

# ECONOMETRIA I - LISTA 3

## MODELO DE REGRESSÃO LINEAR MÚLTIPLA: ESTIMAÇÃO

Mateus Cardoso

22/06/2021

1) (Baseado em Wooldridge, Cap. 3) Quais são as hipóteses de Gauss-Markov para o modelo de regressão linear múltipla?

2) (Baseado em Wooldridge, Cap. 3) Sob quais hipóteses de Gauss-Markov os estimadores de MQO são não viesados?

3) (Baseado em Wooldridge, Cap. 3) Sob quais hipóteses de Gauss-Markov os estimadores de MQO são os melhores estimadores lineares não viesados?

4) (Wooldridge, Cap. 3) Quais dos seguintes itens podem fazer que os estimadores de MQO sejam viesados?

(i) Heteroscedasticidade.

(ii) Omitir uma variável importante.

(iii) Um coeficiente de correlação amostral de 0,95 entre duas variáveis independentes incluídas no modelo.

5) (Wooldridge, Cap. 3) O modelo seguinte é uma versão simplificada do modelo de regressão múltipla usado por Biddle e Hamermesh (1990) para estudar a escolha entre o tempo gasto dormindo e trabalhando e para observar outros fatores que afetam o sono:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u,$$

em que *sleep* e *totwrk* (trabalho total) são mensurados em minutos por semana e *educ* e *age* são mensurados em anos.

(i) Se os adultos escolhem entre dormir e trabalhar, qual é o sinal de  $\beta_1$ ?

- (ii) Que sinais você espera que  $\beta_2$  e  $\beta_3$ , terão?
- (iii) Usando os dados do arquivo SLEEP75, a equação estimada é

$$\widehat{sleep} = 3.638,25 - 0,148totwrk - 11,13educ + 2,20age$$

$$n = 706, \quad R^2 = 0,113.$$

Se alguém trabalha cinco horas a mais por semana, qual é a queda, em minutos, no valor esperado de *sleep*? Esse valor representa uma mudança grande?

- (iv) Discuta o sinal e a magnitude do coeficiente de *educ*.
- (v) Você diria que *totwrk*, *educ* e *age* explicam muito da variação de *sleep*? Quais outros fatores poderiam afetar o tempo gasto dormindo? É provável que sejam correlacionados com *totwrk*?

**6)** (Wooldridge, Cap. 3) Em um estudo que relaciona a nota média em curso superior (*GPA*) ao tempo gasto em várias atividades, você distribui uma pesquisa para vários estudantes. Os estudantes devem responder quantas horas eles despendem, em cada semana, em quatro atividades: estudo (*study*), sono (*sleep*), trabalho (*work*) e lazer (*leisure*). Toda atividade é colocada em uma das quatro categorias, de modo que, para cada estudante, a soma das horas nas quatro atividades deve ser igual a 168.

- (i) No modelo

$$colGPA = \beta_0 + \beta_1study + \beta_2sleep + \beta_3work + \beta_4leisure + u,$$

faz sentido manter *sleep*, *work* e *leisure* fixos, enquanto *study* varia?

- (ii) Explique a razão de esse modelo violar a Hipótese RLM.3.
- (iii) Como você poderia reformular o modelo de modo que seus parâmetros tivessem uma interpretação útil e ele satisfizesse a Hipótese RLM.3?

**7)** (Wooldridge, Cap. 3) Suponha que você tenha interesse em estimar o relacionamento *ceteris paribus* entre  $y$  e  $x_1$ . Para esse propósito você pode coletar dados de duas variáveis de controle,  $x_2$  e  $x_3$ . (Para melhor clareza, você pode entender  $y$  como uma nota do exame final,  $x_1$  como frequência às aulas,  $x_2$  como a nota de média graduação até o semestre anterior, e  $x_3$  como uma nota de teste de aptidão acadêmica ou de teste de avaliação.) Seja  $\tilde{\beta}_1$  a estimativa da regressão simples de  $y$  sobre  $x_1$ , e seja  $\hat{\beta}_1$  a estimativa de regressão múltipla de  $y$  sobre  $x_1$ ,  $x_2$  e  $x_3$ .

- (i) Se  $x_1$  for altamente correlacionada com  $x_2$  e  $x_3$  na amostra e  $x_2$  e  $x_3$  tiverem grandes efeitos parciais em  $y$ , você antecipa que  $\tilde{\beta}_1$  e  $\hat{\beta}_1$ , sejam semelhantes ou muito diferentes? Explique.

- (ii) Se  $x_1$ , for quase não correlacionada com  $x_2$  e  $x_3$ , mas  $x_2$  e  $x_3$  forem altamente correlacionadas,  $\tilde{\beta}_1$  e  $\hat{\beta}_1$  tenderão a ser semelhantes ou muito diferentes? Explique.
- (iii) Se  $x_1$  for altamente correlacionada com  $x_2$  e  $x_3$ , e  $x_2$  e  $x_3$  tiverem pequenos efeitos parciais em  $y$ , você anteciparia que  $\text{ep}(\tilde{\beta}_1)$ <sup>1</sup> ou  $\text{ep}(\hat{\beta}_1)$  será menor? Explique.
- (iv) Se  $x_1$ , for quase não correlacionada com  $x_2$  e  $x_3$ ,  $x_2$  e  $x_3$  tiver grandes efeitos parciais em  $y$ , e  $x_2$  e  $x_3$ , forem altamente correlacionadas, você anteciparia que  $\text{ep}(\tilde{\beta}_1)$  ou  $\text{ep}(\hat{\beta}_1)$  será menor? Explique.

## Exercícios no R

1) (Wooldridge, Cap. 3) O arquivo CEOSAL2 contém dados sobre 177 CEOs e pode ser usado para examinar os efeitos do desempenho da empresa sobre o salário destes executivos.

- (i) Estime um modelo que relacione o salário anual às vendas e ao valor de mercado da companhia. Faça o modelo do tipo de elasticidade constante para ambas as variáveis independentes. Escreva os resultados em forma de equação.
- (ii) Adicione *profits* ao modelo do item (i). Por que essa variável não pode ser incluída em forma logarítmica? Você poderia dizer que essas variáveis de desempenho da empresa explicam a maior parte da variação dos salários dos CEOs?
- (iii) Acrescente a variável *ceoten* ao modelo do item (ii). Qual é o retorno percentual estimado para outro ano de permanência do CEO, mantendo os outros valores fixos?
- (iv) Encontre o coeficiente de correlação amostral entre as variáveis  $\log(\text{mktval})$  e *profits*. Essas variáveis são altamente correlacionadas? O que isso diz sobre os estimadores de MQO?

2) (Wooldridge, Cap. 3) Um problema de interesse de agentes de saúde (e outros) é determinar os efeitos de fumar durante a gravidez sobre a saúde da criança. Uma medida da saúde infantil é o peso ao nascer; um peso muito baixo ao nascer pode colocar o recém-nascido em risco de contrair várias doenças. Como os fatores além do fumo que podem afetar o peso ao nascer são provavelmente relacionados com o ato de fumar, devemos levá-los em conta. Por exemplo, uma renda mais alta geralmente resulta em acesso a um melhor cuidado pré-natal, bem como uma melhor nutrição da mãe. Uma equação que reconhece isso é

$$\text{bwght} = \beta_0 + \beta_1 \text{cigs} + \beta_2 \text{faminc} + u$$

---

<sup>1</sup>ep = Erro Padrão

- (i) Qual é o sinal mais provável para  $\beta_2$ ?
- (ii) Você acha que *cigs* e *faminc* são possivelmente correlacionados? Explique por que a correlação deve ser positiva ou negativa.
- (iii) Agora, estime a equação com e sem *faminc*, usando os dados do arquivo BWGHT. Registre os resultados em forma de equação, incluindo o tamanho da amostra e o R-quadrado. Discuta seus resultados, verificando se adicionar *faminc* altera substancialmente o efeito estimado de *cigs* sobre *bwght*.

## Referências

WOOLDRIDGE, Jeffrey. **Introdução à Econometria:** Uma Abordagem Moderna. 3. ed. São Paulo: Cengage, 2019.