Como já discutimos, presumindo que as não observáveis são, na média, não relacionadas com as variáveis explicativas, é vital para se derivar a primeira propriedade estatística de cada estimador MQO: sua ausência de viés no parâmetro populacional correspondente. É claro, todas as hipóteses anteriores são usadas para demonstrar a ausência de viés.

Hipótese RLM.5 (Homoscedasticidade)

O erro u tem a mesma variância dados quaisquer valores das variáveis explicativas, em outras palavras:

$$\mathrm{Var}(u|x_1, x_2, \ldots, x_k) = \sigma^2.$$

Comparada com a Hipótese RLM.4, a hipótese de homoscedasticidade é de importância secundária; particularmente, a Hipótese RLM.5 não tem influência na ausência de viés das $\hat{\beta}_j$. Ainda assim, a homoscedasticidade tem duas implicações importantes: (1) Podemos derivar fórmulas das variâncias amostrais cujos componentes são fáceis de serem caracterizados; (2) Podemos concluir, sob as hipóteses RLM.1 até a RLM.5 de Gauss-Markov, que os estimadores MQO têm a menor variância entre *todos* os estimadores lineares, não viesados.

Hipótese RLM.6 (Normalidade)

O erro populacional u é *independente* das variáveis explicativas $x_1, x_2, \ldots, x_k$ e é normalmente distribuído com média zero e variância σ^2: $u \sim \mathrm{Normal}(0, \sigma^2)$.

Neste capítulo, adicionamos a Hipótese RLM.6 à distribuição amostral exata das estatísticas t e estatísticas F, de forma que possamos realizar testes de hipóteses. No próximo capítulo, veremos que a RLM.6 pode ser eliminada se tivermos uma amostra de tamanho razoavelmente grande. A Hipótese RLM.6 realmente implica uma propriedade de eficiência mais forte dos MQO: os estimadores MQO têm a menor variância entre *todos* os estimadores não viesados; o grupo de comparação não estará mais restrito a estimadores lineares na $\{y_i: i = 1, 2, \ldots, n\}$.

PROBLEMAS

4.1 Considere uma equação para explicar os salários dos dos diretores executivos (CEOs) em termos das vendas anuais das empresas (*vendas*), dos retornos sobre o patrimônio líquido (*roe*, em forma percentual) e do retorno sobre o capital das empresas (*ros*, na forma percentual):

$$\log(salário) = \beta_0 + \beta_1 \log(vendas) + \beta_2 roe + \beta_3 ros + u.$$

(i) Em termos dos parâmetros do modelo, formule a hipótese nula em que, após controlar *vendas* e *roe*, *ros* não tem efeito sobre o salário dos CEOs. Formule a hipótese alternativa de que um melhor desempenho de mercado das ações aumenta o salário dos diretores executivos.

(ii) Usando os dados em CEOSAL1.RAW, obteve-se a seguinte equação por MQO:

$$\widehat{\log(salário)} = \underset{(0,32)}{4,32} + \underset{(0,035)}{0,280} \log(vendas) + \underset{(0,0041)}{0,0174}\, roe + \underset{(0,00054)}{0,00024}\, ros$$
$$n = 209,\ R^2 = 0,283.$$

Se *ros* aumenta em 50 pontos, qual é a variação percentual prevista em *salário*? Na prática, *raf* tem um efeito grande sobre *salário*?

(iii) Teste a hipótese nula de que *ros* não tem efeito sobre *salário* contra a alternativa de que *ros* tem um efeito positivo. Faça o teste ao nível de significância de 10%.

(iv) Você incluiria *ros* no modelo final que explica a remuneração dos CEOs em termos do desempenho das empresas? Explique.

4.2 Quais dos seguintes itens podem fazer com que as estatísticas t de MQO não sejam válidas (isto é, que elas não tenham distribuições t sob H_0)?

(i) Heteroscedasticidade.

(ii) Um coeficiente de correlação de 0,95 entre duas variáveis independentes que estão no modelo.

(iii) Omitir uma variável explicativa importante.

4.3 No Exemplo 4.7, usamos dados de empresas manufatureiras não sindicalizadas para estimar a relação entre a taxa de rejeição e outras características da firma. Agora, vamos olhar esse exemplo mais de perto e usar todas as empresas disponíveis.

(i) O modelo populacional estimado no Exemplo 4.7 pode ser escrito como

$$\log(rejei) = \beta_0 + \beta_1 hrsemp + \beta_2 \log(vendas) + \beta_3 \log(empreg) + u.$$

Usando as 43 observações disponíveis para 1987, a equação estimada é

$$\widehat{\log(rejei)} = \underset{(4,57)}{11,74} - \underset{(0,019)}{0,042}\, hrsemp - \underset{(0,370)}{0,951} \log(vendas) + \underset{(0,360)}{0,992} \log(empreg)$$
$$n = 43,\ R^2 = 0,310.$$

Compare essa equação com aquela estimada com somente 29 firmas não sindicalizadas na amostra.

(ii) Mostre que o modelo populacional pode também ser escrito como

$$\log(rejei) = \beta_0 + \beta_1 \log(hrsemp) + \beta_2 \log(vendas/empreg) + \theta_3 \log(empreg) + u.$$

em que $\theta_3 = \beta_2 + \beta_3$. [*Aviso*: Lembre-se de que $\log(x_2/x_3) = \log(x_2) - \log(x_3)$.] Interprete a hipótese H_0: $\theta_3 = 0$.

(iii) Quando a equação da parte (ii) é estimada, obtemos

$$\widehat{\log(rejei)} = \underset{(4,57)}{11,74} - \underset{(0,019)}{0,042}\, hrsemp - \underset{(0,370)}{0,951} \log(vendas/empreg) + \underset{(0,205)}{0,041} \log(empreg)$$
$$n = 43,\ R^2 = 0,310.$$

Controlando o treinamento dos trabalhadores e a razão vendas-empregados, as empresas maiores têm taxas de rejeição maiores estatisticamente significantes?

(iv) Teste a hipótese de que um aumento de 1% em *vendas/empreg* está associado a uma queda de 1% na taxa de rejeição.

4.4 A tabela seguinte foi criada ao usar os dados do arquivo CEOSAL2. RAW:

Variável dependente: log(*salário*)			
Variáveis independentes	**(1)**	**(2)**	**(3)**
log(*vendas*)	0,224 (0,027)	0,158 (0,040)	0,188 (0,040)
log(*valmerc*)	—	0,112 (0,050)	0,100 (0,049)
lucrmarg	—	−0,0023 (0,0022)	−0,0022 (0,0021)
permceo	—	—	0,0171 (0,0055)
percomp	—	—	−0,0092 (0,0033)
intercepto	4,94 (0,20)	4,62 (0,25)	4,57 (0,25)
Observações R-quadrado	177 0,281	177 0,304	177 0,353

A variável *valmerc* é o valor de mercado da firma, *lucrmarg* é o lucro como percentagem das vendas, *permceo* corresponde aos anos de trabalho como diretor executivo na atual empresa, e *percomp* é o total de anos na empresa.

(i) Comente os efeitos de *lucrmarg* sobre o salário dos diretores executivos.
(ii) O valor de mercado tem um efeito significante? Explique.
(iii) Interprete os coeficientes de *permceo* e *percomp*. As variáveis são estatisticamente significantes?
(iv) O que você entende do fato de que a permanência muito longa na empresa, mantendo fixos os outros fatores, está associada a salários mais baixos?

4.5 Na Seção 4.5 usamos, como exemplo, o teste da racionalidade da avaliação dos preços de casas. Lá, usamos um modelo log-log em *preço* e *aval* [veja equação (4.47)]. Aqui, vamos usar uma formulação nível-nível.

(i) No modelo de regressão simples

$$preço = \beta_0 + \beta_1 aval + u,$$

a avaliação é racional se $\beta_1 = 1$ e $\beta_0 = 0$. A equação estimada é

$$\widehat{preço} = -14{,}47 + 0{,}976\, aval$$
$$(16{,}27)\quad (0{,}049)$$
$$n = 88,\ \text{SQR} = 165.644{,}51,\ R^2 = 0{,}820.$$

Primeiro, teste a hipótese H_0: $\beta_0 = 0$ contra a hipótese alternativa bilateral. Em seguida, teste H_0: $\beta_1 = 1$ contra a hipótese alternativa bilateral. O que você conclui?

(ii) Para testar a hipótese conjunta $\beta_0 = 0$ e $\beta_1 = 1$, precisamos do SQR do modelo restrito. Isso é igual a calcular $\sum_{i=1}^{n}(preço_i - aval_i)^2$, em que $n = 88$, visto que os resíduos do modelo restrito são exatamente $preço_i - aval_i$. (Nenhuma estimação é necessária para o modelo restrito porque ambos os parâmetros estão especificados sob H_0.) Isso tem como resultado SQR = 209.448,99. Faça o teste F para a hipótese conjunta.

(iii) Agora, teste H_0: $\beta_2 = 0$, $\beta_3 = 0$ e $\beta_4 = 0$ no modelo
$preço = \beta_0 + \beta_1 aval + \beta_2 dimterr + \beta_3 arconstr + \beta_4 qtdorm + u.$

O R-quadrado da estimação desse modelo usando as mesmas 88 residências é 0,829.

(iv) Se a variância de *preço* varia com *aval*, *tamterr*, *arquad* ou *qtdorm*, o que você pode dizer sobre o teste F da parte (iii)?

4.6 A análise de regressão pode ser usada para testar se o mercado usa eficientemente as informações ao avaliar ações. Seja *retorno* o retorno total de possuir ações de uma firma ao longo de um período de quatro anos, do final de 1990 até o final de 1994. A *hipótese de mercados eficientes* diz que esses retornos não devem estar sistematicamente relacionados à informação conhecida em 1990. Se as características conhecidas da empresa no início do período ajudassem a prever os retornos das ações, poderíamos usar essas informações para escolher ações.

Para 1990, seja *dkr* a relação dívida-capital de uma empresa, seja *eps* os ganhos por ação, seja *rendliq* a renda líquida e seja *salário* a remuneração total dos CEOs da empresa.

(i) Usando os dados do arquivo RETURN.RAW, estimou-se a seguinte equação:

$$\widehat{retorno} = -14{,}37 + 0{,}321\, rdc + 0{,}043\, gpa - 0{,}0051\, rendliq + 0{,}0035\, salário$$
$$(6{,}89)\quad (0{,}201)\quad (0{,}078)\quad (0{,}0047)\quad (0{,}0022)$$
$$n = 142,\ R^2 = 0{,}0395.$$

Teste se as variáveis explicativas são conjuntamente significantes ao nível de 5%? Alguma variável explicativa é individualmente significante?

(ii) Agora, nova estimação do modelo usando a forma log para *rendliq* e *salário* forneceu a seguinte equação:

$$\widehat{retorno} = -36{,}30 + 0{,}327\, dkr + 0{,}069\, eps - 4{,}74 \log(rendliq) + 7{,}24 \log(salário)$$
$$(39{,}37)\quad (0{,}203)\quad (0{,}080)\quad (3{,}39)\quad (6{,}31)$$
$$n = 142,\ R^2 = 0{,}0330.$$

Alguma de suas conclusões da parte (i) mudou?

(iii) Nesta amostra, algumas firmas têm zero de dívidas e outras têm ganhos negativos. Deveríamos tentar usar log(*dkr*) ou log(*eps*) para vermos se isso melhorará o ajustamento? Explique
(iv) Em geral, a evidência da previsibilidade dos retornos é forte ou fraca?

4.7 Considere a equação estimada do Exemplo 4.3 que poderia também ser usada para estudar os efeitos de faltar às aulas sobre a GPA em curso superior:

$$\widehat{gradGPA} = 1{,}39 + 0{,}412\, emGPA + 0{,}15\, ACT - 0{,}083\, faltas$$
$$(0{,}33)\quad (0{,}94)\qquad (0{,}011)\qquad (0{,}026)$$
$$n = 141,\ R^2 = 0{,}234.$$

(i) Usando a aproximação normal padronizada, encontre o intervalo de confiança de 95% para β_{emGPA}.
(ii) Você pode rejeitar a hipótese H_0: $\beta_{emGPA} = 0{,}4$ contra a hipótese alternativa bilateral no nível de 5%?
(iii) Você pode rejeitar a hipótese H_0: $\beta_{emGPA} = 1$ contra a hipótese alternativa bilateral no nível de 5%?

4.8 No Problema 3.4, estimamos a equação

$$\widehat{dormir} = 3.638{,}25 - 0{,}148\, trabtot - 11{,}13\, educ + 2{,}20\, idade$$
$$(112{,}28)\quad (0{,}017)\qquad (5{,}88)\qquad (1{,}45)$$
$$n = 706,\ R^2 = 0{,}113,$$

para a qual informamos, agora, os erros-padrão juntamente com as estimativas.

(i) *educ* ou *idade* são individualmente significantes ao nível de 5% contra uma hipótese alternativa bilateral? Mostre como você chegou à resposta.
(ii) Ao retirar *educ* e *idade* da equação, temos

$$\widehat{dormir} = 3.586{,}38 - 0{,}151\, trabtot$$
$$(38{,}91)\quad (0{,}017)$$
$$n = 706,\ R^2 = 0{,}103,$$

É possível afirmar que *educ* e *idade* são conjuntamente significantes na equação original ao nível de 5%? Justifique sua resposta.
(iii) Incluir *educ* e *idade* no modelo afeta muito a relação estimada entre dormir e trabalhar?
(iv) Suponha que a equação de dormir contenha heteroscedasticidade. O que isso significa para os testes calculados nas partes (i) e (ii)?

4.9 As taxas de aluguel são influenciadas pela população de estudantes em uma cidade onde há universidades? Seja *alug* o aluguel médio mensal pago pela unidade alugada em uma determinada cidade, onde há universidades. Seja *pop* o total da população da cidade, *rendmed*, a renda média da cidade e *pctestu*, a população de estudantes como um percentual da população total. Um modelo para testar uma relação é

$$\log(alug) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(rendmed) + \beta_3 pctestu + u.$$

(i) Formule a hipótese nula de que o tamanho da população estudantil relativo à população das cidades não tem efeito *ceteris paribus* sobre os aluguéis mensais. Formule a alternativa de que há um efeito.
(ii) Quais sinais você espera para β_1 e β_2?
(iii) A equação estimada, usando dados de 1990 de 64 cidades com universidades do arquivo RENTAL.RAW, é

$$\widehat{\log(alug)} = 0{,}043 + 0{,}066 \log(pop) + 0{,}507 \log(rendmed) + 0{,}0056\, pctestu$$
$$(0{,}844)\quad (0{,}039)\qquad (0{,}081)\qquad (0{,}0017)$$
$$n = 64,\ R^2 = 0{,}458.$$

O que está errado com a seguinte afirmação: "Um aumento de 10% na população está associado a um aumento de cerca de 6,6% no aluguel"?
(iv) Teste a hipótese formulada na parte (i) no nível de 1%.

4.10 Considere o modelo de regressão múltipla com três variáveis independentes, sob as hipóteses do modelo linear clássico RLM.1 a RLM.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

Você deseja testar a hipótese nula H_0: $\beta_1 - 3\beta_2 = 1$.

(i) Sejam $\hat{\beta}_1$ e $\hat{\beta}_2$ os estimadores de MQO de β_1 e β_2. Encontre $\text{Var}(\hat{\beta}_1 - 3\hat{\beta}_2)$ em termos das variâncias de $\hat{\beta}_1$ e $\hat{\beta}_2$ e a covariância entre eles. Qual é o erro-padrão de $\hat{\beta}_1 - 3\hat{\beta}_2$?
(ii) Escreva a estatística t para testar H_0: $\beta_1 - 3\beta_2 = 1$.
(iii) Defina $\theta_1 = \beta_1 - 3\beta_2$ e $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$. Escreva uma equação de regressão que envolva β_0, θ_1, β_2 e β_3, que permita que você obtenha diretamente $\hat{\theta}_1$ e seu erro-padrão.

4.11 A variável *pdintens* corresponde a gastos com pesquisa e desenvolvimento (P&D) como uma percentagem das vendas. As vendas são mensuradas em milhões de dólares. A variável *lucrmarg* corresponde a lucros como uma percentagem das vendas.
Usando os dados do arquivo RDCHEM.RAW de 32 empresas da indústria química, estimou-se a seguinte equação:

$$\widehat{pdintens} = 0{,}472 + 0{,}321 \log(vendas) + 0{,}050\, lucrmarg$$
$$(1{,}369)\quad (0{,}216)\qquad (0{,}046)$$
$$n = 32,\ R^2 = 0{,}099.$$

(i) Interprete o coeficiente de log(*vendas*). Em particular, se *vendas* aumenta em 10%, qual é a variação percentual estimada em *pdintens*? Esse efeito é economicamente grande?
(ii) Teste a hipótese de que a intensidade de P&D não varia com *vendas* contra a alternativa de que P&D aumenta com as vendas. Teste nos níveis de 5% e 10%.
(iii) Interprete o coeficiente na *lucrmarg*. Ele é economicamente grande?
(iv) *lucrmarg* tem um efeito estatisticamente significante sobre *pdintens*?