sub>i</sub>. Considerando  $\hat{\delta}_1$  depende apenas das variáveis independentes da amostra, podendo ser tratado como fixo, a esperança de  $\hat{\alpha}_1$  é

<sup>1</sup>WOOLDRIDGE, Jeffrey M., **Introductory econometrics: A modern approach**, [s.l.]: Cengage learning, 2015.

$$\begin{aligned}
E(\hat{\alpha}_1) &= E(\hat{\beta}_1 + \hat{\beta}_2 \hat{\delta}_1) \\
&= E(\hat{\beta}_1) + E(\hat{\beta}_2) \hat{\delta}_1 \\
&= \beta_1 + \beta_2 \hat{\delta}_1
\end{aligned} \tag{2}$$

Já o viés de  $\hat{\alpha}_1$ , definido como  $E(\hat{\alpha}_1) - \alpha_1$ , ao substituir  $E(\hat{\alpha}_1)$  pela equação (2)

$$\text{Viés}(\hat{\alpha}_1) = \beta_1 + \beta_2 \hat{\delta}_1 - \alpha_1 \tag{3}$$

Para que não houvesse viés entre  $\hat{\alpha}_1$  e  $\hat{\beta}_1$ ,  $\alpha_1$  deve ser igual a  $\beta_1$ . Porém,  $\alpha_1 = \beta_1 + \beta_2 \hat{\delta}_1$ , o que faz com que seu viés seja  $\beta_2 \hat{\delta}_1$ , justamente o coeficiente omitido na FRP 1. Logo, o coeficiente  $\beta_1$  é mais adequado que o coeficiente  $\alpha_1$  por representar melhor o efeito *ceteris paribus* de uma variação de  $\text{quali\_pub}_i$  em  $\text{taxa\_aprov}_i$ , pois isola melhor, sem a influência da variável omitida, tal efeito. Caso não haja correlação entre  $\text{quali\_pub}_i$  e  $\text{educ\_pais}_i$ , ambas as FRPs poderão ser utilizadas para estimar o efeito *ceteris paribus* da variação de  $\text{quali\_pub}_i$  em  $\text{taxa\_aprov}_i$ .

### 3)

*O que ocorreria caso os resultados da estimação da FRP 1 fossem tomados por base para decidir quanto à implementação da política?*

Caso  $\beta_2$  tenha valores relevantes, possivelmente a taxa de aprovação não terá o efeito estimado pela implementação da política pública ou, no mínimo, será menor do que o esperado.

### 4)

*Uma medida para a discriminação de gênero no mercado de trabalho é dada pela estimativa pontual para o parâmetro  $\delta$  na FRP a seguir, em que  $W_i$  é o salário recebido pelo  $i$ -ésimo indivíduo, “ $\text{educ}_i$ ” capta o nível educacional e  $X$  contém variáveis explicativas socioeconômicas adicionais. Os indivíduos que declararam possuir sexo masculino são indicados com  $d\_mas_i = 1$ , e indivíduos que declararam possuir sexo feminino como  $d\_mas_i = 0$ .*

$$\log(W_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_3 X_i + \delta d\_mas_i$$

*Argunte porque a estimativa pontual de  $\delta$  é uma medida para a porção do diferencial salarial associada à discriminação de gênero. Embase sua resposta no conceito do coeficiente de uma variável binária.*

Uma variável binária, ao captar se um atributo é ou não possuído, indica nesse caso se existe disparidade salarial entre gêneros. Caso a variável  $d\_mas_i$  seja relevante, ou seja, mostre a existência de um diferencial salarial, se observará na regressão uma diferença de nível entre as retas que representam o salário recebido por indivíduos do sexo masculino e feminino, mostrando

que, *ceteris paribus*, existe uma diferença salarial entre os dois gêneros.

5)

Um pesquisador estimou uma FRP em que a variável dependente é o tempo de internação hospitalar, devido a doenças respiratórias, de idosos com pelo menos 65 anos. Foram considerados 763 municípios da Amazônia Legal brasileira. A FPR estimada correspondeu à:  $\text{dias\_leito} = \beta_0 + \beta_1 \text{idade\_média}_i + \beta_2 \text{DCNT}_i + \beta_3 \text{PIB}_i + u_i$ , a qual será referida, doravante, como “equação original”. A primeira explicativa é a idade média da população municipal,  $\text{DNCT} \equiv$  mortalidade devido a doenças crônicas não-transmissíveis, uma medida do nível de saúde da população e  $\text{PIB} \equiv$  produto interno bruto municipal, uma medida para a capacidade de contratar serviços de saúde e comprar medicação. Uma vez obtidos os resultados, o pesquisador decidiu fazer um teste, ampliando a FRP com a inclusão do Índice de Desenvolvimento Humano municipal (IDH). Os resultados seguem na tabela abaixo. Explique porque, mesmo sendo a soma dos quadrados dos resíduos (SQR) inferior na equação ampliada, esta apresentou o mesmo  $R^2$  ajustado.

| Estatística / FRP         | Sem IDH      | Com IDH      |
|---------------------------|--------------|--------------|
| SQE                       | 3.047.142,10 | 3.047.833,81 |
| SQR                       | 393.843,77   | 393.152,06   |
| SQT                       | 3.440.985,87 | 3.440.985,87 |
| Graus de liberdade da SQR | 759          | 758          |
| Graus de liberdade da SQT | 762          | 762          |
| $R^2$ ajustado            | 0,8851       | 0,8851       |

Embora a soma de quadrados do resíduo tenha diminuído, mostrando uma menor diferença entre os valores obtidos na regressão e a média da variável independente, a soma de quadrados do erro, que representa a diferença entre os valores obtidos na regressão e os valores da variável dependente aumentou, levando a um  $R^2$  constante. Isso mostra que a variável IDH não é relevante para compreender o tempo de internação hospitalar, devido a doenças respiratórias, de idosos com pelo menos 65 anos.

6)

Quais dos seguintes itens podem fazer que os estimadores de MQO sejam enviesados? Explique sua escolha.

- a. Heterocedasticidade
- b. Omitir uma variável importante

Em uma regressão linear múltipla, a heterocedasticidade viola uma das hipóteses-base da obtenção da regressão, que é o erro possuir a mesma variância dado qualquer valor das variáveis

explicativas<sup>2</sup>. Já omitir uma variável importante, como demonstrado na questão dois, pode enviesar um estimador de MQO, já que impossibilita a obtenção do efeito *ceteris paribus* a partir do coeficiente estimado.

## Referências

WOOLDRIDGE, Jeffrey M. **Introductory econometrics: A modern approach**. [s.l.]: Cengage learning, 2015.

---

<sup>2</sup>*Ibid.*