**Econometria II
Exercícios para revisão e autoteste**
“Introdução à Econometria”, Jefrey M. Wooldridge
**DADOS EM PAINEL**

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.** Meyer, Viscusi e Durbin (1995) estudaram a extenção do tempo (em semanas), em que um trabalhador acidentado recebe remuneração por conta de indenização trabalhista. Em 1980 houve uma na politica do estado de Kentucky, que aumentou o limite dos ganhos semanais que eram cobertos por essa remuneração. Usando amostras aleatórias, eles iniciaram com uma análise de diferença em diferenças, usando log(*duração*) como a variável dependente. Façamos *apmud* representar uma variável dummy das observações após a mudança da politica e *altrend*, a variável dos trabalhadores de altos rendimentos. Utilizando os dados contidos no arquivo INJURY.RAW, a equação estimada, com os erros padrão entre parênteses, é:

 $\hat{log⁡(duração)}$ = *1,126 + 0,0077apmud + 0,256altrend + 0,191apmud.altrend.* (0,031) (0,0447) (0,047) (0,069)
n = 5.626, R²=0,021.

O que você concluiria do coeficiente e da estatística t de *altrend*?

**2.** Suponha que em uma equação de poupança familiar, dos anos de 1990, 1991 e 1992, definimos kidsit como o número de crianças na família *i* no ano *t*. Se o número de crianças for constante ao longo desse período de três anos na maioria das famílias na amostra, que problemas isso pode causar na estimativa do efeito que o numero de crianças tem sobre a poupança?

**3.** Geronimus e Korenman (1992) usaram pares de irmãs para estudar os efeitos da gradidez na adolescência sobre as consequências econômicas futuras. Quando o resultado é renda em relação as necessidades – algo que depende do número de filhos – , o modelo é

 *log(rendanecfs) = β0 + δ0irmã2 + β1partoadfs + β2idadefs + outros fatores + af + ufs*;

em que f indexa as famílias e s indexa uma irmã dentro da família. A variável de interesse é *partoadfs*, que é uma variável binária igual a um se a irmã *s* na família *f* teve um filho na adolescência, *idadesf* é a idade atual da irmã *s* na família *f*, e *af* é um efeito familiar não observado.
Ao usar o método da diferença, faz sentido a inclusão de variáveis dummy para a etnia da mãe e do pai na equação?

**4.** Utilizando os dados contidos no arquivo KIWLMC.RAW, as seguintes equações foram estimadas usando os anos de 1978 e 1981:

$\hat{log⁡(preço)}$ *= 11,49 - 0,547proxincin + 0,394 a81.proxincin* (0,26) (0,058) (0,080)
n=321, R²=0,220

e

$\hat{log⁡(preço)}$ *= 11,18 - 0,563 a81- 0,403 a81.proxincin* (0,27) (0,044) (0,067)
n=321, R²=0,337

Compara as estimativas do termo de interação de *a81.proxincin* com as da seguinte equação:

$\hat{log⁡(preço)}$ *= 11,29 + 0,457 a81 – 0,340proxincin - 0,063 a81.proxincin* (0,31) (0,045) (0,055) (0,083)
n=321, R²=0,409

Por que as estimativas são diferentes?

**5.** Por que não podemos usar as primeiras diferenças quando temos cortes transversais independentes em dois anos (ao contrário dos dados em painel)?

**6.** Se pensarmos que β1 é positivo na seguinte equação
 txcrimit= *β0 + δ0d87t + β1desempit + ai + uit* (modelo simples de efeitos não observados da taxa de criminalidade de uma cidade em 1982 e 1987, sendo *d87* uma variável dummy para o ano de 1987);
e que Δui e Δdesempi são negativamente correlacionados, qual será o viés no estimador MQO de β1 na equação de primeiras diferenças?

**7.** Suponha que queremos estimar o efeito de diversas variáveis sobre a poupança anual e que temos um conjunto de dados em painel sobre indivíduos coletado em 31 de janeiro de 1990 e 31 de janeiro de 1992. Se incluirmos uma dummy anual para o ano de 1992 e usarmos a primeira diferença, podemos também incluir a idade no modelo original? Explique.

**8.** Em 1985, nem a Flórida nem a Geórgia tinham leis banindo recipientes abertos de bebidas alcoólicas nos compartimento de veículos de passageiros. Em 1990, a Flórida sancionou tal lei, mas a Geórgia não.
a) Suponha que você colete amostras aleatórias da população com idade para dirigir de ambos os estados, de 1985 e 1990. Defina *prisão* como uma variável binária igual à unidade se uma pessoa foi presa por dirigir embriagada durante o ano. Sem controlar quaisquer outros fatores, escreva um modelo de probabilidade linear que possibilite verificar se a lei de recipientes abertos reduziu a probabilidade de alguém ser preso por dirigir embriagado. Que coeficiente em seu modelo mede o efeito da lei?
b) Por que você pode querer controlar outros fatores nesse modelo? Quais poderiam ser esses fatores?

c) Agora, suponha que somente seja possível coletar dados de 1985 e 1990 em nível de municípios dos dois estados. A variável dependente seria a fração dos motoristas habilitados presos por dirigirem embriagados durante o ano. Como essa estrutura de dados difere dos dados em nível individual descritos na parte a)? Que método econométrico você utilizaria?

**9.** Com uma única variável explicativa, a equação usada para obter o estimador intragrupo é
 $ \overbar{y}$i= β0 + βi$\overbar{x}$i + *a*i + $\overbar{u}$i ,
em que a barra superior representa a média ao longo do tempo. Podemos presumir que E(*a*i)=0 por termos incluído um intercepto na equação. Suponha que $\overbar{u}$i seja não correlacionado com $\overbar{x}$i , mas que Cov(xit ,ai­) = σxa para todo t( e i, em razão da amostragem aleatória no corte transversal).
a) Definindo $\tilde{β}$1 como o estimador entre as observações, isto é, o estimador MQO usando as médias temporais, mostre que plim$\tilde{β}$1 = β1 + σxa/Var($\overbar{x}$i), em que limite de probabilidade é definido como N -> ∞.
b) Suponha também que xit, para todo T= 1,2,...,T, seja não correlacionado com a variância constante σx². Mostre que plim$\tilde{β}$1 = β1 + T(σxa/ σx²).
c) Se as variáveis explicativas não forem altamente correlacionadas ao longo do tempo, o que a parte b) sugere quanto à possibilidade de a incossistência no estimador entre as observações ser menor quando existem mais períodos de tempo?

**10.** Em um modelo de efeitos aleatórios, defina o erro composto *vit= ai + uit*, em que *ai* é não correlacionado com *uit* e os *uit* tem variância constante *σu²* e são serialmente não correlacionados. Defina *eit = vit - ʎ* $\overbar{v}$*i*, em que *ʎ* é dado em *ʎ = 1 – [σu²/( σu² + T σa²)]1/2*
a) Mostre que *E(eit)=0.*
b) Mostre que *Var(eit) = σu², t=1,...,T*.
c) Mostre que para *t ≠ s, Cov(eit,eis) = 0*.

**11.** Para determinar os efeitos do desempenho atlético universitário dos candidatos, você coleta dados das inscrições dos candidatos de uma amostra das faculdades da Divisão I dos anos de 1985, 1990 e 1995.
a) Que indicadores de êxito atlético você incluiria em uma equação? Quais seriam alguns dos problemas de cronometragem?
b) Que outros fatores você controlaria na equação?
c) Escreva uma equação que possibilite estimar os efeitos do êxito atlético sobre a mudança percentual nas inscrições. Como você estimaria essa equação? Por que você escolheria esse método?

**12.** Suponha que, para um semestre, você possa coletar os seguintes dados em uma amostra aleatória de calouros e veteranos universitários de cada disciplina: nota padronizada de um exame final, percentagem de frequência as aulas, uma variável dummy indicando se a matéria se enquadra na especialidade do aluno, nota média acumulada antes do inicio do semestre e nota do exame de ingresso (SAT) no curso superior.
a) Por que você classificaria esse conjunto de dados como uma amostra por agrupamento? Quantas observações, aproximadamente, você esperaria para um aluno típico?
b) Se você agrupar todos os dados e usar MQO, o que você está presumindo sobre as características não observadas dos alunos que afetam as taxas de desempenho e de frequência? Que papel, com relação a isso, desempenham a nota de ingresso no curso superior (SAT) e a nota média acumulada antes do inicio do semestre (GPA)?
c) Se você julga que a nota de ingresso no curso superior e a nota média acumulada antes do inicio do semestre não indicam adequadamente a capacidade dos alunos, como você estimaria o efeito da frequência sobre o desempenho no exame final?

**SOLUÇÕES**

1. O coeficiente de *altrend* mostra que, na ausência de qualquer alteração no limite dos ganhos, os trabalhadores com renda mais elevada passam muito mais tempo — cerca de 29,2%, em média — recebendo remuneração por conta de indenização trabalhista.

2. Quer usemos a primeira diferenciação ou a transformação interna, teremos problema com a estimativa do coeficiente de kidsit. Por exemplo, usando a transformação interna, se kidsit não variar para a família i, então, $\ddot{kids}$it = kidsit - $\overbar{kids}$i = 0 para t = 1,2,3. Desde que algumas famílias apresentem variação em kidsit, então, poderemos computar o estimador de efeitos fixos, mas o coeficiente de kids poderá ser estimado de modo muito impreciso. Essa é uma forma de multicolinearidade na estimação de efeitos fixos (ou na estimação de primeira diferenciação).

3. Não, se todas as irmãs de uma mesma família tiverem a mesma mãe e o mesmo pai. Então, como as variáveis referentes à raça dos pais não variarão por irmã, elas seriam eliminadas pela diferenciação.