**Econometria I
Exercícios para revisão e autoteste**
“Introdução à Econometria”, Jefrey M. Wooldridge (1 ao 10)
“Econometria”, Stock e Watson (11)
**VARIÁVEIS BINÁRIAS**

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.** Suponha que, em um estudo que compara resultados de eleições entre candidatos democratas e republicanos, você queira indicar o partido de cada candidato. Um nome como partido será uma boa escolha para uma variável binária nesse caso? Qual seria um nome melhor?

**2.** Utilizando os dados contidos no arquivo SLEEP75.RAW, obtemos a equação estimada

$\hat{dormir}$= *3.840,83 -0,163trabtot -11,71educ -8,70idade +0,128idade² +87,75masculino* (235,11) (0,018) (5,86) (11,21) (0,134) (34,33)
n=706, R²=0,123, $\overbar{R}$²=0,117.

A variável *dormir* é o total de minutos gastos por semana dormindo a noite, *trabtot* é o total de minutos semanais gastos trabalhando, *educ* e *idade* são medidas em anos e masculino é uma variável *dummy* de gênero.
a) Supondo todos os outros fatores iguais, existe evidencia de que os homens durmam mais que as mulheres? O quanto essa evidência é forte?
b) Existe uma relação de substituição estatisticamente significante entre trabalhar e dormir? Qual é a relação de substituição estimada?
c) Que outras regressões você precisa executar para testar a hipótese nula de que, mantendo fixos os outros fatores, a idade não tem efeito sobre dormir?

**3.** As seguintes equações foram estimadas utilizando os dados contidos no arquivo BWGHT.RAW:
$\hat{log⁡(pesonas)}$= *4,66 – 0,0044cigs +0,0093log(rendfam\_) + 0,016ordnas +0,027masculino* (0,22) (0,0009) (0,0059) (0,006) (0,010)  *+0,055branco*(0,013)
n=1.388 ; R²=0,0472

e
$\hat{log⁡(pesonas)}$= *4,65 – 0,0052cigs +0,0110log(rendfam) + 0,017ordnas +0,034masculino* (0,38) (0,0010) (0,0085) (0,006) (0,011)  *+0,045branco -0,0030educmae +0,0032educpai*(0,015) (0,0030) (0,0026)
n=1.191 ; R²=0,0493

As variáveis são definidas, mas adicionamos uma variável *dummy* para o caso de a criança ser do sexo masculino e uma variável *dummy* que indica se a criança é classificada como branca.

a) Na primeira equação, interprete o coeficiente da variável *cigs*. Particularmente, qual é o efeito no peso dos recém-nascidos se a mãe fumar dez ou mais cigarros por dia?
b) Quanto se espera que um recém nascido branco pesará mais que uma criança não branca, mantendo fixos todos os outros fatores na primeira equação? A diferença é estatisticamente significante?
c) Comente sobre o efeito estimado e a significância estatística de *educmae*.
d) Com a informação dada, por que você não terá condições de calcular a estatística F da significância conjunta de *educmae* e *educpai*? O que você teria que fazer para calcular a estatística F?

**4.** Utilizando os dados contidos no arquivo GPA2.RAW, a seguinte equação foi estimada:

$\hat{SAT}$= 1.028,10 +19,30tamclas -2,19tamclas² -45,09feminino -169,81negro
 (6,29) (3,83) (0,53) (4,29) (12,71)
*+62,31feminino.negro*
 (18,15)
n=4.137, R²=0,0858

A variável *SAT* é a pontuação combinada de matemática e habilidade verbal do estudante para ingresso em curso superior (SAT); *tamclas* é o tamanho da classe do aluno no ensino médio, em centenas; *feminino* é uma variável *dummy* para gênero; *negro* é uma variável *dummy* da raça igual a um para negros e zero, caso contrário.
a) Existe forte evidência que *tamclass*² deveria ser incluída no modelo? Desta equação, qual é o tamanho ótimo da classe no ensino médio?
b) Mantendo fixo *tamclass*, qual é a diferença estimada na nota SAT entre mulheres não negras e homens não negros? O quanto é estatisticamente significante essa diferença estimada?
c) Qual é a diferença estimada na SAT entre homens não negros e homens negros? Teste a hipótese nula de que não há diferença entre suas notas, contra a hipótese de que existe uma diferença.
d) Qual é a diferença estimada na nota SAT entre mulheres negras e mulheres não negras? O que você necessitaria fazer para verificar se a diferença é estatisticamente significante?

**5.** Uma equação que explica os salários dos diretores executivos é

$\hat{log⁡(salário)}$*= 4,59 +0,257log(vendas) +0,011roe +0,158financeira +0,181prodcons* (0,30) (0,032) (0,004) (0,089) (0,085) *-0,283servpub*(0,099)
n=209, R²=0,357.

Os dados usados estão no arquivo CEOSAL1.RAW, no qual *financeira, prodcons e servpub* são variáveis binárias que indicam as empresas financeiras, de produtos de consumo e de serviços públicos. O ramo de atividade omitido foi o de transportes.
a) Calcule a diferença percentual aproximada no salário estimado entre os setores de serviços públicos e de transportes, mantendo fixos *vendas* e *roe*. A diferença é estatisticamente significante ao nível de 1%?
b) Use a equação *100.[exp(*$\hat{β}$*1)-1]* para obter a diferença percentual exata n salário estimado entre os setores de serviços públicos e de transportes e compare-a com a resposta obtida na parte a).
c) Qual é a diferença percentual aproximada no salário estimado entre as atividades de produto de consumo e financeiros? Escreva uma equação que possibilite verificar se a diferença é estatisticamente significante.

**6.** Nos dados sobre salários dos jogadores de beisebol encontrados no arquivo MLB1.RAW, os jogadores ocupam uma de seis posições: *pribase, segbase, terbase, interbase, jardext* ou *receptor*. Para possibilitar diferenças salariais entre as posições, com determinados defensores (*jardext*) como o grupo base, quais variáveis *dummy* você incluiria como variáveis independentes?

**7.** Para testar a eficiência de um programa de treinamento de pessoal sobre os subsequentes salários dos trabalhadores, especificamos o modelo

 *log(salárioh) = β0 + β1trein + β2educ + β3exper + u*em que *trein* é uma variável binária igual à unidade se um trabalhador participou do programa. Pense no termo de erro u como contendo a aptidão não observada do trabalhador. Se trabalhadores menos aptos tiverem maior oportunidade de serem selecionados para o programa, o que você pode dizer sobre o provável viés no estimador MQO de *β1*?

**8.** Suponha que você colete dados de uma pesquisa sobre salários, educação, experiência e gênero. Além disso, você solicita informações sobre o uso de maconha. A pergunta original é: “em quantas ocasiões distintas, no mês passado, você fumou maconha?”.
a) Escreva uma equação que permita a você estimar os efeitos do uso de maconha sobre os salários com todos os outros fatores controlados. Você deve ter condições de fazer declarações do tipo “Estima-se que fumar maconha cinco vezes ou mais por mês altera os salários em x%”.
b) Escreva um modelo que permita verificar se o uso de drogas tem efeitos diferentes sobre os salários dos homens e das mulheres. Como você verificaria que não existem diferenças nos efeitos do uso de drogas nos homens e nas mulheres?
c) Suponha que você considere ser melhor avaliar o uso de maconha colocando as pessoas em uma de quatro categorias: não usuário, usuário leve ( um a cinco vezes por mês), usuário moderado (seis a dez vezes por mês), e usuário inveterado (mais de dez vezes por mês). Agora escreva um modelo que permita estimar os efeitos da maconha sobre os salários.
d) Usando o modelo da parte c), explique em detalhes como testar a hipótese nula de que o uso de maconha não tem efeito sobre o salário. Seja bastante especifico e inclua uma relação cuidadoso de graus de liberdade.
e) Quais são alguns dos problemas potenciais de procurar inferência causal utilizando os dados da pesquisa que você coletou?

**9.** Que *d* seja uma variável *dummy* (binária) e que *z* seja uma variável quantitativa. Considere o modelo

 *y = β0 + δ0d + β1z +δ1d.z + u* ;

esta é uma versão geral de um modelo com uma interação entre uma variável *dummy* e uma quantitativa.
a) Como isto não alterará nada importante, defina o erro com valor zero, *u = 0*. Então, quando *d = 0*, podemos escrever o relacionamento entre y e z como a função *f0(z)= β0 + β1z*. Escreva a mesma relação quando *d=1*, em que você deve usar *f1(z)* no lado esquerdo para denotar a função linear de *z.*b) Considerando *δ1*≠ 0 (o que significa que as duas não são paralelas), demonstre que o valor de z\* de tal forma que *f0(z\*)= f1(z\*)* seja z\*=- *δ0*/ *δ1* . Este é o ponto no qual as duas linhas se cruzam. Demonstre que z\* será positivo se, e somente se, *δ0* e *δ1* tiverem sinais opostos.

**10.** Para uma criança i morando em determinada região de ensino, defina voucheri como uma variável dummy igual a um se a criança for selecionada para participar de um programa de bolsas de estudos em uma escola, e defina scorei como a nota da criança em um exame padronizado subsequente. Suponha que a variável de participação, voucheri, seja completamente aleatorizada para que ela seja independente tanto dos fatores observados quanto dos não observados que possam afetar a nota do teste de avaliação.
a) Se você executar uma regressão simples de scorei sobre voucheri usando uma amostra aleatória de tamanho n, o estimador MQO produzirá um estimador não viesado do efeito do programa de bolsas de estudos?
b) Suponha que você possa coletar informações adicionais de perfis familiares tais como renda familiar, estrutura familiar (por exemplo, se a criança mora com os dois pais), e nível de escolaridade dos pais. Você precisará controlar esses fatores para obter um estimador não viesado dos efeitos do programa de bolsas de estudos? Explique.
c) Por que você precisará incluir as variáveis de perfis familiares na regressão? Existe uma situação em que você não incluiria as variáveis de perfis familiares?

**11.** Qual é o efeito da raça sobre a probabilidade de recusa de uma hipoteca, mantendo constante a razão prestação-renda? Para quantificar esse efeito, estimamos uma regressão probit tendo *razão P/R* e *negro* como regressores.

$\hat{P(recusa=1|razão\frac{P}{R}, negro)}$*=* φ*(-2,26 +2,74 razão P/R + 0,71negro)* (0,16) (0,44) (0,083)
Utilize esse modelo para responder às seguintes questões:
a) Um requerente da hipoteca negro possui uma razãp P/R de 0,35. Qual é a probabilidade de que seu pedido seja rejeitado?
b) Suponha que o requerente reduziu sua razão para 0,30. Que efeito isso teria sobre sua probabilidade de ter uma hipoteca recusada?
c) Repita as partes a) e b) para um requerente branco.
d) O efeito marginal da razão P/R sobre a probabilidade de recusa da hipoteca depende da raça? Explique.

**SOLUÇÕES**

1. Não, pois não ficaria claro quando *partido* seria um ou zero. Um nome melhor seria algo como *Dem*, que seria um para os candidatos Democratas, e zero, para os Republicanos. Ou *Rep*, que seria um para os Republicanos, e zero, para os Democratas.

6. Com *jardtext* como o grupo base, incluiríamos as variáveis *dummy pribase, segbase, terbase, interbase* e *receptor*.

11. a) φ(-0,59)= 27,76%
 b) φ(-0,73)=23,27%