**Econometria II
Exercícios para revisão e autoteste**
“Introdução à Econometria”, Jefrey M. Wooldridge
**PROBLEMAS DE ESPECIFICAÇÃO E DE DADOS**

Obs.: os exercícios que indicam ‘arquivos’ para serem resolvidos são do livro do Wooldridge. Os arquivos necessários estão na pasta “Banco de dados Wooldridge”, na área “Programação em R”.

**1.** Seja *educ\** o grau efetivo de escolaridade, medido em anos (que pode ser um número não inteiro), e seja *educ* o número de anos mais elevado de educação formal. Você acha que *educ* e *educ\** são relacionados pelo modelo clássico de erro nas variáveis?

**2.** Suponha que estejamos interessados nos efeitos dos gastos com campanhas eleitorais feitos por candidatos sobre o apoio dos eleitores. Alguns candidatos decidem não apresentar sua candidatura à reeleição. Se somente pudermos coletar os resultados da votação e dos gastos dos candidatos que efetivamente participaram da eleição, existe possibilidade de ocorrer uma seleção endógena da amostra?

**3.** O R-quadrado da estimativa do modelo

log(salário) = β0 + β1log(vendas) + β2log(valmerc) β3lucrmarg + β4perceo + β5percomp + u,

usando os dados contidos no arquivo CEOSAL2.RAW, era R²=0,353 (n=177). Quando *perceo²* e *percomp²* são adicionados, R²=0,375. Existe evidência de má-especificação da forma funcional neste modelo?

**4.** Seja mate10 a porcentagem de aprovação em um teste-padrão de matemática. Estamos interessados em estimar o efeito do gasto por estudante no desempenho em matemática. Um modelo simples é

mate10= β0 + β1log(gasto) + β3log(matricl) + β4pobreza + u,

em que *pobreza* é a percentagem de estudantes vivendo em condições de pobreza.
a) A variável *prgalm* é a porcentagem de estudantes qulificados para o programa de merenda escolar financiado pelo governo federal. Por que ela é uma variável proxy razoável de *pobreza*?
b) A tabela seguinte contém estimativas MQO, com e sem prgalm como uma variável explicativa (variável dependente: mate10)

|  |  |  |
| --- | --- | --- |
| **Variáveis independentes** | **(1)** | **(2)** |
| log(gasto) | 11,12(3,30) | 7,75(3,04) |
| log(matricl) | 0,022(0,615) | -1,26(0,58) |
| prgalm | \_ | -324(0,036) |
| intercepto | -69,24(26,72) | -23,14(24,90) |
| ObservaçõesR-quadrado | 4280,0297 | 4280,1893 |

Explique por que o efeito dos gastos sobre mate10 é menor na coluna (2) do que na coluna (1). O efeito na coluna (2) ainda é estatisticamente maior que zero?

c) Parece que as taxas de aprovação são menores em escolas maiores, com os outros fatores sendo iguais? Explique.
d) Interprete o coeficiente de prgalm na coluna (2).
e) O que você deduz do substancial aumento de R² da coluna (1) para a (2)?

**5.** A equação seguinte explica o número de horas por semana que uma criança passa assistindo televisão, em termos da idade da criança, educação da mãe, educação do pai e número de irmãos:

tvhoras\*= β0 + β1idade + β2idade² + β3educm + β4educp + β5irms + u;

Estamos preocupados com a possibilidade de que tvhoras\* tenha sido medida com erro em nossa pesquisa. Seja tvhoras o número de horas por semana que se gasta assistindo televisão.
a) O que as hipóteses do erro clássico nas variáveis (CEV) requerem nessa aplicação?
b) Você acha que as hipóteses CEV tem possibilidades de se manter? Explique.

**6.** Estima-se um modelo relacionando número de crimes no campus às matriculas de estudantes em um grupo de faculdades. A amostra que usamos não é uma amostra aleatória de faculdades dos Estados Unidos, pois muitas escolas em 1992 não registraram crimes no campus. Você acha que a falnha das faculdades em informar os crimes pode ser vista como uma seleção amostral exógena? Explique.

**7.** No modelo yi= ai + bixi , prove que os MQO estimam consistentemente α e β se ai for não correlacionada com xi e bi for não correlacionada com xi e xi², que são hipóteses mais fracas que E(ai|xi) = E(ai) e E(bi|xi) = E(bi).

**8.** Considere o modelo de regressão simples com erro de medição clássico, y= β0 + β1x\* + u em que temos m medidas na x\*. escreva como zh= x\* + eh, h =1,...,m. Suponha que x\* é não correlacionada com u, e1,...,em=, que os erros de medição são não correlacionados em pares, e têm a mesma variância, σe². Que w=(z1+...+zm)/m seja a média das medidas na x\*, de forma que, de cada observação i, wi=(zi1+...+zim)/m será a média das medidas de m. Que $\overbar{β}$1 seja o estimador MQO da regressão simples yi sobre 1,wi,i=1,...,n, usando uma amostra aleatória de dados.
a) Prove que plim($\overbar{β}$1) = β1$\frac{σ²\_{x\*}}{σ²\_{x\*}+(σ²\_{e}/m) }$.
b) Como a inconsistência na $\overbar{β}$1 se compara com isso quando somente um único indicador está disponível (isto é, m=1)? O que acontece à medida que m cresce? Comente.

**SOLUÇÕES**

1. Não. Se *educ\** for um inteiro — o que significa que alguém não tem nenhuma escolaridade após à série anterior concluída —, o erro de medida será zero. Se *educ\** não for um inteiro, *educ < educ\** e, assim, o erro de medida será negativo. No mínimo, *e1* não pode ter média zero, e *e1* e *educ\** provavelmente são correlacionados.

2. A decisão de um candidato de não concorrer pode estar sistematicamente relacionada com o que ele, ou ela, espera fazer na eleição. Portanto, podemos apenas ter uma amostra de candidatos que, na média, sejam mais fortes que todos os demais possíveis candidatos que poderiam concorrer. Isso resulta em um problema de seleção de amostra se a população de interesse incluir todos os candidatos. Se estivermos apenas interessados nos efeitos dos gastos de campanha sobre os resultados eleitorais dos concorrentes que estão tentando a reeleição, não haverá problema de seleção de amostra.