

PROBLEMAS

7.1 Utilizando os dados contidos no arquivo SLEEP75.RAW (veja também o Problema 3.3), obtenha a equação estimada

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

A variável *dormir* é o total de minutos gastos por semana dormindo durante 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.

- Supondo todos os outros fatores iguais, existe evidência de que os homens durmam mais que as mulheres? O quanto essa evidência é forte?
- Existe uma relação de substituição estatisticamente significativa entre trabalhar e dormir? Qual é a relação de substituição estimada?
- 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?

7.2 As seguintes equações foram estimadas utilizando os dados contidos no arquivo BWGHT.RAW:

$$\widehat{\log(\text{pesonas})} = 4,66 - 0,0044 \text{ cigs} + 0,0093 \log(\text{rendfam}) + 0,016 \text{ ordnas} \\ (0,22) \quad (0,0009) \quad (0,0059) \quad (0,006) \\ + 0,027 \text{ masculino} + 0,055 \text{ branco} \\ (0,010) \quad (0,013) \\ n = 1.388, R^2 = 0,0472$$

e

$$\widehat{\log(\text{pesonas})} = 4,65 - 0,0052 \text{ cigs} + 0,0110 \log(\text{rendfam}) + 0,017 \text{ ordnas} \\ (0,38) \quad (0,0010) \quad (0,0085) \quad (0,006) \\ + 0,034 \text{ masculino} + 0,045 \text{ branco} - 0,0030 \text{ educmae} + 0,0032 \text{ educpai} \\ (0,011) \quad (0,015) \quad (0,0030) \quad (0,0026) \\ n = 1.191, R^2 = 0,0493.$$

As variáveis são definidas como no Exemplo 4.9, 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.

- 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?
- 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 significativa?
- Comente sobre o efeito estimado e a significância estatística de *educmae*.
- Com a informação dada, por que você não terá condições de calcular a estatística *F* da significância conjunta de *educm* e *educpai*? O que você teria que fazer para calcular a estatística *F*?

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

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

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.

- Existe forte evidência que *tamclas*² deveria ser incluída no modelo? Desta equação, qual é o tamanho ótimo da classe no ensino médio?
- Mantendo fixo *tamclas*, qual é a diferença estimada na nota *SAT* entre mulheres não negras e homens não negros? O quanto é estatisticamente significativa essa diferença estimada?
- 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 alternativa de que existe uma diferença.
- 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 significativa?

7.4 Uma equação que explica os salários dos diretores executivos é

$$\widehat{\log(\text{salário})} = 4,59 + 0,257 \log(\text{vendas}) + 0,011 \text{ roe} + 0,158 \text{ financeira} \\ (0,30) \quad (0,032) \quad (0,004) \quad (0,089) \\ + 0,181 \text{ prodcons} - 0,283 \text{ servpub} \\ (0,085) \quad (0,099) \\ n = 209, R^2 = 0,357.$$

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.

- Calcule a diferença percentual aproximada no salário estimado entre os setores de serviços públicos e de transportes, mantendo fixos *ventas* e *roe*. A diferença é estatisticamente significativa ao nível de 1%?
- Use a equação (7.10) para obter a diferença percentual exata no salário estimado entre os setores de serviços públicos e de transportes e compare-a com a resposta obtida na parte (i).
- Qual é a diferença percentual aproximada no salário estimado entre as atividades de produtos de consumo e financeiros? Escreva uma equação que possibilite verificar se a diferença é estatisticamente significativa.

7.5 No Exemplo 7.2 defina *semPC* como uma variável *dummy* igual a um se o aluno não possuir um PC, e zero caso contrário.

- Se *semPC* for usada no lugar de *PC* na equação (7.6), o que acontece com o intercepto na equação estimada? Qual será o coeficiente de *semPC*? (Sugestão: Escreva $PC = 1 - semPC$ e agregue isso na equação $supGPA = \hat{\beta}_0 + \hat{\delta}_0 PC + \hat{\beta}_1 emGPA + \hat{\beta}_2 ACT$.)
- O que acontecerá com o *R*-quadrado se *semPC* for usado em lugar de *PC*?
- As variáveis *PC* e *semPC* deveriam ser incluídas como variáveis independentes no modelo? Explique.

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

$$\log(\text{salário}_i) = \beta_0 + \beta_1 \text{trein} + \beta_2 \text{educ} + \beta_3 \text{exper} + u_i$$

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, e você usar uma análise MQO, o que você pode dizer sobre o provável viés no estimador MQO de β_1 ? (Sugestão: Consulte o Capítulo 3.)

7.7 No exemplo na equação (7.29), suponha que definamos *foraft* como sendo um se a mulher estiver fora da força de trabalho, e zero caso contrário.

- Se fizermos a regressão de *foraft* sobre todas as variáveis independentes na equação (7.29), o que acontecerá com as estimativas do intercepto e da inclinação? (Sugestão: $naft = 1 - foraft$. Agregue essa expressão na equação populacional $naft = \beta_0 + \beta_1 nesprend + \beta_2 educ + \dots$ e reorganize.)
- O que acontecerá com os erros-padrão das estimativas do intercepto e da inclinação?
- O que acontecerá com o *R*-quadrado?

7.8 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?"

- 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*%".
- 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?

- 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.
- Usando o modelo da parte (iii), explique em detalhes como testar a hipótese nula de que o uso de maconha não tem efeito sobre o salário. Seja bastante específico e inclua uma relação cuidadosa de graus de liberdade.
- Quais são alguns dos problemas potenciais de procurar inferência causal utilizando os dados da pesquisa que você coletou?

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

$$y = \beta_0 + \delta_0 d + \beta_1 z + \delta_1 d \cdot z + u$$

esta é uma versão geral de um modelo com uma interação entre uma variável *dummy* e uma quantitativa. [Um exemplo está na equação (7.17).]

- 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 $f_0(z) = \beta_0 + \beta_1 z$. Escreva a mesma relação quando $d = 1$, em que você deve usar $f_1(z)$ no lado esquerdo para denotar a função linear de *z*.
- Considerando $\delta_1 \neq 0$ (o que significa que as duas não são paralelas), demonstre que o valor de z^* de tal forma que $f_0(z^*) = f_1(z^*)$ seja $z^* = -\delta_0/\delta_1$. Este é o ponto no qual as duas linhas se cruzam [como na Figura 7.2(b)]. Demonstre que z^* será positivo se, e somente se, δ_0 e δ_1 tiverem sinais opostos.
- Usando os dados contidos no arquivo TWOYEAR.RAW, a seguinte equação poderá ser estimada:

$$\widehat{\log(\text{salário})} = 2,289 - 0,357 \text{feminino} + 0,50 \text{totgrad} + 0,030 \text{feminino} \cdot \text{totgrad}$$

(0,011) (0,015) (0,003) (0,005)

$$n = 6,763, R^2 = 0,202$$

em que todos os coeficientes e erros-padrão foram arredondados com três casas decimais. Usando esta equação, encontre o valor de *totgrad* de tal forma que os valores preditos de $\log(\text{salário})$ sejam os mesmos tanto para homens como para mulheres.

- Com base na equação na parte (iii), as mulheres podem, de forma realística, obter educação superior de modo que seus ganhos alcancem os dos homens? Explique.

7.10 Para uma criança *i* morando em determinada região de ensino, defina *voucheri* como uma variável *dummy* igual a um se a criança foi 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.

- Se você executar uma regressão simples de *scorei* sobre *voucheri* usando uma amostra aleatória de tamanho *n*, o estimador de MQO produzirá um estimador não viesado do efeito do programa de bolsas de estudos?

- (ii) 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.
- (iii) 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?