

Análise de Regressão com Dados de Corte Transversal

parte 1 do texto aborda a análise de regressão com dados de corte transversal. Ela se apóia na álgebra estudada nos cursos superiores e em conceitos básicos de probabilidade e estatística.

Os Apêndices A, B e C contêm revisões completas sobre esses tópicos.

O Capítulo 2 tem início com o modelo de regressão linear simples, no qual explicamos uma variável em termos de outra. Embora a regressão simples não seja amplamente usada em econometria aplicada, ela é utilizada ocasionalmente e serve como um ponto de partida natural, pois sua álgebra e suas interpretações são relativamente simples.

Os capítulos 3 e 4 cobrem os fundamentos da análise de regressão múltipla, em que permitimos que mais variáveis afetem a variável que estamos tentando explicar. A regressão múltipla é ainda o método mais geralmente usado na pesquisa empírica, de modo que esses capítulos merecem atenção cuidadosa. O capítulo 3 enfatiza a álgebra do método de mínimos quadrados ordinários (MQO), estabelecendo ainda as condições necessárias para que os estimadores MQO sejam não-viesados e também os melhores estimadores lineares não-viesados. O Capítulo 4 trata do importante tópico da inferência estatística.

O Capítulo 5 discute as propriedades referentes às amostras grandes, ou assintóticas, dos estimadores MQO. Essa discussão oferece a justificativa para os procedimentos de inferência do Capítulo 4 quando os erros em um modelo de regressão não são normalmente distribuídos. O capítulo 6 cobre alguns tópicos adicionais da análise de regressão, incluindo questões avançadas sobre forma funcional, transformação dos dados, previsão e grau de ajuste da estimação. O Capítulo 7 explica como a informação qualitativa pode ser incorporada em modelos de regressão múltipla.

O Capítulo 8 ilustra como testar e corrigir o problema da heteroscedasticidade, ou variância não-constante, no termo erro. Mostramos como as estatísticas MQO usuais podem ser ajustadas e também apresentamos uma extensão do método MQO, conhecida como método dos *mínimos quadrados ponderados*, que explica diretamente as diferentes variâncias dos erros. O Capítulo 9 explora o importante problema da correlação entre o termo erro e uma ou mais das variáveis explicativas. Demonstramos como a utilização de uma variável *proxy* pode resolver o problema de variáveis omitidas. Adicionalmente, determinamos o viés e a inconsistência dos estimadores MQO na presença de certos tipos de erros de medida nas variáveis. Diversos problemas de tratamento dos dados são também discutidos, incluindo o problema dos *outliers*.

O Modelo de Regressão Simples

O modelo de regressão simples pode ser usado para estudar a relação entre duas variáveis. Por razões que veremos adiante, o modelo de regressão simples tem limitações enquanto ferramenta geral para a análise empírica. No entanto, às vezes ele é apropriado como ferramenta empírica. Aprender como interpretar o modelo de regressão simples é uma boa prática para estudar a regressão múltipla, o que faremos nos capítulos subsequentes.

2.1 DEFINIÇÃO DO MODELO DE REGRESSÃO SIMPLES

Grande parte da análise econométrica começa com a seguinte premissa: y e x são duas variáveis, representando alguma população, e estamos interessados em “explicar y em termos de x ”, ou em “estudar como y varia com variações em x ”. Discutimos alguns exemplos no Capítulo 1, incluindo: y é a produção de soja, e x , a quantidade de fertilizantes; y é o salário-hora, e x , anos de educação; e y é uma taxa de criminalidade em uma comunidade, e x , o número de policiais.

Ao escrever um modelo que “explicará y em termos de x ”, defrontamo-nos com três questões. Primeira, como nunca há uma relação exata entre duas variáveis, como consideramos outros fatores que afetam y ? Segunda, qual é a relação funcional entre y e x ? E terceira, como podemos estar certos de que estamos capturando uma relação *ceteris paribus* entre y e x (se esse for um objetivo desejado)?

Podemos resolver essas ambigüidades escrevendo uma equação que relaciona y a x . Uma equação simples é

$$y = \beta_0 + \beta_1 x + u. \quad (2.1)$$

A equação (2.1), que supostamente é válida para a população de interesse, define o **modelo de regressão linear simples**. Ela também é chamada *modelo de regressão linear de duas variáveis* ou *modelo de regressão linear bivariada*, pois relaciona as duas variáveis x e y . Vamos discutir, agora, o significado de cada uma das quantidades em (2.1). (A propósito, o termo “regressão” tem origens que não são especialmente importantes para muitas das aplicações econométricas modernas, de modo que não o explicaremos aqui. Veja Stigler (1986) para uma história interessante da análise de regressão.)

Quando relacionadas por (2.1), as variáveis y e x têm vários nomes diferentes, os quais são intercambiáveis, como explicado em seguida. y é chamada a **variável dependente**, a **variável explicada**, a **variável de resposta**, a **variável prevista**, ou o **regressando**. x é chamada a **variável independente**, a **variável explicativa**, a **variável de controle**, a **variável previsora**, ou o **regressor**. (O termo

covariável também é usado para x .) Os termos “variável dependente” e “variável independente” são usados com frequência em econometria. Mas esteja consciente de que o nome “independente” não se refere aqui à noção estatística de independência entre variáveis aleatórias (veja Apêndice B, disponível na página do livro, no site www.thomsonlearning.com.br).

Os termos variáveis “explicada” e “explicativa” são, provavelmente, os mais descritivos. “Resposta” e “controle” são muito usados nas ciências experimentais, em que a variável x está sob o controle do pesquisador. Não usaremos os termos “variável prevista” e “previsora”, embora algumas vezes você os veja no texto. Nossa terminologia para a regressão simples está resumida na Tabela 2.1.

Tabela 2.1

Terminologia para a Regressão Simples

y	x
Variável Dependente	Variável Independente
Variável Explicada	Variável Explicativa
Variável de Resposta	Variável de Controle
Variável Prevista	Variável Previsora
Regressando	Regressor

A variável u , chamada de **termo erro** ou **perturbação** da relação, representa outros fatores, além de x , que afetam y . Uma análise de regressão simples trata, efetivamente, todos os fatores, além de x , que afetam y como não-observados. Você pode pensar em u , convenientemente, como representando o “não-observado”.

A equação (2.1) também trata da questão da relação funcional entre y e x . Se os outros fatores em u são mantidos fixos, de modo que a variação em u é zero, $\Delta u = 0$, então x tem um efeito linear sobre y :

$$\Delta y = \beta_1 \Delta x \text{ se } \Delta u = 0. \quad (2.2)$$

Assim, a variação em y é, simplesmente, β_1 multiplicado pela variação em x . Isso significa que β_1 é o **parâmetro de inclinação** da relação entre y e x , mantendo fixos os outros fatores em u ; ele é de interesse fundamental em economia aplicada. O **parâmetro de intercepto** β_0 também tem seus usos, embora ele raramente seja central para uma análise.

EXEMPLO 2.1

(Produção de Soja e Fertilizantes)

Suponha que a produção de soja seja determinada pelo modelo

$$\text{produção} = \beta_0 + \beta_1 \text{fertilizante} + u, \quad (2.3)$$

EXEMPLO 2.1 (continuação)

de modo que $y = \text{produção}$ e $x = \text{fertilizantes}$. O pesquisador agrícola está interessado no efeito dos fertilizantes sobre a produção, mantendo outros fatores fixos. Esse efeito é dado por β_1 . O termo erro u contém fatores como qualidade da terra, chuva etc. O coeficiente β_1 mede o efeito dos fertilizantes sobre a produção, mantendo outros fatores fixos: $\Delta \text{produção} = \beta_1 \Delta \text{fertilizante}$.

EXEMPLO 2.2**(Uma Equação Simples do Salário)**

Um modelo que relaciona o salário de uma pessoa à educação observada e outros fatores não-observados é

$$\text{saláριο}_h = \beta_0 + \beta_1 \text{educ} + u. \quad (2.4)$$

Se saláριο_h é medido em dólares por hora e educ corresponde a anos de educação formal, β_1 mede a variação no salário-hora dado um ano a mais de educação, mantendo todos os outros fatores fixos. Alguns desses fatores incluem experiência da força de trabalho, aptidão inata, permanência com o empregador atual, ética no trabalho e inumeráveis outras coisas.

A linearidade de (2.1) implica que uma variação de uma unidade em x tem o mesmo efeito sobre y , independentemente do valor inicial de x . Isso é irrealista para muitas aplicações econômicas. Por exemplo, no salário-educação, poderíamos querer considerar retornos *crecentes*: o próximo ano de educação teria, em relação ao anterior, um efeito *maior* sobre os salários. Veremos como considerar tais possibilidades na Seção 2.4.

A questão mais difícil é saber se o modelo (2.1) realmente nos permite tirar conclusões *ceteris paribus* sobre como x afeta y . Acabamos de ver, na equação (2.2), que β_1 mede, *de fato*, o efeito de x sobre y , mantendo todos os outros fatores (em u) fixos. Encerra-se com isso a questão da causalidade? Infelizmente, não. Como podemos esperar aprender algo, em geral, sobre o efeito *ceteris paribus* de x sobre y , mantendo outros fatores fixos, quando estamos ignorando todos aqueles outros fatores?

A Seção 2.5 mostrará que somos capazes de obter estimadores confiáveis de β_0 e β_1 de uma amostra aleatória de dados somente quando fazemos uma hipótese que restrinja a maneira de como o termo não-observável u está relacionado à variável explicativa x . Sem tal restrição, não seremos capazes de estimar o efeito *ceteris paribus*, β_1 . Como u e x são variáveis aleatórias, precisamos de um conceito baseado em probabilidade.

Antes de expormos a hipótese crucial de como x e u são relacionados, podemos sempre fazer uma hipótese sobre u . Se o intercepto β_0 está incluído na equação, nada se perde ao assumir que o valor médio de u na população é zero.

Matematicamente,

$$E(u) = 0. \quad (2.5)$$

A hipótese (2.5) não diz nada sobre a relação entre u e x ; ela simplesmente faz uma afirmação sobre a distribuição dos fatores não-observáveis na população. Usando os exemplos anteriores como ilustração, podemos ver que a hipótese (2.5) não é muito restritiva. No Exemplo 2.1, não perdemos nada ao normalizar os fatores não-observáveis que afetam a produção de soja, tal como a qualidade da terra, para ter uma média zero na população de todos os lotes cultivados. O mesmo é verdadeiro para os fatores não-observáveis do Exemplo 2.2. Sem perda de generalidade, podemos assumir que coisas como a média da aptidão são zero na população de todas as pessoas que trabalham. Se você não está convencido, pode trabalhar com o Problema 2.2 para ver que podemos sempre redefinir o intercepto na equação (2.1) para tornar (2.5) verdadeiro.

Agora, vamos voltar à hipótese crucial concernente à u e x como são relacionados. Uma medida natural de associação entre duas variáveis aleatórias é o *coeficiente de correlação*. (Veja Apêndice B, disponível no site da Thomson, para definição e propriedades.) Se u e x são *não-correlacionados*, logo, enquanto variáveis aleatórias, não são *linearmente* relacionados. Assumir que u e x são não-relacionados requer um caminho longo para definir o sentido em que u e x deveriam ser não-correlacionados na equação (2.1). Mas isso não vai longe o suficiente, pois a correlação mede somente a dependência linear entre u e x . A correlação tem uma característica algo contra-intuitiva: é possível que u seja não-correlacionado com x e seja correlacionado com funções de x , tal como x^2 . (Veja Seção B.4 para uma discussão adicional.) Essa possibilidade não é aceitável para muitos propósitos da regressão, visto que causa problemas para interpretar o modelo e para derivar propriedades estatísticas. Uma hipótese melhor envolve o *valor esperado de u , dado x* .

Como u e x são variáveis aleatórias, podemos definir a distribuição condicional de u , dado qualquer valor de x . Em particular, para qualquer x , podemos obter o valor esperado (ou médio) de u para aquela fatia da população descrita pelo valor de x . A hipótese crucial é que o valor médio de u *não* depende do valor de x . Podemos escrever isso como

$$E(u|x) = E(u) = 0, \quad (2.6)$$

em que a segunda igualdade resulta de (2.5). A primeira igualdade na equação (2.6) é a hipótese nova. Ela diz que, para qualquer valor de x , a média dos fatores não-observáveis é a mesma e, portanto, deve igualar-se ao valor médio de u na população. Quando combinamos a primeira igualdade da equação (2.6) com a hipótese (2.5), obtemos a **hipótese de média condicional zero**.

Vamos ver o que (2.6) acarreta ao exemplo do salário. Para simplificar a discussão, assumamos que u seja o mesmo que aptidão inata. Então, (2.6) requer que o nível médio de aptidão seja o mesmo, independentemente dos anos de educação formal. Por exemplo, se $E(\text{aptidão}|8)$ representa a aptidão média para o grupo de todas as pessoas com oito anos de educação formal, e $E(\text{aptidão}|16)$ representa a aptidão média entre pessoas na população com 16 anos de educação formal, portanto (2.6) implica que essas médias devem ser as mesmas. De fato, o nível de aptidão média deve ser o mesmo para *todos* os níveis de educação. Se, por exemplo, entendemos que a aptidão média aumenta com os anos de educação formal, então (2.6) é falsa. (Isso aconteceria se, em média, pessoas com maior aptidão escolhessem tornar-se mais educadas.) Como não podemos observar aptidão inata, não temos um modo de saber se a aptidão média é ou não a mesma para todos os níveis de educação. Essa é uma questão que devemos resolver antes de aplicar a análise de regressão simples.

No exemplo dos fertilizantes, se as quantidades de fertilizantes são escolhidas independentemente de outras características dos lotes, então (2.6) se sustentará: a qualidade média da terra não dependerá da quantidade de fertilizantes. Entretanto, se mais fertilizantes forem usados em lotes de terra de melhor qualidade, então o valor esperado de u varia com o nível de fertilizantes, e (2.6) não se sustenta.

Suponha que a nota de um exame final (*nota*) dependa da frequência às aulas (*freq*) e de fatores não-observados que afetam o desempenho dos estudantes (tal como a aptidão). Então:

$$nota = \beta_0 + \beta_1 freq + u. \quad (2.7)$$

Em que situação você esperaria que esse modelo satisfaça (2.6)?

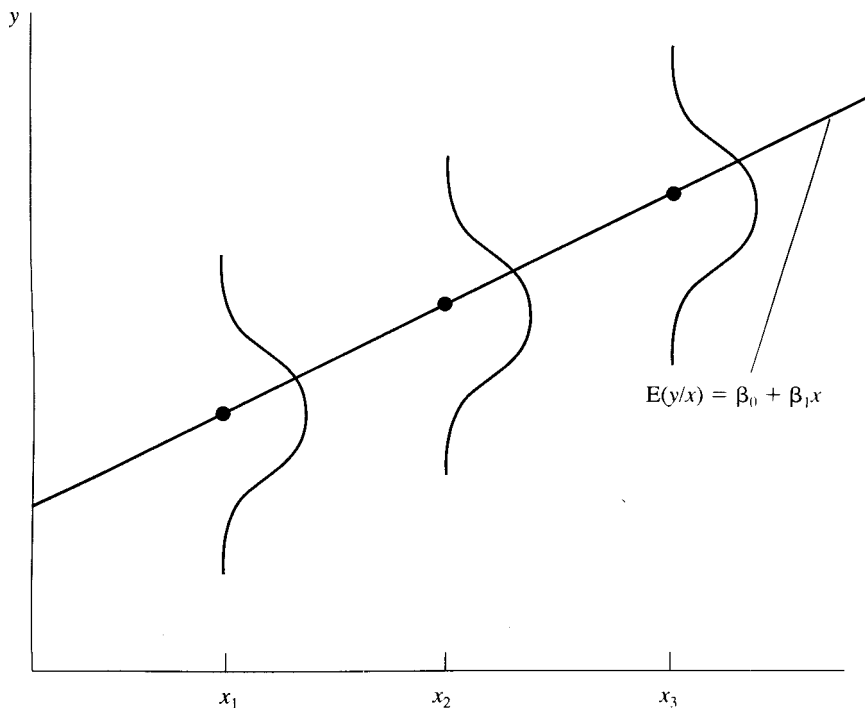
A hipótese (2.6) dá a β_1 outra interpretação que é, freqüentemente, útil. Considerando o valor esperado de (2.1) condicionado a x e usando $E(u|x) = 0$, obtém-se

$$E(y|x) = \beta_0 + \beta_1 x. \quad (2.8)$$

(A equação (2.8) mostra que a **função de regressão populacional (FRP)**, $E(y|x)$, é uma função linear de x . A linearidade significa que o aumento de uma unidade em x faz com que o *valor esperado* de y varie segundo a magnitude de β_1). Para qualquer valor dado de x , a distribuição de y está centrada ao redor de $E(y|x)$, como ilustrado na Figura 2.1.

Figura 2.1

$E(y|x)$ como função linear de x .



Quando (2.6) é verdadeira, é útil dividir y em dois componentes. A parte $\beta_0 + \beta_1 x$ é algumas vezes chamada a *parte sistemática* de y — isto é, a parte de y explicada por x —, e u é chamado a *parte não-sistemática*, ou a parte de y não explicada por x . Usaremos a hipótese (2.6) na próxima seção para encontrar as estimativas de β_0 e β_1 . Essa hipótese também é crucial para a análise estatística na Seção 2.5.

2.2 DERIVAÇÃO DAS ESTIMATIVAS DE MÍNIMOS QUADRADOS ORDINÁRIOS

Agora que discutimos os ingredientes básicos do modelo de regressão simples, trataremos da importante questão de como estimar os parâmetros β_0 e β_1 da equação (2.1). Para tanto, necessitamos de uma amostra da população. Vamos considerar $\{(x_i, y_i): i=1, \dots, n\}$ como uma amostra aleatória de tamanho n da população. Visto que esses dados vêm de (2.1), podemos escrever

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad (2.9)$$

para cada i . Aqui, u_i é o termo erro para a observação i , uma vez que ele contém todos os fatores, além de x_i , que afetam y_i .

Como um exemplo, x_i poderia ser a renda anual e y_i , a poupança anual para a família i durante um determinado ano. Se coletarmos dados de 15 famílias, então $n = 15$. Um gráfico de tal conjunto de dados é dado pela Figura 2.2, juntamente com a função de regressão populacional (necessariamente fictícia).

Devemos decidir como usar esses dados, a fim de obter estimativas do intercepto e da inclinação na regressão populacional da poupança sobre a renda.

Há muitas maneiras de colocar em prática o seguinte procedimento de estimação. Usaremos (2.5) e uma importante implicação da hipótese (2.6): na população, u tem média zero e é não-correlacionada com x . Portanto, vemos que u tem valor esperado zero e que a *covariância* entre x e u é zero:

$$E(u) = 0 \quad (2.10)$$

e

$$\text{Cov}(x, u) = E(xu) = 0, \quad (2.11)$$

onde a primeira igualdade em (2.11) resulta de (2.10). (Veja Seção B.4 do Apêndice B, disponível no site da Thomson, para definição e propriedades da covariância.) Em termos das variáveis observáveis x e y e dos parâmetros desconhecidos β_0 e β_1 , as equações (2.10) e (2.11) podem ser escritas como

$$E(y - \beta_0 - \beta_1 x) = 0 \quad (2.12)$$

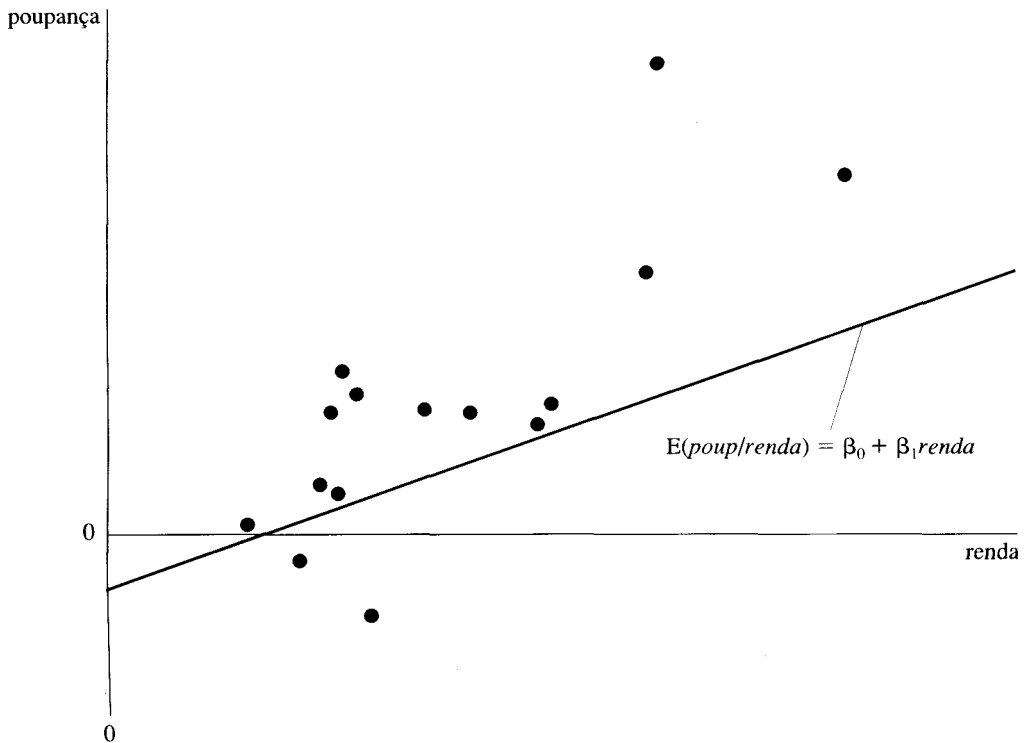
e

$$E[x(y - \beta_0 - \beta_1 x)] = 0, \quad (2.13)$$

respectivamente. As equações (2.12) e (2.13) implicam duas restrições sobre a distribuição de probabilidade conjunta de (x,y) na população. Como há dois parâmetros desconhecidos para estimar, poderíamos esperar que as equações (2.12) e (2.13) pudessem ser usadas para obter bons estimadores de β_0 e β_1 . De fato, elas podem ser usadas. Dada uma amostra de dados, escolhemos as estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ para resolver as equivalências amostrais de (2.12) e (2.13):

Figura 2.2

Gráfico da dispersão de poupança e renda de 15 famílias e a regressão populacional $E(poup|renda) = \beta_0 + \beta_1 \text{renda}$.



$$n^{-1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.14}$$

e

$$n^{-1} \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0. \tag{2.15}$$

Esse é um exemplo da abordagem do *método dos momentos* para a estimação. (Veja a Seção C.4 do Apêndice C, disponível no site da Thomson, para uma discussão das diferentes abordagens de estimação.) Essas equações podem ser resolvidas para $\hat{\beta}_0$ e $\hat{\beta}_1$.

Usando as propriedades básicas do operador somatório a partir do Apêndice A (disponível no site da Thomson), a equação (2.14) pode ser escrita como

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}, \quad (2.16)$$

em que $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ é a média amostral de y_i , e igualmente para \bar{x} . Essa equação nos permite escrever $\hat{\beta}_0$ em termos de $\hat{\beta}_1$, \bar{y} e \bar{x} :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (2.17)$$

Portanto, uma vez que temos a estimativa de inclinação $\hat{\beta}_1$, obtém-se diretamente a estimativa de intercepto $\hat{\beta}_0$, dados \bar{y} e \bar{x} .

Suprimindo o n^{-1} em (2.15) (já que ele não afeta a solução) e inserindo (2.17) em (2.15), obtemos

$$\sum_{i=1}^n x_i [y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i] = 0$$

a qual, após rearranjo, pode ser escrita

$$\sum_{i=1}^n x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^n x_i (x_i - \bar{x}_i).$$

Das propriedades básicas do operador somatório [veja (A.7) e (A.8) disponível no site da Thomson],

$$\sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2 \text{ e } \sum_{i=1}^n x_i (y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y}).$$

Portanto, desde que

$$\sum_{i=1}^n (x_i - \bar{x})^2 > 0, \quad (2.18)$$

a inclinação estimada é

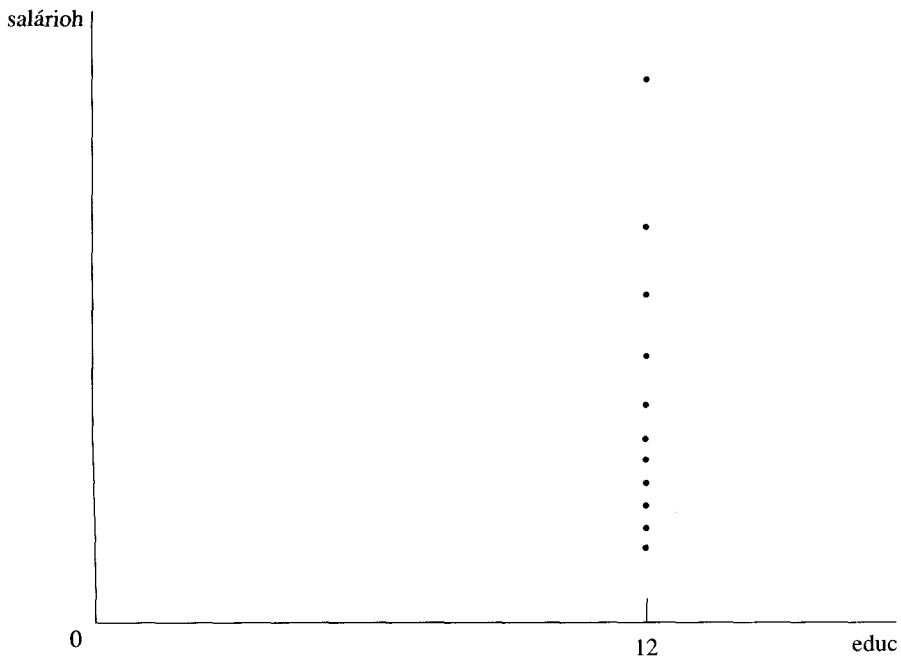
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.19)$$

A equação (2.19) é simplesmente a covariância amostral entre x e y , dividida pela variância amostral de x . (Veja Apêndice C, disponível no site da Thomson. Dividir tanto o numerador como o denominador por $n - 1$ não altera o resultado.) Isso faz sentido, pois β_1 é igual à covariância populacional dividida pela variância de x quando $E(u) = 0$ e $Cov(x, u) = 0$. Uma implicação imediata é que se x e y são positivamente correlacionados na amostra, então $\hat{\beta}_1$ é positivo; se x e y são negativamente correlacionados, então $\hat{\beta}_1$ é negativo.

Embora o método para obter (2.17) e (2.19) decorra de (2.6), a única hipótese necessária para se calcular as estimativas para uma amostra particular é (2.18). Mas essa raramente é uma hipótese: (2.18) é verdadeira sempre que os x_i na amostra não são todos iguais a um mesmo valor. Se (2.18) não se sustentar, então fomos infelizes em obter nossa amostra da população, ou não especificamos um problema interessante (x não varia na população). Por exemplo, se $y = \text{salário}_i$ e $x = \text{educ}$, então (2.18) não se mantém se todos na amostra têm a mesma quantidade de anos de educação formal. (Por exemplo, se todos têm o equivalente ao ensino médio concluído. Veja a Figura 2.3.) Se apenas uma pessoa tem uma quantidade diferente de anos de educação formal, então (2.18) se sustenta, e as estimativas de MQO podem ser calculadas.

Figura 2.3

Gráfico da dispersão de salários e educação, quando $\text{educ}_i = 12$ para todo i .



As estimativas dadas em (2.17) e (2.19) são chamadas de estimativas de **mínimos quadrados ordinários (MQO)** de β_0 e β_1 . Para justificar esse nome, defina, para qualquer $\hat{\beta}_0$ e $\hat{\beta}_1$, um valor estimado para y quando $x = x_i$, tal como

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad (2.20)$$

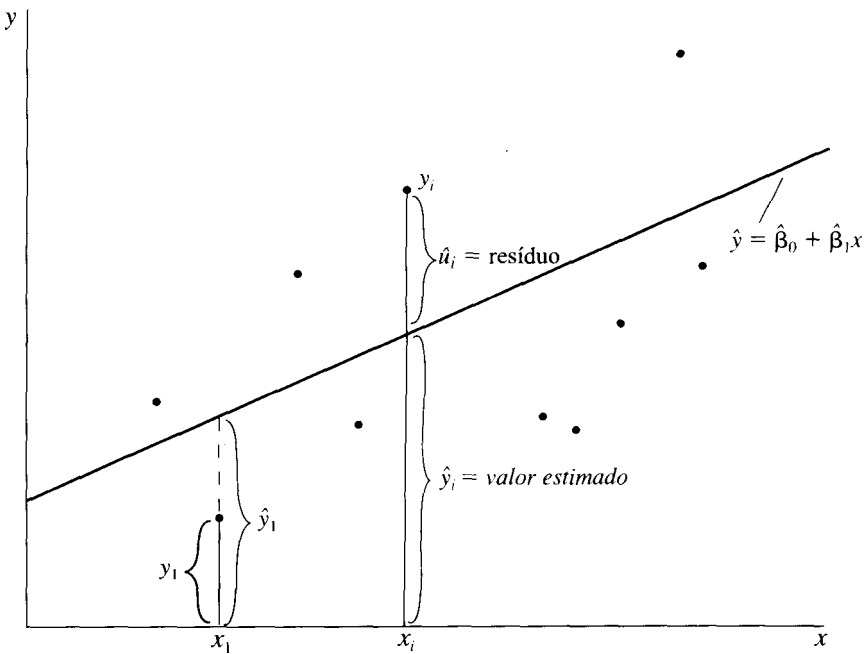
para o intercepto e a inclinação dados. Esse é o valor que prevemos para y quando $x = x_i$. Há um valor estimado para cada observação na amostra. O **resíduo** para a observação i é a diferença entre o valor verdadeiro de y_i e seu valor estimado:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i. \quad (2.21)$$

Novamente, há n desses resíduos. [Eles *não* são iguais aos erros em (2.9), um ponto ao qual retornaremos na Seção 2.5.] Os valores estimados e os resíduos estão indicados na Figura 2.4.

Figura 2.4

Valores estimados e resíduos.



Agora, suponha que escolhemos $\hat{\beta}_0$ e $\hat{\beta}_1$ com a finalidade de tornar a **soma dos resíduos quadrados**,

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \quad (2.22)$$

tão pequena quanto possível. O apêndice deste capítulo mostra que as condições necessárias para $(\hat{\beta}_0, \hat{\beta}_1)$ minimizarem (2.22) são dadas exatamente pelas equações (2.14) e (2.15), sem n^{-1} . As equações (2.14) e (2.15) são frequentemente chamadas de **condições de primeira ordem** para as estimativas de MQO, um termo que vem da otimização utilizada em cálculo (veja o Apêndice A, disponível no site da Thomson). De nossos cálculos anteriores, sabemos que as soluções para as condições de primeira

ordem de MQO são dadas por (2.17) e (2.19). O nome “mínimos quadrados ordinários” vem do fato de que essas estimativas minimizam a soma dos resíduos quadrados.

Quando vemos o método de mínimos quadrados ordinários como um método que minimiza a soma dos resíduos quadrados, é natural perguntar: por que não minimizar alguma outra função dos resíduos, como o valor absoluto dos resíduos? De fato, como discutiremos brevemente na Seção 9.4, minimizar a soma dos valores absolutos dos resíduos é, algumas vezes, muito útil. Mas esse procedimento também tem suas desvantagens. Primeiro, não podemos obter fórmulas para os estimadores resultantes; dado um conjunto de dados, as estimativas devem ser obtidas por rotinas de otimização numérica. Em consequência, a teoria estatística para estimadores que minimizam a soma dos resíduos absolutos é muito complicada. Minimizar outras funções dos resíduos, como a soma de cada resíduo elevado à quarta potência, tem desvantagens similares. (Nunca deveríamos escolher nossos estimadores para minimizar, por exemplo, a soma dos próprios resíduos, pois resíduos grandes em magnitude e com sinais opostos tendem a se cancelar.) Com o método MQO, seremos capazes de derivar, de modo relativamente fácil, inexistência de viés, consistência e outras importantes propriedades estatísticas. E mais, como as equações (2.13) e (2.14) sugerem, e como veremos na Seção 2.5, o método MQO é adequado para estimar os parâmetros que aparecem na função de média condicional (2.8).

Uma vez determinados os estimadores de intercepto e inclinação de MQO, construímos a **reta de regressão de MQO**:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, \quad (2.23)$$

em que $\hat{\beta}_0$ e $\hat{\beta}_1$ foram obtidos ao usar as equações (2.17) e (2.19). A notação \hat{y} — leia-se “y chapéu” — enfatiza que os valores previstos da equação (2.23) são estimativas. O intercepto $\hat{\beta}_0$ é o valor previsto de y quando $x = 0$, embora, em alguns casos, não faça sentido considerar $x = 0$. Nestas situações, $\hat{\beta}_0$ não é, por si mesmo, muito interessante. Ao usar (2.23) para calcular os valores previstos de y para vários valores de x , devemos considerar o intercepto nos cálculos. A equação (2.23) é também chamada **função de regressão amostral (FRA)**, pois ela é a versão estimada da função de regressão populacional $E(y|x) = \beta_0 + \beta_1 x$. É importante lembrar que a FRP é algo fixo, porém desconhecido, na população. Como a FRA é obtida para uma dada amostra de dados, uma amostra nova gerará uma inclinação e um intercepto diferentes na equação (2.23).

Em muitos casos, a estimativa do coeficiente de inclinação, que podemos escrever como

$$\hat{\beta}_1 = \Delta \hat{y} / \Delta x, \quad (2.24)$$

é de interesse fundamental. Ela nos diz o quanto varia \hat{y} quando x aumenta em uma unidade. Equivalentemente,

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x, \quad (2.25)$$

de modo que, dada qualquer variação em x (seja positiva ou negativa), podemos calcular a variação prevista em y .

Agora, vamos apresentar vários exemplos de regressões simples obtidas de dados reais. Em outras palavras, vamos encontrar as estimativas de intercepto e de inclinação a partir das equações (2.17) e (2.19). Como esses exemplos compreendem muitas observações, os cálculos foram feitos usando pro-

gramas econométricos. Neste ponto, não se preocupe muito em interpretar as regressões; elas não estão, necessariamente, revelando uma relação causal. Até aqui, não dissemos nada sobre as propriedades estatísticas do método MQO. Na Seção 2.5, consideraremos as propriedades depois de importamos explicitamente hipóteses sobre a equação do modelo populacional (2.1).

EXEMPLO 2.3

(Salários de Diretores Executivos e Retornos de Ações)

Para a população de diretores executivos, seja y o salário anual (*salário*) em milhares de dólares. Assim, $y = 856,3$ indica um salário anual de \$ 856.300, e $y = 1.452,6$ indica um salário de \$ 1.452.600. Seja x o retorno médio da ação sobre o patrimônio (*rma*), dos três anos anteriores, da empresa do diretor executivo. (O retorno da ação sobre o patrimônio é definido em termos de renda líquida, como uma porcentagem do patrimônio comum.) Por exemplo, se $rma = 10$, então o retorno médio da ação sobre o patrimônio é de 10%.

Para estudar a relação entre essa medida do desempenho das empresas e a remuneração dos seus diretores executivos, postulamos o modelo simples

$$\text{salário} = \beta_0 + \beta_1 rma + u.$$

O parâmetro de inclinação β_1 mede a variação no salário anual, em milhares de dólares, quando o retorno da ação aumenta em um ponto percentual. Como um *rma* mais elevado é melhor para a empresa, esperamos que $\beta_1 > 0$.

O conjunto de dados do arquivo CEOSAL1.RAW contém informações sobre 209 diretores executivos para o ano de 1990; esses dados foram obtidos da revista *Business Week* (6.5.91). Na amostra, o salário médio anual é \$ 1.281.120; sendo que o menor e o maior são \$ 223.000 e \$ 14.822.000, respectivamente. O retorno médio das ações para os anos 1988, 1989 e 1990 é de 17,18%, sendo que os valores menor e maior são de 0,5 e 56,3%, respectivamente.

Usando os dados do arquivo CEOSAL1.RAW, a reta de regressão de MQO que relaciona *salário* a *rma* é

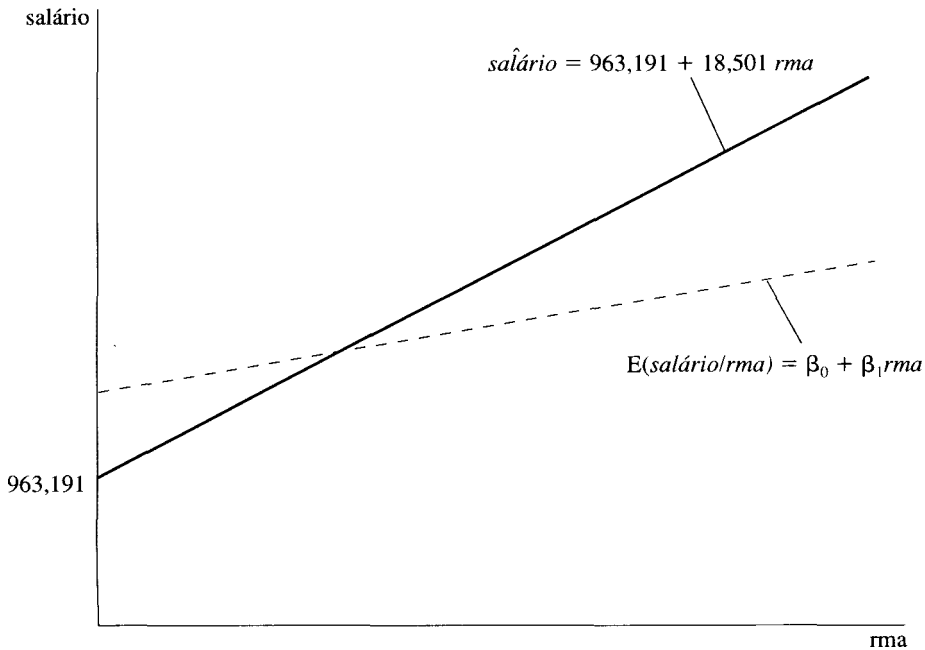
$$\text{salário} = 963,191 + 18,501 rma, \quad (2.26)$$

na qual as estimativas de intercepto e de inclinação foram arredondadas em três casas decimais; usamos "*salário* chapéu" para indicar que essa é uma equação estimada. Como interpretamos a equação? Primeiro, se o retorno da ação é zero, $rma = 0$, então o salário previsto é o intercepto, 963,191, que é igual a \$ 963.191, visto que *salário* é mensurado em milhares. Em seguida, podemos escrever a variação prevista no salário como uma função da variação em *rma*: $\Delta \text{salário} = 18,501(\Delta rma)$. Isso significa que se o retorno da ação aumenta um ponto percentual, $\Delta rma = 1$, então espera-se que *salário* variará cerca de 18,5, ou \$ 18.500. Como (2.26) é uma equação linear, esse valor é a variação estimada independentemente do salário inicial.

Podemos facilmente usar (2.26) para comparar salários previstos para valores diferentes de *rma*. Suponha $rma = 30$. Portanto, $\text{salário} = 963,191 + 18,501(30) = 1.518,221$, o que está pouco acima de \$ 1,5 milhão. Entretanto, isso não significa que um determinado diretor executivo, cuja empresa tenha um $rma = 30$, ganhe \$ 1.518.221. Muitos outros fatores afetam o salário. Essa é somente a nossa previsão a partir da reta de regressão de MQO de (2.26). A reta estimada está representada na Figura 2.5, juntamente com a função de regressão populacional $E(\text{salário}|rma)$. Nunca conheceremos a FRP; assim, não podemos dizer o quão próxima a FRA está da FRP. Outra amostra de dados levará a uma reta de regressão diferente, que pode estar, ou não, mais próxima da reta de regressão populacional.

Figura 2.5

A reta de regressão de MQO $\hat{\text{salário}} = 963,191 + 18,501 \text{ rma}$ e a função de regressão populacional (desconhecida).

**EXEMPLO 2.4****(Salários e Educação)**

Para a população de pessoas na força de trabalho em 1976, seja $y = \text{salário}_h$, em que salário_h é mensurado em dólares por hora. Assim, para uma determinada pessoa, se $\text{salário}_h = 6,75$, o salário-hora é \$ 6,75. Vamos chamar anos de escolaridade formal de $x = \text{educ}$; por exemplo, $\text{educ} = 12$ corresponde ao ensino médio completo (nos Estados Unidos). O salário horário médio na amostra é \$ 5,90, o que equivale, de acordo com o índice de preços ao consumidor dos Estados Unidos, a \$ 16,64 em dólares de 1997.

Usando os dados do arquivo WAGE1.RAW, em que $n = 526$ indivíduos, obtemos a seguinte reta de regressão de MQO (ou função de regressão amostral):

$$\hat{\text{salário}} = -0,90 + 0,54 \text{ educ.} \quad (2.27)$$

Devemos interpretar essa equação com cautela. O intercepto de $-0,90$ significa, literalmente, que uma pessoa sem nenhuma educação formal tem um salário-hora previsto de -90 centavos de dólar por hora. Isso, evidentemente, é tolice. Ocorre que apenas 18 pessoas na amostra de 526 têm menos que oito anos de educação formal. Conseqüentemente, não é surpreendente que a reta de regressão não faça boas previsões para

EXEMPLO 2.4 (continuação)

níveis de educação formal muito baixos. Para uma pessoa com oito anos de educação formal, o salário previsto é $\text{salário} = -0,90 + 0,54(8) = 3,42$, ou \$ 3,42 por hora (em dólares de 1976).

A inclinação estimada em (2.27) implica que um ano a mais de educação formal aumenta o salário horário em 54 centavos de dólar por hora. Portanto, quatro anos a mais de educação formal aumentam o salário horário previsto em $4(0,54) = 2,16$, ou \$ 2,16 por hora. Esses efeitos são razoavelmente grandes. Devido à natureza linear de (2.27), outro ano de educação formal aumenta o salário na mesma quantidade, independentemente do nível inicial de educação. Na Seção 2.4, discutiremos alguns métodos que levam em consideração efeitos marginais não-constantes de nossas variáveis explicativas.

O salário horário estimado em (2.27), quando $\text{educ} = 8$, é \$ 3,42, em dólares de 1976. Qual é esse valor em dólares de 1997? (*Sugestão*: você tem informação suficiente, no Exemplo 2.4, para responder a essa questão.)

EXEMPLO 2.5**(Resultados Eleitorais e Gastos de Campanha)**

O arquivo VOTE1.RAW contém dados sobre resultados eleitorais e gastos de campanha de 173 disputas entre dois partidos, para a *House of Representatives* dos Estados Unidos (equivalente a uma câmara federal). Há dois candidatos em cada disputa: A e B. Seja votoA a percentagem de votos recebida pelo Candidato A e partA a percentagem dos gastos totais de campanha que cabem ao Candidato A. Muitos outros fatores além de partA afetam o resultado eleitoral (incluindo a qualidade dos candidatos e os valores absolutos dos gastos de A e B). No entanto, podemos estimar um modelo de regressão simples para descobrir se gastar mais do que o concorrente implica uma percentagem maior de votos.

A equação estimada usando as 173 observações é

$$\text{voto A} = 26,81 + 0,464 \text{ partA}. \quad (2.28)$$

Isso significa que, se a parte dos gastos do Candidato A aumenta em um ponto percentual, o Candidato A recebe quase meio ponto percentual (0,464) a mais da votação total. Não fica claro se isso revela ou não um efeito causal, mas isso é crível. Se $\text{partA} = 50$, prevê-se que votoA será cerca de 50, ou metade da votação.

Em alguns casos, a análise de regressão não é usada para determinar a causalidade, mas para simplesmente observar se duas variáveis são positiva ou negativamente relacionadas, de modo muito parecido com uma análise padrão de correlação. Um exemplo disso ocorre no Problema 2.12, que pede que você use os dados de Biddle e Hamermesh (1990) referentes ao tempo que se gasta dormindo e trabalhando a fim de investigar a relação entre esses dois fatores.

No Exemplo 2.5, qual é a votação prevista para o Candidato A se $partA = 60$ (que significa 60%)? A resposta parece razoável?

Uma Nota sobre Terminologia

Em muitos casos, indicaremos a estimação de uma relação através de MQO ao escrever uma equação como (2.26), (2.27) ou (2.28). Algumas vezes, por motivo de brevidade, é útil indicar que uma regressão de MQO foi estimada sem realmente escrever a equação. Frequentemente, indicaremos que a equação (2.23) foi obtida por MQO ao dizer que nós *rodamos* a regressão de

$$y \text{ sobre } x, \tag{2.29}$$

ou simplesmente que *regredimos* y sobre x . As posições de y e x em (2.29) indicam qual é a variável dependente e qual é a variável independente: sempre regredimos a variável dependente sobre a variável independente. Para aplicações específicas, substituiremos y e x por seus nomes. Assim, para obter (2.26), regredimos *salário* sobre *rma*, ou, para obter (2.28), regredimos *votoA* sobre *partA*.

Ao usarmos essa terminologia em (2.29), sempre estaremos dizendo que planejamos estimar o intercepto, $\hat{\beta}_0$, juntamente com o coeficiente de inclinação, $\hat{\beta}_1$. Esse caso é apropriado para a maioria das aplicações. Ocasionalmente, podemos querer estimar a relação entre y e x *assumindo* que o intercepto é zero (de modo que $x = 0$ implica $\hat{y} = 0$); cobriremos esse caso, brevemente, na Seção 2.6. A não ser que seja explicitamente dito de outro modo, sempre estimaremos um intercepto juntamente com uma inclinação.

2.3 MECÂNICA DO MÉTODO MQO

Nesta seção, cobriremos algumas propriedades algébricas da reta de regressão de MQO estimada. Talvez, a melhor maneira de pensar nessas propriedades é perceber que elas são características de MQO para uma determinada amostra de dados. Elas podem ser contrastadas com as propriedades *estatísticas* de MQO, o que requer a derivação das características das distribuições amostrais dos estimadores. Discutiremos as propriedades estatísticas na Seção 2.5.

Muitas das propriedades algébricas que derivaremos parecerão triviais. No entanto, ter uma compreensão dessas propriedades ajuda-nos a entender o que acontece com as estimativas de MQO e estatísticas relacionadas quando os dados são manipulados de determinadas maneiras, como quando variam as unidades de medida das variáveis dependente e independentes.

Valores Estimados e Resíduos

Assumimos que as estimativas de intercepto e de inclinação, $\hat{\beta}_0$ e $\hat{\beta}_1$, foram obtidas de uma dada amostra de dados. Dados $\hat{\beta}_0$ e $\hat{\beta}_1$, podemos obter o valor estimado \hat{y}_i para cada observação. [Isso é dado pela equação (2.20).] Por definição, cada valor estimado de \hat{y}_i está sobre a reta de regressão de MQO. O resíduo de MQO associado a cada observação i , \hat{u}_i , é a diferença entre y_i e seu valor estimado, como dado na equação (2.21). Se \hat{u}_i é positivo, a reta subestima y_i ; se \hat{u}_i é negativo, a reta superestima y_i . O caso ideal para a observação i é quando $\hat{u}_i = 0$, mas na maior parte dos casos *todos* os resíduos são

diferentes de zero. Em outras palavras, nenhum dos pontos dos dados deve, realmente, estar sobre a reta de MQO.

EXEMPLO 2.6

(Salário de Diretores Executivos e Retornos de Ações)

A Tabela 2.2 contém uma lista das 15 primeiras observações do conjunto de dados dos salários dos diretores executivos, juntamente com os valores estimados, chamados de *salchapéu*, e os resíduos, estimados de *uchapéu*.

Tabela 2.2

Valores Estimados e Resíduos dos 15 Primeiros Diretores Executivos

<i>nobsd</i>	<i>rma</i>	<i>salário</i>	<i>salchapéu</i>	<i>uchapéu</i>
1	14,1	1.095	1.224,058	-129,0581
2	10,9	1.001	1.164,854	-163,8542
3	23,5	1.122	1.397,969	-275,9692
4	5,9	578	1.072,348	-494,3484
5	13,8	1.368	1.218,508	149,4923
6	20,0	1.145	1.333,215	-188,2151
7	16,4	1.078	1.266,611	-188,6108
8	16,3	1.094	1.264,761	-170,7606
9	10,5	1.237	1.157,454	79,54626
10	26,3	833	1.449,773	-616,7726
11	25,9	567	1.442,372	-875,3721
12	26,8	933	1.459,023	-526,0231
13	14,8	1.339	1.237,009	101,9911
14	22,3	937	1.375,768	-438,7678
15	56,3	2.011	2.004,808	6,191895

Os quatro primeiros diretores executivos têm salários menores do que os previstos a partir da reta de regressão de MQO (2.26); em outras palavras, dado somente o *rma* da empresa, esses diretores executivos ganham menos do que prevemos. Como pode ser visto dos *uchapéus* positivos, o quinto diretor executivo ganha mais do que prevemos a partir da reta de regressão de MQO.

Propriedades Algébricas das Estatísticas de MQO

Há várias propriedades algébricas úteis das estimativas de MQO e das estatísticas a elas associadas. Vamos discutir as três mais importantes.

(1) A soma, e portanto a média amostral, dos resíduos de MQO, é zero. Matematicamente,

$$\sum_{i=1}^n \hat{u}_i = 0. \quad (2.30)$$

Essa propriedade não precisa de prova; ela resulta, imediatamente, da condição de primeira ordem de MQO (2.14), quando lembramos que os resíduos são definidos por $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. Em outras palavras, as estimativas de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ são escolhidas para fazer com que a soma dos resíduos seja zero (para qualquer conjunto de dados). Isso não diz nada sobre o resíduo de qualquer observação i em particular.

(2) A covariância amostral entre os regressores e os resíduos de MQO é zero. Isso resulta da condição de primeira ordem (2.15), que pode ser escrita em termos dos resíduos, como

$$\sum_{i=1}^n x_i \hat{u}_i = 0. \quad (2.31)$$

A média amostral dos resíduos de MQO é zero, de modo que o lado esquerdo de (2.31) é proporcional à covariância amostral entre x_i e \hat{u}_i .

(3) O ponto (\bar{x}, \bar{y}) sempre está sobre a reta de regressão de MQO. Em outras palavras, se considerarmos a equação (2.23) e inserirmos \bar{x} no lugar de x , então o valor estimado é \bar{y} . Isso é exatamente o que a equação (2.16) nos mostrou.

EXEMPLO 2.7

(Salários e Educação)

Para os dados do arquivo em WAGE1.RAW, o salário-hora médio da amostra é 5,90, arredondado para duas casas decimais, e a educação formal média (medida em anos) é 12,56. Se inserirmos $educ = 12,56$ na reta de regressão de MQO (2.27), obtemos $salarioh = -0,90 + 0,54(12,56) = 5,8824$, igual a 5,9 quando arredondamos para uma casa decimal. A razão de esses números não serem exatamente os mesmos é que nós arredondamos o salário-hora e os anos de educação formal médios, assim como as estimativas de intercepto e de inclinação. Se, inicialmente, não tivéssemos arredondado nenhum dos valores, obteríamos respostas mais aproximadas, mas essa prática teria pouco efeito útil.

Ao escrever cada y_i como o seu valor estimado mais seu resíduo, temos outro modo de interpretar uma regressão de MQO. Para cada i , escreva

$$y_i = \hat{y}_i + \hat{u}_i. \quad (2.32)$$

Da propriedade (1), a média dos resíduos é zero; equivalentemente, a média amostral dos valores estimados, \hat{y}_i , é a mesma que a média amostral de y_i , ou $\bar{\hat{y}} = \bar{y}$. Além disso, as propriedades (1) e (2)

podem ser usadas para mostrar que a covariância amostral entre \hat{y}_i e \hat{u}_i é zero. Portanto, podemos ver MQO como um método que decompõe cada y_i em duas partes: um valor estimado e um resíduo. Os valores estimados e os resíduos são não-correlacionados na amostra.

Defina a **soma dos quadrados total (SQT)**, a **soma dos quadrados explicada (SQE)** e a **soma dos quadrados dos resíduos (SQR)** (também conhecida como a soma dos resíduos quadrados), como a seguir:

$$\text{SQT} \equiv \sum_{i=1}^n (y_i - \bar{y})^2. \quad (2.33)$$

$$\text{SQE} \equiv \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (2.34)$$

$$\text{SQR} \equiv \sum_{i=1}^n \hat{u}_i^2. \quad (2.35)$$

SQT é uma medida da variação amostral total em y_i ; isto é, ela mede o quão dispersos estão os y_i na amostra. Se dividirmos SQT por $n - 1$, obteremos a variância amostral de y , como discutido no Apêndice C; no site da Thomson. Semelhantemente, SQE mede a variação amostral em \hat{y}_i (em que usamos o fato de que $\bar{\hat{y}} = \bar{y}$), e SQR mede a variação amostral em \hat{u}_i . A variação total em y pode sempre ser expressa como a soma da variação explicada e da variação não-explicada SQR. Assim,

$$\text{SQT} = \text{SQE} + \text{SQR}. \quad (2.36)$$

Provar (2.36) não é difícil; mas requer o uso de todas as propriedades do operador somatório apresentadas no Apêndice A, no site da Thomson. Escreva

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n [\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 + 2 \sum_{i=1}^n \hat{u}_i (\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \text{SQR} + 2 \sum_{i=1}^n \hat{u}_i (\hat{y}_i - \bar{y}) + \text{SQE}. \end{aligned}$$

Agora, (2.36) é válida se mostrarmos que

$$\sum_{i=1}^n \hat{u}_i (\hat{y}_i - \bar{y}) = 0. \quad (2.37)$$

Mas já dissemos que a covariância amostral entre os resíduos e os valores estimados é zero, e essa covariância é justamente a equação (2.37) dividida por $n - 1$. Conseqüentemente, confirmamos (2.36).

Algumas palavras de precaução sobre SQT, SQE e SQR devem ser mencionadas. Não há concordância uniforme sobre os nomes e abreviações das três quantidades definidas nas equações (2.33), (2.34) e (2.35). A soma dos quadrados total é chamada SQT ou STQ, de modo que aqui não há grandes confusões. Infelizmente, a soma dos quadrados explicada é, às vezes, chamada de “soma dos quadrados da regressão”. Se a esse termo é dado sua abreviação natural, ele pode ser facilmente confundido com o termo “soma dos quadrados dos resíduos”. Alguns programas econométricos referem-se à soma dos quadrados explicada como “soma dos quadrados do modelo”.

Para tornar as coisas ainda piores, a soma dos quadrados dos resíduos é freqüentemente chamada de “soma dos quadrados dos erros”. Esse termo é um tanto inadequado, pois, como veremos na Seção 2.5, os erros e os resíduos são quantidades diferentes. Assim, sempre chamaremos (2.35) de soma dos quadrados dos resíduos ou soma dos resíduos quadrados. Preferimos usar a abreviação SQR para representar a soma dos resíduos quadrados, pois ela é mais comum nos programas econométricos.

Grau de Ajuste

Até aqui, não apresentamos uma maneira de mensurar o quanto bem a variável explicativa ou independente, x , explica a variável dependente, y . Muitas vezes, é útil calcular um número que resume o quão bem a reta de regressão de MQO se ajusta aos dados. Na discussão seguinte, lembre-se de que assumimos estimar o intercepto com a inclinação.

Ao assumirmos que a soma dos quadrados total, SQT, não é igual a zero — o que é verdadeiro, a não ser no evento muito improvável de todos os y_i serem iguais a um mesmo valor —, podemos dividir (2.36) por SQT para obter $1 = \text{SQE}/\text{SQT} + \text{SQR}/\text{SQT}$. O **R-quadrado** da regressão, algumas vezes chamado coeficiente de determinação, é definido como

$$R^2 = \text{SQE}/\text{SQT} = 1 - \text{SQR}/\text{SQT}. \quad (2.38)$$

R^2 é a razão entre a variação explicada e a variação total; assim, ele é interpretado como a *fração da variação amostral em y que é explicada por x* . A segunda equação em (2.38) fornece outra maneira de calcular R^2 .

De (2.36), o valor de R^2 está sempre entre zero e um, visto que SQE não pode ser maior que SQT. Quando interpretamos R^2 , usualmente o multiplicamos por 100 para transformá-lo em percentual: $100 \cdot R^2$ é a *percentagem da variação amostral em y que é explicada por x* .

Se todos os pontos dos dados estiverem sobre a mesma reta, MQO fornece um ajuste perfeito aos dados. Nesse caso, $R^2 = 1$. Um valor de R^2 quase igual a zero indica um ajuste ruim da reta de MQO: muito pouco da variação em y_i é capturado pela variação em \hat{y}_i (que está sobre a reta de regressão de MQO). De fato, pode ser mostrado que R^2 é igual ao *quadrado* do coeficiente de correlação amostral entre y_i e \hat{y}_i . É daí que vem o termo “R-quadrado”. (A letra R era, tradicionalmente, usada para denominar uma estimativa do coeficiente de correlação populacional, e seu uso sobreviveu na análise de regressão.)

EXEMPLO 2.8**(Salário de Diretores Executivos e Retornos de Ações)**

Na regressão de salários de diretores executivos, estimamos a seguinte equação:

$$\widehat{\text{salário}} = 963,191 + 18,501 rma \quad (2.39)$$

$$n = 209, R^2 = 0,0132.$$

Por motivos de clareza, reproduzimos a reta de regressão de MQO e o número de observações. Usando o R -quadrado (arredondado para quatro casas decimais) apresentado para essa equação, podemos ver quanto da variação no salário é, realmente, explicada pelo retorno da ação. A resposta é: não muito. O retorno da ação da empresa explica somente 1,3% da variação nos salários dessa amostra de 209 diretores executivos. Isso significa que 98,7% das variações salariais desses diretores executivos são deixadas sem explicação. Essa falta de poder explicativo não deve ser surpreendente demais, já que muitas outras características, tanto da empresa como do diretor executivo individual, devem influenciar o salário; esses fatores estão, necessariamente, incluídos nos erros de uma análise de regressão simples.

Nas ciências sociais não são incomuns R -quadrados baixos nas equações de regressão, especialmente na análise de corte transversal. Discutiremos essa questão, de modo mais geral, sob a análise de regressão múltipla, mas vale a pena enfatizar agora que um R -quadrado aparentemente baixo não significa, necessariamente, que uma equação de regressão de MQO é inútil. Ainda, é possível que (2.39) seja uma boa estimativa da relação *ceteris paribus* entre *salário* e *rma*; se isso for verdade ou não, *não* depende diretamente da magnitude do R -quadrado. Os estudantes que estão se defrontando com econometria pela primeira vez tendem, ao avaliar equações de regressão, a pôr muito peso na magnitude do R -quadrado. Por enquanto, esteja ciente de que usar o R -quadrado como o principal padrão de medida de sucesso de uma análise econométrica pode levar a confusões.

Algumas vezes, a variável explicativa elucida uma parte substancial da variação amostral na variável dependente.

EXEMPLO 2.9**(Resultados Eleitorais e Gastos de Campanha)**

Na equação de resultados eleitorais (2.28), $R^2 = 0,856$. Assim, a participação dos candidatos nos gastos de campanha explica mais de 85% da variação nos resultados eleitorais nessa amostra. Essa é uma explicação considerável.

2.4 UNIDADES DE MEDIDA E FORMA FUNCIONAL

Duas questões importantes em economia aplicada são: (1) entender como, ao mudar as unidades de medida das variáveis dependente e/ou independente, são afetadas as estimativas de MQO e (2) saber como incorporar, à análise de regressão, formas funcionais populares usadas em economia.

A matemática necessária para uma compreensão completa das questões sobre a forma funcional está revista no Apêndice A, disponível no site da Thomson.

Os Efeitos de Mudanças das Unidades de Medida sobre as Estatísticas de MQO

No Exemplo 2.3, escolhemos mensurar o salário anual em milhares de dólares, e o retorno das ações foi medido como uma percentagem (em vez de um decimal). É crucial saber como *salário* e *rma* são medidos nesse exemplo, a fim de dar sentido às estimativas da equação (2.39).

Devemos também saber que as estimativas de MQO mudam de maneira completamente esperada, quando as unidades de medida das variáveis dependente e independente mudam. No Exemplo 2.3, suponhamos que, em vez de medir o salário em milhares de dólares, nós o medimos em dólares. Seja *salardol* o salário em dólares (*salardol* = 845.761 seria interpretado como \$ 845.761). Evidentemente, *salardol* tem uma relação simples com o salário medido em milhares de dólares: *salardol* = 1.000. *salário*. Não precisamos, realmente, computar a regressão *salardol* sobre *rma* para saber que a equação estimada é:

$$\hat{\text{salardol}} = 963.191 + 18.501 \text{ rma.} \quad (2.40)$$

Obtemos o intercepto e a inclinação em (2.40) ao, simplesmente, multiplicarmos o intercepto e o coeficiente de inclinação em (2.39) por 1.000. Isso dá às equações (2.39) e (2.40) a mesma interpretação. Olhando para (2.40), se *rma* = 0, então *salardol* = 963.191, de modo que o salário previsto é \$ 963.191 [o mesmo valor que obtivemos da equação (2.39)]. Além disso, se *rma* aumenta em um, então o salário previsto aumenta em \$ 18.501; novamente, isso é o que concluímos de nossa análise anterior da equação (2.39).

Em geral, é fácil fazer uma idéia do que acontece às estimativas de intercepto e de inclinação quando se altera a unidade de medida da variável dependente. Se a variável dependente é multiplicada pela constante *c* — o que significa dizer que cada valor na amostra é multiplicado por *c* —, então as estimativas de MQO de intercepto e de inclinação também são multiplicadas por *c*. (Isso assume que nada foi alterado com respeito à variável independente.) No exemplo do salário dos diretores executivos, *c* = 1.000 ao passarmos de *salário* para *salardol*.

Suponha que o salário seja mensurado em centenas de dólares, em vez de milhares de dólares, e o chamemos *salarcent*. Quais serão as estimativas de intercepto e de inclinação na regressão de *salarcent* sobre *rma*?

Também podemos usar o exemplo do salário dos diretores executivos para ver o que acontece quando as unidades de medida da variável independente são mudadas. Defina *rmadec* = *rma*/100 como sendo o equivalente decimal de *rma*; assim, *rmadec* = 0,23 significa um retorno da ação de 23%. A fim de centrarmos o foco na mudança das unidades de medida da variável independente, retornaremos à nossa variável dependente original, *salário*, mensurada em milhares de dólares. Quando regredimos *salário* sobre *rmadec*, obtemos

$$\hat{\text{salário}} = 963,191 + 1.850,1 \text{ rmadec.} \quad (2.41)$$

O coeficiente de *rmadec* é 100 vezes o coeficiente de *rma* em (2.39). Isso é o que deveria ser. Variar *rma* em um ponto percentual é equivalente a $\Delta \text{rmadec} = 0,01$. De (2.41), se $\Delta \text{rmadec} = 0,01$, então

$\Delta\text{salário} = 1.850,1(0,01) = 18,501$, que é igual ao obtido ao se usar (2.39). Observe que, ao passarmos de (2.39) para (2.41), a variável independente foi dividida por 100, e assim a estimativa de inclinação de MQO foi multiplicada por 100, preservando a interpretação da equação. Em geral, se a variável independente é dividida ou multiplicada por alguma constante diferente de zero, c , então o coeficiente de inclinação de MQO é multiplicado ou dividido por c , respectivamente.

O intercepto não mudou em (2.41), pois $r_{\text{madec}} = 0$ ainda corresponde a um retorno zero da ação. Em geral, mudar somente as unidades de medida da variável independente não afeta o intercepto.

Na seção anterior, definimos R -quadrado como uma medida de grau de ajuste para a regressão de MQO. Podemos também questionar o que acontece ao R^2 quando é mudada a unidade de medida da variável independente ou da variável dependente. Sem fazer nenhuma álgebra, deveríamos saber o resultado: o grau de ajuste do modelo não depende das unidades de medida de nossas variáveis. Por exemplo, a quantidade de variação no salário, explicada pelo retorno da ação, não deve depender de o salário ser medido em dólares ou em milhares de dólares, ou de o retorno da ação ser uma porcentagem ou um decimal. Essa intuição pode ser verificada matematicamente: usando a definição de R^2 , pode ser mostrado que R^2 é, de fato, invariante a mudanças nas unidades de y ou x .

Incorporação de Não-Linearidades na Regressão Simples

Até aqui, enfatizamos as relações *lineares* entre as variáveis dependente e independente. Como mencionamos no Capítulo 1, relações lineares não são, em geral, suficientes para todas as aplicações econômicas. Felizmente, é bastante fácil incorporar muitas não-linearidades na análise de regressão simples ao definir apropriadamente as variáveis dependente e independente. Vamos tratar aqui de duas possibilidades que frequentemente aparecem em trabalhos aplicados.

Ao ler trabalhos aplicados nas ciências sociais, com frequência você encontrará equações de regressão em que a variável dependente aparece na forma logarítmica. Por que isso é feito? Lembre-se do exemplo salários-educação, em que regredimos o salário-hora sobre os anos de educação formal. Obtivemos uma estimativa da inclinação de 0,54 [veja a equação (2.27)], o que significa dizer que, para cada ano adicional de educação, é previsto um aumento de 54 centavos de dólar no salário-hora. Devido à natureza linear de (2.27), 54 centavos de dólar é o aumento tanto para o primeiro ano de educação quanto para o vigésimo ano; isso pode não ser razoável.

Suponha, em vez disso, que o aumento percentual no salário é o mesmo, dado um ano a mais de educação formal. O modelo (2.27) não implica um aumento percentual constante: o aumento depende do salário inicial. Um modelo que gera (aproximadamente) um efeito percentual constante é

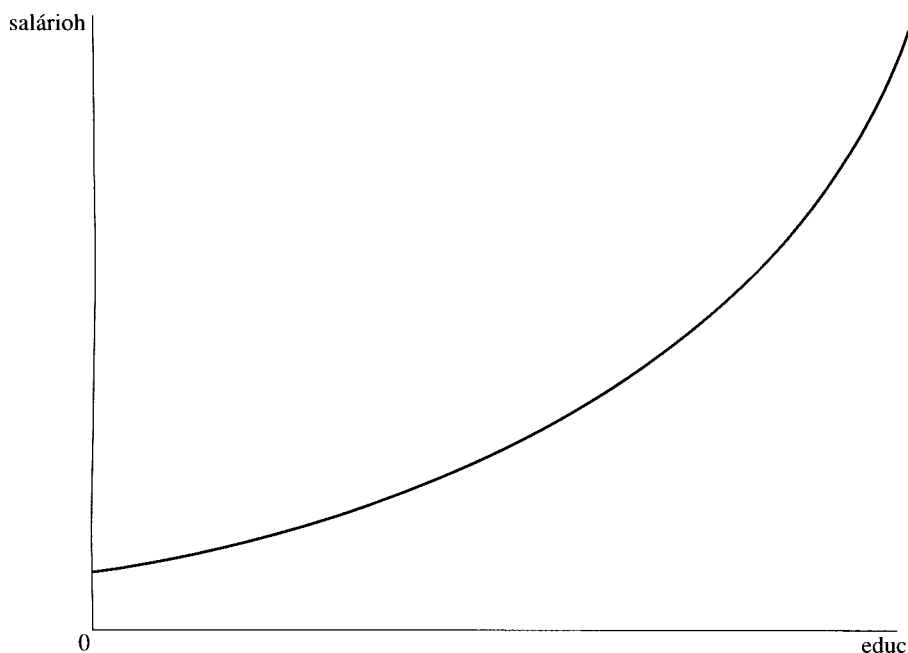
$$\log(\text{salário}_h) = \beta_0 + \beta_1 \text{educ} + u, \quad (2.42)$$

em que $\log(\cdot)$ é o logaritmo *natural*. (Veja Apêndice A, disponível no site da Thomson, para uma revisão sobre logaritmos.) Em particular, se $\Delta u = 0$, então

$$\% \Delta \text{salário}_h \approx (100 \cdot \beta_1) \Delta \text{educ}. \quad (2.43)$$

Figura 2.6

$salárioh = \exp(\beta_0 + \beta_1 educ)$, com $\beta_1 > 0$.



Observe como multiplicamos β_1 por 100 para obter a variação percentual em $salárioh$ dado um ano adicional de educação formal. Como a variação percentual em $salárioh$ é a mesma para cada ano adicional de educação, a variação em $salárioh$, para um ano extra de educação formal, *aumenta* quando a educação formal aumenta; em outras palavras, (2.42) implica um retorno *crecente* da educação formal. Com a exponenciação de (2.42), podemos escrever $salárioh = \exp(\beta_0 + \beta_1 educ + u)$. O gráfico dessa equação aparece na Figura 2.6, com $u = 0$.

Quando se usa a regressão simples, a estimação de um modelo como (2.42) é imediata. Apenas defina a variável dependente, y , como $y = \log(salárioh)$. A variável independente é representada por $x = educ$. A mecânica de MQO é a mesma de antes: as estimativas de intercepto e de inclinação são dadas pelas fórmulas (2.17) e (2.19). Em outras palavras, obtemos $\hat{\beta}_0$ e $\hat{\beta}_1$ da regressão de MQO de $\log(salárioh)$ sobre $educ$.

EXEMPLO 2.10**(Uma Equação do Logaritmo dos Salários-Hora)**

Utilizando os mesmos dados do Exemplo 2.4, mas usando $\log(salárioh)$ como a variável dependente, obtemos a seguinte relação:

$$\log(\hat{salário}) = 0,584 + 0,083 educ \quad (2.44)$$

EXEMPLO 2.10 (continuação)

$$n = 526, R^2 = 0,186.$$

O coeficiente de *educ* tem uma interpretação percentual quando ele é multiplicado por 100: para cada ano adicional de educação formal, *salário_h* aumenta 8,3%. Isso é o que os economistas querem dizer quando se referem ao “retorno de um ano adicional de educação formal”.

É importante lembrar que a principal razão para usar o log de *salário_h* em (2.42) é impor um efeito percentual constante da educação formal sobre *salário_h*. Uma vez obtida a equação (2.42), o log natural de *salário_h* é raramente mencionado. Em particular, não é correto dizer que um ano adicional de educação formal aumenta $\log(\text{salário}_h)$ em 8,3%.

O intercepto em (2.42) não tem muito significado, visto que ele é o $\log(\text{salário}_h)$ previsto quando *educ* = 0. O *R*-quadrado mostra que *educ* explica cerca de 18,6% da variação em $\log(\text{salário}_h)$ (não em *salário*). Finalmente, a equação (2.44) pode não capturar toda a não-linearidade da relação entre salário-hora e escolaridade formal. Se houver “efeitos-diploma”, o décimo segundo ano de educação — formatura do ensino médio nos Estados Unidos — deve ser muito mais valioso que o décimo primeiro ano. No Capítulo 7 aprenderemos como lidar com esse tipo de não-linearidade.

Outro uso importante do log natural está em obter um **modelo de elasticidade constante**.

EXEMPLO 2.11**(Salários de Diretores Executivos e Vendas das Empresas)**

Podemos estimar um modelo de elasticidade constante que relaciona o salário dos diretores executivos às vendas das empresas. O conjunto de dados é o mesmo utilizado no Exemplo 2.3, exceto que agora relacionamos *salário* a *vendas*. Seja *vendas* as vendas anuais das empresas, mensurada em milhões de dólares. Um modelo de elasticidade constante é

$$\log(\hat{\text{salário}}) = \beta_0 + \beta_1 \log(\text{vendas}) + u, \quad (2.45)$$

em que β_1 é a elasticidade de *salário* com respeito a *vendas*. Esse modelo está compreendido no modelo de regressão simples ao se definir a variável dependente como $y = \log(\text{salário})$, e a variável independente como $x = \log(\text{vendas})$. Ao estimar essa equação por MQO, temos

$$\log(\hat{\text{salário}}) = 4,822 + 0,257\log(\text{vendas}) \quad (2.46)$$

$$n = 209, R^2 = 0,211.$$

O coeficiente de $\log(\text{vendas})$ é a elasticidade estimada de *salário* em relação a *vendas*. Ela implica que um aumento de 1% nas vendas das empresas aumenta o salário dos diretores executivos em cerca de 0,257% — a interpretação usual de uma elasticidade.

As duas formas consideradas nesta seção surgirão no restante deste texto. Tratamos aqui de modelos que contêm logaritmos naturais porque eles aparecem muito freqüentemente em trabalhos aplicados. A interpretação desses modelos não será muito diferente no caso da regressão múltipla.

É também útil observar o que acontece às estimativas de intercepto e de inclinação se mudarmos as unidades de medida da variável dependente quando ela aparece na forma logarítmica. Pelo fato de a variação da forma logarítmica aproximar-se de uma variação proporcional, faz sentido que *nada* aconteça com a inclinação. Podemos ver isso ao escrever a variável em uma nova escala como $c_1 y_i$ para cada observação i . A equação original é $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$. Se adicionamos $\log(c_1)$ a ambos os lados da equação, obtemos $\log(c_1) + \log(y_i) = [\log(c_1) + \beta_0] + \beta_1 x_i + u_i$ ou $\log(c_1 y_i) = [\log(c_1) + \beta_0] + \beta_1 x_i + u_i$. (Lembre-se de que a soma dos logs é igual ao log de seus produtos, como mostrado no Apêndice A, disponível no site da Thomson.) Portanto, a inclinação ainda é β_1 , mas o intercepto agora é $\log(c_1) + \beta_0$. Semelhantemente, se a variável independente for $\log(x)$, e mudarmos as unidades de medida de x antes de considerarmos o log, a inclinação permanece a mesma, mas o intercepto muda. Pediremos que você verifique essas asserções no Problema 2.9.

Finalizamos esta subseção resumindo quatro combinações de formas funcionais construídas a partir da variável original ou de seu logaritmo natural. Na Tabela 2.3, x e y representam as variáveis em suas formas originais. O modelo com y como a variável dependente e x como a variável independente é chamado *modelo nível-nível*, pois cada variável aparece em sua forma de nível. O modelo com $\log(y)$ como a variável dependente e x como a variável independente é chamado *modelo log-nível*. Não discutiremos aqui, explicitamente, o modelo *nível-log*, pois ele aparece menos freqüentemente na prática. De qualquer forma, veremos exemplos desse modelo em outros capítulos.

A última coluna na Tabela 2.3 mostra a interpretação de β_1 . No modelo log-nível, $100 \cdot \beta_1$ é algumas vezes chamado **semi-elasticidade** de y em relação a x . Como mencionamos no Exemplo 2.11, no modelo log-log β_1 é a **elasticidade** de y em relação a x . A Tabela 2.3 requer um estudo cuidadoso, já que vamos, com freqüência, nos referir a ela no restante do texto.

O Significado da Regressão “Linear”

O modelo de regressão simples que estudamos neste capítulo também é chamado modelo de regressão *linear simples*. No entanto, como acabamos de ver, o modelo geral também permite certas relações *não-lineares*. Portanto, o que significa “linear” aqui? Você pode observar, ao olhar a equação (2.1), que $y = \beta_0 + \beta_1 x + u$. O importante é que essa equação é linear nos *parâmetros*, β_0 e β_1 . Não há restrições de como y e x se relacionam com as variáveis explicada e explicativa originais de interesse. Como vimos nos Exemplos 2.7 e 2.8, y e x podem ser os logaritmos naturais de variáveis, e isso é muito comum em aplicações. Mas não precisamos parar aqui. Por exemplo, nada nos impede de usar a regressão simples para estimar um modelo tal como $cons = \beta_0 + \beta_1 \sqrt{rend} + u$, em que $cons$ é o consumo anual e $rend$ é a renda anual.

Tabela 2.3

Resumo das Formas Funcionais Envolvendo Logaritmos

<i>Modelo</i>	<i>Variável Dependente</i>	<i>Variável Independente</i>	<i>Interpretação de β_1</i>
nível-nível	y	x	$\Delta y = \beta_1 \Delta x$
nível-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\% \Delta x$
log-nível	$\log(y)$	x	$\% \Delta y = (100\beta_1)\Delta x$
log-log	$\log(y)$	$\log(x)$	$\% \Delta y = \beta_1 \% \Delta x$

Embora a mecânica da regressão simples não dependa de como y e x são definidos, a interpretação dos coeficientes depende, realmente, de suas definições. Para que os trabalhos empíricos sejam bem-sucedidos, é muito mais importante tornar-se proficiente em interpretar coeficientes do que eficiente no cálculo de fórmulas como (2.19). Obteremos muito mais prática em como interpretar as estimativas da reta de regressão de MQO quando estudarmos a regressão múltipla.

Muitos modelos *não podem* ser considerados modelos de regressão linear, porque eles não são lineares em seus parâmetros; um exemplo é $\text{cons} = 1/(\beta_0 + \beta_1 \text{rend}) + u$. A estimação desses modelos leva-nos ao campo âmbito do *modelo de regressão não-linear*, o qual está além do escopo deste texto. Para muitas aplicações, é suficiente escolher um modelo que possa ser expresso dentro do arcabouço da regressão linear.

2.5 VALORES ESPERADOS E VARIÂNCIAS DOS ESTIMADORES DE MQO

Na Seção 2.1 definimos o modelo populacional $y = \beta_0 + \beta_1 x + u$ e afirmamos que a hipótese fundamental para que a análise de regressão simples seja útil é que o valor esperado de u , dado qualquer valor de x , seja zero. Nas seções 2.2, 2.3 e 2.4 discutimos as propriedades algébricas da estimação de MQO. Retornamos agora ao modelo populacional e estudaremos as propriedades *estatísticas* da estimação de MQO. Em outras palavras, veremos agora $\hat{\beta}_0$ e $\hat{\beta}_1$ como *estimadores* dos parâmetros β_0 e β_1 que aparecem no modelo populacional. Isso significa que estudaremos as propriedades das distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$ de diferentes amostras aleatórias da população. (O Apêndice C, disponível no site da Thomson, contém as definições de estimadores e revisões de algumas de suas principais propriedades.)

Inexistência de Viés em MQO

Vamos iniciar estabelecendo a inexistência de viés do método MQO sob um conjunto simples de hipóteses. Para referências futuras, é útil numerar essas hipóteses usando o prefixo “RLS” para regressão linear simples. A primeira hipótese define o modelo populacional.

HIPÓTESE RLS.1 (LINEAR NOS PARÂMETROS)

No modelo populacional, a variável dependente y está relacionada à variável independente x e ao erro (ou perturbação) u como

$$y = \beta_0 + \beta_1 x + u, \quad (2.47)$$

em que β_0 e β_1 são os parâmetros de intercepto e de inclinação populacionais, respectivamente.

Ao especificar o modelo populacional — e para ser realista —, y , x e u são todos vistos como variáveis aleatórias. Discutimos, em alguma extensão, a interpretação desse modelo na Seção 2.1 e demos vários exemplos. Na seção anterior aprendemos que a equação (2.47) não é tão restritiva quanto inicialmente parecia; escolhendo y e x apropriadamente, podemos obter relações não-lineares interessantes (como os modelos de elasticidade constante).

Estamos interessados em usar os dados de y e x para estimar os parâmetros β_0 e, especialmente, β_1 . Assumimos que nossos dados foram obtidos de uma amostra aleatória. (Veja o Apêndice C, disponível no site da Thomson, para uma revisão sobre amostragem aleatória.)

H I P Ó T E S E R L S . 2 (AMOSTRAGEM ALEATÓRIA)

Podemos usar uma amostra aleatória de tamanho n , $\{(x_i, y_i): i = 1, 2, \dots, n\}$, proveniente de um modelo populacional.

Em capítulos posteriores que abordam a análise de séries de tempo e problemas de seleção de amostra, teremos de dar um tratamento ao fato de a hipótese de amostragem aleatória não ser mais válida. Nem todas as amostras de corte transversal podem ser vistas como resultados de amostras aleatórias, mas muitas podem ser assim entendidas.

Podemos escrever (2.47), em termos da amostra aleatória como

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \dots, n, \quad (2.48)$$

em que u_i é o erro ou perturbação da observação i (por exemplo, pessoa i , empresa i , cidade i etc.). Assim, u_i contém os fatores não-observáveis da observação i que afetam y_i . Os u_i não devem ser confundidos com os resíduos, \hat{u}_i , definidos na Seção 2.3. Mais adiante, exploraremos a relação entre os erros e os resíduos. Para interpretar β_0 e β_1 em uma aplicação particular, (2.47) é mais instrutiva, mas (2.48) também é necessária para algumas derivações estatísticas.

A relação (2.48) pode ser colocada em um gráfico para um registro particular dos dados, como mostrado na Figura 2.7.

A fim de obter estimadores não-viesados de β_0 e β_1 , precisamos impor a hipótese de média condicional zero que discutimos, com algum detalhe, na Seção 2.1. Agora, vamos adicioná-la explicitamente à nossa lista de hipóteses.

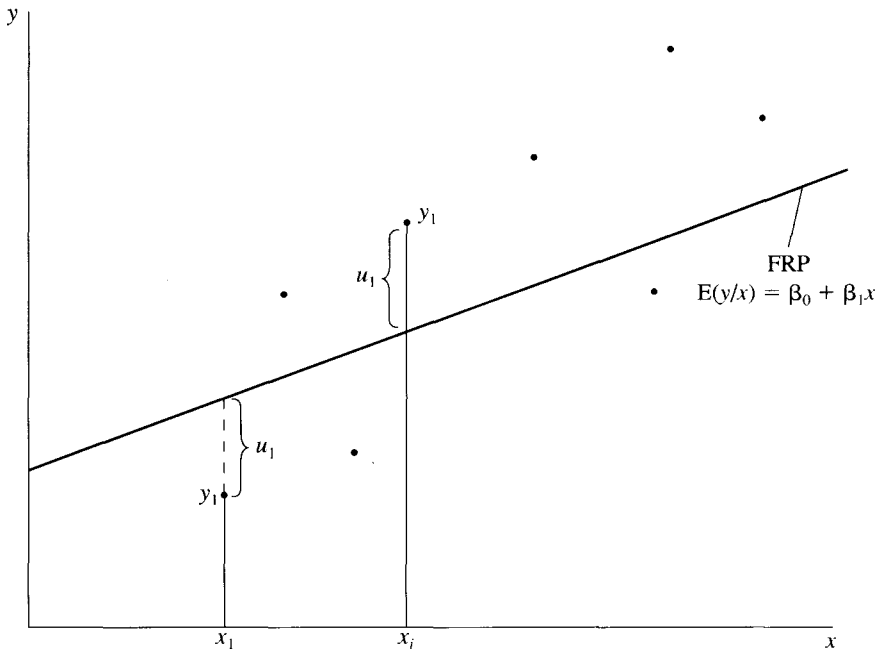
H I P Ó T E S E R L S . 3 (MÉDIA CONDICIONAL ZERO)

$$E(u|x) = 0.$$

Para uma amostra aleatória, essa hipótese implica que $E(u_i|x_i) = 0$, para todo $i = 1, 2, \dots, n$.

Além de restringir a relação entre u e x na população, a hipótese de média condicional zero — juntamente com a hipótese de amostra aleatória — permite uma simplificação técnica conveniente. Em particular, podemos derivar as propriedades estatísticas dos estimadores de MQO como *condicionais* aos valores de x_i em nossa amostra. Tecnicamente, em derivações estatísticas, condicionar aos valores amostrais da variável independente é o mesmo que tratar x_i como *fixo em amostras repetidas*. Esse processo envolve vários passos. Primeiro, escolhemos n valores amostrais para x_1, x_2, \dots, x_n . (Esses valores podem ser repetidos.) Dados esses valores, obtemos uma amostra de y (efetivamente, obtendo uma amostra aleatória de u_i). Em seguida, obtém-se outra amostra de y , usando novamente os mesmos x_1, x_2, \dots, x_n . E assim por diante.

Figura 2.7

Gráfico de $y_i = \beta_0 + \beta_1 x_i + u_i$.

O fixo no cenário de amostras repetidas não é muito realista no contexto não-experimental. Por exemplo, na amostragem de indivíduos do exemplo salários-educação, faz pouco sentido pensar em escolher os valores de *educ* antecipadamente e, em seguida, fazer uma amostra de indivíduos com aqueles níveis particulares de educação formal. A amostragem aleatória, na qual os indivíduos são escolhidos aleatoriamente e seus salários e anos de educação formal são registrados, é um processo representativo de como muitos conjuntos de dados são obtidos para a análise empírica nas ciências sociais. Já que *assumimos* $E(u|x) = 0$, e temos amostragem aleatória, nada se perde nas derivações ao tratar os x_i como não-aleatórios. O perigo é que o fixo na hipótese de amostras repetidas *sempre* implica que u_i e x_i são independentes. Ao decidir quando a análise de regressão simples produzirá estimadores não-viesados, é crucial pensar em termos da hipótese RLS.3.

Visto que concordamos em condicionar as derivações estatísticas aos valores de x_i , precisamos de uma última hipótese para a inexistência de vies.

HIPÓTESE RLS.4 (VARIÇÃO AMOSTRAL NA VARIÁVEL INDEPENDENTE)

Na amostra, as variáveis independentes x_i , $i = 1, 2, \dots, n$, não são todas iguais a uma mesma constante. Isso exige alguma variação em x na população.

Encontramos a hipótese RLS.4 quando derivamos as fórmulas dos estimadores de MQO; ela é equivalente a $\sum_{i=1}^n (x_i - \bar{x})^2 > 0$. Das quatro hipóteses feitas, esta é a menos importante, pois ela essen-

cialmente nunca falha em aplicações interessantes. Se a hipótese RLS.4 não se sustentar, não podemos calcular os estimadores de MQO, o que significa que a análise estatística é irrelevante.

Usando o fato de que $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$ (veja o Apêndice A, no site da Thomson), podemos escrever o estimador de inclinação de MQO na equação (2.19) como

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.49)$$

Como agora estamos interessados no comportamento de $\hat{\beta}_1$ ao longo de todas as amostras possíveis, $\hat{\beta}_1$ é apropriadamente visto como uma variável aleatória.

Podemos escrever $\hat{\beta}_1$ em termos dos coeficientes populacionais e dos erros ao substituir o lado direito de (2.48) em (2.49). Temos

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\text{SQT}_x} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\text{SQT}_x}, \quad (2.50)$$

onde definimos a variação total em x_i como $\text{SQT}_x = \sum_{i=1}^n (x_i - \bar{x})^2$, a fim de simplificar a notação. (Essa expressão não é exatamente a variância amostral de x_i , pois não a dividimos por $n - 1$.) Usando a álgebra do operador somatório, vamos escrever o numerador de $\hat{\beta}_1$ como

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})\beta_0 + \sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i + \sum_{i=1}^n (x_i - \bar{x})u_i \\ = \beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i + \sum_{i=1}^n (x_i - \bar{x})u_i. \end{aligned} \quad (2.51)$$

Como mostrado no Apêndice A, $\sum_{i=1}^n (x_i - \bar{x}) = 0$ e $\sum_{i=1}^n (x_i - \bar{x})x_i = \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2$. Portanto, podemos escrever o numerador de $\hat{\beta}_1$ como $\beta_1 \text{SQT}_x + \sum_{i=1}^n (x_i - \bar{x})u_i$. Escrivendo isso no denominador resulta

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\text{SQT}_x} = \beta_1 + (1/\text{SQT}_x) \sum_{i=1}^n d_i u_i, \quad (2.52)$$

onde $d_i = x_i - \bar{x}$. Vemos agora que o estimador $\hat{\beta}_1$ é igual à inclinação populacional β_1 mais um termo que é uma combinação linear dos erros $\{u_1, u_2, \dots, u_n\}$. Condicionada aos valores de x_i , a aleatoriedade em $\hat{\beta}_1$ deve-se inteiramente aos erros na amostra. O fato de que esses erros sejam, em geral, diferentes de zero é o que faz com que $\hat{\beta}_1$ seja diferente de β_1 .

Ao usar a representação em (2.52), podemos provar a primeira importante propriedade estatística do método MQO.

TEOREMA 2.1 (INEXISTÊNCIA DE VIÉS EM MQO)

Usando as hipóteses RLS.1 a RLS.4,

$$E(\hat{\beta}_0) = \beta_0 \text{ e } E(\hat{\beta}_1) = \beta_1, \quad (2.53)$$

para quaisquer valores de β_0 e β_1 . Em outras palavras, $\hat{\beta}_0$ é não-viesado para β_0 , e $\hat{\beta}_1$ é não-viesado para β_1 .

PROVA: Nesta prova, os valores esperados estão condicionados aos valores amostrais da variável independente. Visto que SQT_x e d_i são funções somente de x_i , eles são não-aleatórios quando condicionais. Portanto, de (2.52), e mantendo o condicionamento a $\{x_1, x_2, \dots, x_n\}$ implícito, temos

$$\begin{aligned} E(\hat{\beta}_1) &= \beta_1 + E[(1/SQT_x) \sum_{i=1}^n d_i u_i] = \beta_1 + (1/SQT_x) \sum_{i=1}^n E(d_i u_i) \\ &= \beta_1 + (1/SQT_x) \sum_{i=1}^n d_i E(u_i) = \beta_1 + (1/SQT_x) \sum_{i=1}^n d_i \cdot 0 = \beta_1, \end{aligned}$$

onde usamos o fato de que o valor esperado de cada u_i (condicional a $\{x_1, x_2, \dots, x_n\}$) é zero sob as hipóteses RLS.2 e RLS.3. Como a inexistência de viés se mantém para qualquer resultado condicionado a $\{x_1, x_2, \dots, x_n\}$, a inexistência de viés também se mantém sem se condicionar a $\{x_1, x_2, \dots, x_n\}$.

A prova para $\hat{\beta}_0$ é agora direta. Obtenha a média de (2.48) através de i para obter $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}$, e insira essa equação na fórmula de $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x} = \beta_0 + (\beta_1 - \hat{\beta}_1) \bar{x} + \bar{u}.$$

Então, condicional aos valores de x_i ,

$$E(\hat{\beta}_0) = \beta_0 + E[(\beta_1 - \hat{\beta}_1) \bar{x}] + E(\bar{u}) = \beta_0 + E[(\beta_1 - \hat{\beta}_1)] \bar{x},$$

já que, pelas hipóteses RLS.2 e RLS.3, $E(\bar{u}) = 0$. No entanto, mostramos que $E(\hat{\beta}_1) = \beta_1$, o que implica que $E[(\hat{\beta}_1 - \beta_1)] = 0$. Assim, $E(\hat{\beta}_0) = \beta_0$. Ambos os argumentos são válidos para quaisquer valores de β_0 e β_1 , e assim estabelecemos a inexistência de viés.

Lembre-se de que a inexistência de viés é uma característica das distribuições amostrais de $\hat{\beta}_1$ e $\hat{\beta}_0$, o que não nos diz nada sobre a estimativa que obtemos de uma dada amostra. Esperamos que, se a amostra que obtemos é de algum modo “típica”, então nossa estimativa deve estar “próxima” do valor populacional. Infelizmente, é sempre possível obter uma amostra ruim que nos dê uma estimativa pontual distante de β_1 , e nós *nunca* saberemos, com certeza, se esse é o caso. Neste ponto, você pode querer revisar o material sobre estimadores não-viesados no Apêndice C (disponível no site da Thomson), especialmente o exercício de simulação da Tabela C.1, o qual ilustra o conceito de inexistência de viés.

Em geral, a inexistência de viés não é válida se qualquer uma das nossas quatro hipóteses não for válida. Isso significa que é importante pensar na veracidade de cada hipótese em uma aplicação particular. Como já discutimos, se a hipótese RLS.4 não for válida, então não seremos capazes de obter as estimativas de MQO. A hipótese RLS.1 requer que y e x sejam linearmente relacionados com uma perturbação adicionada. Certamente, isso pode não ser válido. Mas sabemos também que y e x podem ser escolhidos de forma que possam produzir relações não-lineares interessantes. Estudar com a não validade de (2.47) requer métodos mais avançados, que estão além do escopo deste texto.

Posteriormente, veremos como relaxar a hipótese RLS.2, a hipótese de amostragem aleatória, na análise de séries de tempo. Porém, o que dizer de seu uso na análise de corte transversal? A amostragem aleatória pode não ser válida em um corte transversal quando as amostras não são representativas da população subjacente; de fato, alguns conjuntos de dados são construídos fazendo-se, intencionalmente, amostras de partes diferentes da população. Discutiremos os problemas de amostragem não-aleatória nos Capítulos 9 e 17.

A hipótese na qual devemos nos concentrar agora é a RLS.3. Se RLS.3 se mantém, as estimativas de MQO são não-viesadas. Do mesmo modo, se RLS.3 não se mantém, os estimadores de MQO serão, em geral, viesados. Há maneiras de determinar a direção e o tamanho prováveis do viés, algo que estudaremos no Capítulo 3.

A possibilidade de que x seja correlacionado com u é quase sempre uma preocupação na análise de regressão simples com dados não-experimentais, como indicamos por meio de vários exemplos na Seção 2.1. Usar a regressão simples quando u contém fatores que afetam y e que também estão correlacionados com x pode resultar em *correlação espúria*: isto é, achamos uma relação entre y e x que se deve, em verdade, a outros fatores que afetam y e que também estão correlacionados com x .

EXEMPLO 2.12

(Desempenho em Matemática de Estudantes e o Programa de Merenda Escolar)

Seja *mate10* a porcentagem de alunos do primeiro ano do ensino médio aprovados em um exame de matemática. Suponha que desejamos estimar o efeito do programa de merenda escolar financiado pelo governo sobre o desempenho dos alunos. Esperamos que o programa de merenda tenha um efeito *ceteris paribus* positivo sobre o desempenho: todos os outros fatores permanecendo iguais, se um estudante, bastante pobre para ter regularmente refeições, torna-se qualificado para o programa de merenda escolar, seu desempenho deveria melhorar. Seja *prgalm* a porcentagem de estudantes que estão aptos para participar do programa de merenda escolar. Portanto, o modelo de regressão simples é

$$\text{mate10} = \beta_0 + \beta_1 \text{prgalm} + u, \quad (2.54)$$

em que u contém características da escola e do estudante que afetam o desempenho escolar total. Usando os dados do arquivo de MEAP93.RAW de 408 escolas de Michigan no ano escolar 1992-1993, obtemos

EXEMPLO 2.12 (continuação)

$$\widehat{maie10} = 32,14 - 0,319 \text{ prgalm}$$

$$n = 408, R^2 = 0,171.$$

Essa equação prevê que se a participação dos estudantes no programa de merenda escolar aumenta em dez pontos percentuais, a percentagem de estudantes que passa no exame de matemática *cai* cerca de 3,2 pontos percentuais. Realmente devemos acreditar que a participação maior no programa de merenda escolar causa, *de fato*, um desempenho pior? Muito provavelmente não. Uma explicação melhor é que o termo erro u na Equação (2.54) está correlacionado com prgalm . De fato, u contém fatores como a taxa de pobreza das crianças que freqüentam a escola, que afeta o desempenho dos estudantes e está altamente correlacionada com a qualificação no programa de merenda. Variáveis como qualidade e recursos da escola também estão contidas em u e, provavelmente, estão correlacionados com prgalm . É importante lembrar que a estimativa $-0,319$ é somente para essa amostra particular, mas seu sinal e magnitude nos fazem suspeitar de que u e x sejam correlacionadas, de modo que a regressão linear é viesada.

Além de variáveis omitidas, há outras razões para que x esteja correlacionado com u no modelo de regressão simples. Como essas mesmas questões surgirão na análise de regressão múltipla, postergaremos até lá um tratamento sistemático do problema.

Variâncias dos Estimadores de MQO

Além de saber que a distribuição amostral de $\hat{\beta}_1$ está centrada em torno de β_1 ($\hat{\beta}_1$ é não-viesado), é importante saber o quão distante, em média, podemos esperar que $\hat{\beta}_1$ esteja de β_1 . Entre outras coisas, isso nos permite escolher o melhor estimador entre todos os estimadores não-viesados — ou pelo menos entre uma ampla classe deles. A medida de dispersão da distribuição de $\hat{\beta}_1$ (e $\hat{\beta}_0$) com a qual é mais fácil trabalhar é a variância, ou sua raiz quadrada, o desvio-padrão. (Veja o Apêndice C, disponível no site da Thomson, para uma discussão mais detalhada.)

A variância dos estimadores de MQO pode ser calculada sob as hipóteses RLS.1 a RLS.4. Entretanto, as expressões dessas variâncias são complexas. Em vez disso, vamos adicionar uma hipótese que é tradicional na análise de corte transversal. Essa hipótese afirma que a variância do termo não-observável, u , condicionado a x , é constante. Ela é conhecida como a hipótese de **homoscedasticidade** ou de “variância constante”.

H I P Ó T E S E R L S . 5 (HOMOSCEDASTICIDADE)

$$\text{Var}(u|x) = \sigma^2.$$

Devemos enfatizar que a hipótese de homoscedasticidade é completamente distinta da hipótese de média condicional zero, $E(u|x) = 0$. A hipótese RLS.3 compreende o *valor esperado* de u , enquanto a hipótese RLS.5 diz respeito à *variância* de u (ambos condicionados a x). Lembre-se de que nós estabelecemos a inexistência de viés de MQO sem a hipótese RLS.5: a hipótese de homoscedasticidade não desempenha *qualquer* papel para mostrar que $\hat{\beta}_0$ e $\hat{\beta}_1$ são não-viesados. Adicionamos a hipótese

RLS.5 pois ela simplifica os cálculos da variância de $\hat{\beta}_0$ e $\hat{\beta}_1$ e porque ela implica que o método de mínimos quadrados ordinários tenha certas propriedades de eficiência, algo que veremos no Capítulo 3. Se assumíssemos que u e x são *independentes*, a distribuição de u , dado x , não dependeria de x , e assim $E(u|x) = E(u) = 0$ e $\text{Var}(u|x) = \sigma^2$. No entanto, algumas vezes independência é, uma hipótese forte demais.

Como $\text{Var}(u|x) = E(u^2|x) - [E(u|x)]^2$ e $E(u|x) = 0$, $\sigma^2 = E(u^2|x)$, o que significa que σ^2 também é a esperança *não-condicional* de u^2 . Portanto, $\sigma^2 = E(u^2) = \text{Var}(u)$, pois $E(u) = 0$. Em outras palavras, σ^2 é a variância *não-condicional* de u , e por isso σ^2 é freqüentemente chamado de **variância do erro** ou variância da perturbação. A raiz quadrada de σ^2 , σ , é o desvio-padrão do erro. Um σ grande significa que a distribuição dos fatores não-observáveis que afetam y é mais dispersa.

Freqüentemente, é útil escrever as hipóteses RLS.3 e RLS.5 em termos da média condicional e da variância condicional de y :

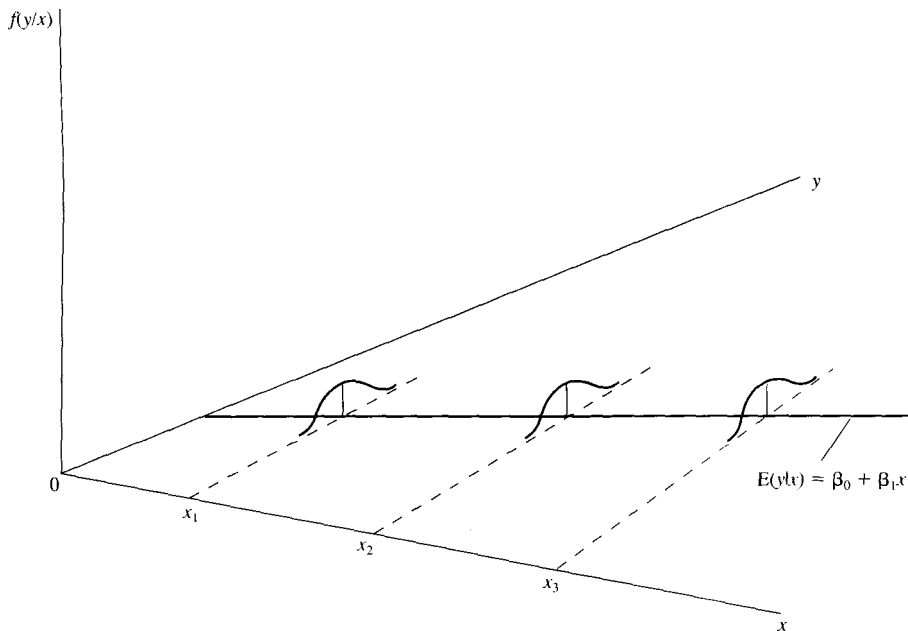
$$E(y|x) = \beta_0 + \beta_1 x. \quad (2.55)$$

$$\text{Var}(y|x) = \sigma^2. \quad (2.56)$$

Em outras palavras, a esperança condicional de y , dado x , é linear em x , mas a variância de y , dado x , é constante. Essa situação está ilustrada na Figura 2.8, em que $\beta_0 > 0$ e $\beta_1 > 0$.

Quando $\text{Var}(u|x)$ depende de x , diz-se que o termo erro apresenta **heteroscedasticidade** (ou variância não-constante). Como $\text{Var}(u|x) = \text{Var}(y|x)$, a heteroscedasticidade está presente sempre que $\text{Var}(y|x)$ é uma função de x .

Figura 2.8
O modelo de regressão simples sob homoscedasticidade.

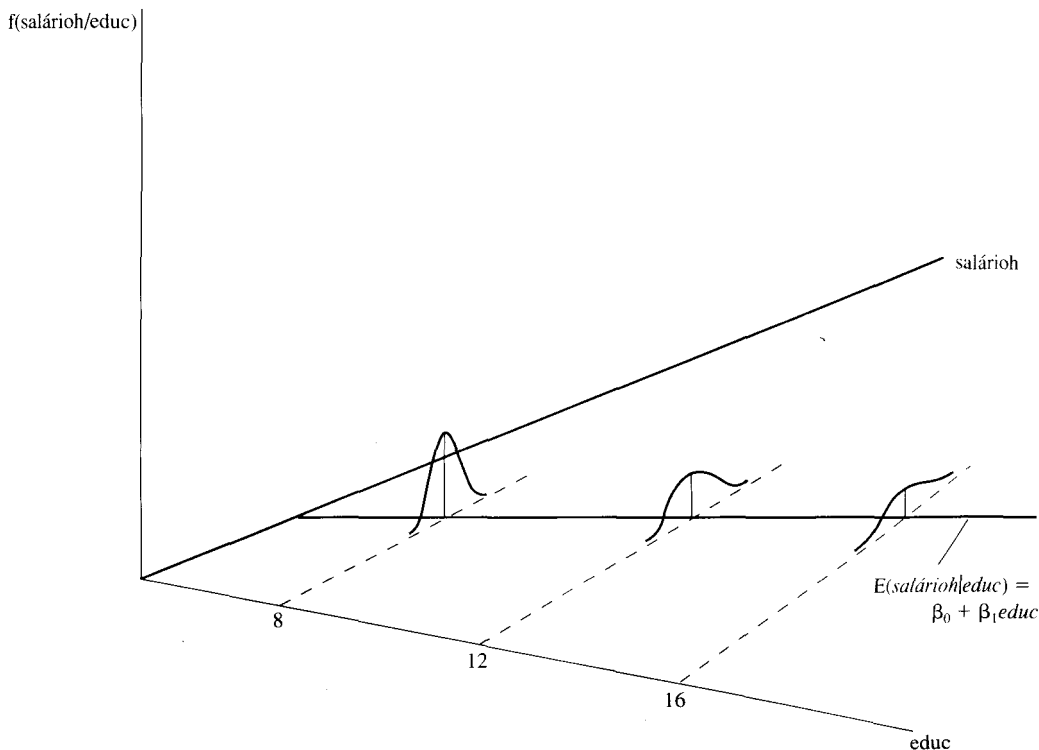


EXEMPLO 2.13**(Heteroscedasticidade em uma Equação de Salários)**

A fim de obter um estimador não-viesado do efeito *ceteris paribus* de *educ* sobre *salárioh*, devemos assumir que $E(u|educ) = 0$, e isso implica $E(\text{salárioh}|educ) = \beta_0 + \beta_1 \text{educ}$. Se também usarmos a hipótese de homoscedasticidade, então $\text{Var}(u|educ) = \sigma^2$ não depende do nível de educação formal, que é o mesmo que assumir que $\text{Var}(\text{salárioh}|educ) = \sigma^2$. Assim, enquanto se deixa o salário-hora médio aumentar com o nível de educação formal – é essa taxa de crescimento que estamos interessados em descrever – assume-se que a *variabilidade* no salário horário em torno de sua média é constante através de todos os níveis de educação formal: isso pode não ser realista. É provável que pessoas com mais tempo de educação formal tenham uma variedade maior de interesses e de oportunidades de trabalho, o que poderia levar a uma variabilidade maior nos níveis de educação formal mais elevados. Pessoas com níveis de educação formal bastante baixos têm muito poucas oportunidades e, freqüentemente, precisam trabalhar recebendo salário mínimo; isso tem o efeito de reduzir a variabilidade salarial nos níveis baixos de educação formal. Essa situação é mostrada na Figura 2.9. Em última análise, se a hipótese RLS.5 se mantém ou não é uma questão empírica. No Capítulo 8 mostraremos como testar a hipótese RLS.5.

Figura 2.9

Var (salárioh/educ) crescendo com educ.



Com a hipótese apropriada de homoscedasticidade, estamos prontos para provar o seguinte:

TEOREMA 2.2 (VARIÂNCIAS AMOSTRAIS EM MQO)

Sob as hipóteses RLS.1 a RLS.5,

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sigma^2/\text{SQT}_x \quad (2.57)$$

e

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.58)$$

as quais estão condicionadas aos valores amostrais $\{x_1, \dots, x_n\}$.

PROVA: Derivaremos a fórmula para $\text{Var}(\hat{\beta}_1)$, deixando a outra derivação como exercício. O ponto de partida é a equação (2.52): $\hat{\beta}_1 = \beta_1 + (1/\text{SQT}_x) \sum_{i=1}^n d_i u_i$. Visto que β_1 é exatamente uma constante, condicional aos x_i , SQT_x e $d_i = x_i - \bar{x}$ também são não-aleatórios. Além disso, como os u_i são variáveis aleatórias independentes para todos os i (por amostragem aleatória), a variância da soma é a soma das variâncias. Usando esses fatos, temos

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= (1/\text{SQT}_x)^2 \text{Var}\left(\sum_{i=1}^n d_i u_i\right) = (1/\text{SQT}_x)^2 \left(\sum_{i=1}^n d_i^2 \text{Var}(u_i)\right) \\ &= (1/\text{SQT}_x)^2 \left(\sum_{i=1}^n d_i^2 \sigma^2\right) \quad [\text{visto que } \text{Var}(u_i) = \sigma^2 \text{ para todos os } i] \\ &= \sigma^2 (1/\text{SQT}_x)^2 \left(\sum_{i=1}^n d_i^2\right) = \sigma^2 (1/\text{SQT}_x)^2 \text{SQT}_x = \sigma^2/\text{SQT}_x \end{aligned}$$

que é o que queríamos provar.

As equações (2.57) e (2.58) são as fórmulas “padrões” para a análise de regressão simples, mas não são válidas na presença de heteroscedasticidade. Isso será importante quando tratarmos dos intervalos de confiança e dos testes de hipóteses na análise de regressão múltipla.

Para a maior parte dos propósitos, estamos interessados em $\text{Var}(\hat{\beta}_1)$. É fácil resumir como essa variância depende da variância do erro, σ^2 , e da variação total em $\{x_1, x_2, \dots, x_n\}$, SQT_x . Primeiro, quanto maior a variância do erro, maior é $\text{Var}(\hat{\beta}_1)$. Isso faz sentido, já que uma variação maior nos fatores

não-observáveis que afetam y faz com que seja mais difícil estimar com precisão β_1 . Em contrapartida, é preferível maior variabilidade na variável independente; quando a variabilidade nos x_i aumenta, a variância de $\hat{\beta}_1$ diminui. Isso também tem um sentido intuitivo, visto que quanto mais dispersa for a amostra das variáveis independentes, mais fácil será descrever a relação entre $E(y|x)$ e x . Isto é, será mais fácil estimar β_1 . Se há pouca variação nos x_i , pode ser difícil estabelecer com precisão como $E(y|x)$ varia com x . Quando o tamanho da amostra cresce, do mesmo modo cresce a variação total nos x_i . Portanto, um tamanho de amostra maior resulta em uma variância menor de $\hat{\beta}_1$.

Essa análise mostra que, se estamos interessados em $\hat{\beta}_1$, e temos uma escolha, então devemos escolher os x_i tão dispersos quanto possível. Às vezes, isso é possível com dados experimentais, mas raramente temos esse luxo nas ciências sociais: usualmente, temos de pegar os x_i que obtemos via amostragem aleatória. Algumas vezes, temos uma possibilidade de obter amostras maiores, embora isso possa ser dispendioso.

Mostre que, ao estimar β_0 , é melhor ter $\bar{x} = 0$. Qual é a $\text{Var}(\hat{\beta}_0)$ nesse caso? [Sugestão: para qualquer amostra de números, $\sum_{i=1}^n x_i^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$, mantendo-se a igualdade somente se $\bar{x} = 0$.]

Para o propósito de construir intervalos de confiança e derivar estatísticas de testes, precisaremos trabalhar com os desvios padrão de $\hat{\beta}_1$ e $\hat{\beta}_0$, $dp(\hat{\beta}_1)$ e $dp(\hat{\beta}_0)$. Lembre-se de que eles são obtidos ao calcular as raízes quadradas das variâncias em (2.57) e (2.58). Particularmente, $dp(\hat{\beta}_1) = \sigma/\sqrt{SQT_x}$, em que σ é a raiz quadrada de σ^2 e $\sqrt{SQT_x}$ é a raiz quadrada de SQT_x .

Estimação da Variância do Erro

As fórmulas em (2.57) e (2.58) permitem-nos isolar os fatores que contribuem para $\text{Var}(\hat{\beta}_1)$ e $\text{Var}(\hat{\beta}_0)$. No entanto, essas fórmulas são desconhecidas, exceto no caso extremamente raro em que σ^2 é conhecido. Não obstante, podemos usar os dados para estimar σ^2 , o qual conseqüentemente nos permite estimar $\text{Var}(\hat{\beta}_1)$ e $\text{Var}(\hat{\beta}_0)$.

Este é um bom momento para enfatizar a diferença entre os *erros* (ou perturbações) e os *resíduos*, já que essa discussão é crucial para a construção de um estimador de σ^2 . A equação (2.48) mostra como escrever o modelo populacional em termos de uma observação extraída aleatoriamente como $y_i = \beta_0 + \beta_1 x_i + u_i$, em que u_i é o erro da observação i . Podemos expressar também y_i em termos de seu valor estimado e do resíduo, como na equação (2.32): $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$. Comparando essas duas equações, vemos que o erro aparece na equação que contém os parâmetros *populacionais*, β_0 e β_1 . Em contrapartida, os resíduos aparecem na equação *estimada* com $\hat{\beta}_0$ e $\hat{\beta}_1$. Os erros nunca são observáveis, enquanto os resíduos são calculados a partir dos dados.

Podemos usar as equações (2.32) e (2.48) para escrever os resíduos como uma função dos erros:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = (\beta_0 - \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i,$$

ou

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i,$$

(2.59)

Embora o valor esperado de $\hat{\beta}_0$ iguale-se a β_0 , e similarmente para $\hat{\beta}_1$, \hat{u}_i não é o mesmo que u_i . A diferença entre eles tem, de fato, um *valor esperado* igual a zero.

Agora que entendemos a diferença entre os erros e os resíduos, podemos retornar para estimar σ^2 . Primeiro, $\sigma^2 = E(u^2)$, de modo que um “estimador” não-viesado de σ^2 é $n^{-1} \sum_{i=1}^n u_i^2$. Infelizmente, esse não é um estimador verdadeiro, pois não observamos os erros u_i . Mas, temos, de fato, as estimativas de u_i , a saber, os resíduos \hat{u}_i de MQO. Se substituirmos os erros pelos resíduos de MQO, temos $n^{-1} \sum_{i=1}^n \hat{u}_i^2 = \text{SQR}/n$. Esse é um estimador verdadeiro, porque ele fornece uma regra computável para qualquer amostra de dados sobre x e y . Uma ligeira desvantagem desse estimador é que ele resulta viesado (embora o viés seja pequeno para n grande). Como é fácil calcular um estimador não-viesado, usamos esse como substituto.

O estimador SQR/n é viesado, essencialmente, porque ele não explica a razão de duas restrições que devem ser satisfeitas pelos resíduos de MQO. Essas restrições são dadas pelas duas condições de primeira ordem de MQO:

$$\sum_{i=1}^n \hat{u}_i = 0, \quad \sum_{i=1}^n x_i \cdot \hat{u}_i = 0. \quad (2.60)$$

Uma maneira de ver essas restrições é esta: se nós conhecemos $n - 2$ dos resíduos, podemos sempre obter os outros dois resíduos usando as restrições implicadas pelas condições de primeira ordem em (2.60). Assim, há somente $n - 2$ **graus de liberdade** nos resíduos de MQO, em oposição a n graus de liberdade nos erros. Se substituíssemos \hat{u}_i por u_i em (2.60), as restrições não mais se manteriam. O estimador não-viesado de σ^2 que utilizaremos faz um ajustamento dos graus de liberdade:

$$\hat{\sigma}^2 = \frac{1}{(n-2)} \sum_{i=1}^n \hat{u}_i^2 = \text{SQR}/(n-2). \quad (2.61)$$

(Esse estimador é, às vezes, denominado por s^2 , mas continuaremos a usar a convenção de colocar “chapéus” sobre os estimadores.)

TEOREMA 2.3 (ESTIMAÇÃO NÃO-VIESADA DE σ^2)

Sob as hipóteses RLS.1 a RLS.5,

$$E(\hat{\sigma}^2) = \sigma^2.$$

PROVA: Se construirmos a média da equação (2.59) para todos os i e usarmos o fato de que os resíduos de MQO têm média igual a zero, temos $0 = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) \bar{x}$; subtraindo essa equação de (2.59), resulta $\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$. Portanto, $\hat{u}_i^2 = (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2(x_i - \bar{x})^2 - 2(u_i - \bar{u})(\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$. A soma ao longo de todos os i resulta na equação $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^n u_i(x_i - \bar{x})$. Agora, o valor esperado do primeiro termo é $(n-1)\sigma^2$, algo que está mostrado no Apêndice C, no site da Thomson.

T E O R E M A 2 . 3 (ESTIMAÇÃO NÃO-VIESADA DE σ^2) (continuação)

O valor esperado do segundo termo é, simplesmente, σ^2 , porque $E[(\hat{\beta}_1 - \beta_1)^2] = \text{Var}(\hat{\beta}_1) = \sigma^2/s_x^2$. Finalmente, o terceiro termo pode ser escrito como: $2(\hat{\beta}_1 - \beta_1)^2 s_x^2$; aplicando as esperanças, resulta em $2\sigma^2$. Colocando esses três termos juntos, obtemos $E\left(\sum_{i=1}^n \hat{u}_i^2\right) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2$, de modo que $E[\text{SQR}/(n-2)] = \sigma^2$.

Se $\hat{\sigma}^2$ for inserido nas fórmulas da variância (2.57) e (2.58), então teremos estimadores não-viesados de $\text{Var}(\hat{\beta}_1)$ e $\text{Var}(\hat{\beta}_0)$. Posteriormente, necessitaremos de estimadores dos desvios padrão de $\hat{\beta}_1$ e $\hat{\beta}_0$, e isso requer estimar σ . O estimador natural de σ é

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \quad (2.62)$$

e é chamado **erro-padrão da regressão (EPR)**. (Outros nomes para $\hat{\sigma}$ são *erro-padrão da estimativa* e *raiz do erro quadrado médio*, mas não os usaremos.) Ainda que $\hat{\sigma}$ não seja um estimador não-viesado de σ , podemos mostrar que ele é um estimador *consistente* de σ (veja Apêndice C, disponível no site da Thomson), e que ele servirá muito bem para nossos propósitos.

A estimativa $\hat{\sigma}$ é interessante, já que ela é uma estimativa do desvio-padrão dos fatores não observáveis que afetam y ; equivalentemente, ela estima o desvio-padrão em y após os efeitos de x terem sido retirados. A maior parte dos programas econométricos informa o valor de $\hat{\sigma}$ juntamente com o R -quadrado, o intercepto, a inclinação e outras estatísticas de MQO (sob um dos vários nomes listados anteriormente). Por enquanto, nosso principal interesse está em usar $\hat{\sigma}$ para estimar os desvios-padrão de $\hat{\beta}_0$ e $\hat{\beta}_1$. Como $\text{dp}(\hat{\beta}_1) = \sigma/s_x$, o estimador natural de $\text{dp}(\hat{\beta}_1)$ é

$$\text{ep}(\hat{\beta}_1) = \hat{\sigma}/s_x = \hat{\sigma} / \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

que é chamado de **erro-padrão** de $\hat{\beta}_1$. Observe que $\text{ep}(\hat{\beta}_1)$ é visto como uma variável aleatória quando pensamos em processar o método MQO usando diferentes amostras de y ; isso ocorre porque σ^2 varia com diferentes amostras. Para uma dada amostra, $\text{ep}(\hat{\beta}_1)$ é um número, exatamente como $\hat{\beta}_1$ é simplesmente um número quando nós o calculamos a partir de dados conhecidos.

Semelhantemente, $\text{ep}(\hat{\beta}_0)$ é obtido de $\text{dp}(\hat{\beta}_0)$ ao substituir σ por $\hat{\sigma}$. O erro-padrão de qualquer estimativa nos dá uma idéia de qual preciso é o estimador. Os erros-padrão desempenham um papel central em todo este texto; nós os usaremos para construir estatísticas de testes e intervalos de confiança para todos os procedimentos econométricos que cobriremos, a partir do Capítulo 4.

2.6 REGRESSÃO ATRAVÉS DA ORIGEM

Em raros casos, desejamos impor a restrição de que, quando $x = 0$, o valor esperado de y é zero. Há certas relações para as quais isso é razoável. Por exemplo, se a renda (x) for zero, então os gastos com o imposto de renda (y) devem ser zero. Além disso, há problemas quando um modelo que originalmente tem um intercepto diferente de zero é transformado em um modelo sem um intercepto.

Formalmente, nós escolhemos agora um estimador da inclinação, que chamaremos de $\tilde{\beta}_1$, e uma reta da forma

$$\tilde{y} = \tilde{\beta}_1 x, \quad (2.63)$$

em que os sinais gráficos do til sobre $\tilde{\beta}_1$ e \tilde{y} são usados para distinguir esse problema do problema muito mais comum de estimar um intercepto juntamente com uma inclinação. Costuma-se chamar (2.63) de **regressão através da origem**, pois a reta (2.63) passa pelo ponto $x = 0, \tilde{y} = 0$. Para obter a estimativa de inclinação em (2.63), nós ainda contamos com o método de mínimos quadrados ordinários, que minimiza, nesse caso, a soma dos resíduos quadrados:

$$\sum_{i=1}^n (y_i - \tilde{\beta}_1 x_i)^2. \quad (2.64)$$

Usando cálculo, pode-se mostrar que $\tilde{\beta}_1$ deve resolver a condição de primeira ordem:

$$\sum_{i=1}^n x_i (y_i - \tilde{\beta}_1 x_i) = 0. \quad (2.65)$$

Dáí, podemos resolver para $\tilde{\beta}_1$:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}, \quad (2.66)$$

desde que nem todos os x_i sejam zero — um caso que excluimos.

Observe como $\tilde{\beta}_1$ se compara com a estimativa de inclinação quando também estimamos o intercepto (em vez de determiná-lo igual a zero). Essas duas estimativas são as mesmas se, e somente se, $\bar{x} = 0$. [Veja a equação (2.49) para $\hat{\beta}_1$.] Obter uma estimativa de β_1 usando a regressão através da origem não é freqüente em trabalhos aplicados, e por boas razões: se o intercepto $\beta_0 \neq 0$, logo $\tilde{\beta}_1$ é um estimador viesado de β_1 . Você será solicitado a provar isso no Problema 2.8.

Neste capítulo introduzimos o modelo de regressão simples e cobrimos suas propriedades básicas. Dada uma amostra aleatória, o método de mínimos quadrados ordinários é usado para estimar os parâmetros de inclinação e de intercepto no modelo populacional. Demonstramos a álgebra da reta de regressão de MQO, incluindo os cálculos dos valores estimados e dos resíduos, e a obtenção das variações previstas na variável dependente a partir de uma dada variação na variável independente. Na Seção 2.4 discutimos duas questões de importância prática: (1) o comportamento das estimativas de MQO quando mudamos as unidades de medida da variável dependente ou da variável independente e (2) o uso do log natural na elaboração de modelos de elasticidade constante e de semi-elasticidade constante.

Na Seção 2.5, mostramos que, sob as quatro hipóteses RLS.1 a RLS.4, os estimadores de MQO são não-viesados. A hipótese fundamental é que o termo erro u tem média zero, dado qualquer valor da variável independente x . Infelizmente, há razões para pensar que isso é falso em muitas aplicações da regressão simples nas ciências sociais, em que os fatores omitidos em u estão frequentemente correlacionados com x . Quando adicionamos a hipótese de que a variância do erro, dado x , é constante, obtemos fórmulas simples das variâncias amostrais dos estimadores de MQO. Como vimos, a variância do estimador de inclinação $\hat{\beta}_1$ cresce quando a variância do erro cresce, e ela decresce quando há mais variação amostral na variável independente. Também derivamos um estimador não-viesado para $\sigma^2 = \text{Var}(u)$.

Na Seção 2.6, discutimos brevemente a regressão através da origem, cujo estimador de inclinação é obtido sob a hipótese de que o intercepto é zero. Às vezes, tal regressão é útil, mas ela não aparece com frequência em trabalhos aplicados.

Temos ainda muito trabalho por fazer. Por exemplo, ainda não sabemos como testar hipóteses sobre os parâmetros populacionais, β_0 e β_1 . Assim, embora saibamos que o método MQO é, sob as hipóteses RLS.1 a RLS.4, não-viesado para os parâmetros populacionais, não temos um modo de fazer inferências sobre a população. Outros tópicos, tais como a eficiência de MQO relativa a outros procedimentos possíveis, também foram omitidos.

As questões de intervalos de confiança, testes de hipóteses e eficiência também são centrais para a análise de regressão múltipla. Como a maneira que construímos os intervalos de confiança e as estatísticas de testes é muito similar para a regressão múltipla — e porque a regressão simples é um caso especial da regressão múltipla —, nosso tempo será mais bem gasto se nos movermos para a regressão múltipla, que é muito mais aplicável que a regressão simples. Nosso propósito, no Capítulo 2, foi fazer você pensar nas questões que surgem na análise econométrica dentro de uma estrutura clara e simples.

2.1 Seja *filhos* o número de filhos de uma mulher e *educ* os anos de educação da mulher. Um modelo simples que relaciona a fertilidade a anos de educação é

$$\textit{filhos} = \beta_0 + \beta_1 \textit{educ} + u,$$

em que u é um erro não-observável.

- (i) Que tipos de fatores estão contidos em u ? É provável que eles estejam correlacionados com o nível de educação?

- (ii) Uma análise de regressão simples mostrará o efeito *ceteris paribus* da educação sobre a fertilidade? Explique.

2.2 No modelo de regressão linear simples $y = \beta_0 + \beta_1 x + u$, suponha que $E(u) \neq 0$. Fazendo $\alpha_0 = E(u)$, mostre que o modelo pode sempre ser reescrito com a mesma inclinação, mas com um novo intercepto e erro, em que o novo erro tem um valor esperado zero.

2.3 A tabela seguinte contém as variáveis *nmgrad* (nota média em curso superior nos Estados Unidos) e *tac* (nota do teste de avaliação de conhecimentos para ingresso em curso superior nos Estados Unidos) com as notas hipotéticas de oito estudantes de curso superior. A nota *nmgrad* está baseada em uma escala de quatro pontos e foi arredondada para um dígito após o ponto decimal. A nota *tac* baseia-se em uma escala de 36 pontos e foi arredondada para um número inteiro.

<i>Estudante</i>	<i>nmgrad</i>	<i>tac</i>
1	2,8	21
2	3,4	24
3	3,0	26
4	3,5	27
5	3,6	29
6	3,0	25
7	2,7	25
8	3,7	30

- (i) Estime a relação entre *nmgrad* e *tac* usando MQO; isto é, obtenha as estimativas de intercepto e de inclinação da equação

$$nmgrad = \hat{\beta}_0 + \hat{\beta}_1 tac.$$

Comente a direção da relação. O intercepto tem uma interpretação útil aqui? Explique. Qual deveria ser o valor previsto de *nmgrad* se a nota *tac* aumentasse em cinco pontos?

- (ii) Calcule os valores estimados e os resíduos de cada observação e verifique que a soma dos resíduos é (aproximadamente) zero.
 (iii) Qual é o valor previsto de *nmgrad* quando *tac* = 20?
 (iv) Quanto da variação de *nmgrad* dos 8 estudantes é explicada por *tac*? Explique.

2.4 Os dados do arquivo BWGHT.RAW contém dados de nascimentos por mulheres nos Estados Unidos. As duas variáveis de interesse são: a variável dependente, peso dos recém-nascidos em onças* (*pesonas*), e a variável explicativa, número médio de cigarros que a mãe fumou por dia durante a gravidez (*cigs*). A seguinte regressão simples foi estimada usando dados de $n = 1.388$ nascimentos:

$$pe\u00f7onas = 119,77 - 0,514 cigs$$

* NT: 1 onça = 31,10 g.

- (i) Qual é o peso de nascimento previsto quando $cigs = 0$? E quando $cigs = 20$ (um maço por dia)? Comente a diferença.
- (ii) O modelo de regressão simples necessariamente captura uma relação causal entre o peso de nascimento da criança e os hábitos de fumar da mãe? Explique.
- (iii) Para prever um peso de nascimento de 125 onças, qual deveria ser a magnitude de $cigs$? Comente.
- (iv) Qual a fração de mulheres na amostra que não fumaram enquanto estiveram grávidas? Isso ajuda a reconciliar sua conclusão da parte (iii)?

2.5 Na função de consumo linear

$$côns = \hat{\beta}_0 + \hat{\beta}_1 rend,$$

a *propensão marginal a consumir* PMgC (estimada) é simplesmente a inclinação $\hat{\beta}_1$, enquanto a *propensão média a consumir* PmeC é $côns/rend = \hat{\beta}_0/rend + \hat{\beta}_1$. Usando as observações de renda e consumo anuais de 100 famílias (ambas medidas em dólares), obteve-se a seguinte equação:

$$côns = -124,84 + 0,853 \text{ rend}$$

$$n = 100, R^2 = 0,692.$$

- (i) Interprete o intercepto dessa equação e comente seu sinal e magnitude.
- (ii) Qual é o consumo previsto quando a renda familiar é \$ 30.000?
- (iii) Com $rend$ sobre o eixo de x , desenhe um gráfico da PMgC e da PmeC estimadas.

2.6 Usando dados de casas vendidas em 1988 em Andover, Massachusetts [Kiel e McClain (1995)], a equação seguinte relaciona os preços das casas ($preço$) à distância de um incinerador de lixo recentemente construído ($dist$):

$$\log(\text{preço}) = 9,40 + 0,312 \log(\text{dist})$$

$$n = 135, R^2 = 0,162.$$

- (i) Interprete o coeficiente de $\log(\text{dist})$. O sinal dessa estimativa é o que você esperava?
- (ii) Você considera que a regressão simples oferece um estimador não-viesado da elasticidade *ceteris paribus* de $preço$ em relação a $dist$? (Pense na decisão da cidade sobre onde colocar o incinerador.)
- (iii) Quais outros fatores relativos a casas afetam seu preço? Eles poderiam estar correlacionados com a distância do incinerador?

2.7 Considere a função de poupança

$$poup = \beta_0 + \beta_1 \text{rend} + u, \quad u = \sqrt{\text{rend}} \cdot e,$$

onde e é uma variável aleatória com $E(e) = 0$ e $\text{Var}(e) = \sigma_e^2$. Assuma que e é independente de $rend$.

- (i) Mostre que $E(u|rend) = 0$, de modo que a hipótese de média condicional zero (hipótese RLS.3) é satisfeita. [Sugestão: se e é independente de $rend$, então $E(e|rend) = E(e)$.]
- (ii) Mostre que $Var(u|rend) = \sigma_e^2 \cdot rend$, de modo que a hipótese de homoscedasticidade RLS.5 é violada. Em particular, a variância de $poup$ aumenta com $rend$. [Sugestão: $Var(e|rend) = Var(e)$, se e e $rend$ são independentes.]
- (iii) Faça uma discussão que sustente a hipótese de que a variância da poupança aumenta com a renda da família.

2.8 Considere o modelo de regressão simples padrão $y = \beta_0 + \beta_1 x + u$, sob as hipóteses RLS.1 a RLS.4. Os estimadores usuais $\hat{\beta}_0$ e $\hat{\beta}_1$ são não-viesados para seus respectivos parâmetros populacionais. Seja $\tilde{\beta}_1$ o estimador de β_1 obtido ao assumir que o intercepto é zero (veja a Seção 2.6).

- (i) Encontre $E(\tilde{\beta}_1)$ em termos de x_i , β_0 e β_1 . Verifique que $\tilde{\beta}_1$ é não-viesado para β_1 quando o intercepto populacional (β_0) é zero. Há outros casos em que $\tilde{\beta}_1$ é não-viesado?
- (ii) Encontre a variância de $\tilde{\beta}_1$. [Sugestão: a variância não depende de β_0 .]
- (iii) Mostre que $Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$. [Sugestão: para qualquer amostra de dados, $\sum_{i=1}^n x_i^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$, com a desigualdade estrita preponderando, a não ser que $\bar{x} = 0$.]
- (iv) Comente a relação entre viés e variância, ao escolher entre $\hat{\beta}_1$ e $\tilde{\beta}_1$.

2.9 (i) Sejam $\hat{\beta}_0$ e $\hat{\beta}_1$ o intercepto e a inclinação da regressão de y_i sobre x_i , usando n observações. Sejam c_1 e c_2 constantes, com $c_2 \neq 0$. Sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de $c_1 y_i$ sobre $c_2 x_i$. Mostre que $\tilde{\beta}_1 = (c_1/c_2) \hat{\beta}_1$ e $\tilde{\beta}_0 = c_1 \hat{\beta}_0$, verificando as observações sobre as unidades de medida da Seção 2.4. [Sugestão: para obter $\tilde{\beta}_1$, insira as transformações de x e y em (2.19). Em seguida, use (2.17) para $\tilde{\beta}_0$, estando seguro de usar as transformações de x e y e a inclinação correta.]

- (ii) Agora, sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ os parâmetros estimados da regressão de $(c_1 + y_i)$ sobre $(c_2 + x_i)$ (sem qualquer restrição sobre c_1 ou c_2). Mostre que $\tilde{\beta}_1 = \hat{\beta}_1$ e $\tilde{\beta}_0 = \hat{\beta}_0 + c_1 - c_2 \hat{\beta}_1$.
- (iii) Agora, sejam $\hat{\beta}_0$ e $\hat{\beta}_1$ as estimativas de MQO da regressão $\log(y_i)$ sobre x_i , para a qual devemos assumir $y_i > 0$ para todo i . Para $c_1 > 0$, sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de $\log(c_1 y_i)$ sobre x_i . Mostre que $\tilde{\beta}_1 = \hat{\beta}_1$ e $\tilde{\beta}_0 = \log(c_1) + \hat{\beta}_0$.
- (iv) Agora, assumindo que $x_i > 0$ para todo i , sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de y_i sobre $\log(c_2 x_i)$. Como $\tilde{\beta}_0$ e $\tilde{\beta}_1$ comparam-se com o intercepto e a inclinação da regressão de y_i sobre $\log(x_i)$?

Minimizando a Soma dos Resíduos Quadrados

Mostramos aqui que as estimativas de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ minimizam a soma dos resíduos quadrados, como afirmado na Seção 2.2. Formalmente, o problema é caracterizar as soluções $\hat{\beta}_0$ e $\hat{\beta}_1$ para o problema de minimização

$$\min_{b_0, b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2,$$

onde b_0 e b_1 são argumentos *dummies* para o problema de otimização; por simplicidade, chame essa função $Q(b_0, b_1)$. De um resultado fundamental do cálculo multivariado (veja Apêndice A, disponível no site de Thomson), uma condição necessária para $\hat{\beta}_0$ e $\hat{\beta}_1$ resolverem o problema de minimização é que as derivadas parciais de $Q(b_0, b_1)$ em relação a b_0 e b_1 devem ser zero quando avaliadas com $\hat{\beta}_0$ e $\hat{\beta}_1$: $\partial Q(\beta_0, \hat{\beta}_1)/\partial b_0 = 0$ e $\partial Q(\hat{\beta}_0, \hat{\beta}_1)/\partial b_1 = 0$. Usando a regra da cadeia do cálculo, essas duas equações tornam-se

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0. \end{aligned}$$

Essas duas equações são exatamente (2.14) e (2.15) multiplicadas por $-2n$ e, portanto, são solucionadas pelos mesmos $\hat{\beta}_0$ e $\hat{\beta}_1$.

Como sabemos que, realmente, minimizamos a soma dos resíduos quadrados? As condições de primeira ordem são necessárias, mas não são suficientes. Uma maneira de verificar que minimizamos a soma dos resíduos quadrados é escrever, para qualquer b_0 e b_1 ,

$$\begin{aligned} Q(b_0, b_1) &= \sum_{i=1}^n [y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i + (\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2 \\ &= \sum_{i=1}^n [\hat{u}_i + (\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 + n(\hat{\beta}_0 - b_0)^2 + (\hat{\beta}_1 - b_1)^2 \sum_{i=1}^n x_i^2 + 2(\hat{\beta}_0 - b_0)(\hat{\beta}_1 - b_1) \sum_{i=1}^n x_i, \end{aligned}$$

onde usamos as equações (2.30) e (2.31). A soma dos resíduos quadrados não depende de b_0 e b_1 , enquanto a soma dos últimos três termos pode ser escrita como

$$\sum_{i=1}^n [(\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2,$$

como pode ser verificado diretamente por álgebra. Visto que essa expressão é uma soma de termos quadrados, ela é maior que zero. Portanto, seu menor valor ocorre quando $b_0 = \hat{\beta}_0$ e $b_1 = \hat{\beta}_1$.

Análise de Regressão Múltipla: Estimação

o Capítulo 2 aprendemos a usar a análise de regressão simples para explicar uma variável dependente y como função de uma única variável independente x . A desvantagem principal de usar a análise de regressão simples em trabalhos empíricos é o fato de ser muito difícil obter conclusões *ceteris paribus* sobre como x afeta y : a hipótese fundamental, RLS.3 — todos os outros fatores que afetam y são não-correlacionados com x —, é freqüentemente irreal.

A **análise de regressão múltipla** é mais receptiva à análise *ceteris paribus*, pois ela nos permite controlar *explicitamente* muitos outros fatores que, de maneira simultânea, afetam a variável dependente. Isso é importante tanto para testar teorias econômicas quanto para avaliar efeitos da política governamental quando devemos nos basear em dados não-experimentais. Como os modelos de regressão múltipla podem acomodar muitas variáveis explicativas que podem estar correlacionadas, podemos esperar inferir causalidade nos casos em que a análise de regressão simples seria enganosa.

Naturalmente, se adicionarmos ao nosso modelo mais fatores que são úteis para explicar y , então mais da variação de y poderá ser explicada. Assim, a análise de regressão múltipla pode ser usada para construir modelos melhores para prever a variável dependente.

Uma vantagem adicional da análise de regressão múltipla é que ela pode incorporar, completamente, relações de formas funcionais gerais. No modelo de regressão simples, somente a função de uma variável explicativa pode aparecer na equação. Como veremos, o modelo de regressão múltipla permite muito mais flexibilidade.

A Seção 3.1 introduz, formalmente, o modelo de regressão múltipla e, mais adiante, discute as vantagens da regressão múltipla sobre a regressão simples. Na Seção 3.2, demonstramos como estimar os parâmetros do modelo de regressão múltipla por meio do método de mínimos quadrados ordinários. Nas seções 3.3, 3.4 e 3.5 descrevemos várias propriedades dos estimadores de MQO, incluindo a inexistência de vies e a eficiência.

O modelo de regressão múltipla ainda é o veículo mais extensamente usado da análise empírica em economia e em outras ciências sociais. Igualmente, o método de mínimos quadrados ordinários é popularmente usado para estimar os parâmetros do modelo de regressão múltipla.

3.1 FUNCIONALIDADE DA REGRESSÃO MÚLTIPLA

Modelo com Duas Variáveis Independentes

Iniciaremos com alguns exemplos simples para mostrar como a análise de regressão múltipla pode ser usada para resolver problemas que não podem ser resolvidos pela regressão simples.

O primeiro exemplo é uma variação simples da equação do salário introduzida no Capítulo 2 para obter o efeito da educação sobre o salário-hora:

$$\text{saláριο} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u, \quad (3.1)$$

em que *exper* representa anos de experiência no mercado de trabalho. Assim, *saláριο* é determinado por duas variáveis explicativas ou independentes, educação e experiência, e por outros fatores não-observados, contidos em *u*. Basicamente, ainda estamos interessados no efeito de *educ* sobre *saláριο*, mantendo fixos todos os outros fatores que afetam *saláριο*; isto é, estamos interessados no parâmetro β_1 .

Comparada com uma análise de regressão simples que relaciona *saláριο* a *educ*, a equação (3.1) remove, efetivamente, *exper* do termo erro e a coloca explicitamente na equação. Como *exper* aparece na equação, seu coeficiente, β_2 , mede o efeito *ceteris paribus* de *exper* sobre *saláριο*, que também é de algum interesse.

Não surpreendentemente, assim como na regressão simples, teremos de fazer hipóteses sobre como *u*, em (3.1), está relacionado às variáveis independentes, *educ* e *exper*. Entretanto, como veremos na Seção 3.2, há uma coisa da qual podemos estar seguros: visto que (3.1) contém a experiência de modo explícito, seremos capazes de mensurar o efeito da educação sobre o salário horário, mantendo a experiência fixa. Na análise de regressão simples — que coloca *exper* no termo erro —, teríamos de assumir que experiência é não-correlacionada com educação, uma hipótese tênue.

Como segundo exemplo, considere o problema de explicar o efeito do gasto público por estudante (*gasto*) sobre a nota média padronizada (*notmed*) do ensino médio. Suponha que a nota média dependa do gasto público, da renda familiar média (*rendfam*) e de outros fatores não-observáveis:

$$\text{notmed} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{rendfam} + u. \quad (3.2)$$

Para o propósito de análise da política governamental, o coeficiente de interesse é β_1 , o efeito *ceteris paribus* de gasto sobre *notmed*. Ao incluir *rendfam* explicitamente no modelo, somos capazes de controlar seu efeito sobre *notmed*. Isso é provavelmente importante, pois a renda familiar média tende a estar correlacionada com o gasto público por estudante: os níveis de gasto público são, frequentemente, determinados tanto por impostos locais sobre a propriedade como sobre a renda. Na análise de regressão simples, *rendfam* estaria incluída no termo erro, que estaria provavelmente correlacionado com *gasto*, fazendo com que o estimador de β_1 de MQO fosse viesado no modelo de duas variáveis.

Nesses dois exemplos similares, mostramos que outros fatores observáveis, além da variável de interesse primordial [*educ* na equação (3.1) e *gasto* na equação (3.2)] podem ser incluídos em um modelo de regressão. Em geral, podemos escrever um modelo com duas variáveis independentes como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad (3.3)$$

em que β_0 é o intercepto, β_1 mede a variação em *y* com relação a x_1 , mantendo fixos outros fatores, e β_2 mede a variação em *y* com relação a x_2 , mantendo outros fatores fixos.

A análise de regressão múltipla também é útil para generalizar relações funcionais entre variáveis. Como exemplo, suponha que o consumo da família (*cons*) é uma função quadrática da renda familiar (*rend*):

$$\text{cons} = \beta_0 + \beta_1 \text{rend} + \beta_2 \text{rend}^2 + u, \quad (3.4)$$

em que u contém outros fatores que afetam o consumo. Nesse modelo, o consumo depende somente de um fator observado, a renda; desse modo, pareceria que ele pode ser tratado dentro do arcabouço da regressão simples. No entanto, esse modelo está fora do padrão da regressão simples, porque ele contém duas funções da renda, rend e rend^2 (e, portanto, três parâmetros: β_0 , β_1 e β_2). Portanto, a função consumo é facilmente escrita como modelo de regressão com duas variáveis, e fazendo $x_1 = \text{rend}$ e $x_2 = \text{rend}^2$.

Mecanicamente, não há *nenhuma* diferença em usar o método de mínimos quadrados ordinários (introduzido na Seção 3.2) para estimar equações tão diferentes como (3.1) e (3.4). Cada uma dessas equações pode ser escrita como (3.3), que é tudo o que importa para os cálculos. Há, entretanto, uma diferença importante em como *interpretar* os parâmetros. Na equação (3.1), β_1 é o efeito *coeteris paribus* de *educ* sobre *salário*. O parâmetro β_1 não tem tal interpretação em (3.4). Em outras palavras, não faz sentido medir o efeito de *rend* sobre *cons* mantendo, ao mesmo tempo, rend^2 fixo, porque se *rend* varia, então rend^2 deve variar! Em vez disso, a variação no consumo com respeito à variação na renda — a propensão marginal a consumir — é aproximada por

$$\frac{\Delta \text{cons}}{\Delta \text{rend}} \approx \beta_1 + 2\beta_2 \text{rend}$$

Para os cálculos necessários quanto a derivação dessa equação, veja o Apêndice A, disponível na página do livro, no site www.thomsonlearning.com.br. Em outras palavras, o efeito marginal da renda sobre o consumo depende tanto de β_2 como de β_1 e do nível de renda. Esse exemplo mostra que, em qualquer aplicação particular, as definições das variáveis independentes são cruciais. Mas, para o desenvolvimento teórico da regressão múltipla, podemos ser vagos com relação a tais detalhes. No Capítulo 6 estudaremos exemplos como esse de forma mais completa.

No modelo com duas variáveis independentes, a hipótese fundamental sobre como u está relacionado a x_1 e x_2 é

$$E(u|x_1, x_2) = 0. \quad (3.5)$$

A interpretação da condição (3.5) é similar à interpretação da hipótese RLS.3 da análise de regressão simples. Ela significa que, para qualquer valor de x_1 e x_2 na população, o fator não-observável médio é igual a zero. Assim como na regressão simples, a parte importante da hipótese é que o valor esperado de u é o mesmo para todas as combinações de x_1 e x_2 ; dizer que esse valor comum é zero está longe de ser uma hipótese, desde que o intercepto β_0 esteja incluído no modelo (veja a Seção 2.1).

Como podemos interpretar a hipótese de média condicional zero no exemplo anterior? Na equação (3.1), a hipótese é $E(u|\text{educ}, \text{exper}) = 0$. Isso implica que outros fatores que afetam *salário* não estão, em média, relacionados a *educ* e *exper*. Portanto, se entendermos que aptidão inata é parte de u , então precisaremos que os níveis médios de aptidão sejam os mesmos em todas as combinações de educação e experiência na população que trabalha. Isso pode ou não ser verdadeiro, mas, como veremos na Seção 3.3, essa é a questão que precisamos fazer a fim de determinar se o método de mínimos quadrados ordinários produz estimadores não-viesados.

O exemplo que mede o desempenho dos estudantes [equação (3.2)] é similar à equação do salário. A hipótese de média condicional zero é $E(u|\text{gasto}, \text{rendfam}) = 0$, o que significa que os outros fato-

res que afetam as notas — características das escolas e dos estudantes — são, em média, não-relacionados aos gastos públicos por estudante e à renda familiar média.

Um modelo simples para explicar as taxas de homicídio nas cidades (*taxhom*) em termos da probabilidade de condenação (*prcond*) e da duração média da sentença (*sentmed*) é

$$\text{taxhom} = \beta_0 + \beta_1 \text{prcond} + \beta_2 \text{sentmed} + u.$$

Que fatores estão contidos em u ? Você entende ser provável que a hipótese (3.5) se mantenha?

Quando aplicada à função quadrática do consumo em (3.4), a hipótese da média condicional zero tem uma interpretação ligeiramente diferente. A equação (3.5), escrita literalmente, é $E(u|rend,rend^2) = 0$. Como $rend^2$ é conhecido quando se conhece $rend$, é redundante incluir $rend^2$ na esperança: $E(u|rend,rend^2) = 0$ é o mesmo que $E(u|rend) = 0$. Não há problema em colocar $rend^2$ junto com $rend$ na esperança quando expressamos a hipótese, mas $E(u|rend) = 0$ é mais conciso.

Modelo com k Variáveis Independentes

Como estamos no contexto da regressão múltipla, não há necessidade de ficarmos com duas variáveis independentes. A análise de regressão múltipla permite que muitos fatores observados afetem y . No exemplo do salário, poderíamos também incluir semanas de treinamento de trabalho, anos de permanência com o empregador atual, medidas de aptidão e mesmo variáveis demográficas, como o número de irmãos ou a educação da mãe. No exemplo do gasto público por estudante, poderiam ser incluídos variáveis adicionais que medissem a qualidade dos professores e o tamanho das escolas.

O **modelo de regressão linear múltipla** geral (também chamado modelo de regressão múltipla) pode ser escrito, na população, como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u, \quad (3.6)$$

onde β_0 é o **intercepto**, β_1 é o parâmetro associado a x_1 , β_2 é o parâmetro associado a x_2 , e assim por diante. Como há k variáveis independentes e um intercepto, a equação (3.6) contém $k + 1$ parâmetros (desconhecidos) populacionais. Para simplificar, algumas vezes vamos nos referir aos outros parâmetros que não o intercepto como **parâmetros de inclinação**, ainda que, literalmente, nem sempre eles tenham esse significado. [Veja a equação (3.4), em que nenhum dos parâmetros, β_1 e β_2 , é, por si mesmo, uma inclinação, mas juntos determinam a inclinação da relação entre o consumo e a renda.]

A terminologia da regressão múltipla é similar àquela da regressão simples e é apresentada na Tabela 3.1. Exatamente como na regressão simples, a variável u é o **termo erro** ou **perturbação**. Ele contém outros fatores, além de x_1, x_2, \dots, x_k , que afetam y . Não importa quantas variáveis explicativas incluímos em nosso modelo, pois sempre haverá fatores que não podemos incluir, e eles estão contidos, coletivamente, em u .

Tabela 3.1

Terminologia para a Regressão Múltipla

y	x_1, x_2, \dots, x_k
Variável Dependente	Variáveis Independentes
Variável Explicada	Variáveis Explicativas
Variável de Resposta	Variáveis de Controle
Variável Prevista	Variáveis Previsoras
Regressando	Regressores

Ao aplicar o modelo de regressão múltipla geral, devemos saber como interpretar os parâmetros. Agora e nos capítulos subsequentes vamos adquirir bastante prática, mas é útil, neste ponto, relembrarmos algumas coisas que já sabemos. Suponha que os salários (*salário*) dos diretores executivos estejam relacionados às vendas das empresas (*vendas*) e à permanência dos diretores executivos nas empresas (*permceo*) pela equação

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 \text{permceo} + \beta_3 \text{permceo}^2 + u. \quad (3.7)$$

Essa equação enquadra-se no modelo de regressão múltipla (com $k = 3$) ao definirmos $y = \log(\text{salário})$, $x_1 = \log(\text{vendas})$, $x_2 = \text{permceo}$ e $x_3 = \text{permceo}^2$. Como sabemos do Capítulo 2, o parâmetro β_1 é a elasticidade (*ceteris paribus*) de *salário* em relação a *vendas*. Se $\beta_3 = 0$, então $100 \beta_2$ é, aproximadamente, o aumento percentual *ceteris paribus* em *salário* quando *permceo* aumenta em um ano. Quando $\beta_3 \neq 0$, o efeito de *permceo* sobre *salário* é mais complicado. Postergaremos até o Capítulo 6 um tratamento detalhado de modelos com termos quadráticos.

A equação (3.7) fornece um lembrete importante sobre a análise de regressão múltipla. O termo “linear” na expressão “modelo de regressão linear múltipla” significa que a equação (3.6) é linear nos parâmetros, β_j . A equação (3.7) é um exemplo de modelo de regressão múltipla que, ao mesmo tempo, é linear nos β_j e é uma relação não-linear entre *salário* e as variáveis *vendas* e *permceo*. Muitas aplicações da regressão múltipla envolvem relações não-lineares entre as variáveis subjacentes.

É fácil expressar a hipótese essencial para o modelo de regressão múltipla geral em termos de uma esperança condicional:

$$E(u|x_1, x_2, \dots, x_k) = 0. \quad (3.8)$$

No mínimo, a equação (3.8) requer que todos os fatores no termo erro não-observado sejam não-correlacionados com as variáveis explicativas. Ela também significa que consideramos corretamente a relação funcional entre as variáveis explicada e as explicativas. Qualquer problema que faça com que u seja correlacionado com qualquer variável independente faz com que (3.8) não seja válida. Na Seção 3.3 mostraremos que a hipótese (3.8) implica que o método MQO é não-viesado e derivaremos o viés que surge quando uma variável-chave for omitida da equação. Nos capítulos 15 e 16 estudaremos outras razões que podem fazer com que (3.8) não seja válida e mostraremos o que pode ser feito nesses casos.

3.2 MECÂNICA E INTERPRETAÇÃO DOS MÍNIMOS QUADRADOS ORDINÁRIOS

Vamos resumir, agora, algumas características computacionais e algébricas do método de mínimos quadrados ordinários, quando ele se aplica a um conjunto particular de dados. Discutiremos também como interpretar a equação estimada.

Obtenção das Estimativas de MQO

Vamos considerar, primeiramente, a estimação do modelo com duas variáveis independentes. A equação de MQO estimada é escrita de forma similar ao caso da regressão simples:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2, \quad (3.9)$$

onde $\hat{\beta}_0$ é a estimativa de β_0 , $\hat{\beta}_1$ é a estimativa de β_1 , e $\hat{\beta}_2$ é a estimativa de β_2 . Porém, como obtemos $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$? O método de **mínimos quadrados ordinários** escolhe as estimativas que minimizam a soma dos resíduos quadrados. Isto é, dadas n observações de y , x_1 e x_2 , $\{(x_{i1}, x_{i2}, y_i): i = 1, 2, \dots, n\}$, as estimativas $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$ são escolhidas, simultaneamente, para fazer com que a expressão

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 \quad (3.10)$$

tenha o menor tamanho possível.

A fim de entender o que o método MQO está fazendo, é importante dominar o significado da indexação das variáveis independentes em (3.10). As variáveis independentes têm, aqui, dois subscritos: i seguido de 1 ou 2. O subscrito i refere-se ao número da observação. Assim, a soma em (3.10) contempla todas as observações de $i = 1$ a n . O segundo índice é simplesmente um método para distinguir as diferentes variáveis independentes. No exemplo que relaciona *salário* a *educ* e *exper*, $x_{i1} = educ_i$ é a educação formal da pessoa i na amostra, e $x_{i2} = exper_i$ é a experiência da pessoa i . A soma dos resíduos quadra-

dos na equação (3.10) é $\sum_{i=1}^n (\text{salário}_i - \hat{\beta}_0 - \hat{\beta}_1 \text{educ}_i - \hat{\beta}_2 \text{exper}_i)^2$. No que vem a seguir, o subscrito i é reservado para indexar o número da observação. Se escrevermos x_{ij} , então isso significa a i -ésima observação da j -ésima variável independente. (Alguns autores preferem mudar a ordem do número da observação e do número da variável, de modo que x_{1i} é a observação i da variável um. Mas isso é apenas um problema de gosto notacional.)

No caso geral com k variáveis independentes, procuramos estimar $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ na equação

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (3.11)$$

As $k + 1$ estimativas de MQO delas foram escolhidas para minimizar a soma dos resíduos quadrados:

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2. \quad (3.12)$$

Esse problema de minimização pode ser resolvido usando cálculo multivariado (veja o Apêndice 3A, disponível no site da Thomson. Isso leva a $k + 1$ equações lineares com $k + 1$ estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ desconhecidos:

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ \vdots & \\ \sum_{i=1}^n x_{ik} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0. \end{aligned} \tag{3.13}$$

Essas equações são, chamadas freqüentemente de **condições de primeira ordem** de MQO. Assim, como no modelo de regressão simples da Seção 2.2, as condições de primeira ordem de MQO podem ser obtidas pelo método dos momentos: sob a hipótese (3.8), $E(u) = 0$ e $E(x_j u) = 0$, onde $j = 1, 2, \dots, k$. As equações em (3.13) são contrapartidas amostrais desses momentos da população, embora tenhamos omitido a divisão pelo tamanho da amostra n .

Mesmo para n e k de tamanhos moderados, resolver as equações em (3.13) fazendo os cálculos manualmente é tedioso. Não obstante, computadores modernos que processam programas padrões de estatística e econometria podem resolver essas equações com n e k grandes muito rapidamente.

Há somente um pequeno aviso: devemos assumir que as equações em (3.13) podem ser resolvidas *unicamente* para os $\hat{\beta}_j$. Por enquanto, assumimos apenas isso, como é usualmente o caso em modelos bem definidos. Na Seção 3.3 formulamos a hipótese necessária para a existência de estimativas de MQO únicas (veja a hipótese RLM.4).

Como na análise de regressão simples, a equação (3.11) é chamada **reta de regressão de MQO** ou a **função de regressão amostral (FRA)**. Chamaremos $\hat{\beta}_0$ a **estimativa de intercepto de MQO** e $\hat{\beta}_1, \dots, \hat{\beta}_k$ de **estimativas de inclinação de MQO** (correspondentes às variáveis independentes x_1, x_2, \dots, x_k).

A fim de indicar que uma regressão de MQO foi computada, escreveremos a equação (3.11) com y e x_1, \dots, x_k substituídos pelos seus nomes de variável (tais como *salário*, *educ* e *exper*), ou diremos que “rodamos uma regressão de MQO de y sobre x_1, x_2, \dots, x_k ou que regredimos y sobre x_1, x_2, \dots, x_k ”. Essas expressões são modos de dizer que o método de mínimos quadrados ordinários foi usado para obter a equação de MQO (3.11). A não ser que afirmemos explicitamente, sempre estimaremos um intercepto juntamente com as inclinações.

Interpretação da Equação de Regressão de MQO

Mais importante que os detalhes subjacentes à computação dos $\hat{\beta}_j$ é a *interpretação* da equação estimada. Iniciaremos com o caso de duas variáveis independentes:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \tag{3.14}$$

O intercepto $\hat{\beta}_0$ na equação (3.14) é o valor previsto de y quando $x_1 = 0$ e $x_2 = 0$. Às vezes, colocar x_1 e x_2 iguais a zero é um cenário interessante; em outros casos, isso não fará sentido. Não obstante, para obter uma previsão de y a partir da reta de regressão de MQO, o intercepto sempre é necessário, como (3.14) deixa claro.

As estimativas $\hat{\beta}_1$ e $\hat{\beta}_2$ têm interpretações de **efeito parcial**, ou *ceteris paribus*. Da equação (3.14), temos

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2,$$

de modo que podemos obter a variação prevista em y dadas as variações em x_1 e x_2 . (Observe que o intercepto não tem nada a ver com as variações em y .) Em particular, quando x_2 é mantido fixo, de modo que $\Delta x_2 = 0$, então

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1,$$

mantendo x_2 fixo. O ponto fundamental é que, ao incluir x_2 no nosso modelo, obtemos um coeficiente de x_1 com uma interpretação *ceteris paribus*. Essa é a razão de a análise de regressão múltipla ser tão útil. Semelhantemente,

$$\Delta \hat{y} = \hat{\beta}_2 \Delta x_2,$$

mantendo x_1 fixo.

EXEMPLO 3.1

(Determinantes da Nota Média em Curso Superior nos Estados Unidos)

As variáveis do arquivo GPA1.RAW incluem a nota média em um curso superior (*nmgrad*), a nota média do ensino médio (*nmem*) e a nota do teste de avaliação de conhecimentos para ingresso em curso superior (*tac*) para uma amostra de 141 estudantes de uma grande universidade dos Estados Unidos; tanto *nmgrad* como *nmem* estão baseados em uma escala de quatro pontos. Obtemos a seguinte reta de regressão de MQO para estimar *nmgrad* a partir de *nmem* e *tac*:

$$nmgrad = 1,29 + 0,453 nmem + 0,0094 tac. \quad (3.15)$$

Como interpretamos essa equação? Primeiro, o intercepto de 1,29 é o valor previsto de *nmgrad* se tanto *nmem* como *tac* forem iguais a zero. Como ninguém que frequenta um curso superior teve nota média no ensino médio igual a zero ou uma nota no teste de ingresso no curso superior igual a zero, o intercepto nessa equação não é, por si mesmo, significativo.

As estimativas mais interessantes são os coeficientes de inclinação de *nmem* e *tac*. Como esperamos, há uma relação parcial positiva entre *nmgrad* e *nmem*: mantendo *tac* fixo, um ponto adicional em *nmem* está associado a 0,453 de um ponto em *nmgrad*, ou quase meio ponto. Em outras palavras, se escolhermos dois estudantes, A e B, e esses estudantes tiverem a mesma nota *tac*, mas *nmem* do estudante A é um ponto maior que a *nmem* do estudante B, prevemos que o estudante A tem *nmgrad*

EXEMPLO 3.1 (continuação)

0,453 maior que *nmgrad* do estudante B. (Isso não diz nada sobre quaisquer duas pessoas reais, mas é a nossa melhor previsão.)

O sinal *tac* implica que, mantendo *nmem* fixo, uma variação de 10 pontos na nota em *tac* — uma variação muito grande, visto que a nota média na amostra é de cerca de 24, com um desvio-padrão menor que três — afeta *nmgrad* em menos de um décimo de um ponto. Esse é um efeito pequeno e sugere que, uma vez considerada o *nmem*, a nota do *tac* não é um forte preditor de *nmgrad*. (Naturalmente, há muitos outros fatores que contribuem para *nmgrad*, mas aqui estamos enfatizando as estatísticas disponíveis de estudantes do ensino médio.) Posteriormente, após discutirmos a inferência estatística, mostraremos que o coeficiente de *tac* não é somente pequeno na prática, mas ele também é estatisticamente não significativo.

Se colocarmos o foco na análise de regressão simples relacionando somente *nmgrad* e *tac*, obtemos

$$nmgrad = 2,40 + 0,0271 tac;$$

assim, o coeficiente *tac* é quase três vezes maior que a estimativa em (3.15). No entanto, essa equação não nos permite comparar duas pessoas com o mesmo *nmem*; ela corresponde a um experimento diferente. Mais adiante, falaremos mais sobre as diferenças entre as regressões múltipla e simples.

O caso com mais de duas variáveis independentes é similar. A reta de regressão de MQO é

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (3.16)$$

Escrita em termos de variações,

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \dots + \hat{\beta}_k \Delta x_k. \quad (3.17)$$

O coeficiente de x_1 mede a variação em \hat{y} devido a um aumento de uma unidade em x_1 , mantendo todas as outras variáveis independentes fixas. Isto é,

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1, \quad (3.18)$$

mantendo x_2, x_3, \dots, x_k fixos. Assim, controlamos as variáveis x_2, x_3, \dots, x_k ao estimar o efeito de x_1 sobre y . Os outros coeficientes têm uma interpretação similar.

O exemplo seguinte contém três variáveis independentes.

EXEMPLO 3.2**(Equação do Salário Horário)**

Usando as 526 observações de trabalhadores do arquivo WAGE1.RAW, incluímos *educ* (anos de educação formal), *exper* (anos de experiência no mercado de trabalho) e *perm* (anos com o empregador atual) na equação que explica $\log(\text{salário})$. A equação estimada é

EXEMPLO 3.2 (continuação)

$$\log(\text{saláριο}) = 0,284 + 0,092 \text{ educ} + 0,0041 \text{ exper} + 0,022 \text{ perm}. \quad (3.19)$$

Como no caso da regressão simples, os coeficientes têm uma interpretação de percentagem. A única diferença é que eles também têm uma interpretação *ceteris paribus*. O coeficiente 0,092 significa que, mantendo *exper* e *perm* fixos, um ano a mais de educação formal aumenta o valor esperado de $\log(\text{saláριο})$ em 0,092, o que se traduz em um aumento aproximado de 9,2% [$100(0,092)$] em *saláριο*. Alternativamente, se considerarmos duas pessoas com os mesmos níveis de experiência e permanência no trabalho, o coeficiente de *educ* é a diferença proporcional no salário horário previsto quando seus níveis de educação diferem em um ano. Essa medida de retorno da educação mantém fixos ao menos dois importantes fatores de produtividade; saber se ela é uma boa estimativa do retorno *ceteris paribus* de mais um ano de educação formal requer que estudemos as propriedades estatísticas de MQO (veja a Seção 3.3).

Sobre o Significado de “Manter Outros Fatores Fixos” na Regressão Múltipla

Como a interpretação de efeito parcial dos coeficientes de inclinação na análise de regressão múltipla pode causar alguma confusão, assim, vamos tentar impedir o surgimento desse problema agora.

No Exemplo 3.1, observamos que o coeficiente *tac* mede a diferença prevista em *nmgrad*, mantendo *nmem* fixo. O poder da análise de regressão múltipla é que ela proporciona uma interpretação *ceteris paribus* mesmo que os dados *não* sejam coletados de uma maneira *ceteris paribus*. Ao dar ao coeficiente de *tac* uma interpretação de efeito parcial, pode parecer que, realmente, saímos a campo e extraímos amostras compostas de pessoas com a mesma *nmem* e, possivelmente, com diferentes notas do *tac*. Isso não é verdade. Os dados são uma amostra aleatória de uma universidade grande: não há restrições colocadas sobre os valores amostrais de *nmem* ou *tac* na obtenção dos dados. De fato, raramente temos o luxo de manter certas variáveis fixas na obtenção de nossa amostra. Se pudéssemos coletar uma amostra de indivíduos com a mesma *nmem*, então poderíamos realizar uma análise de regressão simples relacionando *nmgrad* a *tac*. A regressão múltipla nos permite, efetivamente, simular essa situação sem restringir os valores de quaisquer variáveis independentes.

O poder que a análise de regressão múltipla tem é que ela nos permite fazer, em ambientes não-experimentais, o que os cientistas naturais são capazes de fazer em um ambiente controlado de laboratório: manter outros fatores fixos.

Varição de mais de uma Variável Independente Simultaneamente

Às vezes, queremos variar mais que uma variável independente ao mesmo tempo para encontrar o efeito resultante sobre a variável dependente. Isso é facilmente feito usando a equação (3.17). Por exemplo, na equação (3.19), podemos obter o efeito estimado sobre *saláριο* quando um indivíduo permanece na mesma empresa por mais um ano: ambos *exper* (experiência geral da força de trabalho) e *perm* aumentam em um ano. O efeito total (mantendo *educ* fixo) é

$$\Delta \log(\hat{\text{saláριο}}) = 0,0041 \Delta \text{exper} + 0,022 \Delta \text{perm} = 0,0041 + 0,022 = 0,0261,$$

ou cerca de 2,6%. Como *exper* e *perm* aumentam, cada um, em um ano, apenas somamos os coeficientes de *exper* e *perm* e multiplicamos por 100 para converter o efeito em uma percentagem.

Valores Estimados e Resíduos de MQO

Após obter a reta de regressão de MQO (3.11), podemos obter um *valor ajustado* ou *predito* para cada observação. Para a observação *i*, o valor ajustado é simplesmente

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}, \quad (3.20)$$

que é exatamente o valor previsto obtido inserindo os valores das variáveis independentes da observação *i* na equação (3.11). Ao obter os valores estimados, não devemos esquecer do intercepto; de outro modo, a resposta pode ser muito equivocada. Como exemplo, se em (3.15) $nmem_i = 3,5$ e $tac_i = 24$, $nmgrad_i = 1,29 + 0,453(3,5) + 0,0094(24) = 3,101$ (arredondado em três casas após o decimal).

Normalmente, para qualquer observação *i*, o valor real y_i não se iguala ao valor previsto, \hat{y}_i . O método de MQO minimiza o erro quadrado médio de previsão, que não diz nada sobre o erro de previsão de qualquer observação específica. O **resíduo** da observação *i* é definido exatamente como no caso da regressão simples,

$$\hat{u}_i = y_i - \hat{y}_i. \quad (3.21)$$

Há um resíduo para cada observação. Se $\hat{u}_i > 0$, então \hat{y}_i está abaixo de y_i , o que significa que, para essa observação, y_i é subestimado. Se $\hat{u}_i < 0$, então $y_i < \hat{y}_i$, e y_i é superestimado.

Os valores estimados de MQO e os resíduos têm algumas propriedades importantes que são extensões imediatas do caso da variável única:

1. A média amostral dos resíduos é zero.
2. A covariância amostral entre cada variável independente e os resíduos de MQO é zero. Conseqüentemente, a covariância amostral entre os valores estimados de MQO e os resíduos de MQO é zero.
3. O ponto $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k, \bar{y})$ está sempre sobre a reta de regressão de MQO: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \dots + \hat{\beta}_k \bar{x}_k$.

As duas primeiras propriedades são conseqüências imediatas do conjunto de equações usadas para obter as estimativas de MQO. A primeira equação em (3.13) diz que a soma dos resíduos é zero. As equações restantes são da forma $\sum_{i=1}^n x_{ij} \hat{u}_i = 0$, implicando que cada variável independente tem covariância amostral zero com \hat{u}_i . A Propriedade 3 decorre imediatamente da Propriedade 1.

No Exemplo 3.1, a reta estimada de MQO que explica *nmgrad* em termos de *nmem* é

$$nmgrad = 1,29 + 0,453 nmem + 0,0094 tac.$$

Se *nmem* médio é de cerca de 3,4 e a nota média do *tac* está em torno de 24,2, qual é o *nmgrad* médio na amostra?

Interpretação de “Parcialidade” da Regressão Múltipla

Ao aplicar o método MQO, não precisamos saber as fórmulas explícitas dos $\hat{\beta}_j$ que solucionam o sistema de equações em (3.13). Entretanto, para certas derivações, precisamos de fórmulas explícitas dos $\hat{\beta}_j$. Essas fórmulas também ajudam a esclarecer o funcionamento de MQO.

Considere, uma vez mais, o caso com $k = 2$ variáveis independentes, em que $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Para uma idéia mais concreta, vamos enfatizar $\hat{\beta}_1$. Um modo de expressar $\hat{\beta}_1$ é

$$\hat{\beta}_1 = \left(\sum_{i=1}^n \hat{r}_{i1} y_i \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right), \quad (3.22)$$

onde os \hat{r}_{i1} são os resíduos de MQO de uma regressão simples de x_1 sobre x_2 , usando a amostra à mão. Regredimos nossa primeira variável independente, x_1 , sobre nossa segunda variável independente, x_2 , e, em seguida, obtemos os resíduos (y não desempenha nenhum papel aqui). A equação (3.2) mostra que podemos, portanto, fazer uma regressão simples de y sobre \hat{r}_{i1} para obter $\hat{\beta}_1$. (Observe que os resíduos \hat{r}_{i1} têm uma média amostral zero, e assim $\hat{\beta}_1$ é a estimativa de inclinação usual da regressão simples.)

A representação da equação (3.22) dá outra demonstração da interpretação do efeito parcial de $\hat{\beta}_1$. Os resíduos \hat{r}_{i1} são a parte de x_{i1} que é não-correlacionada com x_{i2} . Outro modo de dizer isso é que \hat{r}_{i1} é x_{i1} após o efeito de x_{i2} ter sido *isolado*, ou *deduzido*. Assim, $\hat{\beta}_1$ mede a relação amostral entre y e x_1 após x_2 ter sido imparcializado.

Na análise de regressão simples, não há parcialização de outras variáveis, porque outras variáveis não estão incluídas na regressão. O Problema 3.17 o conduzirá, passo a passo, pelo processo de parcialização usando os dados de salários do Exemplo 3.2. Para propósitos práticos, o importante é que $\hat{\beta}_1$, na equação $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, mede a variação em y , dado um aumento de uma unidade em x_1 , mantendo x_2 fixo.

No modelo geral com k variáveis explicativas, $\hat{\beta}_1$ pode ainda ser escrito como na equação (3.22), mas os resíduos \hat{r}_{i1} vêm da regressão de x_1 sobre x_2, \dots, x_k . Assim, $\hat{\beta}_1$ mede o efeito de x_1 sobre y após x_2, \dots, x_k terem sido isolados ou deduzidos.

Comparação das Estimativas das Regressões Simples e Múltipla

Há dois casos especiais em que tanto a regressão simples de y sobre x_1 como a regressão de y sobre x_1 e x_2 produzirão a mesma estimativa de MQO de x_1 . Para maior precisão, escreva a regressão simples de y sobre x_1 como $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, e escreva a regressão múltipla como $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Sabemos que o coeficiente da regressão simples $\tilde{\beta}_1$ não se iguala, geralmente, ao coeficiente da regressão múltipla $\hat{\beta}_1$. Acontece que há uma relação simples entre $\tilde{\beta}_1$ e $\hat{\beta}_1$, que permite comparações interessantes entre as regressões simples e múltipla:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1,$$

em que $\tilde{\delta}_1$ é o coeficiente de inclinação da regressão simples de x_{i2} sobre x_{i1} , $i = 1, \dots, n$. Essa equação mostra como $\tilde{\beta}_1$ difere do efeito parcial de x_1 sobre \hat{y} . O termo que pode causar confusão é o efeito parcial de x_2 sobre \hat{y} multiplicado pela inclinação da regressão amostral de x_2 sobre x_1 . (Veja a Seção 3A.4, no apêndice do capítulo, disponível no site da Thomson, para uma verificação mais geral.)

A relação entre $\tilde{\beta}_1$ e $\hat{\beta}_1$ também mostra que há dois casos distintos em que eles são iguais:

1. O efeito parcial de x_2 sobre \hat{y} é zero na amostra, isto é, $\hat{\beta}_2 = 0$.
2. x_1 e x_2 são não-correlacionados na amostra, isto é, $\tilde{\delta}_1 = 0$.

Ainda que as estimativas das regressões múltipla e simples quase nunca sejam idênticas, podemos usar a fórmula anterior para caracterizar o motivo pelo qual elas deveriam ser muito diferentes ou bastante similares. Por exemplo, se $\hat{\beta}_2$ é pequeno, deveríamos esperar que as estimativas das regressões múltipla e simples de β_1 sejam semelhantes. No Exemplo 3.1, a correlação amostral entre *nmem* e *tac* é de cerca de 0,346, o que não é uma correlação trivial. Porém, o coeficiente de *tac* é razoavelmente pequeno. Não é surpreendente descobrir que a regressão de *nmgrad* sobre *nmem* produz uma estimativa de inclinação de 0,482, não é muito diferente da estimativa de 0,453 em (3.15).

EXEMPLO 3.3

(Participação nos Planos de Pensão 401(k))

Vamos usar os dados do arquivo em 401K.RAW para estimar o efeito de uma taxa de contribuição para um plano (*taxcont*) sobre a taxa de participação (*taxap*) dos trabalhadores nos planos de pensão de contribuição definidos existentes nos Estados Unidos. A taxa de contribuição é a quantidade com a qual a firma contribui para um fundo de trabalhadores, para cada dólar de contribuição do trabalhador (até certo limite); assim, *taxcont* = 0,75 significa que a firma contribui com 75 centavos de dólar para cada dólar contribuído pelo trabalhador. A taxa de participação é a percentagem de trabalhadores habilitados a ter uma conta no plano de pensão. A variável *idade* é a idade do plano de pensão. Há 1.534 planos no banco de dados, a *taxap* média é 87,36, a *taxcont* média é 0,732 e a *idade* média é 13,2.

Regredindo *taxap* sobre *taxcont* e *idade* resulta na equação

$$\hat{taxap} = 80,12 + 5,52 \text{ taxcont} + 0,243 \text{ idade.} \quad (3.23)$$

Assim, *taxcont* e *idade* têm os efeitos esperados. O que aconteceria se não controlássemos a variável *idade*? O efeito estimado de *idade* não é trivial, e portanto poderíamos esperar uma variação grande no efeito estimado de *taxcont* se *idade* fosse excluída da regressão. Entretanto, a regressão simples de *taxap* sobre *taxcont* produz $\hat{taxap} = 83,08 + 5,86 \text{ taxcont}$. A estimativa de regressão simples do efeito de *taxcont* sobre *taxap* é, claramente, diferente da estimativa de regressão múltipla, mas a diferença não é muito grande. (A estimativa da regressão simples é somente cerca de 6,2% maior que a estimativa da regressão múltipla.) Isso pode ser explicado pelo fato de a correlação amostral entre *taxcont* e *idade* ser somente de 0,12.

No caso com k variáveis independentes, a regressão simples de y sobre x_1 e a regressão múltipla de y sobre x_1, x_2, \dots, x_k produzem uma estimativa idêntica de x_1 somente se: (1) os coeficientes de MQO de x_2 até x_k forem todos zero ou (2) x_1 for não-correlacionado com cada um dos x_2, \dots, x_k . Na prática, nenhuma dessas possibilidades é muito provável. Porém, se os coeficientes de x_2 até x_k forem pequenos, ou as correlações amostrais entre x_1 e as outras variáveis independentes forem pouco substanciais, então as estimativas das regressões simples e múltiplas do efeito de x_1 sobre y podem ser similares.

Grau de Ajuste

Assim como na regressão simples, podemos definir a **Soma dos Quadrados Total (SQT)**, a **Soma dos Quadrados Explicada (SQE)** e a **Soma dos Quadrados dos Resíduos** ou **Soma dos Resíduos Quadrados (SQR)** como

$$\text{SQT} \equiv \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.24)$$

$$\text{SQE} \equiv \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.25)$$

$$\text{SQR} \equiv \sum_{i=1}^n \hat{u}_i^2. \quad (3.26)$$

Usando o mesmo argumento utilizado no caso da regressão simples, podemos mostrar que

$$\text{SQT} = \text{SQE} + \text{SQR}. \quad (3.27)$$

Em outras palavras, a variação total em $\{y_i\}$ é a soma das variações totais em $\{\hat{y}_i\}$ e em $\{\hat{u}_i\}$.

Assumindo que a variação total em y não é zero — como é o caso, a não ser que y_i seja constante na amostra —, podemos dividir (3.27) por SQT para obter

$$\text{SQR/SQT} + \text{SQE/SQT} = 1.$$

Exatamente como no caso da regressão simples, o R -quadrado é definido como

$$R^2 \equiv \text{SQE/SQT} = 1 - \text{SQR/SQT}, \quad (3.28)$$

e é interpretado como a proporção da variação amostral em y_i que é explicada pela reta de regressão de MQO. Por definição, R^2 é um número entre zero e um.

Pode-se também mostrar que R^2 é igual ao quadrado do coeficiente de correlação entre os valores reais y_i real e os valores estimados \hat{y}_i ajustado. Isto é,

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \right)} \quad (3.29)$$

[Inserimos a média de \hat{y}_i em (3.29) por coerência com a fórmula do coeficiente de correlação; sabemos que essa média é igual a \bar{y} , porque a média amostral dos resíduos é zero e $y_i = \hat{y}_i + \hat{u}_i$.]

Um fato importante sobre R^2 é que ele nunca diminui, e geralmente aumenta, quando outra variável independente é adicionada à regressão. Esse fato algébrico ocorre por definição, pois a

soma dos resíduos quadrados nunca aumenta quando regressores adicionais são acrescentados ao modelo.

O fato de que R^2 nunca diminui quando *qualquer* variável for adicionada a uma regressão faz dele um instrumento fraco para decidir se uma variável ou diversas variáveis deveriam ser adicionadas ao modelo. O fator que deve determinar se uma variável explicativa pertence a um modelo é se a variável explicativa tem, na *população*, um efeito parcial sobre y diferente de zero. No Capítulo 4, quando cobrirmos a inferência estatística, mostraremos como testar essa hipótese. Veremos também que, quando usado apropriadamente, R^2 permite-nos *testar* um grupo de variáveis com a finalidade de ver se ele é importante para explicar y . Por enquanto, usaremos R^2 como uma medida do grau de ajuste para um dado modelo.

EXEMPLO 3.4

(Determinantes de *nmgrad*)

Da regressão de *nmgrad* que fizemos anteriormente, a equação com R^2 é

$$\begin{aligned} nmgrad &= 1,29 + 0,453 nmem + 0,0094 tac \\ n &= 141, R^2 = 0,176. \end{aligned}$$

Isso significa que *nmem* e *tac* explicam, juntos, cerca de 17,6% da variação em *nmgrad* da amostra de estudantes. Isso pode não parecer uma porcentagem alta, mas devemos nos lembrar de que há muitos outros fatores — incluindo formação da família, personalidade, qualidade da educação do ensino médio, afinidade com o curso escolhido — que contribuem para o desempenho dos estudantes. Se *nmem* e *tac* explicassem quase toda a variação em *nmgrad*, então o desempenho no curso superior seria predeterminado pelo desempenho no ensino médio!

EXEMPLO 3.5

(Explicando os Registros de Prisões)

O arquivo CRIME1.RAW contém dados de prisões durante o ano de 1986 e outras informações sobre 2.725 homens nascidos em 1960 ou 1961 na Califórnia. Cada homem na amostra foi preso pelo menos uma vez antes de 1986. A variável *npre86* é o número de vezes que determinado homem foi preso em 1986: ela é zero para muitos homens da amostra (72,29%), e varia de 0 a 12. (A porcentagem de homens presos uma vez em 1986 foi de 20,51%.) A variável *pcond* é a proporção (não a porcentagem) de prisões anteriores a 1986 que levaram à condenação, *sentmed* é a duração média da sentença cumprida por condenação prévia (zero para muitas pessoas), *ptemp86* são os meses passados na prisão em 1986 e *empr86* é o número de trimestres durante o qual determinado homem ficou empregado em 1986 (de zero a quatro).

Um modelo linear explicando as prisões é

$$npre86 = \beta_0 + \beta_1 pcond + \beta_2 sentmed + \beta_3 ptemp86 + \beta_4 empr86 + u,$$

em que *pcond* é uma variável (*proxy*) da probabilidade de um homem ser condenado por um crime, e *sentmed* é uma medida do rigor esperado da pena, em caso de condenação. A variável *ptemp86* captura o efeito de confinamento do crime: se um indivíduo está na prisão, ele não pode ser preso por um crime fora da prisão. As oportunidades no mercado de trabalho são capturadas grosseiramente por *empr86*.

Primeiro, estimamos o modelo sem a variável *sentmed*. Obtemos

EXEMPLO 3.5 (continuação)

$$\begin{aligned} n\hat{p}re86 &= 0,712 - 0,150 pcond - 0,034 ptemp86 - 0,104 empr86 \\ n &= 2.725, R^2 = 0,0413. \end{aligned}$$

Essa equação diz que, como um grupo, as três variáveis $pcond$, $ptemp86$ e $empr86$ explicam cerca de 4,1% da variação em $npre86$.

Cada um dos coeficientes de inclinação de MQO tem o sinal esperado. Um aumento na proporção de condenações diminui o número previsto de prisões. Se aumentarmos $pcond$ em 0,50 (um aumento grande na probabilidade de condenação), então, mantendo os outros fatores fixos, $\Delta n\hat{p}re86 = -0,150(0,50) = -0,075$. Isso pode parecer pouco usual, porque uma prisão não pode ser uma fração. No entanto, podemos usar esse valor para obter a variação prevista das prisões esperadas de um grande grupo de homens. Por exemplo, entre cem homens, a queda esperada de prisões quando $pcond$ aumenta em 0,50 é $-7,5$.

Semelhantemente, um período de prisão mais longo leva a um número previsto menor de prisões. De fato, se $ptemp86$ aumenta de 0 para 12, as prisões previstas para um determinado homem diminuem em $0,034(12) = 0,408$. Um trimestre a mais no qual o emprego legal é informado diminui as prisões esperadas em 0,104, o que significaria 10,4 prisões entre cem homens.

Se $sentmed$ for adicionado ao modelo, sabemos que R^2 aumentará. A equação estimada é

$$\begin{aligned} n\hat{p}re86 &= 0,707 - 0,151 pcond + 0,0074 sentmed - 0,037 ptemp86 - 0,103 empr86 \\ n &= 2.725, R^2 = 0,0422. \end{aligned}$$

Assim, ao adicionar a variável sentença média, R^2 aumenta de 0,0413 para 0,0422, um efeito praticamente insignificante. O sinal do coeficiente de $sentmed$ também é inesperado: ele diz que uma duração mais longa da sentença média aumenta a atividade criminal.

O Exemplo 3.5 merece uma palavra final de cautela. O fato de as quatro variáveis explicativas incluídas na segunda regressão explicarem somente 4,2% da variação em $npre86$ não necessariamente significa que a equação é inútil. Ainda que, coletivamente, essas variáveis não expliquem muito da variação nas prisões, é possível que as estimativas de MQO sejam estimativas confiáveis dos efeitos *ceteris paribus* de cada variável independente sobre $npre86$. Como veremos, se esse for o caso, isso não depende, diretamente, do tamanho do R^2 . Em geral, um R^2 baixo indica que é difícil prever resultados individuais sobre y com muita precisão, algo que estudaremos com mais detalhes no Capítulo 6. No exemplo da prisão, o R^2 pequeno reflete algo sobre o qual suspeitamos nas ciências sociais: geralmente, é muito difícil prever o comportamento individual.

Regressão através da Origem

Algumas vezes, uma teoria econômica, ou o senso comum, sugere que β_0 deveria ser zero, e por isso devemos mencionar, brevemente, a estimação de MQO quando o intercepto é zero. Especificamente, vamos agora buscar uma equação da forma

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \dots + \tilde{\beta}_k x_k, \quad (3.30)$$

em que o símbolo gráfico “ \sim ” sobre as estimativas é utilizado para distingui-las das estimativas de MQO obtidas juntamente com o intercepto [como em (3.11)]. Em (3.30), quando $x_1 = 0$, $x_2 = 0$, ..., $x_k = 0$, o

valor previsto é zero. Nesse caso, diz-se que $\tilde{\beta}_1, \dots, \tilde{\beta}_k$ são as estimativas de MQO da regressão de y sobre x_1, x_2, \dots, x_k através da origem.

As estimativas de MQO em (3.30), como sempre, minimizam a soma dos resíduos quadrados, mas com o intercepto igualado a zero. Você deve estar prevenido de que as propriedades de MQO que derivamos anteriormente não se mantêm mais para a regressão através da origem. Em particular, os resíduos de MQO não têm mais uma média amostral zero. Além disso, se R^2 for definido como $1 - \text{SQR}/\text{SQT}$, em que SQT está dado em (3.24) e SQR é agora $\sum_{i=1}^n (y_i - \tilde{\beta}_1 x_{i1} - \dots - \tilde{\beta}_k x_{ik})^2$, então R^2 pode ser, de fato, negativo. Isso significa que a média amostral, \bar{y} , “explica” mais da variação em y_i do que as variáveis explicativas.

Devemos incluir um intercepto na regressão ou devemos concluir que as variáveis explicativas explicam fracamente y . A fim de sempre ter um R -quadrado não-negativo, alguns economistas preferem calcular R^2 como o quadrado do coeficiente de correlação entre os valores reais e estimados de y , como em (3.29). (Nesse caso, o valor estimado médio deve ser calculado diretamente, já que ele não é mais igual a \bar{y} .) Entretanto, não há um conjunto de regras sobre como calcular o R -quadrado para a regressão através da origem.

Uma desvantagem séria com a regressão através da origem é que, se o intercepto β_0 for diferente de zero no modelo populacional, então os estimadores dos parâmetros de inclinação serão viesados. O viés pode ser severo em alguns casos. O custo de estimar um intercepto quando β_0 é realmente zero é que as variâncias dos estimadores de inclinação de MQO são maiores.

3.3 O VALOR ESPERADO DOS ESTIMADORES DE MQO

Vamos nos voltar, agora, para as propriedades estatísticas do método de MQO, para estimar os parâmetros de um modelo da população subjacente. Nesta seção, derivamos o valor esperado dos estimadores de MQO. Em particular, formulamos e discutimos quatro hipóteses, que são extensões diretas das hipóteses do modelo de regressão simples, sob as quais os estimadores de MQO são estimadores não-viesados dos parâmetros da população. Também obtemos explicitamente o viés em MQO, quando uma variável importante for omitida da regressão.

Você deve lembrar que propriedades estatísticas não têm nada a ver com uma amostra particular, mas sim, mais precisamente, com a propriedade dos estimadores quando a amostragem aleatória é feita repetidamente. Assim, as seções 3.3, 3.4 e 3.5 são um pouco abstratas. Apesar de darmos exemplos de derivação do viés de modelos específicos, não é significativo falar sobre as propriedades estatísticas de um conjunto de estimativas de uma única amostra.

A primeira hipótese que vamos fazer define, simplesmente, o modelo de regressão linear múltipla (RLM).

HIPÓTESE RLM . 1 (LINEAR NOS PARÂMETROS)

O modelo na população pode ser escrito como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u, \quad (3.31)$$

em que $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros desconhecidos (constantes) de interesse, e u é um erro aleatório não-observável ou um termo de perturbação aleatória.

A equação (3.31) especifica, formalmente, o **modelo populacional**, algumas vezes chamado **modelo verdadeiro**, para considerar a possibilidade de podermos estimar um modelo diferente de (3.31). A característica fundamental é que o modelo é linear nos parâmetros $\beta_0, \beta_1, \dots, \beta_k$. Como sabemos, (3.31) é bastante flexível, pois y e as variáveis independentes podem ser funções arbitrárias de variáveis subjacentes de interesse, como os logaritmos naturais e os quadrados [veja, por exemplo, a equação (3.7)].

HIPÓTESE RLM. 2 (AMOSTRAGEM ALEATÓRIA)

Temos uma amostra aleatória de n observações, $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) : i = 1, 2, \dots, n\}$, do modelo populacional descrito por (3.31).

Às vezes, precisamos escrever a equação de uma observação particular i : para uma observação extraída aleatoriamente da população, temos

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i. \quad (3.32)$$

Lembre-se de que i refere-se à observação, enquanto o segundo subscrito em x é o número da variável. Por exemplo, podemos escrever uma equação do salário de diretores executivos para um diretor executivo específico particular i como

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}_i) + \beta_2 \text{permceo}_i + \beta_3 \text{perceo}_i^2 + u_i. \quad (3.33)$$

O termo u_i contém os fatores não-observáveis para o diretor executivo i que afetam seu salário. Nas aplicações, é usualmente mais fácil escrever o modelo na forma populacional, como em (3.31). Ela é menos desordenada e enfatiza que estamos interessados em estimar a relação populacional.

À luz do modelo (3.31), os estimadores de MQO $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ da regressão de y sobre x_1, \dots, x_k são agora considerados estimadores de $\beta_0, \beta_1, \dots, \beta_k$. Vimos, na Seção 3.2, que MQO escolhe as estimativas de uma amostra particular, de modo que os resíduos sejam, em média, iguais a zero e a correlação amostral entre cada variável independente e os resíduos seja zero. Para que MQO seja não-viesado, é preciso que a versão *populacional* dessa condição seja verdadeira.

HIPÓTESE RLM. 3 (MÉDIA CONDICIONAL ZERO)

O erro u tem um valor esperado igual a zero, dados quaisquer valores das variáveis independentes. Em outras palavras,

$$E(u|x_1, x_2, \dots, x_k) = 0. \quad (3.34)$$

Uma maneira como a hipótese RLM.3 pode ser violada é quando a relação funcional entre as variáveis explicadas e explicativas está mal-especificada na equação (3.31): por exemplo, se esquece-

mos de incluir o termo quadrático $rend^2$ na função consumo $cons = \beta_0 + \beta_1 rend + \beta_2 rend^2 + u$ quando estimamos o modelo. Outra forma funcional mal-especificada ocorre quando usamos o nível da variável e , de fato, é o log da variável que aparece no modelo populacional, ou vice-versa. Por exemplo, se o modelo verdadeiro tiver $\log(salário)$ como variável dependente, mas usarmos $salário$ como variável dependente em nossa análise de regressão, então os estimadores serão viesados. Intuitivamente, isso deveria ser muito claro. No Capítulo 9 discutiremos maneiras de detectar a má especificação da forma funcional.

Omitir um fator importante que está correlacionado com qualquer uma das variáveis x_1, x_2, \dots, x_k faz com que a hipótese RLM.3 também não se sustente. Com a análise de regressão múltipla, somos capazes de incluir muitos fatores entre as variáveis explicativas e, por isso, variáveis omitidas são menos prováveis de serem um problema na análise de regressão múltipla do que na análise de regressão simples. Não obstante, em qualquer aplicação, há sempre muitos fatores que, devido à limitação de dados ou à ignorância deles, não somos capazes de incluir. Se acharmos que esses fatores devem ser controlados e que eles estão correlacionados com uma ou mais variáveis independentes, então a hipótese RLM.3 será violada. Posteriormente, derivaremos esse viés.

Há outros modos pelos quais u pode estar correlacionado com uma variável explicativa. No Capítulo 15 discutiremos o problema do erro de medida em uma variável explicativa. No Capítulo 16 cobriremos o problema, conceitualmente mais difícil, em que uma ou mais variáveis explicativas é determinada conjuntamente com y . Vamos postergar nosso estudo desses problemas até que tenhamos um domínio firme da análise de regressão múltipla sob um conjunto ideal de hipóteses.

Quando a hipótese RLM.3 se mantém, dizemos freqüentemente que temos **variáveis explicativas exógenas**. Se x_j for correlacionado com u por alguma razão, então se diz que x_j é uma **variável explicativa endógena**. Os termos “exógena” e “endógena” originaram-se da análise de equações simultâneas (veja Capítulo 16), mas o significado do termo “variável explicativa endógena” evoluiu e passou a incluir qualquer caso em que uma variável explicativa pode estar correlacionada com o termo erro.

A última hipótese de que precisamos para mostrar que MQO é não-viesado assegura que os estimadores de MQO são, realmente, bem definidos. Para a regressão simples, precisamos assumir que a única variável independente não era constante na amostra. A hipótese correspondente para a análise de regressão múltipla é mais complicada.

H I P Ó T E S E R L M . 4 (COLINEARIDADE NÃO PERFEITA)

Na amostra (e, portanto, na população), nenhuma das variáveis independentes é constante, e não há relações *lineares exatas* entre as variáveis independentes.

A hipótese de colinearidade não perfeita somente diz respeito às variáveis independentes. Estudantes de econometria iniciantes tendem a confundir as hipóteses RLM.4 e RML.3, de modo que enfatizamos aqui que RLM.4 não diz *nada* sobre a relação entre u e as variáveis explicativas.

A hipótese RLM.4 é mais complicada que sua contrapartida na regressão simples, pois agora devemos examinar as relações entre todas as variáveis independentes. Se uma variável independente em (3.31) é uma combinação linear exata de outras variáveis independentes, dizemos que o modelo sofre de **colinearidade perfeita**, e ele não pode ser estimado por MQO.

É importante observar que a hipótese RLM.4 permite, *de fato*, que as variáveis independentes sejam correlacionadas; elas apenas não podem ser correlacionadas *perfeitamente*. Se não permitíssemos qualquer correlação entre as variáveis independentes, então a regressão múltipla não seria muito útil para a análise econométrica. Por exemplo, no modelo que relaciona notas de estudantes aos gastos com educação e à renda familiar,

$$\text{notmed} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{rendfam} + u,$$

esperamos, com certeza, que *gasto* e *rendfam* sejam correlacionados: distritos escolares com rendas familiares médias altas tendem a gastar mais em educação por estudante. De fato, a principal motivação para incluir *rendfam* na equação é que suspeitamos que ela seja correlacionada com *gasto*, e, desse modo, gostaríamos de mantê-la fixa na análise. A hipótese RLM.4 somente exclui a correlação perfeita entre *gasto* e *rendfam* em nossa amostra. Teríamos muito azar se obtivéssemos uma amostra em que os gastos por estudante fossem perfeitamente correlacionados com a renda familiar média. Porém, espera-se que haja alguma correlação — talvez uma quantidade substancial — e certamente ela é permitida.

A maneira mais simples como duas variáveis independentes podem ser perfeitamente correlacionadas é quando uma variável é um múltiplo constante da outra. Isso pode acontecer quando um pesquisador, inadvertidamente, coloca a mesma variável medida em unidades diferentes dentro da equação de regressão. Por exemplo, ao estimar a relação entre consumo e renda, não faz sentido incluir como variáveis independentes a renda mensurada em dólares e a renda mensurada em milhares de dólares. Uma delas é redundante. Que sentido faria manter a renda mensurada em dólares fixa, enquanto a renda mensurada em milhares de dólares varia?

Já sabemos que diferentes funções não-lineares da mesma variável *podem* aparecer entre os regressores. Por exemplo, o modelo $\text{cons} = \beta_0 + \beta_1 \text{rend} + \beta_2 \text{rend}^2 + u$ não viola a hipótese RLM.4: ainda que $x_2 = \text{rend}^2$ seja uma função exata de $x_1 = \text{rend}$, rend^2 não é uma função *linear* de *rend*. Incluir rend^2 no modelo é uma maneira útil de generalizar a forma funcional, diferentemente de incluir a renda mensurada em dólares e em milhares de dólares.

O senso comum nos diz para não incluir a mesma variável explicativa medida em diferentes unidades na mesma equação de regressão. Há também maneiras mais sutis de uma variável independente poder ser um múltiplo de outra. Suponha que gostaríamos de estimar uma extensão da função de consumo de elasticidade constante. Poderia parecer natural especificar um modelo tal como

$$\log(\text{cons}) = \beta_0 + \beta_1 \log(\text{rend}) + \beta_2 \log(\text{rend}^2) + u, \quad (3.35)$$

em que $x_1 = \log(\text{rend})$ e $x_2 = \log(\text{rend}^2)$. Usando as propriedades básicas do log natural (veja o Apêndice A, disponível no site de Thomson), $\log(\text{rend}^2) = 2 \cdot \log(\text{rend})$. Isto é, $x_2 = 2x_1$, e naturalmente isso é válido para todas as observações na amostra. Isso viola a hipótese RLM.4. Em vez disso, deveríamos incluir $[\log(\text{rend})]^2$, e não $\log(\text{rend}^2)$, juntamente com $\log(\text{rend})$. Essa é uma extensão prudente do modelo de elasticidade constante, e veremos como interpretar tais modelos no Capítulo 6.

Outra maneira de as variáveis independentes serem perfeitamente colineares ocorre quando uma variável independente pode ser expressa como uma função linear exata de duas ou mais das outras variáveis independentes. Por exemplo, suponha que queremos estimar o efeito dos gastos de campanha sobre os resultados da campanha. Por simplicidade, assumamos que cada eleição tem dois candidatos. Seja *votoA* a percentagem de votos do Candidato A; seja *gastoA* os gastos de campanha do Candidato A; seja *gastoB* os gastos de campanha do Candidato B; e seja *totalgasto* os gastos totais de campanha; todas as últimas três variáveis são medidas em dólares. Pode parecer natural especificar o modelo como

$$\text{votoA} = \beta_0 + \beta_1 \text{gastoA} + \beta_2 \text{gastoB} + \beta_3 \text{totalgasto} + u, \quad (3.36)$$

a fim de isolar os efeitos dos gastos de cada candidato e da quantidade total de gastos. No entanto, esse modelo viola a hipótese RLM.4, porque $x_3 = x_1 + x_2$ por definição. Tentar interpretar essa equação ao estilo *ceteris paribus* revela o problema. Supõe-se que o parâmetro de β_1 na equação (3.36) meça o efeito de um aumento de um dólar nos gastos do Candidato A sobre os votos do Candidato A, mantendo os gastos do Candidato B e os gastos totais fixos. Isso é uma tolice, pois, se *gastoB* e *totalgasto* forem mantidos fixos, não podemos aumentar *gastoA*.

A solução para a colinearidade perfeita em (3.36) é simples: retire qualquer uma das três variáveis do modelo. Provavelmente, tiraríamos *totalgasto*, e conseqüentemente o coeficiente de *gastoA* mensuraria o efeito de aumentar os gastos de A sobre a percentagem de votos recebidos por A, mantendo os gastos de B fixos.

O exemplo anterior mostra que a hipótese RLM.4 pode ser violada se não formos cuidadosos ao especificar nosso modelo. Essa hipótese também não se mantém se o tamanho da amostra, n , é muito pequeno em relação ao número de parâmetros que são estimados. No modelo de regressão geral da equação (3.31), há $k + 1$ parâmetros, e RLM.4 não se mantém se $n < k + 1$. Intuitivamente, isso faz sentido: para estimar $k + 1$ parâmetros, necessitamos de pelo menos $k + 1$ observações. Não surpreendentemente, é melhor ter tantas observações quanto possível, algo que veremos em nossos cálculos da variância na Seção 3.4.

No exemplo anterior, se usarmos como variáveis explicativas *gastoA*, *gastoB* e *partA*, em que $partA = 100 \cdot (gastoA/totalgasto)$ é a participação percentual dos gastos totais de campanha feitos pelo Candidato A, isso viola a hipótese RLM.4?

Se o modelo for cuidadosamente especificado e $n \geq k + 1$, a hipótese RLM.4 pode não se manter em casos raros devido a um azar ao coletar a amostra. Por exemplo, em uma equação de salários, com educação e experiência como variáveis, poderíamos obter uma amostra aleatória em que cada indivíduo tivesse exatamente duas vezes mais educação que anos de experiência. Esse cenário faria com que a hipótese RLM.4 falhasse, mas isso pode ser considerado muito improvável, a não ser que tenhamos um tamanho de amostra extremamente pequeno.

Agora, estamos prontos para mostrar que, sob essas quatro hipóteses da regressão múltipla, os estimadores de MQO são não-viesados. Como no caso da regressão simples, as esperanças estão condicionadas aos valores das variáveis independentes da amostra, mas não mostraremos esse condicionamento explicitamente.

T E O R E M A 3 . 1 (INEXISTÊNCIA DE VIÉS DE MQO)

Sob as hipóteses RLM.1 a RLM.4,

$$E(\hat{\beta}_j) = \beta_j, j = 0, 1, \dots, k, \quad (3.37)$$

para qualquer valor do parâmetro populacional β_j . Em outras palavras, os estimadores de MQO são estimadores não-viesados dos parâmetros da população.

Em nossos exemplos empíricos anteriores, a hipótese RLM.4 foi satisfeita (visto que fomos capazes de calcular as estimativas de MQO). Além disso, em sua maior parte, as amostras são aleatoriamente escolhidas de uma população bem-definida. Se acreditamos que os modelos especificados estão corretos sob a hipótese fundamental RLM.3, então podemos concluir que MQO é não-viesado nesses exemplos.

Como estamos nos aproximando do ponto em que podemos usar a regressão múltipla no trabalho empírico é útil lembrar o significado de inexistência de viés. É tentador, nos exemplos como o da equação do salário em (3.19), dizer algo como “9,2% é uma estimativa não-viesada do retorno da educação”. Como sabemos, uma estimativa não pode ser viesada: uma estimativa é um número fixo, obtido de uma amostra particular, usualmente diferente do parâmetro populacional. Quando dizemos que MQO é não-viesado sob as hipóteses RLM.1 a RLM.4, estamos dizendo que o *procedimento* pelo qual as estimativas de MQO foram obtidas é não-viesado, e tal procedimento é visto como algo aplicado em todas as amostras aleatórias possíveis. Esperamos que tenhamos obtido uma amostra que nos dê uma estimativa próxima do valor da população, mas, infelizmente, isso não pode ser garantido.

Inclusão de Variáveis Irrelevantes em um Modelo de Regressão

Uma questão que podemos dispensar com rapidez razoável é a **inclusão de uma variável irrelevante** ou a **superespecificação do modelo** na análise de regressão múltipla. Isso significa que uma (ou mais) das variáveis independentes está incluída no modelo, embora ela não tenha efeito parcial sobre y na população. (Isto é, seu coeficiente populacional é zero.)

Para ilustrar a questão, suponha que especificamos o modelo como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u, \quad (3.38)$$

e esse modelo satisfaz as hipóteses RLM.1 a RLM.4. Entretanto, x_3 não tem efeito sobre y após x_1 e x_2 terem sido controlados, o que significa que $\beta_3 = 0$. A variável x_3 pode ou não ser correlacionada com x_1 e x_2 , o que importa é que, uma vez que x_1 e x_2 estejam controlados, x_3 não tem efeito sobre y . Em termos de esperanças condicionais, $E(y|x_1, x_2, x_3) = E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Como não sabemos que $\beta_3 = 0$, somos inclinados a estimar a equação com x_3 :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3, \quad (3.39)$$

Incluimos a variável irrelevante, x_3 , em nossa regressão. Qual é o efeito de incluir x_3 em (3.39), quando seu coeficiente no modelo populacional (3.38) é zero? Em termos da inexistência de viés de $\hat{\beta}_1$ e $\hat{\beta}_2$, não há *nenhum efeito*. Essa conclusão não requer nenhuma derivação especial, já que ela decorre imediatamente do Teorema 3.1. Lembre-se, a inexistência de viés significa $E(\hat{\beta}_j) = \beta_j$ para *qualquer* valor de β_j , incluindo $\beta_j = 0$. Assim, podemos concluir que $E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_2) = \beta_2$ e $E(\hat{\beta}_3) = 0$ (para quaisquer valores de β_0 , β_1 e β_2). Mesmo que $\hat{\beta}_3$, por si mesmo, nunca seja exatamente zero, seu valor médio obtido de muitas amostras aleatórias será zero.

A conclusão do exemplo anterior é muito mais geral: incluir uma ou mais variáveis irrelevantes no modelo de regressão múltipla, ou superespecificar o modelo, não afeta a inexistência de viés dos estimadores de MQO. Isso significa que incluir variáveis irrelevantes é inócua? Não. Como veremos na Seção 3.4, incluir variáveis irrelevantes pode ter efeitos indesejáveis sobre as *variâncias* dos estimadores de MQO.

Viés de Variável Omitida: O Caso Simples

Suponha agora que, em vez de incluir uma variável irrelevante, omitimos uma variável que, realmente, pertence ao modelo verdadeiro (ou populacional). Isso é freqüentemente chamado problema de **excluir uma variável relevante** ou de **subespecificar o modelo**. No Capítulo 2, e anteriormente neste capítulo, afirmamos que esse problema geralmente faz com que os estimadores de MQO sejam viesados. Agora é o momento de mostrar isso explicitamente e, não menos importante, derivar a direção e o tamanho do viés.

Derivar o viés causado ao omitir uma variável importante é um exemplo de **análise de má-especificação**. Iniciaremos com o caso em que o modelo populacional verdadeiro tem duas variáveis explicativas e um termo erro:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad (3.40)$$

e assumimos que esse modelo satisfaz as hipóteses RLM.1 a RLM.4.

Suponha que nosso interesse primordial esteja em β_1 , o efeito parcial de x_1 sobre y . Por exemplo, y é o salário horário (ou log do salário horário), x_1 é educação e x_2 é uma medida de aptidão inata. A fim de obter um estimador não-viesado de β_1 , *deveríamos* computar a regressão de y sobre x_1 e x_2 (o que resulta em estimadores não-viesados de β_0 , β_1 e β_2). Entretanto, devido à nossa ignorância ou indisponibilidade de dados, estimamos o modelo *excluindo* x_2 . Em outras palavras, executamos somente uma regressão simples de y sobre x_1 , obtendo a equação

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1. \quad (3.41)$$

Usamos o símbolo gráfico “~” em vez de “^” para enfatizar que $\tilde{\beta}_1$ vem de um modelo subespecificado.

Ao aprender, pela primeira vez, o problema de variável omitida, pode ser difícil para o estudante distinguir entre o modelo verdadeiro subjacente, (3.40) nesse caso, e o modelo que realmente estimamos, capturado pela regressão em (3.41). Pode parecer bobagem omitir a variável x_2 se ela pertence ao modelo, mas freqüentemente não temos escolha. Por exemplo, suponha que *saláριοh* seja determinado pela equação

$$\text{saláριοh} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{aptid} + u. \quad (3.42)$$

Como a aptidão não é observada, estimamos, em vez disso, o modelo

$$\text{saláριοh} = \beta_0 + \beta_1 \text{educ} + v,$$

onde $v = \beta_2 \text{aptid} + u$. O estimador de β_1 da regressão simples de *saláριοh* sobre *educ* é o que estamos chamando $\tilde{\beta}_1$.

Vamos derivar o valor esperado de $\tilde{\beta}_1$ condicionado aos valores amostrais de x_1 e x_2 . Derivar essa esperança não é difícil, pois $\tilde{\beta}_1$ é exatamente o estimador de inclinação de MQO de uma regressão simples, e já estudamos esse estimador extensivamente no Capítulo 2. A diferença aqui é que devemos analisar suas propriedades quando o modelo de regressão simples é mal-especificado devido a uma variável omitida.

Da equação (2.49), podemos expressar $\tilde{\beta}_1$ como

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}. \quad (3.43)$$

O próximo passo é o mais importante. Visto que (3.40) é o modelo verdadeiro, escrevemos y_i para cada observação i como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \quad (3.44)$$

(e não $y_i = \beta_0 + \beta_1 x_{i1} + u_i$, já que o modelo verdadeiro contém x_2). Seja SQT_1 o denominador em (3.43). Se inserirmos em (3.43) o y_i de (3.44), o numerador em (3.43) passa a ser

$$\begin{aligned} & \sum_{i=1}^n (x_{i1} - \bar{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i) \\ &= \beta_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2} + \sum_{i=1}^n (x_{i1} - \bar{x}_1)u_i \\ &= \beta_1 SQT_1 + \beta_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2} + \sum_{i=1}^n (x_{i1} - \bar{x}_1)u_i. \end{aligned} \quad (3.45)$$

Se dividirmos (3.45) por SQT_1 , considerarmos a esperança condicionada aos valores das variáveis independentes e usarmos $E(u_i) = 0$, obteremos

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}. \quad (3.46)$$

Assim, $E(\tilde{\beta}_1)$ não é, geralmente, igual a β_1 : $\tilde{\beta}_1$ é viesado para β_1 .

A razão que multiplica β_2 em (3.46) tem uma interpretação simples: ela é exatamente o coeficiente de inclinação da regressão de x_2 sobre x_1 , utilizando nossa amostra de variáveis independentes, que pode ser escrita como

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1. \quad (3.47)$$

Como estamos condicionados aos valores amostrais de ambas as variáveis independentes, $\tilde{\delta}_1$ não é aleatório aqui. Portanto, podemos escrever (3.46) como

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1, \quad (3.48)$$

o que implica que o viés em $\tilde{\beta}_1$ é $E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$. Essa expressão é chamada freqüentemente de **viés de variável omitida**.

Da equação (3.48), vemos que há dois casos em que $\tilde{\beta}_1$ é não-viesado. O primeiro é bastante óbvio: se $\beta_2 = 0$ — de modo que x_2 não aparece no modelo verdadeiro (3.40) —, então $\tilde{\beta}_1$ é não-viesado. Já sabemos isso da análise de regressão simples do Capítulo 2. O segundo caso é mais interessante. Se $\tilde{\delta}_1 = 0$, então $\tilde{\beta}_1$ é não-viesado para β_1 , mesmo se $\beta_2 \neq 0$.

Como $\tilde{\delta}_1$ é a covariância amostral entre x_1 e x_2 sobre a variância amostral de x_1 , $\tilde{\delta}_1 = 0$ se, e somente se, x_1 e x_2 forem não-correlacionados na amostra. Assim, temos a importante conclusão de que, se x_1 e x_2 forem não-correlacionados na amostra, então $\tilde{\beta}_1$ é não-viesado. Isso não é surpreendente: na Seção 3.2, mostramos que o estimador da regressão simples $\tilde{\beta}_1$ e o estimador da regressão múltipla $\hat{\beta}_1$ são iguais quando x_1 e x_2 forem não-correlacionados na amostra. [Podemos também mostrar que $\tilde{\beta}_1$ é não-viesado sem condicionar a x_{i2} se $E(x_2|x_1) = E(x_2)$; então, para a estimação de β_1 , deixar x_2 no termo erro não viola a hipótese de média condicional zero do erro, uma vez que ajustamos o intercepto.]

Quando x_1 e x_2 forem correlacionados, $\tilde{\delta}_1$ tem o mesmo sinal da correlação entre x_1 e x_2 : $\tilde{\delta}_1 > 0$ se x_1 e x_2 forem positivamente correlacionados, e $\tilde{\delta}_1 < 0$ se x_1 e x_2 forem negativamente correlacionados. O sinal do viés em $\tilde{\beta}_1$ depende tanto do sinal de β_2 como de $\tilde{\delta}_1$ e está resumido na Tabela 3.2 para os quatro casos possíveis quando há viés. A Tabela 3.2 justifica um estudo cuidadoso. Por exemplo, o viés em $\tilde{\beta}_1$ é positivo se $\beta_2 > 0$ (x_2 tem um efeito positivo sobre y) e x_1 e x_2 são positivamente correlacionados; o viés é negativo se $\beta_2 > 0$ e x_1 e x_2 são negativamente correlacionados, e assim por diante.

Tabela 3.2

Sumário do Viés em $\tilde{\beta}_1$ quando x_2 é Omitida na Estimação da Equação (3.40)

	Corr(x_1, x_2) > 0	Corr(x_1, x_2) < 0
$\beta_2 > 0$	viés positivo	viés negativo
$\beta_2 < 0$	viés negativo	viés positivo

A Tabela 3.2 resume a direção do viés, mas o tamanho do viés também é muito importante. Um viés pequeno de qualquer dos dois sinais não precisa ser causa de preocupação. Por exemplo, se o retorno da educação formal é de 8,6% na população e o viés do estimador de MQO é de 0,1% (um décimo de um ponto percentual), então não precisaríamos ficar muito preocupados. De outro lado, um viés da ordem de três pontos percentuais seria muito mais sério. O tamanho do viés é determinado pelos tamanhos de β_2 e $\tilde{\delta}_1$.

Na prática, como β_2 é um parâmetro populacional desconhecido, não podemos estar certos se β_2 é positivo ou negativo. Entretanto, temos geralmente uma boa idéia sobre a direção do efeito parcial de x_2 sobre y . Além disso, ainda que o sinal da correlação entre x_1 e x_2 não possa ser conhecido se x_2 não é observado, em muitos casos, podemos fazer uma suposição criteriosa sobre se x_1 e x_2 são positiva ou negativamente correlacionados.

Na equação do salário (3.42), por definição, mais aptidão conduz a uma produtividade maior e, portanto, a salários maiores: $\beta_2 > 0$. Há também razões para acreditar que *educ* e *aptid* sejam positivamente correlacionados: em média, indivíduos com mais aptidão inata escolhem níveis maiores de educação formal. Assim, as estimativas de MQO da equação de regressão simples do $\text{salário} = \beta_0 + \beta_1 \text{educ} + v$ são, em média, muito grandes. Isso não significa que a estimativa obtida de nossa amostra seja enorme. Somente podemos dizer que, se coletarmos muitas amostras aleatoriamente e obtivermos as estimativas da regressão simples a cada vez, a média dessas estimativas será maior que β_1 .

EXEMPLO 3.6

(Equação do Salário Horário)

Suponha que o modelo $\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{aptid} + u$ satisfaça as hipóteses RLM.1 a RLM.4. O conjunto de dados no arquivo WAGE1.RAW não contém dados sobre aptidão, de modo que estimamos β_1 a partir da regressão simples

$$\log(\tilde{\text{salário}}) = 0,584 + 0,083 \text{ educ} \\ n = 526, R^2 = 0,186.$$

Esse é somente o resultado de uma única amostra, de modo que não podemos dizer que 0,083 é maior que β_1 ; o retorno verdadeiro da educação poderia ser menor ou maior que 8,3% (nunca saberemos com certeza). Entretanto, sabemos que a média dos estimadores de todas as amostras aleatórias seria bastante grande.

Como um segundo exemplo, suponha que, no nível fundamental do ensino, a nota média dos estudantes de um exame padronizado seja determinado por

$$\text{notmed} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{taxpob} + u,$$

em que *gasto* é o gasto público por estudante, e *taxpob* é a taxa de pobreza das crianças da escola. Usando dados do distrito da escola, temos somente observações da percentagem de estudantes com uma nota de aprovação e gastos públicos por estudante; não temos informações sobre taxas de pobreza. Assim, estimamos β_1 a partir da regressão simples de *notmed* sobre *gasto*.

Podemos obter, uma vez mais, o viés provável em $\tilde{\beta}_1$. Primeiro, β_2 é provavelmente negativo: há ampla evidência de que crianças que vivem na pobreza têm, em média, notas mais baixas em testes padronizados. Segundo, o gasto público médio por estudante é, provavelmente, negativamente correlacionado com a taxa de pobreza: quanto maior a taxa de pobreza menor o gasto público médio por estudante, de modo que $\text{Corr}(x_1, x_2) < 0$. De acordo com a Tabela 3.2, $\tilde{\beta}_1$ terá um viés positivo. Essa observação tem implicações importantes. Pode ser que o efeito verdadeiro do gasto público fosse zero: isto é, $\beta_1 = 0$. Entretanto, a estimativa de β_1 da regressão simples será, geralmente, maior que zero, e isso poderia nos levar a concluir que os gastos públicos são importantes quando eles não são.

Ao ler e ao fazer trabalhos empíricos em economia, é importante dominar a terminologia associada aos estimadores viesados. No contexto de omissão de uma variável do modelo (3.40), se $E(\tilde{\beta}_1) > \beta_1$, então dizemos que $\tilde{\beta}_1$ tem um **viés para cima**. Quando $E(\tilde{\beta}_1) < \beta_1$, $\tilde{\beta}_1$ tem um **viés para baixo**. Essas definições são as mesmas, seja β_1 positivo ou negativo. A expressão **viesado para zero** refere-se aos casos em que $E(\tilde{\beta}_1)$ está mais próxima de zero do que de β_1 . Portanto, se β_1 for positivo, $\tilde{\beta}_1$ será viesado para zero se ele tiver um viés para baixo. De outro lado, se $\beta_1 < 0$, $\tilde{\beta}_1$ será viesado para zero se ele tiver um viés para cima.

Viés de Variável Omitida: Casos mais Gerais

É mais difícil derivar o sinal do viés de variável omitida quando há múltiplos regressores no modelo estimado. Devemos lembrar que a correlação entre uma única variável explicativa e o erro resulta, geralmente, em *todos* os estimadores de MQO serem viesados. Por exemplo, suponha que o modelo populacional

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \quad (3.49)$$

satisfaça as hipóteses RLM.1 a RLM.4. No entanto, omitimos x_3 e estimamos o modelo como

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2. \quad (3.50)$$

Agora suponha que x_2 e x_3 sejam não-correlacionados, mas que x_1 é correlacionado com x_3 . Em outras palavras, x_1 é correlacionado com a variável omitida, mas x_2 não é. É tentador pensar que, embora provavelmente $\tilde{\beta}_1$ seja viesado com base na derivação da subseção anterior, $\tilde{\beta}_2$ seja não-viesado, pois x_2 é não-correlacionado com x_3 . Infelizmente, esse *não é*, geralmente, o caso: normalmente, tanto $\tilde{\beta}_1$ como $\tilde{\beta}_2$ serão viesados. A única exceção a isso ocorre quando x_1 e x_2 também são não-correlacionados.

Mesmo em um modelo razoavelmente simples como o apresentado, pode ser difícil obter a direção do viés em $\tilde{\beta}_1$ e $\tilde{\beta}_2$. Isso se deve ao fato de que x_1 , x_2 e x_3 podem estar correlacionados aos pares. Entretanto, uma aproximação é, freqüentemente, útil na prática. Se assumirmos que x_1 e x_2 são não-correlacionados, podemos estudar o viés em $\tilde{\beta}_1$ como se x_2 estivesse ausente dos modelos populacional e estimado. De fato, quando x_1 e x_2 são não-correlacionados, pode-se mostrar que

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i3}}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

Essa equação é exatamente igual a (3.46), mas β_3 substitui β_2 , e x_3 substitui x_2 . Portanto, o viés em $\tilde{\beta}_1$ é obtido ao se substituir β_2 por β_3 e x_2 por x_3 na Tabela 3.2. Se $\beta_3 > 0$ e $\text{Corr}(x_1, x_3) > 0$, o viés em $\tilde{\beta}_1$ é positivo, e assim por diante.

Como um exemplo, suponha que acrescentamos *exper* ao modelo do salário:

$$\text{salário}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{aptid}_i + u_i$$

Se *aptid* for omitida do modelo, ambos os estimadores de β_1 e β_2 serão viesados, mesmo se assumirmos que *exper* é não-correlacionado com *aptid*. Estamos principalmente interessados no retorno da educação formal, de modo que seria bom se pudéssemos concluir que $\tilde{\beta}_1$ tem um viés para cima ou para baixo devido à omissão da aptidão. Essa conclusão não é possível sem hipóteses adicionais. Como uma *aproximação*, suponhamos que, além de *exper* e *aptid* serem não-correlacionadas, *educ* e *exper* também sejam não-correlacionadas. (Na realidade, elas são negativamente correlacionadas.) Como $\beta_3 > 0$ e *educ* e *aptid* são positivamente correlacionadas, $\tilde{\beta}_1$ teria um viés para cima, exatamente como se *exper* não estivesse no modelo.

O raciocínio usado no exemplo anterior é, muitas vezes, compreendido como um guia aproximação para obter o viés provável dos estimadores em modelos mais complicados. Geralmente, o foco está

na relação entre uma variável explicativa particular, por exemplo x_1 , e o fator omitido fundamental. Estritamente falando, ignorar todas as outras variáveis explicativas é uma prática válida somente quando cada uma delas é não-correlacionada com x_1 , mas essa ainda é uma orientação útil. O Apêndice 3A (disponível no site de Thomson) contém uma análise mais cuidadosa do viés de variável omitida com múltiplas variáveis explicativas.

3.4 A VARIÂNCIA DOS ESTIMADORES DE MQO

Obteremos, agora, a variância dos estimadores de MQO, de modo que, além de conhecermos as tendências centrais dos $\hat{\beta}_j$, também teremos uma medida da dispersão de sua distribuição amostral. Antes de encontrarmos as variâncias, vamos adicionar uma hipótese de homoscedasticidade, como no Capítulo 2. Fazemos isso por duas razões. Primeira, ao impor a hipótese de variância constante do erro, as fórmulas são simplificadas. Segunda, veremos na Seção 3.5 que MQO tem uma propriedade importante de eficiência se acrescentamos a hipótese de homoscedasticidade.

No arcabouço da regressão múltipla, a homoscedasticidade é expressa como a seguir:

HIPÓTESE RLM.5 (HOMOSCEDASTICIDADE)

$$\text{Var}(u|x_1, \dots, x_k) = \sigma^2.$$

A hipótese RLM.5 significa que a variância do termo erro, u , condicionada às variáveis explicativas, é a *mesma* para todas as combinações de resultados das variáveis explicativas. Se essa hipótese é violada, o modelo exibe heteroscedasticidade, exatamente como no caso de duas variáveis.

Na equação

$$\text{salário}_i = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u,$$

a homoscedasticidade requer que a variância do erro não-observado u não dependa dos níveis de educação, experiência ou permanência. Isto é,

$$\text{Var}(u|\text{educ}, \text{exper}, \text{perm}) = \sigma^2.$$

Se a variância varia com qualquer uma das três variáveis explicativas, então a heteroscedasticidade está presente.

As hipóteses RLM.1 a RLM.5 são, em conjunto, conhecidas como as **hipóteses de Gauss-Markov** (para a regressão de corte transversal). Até agora, nossas asserções sobre as hipóteses são adequadas somente quando aplicadas à análise de corte transversal com amostragem aleatória. Como veremos, as hipóteses de Gauss-Markov para a análise de séries de tempo — e para outras situações, como a análise de dados de painel — são mais difíceis de se manterem, embora haja muitas similaridades.

Na discussão a seguir, usaremos o símbolo \mathbf{x} para representar o conjunto de todas as variáveis independentes, (x_1, \dots, x_k) . Assim, na regressão do salário horário com *educ*, *exper* e *perm* como variáveis independentes, $\mathbf{x} = (\text{educ}, \text{exper}, \text{perm})$. Conseqüentemente, podemos escrever as hipóteses RLM.1 e RLM.3 como

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_3 x_3,$$

e a hipótese RLM.5 é a mesma que $\text{Var}(y|x) = \sigma^2$. Expressar as hipóteses desse modo ilustra como a hipótese RLM.5 difere muito da hipótese RLM.3. Esta diz que o valor esperado de y , dado \mathbf{x} , é linear nos parâmetros, mas ele certamente depende de x_1, x_2, \dots, x_k . A hipótese RLM.5 diz que a variância de y , dado \mathbf{x} , não depende dos valores das variáveis independentes.

Podemos obter, agora, as variâncias dos $\hat{\beta}_j$, que uma vez mais, estão condicionadas aos valores amostrais das variáveis independentes. A prova está no apêndice deste capítulo.

T E O R E M A 3 . 2 (VARIÂNCIAS AMOSTRAIS DOS ESTIMADORES DE INCLINAÇÃO DE MQO)

Sob as hipóteses RLM.1 a RLM.5, condicionadas aos valores amostrais das variáveis independentes,

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}, \quad (3.51)$$

para $j = 1, 2, \dots, k$, em que $\text{SQT}_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ é a variação amostral total em x_j , e R_j^2 é o R -quadrado da regressão x_j sobre todas as outras variáveis independentes (incluindo um intercepto).

Antes de estudarmos a equação (3.51) mais detalhadamente, é importante saber que todas as hipóteses de Gauss-Markov são usadas na obtenção dessa fórmula. Embora não necessitemos da hipótese de homoscedasticidade para concluir que MQO é não-viesado, precisamos dela para validar a equação (3.51).

O tamanho de $\text{Var}(\hat{\beta}_j)$ é importante na prática. Uma variância maior significa um estimador menos preciso, e isso se traduz em intervalos de confiança maiores e testes de hipóteses menos acurados (como veremos no Capítulo 4). Na próxima subseção, discutiremos os elementos que compreendem (3.51).

Os Componentes das Variâncias de MQO: Multicolinearidade

A equação (3.51) mostra que a variância de $\hat{\beta}_j$ depende de três fatores: σ^2 , SQT_j e R_j^2 . Lembre-se de que o índice j representa simplesmente qualquer uma das variáveis independentes (como a educação ou a taxa de pobreza). Agora, vamos considerar cada um dos fatores que afetam $\text{Var}(\hat{\beta}_j)$.

A VARIÂNCIA DO ERRO, σ^2 . Da equação (3.51), um σ^2 maior significa variâncias maiores dos estimadores de MQO. Isso não é totalmente surpreendente: mais “ruído” na equação (um σ^2 maior) torna mais difícil estimar o efeito parcial de qualquer uma das variáveis independentes sobre y , e isso é refletido nas variâncias maiores dos estimadores de inclinação de MQO. Visto que σ^2 é uma característica da população, ele não tem nada a ver com o tamanho da amostra. Ele é o componente de (3.51) que é desconhecido. Veremos mais adiante como obter um estimador não-viesado de σ^2 .

Para uma dada variável dependente y , há de fato somente uma maneira de reduzir a variância do erro, que é adicionar mais variáveis explicativas à equação (retirar alguns fatores do termo erro). Isso nem sempre é possível, nem sempre é desejável, por razões discutidas posteriormente neste capítulo.

A VARIAÇÃO AMOSTRAL TOTAL EM x_j , SQT_j . Da equação (3.51), vemos que quanto maior a variação total em x_j , menor é $\text{Var}(\hat{\beta}_j)$. Assim, tudo o mais sendo igual para estimar $\hat{\beta}_j$ preferimos ter tanta variação amostral em x_j quanto possível. Já descobrimos isso no caso da regressão simples, no Capítulo 2. Embora raramente seja possível escolher os valores amostrais das variáveis independentes, há uma maneira de aumentar a variação amostral em cada uma das variáveis independentes: aumentar o tamanho da amostra. De fato, na amostragem aleatória de uma população, SQT_j aumenta sem limite quando o tamanho da amostra torna-se maior. Esse é o componente da variância que depende sistematicamente do tamanho da amostra.

Quando SQT_j é pequeno, $\text{Var}(\hat{\beta}_j)$ pode ficar muito grande, mas um SQT_j pequeno não é uma violação da hipótese RLM.4. Tecnicamente, quando SQT_j tende a zero, $\text{Var}(\hat{\beta}_j)$ aproxima-se do infinito. O caso extremo de nenhuma variação amostral em x_j , $SQT_j = 0$, não é permitido pela hipótese RLM.4.

AS RELAÇÕES LINEARES ENTRE AS VARIÁVEIS INDEPENDENTES, R_j^2 . O termo R_j^2 na equação (3.51) é o mais difícil dos três componentes de se entender. Esse termo não aparece na análise de regressão simples porque há somente uma variável independente em tal caso. É importante compreender que esse R -quadrado é distinto do R -quadrado da regressão de y sobre x_1, x_2, \dots, x_k : R_j^2 é obtido de uma regressão que envolve somente as variáveis independentes do modelo original, em que x_j desempenha o papel de uma variável dependente.

Considere, primeiro, o caso $k = 2$: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Então, $\text{Var}(\hat{\beta}_1) = \sigma^2 / [SQT_1(1 - R_1^2)]$, em que R_1^2 é o R -quadrado da regressão simples de x_1 sobre x_2 (e um intercepto, como sempre). Como o R -quadrado mede o grau de ajuste, um valor de R_1^2 próximo de um indica que x_2 explica bastante da variação de x_1 na amostra. Isso significa que x_1 e x_2 são altamente correlacionados.

Quando R_1^2 cresce em direção a um, $\text{Var}(\hat{\beta}_1)$ torna-se maior. Assim, um grau elevado de relação linear entre x_1 e x_2 pode levar a variâncias grandes dos estimadores de inclinação de MQO. (Um argumento similar se aplica a $\hat{\beta}_2$.) Veja a Figura 3.1 para a relação entre $\text{Var}(\hat{\beta}_1)$ e o R -quadrado da regressão de x_1 sobre x_2 .

No caso geral, R_j^2 é a proporção da variação total de x_j que pode ser explicada pelas outras variáveis independentes que aparecem na equação. Para dados σ^2 e SQT_j , a menor $\text{Var}(\hat{\beta}_j)$ é obtida quando $R_j^2 = 0$, que ocorre se, e somente se, x_j tem correlação amostral zero com cada uma das outras variáveis independentes. Esse é o melhor caso para estimar β_j , mas é raramente encontrado.

O outro caso extremo, $R_j^2 = 1$, é excluído pela hipótese RLM.4, pois $R_j^2 = 1$ significa que, na amostra, x_j é uma combinação linear perfeita de algumas das outras variáveis independentes da regressão. Um caso mais relevante é quando R_j^2 está “próximo” de um. Da equação (3.51) e da Figura 3.1, vemos que isso pode fazer com que $\text{Var}(\hat{\beta}_j)$ seja grande: $\text{Var}(\hat{\beta}_j) \rightarrow \infty$ quando $R_j^2 \rightarrow 1$. Correlação alta (mas não perfeita) entre duas ou mais variáveis independentes é chamada **multicolinearidade**.

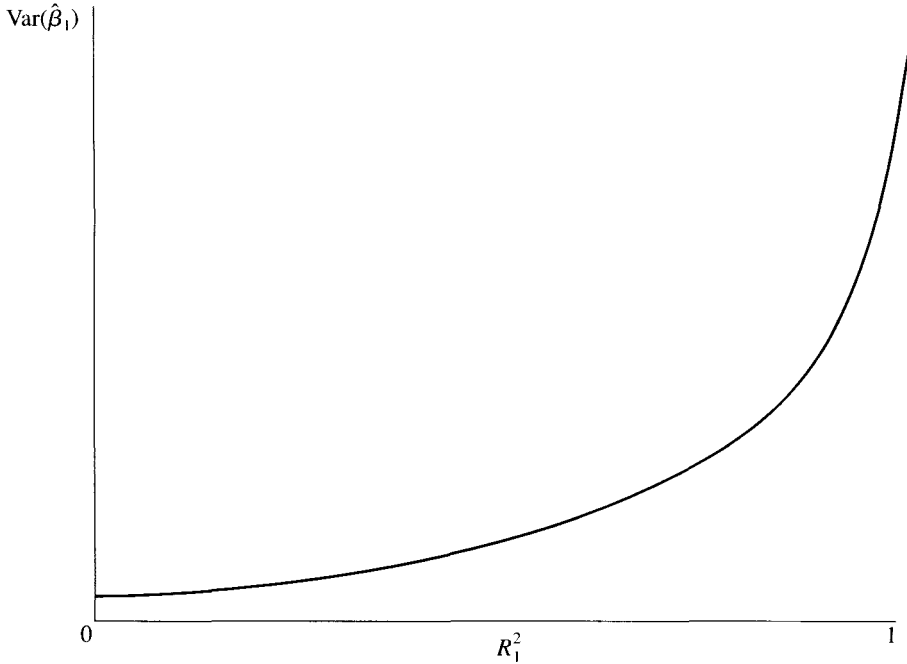
Antes de discutirmos mais a questão da multicolinearidade, é importante que uma coisa esteja bem clara: um caso em que R_j^2 está próximo de um não é uma violação da hipótese RLM.4.

Como a multicolinearidade não viola nenhuma de nossas hipóteses, o “problema” da multicolinearidade não é, de fato, bem definido. Ao dizer que a multicolinearidade surge ao estimarmos β_j , quando R_j^2 está “próximo” de um, colocamos “próximo” dentro de aspas porque não há um número absoluto que podemos citar para concluir que a multicolinearidade é um problema. Por exemplo, $R_j^2 = 0,9$ significa que 90% da variação amostral em x_j pode ser explicada pelas outras variáveis independentes do modelo de regressão. Inquestionavelmente, isso significa que x_j tem uma forte relação linear com as outras variáveis independentes. No entanto, se isso se traduz em uma $\text{Var}(\hat{\beta}_j)$ que é grande demais para ser útil, depende dos tamanhos de σ^2 e SQT_j . Como veremos no Capítulo

4, para a inferência estatística, o que essencialmente importa é quanto $\hat{\beta}_j$ é grande com relação a seu desvio-padrão.

Figura 3.1

$\text{Var}(\hat{\beta}_j)$ como uma função de R_j^2 .



Assim como um valor grande de R_j^2 pode causar uma $\text{Var}(\hat{\beta}_j)$ grande, um valor pequeno de SQT_j também pode fazer com que $\text{Var}(\hat{\beta}_j)$ seja grande. Portanto, um tamanho pequeno da amostra pode também levar a variâncias amostrais grandes. Preocupar-se com graus elevados de correlação entre variáveis independentes da amostra não é, de fato, diferente de se preocupar com um tamanho pequeno da amostra: ambos funcionam para aumentar $\text{Var}(\hat{\beta}_j)$.

O famoso econometrista da Universidade de Wisconsin Arthur Goldberger, reagindo à obsessão dos econométricos pela multicolinearidade, criou (jocosamente) o termo **micronumerosidade**, que ele define como o “problema do tamanho pequeno da amostra”. [Para uma discussão interessante sobre multicolinearidade e micronumerosidade, veja Goldberger (1991).]

Embora o problema da multicolinearidade não possa ser claramente definido, uma coisa é clara: tudo mais sendo igual, para estimar β_j , é melhor ter menos correlação entre x_j e as outras variáveis independentes. Essa observação muitas vezes leva a uma discussão de como “resolver” o problema da multicolinearidade. Nas ciências sociais, em que somos geralmente coletores passivos de dados, não há uma boa maneira de reduzir as variâncias dos estimadores não-viesados que não seja coletar mais dados. Para um determinado conjunto de dados, podemos tentar, num esforço para reduzir a multicolinearidade, suprimir outras variáveis independentes do modelo. Infelizmente, suprimir uma variável que pertence ao modelo populacional pode levar viés, como vimos na Seção 3.3.

Neste ponto, talvez um exemplo ajude a esclarecer algumas das questões aqui levantadas relativas à multicolinearidade. Suponha que estamos interessados em estimar o efeito de várias categorias de

despesas de escolas sobre o desempenho de estudantes. É provável que as despesas com salários de professores, materiais institucionais, atletismo etc. estejam altamente correlacionadas: escolas mais ricas tendem a gastar mais com tudo, e escolas mais pobres gastam menos com tudo. Não surpreendentemente, pode ser difícil estimar o efeito de qualquer categoria de despesa específica sobre o desempenho dos estudantes quando há pouca variação em uma categoria que não pode ser, em grande medida, explicada por variações das outras categorias de despesas (isso leva a um R_j^2 alto para cada uma das variáveis de despesas). Esses problemas de multicolinearidade podem ser mitigados ao coletar mais dados mas assim, em certo sentido, nós mesmos nos impusemos o problema: estamos formulando questões que podem ser sutis demais para que os dados disponíveis as respondam com alguma precisão. Provavelmente, podemos fazer algo muito melhor mudando o escopo da análise e agrupando todas as categorias de despesa em uma única categoria, desde que não mais estivéssemos tentando estimar o efeito parcial de cada categoria separadamente.

Outro ponto importante é que um elevado grau de correlação entre certas variáveis independentes pode ser irrelevante no que diz respeito a quão bem podemos estimar outros parâmetros do modelo. Por exemplo, considere um modelo com três variáveis independentes:

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

em que x_2 e x_3 são altamente correlacionados. Então, $\text{Var}(\hat{\beta}_2)$ e $\text{Var}(\hat{\beta}_3)$ podem ser grandes. Mas o valor da correlação entre x_2 e x_3 não tem efeito direto sobre $\text{Var}(\hat{\beta}_1)$. De fato, se x_1 é não-correlacionado com x_2 e x_3 , então $R_1^2 = 0$ e $\text{Var}(\hat{\beta}_1) = \sigma^2/\text{SQT}_1$, independentemente da quantia de correlação existir entre x_2 e x_3 . Se β_1 é o parâmetro de interesse, realmente não devemos nos preocupar com o valor da correlação entre x_2 e x_3 .

Suponha que você postula um modelo que explica a nota do exame final em termos da frequência às aulas. Assim, a variável dependente é a nota do exame final, e a principal variável explicativa é o número de aulas frequentadas. A fim de controlar as aptidões dos estudantes e pelos esforços fora da sala de aula, você inclui entre as variáveis explicativas a nota acumulada durante todo o curso, a nota do teste de avaliação de conhecimentos para ingresso em curso superior e as medidas do desempenho do estudante no ensino médio. Alguém diz: "Você não pode esperar aprender nada com esse exercício, pois todas essas variáveis são, provavelmente, altamente colineares". Qual seria sua resposta?

A observação anterior é importante porque os economistas frequentemente incluem muitas variáveis de controle a fim de isolar o efeito causal de uma variável particular. Por exemplo, ao olhar para a relação entre as taxas de aprovação de empréstimos e a percentagem de minorias em uma região, poderíamos incluir variáveis como renda média, valor médio das moradias, medidas de inadimplência, e assim por diante, pois esses fatores precisam ser considerados a fim de se extrair conclusões causais sobre a discriminação. Renda, preços de moradia e inadimplência são, geralmente, altamente correlacionados entre si. No entanto, correlações altas entre essas variáveis não tornam mais difícil determinar os efeitos da discriminação.

Variâncias em Modelos Mal Especificados

A escolha de incluir ou não uma variável particular em um modelo de regressão pode ser feita ao analisar o dilema entre viés e variância. Na Seção 3.3 derivamos o viés produzido pela omissão de uma variável relevante quando o modelo verdadeiro contém duas variáveis explicativas. Vamos continuar a análise desse modelo comparando as variâncias dos estimadores de MQO.

Escreva o modelo populacional verdadeiro que satisfaz as hipóteses de Gauss-Markov, como

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

Consideremos dois estimadores de β_1 . O estimador $\hat{\beta}_1$ é proveniente da regressão múltipla

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \quad (3.52)$$

Em outras palavras, incluímos x_2 , juntamente com x_1 , no modelo de regressão. O estimador $\tilde{\beta}_1$ é obtido ao omitir x_2 do modelo e computando uma regressão simples de y sobre x_1 :

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1. \quad (3.53)$$

Quando $\beta_2 \neq 0$, a equação (3.53) exclui uma variável relevante do modelo e, como vimos na Seção 3.3, isso induz um viés em $\tilde{\beta}_1$, a não ser que x_1 e x_2 sejam não-correlacionados. De outro lado, $\hat{\beta}_1$ é não-viesado para β_1 , para qualquer valor de β_2 , incluindo $\beta_2 = 0$. Segue-se que, se o viés for usado como único critério, $\hat{\beta}_1$ é preferível a $\tilde{\beta}_1$.

A conclusão de que $\hat{\beta}_1$ é sempre preferível a $\tilde{\beta}_1$ não se sustenta quando trazemos a variância para dentro da análise. Condicionando aos valores de x_1 e x_2 na amostra, temos, de (3.51),

$$\text{Var}(\hat{\beta}_1) = \sigma^2 / [\text{SQT}_1(1 - R_1^2)], \quad (3.54)$$

em que SQT_1 é a variação total em x_1 , e R_1^2 é o R -quadrado da regressão de x_1 sobre x_2 . Além disso, uma simples modificação da prova para a regressão de duas variáveis do Capítulo 2 mostra que

$$\text{Var}(\tilde{\beta}_1) = \sigma^2 / \text{SQT}_1. \quad (3.55)$$

Comparar (3.55) a (3.54) mostra que $\text{Var}(\tilde{\beta}_1)$ é sempre *menor* que $\text{Var}(\hat{\beta}_1)$, a menos que x_1 e x_2 sejam não-correlacionados na amostra, caso em que os dois estimadores $\tilde{\beta}_1$ e $\hat{\beta}_1$ são os mesmos. Ao assumir que x_1 e x_2 são não-correlacionados, podemos ter as seguintes conclusões:

1. Quando $\beta_2 \neq 0$, $\tilde{\beta}_1$ é viesado, $\hat{\beta}_1$ é não-viesado e $\text{Var}(\tilde{\beta}_1) < \text{Var}(\hat{\beta}_1)$.
2. Quando $\beta_2 = 0$, $\tilde{\beta}_1$ e $\hat{\beta}_1$ são ambos não-viesados e $\text{Var}(\tilde{\beta}_1) < \text{Var}(\hat{\beta}_1)$.

Da segunda conclusão, é claro que $\tilde{\beta}_1$ é preferido se $\beta_2 = 0$. Intuitivamente, se x_2 não tem um efeito parcial sobre y , incluí-lo no modelo pode somente exacerbar o problema da multicolinearidade, o que

leva a um estimador menos eficiente de β_1 . O custo de incluir uma variável irrelevante no modelo é uma variância maior do estimador de β_1 .

O caso em que $\beta_2 \neq 0$ é mais difícil. Omitindo x_2 do modelo leva a um estimador viesado de β_1 . Tradicionalmente, econométristas têm sugerido comparar o tamanho provável do viés devido à omissão de x_2 com a redução na variância — resumida no tamanho de R_1^2 — para decidir se x_2 deve ser incluído. Entretanto, quando $\beta_2 \neq 0$, há duas razões favoráveis para incluir x_2 no modelo. A mais importante delas é que qualquer viés em $\hat{\beta}_1$ não diminui quando o tamanho da amostra cresce; de fato, o viés não segue, necessariamente, qualquer padrão. Portanto, podemos em geral pensar o viés como mais ou menos o mesmo para qualquer tamanho de amostra. De outro lado, $\text{Var}(\hat{\beta}_1)$ e $\text{Var}(\tilde{\beta}_1)$ tendem a zero quando n torna-se grande, o que significa que a multicolinearidade induzida pela adição de x_2 torna-se menos importante quando o tamanho da amostra cresce. Em amostras grandes, preferiríamos $\hat{\beta}_1$.

A outra razão para preferir $\hat{\beta}_1$ é mais sutil. A fórmula da variância em (3.55) está condicionada aos valores de x_1 e x_2 na amostra, o que oferece o melhor cenário para $\tilde{\beta}_1$. Quando $\beta_2 \neq 0$, a variância de $\tilde{\beta}_1$ condicionada somente a x_1 é maior que aquela apresentada em (3.55). Intuitivamente, quando $\beta_2 \neq 0$ e x_2 é excluído do modelo, a variância do erro aumenta porque o erro efetivamente contém parte de x_2 . Mas, (3.55) ignora o aumento da variância do erro porque ela trata ambos os regressores como não-aleatórios. Uma discussão completa de quais variáveis independentes deveriam ser condicionadas nos desviaria demais de nosso caminho. É suficiente dizer que (3.55) é bastante generosa quando ela aparece para medir a precisão de $\tilde{\beta}_1$.

Estimação de σ^2 : Os Erros-Padrão dos Estimadores de MQO

Vamos, agora, mostrar como escolher um estimador não-viesado de σ^2 , que nos permitirá obter estimadores não-viesados de $\text{Var}(\hat{\beta}_j)$.

Como $\sigma^2 = E(u^2)$, um “estimador” não-viesado de σ^2 é a média amostral dos erros quadrados: $n^{-1} \sum_{i=1}^n u_i^2$. Infelizmente, esse não é um estimador verdadeiro, pois não observamos os u_i . Não obstante, lembre-se de que os erros podem ser escritos como $u_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik}$, e a razão real de não observarmos os u_i é que não conhecemos os β_j . Quando substituímos cada β_j por seu estimador de MQO, obtemos os resíduos de MQO:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik}.$$

Parece natural estimar σ^2 ao substituir u_i por \hat{u}_i . No caso da regressão simples, vimos que isso leva a um estimador viesado. O estimador não-viesado de σ^2 no caso geral da regressão múltipla é

$$\hat{\sigma}^2 = \left(\sum_{i=1}^n \hat{u}_i^2 \right) / (n - k - 1) = \text{SQR} / (n - k - 1). \quad (3.56)$$

Já encontramos esse estimador no caso $k = 1$ da regressão simples.

O termo $n - k - 1$ em (3.56) representa os **graus de liberdade (gl)** do problema geral de MQO com n observações e k variáveis independentes. Como há $k + 1$ parâmetros em um modelo de regressão com k variáveis independentes e um intercepto, podemos escrever

$$\begin{aligned}
 gl &= n - (k + 1) \\
 &= (\text{número de observações}) - (\text{número de parâmetros estimados})
 \end{aligned}
 \tag{3.57}$$

Essa é a maneira mais fácil de calcular os graus de liberdade em uma aplicação particular: conte o número de parâmetros, incluindo o intercepto, e subtraia esse valor do número de observações. (No caso raro em que o intercepto não é estimado, o número de parâmetros diminui em um.)

Tecnicamente, a divisão por $n - k - 1$ em (3.56) é proveniente do fato de que o valor esperado da soma dos resíduos quadrados é $E(\text{SQR}) = (n - k - 1) \sigma^2$. Intuitivamente, podemos entender porque o ajustamento de graus de liberdade é necessário ao retornarmos às condições de primeira ordem dos estimadores de MQO. Elas podem ser escritas como $\sum_{i=1}^n u_i = 0$ e $\sum_{i=1}^n x_{ij} \hat{u}_i = 0$, onde $j = 1, 2, \dots, k$. Assim, na obtenção dos estimadores de MQO, $k + 1$ restrições são impostas sobre os resíduos de MQO. Isso significa que, dados $n - (k + 1)$ dos resíduos, os $k + 1$ resíduos restantes são conhecidos: há somente $n - (k + 1)$ graus de liberdade nos resíduos. (Isso pode ser contrastado com os erros u_i , os quais têm n graus de liberdade na amostra.)

Para referência, vamos resumir essa discussão com o Teorema 3.3. Provamos esse teorema para o caso da análise de regressão simples no Capítulo 2 (veja o Teorema 2.3). (No Apêndice E, disponível no site de Thomson, é dada uma prova geral, que requer álgebra matricial.)

TEOREMA 3.3 (ESTIMADOR NÃO-VIESADO DE σ^2)

Sob as hipóteses de Gauss-Markov RLM.1 a RLM.5, $E(\hat{\sigma}^2) = \sigma^2$.

A raiz quadrada positiva de $\hat{\sigma}^2$, denominada $\hat{\sigma}$, é chamada **Erro-Padrão da Regressão (EPR)**. O EPR é um estimador do desvio-padrão do termo erro. Essa estimativa é usualmente informada pelos programas de regressão, embora ela seja chamada de nomes diferentes pelos diferentes programas. (Além de EPR, $\hat{\sigma}$ também é chamado *erro-padrão da estimativa* e a *raiz do erro quadrado médio*.)

Observe que $\hat{\sigma}$ pode diminuir ou aumentar quando outra variável independente é acrescentada a uma regressão (para uma dada amostra). Isso ocorre pois, embora SQR deva cair quando outra variável explicativa é adicionada, os graus de liberdade também diminuem em um. Como SQR está no numerador e gl está no denominador, não podemos dizer, de antemão, qual efeito prevalecerá.

Para construir intervalos de confiança e conduzir testes no Capítulo 4, precisaremos estimar o **desvio-padrão de $\hat{\beta}_j$** , que é exatamente a raiz quadrada da variância:

$$dp(\hat{\beta}_j) = \sigma / [\text{SQT}_j(1 - R_j^2)]^{1/2}.$$

Como σ é desconhecido, ele é substituído pelo seu estimador, $\hat{\sigma}$. Isso nos dá o **erro-padrão de $\hat{\beta}_j$** :

$$ep(\hat{\beta}_j) = \hat{\sigma} / [\text{SQT}_j(1 - R_j^2)]^{1/2}. \tag{3.58}$$

Assim como as estimativas de MQO podem ser obtidas para qualquer amostra dada, os erros-padrão também podem. Como $ep(\hat{\beta}_j)$ depende de $\hat{\sigma}$, o erro-padrão tem uma distribuição amostral, que desempenhará um papel importante no Capítulo 4.

Devemos enfatizar algo sobre os erros-padrão. Como (3.58) é obtido diretamente da fórmula da variância em (3.51), e como (3.51) se apóia na hipótese de homoscedasticidade RLM.5, a fórmula do erro-padrão em (3.58) *não* é um estimador válido de $\text{dp}(\hat{\beta}_j)$ se os erros exibem heteroscedasticidade. Assim, enquanto a presença de heteroscedasticidade não causa viés em $\hat{\beta}_j$, ela leva viés da fórmula usual da $\text{Var}(\hat{\beta}_j)$, o que invalida, portanto, os erros-padrão. Isso é importante porque qualquer programa de regressão calcula (3.58) como o erro-padrão básico de cada coeficiente (com uma interpretação um pouco diferente para o intercepto). Se suspeitarmos de heteroscedasticidade, então os erros-padrão de MQO “habituais” não são válidos, e alguma ação corretiva deve ser tomada. No Capítulo 8, veremos quais métodos estão disponíveis para trabalhar com a heteroscedasticidade.

3.5 EFICIÊNCIA DE MQO: O TEOREMA DE GAUSS-MARKOV

Nesta seção, apresentaremos e discutiremos o importante **Teorema de Gauss-Markov**, que justifica o uso do método de MQO em vez de usar uma variedade de estimadores concorrentes. Já conhecemos uma justificativa para MQO: sob as hipóteses RLM.1 a RLM.4, MQO é não-viesado. Entretanto, há *muitos* estimadores não-viesados de β_j sob essas hipóteses (por exemplo, veja o Problema 3.12). Poderia haver outros estimadores não-viesados com variâncias menores que as dos estimadores de MQO?

Se limitarmos apropriadamente a classe de estimadores concorrentes, podemos mostrar que MQO é o melhor dentro de sua classe. Especificamente, argumentamos que, sob as hipóteses RLM.1 a RLM.5, o estimador de MQO $\hat{\beta}_j$ para β_j é o **melhor estimador linear não-viesado** (*Best Linear Unbiased Estimator* — *BLUE*). A fim de formular o teorema, precisamos entender cada componente da sigla “BLUE”. Primeiro, sabemos o que é um estimador: ele é uma regra que pode ser aplicada a qualquer amostra de dados para produzir uma estimativa. Também sabemos o que é um estimador não-viesado: no contexto corrente, um estimador, por exemplo $\tilde{\beta}_j$, de β_j é um estimador não-viesado de β_j se $E(\tilde{\beta}_j) = \beta_j$ para qualquer $\beta_0, \beta_1, \dots, \beta_k$.

E o que dizer sobre o significado do termo “linear”? No contexto atual, um estimador $\tilde{\beta}_j$ de β_j é linear se, e somente se, ele puder ser expresso como uma função linear dos dados da variável dependente:

$$\tilde{\beta}_j = \sum_{i=1}^n w_{ij} y_i, \quad (3.59)$$

em que cada w_{ij} pode ser uma função dos valores amostrais de todas as variáveis independentes. Os estimadores de MQO são lineares, como pode ser visto na equação (3.22).

Finalmente, como definir “o melhor”? Para o teorema corrente, o melhor é definido como a *variância menor*. Dados dois estimadores não-viesados, é lógico preferir aquele com a variância menor (veja o Apêndice C, disponível no site de Thomson).

Agora, vamos chamar de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ os estimadores de MQO do modelo (3.31) sob as hipóteses RLM.1 a RLM.5. O Teorema de Gauss-Markov diz que, para qualquer estimador $\tilde{\beta}_j$ que é linear e não-viesado, $\text{Var}(\tilde{\beta}_j) \geq \text{Var}(\hat{\beta}_j)$, e a desigualdade é geralmente estrita. Em outras palavras, na classe dos estimadores lineares não-viesados, MQO tem a menor variância (sob as cinco hipóteses de Gauss-Markov). De fato, o teorema diz mais do que isso. Se desejarmos estimar qualquer função linear de β_j , a combinação linear correspondente dos estimadores de MQO alcança a menor variância entre todos os estimadores não-viesados. Vamos concluir com um teorema, o provado no Apêndice 3A, disponível no site da Thomson.

TEOREMA 3.4 (TEOREMA DE GAUSS-MARKOV)

Sob as hipóteses RLM.1 a RLM.5, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ são os melhores estimadores lineares não-viesados (BLUEs) de $\beta_0, \beta_1, \dots, \beta_k$, respectivamente.

É por causa desse teorema que as hipóteses RLM.1 a RLM.5 são conhecidas como as hipóteses de Gauss-Markov (para a análise de corte transversal).

A importância do Teorema de Gauss-Markov é que, quando um conjunto padrão de hipóteses se mantém, não necessitamos procurar por estimadores não-viesados alternativos da forma expressa em (3.59): nenhum será melhor que MQO. Equivalentemente, se somos apresentados a um estimador que é tanto linear como não-viesado, então sabemos que a variância desse estimador é pelo menos tão grande quanto a variância de MQO; nenhum cálculo adicional é necessário para mostrar isso.

Para nossos propósitos, o Teorema 3.4 justifica o uso de MQO para estimar modelos de regressão múltipla. Se qualquer uma das hipóteses de Gauss-Markov for violada, o teorema não é mais válido. Já sabemos que a falha da hipótese de média condicional zero (hipótese RLM.3) faz com que MQO seja viesado, de modo que o Teorema 3.4 também deixa de ser válido. Também sabemos que a heteroscedasticidade (falha da hipótese RLM.5) não faz com que MQO seja viesado. Entretanto, MQO não tem mais a menor variância entre os estimadores lineares não-viesados na presença da heteroscedasticidade. No Capítulo 8, analisaremos um estimador que melhora MQO quando conhecemos o tipo da heteroscedasticidade.

1. O modelo de regressão múltipla nos permite, efetivamente, manter os outros fatores fixos ao examinarmos os efeitos de uma variável independente particular sobre a variável dependente. Ele permite, explicitamente, que as variáveis sejam correlacionadas.
2. Embora o modelo seja linear em seus *parâmetros*, ele pode ser usado para modelar relações não-lineares ao se escolher, apropriadamente, as variáveis dependente e independente.
3. O método de mínimos quadrados ordinários é facilmente aplicado para estimar o modelo de regressão múltipla. Cada estimativa de inclinação mede o efeito parcial da variável independente correspondente sobre a variável dependente, mantendo todas as outras variáveis independentes fixas.
4. R^2 é a proporção da variação amostral da variável dependente explicada pelas variáveis independentes, e é utilizado como uma medida do grau de ajuste. É importante não dar importância demais ao valor do R^2 na avaliação de modelos econométricos.
5. Sob as primeiras quatro hipóteses de Gauss-Markov (RLM.1 a RLM.4), os estimadores de MQO são não-viesados. Isso implica que incluir uma variável irrelevante em um modelo não tem nenhum efeito sobre a inexistência de viés dos estimadores de intercepto e de inclinação. De outro lado, omitir uma variável importante faz com que MQO seja viesado. Em muitas circunstâncias, a direção do viés pode ser determinada.
6. Sob as cinco hipóteses de Gauss-Markov, a variância de um estimador de inclinação de MQO é dada por $\text{Var}(\hat{\beta}_j) = \sigma^2 / [\text{SQT}_j(1 - R_j^2)]$. Quando a variância do erro σ^2 cresce, o mesmo ocorre com $\text{Var}(\hat{\beta}_j)$, enquanto $\text{Var}(\hat{\beta}_j)$ diminui quando a variação amostral em x_j , SQT_j , aumenta. O termo R_j^2 mede a magnitude da colinearidade entre x_j e as outras variáveis explicativas. Quando R_j^2 aproxima-se de um, $\text{Var}(\hat{\beta}_j)$ é ilimitada.

7. Adicionar uma variável irrelevante a uma equação geralmente aumenta as variâncias dos demais estimadores de MQO, por causa da multicolinearidade.
8. Sob as hipóteses de Gauss-Markov (RLM.1 a RLM.5), os estimadores de MQO são os melhores estimadores lineares não-viesados (BLUE).

3.1 Usando os dados do arquivo GPA2.RAW sobre 4.137 estudantes de curso superior nos Estados Unidos, estimou-se a seguinte equação por MQO:

$$nmgrad = 1,392 - 0,0135 emperc + 0,00148 sat$$

$$n = 4.137, R^2 = 0,273,$$

em que *nmgrad* é mensurada em uma escala de quatro pontos, *emperc* é o percentil da turma de formados do ensino médio (definido de modo que, por exemplo, *emperc* = 5 significa os cinco por cento melhores da sala), e *sat* é uma nota média ponderada de matemática e habilidade verbal do estudante para ingresso em curso superior.

- (i) Por que faz sentido que o coeficiente de *emperc* seja negativo?
- (ii) Qual é o valor previsto de *nmgrad* quando *emperc* = 20 e *sat* = 1.050?
- (iii) Suponha que dois alunos do ensino médio, A e B, estejam no mesmo percentil no ensino médio, mas a nota *sat* do Estudante A foi 140 pontos maior (cerca de um desvio-padrão na amostra). Qual é a diferença prevista em *nmgrad* para esses dois estudantes? A diferença é grande?
- (iv) Mantendo *emperc* fixo, que diferença na nota *sat* levaria a uma diferença prevista de *nmgrad* de 0,50? Comente sua resposta.

3.2 Os dados do arquivo WAGE2.RAW, sobre homens que trabalham, foram utilizados para estimar a seguinte equação:

$$educ = 10,36 - 0,094 irms + 0,131 educm + 0,210 educp$$

$$n = 722, R^2 = 0,214,$$

em que *educ* é anos de escolaridade formal, *irms* é o número de irmãos, *educm* é anos de escolaridade formal da mãe e *educp* é anos de escolaridade formal do pai.

- (i) *irms* tem o efeito esperado? Explique. Mantendo *educm* e *educp* fixos, em quanto deveria *irms* aumentar para reduzir os anos previstos da educação formal em um ano? (Uma resposta incompleta é aceitável aqui.)
- (ii) Discuta a interpretação do coeficiente de *educm*.
- (iii) Suponha que o Homem A não tenha irmãos, e sua mãe e seu pai tenham, cada um, 12 anos de educação formal. Suponha também que o Homem B não tenha irmãos, e sua mãe e seu pai tenham, cada um, 16 anos de educação formal. Qual é a diferença prevista em anos de educação formal entre B e A?

3.3 O modelo seguinte é uma versão simplificada do modelo de regressão múltipla usado por Biddle e Hamermesh (1990) para estudar a escolha entre o tempo gasto dormindo e trabalhando e para observar outros fatores que afetam o sono:

$$dormir = \beta_0 + \beta_1 trabtot + \beta_2 educ + \beta_3 idade + u,$$

em que *dormir* e *trabtot* (trabalho total) são mensurados em minutos por semana e *educ* e *idade* são mensurados em anos. (Veja também o Problema 2.12.)

- (i) Se os adultos escolhem entre dormir e trabalhar, qual é sinal de β_1 ?
- (ii) Que sinais você espera que β_2 e β_3 terão?
- (iii) Usando os dados do arquivo SLEEP75.RAW, a equação estimada é

$$\begin{aligned} \hat{dormir} &= 3.638,25 - 0,148 trabtot - 11,13 educ + 2,20 idade \\ n &= 706, R^2 = 0,113. \end{aligned}$$

Se alguém trabalha 5 horas a mais por semana, qual é a queda, em minutos, no valor esperado de dormir? Esse valor representa uma escolha grande?

- (iv) Discuta o sinal e a magnitude do coeficiente de *educ*.
- (v) Você diria que *trabtot*, *educ* e *idade* explicam muito da variação de *dormir*? Quais outros fatores poderiam afetar o tempo gasto dormindo? É provável que eles sejam correlacionados com *trabtot*?

3.4 O salário inicial (mediano) para recém-formados em direito é determinado pela equação

$$\log(\hat{\text{salárioim}}) = \beta_0 + \beta_1 lsat + \beta_2 nmgrad + \beta_3 \log(\text{volbib}) + \beta_4 \log(\text{custo}) + \beta_5 rank + u,$$

em que *lsat* é a nota mediana do *lsat* (nota de ingresso no curso de direito) dos recém-formados, *nmgrad* é a nota mediana dos recém-formados nas disciplinas do curso de direito, *volbib* é o número de volumes da biblioteca da escola de direito, *custo* é o custo anual da escola de direito e *rank* é a classificação da escola de direito (com *rank* = 1 sendo o melhor posto de classificação).

- (i) Explique a razão de esperarmos $\beta_5 \leq 0$.
- (ii) Quais são os sinais que você espera para os outros parâmetros de inclinação? Justifique sua resposta.
- (iii) Utilizando os dados do arquivo LAWSCH85.RAW, a equação estimada é

$$\begin{aligned} \log(\hat{\text{salárioim}}) &= 8,34 + 0,0047 lsat + 0,248 nmgrad + 0,095 \log(\text{volbib}) \\ &\quad + 0,038 \log(\text{custo}) - 0,0033 rank \\ n &= 136, R^2 = 0,842. \end{aligned}$$

Qual é a diferença *ceteris paribus* prevista no salário para as escolas com um *nmgrad* mediano diferente em um ponto? (Descreva sua resposta em percentual.)

- (iv) Interprete o coeficiente da variável $\log(\text{volbib})$.
- (v) Você diria que é melhor freqüentar uma escola de direito que tem uma classificação melhor? Qual é a diferença no salário inicial esperado para uma escola que tem uma classificação igual a 20?

3.5 Em um estudo que relaciona a nota média em curso superior (*nmgrad*) ao tempo gasto em várias atividades, você distribuiu uma pesquisa para vários estudantes. Os estudantes devem responder quan-

tas horas eles despendem, em cada semana, em quatro atividades: estudo, sono, trabalho e lazer. Toda atividade é colocada em uma das quatro categorias, de modo que, para cada estudante, a soma das horas nas quatro atividades deve ser igual a 168.

(i) No modelo

$$nmgrad = \beta_0 + \beta_1 \text{estudar} + \beta_2 \text{dormir} + \beta_3 \text{trabalhar} + \beta_4 \text{lazer} + u,$$

faz sentido manter *dormir*, *trabalhar* e *lazer* fixos, enquanto *estudar* varia?

- (ii) Explique a razão de esse modelo violar a hipótese RLM.4.
 (iii) Como você poderia reformular o modelo, de modo que seus parâmetros tivessem uma interpretação útil e ele satisfizesse a hipótese RLM.4?

3.6 Considere o modelo de regressão múltipla contendo três variáveis independentes, sob as hipóteses RLM.1 a RLM.4:

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

Você está interessado em estimar a soma dos parâmetros de x_1 e x_2 ; chame-a de $\theta_1 = \beta_1 + \beta_2$. Mostre que $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$ é um estimador não-viesado de θ_1 .

3.7 Quais dos seguintes itens podem fazer com que os estimadores de MQO sejam viesados?

- (i) Heteroscedasticidade.
 (ii) Omitir uma variável importante.
 (iii) Um coeficiente de correlação amostral de 0,95 entre duas variáveis independentes incluídas no modelo.

3.8 Suponha que a produtividade média do trabalhador da indústria (*prodmed*) dependa de dois fatores — horas médias de treinamento do trabalhador (*treinmed*) e aptidão média do trabalhador (*aptidmed*):

$$prodmed = \beta_0 + \beta_1 \text{treinmed} + \beta_2 \text{aptidmed} + u.$$

Assuma que essa equação satisfaça as hipóteses de Gauss-Markov. Se um subsídio foi dado às empresas cujos trabalhadores têm uma aptidão menor do que a média, de modo que *treinmed* e *aptidmed* sejam negativamente correlacionados, qual é o provável viés em $\hat{\beta}_1$ obtido da regressão simples de *prodmed* sobre *treinmed*?

3.9 A equação seguinte descreve o preço mediano das residências de uma comunidade em termos da quantidade de poluição (*oxn*, de óxido nitroso) e do número médio de cômodos nas residências da comunidade (*comods*):

$$\log(\text{preço}) = \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \text{comods} + u.$$

- (i) Quais são os prováveis sinais de β_1 e β_2 ? Qual é a interpretação de β_1 ? Explique.
 (ii) Por que *oxn* [ou, mais precisamente, $\log(\text{oxn})$] e *comods* deveriam ser negativamente correlacionados? Se esse é o caso, a regressão simples de $\log(\text{preço})$ sobre $\log(\text{oxn})$ produz um estimador viesado para cima ou para baixo de β_1 ?

(iii) Utilizando os dados do arquivo HPRICE2.RAW foram estimadas as seguintes equações:

$$\log(\widehat{preço}) = 11,71 - 1,043 \log(oxn), n = 506, R^2 = 0,264.$$

$$\log(\widehat{preço}) = 9,23 - 0,718 \log(oxn) + 0,306 \text{ comods}, n = 506, R^2 = 0,514.$$

A relação entre as estimativas da elasticidade do *preço* das regressões simples e múltipla é a que você previu, tomando como base suas respostas na parte (ii)? Pode-se dizer que $-0,718$ está claramente mais próximo da elasticidade verdadeira que $-1,043$?

3.10 Suponha que o modelo populacional que determina y seja:

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

e esse modelo satisfaz as hipóteses de Gauss-Markov. Entretanto, estimamos o modelo que omite x_3 . Sejam $\tilde{\beta}_0$, $\tilde{\beta}_1$ e $\tilde{\beta}_2$ os estimadores de MQO da regressão de y sobre x_1 e x_2 . Mostre que o valor esperado de $\tilde{\beta}_1$ (dados os valores das variáveis independentes da amostra) é

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n \hat{r}_{i1} x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2},$$

em que os \hat{r}_{i1} são os resíduos de MQO da regressão de x_1 sobre x_2 . [Sugestão: a fórmula de $\tilde{\beta}_1$ é proveniente da equação (3.22). Coloque $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ nessa equação. Após alguma álgebra, aplique o operador esperança, tratando x_{i3} e \hat{r}_{i1} como não-aleatórios.]

3.11 A seguinte equação representa os efeitos das receitas totais de impostos sobre o crescimento subsequente do emprego para a população de municípios dos Estados Unidos:

$$\text{cresc} = \beta_0 + \beta_1 \text{parc}_p + \beta_2 \text{parc}_r + \beta_3 \text{parc}_v + \text{outros fatores},$$

em que *cresc* é a variação percentual do emprego de 1980 a 1990, enquanto o total das receitas de impostos tem a seguinte distribuição: parc_p é a parcela dos impostos sobre a propriedade, parc_r é a parcela das receitas de impostos sobre a renda e parc_v é a parcela das receitas de impostos sobre as vendas. Todas essas variáveis estão mensuradas em 1980. A parcela omitida, parc_t , inclui taxas e impostos variados. Por definição, as quatro parcelas somam um. Outros fatores incluiriam despesas com educação, infra-estrutura, e assim por diante (todos mensurados em 1980).

- (i) Por que devemos omitir uma das variáveis de parcela de impostos da equação?
- (ii) Dê uma interpretação cuidadosa de β_1 .

3.12 (i) Considere o modelo de regressão simples $y = \beta_0 + \beta_1 x + u$, sob as primeiras quatro hipóteses de Gauss-Markov. Para alguma função $g(x)$, por exemplo, $g(x) = x^2$ ou $g(x) = \log(1 + x^2)$, defina $z_i = g(x_i)$. Defina um estimador de inclinação como

$$\tilde{\beta}_1 = \left(\sum_{i=1}^n (z_i - \bar{z}) y_i \right) / \left(\sum_{i=1}^n (z_i - \bar{z}) x_i \right).$$

Mostre que $\tilde{\beta}_1$ é linear e não-viesado. Lembre-se: como $E(u|x) = 0$, você pode tratar tanto x_i como z_i como não-aleatórios em sua derivação.

(ii) Acrescente a hipótese de homoscedasticidade, RLM.5. Mostre que

$$\text{Var}(\tilde{\beta}_1) = \sigma^2 \left(\sum_{i=1}^n (z_i - \bar{z})^2 \right) / \left(\sum_{i=1}^n (z_i - \bar{z}) x_i \right)^2.$$

(iii) Mostre diretamente que, sob as hipóteses de Gauss-Markov, $\text{Var}(\hat{\beta}_1) \leq \text{Var}(\tilde{\beta}_1)$, em que $\hat{\beta}_1$ é o estimador de MQO. [Sugestão: a desigualdade de Cauchy-Schwartz do Apêndice B (disponível no site da Thomson) implica que

$$\left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x}) \right)^2 \leq \left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})^2 \right) \left(n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right);$$

Observe que podemos retirar \bar{x} da covariância amostral.]

3A.1 Derivação das Condições de Primeira Ordem da Equação (3.13)

A análise é muito similar à do caso da regressão simples. Devemos caracterizar as soluções para o problema

$$\min_{b_0, b_1, \dots, b_k} \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2.$$

Considerando as derivadas parciais em relação a cada um dos b_j (veja o Apêndice A, disponível no site da Thomson), avaliando-as nas soluções e igualando-as a zero resulta

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ -2 \sum_{i=1}^n x_{ij} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0, \text{ para todo } j = 1, \dots, k. \end{aligned}$$

Cancelando -2 obtemos as condições de primeira ordem em (3.13).

3A.2 Derivação da Equação (3.22)

Para derivar (3.22), escreva x_{i1} em termos de seus valores estimados e seus resíduos a partir da regressão de x_1 sobre x_2, \dots, x_k : $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$, para todo $i = 1, \dots, n$. Agora, insira essa expressão na segunda equação de (3.13):

$$\sum_{i=1}^n (\hat{x}_{i1} + \hat{r}_{i1})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0 \tag{3.60}$$

Pela definição do resíduo de MQO \hat{u}_i , como \hat{x}_{i1} é exatamente uma função linear das variáveis explicativas x_{i2}, \dots, x_{ik} , segue que $\sum_{i=1}^n \hat{x}_{i1} \hat{u}_i = 0$. Portanto, a equação (3.60) pode ser expressa como

$$\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0. \quad (3.61)$$

Como os \hat{r}_{i1} são os resíduos da regressão de x_1 sobre x_2, \dots, x_k , $\sum_{i=1}^n x_{ij} \hat{r}_{i1} = 0$, para todo $j = 2, \dots, k$. Portanto, (3.61) é equivalente a $\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_1 x_{i1}) = 0$. Finalmente, usamos o fato de que $\sum_{i=1}^n \hat{x}_{i1} \hat{r}_{i1} = 0$, o que significa que $\hat{\beta}_1$ soluciona

$$\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_1 \hat{r}_{i1}) = 0.$$

Agora, com um pouco de álgebra chegamos a (3.22), desde que, evidentemente, $\sum_{i=1}^n \hat{r}_{i1}^2 > 0$; isso é garantido pela hipótese RLM.4.

3A.3 Prova do Teorema 3.1

Vamos provar o Teorema 3.1 para $\hat{\beta}_1$; a prova para os outros parâmetros de inclinação é, virtualmente, idêntica. (Veja o Apêndice E, disponível no site da Thomson para uma prova mais sucinta utilizando matrizes.) Sob a hipótese RLM.4, os estimadores de MQO existem, e podemos escrever $\hat{\beta}_1$ como em (3.22). Sob a hipótese RLM.1, podemos escrever y_i como em (3.32); substitua-o pelo y_i de (3.22). Então, usando $\sum_{i=1}^n \hat{r}_{i1} = 0$, $\sum_{i=1}^n x_{j1} \hat{r}_{i1} = 0$, para todo $j = 2, \dots, k$ e $\sum_{i=1}^n x_{j1} \hat{r}_{i1} = \sum_{i=1}^n \hat{r}_{i1}^2$, temos

$$\hat{\beta}_1 = \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} u_i \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right). \quad (3.62)$$

Agora, sob as hipóteses RLM.2 e RLM.3, o valor esperado de cada u_i , dadas todas as variáveis independentes na amostra, é zero. Como os \hat{r}_{i1} são justamente funções das variáveis independentes da amostra, segue-se que

$$\begin{aligned} E(\hat{\beta}_1 | \mathbf{X}) &= \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} E(u_i | \mathbf{X}) \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) \\ &= \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} \cdot 0 \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) = \beta_1, \end{aligned}$$

em que \mathbf{X} representa os dados de todas as variáveis independentes, e $E(\hat{\beta}_1 | \mathbf{X})$ é o valor esperado de $\hat{\beta}_1$, dado x_{i1}, \dots, x_{ik} , para todo $i = 1, \dots, n$. Isso completa a prova.

3A.4 Viés de Variável Omitida no Modelo Geral

Podemos derivar o viés de variável omitida no modelo geral da equação (3.31) sob as quatro primeiras hipóteses de Gauss-Markov. Em particular, sejam $\hat{\beta}_j, j = 0, 1, \dots, k$ os estimadores de MQO da regressão ao se usar o conjunto completo de variáveis explicativas. Sejam $\tilde{\beta}_j, j = 0, 1, \dots, k - 1$ os estimadores de MQO da regressão que omite x_k . Sejam $\tilde{\delta}_j, j = 1, \dots, k - 1$ os coeficientes de inclinação de x_j da regressão auxiliar de x_{ik} sobre $x_{i1}, x_{i2}, \dots, x_{i,k-1}, i = 1, \dots, n$. Um fato útil é que

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j, \tag{3.63}$$

Isso mostra explicitamente que, quando não controlamos x_k na regressão, o efeito parcial estimado de x_j é igual ao efeito parcial quando incluímos x_k mais o efeito parcial de x_k sobre \hat{y} vezes a relação parcial entre a variável omitida, x_k , e $x_j, j < k$. Condicionado ao conjunto inteiro de variáveis explicativas, \mathbf{X} , sabemos que os $\hat{\beta}_j$ são todos não-viesados para os correspondentes $\beta_j, j = 1, \dots, k$. Além disso, como $\tilde{\delta}_j$ é exatamente uma função de \mathbf{X} , temos

$$\begin{aligned} E(\tilde{\beta}_j | \mathbf{X}) &= E(\hat{\beta}_j | \mathbf{X}) + E(\hat{\beta}_k | \mathbf{X}) \tilde{\delta}_j \\ &= \beta_j + \beta_k \tilde{\delta}_j. \end{aligned} \tag{3.64}$$

A equação (3.64) mostra que $\tilde{\beta}_j$ é viesado para β_j , a menos que $\beta_k = 0$ — caso em que x_k não tem efeito parcial na população —, ou $\tilde{\delta}_j$ é igual a zero, o que significa que x_{ik} e x_{ij} são parcialmente não-correlacionados na amostra. A chave para obter a equação (3.64) é a equação (3.63). Para mostrar a equação (3.63), podemos usar a equação (3.22) várias vezes. Por simplicidade, vamos olhar para $j = 1$. Agora, $\tilde{\beta}_1$ é o coeficiente de inclinação da regressão simples de y_i sobre $\tilde{r}_{i1}, i = 1, \dots, n$, em que os \tilde{r}_{i1} são os resíduos de MQO da regressão de x_{i1} sobre $x_{i2}, x_{i3}, \dots, x_{i,k-1}$. Considere o numerador da expressão de $\tilde{\beta}_1: \sum_{i=1}^n \tilde{r}_{i1} y_i$. Para cada i , podemos escrever $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik} + \hat{u}_i$ e colocar em y_i . Agora, pelas propriedades dos resíduos de MQO, os \tilde{r}_{i1} têm média amostral zero e são não-correlacionados com $x_{i2}, x_{i3}, \dots, x_{i,k-1}$ na amostra. Semelhantemente, os \hat{u}_i têm média amostral zero e correlação amostral zero com $x_{i1}, x_{i2}, \dots, x_{ik}$. Segue-se que os \tilde{r}_{i1} e \hat{u}_i são não-correlacionados na amostra (visto que os \tilde{r}_{i1} são exatamente uma combinação linear de $x_{i1}, x_{i2}, \dots, x_{i,k-1}$). Assim

$$\sum_{i=1}^n \tilde{r}_{i1} y_i = \hat{\beta}_1 \left(\sum_{i=1}^n \tilde{r}_{i1} x_{i1} \right) + \hat{\beta}_k \left(\sum_{i=1}^n \tilde{r}_{i1} x_{ik} \right). \tag{3.65}$$

Agora, $\sum_{i=1}^n \tilde{r}_{i1} x_{i1} = \sum_{i=1}^n \tilde{r}_{i1}^2$, que é também o denominador de $\tilde{\beta}_1$. Portanto, mostramos que

$$\begin{aligned} \tilde{\beta}_1 &= \hat{\beta}_1 + \hat{\beta}_k \left(\frac{\sum_{i=1}^n \tilde{r}_{i1} x_{ik}}{\sum_{i=1}^n \tilde{r}_{i1}^2} \right) \\ &= \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1. \end{aligned}$$

Essa é a relação que queríamos mostrar.

3A.5 Prova do Teorema 3.2

Novamente, vamos provar o teorema para $j = 1$. Escreva $\hat{\beta}_1$ como na equação (3.62). Agora, sob RLM.5, $\text{Var}(u_i | \mathbf{X}) = \sigma^2$, para todo $i = 1, \dots, n$. Sob amostragem aleatória, os u_i são independentes, mesmo condicionados a \mathbf{X} , e os \hat{r}_{i1} são não-aleatórios condicionados a \mathbf{X} . Portanto,

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | \mathbf{X}) &= \left(\sum_{i=1}^n \hat{r}_{i1}^2 \text{Var}(u_i | \mathbf{X}) \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right)^2 \\ &= \left(\sum_{i=1}^n \hat{r}_{i1}^2 \sigma^2 \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right)^2 = \sigma^2 / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right). \end{aligned}$$

Agora, visto que $\sum_{i=1}^n \hat{r}_{i1}^2$ é a soma dos quadrados dos resíduos da regressão de x_1 sobre x_2, \dots, x_k , $\sum_{i=1}^n \hat{r}_{i1}^2 = \text{SQT}_1(1 - R_1^2)$. Isso completa a prova.

3A.6 Prova do Teorema 3.4

Mostramos que, para qualquer outro estimador linear não-viesado $\tilde{\beta}_1$ de β_1 , $\text{Var}(\tilde{\beta}_1) \geq \text{Var}(\hat{\beta}_1)$, em que $\hat{\beta}_1$ é o estimador de MQO. Não se perde generalidade ao jogarmos o foco em $j = 1$.

Para $\tilde{\beta}_1$ como na equação (3.59), podemos inserir em y_i para obter

$$\tilde{\beta}_1 = \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik} + \sum_{i=1}^n w_{i1} u_i.$$

Agora, visto que os w_{i1} são funções de x_{ij} ,

$$\begin{aligned} E(\tilde{\beta}_1 | \mathbf{X}) &= \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik} + \sum_{i=1}^n w_{i1} E(u_i | \mathbf{X}) \\ &= \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik} \end{aligned}$$

porque $E(u_i | \mathbf{X}) = 0$, para todo $i = 1, \dots, n$, sob RLM.2 e RLM.3. Portanto, para $E(\tilde{\beta}_1 | \mathbf{X})$ igualar-se a β_1 para quaisquer valores dos parâmetros, devemos ter

$$\sum_{i=1}^n w_{i1} = 0, \quad \sum_{i=1}^n w_{i1} x_{i1} = 1, \quad \sum_{i=1}^n w_{i1} x_{ij} = 0, \quad j = 2, \dots, k. \quad (3.66)$$

Agora, sejam \hat{r}_{i1} os resíduos da regressão de x_{i1} sobre x_{i2}, \dots, x_{ik} . Então, de (3.66), segue-se que

$$\sum_{i=1}^n w_{i1} \hat{r}_{i1} = 1 \quad (3.67)$$

visto que $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$ e $\sum_{i=1}^n w_{i1} \hat{x}_{i1} = 0$. Agora, considere a diferença entre $\text{Var}(\tilde{\beta}_1 | \mathbf{X})$ e $\text{Var}(\hat{\beta}_1 | \mathbf{X})$ sob RLM.1 a RLM.5:

$$\sigma^2 \sum_{i=1}^n w_{i1}^2 - \sigma^2 \left/ \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) \right. \quad (3.68)$$

Por causa de (3.67), podemos escrever a diferença em (3.68), sem σ^2 , como

$$\sum_{i=1}^n w_{i1}^2 - \left(\sum_{i=1}^n w_{i1} \hat{r}_{i1} \right)^2 \left/ \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) \right. \quad (3.69)$$

No entanto, (3.69) é simplesmente

$$\sum_{i=1}^n (w_{i1} - \hat{\gamma}_1 \hat{r}_{i1})^2, \quad (3.70)$$

em que $\hat{\gamma}_1 = \left(\sum_{i=1}^n w_{i1} \hat{r}_{i1} \right) \left/ \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) \right.$, como pode ser visto ao se elevar cada termo em (3.70) ao quadrado, somando e cancelando os termos. Como (3.70) é exatamente a soma dos resíduos quadrados da regressão simples de w_{i1} sobre \hat{r}_{i1} — lembre-se de que a média amostral de \hat{r}_{i1} é zero —, (3.70) deve ser não-negativo. Isso completa a prova.

Análise de Regressão Múltipla: Problemas Adicionais

Este capítulo articula vários problemas da análise de regressão múltipla que não foram tratados convenientemente nos capítulos anteriores. Estes tópicos não são tão fundamentais quanto os discutidos nos capítulos 3 e 4, mas são importantes para a aplicação da regressão múltipla em uma ampla gama de problemas empíricos.

6.1 EFEITOS DA DIMENSÃO DOS DADOS NAS ESTATÍSTICAS MQO

No Capítulo 2, sobre regressão bivariada, discutimos de forma sucinta os efeitos da mudança nas unidades de medida sobre os interceptos e as estimativas de inclinação do MQO. Também mostramos que a mudança nas unidades de medida não afeta o R -quadrado. Agora retornaremos ao problema da dimensão dos dados e examinaremos o efeito do redimensionamento das variáveis dependente ou independente sobre os erros-padrão, estatísticas t , estatísticas F e intervalos de confiança.

Veremos que tudo o que esperamos acontecer, acontecerá. Quando as variáveis são redimensionadas, os coeficientes, erros-padrão, intervalos de confiança, estatísticas t e F mudam de tal maneira que preservam todos os efeitos mensurados e os resultados dos testes. Embora isso não seja uma grande surpresa – aliás, ficaríamos muito preocupados se não fosse assim – é útil ver o que ocorre explicitamente. Muitas vezes, o redimensionamento de dados é usado com finalidade cosmética, tal como reduzir o número de zeros depois da vírgula, em um coeficiente estimado. Escolhendo-se criteriosamente as unidades de medida, podemos melhorar a aparência de uma equação estimada sem alterar nada que seja essencial.

Poderíamos tratar deste problema de maneira generalizada, mas ele é mais bem ilustrado com exemplos. Da mesma forma, será de pouca valia neste ponto introduzirmos uma notação abstrata.

Começamos com uma equação relacionando o peso dos recém-nascidos com o hábito de fumar e a renda familiar:

$$pesônas = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 rendfam \quad (6.1)$$

onde $pesônas$ é o peso dos recém-nascidos, em onças, $cigs$ é o número médio de cigarros que a mãe fumou por dia durante a gravidez, e $rendfam$ é a renda anual familiar, em milhares de dólares. As estimativas desta equação, obtidas utilizando dados contidos no arquivo BWGHT.RAW, são dadas na primeira coluna da Tabela 6.1. Os erros-padrão estão relacionados entre parênteses. A estimativa de $cigs$ mostra que se uma mulher fumar cinco ou mais cigarros por dia, o peso previsto dos recém-

nascidos deve estar em torno de $0,4634(5) = 2,317$ onças a menos. A estatística t de *cigs* é $-5,06$, de modo que a variável é estatisticamente bastante significativa.

Tabela 6.1

Efeitos da Dimensão dos Dados

Variável Dependente	(1) <i>pesonas</i>	(2) <i>pesonaslb</i>	(3) <i>pesonas</i>
Variáveis Independentes			
<i>cigs</i>	-0,4634 (0,0916)	-0,0289 (0,0057)	_____
<i>maços</i>	_____	_____	-9,268 (1,832)
<i>rendfam</i>	0,0927 (0,0292)	0,0058 (0,0018)	0,0927 (0,0292)
<i>intercepto</i>	116,974 (1,049)	7,3109 (0,0656)	116,974 (1,049)
Observações	1.388	1.388	1.388
R-quadrado	0,0298	0,0298	0,0298
SQR	557.485,51	2.177,6778	557.485,51
EPR	20,063	1,2539	20,063

Agora, suponha que decidimos medir o peso dos recém-nascidos em libras, em vez de onças. Façamos $pesonaslb = pesonas/16$ ser o peso dos recém-nascidos em libras. O que acontece com nossas estatísticas MQO se usarmos essa variável dependente em nossa equação? É fácil verificar o efeito no coeficiente da estimativa pela simples manipulação da equação (6.1). Divida a equação inteira por 16:

$$pesonas/16 = \hat{\beta}_0/16 + (\hat{\beta}_1/16)cigs + (\hat{\beta}_2/16)rendfam.$$

Como o termo da esquerda é o peso dos recém-nascidos em libras, segue-se que cada novo coeficiente corresponderá ao coeficiente antigo dividido por 16. Para verificar isso, a regressão de *pesonaslb* sobre *cigs* e *rendfam* está registrada na coluna (2) da Tabela 6.1. Até quatro dígitos, o intercepto e as inclinações da coluna (2) são exatamente os da coluna (1) divididos por 16. Por exemplo, o coeficiente de *cigs* é agora $-0,0289$; isso significa que, se *cigs* fosse cinco vezes mais alto, o peso de nascimento seria $0,0289(5) = 0,1445$ libras mais baixo. Em termos de onças, temos $0,1445(16) = 2,312$, que é um pouco diferente dos 2,317 que obtivemos anteriormente devido ao erro de arredondamento. A questão importante é que, uma vez que os efeitos tenham sido transformados nas mesmas unidades, obtemos exatamente a mesma resposta, independentemente de como a variável dependente seja medida.

E quanto à significância estatística? Como esperado, a alteração da variável dependente de onças para libras não tem efeito sobre o quanto são estatisticamente importantes as variáveis independentes. Os erros-padrão na coluna (2) são 16 vezes menores que os da coluna (1). Alguns cálculos rápidos mostram que as estatísticas t na coluna (2) são, realmente, idênticas às da coluna (1). Os pontos extremos dos intervalos de confiança na coluna (2) são exatamente os pontos extremos na coluna (1) divididos por 16. Isso ocorre porque os ICs mudam pelos mesmos fatores dos erros-padrão. (Lembre-se de que o IC de 95% neste caso é $\hat{\beta}_j \pm 1,96 \text{ ep}(\hat{\beta}_j)$).

Em termos de grau de ajuste, os R -quadrados das duas regressões são idênticos, como esperado. Observe que a soma dos resíduos quadrados, SQR , e o erro-padrão da regressão, EPR , diferem nas equações. Essas diferenças são facilmente explicadas. Seja \hat{u}_i o resíduo da observação i na equação original (6.1). Então, quando *pesonaslbs* é a variável dependente, o resíduo é simplesmente $\hat{u}_i/16$. Assim, o resíduo *quadrado* na segunda equação é $(\hat{u}_i/16)^2 = \hat{u}_i^2/256$. Essa é a razão pela qual a soma dos resíduos quadrados na coluna (2) é igual à SQR na coluna (1) dividida por 256.

Como $EPR = \hat{\sigma} = \sqrt{SQR/(n-k-1)} = \sqrt{SQR/1.385}$, SQR na coluna (2) é 16 vezes menor do que na coluna (1). Outra maneira de ver isso é que o erro na equação com *pesonaslb* como a variável dependente tem um desvio-padrão 16 vezes menor do que o desvio-padrão do erro original. Isso não significa que tenhamos reduzido o erro ao alterarmos a maneira pela qual o peso dos recém-nascidos é medido: o EPR menor simplesmente reflete uma diferença nas unidades de medida.

Continuando, retornemos à unidade de medida original da variável dependente: *pesonas* é medido em onças. Vamos alterar a unidade de medida de uma das variáveis independentes, *cigs*. Defina *maços* como sendo a quantidade de maços de cigarros fumados por dia. Assim, $\text{maços} = \text{cigs}/20$. Agora, o que acontece com os coeficientes e outras estatísticas MQO? Dessa forma, podemos escrever

$$\text{pesonas} = \hat{\beta}_0 + (20\hat{\beta}_1)(\text{cigs}/20) + \hat{\beta}_2 \text{rendfam} = \hat{\beta}_0 + (20\hat{\beta}_1)\text{maços} + \hat{\beta}_2 \text{rendfam}.$$

Portanto, o intercepto e o coeficiente de inclinação de *rendfam* não se alteraram, mas o coeficiente de *maços* é 20 vezes o de *cigs*. Isso é intuitivamente atraente. Os resultados da regressão de *pesonas* sobre *maços* e *rendfam* estão na coluna (3) da Tabela 6.1. A propósito, lembre-se de que não teria sentido incluir tanto *cigs* como *maços* na mesma equação; isso induziria à multicolinearidade perfeita e não teria nenhum significado interessante.

Na equação original sobre o peso dos recém-nascidos (6.1), suponha que *rendfam* seja medida em dólares em lugar de milhares de dólares. Desse modo, defina a variável $\text{rendfamdol} = 1.000 \cdot \text{rendfam}$. Como mudam as estatísticas MQO quando *rendfamdol* substitui *rendfam*? Para o propósito de apresentar os resultados da regressão, você acha melhor medir a renda em dólares ou em milhares de dólares?

Além do coeficiente de *maços*, existe outra estatística na coluna (3) que difere da mostrada na coluna (1): o erro-padrão de *maços* é 20 vezes maior que o de *cigs* na coluna (1). Isso significa que a estatística t para verificar a significância do hábito de fumar é a mesma, quer ele seja medido em cigarros ou em maços. Isso é natural.

O exemplo anterior explica claramente a maioria das possibilidades que surgem quando a variável dependente e as variáveis independentes são redimensionadas. O redimensionamento muitas

vezes é feito com os valores monetários em economia, especialmente quando os montantes são muito grandes.

No Capítulo 2, argumentamos que, se a variável dependente aparecer na forma logarítmica, a alteração na unidade de medida não afetará o coeficiente de inclinação. Isso também acontece aqui: a alteração na unidade de medida da variável dependente, quando aparece na forma logarítmica, não afeta qualquer das estimativas de inclinação. Isso resulta do simples fato de que $\log(c_1 y_i) = \log(c_1) + \log(y_i)$ para qualquer constante $c_1 > 0$. O novo intercepto será $\log(c_1) + \hat{\beta}_0$. De forma semelhante, a alteração da unidade de medida de qualquer x_j , onde $\log(x_j)$ aparece na regressão, afeta somente o intercepto. Isso corresponde ao que conhecemos sobre alterações em porcentagens e, em particular, em elasticidades: elas não sofrem alterações quando mudam as unidades de medida de y ou de x_j . Por exemplo, se tivéssemos especificado a variável dependente em (6.1) como $\log(\text{pesonas})$, estimássemos a equação, e depois a tivéssemos reestimado com $\log(\text{pesonas}lb)$ como a variável dependente, os coeficientes de $cigs$ e $rendfam$ seriam os mesmos em ambas as regressões; somente o intercepto seria diferente.

Os Coeficientes Beta

Algumas vezes, em aplicações econométricas, uma variável-chave é medida em uma dimensão de difícil interpretação. Economistas especializados na área de trabalho frequentemente incluem a pontuação de testes de conhecimentos em equações salariais, e a dimensão em que tais testes são registrados muitas vezes é arbitrária e de difícil interpretação (pelo menos para os economistas!). Em quase todos os casos estamos interessados em saber como a pontuação de um indivíduo em particular se compara com a população. Assim, em lugar de perguntarmos a respeito do efeito sobre o salário por hora se, digamos, a pontuação do teste for dez pontos mais alta, faz mais sentido perguntar o que acontece quando a pontuação do teste for um desvio-padrão mais alto.

Nada impede que vejamos o que acontece com a variável dependente quando uma variável independente em um modelo estimado aumenta certo número de desvios-padrão, supondo que tenhamos obtido o desvio-padrão da amostra (o que é fácil na maioria dos programas de regressão). Geralmente, essa é uma boa idéia. Assim, por exemplo, quando observamos o efeito de uma pontuação de teste padronizada, como o SAT (nota de ingresso em curso superior nos Estados Unidos), sobre a nota média em curso superior, podemos encontrar o desvio-padrão de SAT e verificar o que acontece quando essa pontuação aumenta em um ou dois desvios-padrão.

Algumas vezes é útil obter resultados de regressão quando todas as variáveis envolvidas, a dependente e todas as independentes, tenham sido padronizadas. Uma variável é padronizada em uma amostra pela subtração de sua média e dividindo o resultado por seu desvio-padrão (veja Apêndice C disponível no site do livro, no site www.thomsonlearning.com.br). Isso significa que computamos a transformação z de cada variável na amostra. Depois, fazemos a regressão usando os valores de z .

Por que a padronização é útil? É mais fácil começarmos com a equação MQO original, com as variáveis em suas formas originais:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik} + \hat{u}_i \quad (6.2)$$

Incluimos o subscrito de observação i para enfatizar que nossa padronização é aplicada a todos os valores da amostra. Agora, se ao calcularmos a média de (6.2), usarmos o fato de que \hat{u}_i tem uma média de amostra zero, e subtrairmos o resultado de (6.2), temos

$$y_i - \bar{y} = \hat{\beta}_1(x_{i1} - \bar{x}_1) + \hat{\beta}_2(x_{i2} - \bar{x}_2) + \dots + \hat{\beta}_k(x_{ik} - \bar{x}_k) + \hat{u}_i$$

Em seguida, definamos $\hat{\sigma}_y$ como o desvio-padrão da amostra da variável dependente, $\hat{\sigma}_1$ como o dp da amostra da x_1 , $\hat{\sigma}_2$ como o dp da amostra de x_2 , e assim sucessivamente. Agora, um pouco de álgebra produz a equação

$$\begin{aligned} (y_i - \bar{y})/\hat{\sigma}_y &= (\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1[(x_{i1} - \bar{x}_1)/\hat{\sigma}_1] + \dots \\ &+ (\hat{\sigma}_k/\hat{\sigma}_y)\hat{\beta}_k[(x_{ik} - \bar{x}_k)/\hat{\sigma}_k] + (\hat{u}_i/\hat{\sigma}_y). \end{aligned} \quad (6.3)$$

Cada variável em (6.3) foi padronizada pela substituição de suas médias por seus valores de z , e isso resultou em novos coeficientes de inclinação. Por exemplo, o coeficiente de inclinação de $(x_{i1} - \bar{x}_1)/\hat{\sigma}_1$ é $(\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1$. Isso é simplesmente o coeficiente original, $\hat{\beta}_1$, multiplicado pela razão do desvio-padrão de x_1 sobre o desvio-padrão de y . O intercepto simplesmente desapareceu.

É útil reescrever (6.3), eliminando o subscrito i , como

$$z_y = \hat{b}_1 z_{x_1} + \hat{b}_2 z_{x_2} + \dots + \hat{b}_k z_{x_k} + \text{erro}, \quad (6.4)$$

onde z_y é o valor de z de y , z_{x_j} é o valor de z de x_j , e assim por diante. Os novos coeficientes são

$$\hat{b}_j = (\hat{\sigma}_j/\hat{\sigma}_y)\hat{\beta}_j \text{ para } j = 1, \dots, k. \quad (6.5)$$

Esses \hat{b}_j são tradicionalmente chamados de **coeficientes padronizados** ou **coeficientes beta**. (Esta última denominação é mais comum, mas um pouco inadequada, já que temos usado o beta chapéu para representar as estimativas MQO usuais.)

Os coeficientes beta recebem seus interessantes significados a partir da equação (6.4): Se x_1 aumentar em um desvio-padrão, \hat{y} , então, será alterado em \hat{b}_1 desvios-padrão. Assim, estamos medindo os efeitos não em termos das unidades originais de y ou de x_j , mas em unidades de desvios-padrão. Como isso torna a dimensão dos regressores irrelevante, essa equação coloca as variáveis explicativas em pé de igualdade. Em uma equação MQO padrão, não é possível simplesmente verificar o tamanho dos diferentes coeficientes e concluir que a variável explicativa com o maior coeficiente é “a mais importante”. Acabamos de ver que a magnitude dos coeficientes pode ser mudada à vontade pela alteração das unidades de medida das variáveis x_j . Mas, quando cada x_j é padronizado, a comparação das magnitudes dos coeficientes beta resultantes é mais convincente.

Para obter os coeficientes beta, podemos sempre padronizar y, x_1, \dots, x_k e em seguida computar a regressão MQO do valor de z de y sobre os valores de z de x_1, \dots, x_k – no qual não é necessário incluir um intercepto, já que ele será zero. Isso pode ser tedioso com muitas variáveis independentes. Alguns programas econométricos produzem coeficientes beta com um simples comando. O exemplo seguinte ilustra o uso de coeficientes beta.

EXEMPLO 6.1**(Efeitos da Poluição sobre os Preços de Imóveis)**

Utilizamos os dados do Exemplo 4.5 (do arquivo HPRICE2.RAW) para ilustrar o uso de coeficientes beta. Lembre-se de que a principal variável independente é *oxn*, uma medida do óxido nitroso no ar em cada comunidade. Uma maneira de entender o tamanho do efeito da poluição – sem entrar na questão científica do efeito do óxido de nitrogênio sobre a qualidade do ar é computar os coeficientes beta. (O Exemplo 4.5 contém um método alternativo: obtivemos uma elasticidade-preço em relação a *oxn* usando *preço* e *oxn* em forma logarítmica.)

A equação populacional é o modelo nível-nível

$$preço = \beta_0 + \beta_1 oxn + \beta_2 crime + \beta_3 comods + \beta_4 dist + \beta_5 razestud + u,$$

onde todas as variáveis exceto *crime* foram definidas no Exemplo 4.5; *crime* é o número de crimes registrados *per capita*. Os coeficientes beta aparecem na seguinte equação (portanto cada variável foi convertida ao seu valor de *z*):

$$zpreço = -0,340 zoxn - 0,143 zcrime + 0,514 zcomods - 0,235 zdist - 0,270 zrazestud.$$

Esta equação mostra que o aumento de um desvio-padrão em *oxn* reduz o preço em 0,34 desvio-padrão; o aumento de um desvio-padrão em *crime* reduz o preço em 0,14 desvio-padrão. Assim, o mesmo movimento relativo da poluição na população tem um efeito maior sobre os preços dos imóveis do que o da criminalidade. O tamanho do imóvel, medido pelo número de cômodos (*comods*), tem o maior efeito padronizado. Se quisermos saber os efeitos de cada variável independente sobre o valor da mediana dos preços dos imóveis, teremos que usar as variáveis não padronizadas.

O uso de variáveis padronizadas ou não padronizadas não afetará a significância estatística: as estatísticas *t* serão as mesmas, em ambos os casos.

6.2 UM POUCO MAIS SOBRE A FORMA FUNCIONAL

Em vários dos exemplos anteriores, encontramos o artifício mais comum em econometria para permitir relações não lineares entre a variável explicada e as variáveis explicativas: o uso de logaritmos das variáveis dependentes ou independentes. Também vimos modelos contendo os quadrados de algumas variáveis explicativas, mas ainda precisamos discutir sobre um tratamento sistemático desses tópicos. Nesta seção, trataremos de algumas variações e extensões sobre formas funcionais que surgem freqüentemente em trabalhos aplicados.

Um pouco mais sobre o Uso de Formas Funcionais Logarítmicas

Começamos revendo como interpretar os parâmetros no modelo

$$\log(preço) = \beta_0 + \beta_1 \log(oxn) + \beta_2 comods + u, \quad (6.6)$$

onde essas variáveis são as mesmas do Exemplo 4.5. Lembre-se de que em todo o texto $\log(x)$ é o log natural de x . O coeficiente β_1 é a elasticidade do *preço* em relação a *oxn* (poluição). O coeficiente β_2 é a mudança em $\log(\textit{preço})$, quando $\Delta\textit{comods} = 1$; como vimos muitas vezes, quando multiplicada por 100, essa é a percentagem aproximada de mudança em *preço*. Lembre-se de que $100 \cdot \beta_2$ é algumas vezes chamado de semi-elasticidade do *preço* em relação a *comods*.

Quando estimamos utilizando os dados do arquivo HPRICE2.RAW, obtemos

$$\begin{aligned} \log(\hat{\textit{preço}}) &= 9,23 - 0,718 \log(\textit{oxn}) + 0,306 \textit{comods} \\ &\quad (0,19) \quad (0,66) \quad (0,019) \qquad \qquad \qquad \mathbf{(6.7)} \\ n &= 506, R_2 = 0,514. \end{aligned}$$

Assim, quando *oxn* aumenta em 1%, *preço* cai em 0,718%, mantendo-se apenas *comods* fixo. Quando *comods* aumenta em um, *preço* aumenta em aproximadamente $100(0,306) = 30,6\%$.

A estimativa de que um cômodo a mais aumenta o preço em cerca de 30,6% acaba por ser de certa forma impreciso para esta aplicação. O erro de aproximação ocorre porque, como a mudança em $\log(y)$ se torna cada vez maior, a aproximação $\% \Delta y \approx 100 \cdot \Delta \log(y)$ se mostra cada vez mais imprecisa. Felizmente, existe um cálculo simples para computar a percentagem exata de mudança.

Para descrever o procedimento, consideremos o modelo estimado de forma geral

$$\hat{\log}(y) = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2.$$

(A inclusão de variáveis independentes adicionais não altera o procedimento.) Agora, fixando x_1 , temos $\Delta \hat{\log}(y) = \hat{\beta}_2 \Delta x_2$. O uso de simples propriedades algébricas das funções exponenciais e logarítmicas produz a percentagem exata de mudança no y estimado como

$$\% \hat{\Delta} y = 100 \cdot [\exp(\hat{\beta}_2 \Delta x_2) - 1], \qquad \qquad \qquad \mathbf{(6.8)}$$

onde a multiplicação por 100 transforma a mudança proporcional em uma mudança percentual. Quando $\Delta x_2 = 1$,

$$\% \hat{\Delta} y = 100 \cdot [\exp(\hat{\beta}_2) - 1]. \qquad \qquad \qquad \mathbf{(6.9)}$$

Aplicada ao exemplo dos preços dos imóveis com $x_2 = \textit{comods}$ e $\hat{\beta}_2 = 0,306$, $\% \Delta \textit{preço} = 100[\exp(0,306) - 1] = 35,8\%$, que é visivelmente maior do que a percentagem aproximada de mudança, 30,6%, obtida diretamente de (6.7). (A propósito, esse não é um estimador não-viesado, pois $\exp(\cdot)$ é uma função não-linear; ele é, porém, um estimador consistente de $100[\exp(\beta_2) - 1]$. Isso é assim porque o limite de probabilidade é calculado por meio de funções contínuas, enquanto o operador valor esperado não é calculado dessa forma. Veja Apêndice C, no site da Thomson.)

O ajuste na equação (6.8) não é tão crucial para pequenas mudanças percentuais. Por exemplo, quando incluímos na equação (6.7) a relação aluno-professor, seu coeficiente estimado é $-0,052$, o que significa que, se *razestud* aumentar em um, *preço* diminui em aproximadamente 5,2%. A mudança proporcional exata é $\exp(-0,052) - 1 \approx -0,051$, ou $-5,1\%$. De outro lado, se aumentarmos *razestud* em

cinco, então a mudança percentual aproximada em *preço* será -26% , enquanto a mudança exata obtida da equação (6.8) é $100[\exp(-0,26) - 1] \approx -22,9\%$.

Vimos que o uso de logs naturais leva a coeficientes com interpretações interessantes e podemos ignorar o fato de as unidades de medida das variáveis aparecerem em forma logarítmica, pois os coeficientes de inclinação são invariantes em relação a redimensionamentos. Existem várias outras razões pelas quais os logs são tão usados em trabalhos aplicados. Em primeiro lugar, quando $y > 0$, os modelos que usam $\log(y)$ como a variável dependente geralmente satisfazem as hipóteses do MLC mais apropriadamente do que os modelos que usam o nível de y . Variáveis estritamente positivas frequentemente possuem distribuições condicionais que são heteroscedásticas ou concentradas; o uso do log pode aliviar, se não eliminar, ambos os problemas.

Além disso, o uso de logs normalmente estreita a amplitude dos valores das variáveis, em alguns casos em quantidade considerável. Isso torna as estimativas menos sensíveis a observações díspares (ou extremas) na variável dependente ou nas variáveis independentes. Abordaremos a questão das observações extremas no Capítulo 9.

Existem algumas regras práticas padronizadas para o uso de logs, embora nenhuma definitiva. Quando a variável é um valor monetário positivo, ele frequentemente é transformado em log. Temos visto isso para variáveis como salários, vendas de empresas e valores de mercado das empresas. Variáveis como população, número total de empregados e matrículas escolares frequentemente aparecem em forma logarítmica; elas têm a característica comum de serem grandes valores inteiros.

Variáveis que são medidas em anos — como educação, experiência, tempo de permanência, idade etc. — normalmente aparecem em sua forma original. Uma variável que seja uma proporção ou uma porcentagem — como a taxa de desemprego, a taxa de participação em planos de aposentadoria, a taxa de estudantes aprovados em um exame padronizado e a taxa de detenção sobre crimes registrados — pode aparecer tanto em sua forma original como logarítmica, embora haja uma tendência em usá-la em forma de nível. Isso se deve ao fato de que quaisquer coeficientes de regressão envolvendo a variável *original* — seja ela a variável dependente ou independente — terão uma interpretação de mudança de *pontos percentuais*. (Veja Apêndice A, no site da Thomson, para uma revisão sobre a distinção entre mudança percentual e mudança de pontos percentuais.) Se usarmos, digamos, $\log(\textit{desemp})$ em uma regressão, onde *desemp* é a porcentagem de indivíduos desempregados, precisamos ter muito cuidado para distinguir entre uma mudança de pontos percentuais e uma mudança percentual. Lembre-se, quando *desemp* aumenta de oito para nove, isso é um acréscimo de um ponto percentual, equivalente a um incremento de $12,5\%$ sobre o nível de desemprego inicial. O uso do log significa que queremos saber a mudança percentual da taxa de desemprego: $\log(9) - \log(8) \approx 0,118$ ou $11,8\%$, que é a aproximação logarítmica do aumento efetivo de $12,5\%$.

Suponha que o número anual de prisões por direção de veículo sob embriaguez¹ seja determinado por

$$\log(\textit{prisões}) = \beta_0 + \beta_1 \log(\textit{pop}) + \beta_2 \textit{idade16_25} + \textit{outros fatores},$$

onde *idade16_25* é a proporção da população entre 16 e 25 anos de idade. Mostre que β_2 tem a seguinte interpretação (*ceteris paribus*): ela é a mudança percentual em *prisões* quando a porcentagem da população com idade entre 16 e 25 anos aumenta em um *ponto percentual*.

¹ NE: Nos Estados Unidos.

Uma limitação do log é que ele não pode ser usado, caso uma variável assumia valor zero ou negativo. Em casos nos quais a variável y não seja negativa, mas pode assumir o valor 0, $\log(1 + y)$ é algumas vezes usado. As interpretações de mudança percentual são, em geral, estritamente preservadas, exceto para mudanças começando em $y = 0$ (em que a percentagem de mudança não é sequer definida). Geralmente, usar $\log(1 + y)$ e depois interpretar as estimativas como se a variável fosse $\log(y)$ é aceitável quando os dados em y não são dominados por zeros. Um exemplo pode ser o de y representar horas de treinamento por funcionário da população industrial, quando uma grande fração das empresas oferece treinamento a, pelo menos, um de seus funcionários.

Uma desvantagem de usar uma variável dependente na forma logarítmica está na dificuldade de se prever a variável original. O modelo original nos permite prever $\log(y)$, e não y . No entanto, é razoavelmente fácil transformar uma previsão de $\log(y)$ em uma previsão de y (veja Seção 6.4). Uma questão relacionada é que *não* é válido comparar R -quadrados de modelos nos quais y é a variável dependente em um caso e $\log(y)$ é a variável dependente no outro. Essas medidas explicam variações em diferentes variáveis. Discutimos como computar medidas comparáveis de graus de ajuste na Seção 6.4.

Modelos com Funções Quadráticas

As **funções quadráticas** também são usadas com bastante frequência em economia aplicada para capturar efeitos marginais crescentes ou decrescentes. Seria interessante rever as propriedades das funções quadráticas no Apêndice A, no site da Thomson.

No caso mais simples, y depende de um único fator observado x , mas de uma forma quadrática:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u.$$

Por exemplo, considere $y = \text{salário}$ e $x = \text{exper.}$ Como discutimos no Capítulo 3, esse modelo não se enquadra na análise de regressão simples, mas é facilmente trabalhado em regressão múltipla.

É importante lembrar que β_1 não mede a mudança em y em relação a x ; não faz sentido manter x^2 fixo quando se altera x . Se escrevermos a equação estimada como

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2, \quad (6.10)$$

teremos a aproximação

$$\Delta \hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x) \Delta x, \text{ e assim } \Delta \hat{y} / \Delta x \approx \hat{\beta}_1 + 2\hat{\beta}_2 x. \quad (6.11)$$

Isso nos mostra que a inclinação da relação entre x e y depende do valor de x ; a inclinação estimada é $\hat{\beta}_1 + 2\hat{\beta}_2 x$. Se inserirmos $x = 0$, veremos que $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$. Para outras mudanças no valor de x , o segundo termo, $2\hat{\beta}_2 x$, deve ser levado em conta.

Se estivermos interessados em somente computar a mudança prevista em y dado um valor inicial de x e uma mudança de x , poderíamos usar (6.10) diretamente: não há nenhuma razão para usar cálculos de aproximação. Porém, normalmente estamos mais interessados em resumir rapidamente o efeito de x em y , e a interpretação de $\hat{\beta}_1$ e de $\hat{\beta}_2$ na equação (6.11) fornece esse resumo. Em geral, podemos

inserir o valor médio de x na amostra, ou outros valores de interesse, como a mediana ou os valores dos quartis inferior ou superior de x .

Em muitas aplicações, $\hat{\beta}_1$ é positivo, e $\hat{\beta}_2$ é negativo. Por exemplo, utilizando os dados de salários contidos no arquivo WAGE1.RAW, obtemos

$$\begin{aligned} \text{saláριο} &= 3,73 + 0,298 \text{ exper} - 0,0061 \text{ exper}^2 \\ &\quad (0,35) \quad (0,041) \quad (0,0009) \qquad \qquad \qquad \mathbf{(6.12)} \\ n &= 526, R^2 = 0,093. \end{aligned}$$

A equação estimada sugere que exper tem um efeito de redução sobre saláριο . O primeiro ano de experiência vale aproximadamente 30 centavos de dólar por hora (0,298 dólares). O segundo ano de experiência vale menos [cerca de $0,298 - 2(0,0061)(1) \approx 0,286$, ou 28,6 centavos de dólar, de acordo com a aproximação em (6.11) com $x = 1$]. Aumentando de 10 para 11 os anos de experiência, a previsão de aumento do salário-hora é de cerca de $0,298 - 2(0,0061)(10) \approx 0,176$ ou 17,6 centavos de dólar. E assim por diante.

Quando o coeficiente de x é positivo e o coeficiente de x^2 é negativo, a função quadrática tem um formato parabólico. Sempre existe um valor positivo de x , no qual o efeito de x sobre y é zero; antes desse ponto, x tem um efeito positivo sobre y ; após esse ponto, x tem um efeito negativo sobre y . Na prática, pode ser importante saber onde fica esse ponto crítico.

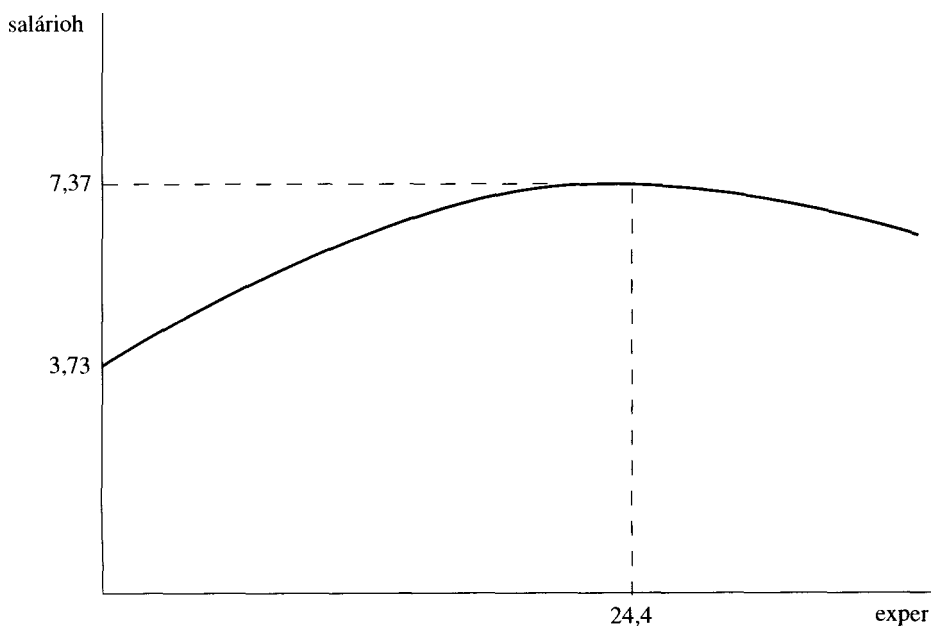
Na equação estimada (6.10) com $\hat{\beta}_1 > 0$ e $\hat{\beta}_2 < 0$, esse ponto crítico (ou o máximo da função) é sempre alcançado na relação entre o coeficiente de x e duas vezes o valor absoluto do coeficiente de x^2 :

$$x^* = |\hat{\beta}_1 / (2\hat{\beta}_2)|. \qquad \qquad \qquad \mathbf{(6.13)}$$

No exemplo dos salários, $x^* = \text{exper}^* \approx 0,298 / [2(0,0061)] \approx 24,4$. (Observe como simplesmente eliminamos o sinal de menos em $-0,0061$ ao fazermos esse cálculo.) Esta relação quadrática está ilustrada na Figura 6.1.

Na equação dos salários (6.12), o retorno da experiência passa a ser zero por volta dos 24,4 anos. O que devemos concluir disso? Existem, pelo menos, três explicações possíveis. Primeiro, pode ser que poucas pessoas na amostra tenham mais de 24 anos de experiência, e assim a parte da curva à direita de 24 pode ser ignorada. A consequência de se usar uma função quadrática para capturar efeitos decrescentes é que a partir de certo ponto ela acabará fazendo um movimento inverso. Se esse ponto estiver além de uma pequena percentagem das pessoas na amostra, isso não será motivo para grande preocupação. Mas no conjunto de dados do arquivo WAGE1.RAW, cerca de 28% das pessoas na amostra têm mais de 24 anos de experiência; essa é uma percentagem alta demais para se ignorar.

É possível que o retorno de exper realmente se torne negativo em algum ponto, mas é difícil acreditar que isso aconteça aos 24 anos de experiência. Uma possibilidade mais provável é que o efeito estimado de exper sobre saláριο seja viesado, por não termos controlado outros fatores ou porque a relação funcional entre saláριο e exper na equação (6.12) não está totalmente correta. O Problema 6.9 pede que você explore essa possibilidade controlando a educação, além de usar $\log(\text{saláριο})$ como a variável dependente.

Figura 6.1Relação quadrática entre *salárioh* e *exper*.

Quando um modelo tem uma variável dependente na forma logarítmica e uma variável explicativa como uma função quadrática, é necessário certo cuidado para fazer uma boa interpretação. O exemplo seguinte também mostra que a função quadrática pode ter um formato em U, em vez de uma forma parabólica. A forma em U surge na equação (6.10) quando $\hat{\beta}_1$ é negativo e $\hat{\beta}_2$ é positivo; isso captura um efeito crescente de x sobre y .

EXEMPLO 6.2**(Efeitos da Poluição sobre os Preços dos Imóveis)**

Modificamos o modelo dos preços dos imóveis do Exemplo 4.5 para incluir um termo quadrático em *comods*:

$$\begin{aligned} \log(\text{preço}) = & \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \log(\text{dist}) + \beta_3 \text{comods} \\ & + \beta_4 \text{comods}^2 + \beta_5 \text{razestud} + u. \end{aligned} \quad (6.14)$$

O modelo estimado utilizando os dados contidos no arquivo HPRICE2.RAW é

$$\begin{aligned} \log(\text{preço}) = & 13,39 - 0,902 \log(\text{oxn}) - 0,087 \log(\text{dist}) \\ & (0,57) \quad (0,115) \quad (0,043) \end{aligned}$$

EXEMPLO 6.2 (continuação)

$$- 0,545 \text{ comods} + 0,062 \text{ comods}^2 - 0,48 \text{ razestud}$$

$$(0,165) \quad (0,013) \quad (0,006)$$

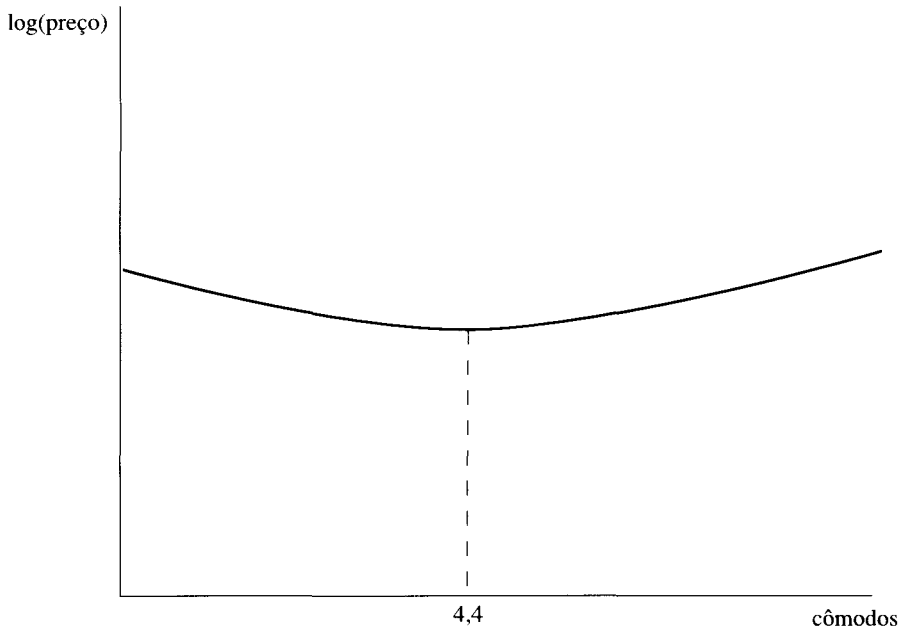
$$n = 506, R^2 = 0,603.$$

O termo quadrático comods^2 tem uma estatística t em torno de 4,77 e, portanto, é estatisticamente bastante significativa. Mas o que é possível afirmar sobre a interpretação do efeito de comods sobre $\log(\text{preço})$? Inicialmente, o efeito parece ser estranho. Como o coeficiente de comods é negativo e o coeficiente de comods^2 é positivo, a equação literalmente sugere que, com valores baixos de comods , um cômodo adicional tem um efeito *negativo* sobre $\log(\text{preço})$. Em algum ponto, o efeito se torna positivo, e a forma quadrática significa que a semi-elasticidade de preço em relação a comods cresce na mesma proporção do crescimento de comods . Esta situação é mostrada na Figura 6.2.

Obtemos o valor do ponto crítico de comods usando a equação (6.13) (embora $\hat{\beta}_1$ seja negativo e $\hat{\beta}_2$ seja positivo). O valor absoluto do coeficiente de comods , 0,545, dividido pelo dobro do coeficiente de comods^2 , 0,062, resulta em $\text{comods}^* = 0,545/[2(0,062)] \approx 4,4$; este ponto está marcado na Figura 6.2.

Figura 6.2

$\log(\text{preço})$ como uma função quadrática de comods .



EXEMPLO 6.2 (continuação)

Será que podemos acreditar que se iniciarmos com três cômodos e aumentarmos para quatro isso efetivamente reduzirá o valor esperado do imóvel? Provavelmente não. Acontece que somente cinco das 506 comunidades na amostra possuem imóveis com média de 4,4 cômodos ou menos, cerca de 1% da amostragem. Isso é tão pequeno que a função quadrática à esquerda de 4,4 pode, para fins práticos, ser ignorada. À direita de 4,4, vemos que a adição de outro cômodo tem um efeito crescente na mudança percentual no preço:

$$\Delta \log(\hat{\text{preço}}) \approx \{-0,545 + 2(0,062)\text{comods}\} \Delta \text{comods}$$

e assim

$$\begin{aligned} \% \Delta \hat{\text{preço}} &\approx 100\{-0,545 + 2(0,062)\text{comods}\} \Delta \text{comods} \\ &= (-54,5 + 12,4\text{comods}) \Delta \text{comods}. \end{aligned}$$

Portanto, um aumento em *comods*, digamos de cinco para seis, aumenta o preço em aproximadamente $-54,5 + 12,4(5) = 7,5\%$; o aumento de seis para sete aumenta o preço em aproximadamente $-54,5 + 12,4(6) = 19,9\%$. Esse é um efeito crescente bastante forte.

Existem muitas outras possibilidades de usar funções quadráticas juntamente com logaritmos. Por exemplo, uma extensão de (6.14) que permita uma elasticidade não-constante entre *preço* e *oxn* é

$$\begin{aligned} \log(\text{preço}) &= \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 [\log(\text{oxn})]^2 \\ &+ \beta_3 \text{crime} + \beta_4 \text{comods} + \beta_5 \text{comods}^2 + \beta_6 \text{razestud} + u. \end{aligned} \quad (6.15)$$

Se $\beta_2 = 0$, β_1 será a elasticidade do *preço* em relação a *oxn*. Caso contrário, essa elasticidade dependerá do nível de *oxn*. Para verificar isso, podemos combinar os argumentos dos efeitos parciais nos modelos quadrático e logarítmico para mostrar que

$$\% \Delta \text{preço} \approx [\beta_1 + 2\beta_2 \log(\text{oxn})] \% \Delta \text{oxn}; \quad (6.16)$$

portanto, a elasticidade do *preço* em relação a *oxn* é $\beta_1 + 2\beta_2 \log(\text{oxn})$, de forma que ela depende de $\log(\text{oxn})$.

Finalmente, outros termos polinomiais podem ser incluídos nos modelos de regressão. Certamente a função quadrática é vista com mais freqüência, mas um termo cúbico ou até de quarta potência aparece de vez em quando. Uma forma funcional freqüentemente razoável de uma função de custo total é

$$\text{custo} = \beta_0 + \beta_1 \text{quantidade} + \beta_2 \text{quantidade}^2 + \beta_3 \text{quantidade}^3 + u.$$

Não é complicado estimar este modelo. A interpretação dos parâmetros é mais trabalhosa (embora objetiva com o uso de cálculo infinitesimal); não estudaremos este modelo com mais detalhes.

Modelos com Termos de Interação

Algumas vezes, é natural que o efeito parcial, a elasticidade, ou a semi-elasticidade da variável dependente, em relação a uma variável explicativa, dependa da magnitude de *outra* variável explicativa. Por exemplo, no modelo

$$\text{preço} = \beta_0 + \beta_1 \text{arquad} + \beta_2 \text{qtdorm} + \beta_3 \text{arquad} \cdot \text{qtdorm} + \beta_4 \text{banhos} + u,$$

o efeito parcial de *qtdorm* sobre *preço* (mantendo fixas todas as outras variáveis) é

$$\frac{\Delta \text{preço}}{\Delta \text{qtdorm}} = \beta_2 + \beta_3 \text{arquad}. \quad (6.17)$$

Se $\beta_3 > 0$, (6.17) sugere que um quarto a mais produz um aumento maior no preço dos imóveis maiores. Em outras palavras, existe um **efeito de interação** entre a área do imóvel e o número de quartos. Ao resumirmos o efeito de *qtdorm* sobre *preço*, devemos avaliar (6.17) quanto aos valores de interesse de *arquad*, como o valor médio, ou os quartis inferior ou superior na amostra. Se β_3 é zero ou não, é algo que podemos verificar facilmente.

Pode ser complicado interpretar os parâmetros das variáveis originais quando incluímos um termo de interação. Por exemplo, na equação anterior sobre preços de imóveis, a equação (6.17) mostra que β_2 é o efeito de *qtdorms* sobre *preço* para um preço com zero de área construída! Esse efeito, obviamente, não é muito interessante. Em vez disso, devemos ser cuidadosos ao inserirmos valores de interesse da área do imóvel, como o valor médio ou a mediana da amostra, na versão estimada da equação (6.17).

Freqüentemente, é vantajoso reparametrizar um modelo para que os coeficientes das variáveis originais tenham significados interessantes. Considere um modelo com duas variáveis explicativas e uma interação:

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

Como acabamos de mencionar, β_2 é o efeito parcial de x_2 quando $x_1 = 0$. Muitas vezes, isso não é de interesse. Em vez disso, podemos reparametrizar o modelo como

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u,$$

onde μ_1 é a média populacional de x_1 e μ_2 é a média populacional de x_2 . Podemos facilmente ver que agora o coeficiente de x_2 , δ_2 , é o efeito parcial de x_2 sobre y no valor médio de x_1 . (Multiplicando a interação na segunda equação e comparando os coeficientes, podemos mostrar com facilidade que $\delta_2 = \beta_2 + \beta_3 \mu_1$. O parâmetro δ_1 tem uma interpretação semelhante.) Portanto, se subtrairmos as médias das variáveis — na prática, elas seriam, tipicamente, as médias da amostra — antes de criarmos o termo de interação, os coeficientes das variáveis originais terão uma interpretação útil. E mais, obteremos imediatamente os erros-padrão dos efeitos parciais ao nível dos valores médios. Nada nos impede de substituir μ_1 ou μ_2 por outros valores das variáveis explicativas que possam ser de interesse. O exemplo seguinte ilustra como podemos usar os termos de interação.

EXEMPLO 6.3**(Efeitos da Freqüência Escolar no Desempenho de Exames Finais)**

Um modelo para explicar o resultado padronizado de um exame final (*respad*) em termos da taxa de freqüência escolar, da nota média anterior* ao curso superior até o penúltimo semestre, e da nota do teste de avaliação de conhecimentos para ingresso em curso superior é

$$\begin{aligned} \text{respad} = \beta_0 + \beta_1 \text{taxafreq} + \beta_2 \text{nmgradp} + \beta_3 \text{tac} + \beta_4 \text{nmgradp}^2 \\ + \beta_5 \text{tac}^2 + \beta_6 \text{nmgradp} \cdot \text{taxafreq} + u. \end{aligned} \quad (6.18)$$

(Utilizamos o resultado padronizado do exame pelos motivos discutidos na Seção 6.1: é mais fácil interpretar o desempenho de um aluno em relação ao restante da classe.) Além dos termos quadráticos em *nmgradp* e *tac*, este modelo inclui uma interação entre *nmgradp* e *taxafreq*. A idéia é que a freqüência às aulas pode ter um efeito diferente nos alunos que tiveram desempenhos diferentes no passado, como medido por *nmgradp*. Estamos interessados nos efeitos da freqüência sobre as notas do exame final: $\Delta \text{respad} / \Delta \text{taxafreq} = \beta_1 + \beta_6 \text{nmgradp}$.

Usando as 680 observações do arquivo ATTEND.RAW para estudantes da área de economia, a equação estimada é

$$\begin{aligned} \widehat{\text{respad}} = 2,05 - 0,0067 \text{taxafreq} - 1,63 \text{nmgradp} - 0,128 \text{tac} \\ (1,36) \quad (0,0102) \quad (0,48) \quad (0,098) \\ + 0,296 \text{nmgradp}^2 + 0,0045 \text{tac}^2 + 0,0056 \text{nmgradp} \cdot \text{taxafreq} \quad (6.19) \\ (0,101) \quad (0,0022) \quad (0,0043) \\ n = 680, R^2 = 0,229, \bar{R}^2 = 0,222. \end{aligned}$$

Devemos interpretar essa equação com extremo cuidado. Se simplesmente olharmos o coeficiente de *taxafreq*, concluiremos de forma errônea que a freqüência tem um efeito negativo na nota do exame final. Porém, esse coeficiente supostamente mede o efeito quando *nmgradp* = 0, o que não é interessante (nessa amostra, a menor nota média do ensino médio é cerca de 0,86). Também devemos ter cuidado para não examinarmos separadamente as estimativas de β_1 e β_6 e concluirmos que, como cada estatística *t* é não significativa, não podemos rejeitar $H_0: \beta_1 = 0, \beta_6 = 0$. Aliás, o *p*-valor do teste *F* dessa hipótese conjunta é 0,014, de modo que com certeza rejeitamos H_0 ao nível de 5%. Este é um bom exemplo de como o exame em separado de estatísticas *t*, quando estamos testando uma hipótese conjunta, pode nos levar a equívocos.

Como devemos estimar o efeito parcial de *taxafreq* sobre *respad*? Devemos inserir valores de interesse de *nmgradp* para obter o efeito parcial. O valor médio de *nmgradp* na amostra é 2,59, de modo que nesse valor médio o efeito de *taxafreq* sobre *respad* é $-0,0067 + 0,0056(2,59) \approx 0,0078$. Qual o significado disso? Como *taxafreq* é medida como um percentual, isso significa que um aumento de dez pontos percentuais em *taxafreq* aumenta *respad* em 0,078 desvios-padrão da nota média do exame final.

Como podemos dizer se a estimativa 0,0078 é estatisticamente diferente de zero? Temos que computar novamente a regressão, substituindo *nmgradp* · *taxafreq* por $(\text{nmgradp} - 2,59) \cdot \text{taxafreq}$. Isso produz, como o novo coeficiente de *taxafreq*, o efeito estimado quando *nmgradp* = 2,59, juntamente com seu erro-padrão; nada mais é alterado na regressão. (Descrevemos esse mecanismo na Seção 4.4.) A execução dessa nova regressão fornece o erro-padrão de $\hat{\beta}_1 + \hat{\beta}_6(2,59) = 0,0078$ como 0,0026, o que produz

EXEMPLO 6.3 (continuação)

$t = 0,0078/0,026 = 3$. Portanto, na *nmgradp* média, concluímos que a taxa de frequência às aulas tem um efeito positivo estatisticamente significante nas notas do exame final.

É ainda mais complicado encontrar o efeito de *nmgradp* sobre *respad*, devido ao termo quadrático $nmgradp^2$. Para encontrar o efeito no valor médio de *nmgradp* e na taxa média de frequência, 0,82, substituímos $nmgradp^2$ por $(nmgradp - 2,59)^2$ e $nmgradp \cdot taxafreq$ por $nmgradp \cdot (taxafreq - 0,82)$. O coeficiente de *nmgradp* se tornará o efeito parcial nos valores médios e obteremos seu erro-padrão. (Veja o Problema 6.14.)

Se adicionarmos o termo $\beta_7 taxafreq$ à equação (6.18), qual será o efeito parcial de *taxafreq* sobre *respad*?

6.3 UM POUCO MAIS SOBRE O GRAU DE AJUSTE E A SELEÇÃO DE REGRESSORES

Até agora, não dedicamos muita atenção ao tamanho do R^2 na avaliação de nossos modelos de regressão, porque estudantes iniciantes tendem a colocar muito peso no R -quadrado. Como em breve veremos, a seleção de um conjunto de variáveis explicativas com base no tamanho do R -quadrado pode levar a modelos absurdos. No Capítulo 10 descobriremos que R -quadrados obtidos de regressões de séries temporais podem ser artificialmente altos e podem resultar em conclusões enganosas.

Nada nas hipóteses do modelo linear clássico exige que o R^2 esteja acima de qualquer valor em particular; o R^2 é simplesmente uma estimativa do quanto da variação em y é explicado por x_1, x_2, \dots, x_k na população. Vimos várias regressões que tinham R -quadrados bastante pequenos. Embora isso signifique que não tenhamos avaliado vários fatores que afetam y , isso não quer dizer que os fatores em u sejam correlacionados com as variáveis independentes. A hipótese de média condicional zero RLM.3 é a que determina se obteremos estimadores não-viesados dos efeitos *ceteris paribus* das variáveis independentes, e o tamanho do R -quadrado não tem influência direta nisso.

Um R -quadrado pequeno sugere que a variância do erro é grande em relação à variância de y , o que significa que podemos ter muito trabalho para estimar β_j com precisão. Porém, lembre-se: vimos na Seção 3.4 que uma variância grande do erro pode ser compensada por uma amostra de tamanho grande: se tivermos dados suficientes, podemos ter condições de estimar com precisão os efeitos parciais, mesmo que não tenhamos controlado muitos dos fatores não-observados. Se podemos ou não obter estimativas suficientemente precisas, depende da aplicação que estamos pesquisando. Por exemplo, suponha que alguns alunos ingressantes de uma grande universidade recebam, aleatoriamente, subsídios para a compra de computadores. Se o montante do subsídio for determinado de forma realmente aleatória, podemos estimar o efeito *ceteris paribus* do montante do subsídio sobre a nota de aproveitamento dos alunos, com o uso de uma análise de regressão simples. (Devido à atribuição aleatória, todos os outros fatores que afetam a nota de aproveitamento seriam não-correlacionados com o montante do subsídio.) Parece provável que o montante de subsídio explique pouco da variação na nota de aproveitamento, de modo que o R -quadrado de tal regressão provavelmente será muito pequeno. Mas se tivermos uma amostra de grande tamanho, ainda poderemos ter condições de obter uma estimativa razoavelmente precisa do efeito do subsídio.

Lembre-se, porém, de que a *mudança* relativa no R -quadrado, quando variáveis são adicionadas à equação, é muito útil: a estatística F em (4.41) para testar a significância conjunta depende de forma crucial da diferença nos R -quadrados entre o modelo sem restrições e o modelo restrito.

O R -Quadrado Ajustado

A maioria dos programas econométricos registra, juntamente com o R -quadrado, uma estatística chamada **R -quadrado ajustado**. Como o R -quadrado ajustado é descrito em muitos trabalhos aplicados, e como ele tem algumas características úteis, trataremos dele nesta subseção.

Para verificar como o R -quadrado usual pode ser ajustado, é útil escrevê-lo como

$$R^2 = 1 - (\text{SQR}/n)/(\text{SQT}/n), \quad (6.20)$$

onde SQR é a soma dos resíduos quadrados e SQT é a soma dos quadrados total; comparada com a equação (3.28), tudo o que fizemos foi dividir tanto SQR como SQT por n . Essa expressão revela o que R^2 está realmente estimando. Defina σ_y^2 como a variância populacional de y e faça com que σ_u^2 represente a variância populacional do termo erro, u . (Até agora temos usado σ^2 para representar σ_u^2 , mas é vantajoso ser mais específico neste caso.) O **R -quadrado da população** é definido como $1 - \sigma_u^2/\sigma_y^2$; essa é a proporção da variação em y na população explicada pelas variáveis independentes. Isso é o que, supostamente, R^2 deve estar estimando.

O R^2 estima σ_u^2 por SQR/n , que sabemos ser viesado. Então, por que não substituir SQR/n por $\text{SQR}/(n - k - 1)$? Além disso, podemos usar $\text{SQT}/(n - 1)$ em lugar de SQT/n , já que o primeiro é o estimador não-viesado de σ_y^2 . Usando esses estimadores, chegamos ao R -quadrado ajustado:

$$\begin{aligned} \bar{R}^2 &= 1 - [\text{SQR}/(n - k - 1)]/[\text{SQT}/(n - 1)] \\ &= 1 - \hat{\sigma}^2/[\text{SQT}/(n - 1)], \end{aligned} \quad (6.21)$$

já que $\hat{\sigma}^2 = \text{SQR}/(n - k - 1)$. Devido à notação usada para representar o R -quadrado ajustado, ele é, algumas vezes, chamado de *R -barra-quadrado*.

O R -quadrado ajustado algumas vezes é chamado de *R -quadrado corrigido*, mas esse não é um bom nome, pois sugere que \bar{R}^2 é de alguma forma melhor que R^2 como um estimador do R -quadrado da população. Infelizmente, \bar{R}^2 não é reconhecido, de forma geral, como um melhor estimador. É tentador imaginar que \bar{R}^2 corrige o viés de R^2 na estimativa do R -quadrado da população, mas ele não faz isso: a razão de dois estimadores não-viesados não é um estimador não-viesado.

O ponto mais atraente do \bar{R}^2 é que ele impõe uma penalidade à inclusão de variáveis independentes adicionais em um modelo. Sabemos que R^2 nunca pode diminuir quando uma nova variável independente é incluída em uma equação de regressão: isso ocorre porque SQR nunca aumenta (e normalmente diminui) quando novas variáveis independentes são adicionadas. Mas a fórmula do \bar{R}^2 mostra que ele depende explicitamente de k , o número de variáveis independentes. Se uma variável independente for adicionada a uma regressão, SQR diminui, mas o mesmo acontece com os gl na regressão, $n - k - 1$. Portanto, $\text{SQR}/(n - k - 1)$ pode aumentar ou diminuir quando uma nova variável independente é adicionada a uma regressão.

Um fato algébrico interessante é o seguinte: se adicionarmos uma nova variável independente a uma equação de regressão, R^2 aumenta se, e somente se, a estatística t da nova variável for maior que um em valor absoluto. (Uma extensão disto é que \bar{R}^2 aumenta quando um grupo de variáveis é adicionado a uma regressão se, e somente se, a estatística F da significância conjunta das novas variáveis for maior

que a unidade.) Assim, vemos imediatamente que usar o \bar{R}^2 para decidir se determinada variável independente (ou conjunto de variáveis) pertence a um modelo nos fornece uma resposta diferente daquelas fornecidas pelos testes usuais t ou F (já que uma estatística t ou F igual à unidade não é estatisticamente significativa aos níveis tradicionais de significância).

Algumas vezes é útil termos uma fórmula do \bar{R}^2 em termos de R^2 . A álgebra simples mostra que

$$\bar{R}^2 = 1 - (1 - R^2)(n - 1)/(n - k - 1). \quad (6.22)$$

Por exemplo, se $R^2 = 0,30$, $n = 51$, e $k = 10$, então $\bar{R}^2 = 1 - 0,70(50)/40 = 0,125$. Assim, para n pequeno e k grande, \bar{R}^2 pode estar substancialmente abaixo de R^2 . De fato, se o R -quadrado normal for pequeno, e $n - k - 1$ for pequeno, \bar{R}^2 pode, na realidade, ser negativo! Por exemplo, podemos considerar $R^2 = 0,10$, $n = 51$, e $k = 10$ para verificar que $\bar{R}^2 = -0,125$. Um \bar{R}^2 negativo indica uma adaptação muito pobre do modelo relativamente ao número de graus de liberdade.

O R -quadrado ajustado algumas vezes é descrito junto com o R -quadrado habitual em regressões, e algumas vezes o \bar{R}^2 é descrito em lugar do R^2 . É importante lembrar que é o R^2 , e não o \bar{R}^2 , que aparece na estatística F em (4.41). A mesma fórmula com \bar{R}_r^2 e \bar{R}_{ir}^2 não é válida.

O Uso do R -quadrado Ajustado para a Escolha entre Modelos Não-Aninhados

Na Seção 4.5 aprendemos como calcular uma estatística F para testar a significância conjunta de um grupo de variáveis; isso nos possibilita decidir, em um nível particular de significância, se pelo menos uma variável no grupo afeta a variável dependente. Esse teste não nos permite decidir *qual* das variáveis tem um efeito. Em alguns casos, queremos escolher um modelo sem variáveis independentes redundantes, e o R -quadrado ajustado pode nos ajudar nessa tarefa.

No exemplo dos salários dos jogadores da principal liga de beisebol na Seção 4.5, vimos que nem *hrunano* nem *rebrunano* eram individualmente significantes. Essas duas variáveis são altamente correlacionadas, de modo que podemos querer optar entre os modelos

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \beta_4 \text{hrunano} + u$$

e

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \beta_4 \text{rebrunano} + u.$$

Esses dois exemplos são **modelos não-aninhados**, pois nenhuma equação é um caso especial da outra. A estatística F , que estudamos no Capítulo 4, nos permite testar somente modelos *aninhados*: um modelo (o modelo restrito) é um caso especial do outro modelo (o modelo sem restrições). Veja as equações (4.32) e (4.28) para exemplos dos modelos restritos e sem restrições. Uma possibilidade é criar um modelo combinado que contenha *todas* as variáveis explicativas dos modelos originais e depois testar cada modelo contra o modelo geral usando o teste F . O problema deste processo é que ambos os modelos poderão ser rejeitados, ou nenhum modelo poderá ser rejeitado (como acontece com o exemplo dos salários dos jogadores da principal liga de beisebol na Seção 4.5). Assim, esse processo nem sempre fornece uma maneira de fazermos a distinção entre modelos com regressores não-aninhados.

Na regressão dos salários dos jogadores de beisebol, o \bar{R}^2 da regressão contendo *hrunano* é 0,6211 e o \bar{R}^2 da regressão contendo *rebrunano* é 0,6226. Portanto, com base no R -quadrado ajustado, existe uma preferência muito pequena para o modelo com *rebrunano*. Mas a diferença, na prática, é muito pequena, e podemos obter uma resposta diferente controlando algumas das variáveis no Problema 4.16. (Como ambos os modelos não-aninhados contêm cinco parâmetros, o R -quadrado habitual pode ser usado para fornecer a mesma conclusão.)

A comparação dos \bar{R}^2 para optarmos entre os diferentes conjuntos não-aninhados de variáveis independentes pode ser de grande valia quando essas variáveis representam formas funcionais diferentes. Considere dois modelos relacionando a intensidade de P&D às vendas de uma empresa:

$$pdintens = \beta_0 + \beta_1 \log(vendas) + u. \quad (6.23)$$

$$pdintens = \beta_0 + \beta_1 vendas + \beta_2 vendas^2 + u. \quad (6.24)$$

O primeiro modelo captura um rendimento decrescente pela inclusão de *vendas* na forma logarítmica; o segundo modelo faz isso com o uso de um termo quadrático. Assim, o segundo modelo contém um parâmetro a mais que o primeiro.

Quando as equações são estimadas usando as 32 observações das empresas de produtos químicos do arquivo RDCHEM.RAW, o R^2 é 0,061, e o R^2 da equação (6.24) é 0,148. Portanto, parece que a função quadrática faz um ajuste muito melhor. Mas uma comparação dos R -quadrados habituais com o primeiro modelo é injusta, porque ele contém um parâmetro a menos que a equação (6.24). Isto é, (6.23) é um modelo mais parcimonioso que (6.24).

Tudo o mais igual, modelos mais simples são melhores. Como o R -quadrado habitual não penaliza modelos mais complicados, é melhor usar o \bar{R}^2 . O \bar{R}^2 de (6.23) é 0,030, enquanto o \bar{R}^2 de (6.24) é 0,090. Portanto, mesmo depois dos ajustes das diferenças nos graus de liberdade, o modelo quadrático é o melhor. O modelo quadrático também é o preferido quando margens de lucro são incluídas em cada regressão.

Existe uma limitação importante no uso do \bar{R}^2 para escolher entre modelos não-aninhados: não podemos usá-lo para a escolha entre diferentes formas funcionais da variável dependente. Isso é uma pena, pois muitas vezes queremos saber se y ou $\log(y)$ (ou talvez alguma outra transformação) deve ser usada como a variável dependente com base no grau de ajuste. Mas nem o R^2 nem o \bar{R}^2 podem ser usados para esse fim. A razão é simples: esses R -quadrados medem a proporção explicada do total da variação de qualquer variável dependente que estejamos usando na regressão, e diferentes funções da variável dependente terão diferentes montantes de variação a ser explicadas. Por exemplo, as variações totais em y e $\log(y)$ não são as mesmas. A comparação dos R -quadrados ajustados dessas regressões com essas diferentes formas das variáveis dependentes não nos dá nenhuma informação sobre qual modelo se adapta melhor; eles estimam duas variáveis dependentes separadas.

Explique por que escolher um modelo maximizando \bar{R}^2 ou minimizando $\hat{\sigma}$ (o erro-padrão da regressão) é a mesma coisa.

EXEMPLO 6.4**(Remuneração de Diretores Executivos e Desempenho de Empresas)**

Considere dois modelos estimados relacionando a remuneração de diretores executivos ao desempenho de empresas:

$$\begin{aligned} \widehat{\text{salário}} &= 830,63 + 0,0163 \text{ vendas} + 19,63 \text{ rma} \\ &\quad (223,90) \quad (0,0089) \quad (11,08) \\ n &= 209, R^2 = 0,029, \bar{R}^2 = 0,20 \end{aligned} \quad (6.25)$$

e

$$\begin{aligned} \ln \widehat{\text{salário}} &= 4,36 + 0,275 \ln \text{vendas} + 0,0179 \text{ rma} \\ &\quad (0,29) \quad (0,033) \quad (0,0040) \\ n &= 209, R^2 = 0,282, \bar{R}^2 = 0,275, \end{aligned} \quad (6.26)$$

onde *rma* é o retorno das ações, discutido no Capítulo 2. Para simplificar, *lsalário* e *lvendas* representam os logs naturais de *salário* e *vendas*. Já sabemos como interpretar essas diferentes equações estimadas. Mas podemos dizer se um modelo ajusta os dados melhor que o outro?

O *R*-quadrado da equação (6.25) mostra que *vendas* e *rma* explicam somente cerca de 2,9% da variação do salário dos diretores executivos na amostra. Tanto *vendas* como *rma* têm significância estatística marginal.

A equação (6.26) mostra que $\log(\text{vendas})$ e *rma* explicam cerca de 28,2% da variação do $\log(\text{salário})$. Em termos de grau de ajuste, esse *R*-quadrado bem mais alto parece sugerir que o modelo (6.26) é bem melhor, mas esse não é necessariamente o caso. A soma dos quadrados total de *salário* na amostra é 391.732.982, enquanto a soma dos quadrados total de $\log(\text{salário})$ é somente 66,72. Assim, há muito menos variação em $\log(\text{salário})$ que precisa ser explicada.

Neste ponto, podemos usar outros recursos além do R^2 e do \bar{R}^2 para optar entre esses modelos. Por exemplo, $\log(\text{vendas})$ e *rma* são muito mais significantes, estatisticamente, em (6.26) do que são *vendas* e *rma* em (6.25), e os coeficientes em (6.26) provavelmente são de maior interesse. Para termos certeza, porém, precisaremos fazer uma comparação válida dos graus de ajuste.

Na Seção 6.4 forneceremos um indicador que efetivamente nos permita comparar modelos nos quais y aparece tanto na forma em nível como na forma logarítmica.

O Controle de muitos Fatores na Análise de Regressão

Em muitos dos exemplos que tratamos, e certamente em nossa discussão sobre o viés de variáveis omitidas no Capítulo 3, temos nos preocupado com a omissão de fatores importantes em modelos que possam estar correlacionados com as variáveis independentes. Também é possível controlarmos grande quantidade de variáveis em uma análise de regressão.

Se enfatizarmos exageradamente o grau de ajuste, estaremos nos propondo a controlar fatores em um modelo de regressão que não deveriam ser controlados. Para evitar este equívoco, precisamos nos lembrar da interpretação *ceteris paribus* de modelos de regressão múltipla.

Para ilustrar esse problema, suponha que estejamos fazendo um estudo para avaliar o impacto dos impostos estaduais sobre a cerveja em acidentes fatais de trânsito. A idéia é que um imposto mais elevado sobre a cerveja reduzirá o consumo de bebidas alcoólicas e, da mesma forma, o hábito de dirigir sob embriaguez, resultando em menos acidentes fatais de trânsito. Para medirmos o efeito *ceteris paribus* dos impostos sobre esses acidentes, podemos modelar *fatalidades* como uma função de diversos fatores, inclusive o *imposto* sobre a cerveja:

$$fatalidades = \beta_0 + \beta_1 imposto + \beta_2 milhas + \beta_3 perc masc + \beta_4 perc16_21 + \dots,$$

onde *milhas* é o total de milhas dirigidas, *perc masc* é a percentagem masculina da população do Estado, e *perc16_21* é a percentagem da população entre 16 e 21 anos de idade, e assim por diante. Observe que não incluímos uma variável medindo o consumo *per capita* de cerveja. Estaremos cometendo um erro de variáveis omitidas? A resposta é não. Se controlarmos o consumo de cerveja nessa equação, de que forma o imposto sobre cerveja afetará as fatalidades no trânsito? Na equação

$$fatalidades = \beta_0 + \beta_1 imposto + \beta_2 conscerv + \dots,$$

β_1 mede a diferença nas fatalidades devido ao aumento de um ponto percentual no imposto, mantendo *conscerv* fixo. É difícil de entender por que isso seria de interesse. Não deveríamos controlar as diferenças de *conscerv* entre os Estados, a menos que quiséssemos verificar algum tipo de efeito indireto do imposto sobre a cerveja. Outros fatores, como a distribuição por sexo e idade, deveriam ser controlados.

A questão de decidir se devemos ou não controlar certos fatores nem sempre é bem definida. Por exemplo, Betts (1995) estuda o efeito da qualidade do ensino médio sobre a renda subsequente. Ele salienta que, se qualidade melhor de ensino resulta em mais educação, então controlar a educação na regressão juntamente com avaliação da qualidade subestimar o retorno da qualidade. Betts faz a análise com e sem anos de escolaridade na equação para obter uma gama de efeitos estimados da qualidade de ensino.

Para verificar claramente como a ênfase em *R*-quadrados altos pode criar problemas, considere o exemplo do preço dos imóveis da Seção 4.5 que ilustra a verificação de múltiplas hipóteses. Naquele caso, queríamos verificar a racionalidade da avaliação dos preços dos imóveis. Fizemos a regressão de $\log(\text{preço})$ sobre $\log(\text{aval})$, $\log(\text{tamterr})$, $\log(\text{arquad})$, e qtdorm e verificamos se as três últimas variáveis tinham coeficientes populacionais zero enquanto $\log(\text{aval})$ tinha um coeficiente unitário. Entretanto, se quisermos estimar um modelo de preço hedônico, como no Exemplo 4.8, onde os valores marginais de várias características dos imóveis são obtidos? Devemos incluir $\log(\text{aval})$ na equação? O *R*-quadrado ajustado da regressão com $\log(\text{aval})$ é 0,762, enquanto o *R*-quadrado ajustado sem ele é 0,630. Com base somente no grau de ajuste, devemos incluir $\log(\text{aval})$. Contudo, isso será incorreto se nossa meta for determinar os efeitos do tamanho da propriedade, área construída e número de quartos nos valores dos imóveis. A inclusão de $\log(\text{aval})$ na equação equivale a manter um indicador de valor fixo e indagar quanto a adição de um quarto alterará outro indicador de valor. Essa medida não faz sentido na avaliação das características dos imóveis.

Se lembrarmos que modelos diferentes servem a propósitos diferentes, e nos concentrarmos na interpretação *ceteris paribus* da regressão, não incluiremos os fatores errados em um modelo de regressão.

A Adição de Regressores para Reduzir a Variância do Erro

Acabamos de ver alguns exemplos nos quais certas variáveis independentes não devem ser incluídas em um modelo de regressão, mesmo que elas sejam correlacionadas com a variável dependente. Do Capítulo 3, sabemos que a adição de uma nova variável independente em uma regressão pode exa-

cerbar o problema da multicolinearidade. De outro lado, como estamos retirando algo do termo erro, a adição de uma variável geralmente reduz a variância do erro. De forma geral, não podemos saber que efeito será dominante.

Porém, há um caso que é óbvio: devemos sempre incluir variáveis independentes que afetem y e que sejam *não-correlacionadas* com todas as variáveis independentes de interesse. A razão para esta inclusão é simples: a adição dessa variável não induz multicolinearidade na população (e, portanto, a multicolinearidade na amostra deve ser desprezível), mas reduzirá a variância do erro. Em amostras de tamanho grande, os erros-padrão de todos os estimadores MQO serão reduzidos.

Como exemplo, considere estimar a demanda individual por cerveja como uma função do preço médio da cerveja no município. Pode ser razoável assumir que as características individuais sejam não-correlacionadas com os preços em nível de municípios, e assim uma regressão simples do consumo de cerveja sobre o preço nos municípios seria suficiente para estimar o efeito do preço sobre a demanda individual. Entretanto, é possível obter uma estimativa mais precisa da elasticidade-preço da demanda por cerveja com a inclusão de características individuais, como a idade e o grau de escolaridade. Se esses fatores afetarem a demanda e forem não-correlacionados com o preço, o erro-padrão do coeficiente do preço será menor, pelo menos em amostras grandes.

Como segundo exemplo, considere o subsídio para equipamentos de computação dado no início da Seção 6.3. Se, além da variável do subsídio, controlarmos outros fatores que possam explicar a nota média em curso superior, poderemos provavelmente conseguir uma estimativa mais precisa do efeito do subsídio. Variáveis indicadoras da nota média no ensino médio, a classificação da instituição, as pontuações sat e tac e os antecedentes familiares são bons candidatos. Como os montantes do subsídio são determinados aleatoriamente, todas as variáveis de controle adicionais serão não-correlacionadas com o montante de subsídio; nessa amostra, a multicolinearidade entre o montante do subsídio e as outras variáveis independentes deve ser mínima. Porém, a adição de controles extras pode reduzir significativamente a variância do erro, conduzindo a uma estimativa mais precisa do efeito do subsídio. Lembre-se, neste caso, de que o problema não é a inexistência de viés: obteremos um estimador não-viesado e consistente, quer incluamos ou não as variáveis de desempenho no ensino médio e de antecedentes familiares. O problema está na obtenção de um estimador com uma menor variância amostral.

Infelizmente, casos em que temos informações sobre as variáveis explicativas adicionais que sejam não-correlacionadas com as variáveis explicativas de interesse são raros no campo das ciências sociais. Porém, vale a pena lembrar que, quando essas variáveis estão disponíveis, elas poderão ser incluídas em um modelo para reduzir a variância do erro sem induzir multicolinearidade.

6.4 PREVISÃO E ANÁLISE DE RESÍDUOS

No Capítulo 3 definimos os valores previstos ou estimados do MQO e os resíduos do MQO. As **previsões** certamente são úteis, mas estão sujeitas à variação amostral, já que elas são obtidas com o uso dos estimadores MQO. Assim, nesta seção, mostramos como obter intervalos de confiança de previsões da linha de regressão MQO.

Sabemos, dos capítulos 3 e 4, que os resíduos são usados para obter a soma dos resíduos quadrados e o R -quadrado, de modo que eles são importantes para o grau de ajuste e os testes de hipóteses.

Algumas vezes, os economistas estudam os resíduos de uma observação específica para obter informações sobre os indivíduos (ou empresas, imóveis etc.) na amostra.

Intervalos de Confiança de Previsões

Suponha que tenhamos estimado a equação

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (6.27)$$

Quando inserimos valores específicos das variáveis independentes nessa equação, obtemos uma previsão de y , que é uma estimativa do *valor esperado* de y , dados os valores específicos das variáveis explicativas. Para enfatizar, sejam c_1, c_2, \dots, c_k valores particulares de cada uma das k variáveis independentes; elas poderão ou não corresponder a um ponto efetivo dos dados em nossa amostra. O parâmetro que gostaríamos de estimar é

$$\begin{aligned} \theta_0 &= \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k \\ &= E(y | x_1 = c_1, x_2 = c_2, \dots, x_k = c_k). \end{aligned} \quad (6.28)$$

O estimador de θ_0 é

$$\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \dots + \hat{\beta}_k c_k. \quad (6.29)$$

Na prática, isso é fácil de ser computado. Entretanto, se o que quisermos for um indicador da incerteza nesse valor previsto? É natural construir um intervalo de confiança de θ_0 que seja centrado em $\hat{\theta}_0$.

Para obter um intervalo de confiança de θ_0 precisamos de um erro-padrão de $\hat{\theta}_0$. Então, com um grande gl , poderemos construir um intervalo de confiança de 95% utilizando a regra prática $\hat{\theta}_0 \pm 2 \cdot \text{ep}(\hat{\theta}_0)$. (Como sempre, podemos usar os percentis exatos em uma distribuição t .)

Como obtemos o erro-padrão de $\hat{\theta}_0$? Este é o mesmo problema que encontramos na Seção 4.4: precisamos obter um erro-padrão de uma combinação linear dos estimadores MQO. Aqui, o problema é ainda mais complicado, pois todos os estimadores MQO geralmente aparecem em $\hat{\theta}_0$ (a menos que algum c_j seja zero). No entanto, o mesmo truque que usamos na Seção 4.4 funcionará aqui. Escreva $\beta_0 = \theta_0 - \beta_1 c_1 - \dots - \beta_k c_k$ e agregue isso à equação

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

para obter

$$y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + \dots + \beta_k(x_k - c_k) + u. \quad (6.30)$$

Em outras palavras, subtraímos o valor c_j de cada observação de x_j , e depois computamos a regressão de

$$y_i \text{ sobre } (x_{i1} - c_1), \dots, (x_{ik} - c_k), i = 1, 2, \dots, n. \quad (6.31)$$

O valor previsto em (6.29) e, mais importante, seu erro-padrão, são obtidos do *intercepto* (ou constante) na regressão (6.31).

Como exemplo, obtemos um intervalo de confiança de uma previsão a partir de uma regressão de nota média em curso superior, das quais usamos informações do ensino médio.

EXEMPLO 6.5**(Intervalo de Confiança de $nmgrad$ Previsto)**

Utilizando os dados contidos no arquivo GPA2.RAW, obtemos a seguinte equação para prever $nmgrad$:

$$\begin{aligned} nmgrad = & 1,493 + 0,00149 sat - 0,01386 emperc \\ & (0,075) \quad (0,00007) \quad (0,00056) \\ & - 0,06088 tamclas + 0,00546 tamclas^2 \\ & (0,1650) \quad (0,00227) \end{aligned} \quad (6.32)$$

$$n = 4.137, R^2 = 0,278, \bar{R}^2 = 0,277, \hat{\sigma} = 0,560,$$

onde apresentamos as estimativas com várias casas decimais para reduzir o erro de arredondamento. Qual a previsão de $nmgrad$ quando $sat = 1.200$, $emperc = 30$ e $tamclas = 5$ (o que significaria 500)? Isso é fácil de ser obtido, incorporando esses valores na equação (6.32): $nmgrad = 2,70$ (arredondado para duas casas decimais). Infelizmente, não podemos usar diretamente a equação (6.32) para obter um intervalo de confiança da $nmgrad$ esperada com os valores dados das variáveis independentes. Uma maneira simples de obter um intervalo de confiança é definir um novo conjunto de variáveis independentes: $sat0 = sat - 1.200$, $emperc0 = emperc - 30$, $tamclas0 = tamclas - 5$ e $tamclasquad0 = tamclas^2 - 25$. Quando fazemos a regressão de $nmgrad$ sobre essas novas variáveis independentes, obtemos

$$\begin{aligned} nmgrad = & 2,700 + 0,00149 sat0 - 0,01386 emperc0 \\ & (0,020) \quad (0,00007) \quad (0,00056) \\ & - 0,06088 tamclas0 + 0,00546 tamclasquad0 \\ & (0,1650) \quad (0,00227) \end{aligned}$$

$$n = 4.137, R^2 = 0,278, \bar{R}^2 = 0,277, \hat{\sigma} = 0,560.$$

A única diferença entre esta regressão e aquela em (6.32) é o intercepto, que é a previsão que queremos, juntamente com seu erro-padrão, 0,020. Não é por acidente que os coeficientes de inclinação, seus erros-padrão, R -quadrado etc. são os mesmos de antes; esse fato fornece uma maneira de verificarmos se foram feitas as transformações adequadas. Podemos construir com facilidade um intervalo de confiança de 95% da nota média esperada: $2,70 \pm 1,96(0,020)$ ou em torno de 2,66 a 2,74. Este intervalo de confiança é suficientemente estreito devido ao tamanho bastante grande da amostra.

Como a variância do estimador do intercepto é a menor quando cada variável explicativa tem média amostral zero (veja a Questão 2.5 para o caso da regressão simples), segue da regressão em (6.31) que a variância da previsão nos valores médios de x_j (isto é, $c_j = \bar{x}_j$ para todo j) é a menor. Este resultado não é tão surpreendente, já que o ponto de maior confiabilidade em nossa linha de regressão está próximo ao centro dos dados. Na medida em que os valores de c_j se afastam de \bar{x}_j , $\text{Var}(\hat{y})$ se torna cada vez maior.

O método anterior nos possibilita colocar um intervalo de confiança em torno da estimativa MQO de $E(y|x_1, \dots, x_k)$ para quaisquer valores das variáveis explicativas. Em outras palavras, obtemos um

intervalo de confiança do valor médio de y da subpopulação com determinado conjunto de covariadas. Entretanto, um intervalo de confiança da média pessoal na subpopulação não é a mesma coisa que um intervalo de confiança de uma unidade particular (indivíduo, família, empresa etc.) da população. Na formação de um intervalo de confiança de um resultado desconhecido de y , devemos avaliar outra fonte muito importante de variação: a variância no erro não observado, que registra nosso desconhecimento dos fatores não observados que afetam y .

Seja y^0 o valor para o qual gostaríamos de construir um intervalo de confiança, que algumas vezes chamamos de **intervalo de previsão**. Por exemplo, y^0 poderia representar uma pessoa ou uma empresa que não esteja em nossa amostra original. Façamos x_1^0, \dots, x_k^0 serem os novos valores das variáveis independentes, que assumimos observar, e u^0 ser o erro não observado. Portanto, temos

$$y^0 = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \dots + \beta_k x_k^0 + u^0. \quad (6.33)$$

Como antes, nossa melhor previsão de y^0 é o valor esperado de y^0 , dadas as variáveis explicativas que estimamos da linha de regressão MQO: $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \hat{\beta}_2 x_2^0 + \dots + \hat{\beta}_k x_k^0$. O **erro de previsão** com o uso de \hat{y}_0 para prever y^0 é

$$\hat{e}^0 = y^0 - \hat{y}_0 = (\beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0) + u^0 - \hat{y}_0. \quad (6.34)$$

Agora, $E(\hat{y}_0) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x_1^0 + E(\hat{\beta}_2)x_2^0 + \dots + E(\hat{\beta}_k)x_k^0 = \beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0$, porque os $\hat{\beta}_j$ são não-viesados. (Como antes, essas expectativas são todas condicionais aos valores amostrais das variáveis independentes.) Como u^0 tem média zero, $E(\hat{e}^0) = 0$. Mostramos que o erro de previsão esperado é zero.

Ao encontrar a variância de \hat{e}^0 , observe que u^0 é não-correlacionado com cada $\hat{\beta}_j$, porque u^0 é não-correlacionado com os erros na amostra usada para a obtenção de $\hat{\beta}_j$. Pelas propriedades básicas da covariância (veja Apêndice B, no site da Thomson), u^0 e \hat{y}_0 são não-correlacionados. Portanto, a **variância do erro de previsão** (condicional a todos os valores das variáveis independentes incluídas na amostra) é a soma das variâncias:

$$\text{Var}(\hat{e}^0) = \text{Var}(\hat{y}_0) + \text{Var}(u^0) = \text{Var}(\hat{y}_0) + \sigma^2, \quad (6.35)$$

onde $\sigma^2 = \text{Var}(u^0)$ é a variância do erro. Existem duas fontes de variância em \hat{e}^0 . A primeira é o erro de amostragem em \hat{y}_0 , que surge por termos estimado β_j . Como cada $\hat{\beta}_j$ tem uma variância proporcional a $1/n$, na qual n é o tamanho da amostra, $\text{Var}(\hat{y}_0)$ é proporcional a $1/n$. Isso significa que, para amostras grandes, $\text{Var}(\hat{y}_0)$ pode ser muito pequena. Em contraposição, σ^2 é a variância do erro na população; ela não muda com o tamanho da amostra. Em muitos exemplos, σ^2 será o termo dominante em (6.35).

Sob as hipóteses do modelo linear clássico, $\hat{\beta}_j$ e u^0 são normalmente distribuídos, e assim \hat{e}^0 também é normalmente distribuído (condicional a todos os valores amostrais das variáveis explicativas). Anteriormente, descrevemos como obter um estimador não-viesado de $\text{Var}(\hat{y}_0)$, e obtivemos nosso estimador não-viesado de σ^2 no Capítulo 3. Com o uso desses estimadores, podemos definir o erro-padrão de \hat{e}^0 como

$$ep(\hat{e}^0) = \{[ep(\hat{y}^0)]^2 + \hat{\sigma}^2\}^{1/2}. \quad (6.36)$$

Utilizando o mesmo raciocínio para as estatísticas t de $\hat{\beta}_j$, $\hat{e}^0 / ep(\hat{e}^0)$ tem uma distribuição t com $n - (k + 1)$ graus de liberdade. Portanto,

$$P[-t_{0,025} \leq \hat{e}^0 / ep(\hat{e}^0) \leq t_{0,025}] = 0,95,$$

onde $t_{0,025}$ é o 97,5º percentil na distribuição t_{n-k-1} . Para $n - k - 1$ grande, lembre-se de que $t_{0,025} \approx 1,96$. Inserindo $\hat{e}^0 = y^0 - \hat{y}^0$ e fazendo a reordenação, obtemos um intervalo de previsão de 95% para y^0 :

$$\hat{y}^0 \pm t_{0,025} \cdot ep(\hat{e}^0); \quad (6.37)$$

como sempre, exceto para gl pequeno, uma boa regra prática é $\hat{y}^0 \pm 2ep(\hat{e}^0)$. Isso é mais amplo que o próprio intervalo de confiança de \hat{y}^0 , devido a $\hat{\sigma}^2$ em (6.36); normalmente ela é muito mais ampla para refletir os fatores em u^0 que não tenhamos controlado.

EXEMPLO 6.6

(Intervalo de Confiança de Notas Médias Futuras)

Suponha que desejamos um IC de 95% de *nmgrad* futuro de um aluno do ensino médio com *sat* = 1.200, *emperc* = 30 e *tamclas* = 5. No Exemplo 6.5 obtivemos um intervalo de confiança de 95% da *média* da nota média entre todos os alunos com as características particulares *sat* = 1.200, *emperc* = 30 e *tamclas* = 5. Agora, queremos um intervalo de confiança de 95% de qualquer aluno que *especificamente* tenha essas características. O intervalo de previsão de 95% deve registrar a variação na característica individual, não-observada, que afeta o desempenho universitário. Temos tudo que é preciso para obter um IC de *nmgrad*. Sabemos que $ep(\hat{y}^0) = 0,020$ e $\hat{\sigma} = 0,560$ e, portanto, de (6.36), $ep(\hat{e}^0) = [(0,020)^2 + (0,560)^2]^{1/2} \approx 0,560$. Observe o quanto $ep(\hat{y}^0)$ é pequeno em relação a $\hat{\sigma}$: virtualmente, toda a variação em \hat{e}^0 vem da variação em u^0 . O IC de 95% é $2,70 \pm 1,96(0,560)$ ou está entre 1,60 e 3,80. Este é um intervalo de confiança enorme, e mostra que, com base nos fatores que incluímos na regressão, não podemos definir com clareza a futura nota de graduação de determinado indivíduo. (Em certo sentido, isso é bom, por significar que a classificação no curso médio e o desempenho no teste de aptidão acadêmica não predeterminam o desempenho de alguém na faculdade.) Evidentemente, as características não observadas variam amplamente de um indivíduo para o outro com as mesmas notas no teste de aptidão acadêmica e na classificação no curso médio observadas.

Análise de Resíduos

Algumas vezes, é útil examinar as observações individuais para verificar se o valor efetivo da variável dependente está acima ou abaixo do valor previsto; isto é, examinar os resíduos das observações individuais. Este processo é chamado **análise de resíduos**. Os economistas são conhecidos por examinarem os resíduos de uma regressão para auxiliá-los, por exemplo, na compra de um imóvel.

O exemplo seguinte sobre preços de imóveis ilustra a análise de resíduos. Os preços dos imóveis estão relacionados a várias características observadas do imóvel. Podemos relacionar todas as características que julgarmos importantes, como tamanho, número de quartos, número de banheiros, e assim por diante. Podemos usar uma amostra de imóveis para estimar o relacionamento entre o preço e as características, e terminamos obtendo um valor previsto e um valor real de cada imóvel. Então, podemos construir os resíduos, $\hat{u}_i = y_i - \hat{y}_i$. O imóvel com o maior resíduo negativo é, pelo menos com base nos fatores que estamos controlando, o mais barato em relação às suas características observadas. É claro que um preço de venda substancialmente inferior ao seu preço previsto poderia indicar alguma característica indesejável do imóvel que deixamos de avaliar, e que portanto está contido no erro não-observado. Além da obtenção da previsão e do resíduo, também faz sentido computar o intervalo de confiança de qual seria o preço de venda do imóvel no futuro, utilizando o método descrito na equação (6.37).

Utilizando os dados contidos no arquivo HPRICE1.RAW, computamos a regressão de *preço* sobre *tamterr*, *arquad* e *qtdorm*. Na amostra de 88 imóveis, o resíduo mais negativo é $-120,206$, do 81º imóvel. Portanto, o preço pedido por esse imóvel está US\$120.206,00 abaixo de seu preço previsto.

Existem muitos outros usos da análise de resíduos. Uma maneira de classificar as faculdades de direito é fazer a regressão da mediana dos salários iniciais sobre uma variedade de características dos alunos (como a mediana das notas de ingresso nos cursos para novos alunos, a mediana das notas médias de graduação para novos alunos etc.) e obter um valor previsto e um resíduo de cada faculdade de direito. A faculdade de direito com o maior resíduo terá o maior valor agregado previsto. (Naturalmente, ainda existirá muita incerteza sobre como o salário inicial de um indivíduo se compararia com a mediana geral de uma faculdade de direito.) Esses resíduos poderão ser usados juntamente com mensalidades cobradas pelas faculdades de direito para determinarmos o valor mais vantajoso; isso exigirá um desconto apropriado dos ganhos futuros.

A análise de resíduos também tem participação em decisões judiciais. Um artigo publicado no jornal *The New York Times* intitulado “Juiz Diz que Pobreza de Alunos, e Não a Segregação, Prejudica Aproveitamento Escolar” (28.06.95) descreve um importante processo legal. A questão era se o fraco desempenho nos exames padronizados do Distrito Escolar de Hartford, em relação ao desempenho nos distritos vizinhos, era devido à baixa qualidade de ensino nas escolas altamente segregadas. O juiz concluiu que “a disparidade nas notas de aproveitamento escolar não indica que Hartford esteja fazendo um trabalho inadequado ou insuficiente na educação de seus alunos ou que suas escolas sejam deficientes, pois as notas de aproveitamento previstas com base em relevantes fatores socioeconômicos estão próximas dos níveis esperados”. Esta conclusão foi, quase com certeza, baseada em uma análise de regressão das notas de aproveitamento médias ou de suas medianas sobre as características socioeconômicas de vários distritos escolares de Connecticut. A conclusão do juiz sugere que, considerando o nível de pobreza dos alunos das escolas de Hartford, as notas de aproveitamento efetivas dos alunos eram semelhantes às previstas em uma análise de regressão: o resíduo de Hartford não era suficientemente negativo para se concluir que as escolas em si mesmas eram responsáveis pelas baixas notas de aproveitamento escolar.

Como você poderia usar a análise de resíduos para determinar quais atores cinematográficos são remunerados em níveis acima da bilheteria dos filmes em que atuam?

Previsão de y quando a Variável Dependente é $\log(y)$

Como a transformação do log natural é usada com tanta frequência na variável dependente em economia empírica, dedicamos esta subseção ao problema de prognosticar y quando a variável dependente é $\log(y)$. Como um subproduto, obteremos um indicador de grau de ajuste do modelo log que possa ser comparado com o R -quadrado do modelo em nível.

Para obter uma previsão é útil definirmos $\log y = \log(y)$; isso realça o fato de que é o log de y que será previsto no modelo

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u. \quad (6.38)$$

Nesta equação os x_j poderão ser transformações de outras variáveis; por exemplo, poderíamos ter $x_1 = \log(\text{vendas})$, $x_2 = \log(\text{valmerc})$, $x_3 = \text{permceo}$ no exemplo do salário dos diretores executivos.

Dados os estimadores MQO, sabemos como prever $\log y$ para qualquer valor das variáveis independentes:

$$\hat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (6.39)$$

Agora, como o exponencial desfaz o log, nossa primeira suposição para prever y é simplesmente exponenciar o valor previsto de $\log(y)$: $\hat{y} = \exp(\hat{\log y})$. Isso não funciona; aliás, isso sistematicamente *subestimar*á o valor esperado de y . De fato, se o modelo (6.38) obedecer às hipóteses do modelo linear clássico RLM.1 até RLM.6, pode ser demonstrado que

$$E(y|\mathbf{x}) = \exp(\sigma^2/2) \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

onde \mathbf{x} representa as variáveis independentes e σ^2 é a variância de u . [Se $u \sim \text{Normal}(0, \sigma^2)$, o valor esperado de $\exp(u)$ será $\exp(\sigma^2/2)$]. Esta equação mostra que um ajuste simples é necessário para prevermos y :

$$\hat{y} = \exp(\hat{\sigma}^2/2) \exp(\hat{\log y}), \quad (6.40)$$

onde $\hat{\sigma}^2$ é simplesmente o estimador não-viesado de σ^2 . Como $\hat{\sigma}$, o erro-padrão da regressão, é sempre conhecido, a obtenção de valores previstos de y será fácil. Como $\hat{\sigma}^2 > 0$, $\exp(\hat{\sigma}^2/2) > 1$. Para um $\hat{\sigma}^2$ grande, esse fator de ajuste poderá ser substancialmente maior que a unidade.

A previsão em (6.40) não é não-viesada, mas consistente. Não existem previsões não-viesadas de y , e em muitos casos (6.40) funciona bem. Porém, ela depende da normalidade do termo erro, u . No Capítulo 5, mostramos que o MQO possui propriedades desejáveis, mesmo quando u não for normalmente distribuído. Portanto, é vantajoso ter uma previsão que não dependa da normalidade. Se simplesmente assumirmos que u é independente das variáveis explicativas, teremos

$$E(y|\mathbf{x}) = \alpha_0 \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k), \quad (6.41)$$

onde α_0 é o valor esperado de $\exp(u)$, que deve ser maior que a unidade.

Dada uma estimativa de $\hat{\alpha}_0$, podemos prever y como

$$\hat{y} = \hat{\alpha}_0 \exp(\hat{l}ogy), \quad (6.42)$$

que mais uma vez simplesmente requer que façamos a exponenciação do valor previsto do modelo \log e que multipliquemos o resultado por $\hat{\alpha}_0$.

O resultado é que um estimador consistente de $\hat{\alpha}_0$ é facilmente obtido.

PREVISÃO DE y QUANDO A VARIÁVEL DEPENDENTE É $\log(y)$:

- (i) Obtenha os valores estimados $\hat{l}ogy_i$ da regressão de $\log y$ sobre x_1, \dots, x_k .
- (ii) Para cada observação i , crie $\hat{m}_i = \exp(\hat{l}ogy_i)$.
- (iii) Agora, faça a regressão de y sobre a variável única \hat{m} sem um intercepto; isto é, faça uma regressão simples passando pela origem. O coeficiente de \hat{m} , o único coeficiente que existe, é a estimativa de α_0 .

Quando $\hat{\alpha}_0$ for obtido, ele poderá ser usado juntamente com previsões de $\log y$ para prever y . Os passos são os seguintes:

- (i) Para valores dados de x_1, x_2, \dots, x_k , obtenha $\hat{l}ogy$ de (6.39).
- (ii) Obtenha a previsão de \hat{y} de (6.42).

EXEMPLO 6.7

(Previsão de Salários de Diretores Executivos)

O modelo de interesse é

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 \log(\text{valmerc}) + \beta_3 \text{permceo} + u,$$

de forma que β_1 e β_2 são elasticidades e $100 \cdot \beta_3$ é uma semi-elasticidade. A equação estimada utilizando os dados contidos no arquivo CEOSAL2.RAW é

$$\begin{aligned} \hat{Isalário} = & 4,504 + 0,163 \hat{l}vendas + 0,109 \hat{l}valmerc + 0,0117 \text{permceo} \\ & (0,257) \quad (0,039) \quad (0,050) \quad (0,0053) \end{aligned} \quad (6.43)$$

$$n = 177, R^2 = 0,318,$$

onde, para maior clareza, $\hat{Isalário}$ representa o log de salário e, de forma semelhante, $\hat{l}vendas$ e $\hat{l}valmerc$ representam o log de vendas e valmerc . A seguir, obtemos $\hat{m}_i = \exp(\hat{Isalário}_i)$ de cada observação na amostra. A regressão de salário sobre \hat{m} (sem uma constante) produz $\hat{\alpha}_0 \approx 1,117$.

Podemos usar esse valor de $\hat{\alpha}_0$ juntamente com (6.43) para prever salário de qualquer valor de vendas , valmerc e permceo . Encontremos a previsão de $\text{vendas} = 5.000$ (que significa 5 bilhões de dólares, já que vendas está em milhões de dólares), $\text{valmerc} = 10.000$ (10 bilhões de dólares), e $\text{permceo} = 10$. Da equação (6.43), a previsão de $\hat{Isalário}$ é $4,504 + 0,163 \cdot \log(5.000) + 0,109 \cdot \log(10.000) + 0,0117(10) \approx 7,013$. O salário previsto é, portanto, $1,117 \cdot \exp(7,013) \approx 1.240,967$, ou 1.240.967 dólares. Se esquecermos de multiplicar por $\hat{\alpha}_0 = 1,117$, obteremos uma previsão de 1.110.983 dólares.

Podemos usar o método anterior de obter previsões para determinar o quanto o modelo com $\log(y)$ como variável dependente explica bem a variável y . Já temos indicadores para modelos quando y é a variável dependente: o R -quadrado e o R -quadrado ajustado. O objetivo é encontrar um bom indicador de grau de ajuste no modelo $\log(y)$ que possa ser comparado com um R -quadrado de um modelo no qual y é a variável dependente.

Existem várias maneiras de encontrarmos esse indicador, mas apresentamos um método que é fácil de ser implementado. Após computar a regressão de y sobre \hat{m} passando pela origem no passo (iii), obtemos os valores estimados dessa regressão, $\hat{y}_i = \hat{\alpha}_0 \hat{m}_i$. Em seguida, encontramos a correlação amostral entre \hat{y}_i e o verdadeiro y_i na amostra. O quadrado dessa correlação amostral pode ser comparado com o R -quadrado que obtemos com o uso de y como a variável dependente em um modelo de regressão linear. Lembre-se de que o R -quadrado na equação estimada

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

é exatamente a correlação quadrada entre y_i e \hat{y}_i (veja a Seção 3.2).

EXEMPLO 6.8

(Previsão de Salários de Diretores Executivos)

Após o passo (iii) no procedimento precedente, obtemos os valores estimados $\text{salário} = \hat{\alpha}_0 \hat{m}_i$. A correlação simples entre salário_i e salário na amostra é 0,493; o quadrado desse valor está por volta de 0,243. Esse é o nosso indicador de quanto da variação de salário é explicada pelo modelo log; não se trata do R -quadrado de (6.43), que é 0,318.

Suponha que estimemos um modelo com todas as variáveis em nível:

$$\text{salário} = \beta_0 + \beta_1 \text{vendas} + \beta_2 \text{valmerc} + \beta_3 \text{permceo} + u.$$

O R -quadrado obtido estimando esse modelo, usando as mesmas 177 observações, é 0,201. Assim, o modelo log explica mais da variação em salário , e assim é o nosso preferido com base no grau de ajuste. O modelo log também é o escolhido porque parece ser mais realista e seus parâmetros são mais fáceis de ser interpretados.

Neste capítulo, tratamos de alguns tópicos importantes sobre a análise de regressão múltipla.

A Seção 6.1 mostrou que uma mudança nas unidades de medida de uma variável independente altera o coeficiente do MQO da maneira esperada: se x_j for multiplicado por c , seu coeficiente será dividido por c . Se a variável dependente é multiplicada por c , todos os coeficientes de MQO são multiplicados por c . Nem a estatística t nem a F são alteradas pela mudança das unidades de medida de quaisquer variáveis.

Discutimos os coeficientes beta, que medem os efeitos das variáveis independentes sobre a variável dependente em unidades de desvios-padrão. Os coeficientes beta são obtidos de uma regressão

MQO padrão depois de as variáveis dependente e independentes terem sido transformadas em valores padronizados.

Como vimos em vários exemplos, a forma funcional logarítmica produz coeficientes com interpretações de efeitos percentuais. Comentamos sobre suas vantagens adicionais na Seção 6.2. Também vimos como computar o efeito percentual exato quando um coeficiente em um modelo log-nível é grande. Modelos com termos quadráticos consideram efeitos marginais decrescentes ou crescentes. Modelos com interações possibilitam que o efeito marginal de uma variável explicativa dependa do nível de outra variável explicativa.

Introduzimos o R -quadrado ajustado, \bar{R}^2 , como uma alternativa ao R -quadrado habitual para medir o grau de ajuste. Enquanto o R^2 nunca pode cair quando outra variável é adicionada na regressão, o \bar{R}^2 penaliza o número de regressores e pode cair quando uma variável independente é adicionada. Isso faz do \bar{R}^2 o preferido para a opção entre modelos não-aninhados com diferentes quantidades de variáveis explicativas. Nem o R^2 nem o \bar{R}^2 podem ser usados para comparar modelos com variáveis dependentes diferentes. No entanto, é bastante fácil obter indicadores de graus de ajuste para optarmos entre y e $\log(y)$ como a variável dependente, como mostrado na Seção 6.4.

Na Seção 6.3 discutimos o problema de certa forma sutil de dependermos demasiadamente do R^2 ou do \bar{R}^2 para chegarmos a um modelo final: é possível controlarmos grandes quantidades de fatores em um modelo de regressão. Por essa razão, é importante pensar à frente sobre a especificação de modelos, particularmente sobre a natureza *ceteris paribus* da equação de regressão múltipla. Variáveis explicativas que afetem y e sejam não-correlacionadas com todas as outras variáveis explicativas podem ser usadas para reduzir a variância do erro sem induzir multicolinearidade.

Na Seção 6.4 demonstramos como obter um intervalo de confiança de uma previsão feita de uma linha de regressão MQO. Também mostramos como um intervalo de confiança pode ser construído para um valor futuro e desconhecido de y .

Ocasionalmente, queremos prever quando $\log(y)$ é usado como a variável dependente em um modelo de regressão. A Seção 6.4 explica esse método simples. Finalmente, algumas vezes estamos interessados em conhecer o sinal e a magnitude dos resíduos de observações específicas. A análise de resíduos pode ser usada para determinarmos se elementos específicos da amostra possuem valores previstos que estejam bem acima, ou bem abaixo, dos verdadeiros resultados.

6.1 A seguinte equação foi estimada utilizando os dados contidos no arquivo CEOSAL1.RAW:

$$\log(\widehat{\text{salário}}) = 4,322 + 0,276 \log(\widehat{\text{vendas}}) + 0,0215 \text{ rma} - 0,00008 \text{ rma}^2$$

$$(0,324) \quad (0,33) \qquad \qquad (0,129) \qquad \qquad (0,00026)$$

$$n = 209, R^2 = 0,282.$$

Esta equação permite que rma tenha um efeito decrescente sobre $\log(\widehat{\text{salário}})$. Essa generalidade é necessária? Justifique.

6.2 Sejam $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ as estimativas MQO da regressão de y_i sobre x_{i1}, \dots, x_{ik} , $i = 1, 2, \dots, n$. Para constantes diferentes de zero c_1, \dots, c_k , argumente que o intercepto e as inclinações MQO da regressão de $c_0 y_i$ sobre $c_1 x_{i1}, \dots, c_k x_{ik}$, $i = 1, 2, \dots, n$ são dados por $\hat{\beta}_0 = c_0 \hat{\beta}_0$, $\hat{\beta}_1 = (c_0/c_1) \hat{\beta}_1, \dots, \hat{\beta}_k = (c_0/c_k)$

$\hat{\beta}_k$. (Sugestão: Use o fato de que $\hat{\beta}_j$ soluciona as condições de primeira ordem em (3.13), e que $\tilde{\beta}_j$ deve solucionar as condições de primeira ordem envolvendo as variáveis dependente e independentes redimensionadas.)

6.3 Utilizando os dados contidos no arquivo RDCHEM.RAW, a seguinte equação foi obtida por MQO:

$$\begin{aligned} pdi\hat{n}tens &= 2,613 + 0,00030 vendas - 0,0000000070 vendas^2 \\ &\quad (0,429) \quad (0,00014) \quad (0,0000000037) \\ n &= 32, R^2 = 0,1484. \end{aligned}$$

- (i) Em que ponto o efeito marginal de *vendas* sobre *pdintens* se torna negativo?
- (ii) Você manteria o termo quadrático no modelo? Explique.
- (iii) Defina *vendasbil* como vendas expressas em bilhões de dólares: $vendasbil = vendas/1.000$. Reescreva a equação com *vendasbil* e $vendasbil^2$ como as variáveis independentes. Certifique-se de descrever os erros-padrão e o *R*-quadrado. [Sugestão: Observe que $vendasbil^2 = vendas^2/(1.000)^2$.]
- (iv) Com o propósito de descrever os resultados, qual equação você prefere?

6.4 O seguinte modelo permite que o retorno da educação dependa da educação total dos pais, chamada *edupais*:

$$\log(\text{salário}) = \beta_0 + \beta_1 educ + \beta_2 educ \cdot edupais + \beta_3 exper + \beta_4 perm + u.$$

- (i) Mostre que, em forma decimal, o retorno de mais um ano de educação nesse modelo é

$$\Delta \log(\text{salário}) / \Delta educ = \beta_1 + \beta_2 edupais.$$

Que sinal você espera para β_2 ? Por quê?

- (ii) Utilizando os dados contidos no arquivo WAGE2.RAW, a equação estimada é

$$\begin{aligned} \log(\hat{\text{salário}}) &= 5,65 + 0,047 educ + 0,00078 educ \cdot edupais + \\ &\quad (0,13) \quad (0,010) \quad (0,00021) \\ &\quad 0,019 exper + 0,010 perm \\ &\quad (0,004) \quad (0,003) \\ n &= 722, R^2 = 0,169. \end{aligned}$$

(Somente 722 observações contêm todas as informações sobre a educação dos pais.) Interprete o coeficiente do termo de interação. Pode ser interessante escolher dois valores específicos para *edupais*, por exemplo, $edupais = 32$ se ambos tiverem educação superior, ou $edupais = 24$ se ambos tiverem educação de nível médio — e comparar o retorno estimado de *educ*.

(iii) Quando *edupais* é adicionada como uma variável separada na equação, obtemos:

$$\begin{aligned} \log(\widehat{\text{salário}}) = & 4,94 + 0,097 \text{ educ} + 0,033 \text{ edupais} + 0,0016 \text{ educ} \cdot \text{edupais} \\ & (0,38) \quad (0,027) \quad (0,017) \quad (0,0012) \\ & + 0,020 \text{ exper} + 0,010 \text{ perm} \\ & (0,004) \quad (0,003) \\ & n = 722, R^2 = 0,174. \end{aligned}$$

O retorno da educação agora depende positivamente da educação dos pais? Teste a hipótese nula de que o retorno da educação não depende da educação dos pais.

6.5 No Exemplo 4.2, no qual a percentagem de alunos aprovados em um exame de matemática do 10^o ano (*mate10*) é a variável dependente, faz sentido incluir *ciencia11* — a percentagem de alunos do 11^o ano aprovados em um exame de ciências — como uma variável explicativa adicional?

6.6 Quando *taxafreq*² e *tac* · *taxafreq* são adicionadas à equação (6.19), o *R*-quadrado passa a ser 0,232. Esses termos adicionais são conjuntamente significantes no nível de 10%? Você os incluiria no modelo?

6.7 As três seguintes equações foram estimadas utilizando-se as 1,534 observações contidas no arquivo 401K.RAW.

$$\begin{aligned} \widehat{taxap} = & 80,29 + 5,44 \text{ taxcomp} + 0,269 \text{ idade} - 0,00013 \text{ totemp} \\ & (0,78) \quad (0,52) \quad (0,045) \quad (0,00004) \\ & R^2 = 0,100, \bar{R}^2 = 0,098. \end{aligned}$$

$$\begin{aligned} \widehat{taxap} = & 97,32 + 5,02 \text{ taxcomp} + 0,314 \text{ idade} - 2,66 \log(\text{totemp}) \\ & (1,95) \quad (0,51) \quad (0,044) \quad (0,28) \\ & R^2 = 0,144, \bar{R}^2 = 0,142. \end{aligned}$$

$$\begin{aligned} \widehat{taxap} = & 80,62 + 5,34 \text{ taxcomp} + 0,290 \text{ idade} - 0,00043 \text{ totemp} \\ & (0,78) \quad (0,52) \quad (0,045) \quad (0,00009) \\ & + 0,0000000039 \text{ totemp}^2 \\ & (0,0000000010) \\ & R^2 = 0,108, \bar{R}^2 = 0,106. \end{aligned}$$

Qual desses três modelos você prefere? Por quê?