CAPÍTULO 4

Regressão Linear com um Regressor

Um estado implementa novas punições severas para motoristas bêbados; qual é o efeito disso sobre os acidentes fatais nas estradas? Uma diretoria regional de ensino reduz o tamanho de suas turmas do ensino fundamental; qual é o efeito disso sobre as pontuações dos alunos nos exames nacionais? Você conclui com sucesso mais um ano de estudos na universidade; qual é o efeito disso sobre seu salário futuro?

As três perguntas dizem respeito ao efeito desconhecido da variação de uma variável, X — seja X as punições severas por dirigir embriagado, o tamanho da turma ou os anos de instrução — sobre outra variável, Y — seja Y as mortes nas estradas, a pontuação dos alunos nos exames ou o salário.

Este capítulo apresenta o modelo de regressão linear que relaciona uma variável, X, a outra, Y. Esse modelo postula uma relação linear entre X e Y; a declividade da reta que relaciona X e Y é o efeito da variação de uma unidade em X sobre Y. Assim como a média de Y é uma característica desconhecida da distribuição da população de Y, a declividade da reta que relaciona X e Y é uma característica desconhecida da distribuição conjunta da população de X e Y. O problema econométrico é estimar essa declividade — isto é, estimar o efeito de uma variação de uma unidade em X sobre Y — utilizando uma amostra de dados para essas duas variáveis.

O capítulo descreve métodos para realizar inferências estatísticas sobre esse modelo de regressão utilizando uma amostra aleatória de dados para X e Y. Por exemplo, utilizando dados sobre tamanhos de turma e pontuação nos exames de diferentes diretorias regionais de ensino, mostramos como estimar o efeito esperado da redução dos tamanhos de turma em, digamos, um aluno por turma, sobre a pontuação nos exames. A declividade e o intercepto da reta que relaciona X e Y podem ser estimados por um método chamado de mínimos quadrados ordinários (MQO). Além disso, o estimador de MQO pode ser utilizado para testar hipóteses sobre o valor da população da declividade — por exemplo, testar a hipótese de que a redução no tamanho da turma não tem nenhum efeito sobre a pontuação nos exames — e para criar intervalos de confiança para a declividade.

4.1 Modelo de Regressão Linear

A superintendente de uma diretoria de ensino fundamental nos Estados Unidos precisa decidir se deve contratar mais professores e quer sua ajuda. Se contratar os professores, reduzirá o número de alunos por professor (a razão aluno-professor) em dois. Ela está diante de um dilema. Os pais querem turmas menores para que seus filhos possam receber uma atenção mais individualizada. Porém, contratar mais professores significa gastar mais dinheiro, o que não agrada àqueles que pagam a conta! Então ela pergunta: se reduzir o tamanho das turmas, qual será o efeito dessa redução sobre o desempenho dos alunos?

Em muitas diretorias de ensino, o desempenho dos alunos é medido por exames nacionais e a situação profissional ou o salário de alguns diretores podem depender em parte do sucesso dos alunos nesses exames. Portanto, vamos tornar a pergunta da superintendente mais específica: se ela reduzir o tamanho médio das turmas em dois alunos, qual será o efeito dessa redução sobre a pontuação nos exames nacionais em sua diretoria?

Uma resposta precisa para essa pergunta requer uma afirmação quantitativa sobre variações. Se a superintendente *mudasse* o tamanho das turmas em determinado montante, que *variação* ela esperaria na pontuação nos exames nacionais? Podemos escrever isso como uma relação matemática utilizando a letra grega beta, $\beta_{TamTurma}$, em que o subscrito "TamTurma" separa o efeito da variação no tamanho da turma dos demais efeitos.

Portanto,

$$\beta_{TamTurma} = \frac{\text{variação na pontuação na prova}}{\text{variação no tamanho da turma}} = \frac{\Delta \text{PontExame}}{\Delta \text{TamTurma}}, \tag{4.1}$$

onde a letra grega Δ (delta) significa "variação em". Isto é, $\beta_{TamTurma}$ é a variação na pontuação da prova resultante da variação no tamanho da turma dividida pela variação no tamanho da turma.

Se você tivesse sorte suficiente para saber o valor de $\beta_{TamTurma}$, seria capaz de dizer à superintendente que a diminuição do tamanho da turma em um aluno mudaria a pontuação nos exames por toda a diretoria regional em $\beta_{TamTurma}$. Você também poderia responder à verdadeira pergunta da superintendente, que diz respeito à variação de dois alunos por turma no tamanho da turma. Para fazer isso, reorganize a Equação (4.1) de modo que

$$\Delta PontExame = \beta_{TamTurma} \times \Delta TamTurma. \tag{4.2}$$

Suponha que $\beta_{TamTurma} = -0{,}6$. Assim, uma redução de dois alunos por turma no tamanho da turma resultaria em uma variação prevista na pontuação nos exames de $(-0{,}6) \times (-2) = 1{,}2$; isto é, sua previsão é de que a pontuação nos exames *aumentaria* em 1,2 ponto como resultado da *redução* de dois alunos por turma nos tamanhos de turma.

A Equação (4.1) é a definição da declividade de uma linha reta que relaciona pontuação nos exames e tamanho da turma. Essa linha reta pode ser expressa como

$$PontExame = \beta_0 + \beta_{TamTurma} \times TamTurma, \tag{4.3}$$

onde β_0 é o intercepto dessa linha reta e, como antes, $\beta_{TamTurma}$ é a declividade. Segundo a Equação (4.3), se você conhecesse β_0 e $\beta_{TamTurma}$, não somente seria capaz de determinar, em uma diretoria, a *variação* na pontuação nos exames associada a uma *variação* no tamanho da turma, mas também seria capaz de prever a pontuação média nos exames para um dado tamanho de turma.

Quando você propõe a Equação (4.3) para a superintendente, ela lhe diz que há algo errado com essa fórmula. Ela ressalta que o tamanho da turma é apenas um dos vários aspectos do ensino fundamental e que duas diretorias com o mesmo tamanho de turma registram pontuações diferentes nos exames por muitas razões. Uma diretoria pode ter professores melhores ou adotar livros didáticos melhores. Duas diretorias com tamanhos de turma, professores e livros compatíveis ainda podem ter populações de alunos muito diferentes; pode ser que uma diretoria tenha mais alunos imigrantes (e portanto menos alunos que tenham o inglês como língua materna) ou alunos provenientes de famílias mais ricas. Finalmente, ela ressalta que, mesmo que duas diretorias sejam iguais sob todos esses aspectos, elas podem registrar pontuações diferentes nos exames por razões essencialmente aleatórias que dizem respeito ao desempenho de cada aluno no dia do exame. Ela está certa, obviamente; por todas essas razões, a Equação (4.3) não será válida com exatidão para todas as diretorias. Em vez disso, deve ser vista como uma afirmação sobre uma relação válida *em média* para a população de diretorias.

Uma versão dessa relação linear que vale para *cada* diretoria deve incluir esses outros fatores que influenciam a pontuação nos exames, incluindo as características exclusivas de cada diretoria (qualidade de seus professores, situação econômica de seus alunos, o fator sorte dos alunos no dia do exame etc.). Um enfoque seria enumerar os fatores mais importantes e introduzi-los explicitamente na Equação (4.3) (retornaremos a essa idéia no Capítulo 5). Por ora, entretanto, simplesmente agrupamos esses "outros fatores" e escrevemos a relação para dada diretoria como

$$PontExame = \beta_0 + \beta_{TamTurma} \times TamTurma + \text{outros fatores}. \tag{4.4}$$

Assim, a pontuação nos exames para a diretoria é escrita em termos de um componente $\beta_0 + \beta_{TamTurma} \times TamTurma$, que representa o efeito médio do tamanho da turma sobre as pontuações na população de diretorias regionais de ensino, e um segundo componente que representa todos os outros fatores.

Embora esta discussão tenha se concentrado na pontuação nos exames e no tamanho da turma, a idéia expressa na Equação (4.4) é muito mais geral, portanto é útil introduzir uma notação mais geral. Suponha que você tenha uma amostra de n diretorias. Seja Y_i a pontuação média nos exames da i-ésima diretoria, seja X_i o tamanho médio da turma na i-ésima diretoria e seja u_i os outros fatores que influenciam a pontuação nos exames da i-ésima diretoria. Desse modo, a Equação (4.4) pode ser escrita de forma mais geral como

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \tag{4.5}$$

para cada diretoria, isto é, $i = 1, ..., n$, onde β_0 é o intercepto dessa reta e β_1 é a declividade. (A notação geral "β_1" é utilizada para declividade na Equação (4.5) no lugar de "$\beta_{TamTurma}$" porque essa equação está escrita em termos de uma variável geral X_i.)

A Equação (4.5) é o **modelo de regressão linear com um único regressor**, em que Y é a **variável dependente** e X é a **variável independente**, ou **regressor**.

A primeira parte da Equação (4.5), $\beta_0 + \beta_1 X_i$, é a **reta de regressão da população** ou **função de regressão da população**. Essa é a relação válida entre Y e X em média ao longo da população. Portanto, se você conhece o valor de X, de acordo com essa reta de regressão da população, prevê que o valor da variável dependente, Y, é $\beta_0 + \beta_1 X$.

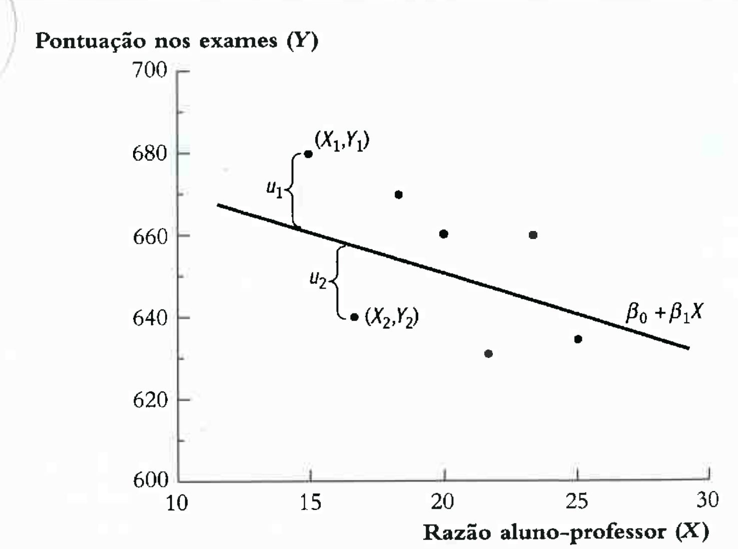
O **intercepto** β_0 e a **declividade** β_1 são os **coeficientes** da reta de regressão da população, também conhecidos como **parâmetros** da reta de regressão da população. A declividade β_1 é a variação em Y associada a uma variação unitária em X. O intercepto é o valor da reta de regressão da população quando $X = 0$; é o ponto em que a reta de regressão da população cruza o eixo Y. Em algumas aplicações econométricas, como a aplicação da Seção 4.7, o intercepto tem uma interpretação econômica significativa. Em outras aplicações, entretanto, o intercepto não tem nenhum significado em termos do mundo real; por exemplo, quando X é o tamanho da turma, a rigor o intercepto é o valor previsto da pontuação nos exames quando não há nenhum aluno na turma! Quando o intercepto não tem nenhum significado em termos do mundo real, é melhor considerá-lo matematicamente como o coeficiente que determina o nível da reta de regressão.

O termo u_i na Equação (4.5) é o **termo de erro**. Este termo incorpora todos os fatores responsáveis pela diferença entre a pontuação média nos exames da i-ésima diretoria e o valor previsto pela reta de regressão da população. Esse termo de erro contém todos os outros fatores além de X que determinam o valor da variável dependente, Y, para uma observação específica, i. No exemplo do tamanho da turma, esses outros fatores incluem todas as características exclusivas da i-ésima diretoria que afetam o desempenho de seus alunos no exame, incluindo a qualidade dos professores, a situação econômica dos alunos, a sorte e até mesmo os erros na correção da prova.

O modelo de regressão linear e sua terminologia estão resumidos no Conceito-Chave 4.1.

FIGURA 4.1 Gráfico de Dispersão da Pontuação nos Exames *versus* Razão Aluno-Professor (Dados Hipotéticos)

O gráfico de dispersão mostra observações hipotéticas de sete diretorias regionais de ensino. A reta de regressão da população é $\beta_0 + \beta_1 X$. A distância vertical entre o i-ésimo ponto e a reta de regressão da população é $Y_i - (\beta_0 + \beta_1 X_i)$, que é o termo de erro da população u_i para a i-ésima observação.



Terminologia para o Modelo de Regressão Linear com um Único Regressor

Conceito-Chave 4.1

O modelo de regressão linear é:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

onde:

o subscrito i representa o número de observações, $i = 1, ..., n$;

Y_i é a *variável dependente*, o *regressando*, ou simplesmente a *variável do lado esquerdo*;

X_i é a *variável independente*, o *regressor*, ou simplesmente a *variável do lado direito*;

$\beta_0 + \beta_1 X$ é a *reta de regressão da população* ou *função de regressão da população*;

β_0 é o *intercepto* da reta de regressão da população;

β_1 é a *declividade* da reta de regressão da população e

u_i é o *termo de erro*.

A Figura 4.1 resume o modelo de regressão linear com um único regressor para sete observações hipotéticas de pontuação nos exames (Y) e tamanho de turma (X). A reta de regressão da população é $\beta_0 + \beta_1 X$. A reta de regressão da população tem declividade para baixo, isto é, $\beta_1 < 0$, o que significa que diretorias com uma razão aluno-professor mais baixa (turmas menores) tendem a obter pontuações mais altas nos exames. O intercepto β_0 possui um significado matemático como o valor em que o eixo Y cruza a reta de regressão da população, mas, como mencionado anteriormente, não tem neste exemplo nenhum significado em termos do mundo real.

Como os outros fatores que determinam o desempenho nos exames, as observações hipotéticas da Figura 4.1 não estão exatamente sobre a reta de regressão da população. Por exemplo, o valor de Y para a diretoria nº 1, Y_1, está acima da reta de regressão da população. Isso significa que a pontuação nos exames da diretoria nº 1 foi melhor do que o previsto pela reta de regressão da população, de modo que o termo de erro para essa diretoria, u_1, é positivo. A Y_2, por sua vez, está abaixo da reta de regressão da população, de modo que a pontuação nos exames dessa diretoria foi pior do que o previsto, com $u_2 < 0$.

Agora volte a seu problema como consultor da superintendente: Qual é o efeito esperado da redução da razão aluno-professor em dois alunos por professor sobre a pontuação nos exames? A resposta é fácil: a variação esperada é $(-2) \times \beta_{TamTurma}$. Mas qual é o valor de $\beta_{TamTurma}$?

4.2 Estimando os Coeficientes do Modelo de Regressão Linear

Em uma situação prática, tal como a aplicação para o tamanho da turma e a pontuação nos exames, o intercepto β_0 e a declividade β_1 da reta de regressão da população são desconhecidos. Portanto, devemos utilizar dados para estimá-los.

Esse problema de estimação é semelhante a outros que você enfrentou em estatística. Por exemplo, suponha que você queira comparar os salários médios de homens e mulheres recém-formados. Embora os salários médios da população sejam desconhecidos, podemos estimar as médias da população utilizando uma amostra aleatória de homens e mulheres com curso superior. Assim, o estimador natural dos salários médios da população desconhecidos para as mulheres é o salário médio das mulheres com curso superior da amostra.

A mesma idéia se estende para o modelo de regressão linear. Não sabemos o valor da população de $\beta_{TamTurma}$, a declividade da reta de regressão da população desconhecida que relaciona X (tamanho da turma) e Y (pontuação nos exames). Mas, assim como foi possível conhecer a média da população utilizando uma amostra de dados selecionados dessa população, também é possível conhecer a declividade da população $\beta_{TamTurma}$ utilizando uma amostra de dados.

Os dados que analisamos aqui consistem da pontuação nos exames e do tamanho das turmas em 1998 de 420 diretorias regionais de ensino da Califórnia, que cuidam da pré-escola à oitava série. Em cada diretoria, a pontuação nos exames consiste na média das pontuações em leitura e matemática dos alunos da quinta série. O tamanho da turma pode ser medido de várias formas. A medida utilizada aqui, uma das mais abrangentes, é o número de alunos na diretoria dividido pelo número de professores, isto é, a razão aluno-professor na diretoria. Esses dados estão descritos de maneira mais detalhada no Apêndice 4.1.

A Tabela 4.1 resume as distribuições da pontuação nos exames e do tamanho das turmas nessa amostra. A razão aluno-professor média é de 19,6 alunos por professor e o desvio padrão é de 1,9 aluno por professor. O 10º percentil da distribuição da razão aluno-professor é 17,3 (isto é, apenas 10 por cento das diretorias têm razões aluno-professor inferiores a 17,3), ao passo que a diretoria do 90º percentil tem uma razão aluno-professor de 21,9.

A Figura 4.2 mostra um gráfico de dispersão dessas 420 observações sobre pontuação nos exames e razão aluno-professor. A correlação da amostra é de –0,23, o que indica uma relação negativa fraca entre as duas variáveis. Embora turmas maiores dessa amostra tendam a apresentar pontuações menores nos exames, existem outros determinantes da pontuação que impedem que as observações estejam perfeitamente sobre uma reta.

A despeito da correlação pequena, se fosse possível desenhar uma linha reta ao longo desses dados, então a declividade dessa reta seria uma estimativa de $\beta_{TamTurma}$ com base nesses dados. Uma forma de desenhar a reta seria pegar um lápis e uma régua e encontrar a melhor reta possível usando o "golpe de vista". Embora esse método seja fácil, ele não é científico e pessoas diferentes podem criar retas estimadas diferentes.

Como, então, você pode escolher entre as várias retas possíveis? A forma mais comum é, de longe, escolher a reta que produz o ajuste de "mínimos quadrados" a esses dados, isto é, utilizar o estimador de mínimos quadrados ordinários (MQO).*

O Estimador de Mínimos Quadrados Ordinários

O estimador de MQO escolhe os coeficientes de regressão de modo que a reta de regressão estimada seja a mais próxima possível dos dados observados, em que a proximidade é medida pela soma dos quadrados dos erros de previsão de Y dado X.

Conforme discutido na Seção 3.1, a média da amostra, $\overline{Y}$, é o estimador de mínimos quadrados da média da população $E(Y)$; isto é, $\overline{Y}$ minimiza o total dos quadrados dos erros de estimação $\sum_{i=1}^{n}(Y_i - m)^2$ entre todos os estimadores m possíveis (veja a Expressão (3.2)).

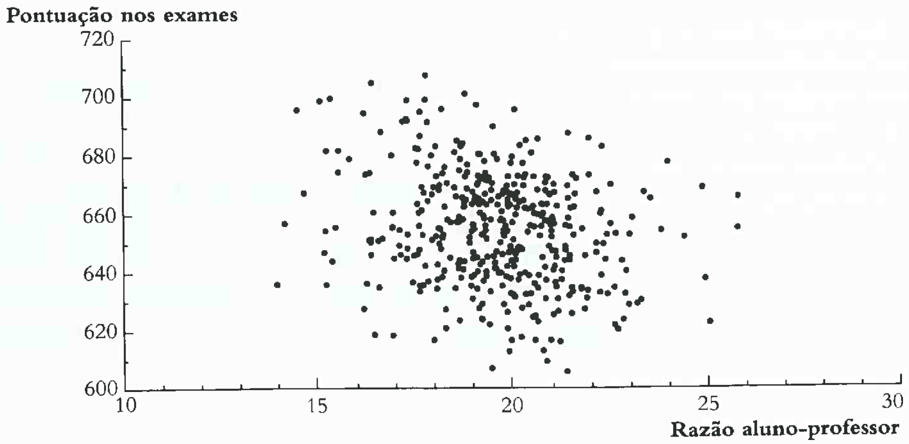
TABELA 4.1 Resumo da Distribuição da Razão Aluno-Professor e Pontuação nos Exames da 5ª Série para 420 Diretorias K-8 na Califórnia em 1998

	Média	Desvio padrão	Percentil 10%	25%	40%	50% (mediana)	60%	75%	90%
Razão aluno-professor	19,6	1,9	17,3	18,6	19,3	19,7	20,1	20,9	21,9
Pontuação nos exames	654,2	19,1	630,4	640,0	649,1	654,5	659,4	666,7	679,1

* O termo "mínimos quadrados ordinários" é muito utilizado em economometria. A tradução mais correta seria "mínimos quadrados comuns", que não traz a conotação negativa que o termo "ordinário" possui em português (N. do R.T.).

FIGURA 4.2 Gráfico de Dispersão da Pontuação nos Exames *versus* Razão Aluno-Professor (Dados da Diretoria Regional de Ensino da Califórnia)

Dados de 420 diretorias regionais de ensino da Califórnia. Existe uma relação negativa fraca entre a razão aluno-professor e a pontuação nos exames: a correlação da amostra é –0,23.



O estimador de MQO estende essa idéia para o modelo de regressão linear. Sejam b_0 e b_1 alguns estimadores de β_0 e β_1. A reta de regressão baseada nesses estimadores é $b_0 + b_1X$, de modo que o valor de Y_i previsto utilizando essa reta é $b_0 + b_1X_i$. Portanto, o erro feito na previsão da i-ésima observação é $Y_i - (b_0 + b_1X_i) = Y_i - b_0 - b_1X_i$). A soma dos quadrados dos erros de previsão para todas as n observações é

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2. \tag{4.6}$$

A soma dos quadrados dos erros para o modelo de regressão linear na Expressão (4.6) é a extensão da soma dos quadrados dos erros para o problema de se estimar a média na Expressão (3.2). Na verdade, se não há nenhum regressor, b_1 não entra na Expressão (4.6) e os dois problemas são idênticos, exceto pela notação diferente (m na Expressão (3.2), b_0 na Expressão (4.6)). Assim como há um único estimador, $\overline{Y}$, que minimiza a Expressão (3.2), há um único par de estimadores de β_0 e β_1 que minimizam a Expressão (4.6).

Os estimadores do intercepto e da declividade que minimizam a soma dos quadrados dos erros na expressão (4.6) são chamados de **estimadores de mínimos quadrados ordinários (MQO)** de β_0 e β_1.

MQO possui sua própria notação e terminologia especial. O estimador de MQO de β_0 é representado por $\hat{\beta}_0$ e o estimador de MQO de β_1 é representado por $\hat{\beta}_1$. A **reta de regressão de MQO** é a linha reta construída utilizando os estimadores de MQO, isto é, $\hat{\beta}_0 + \hat{\beta}_1X$. O **valor previsto** de Y_i dado X_i, com base na reta de regressão de MQO, é $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_i$. O **resíduo** para a i-ésima observação é a diferença entre Y_i e seu valor previsto, isto é, o resíduo é $\hat{u}_i = Y_i - \hat{Y}_i$.

Você poderia calcular os estimadores de MQO, $\hat{\beta}_0$ e $\hat{\beta}_1$, experimentando diferentes valores de b_0 e b_1 repetidamente até encontrar aqueles que minimizam o total dos quadrados dos erros na Expressão (4.6); eles são as estimativas de mínimos quadrados. Esse método, contudo, seria bastante cansativo. Felizmente, há fórmulas, derivadas pela minimização da Expressão (4.6) utilizando cálculo, que agilizam o cálculo dos estimadores de MQO.

As fórmulas de MQO e a terminologia estão reunidas no Conceito-Chave 4.2. Essas fórmulas são implementadas em praticamente todos os programas estatísticos e de planilha. Essas fórmulas são derivadas no Apêndice 4.2.

Estimativas de MQO da Relação entre Pontuação nos Exames e Razão Aluno-Professor

Quando MQO é utilizado para estimar uma reta que relaciona a razão aluno-professor à pontuação nos exames utilizando as 420 observações da Figura 4.2, a declividade estimada é de –2,28 e o intercepto estimado é de 698,9. Assim, a reta de regressão de MQO para essas 420 observações é

$$\widehat{PontExame} = 698{,}9 - 2{,}28 \times RAP, \tag{4.7}$$

Estimador de MQO, Valores Previstos e Resíduos

Conceito-Chave 4.2

Os estimadores de MQO da declividade β_1 e do intercepto β_0 são:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.8}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.9}$$

Os valores previstos de MQO $\hat{Y}_i$ e os resíduos $\hat{u}_i$ são:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_i,\ i = 1, \ldots, n \tag{4.10}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{4.11}$$

O intercepto estimado ($\hat{\beta}_0$), a declividade ($\hat{\beta}_1$) e o resíduo ($\hat{u}_i$) são calculados a partir de uma amostra de n observações de X_i e Y_i, $i = 1, \ldots, n$. Essas são estimativas dos verdadeiros e desconhecidos intercepto da população β_0, da declividade β_1 e do termo de erro u_i verdadeiros desconhecidos.

em que *PontExame* é a pontuação média nos exames na diretoria e *RAP* é a razão aluno-professor. O símbolo ^ sobre *PontExame* na Equação (4.7) indica que esse é o valor previsto com base na reta de regressão de MQO. A Figura 4.3 mostra a reta de regressão de MQO superposta ao gráfico de dispersão dos dados mostrados anteriormente na Figura 4.2.

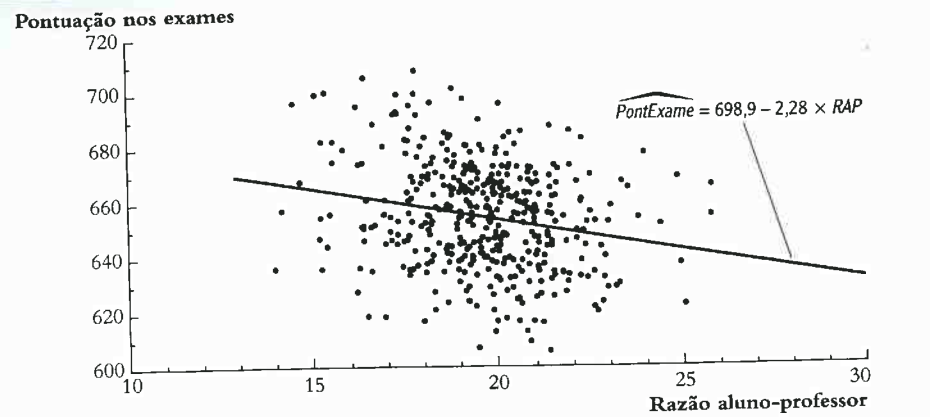
A declividade de –2,28 significa que um aumento de um aluno por turma na razão aluno-professor está, em média, associado a uma queda de 2,28 pontos na pontuação nos exames da diretoria como um todo. Uma redução de dois alunos por turma na razão aluno-professor está, em média, associada a um aumento de 4,56 pontos (= –2 × (–2,28)). A declividade negativa indica que uma quantidade maior de alunos por professor (turmas maiores) está associada a um desempenho pior nos exames.

Agora é possível prever a pontuação nos exames da diretoria como um todo para dado valor da razão aluno-professor. Por exemplo, para uma diretoria com 20 alunos por professor, a pontuação prevista nos exames é de 698,9 – 2,28 × 20 = 653,3. É óbvio que essa previsão não estará correta com exatidão em virtude de outros fatores que determinam o desempenho de uma diretoria. Mas a reta de regressão fornece uma previsão (a previsão de MQO) de quais seriam as pontuações nos exames para essa diretoria com base em sua razão aluno-professor, excluídos os demais fatores. Essa estimativa da declividade é grande ou pequena? Para responder a essa questão, voltemos ao problema da superintendente. Lembre-se de que ela está considerando a contratação de um número suficiente de professores que reduza a razão aluno-professor em 2. Suponha que sua diretoria esteja na mediana das diretorias da Califórnia. Na Tabela 4.1, a razão mediana aluno-professor é de 19,7 e a pontuação mediana nos exames é de 654,5. Uma redução de dois alunos por turma, de 19,7 para 17,7, moveria a razão aluno-professor de sua diretoria do 50º percentil para um ponto muito próximo do 10º percentil. É uma variação considerável, e ela precisaria contratar muitos professores novos. Como isso afetaria a pontuação nos exames?

Com base na Equação (4.7), a previsão é de que uma redução de 2 alunos por turma aumente a pontuação nos exames em aproximadamente 4,6 pontos; se a pontuação nos exames da diretoria da superintendente está na mediana, 654,5, ela deve aumentar para 659,1. Essa melhora é grande ou pequena? Segundo a Tabela 4.1, a melhora deslocaria sua diretoria da mediana para um ponto próximo do 60º percentil. Assim, uma redução no tamanho da turma que colocaria sua diretoria próxima dos 10 por cento com as menores turmas deslocaria sua pontuação nos exames do 50º para o 60º percentil. De acordo com essas estimativas, ao menos, diminuir a razão aluno-professor em um montante grande (2 alunos por professor) ajudaria e poderia valer a pena, dependendo de sua situação orçamentária, mas não seria uma panacéia.

FIGURA 4.3 Reta de Regressão Estimada para os Dados da Califórnia

A reta de regressão estimada mostra uma relação negativa entre pontuação nos exames e a razão aluno-professor. Se o tamanho das turmas for reduzido em um aluno, a regressão estimada prevê que a pontuação nos exames aumentará em 2,28 pontos.



E se a superintendente estivesse considerando uma variação bem mais radical, tal como reduzir a razão aluno-professor de 20 alunos por professor para 5? Infelizmente, as estimativas da Equação (4.7) não seriam muito úteis para ela. Essa regressão foi estimada utilizando os dados da Figura 4.2 e, como a figura mostra, a menor razão aluno-professor nesses dados é 14. Esses dados não contêm nenhuma informação sobre o comportamento das diretorias com turmas extremamente pequenas, de modo que somente esses dados não são uma base confiável para prever o efeito de uma variação radical para uma razão aluno-professor extremamente pequena.

O "Beta" de uma Ação

Uma idéia fundamental das finanças modernas é a de que um investidor precisa de incentivo financeiro para assumir um risco. Dito de outra forma, o retorno esperado[1] de um investimento com risco, R, deve exceder o retorno de um investimento seguro, ou sem risco, R_f. Portanto, o excesso de retorno esperado, $R - R_f$, de um investimento com risco, como a posse de ações de uma empresa, deve ser positivo.

A princípio pode parecer que o risco de uma ação deve ser medido por sua variância. Entretanto, muito desse risco pode ser reduzido mantendo-se outras ações em uma carteira, isto é, diversificando seus ativos financeiros. Isso significa que a forma certa de medir o risco de uma ação não é por sua *variância*, mas sim por sua *co-variância* com o mercado.

O Modelo de Precificação de Ativos de Capital (Capital Asset Pricing Model — CAPM) formaliza essa idéia. Segundo o CAPM, o excesso de retorno esperado de um ativo é proporcional ao excesso de retorno esperado de uma carteira com todos os ativos disponíveis (a "carteira do mercado"). Isto é, o CAPM diz que

$$R - R_f = \beta(R_m - R_f), \qquad (4.12)$$

onde R_m é o retorno esperado da carteira do mercado e β é o coeficiente da regressão da população de $R - R_f$ sobre $R_m - R_f$. Na prática, o retorno sem risco é freqüentemente considerado a taxa de juros da dívida de curto prazo do governo dos Estados Unidos. De acordo com o CAPM, uma ação com $\beta < 1$ oferece menos risco do que a carteira do mercado e, portanto, possui um excesso de retorno esperado mais baixo do que a carteira do mercado. Uma ação com $\beta > 1$, por sua vez, oferece mais risco do que a carteira do mercado e dessa forma exige um excesso de retorno esperado mais alto.

O "beta" de uma ação tornou-se o carro-chefe da indústria de investimentos e você pode obter a estimativa do β de centenas de ações em sites de empresas de investimentos. Esses βs são normalmente estimados por regressão de MQO do excesso de retorno efetivo da ação contra o excesso de retorno efetivo de um índice de mercado abrangente.

A tabela abaixo fornece os βs estimados para seis ações norte-americanas. Empresas que fabricam bens de consumo de baixo risco como a Kellogg têm βs baixos; ações de alta tecnologia com risco como a Microsoft têm βs altos.

Companhia	β Estimado
Kellogg (cereal matinal)	0,24
Waste Management (coleta de lixo)	0,38
Sprint (operadora telefônica de longa distância)	0,59
Walmart (varejista de descontos)	0,89
Barnes and Noble (rede de livrarias)	1,03
Best Buy (varejista de equipamentos eletrônicos)	1,80
Microsoft (software)	1,83

Fonte: Yahoo.com

1 O retorno sobre um investimento é a variação em seu preço acrescida de qualquer dividendo do investimento como uma porcentagem de seu preço inicial. Por exemplo, uma ação comprada em 1º de janeiro por US$ 100, que pagou um dividendo de US$ 2,50 durante o ano e foi vendida em 31 de dezembro por US$ 105, tem um retorno de R = [(US$ 105 – 100) + US$ 2,50] / US$ 100 = 7,5 por cento.

Por que Utilizar o Estimador de MQO?

Existem tanto motivos práticos quanto teóricos para utilizar os estimadores de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$. Como MQO é o método dominante utilizado na prática, ele tornou-se a linguagem comum para a análise de regressão em economia, finanças (veja o quadro na p. 70) e ciências sociais em geral. Apresentar resultados utilizando MQO (ou suas variantes, que serão discutidas mais adiante neste livro) indica que você está "falando a mesma língua" de outros economistas e estatísticos. As fórmulas de MQO estão incorporadas em praticamente todos os programas estatísticos e planilhas eletrônicas para computador, o que facilita o uso do MQO.

Os estimadores de MQO também possuem propriedades teóricas desejáveis. Por exemplo, a média de amostra $\overline{Y}$ é um estimador não viesado da média $E(Y)$, isto é, $E(\overline{Y}) = \mu_Y$; $\overline{Y}$ é um estimador consistente de μ_Y; e para amostras grandes a distribuição amostral de $\overline{Y}$ é aproximadamente normal (veja a Seção 3.1). Os estimadores de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ também possuem essas propriedades. Sob um conjunto geral de hipóteses (especificadas na Seção 4.3), $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores não viesados e consistentes de β_0 e β_1 e sua distribuição amostral é aproximadamente normal. Esses resultados são discutidos na Seção 4.4.

Outra propriedade teórica desejável de $\overline{Y}$ é que ele seja eficiente entre os estimadores que são funções lineares de $Y_1, ..., Y_n$: ele apresenta a menor variância de todos os estimadores que são médias ponderadas de $Y_1, ..., Y_n$ (veja a Seção 3.1). Um resultado similar também é verdadeiro para o estimador de MQO, mas esse resultado requer uma hipótese adicional além daquelas especificadas na Seção 4.3, de modo que adiamos sua discussão para a Seção 4.9.

4.3 Hipóteses de Mínimos Quadrados

Nesta seção, apresentamos um conjunto de três hipóteses para o modelo de regressão linear e o esquema de amostragem em que MQO fornece um estimador apropriado dos coeficientes de regressão desconhecidos β_0 e β_1. Inicialmente, essas hipóteses podem parecer abstratas. Entretanto, elas possuem interpretações naturais e entendê-las é essencial para saber quando o MQO fornecerá — ou não — estimativas úteis dos coeficientes de regressão.

Hipótese nº 1: a Distribuição Condicional de u_i Dado X_i Possui uma Média Igual a Zero

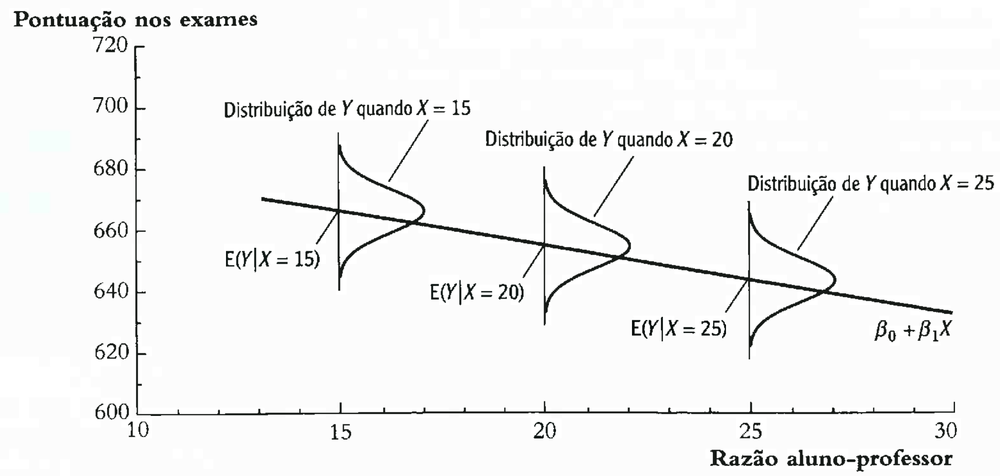
A primeira **hipótese dos mínimos quadrados** é que a distribuição condicional de u_i dado X_i possui média zero. Essa hipótese é uma declaração matemática formal sobre os "outros fatores" contidos em u_i e afirma que esses outros fatores não têm relação com X_i no sentido de que, dado um valor de X_i, a média da distribuição desses outros fatores é zero.

Isso está ilustrado na Figura 4.4. A regressão da população é a relação válida em média entre o tamanho da turma e a pontuação nos exames na população e o termo de erro u_i representa os outros fatores que levam a pontuação nos exames em dada diretoria a diferir da previsão baseada na reta de regressão da população. Conforme mostrado na Figura 4.4, em dado valor de tamanho de turma, digamos 20 alunos por turma, algumas vezes outros fatores levam a um desempenho melhor do que o previsto ($u_i > 0$) e outras vezes a um desempenho pior ($u_i < 0$), mas, na média, ao longo da população, a previsão está correta. Em outras palavras, dado $X_i = 20$, a média da distribuição de u_i é zero. Isso é mostrado na Figura 4.4 como a distribuição de u_i centrada em torno da reta de regressão da população em $X_i = 20$, e, de modo mais geral, em outros valores x de X_i também. Dito de outra forma, a distribuição de u_i, condicional a $X_i = x$, possui uma média zero; expressando matematicamente, $E(u_i | X_i = x) = 0$ ou, em uma notação mais simples, $E(u_i | X_i) = 0$.

Conforme mostrado na Figura 4.4, a hipótese de que $E(u_i | X_i) = 0$ é equivalente à suposição de que a reta de regressão da população é a média condicional de Y_i dado X_i (uma prova matemática disso está no Exercício 4.3).

Correlação e média condicional. Lembre-se da Seção 2.3, em que você viu que, se a média condicional de uma variável aleatória dada a outra é zero, as duas variáveis aleatórias possuem co-variância zero e desse modo são

FIGURA 4.4 Distribuições de Probabilidade Condicional e Reta de Regressão da População



A figura mostra a probabilidade condicional da pontuação nos exames para diretorias com tamanho de turma de 15, 20 e 25 alunos. A média da distribuição condicional da pontuação nos exames, dada a razão aluno-professor, $E(Y|X)$, é a reta de regressão da população $\beta_0 + \beta_1 X$. Para dado valor de X, Y é distribuída em torno da reta de regressão, e o erro, $u = Y - (\beta_0 + \beta_1 X)$, possui média condicional zero para todos os valores de X.

não-correlacionadas (veja a Equação (2.25)). Portanto, a hipótese de média condicional $E(u_i | X_i) = 0$ implica que X_i e u_i são não-correlacionados, ou $\text{corr}(X_i, u_i) = 0$. Como a correlação é uma medida de associação linear, essa implicação não é válida no sentido inverso. Mesmo que X_i e u_i sejam não-correlacionados, a média condicional de u_i dado X_i pode ser diferente de zero. Entretanto, se X_i e u_i são correlacionados, então $E(u_i | X_i)$ deve ser diferente de zero. Portanto, com freqüência é conveniente discutir a hipótese de média condicional em termos da possível correlação entre X_i e u_i. Se X_i e u_i são correlacionados, a hipótese de média condicional é violada.

Hipótese nº 2: (X_i, Y_i), $i = 1, \ldots, n$ São Independente e Identicamente Distribuídas

A segunda hipótese dos mínimos quadrados é de que (X_i, Y_i), $i = 1, \ldots, n$ são independente e identicamente distribuídas (i.i.d.) ao longo das observações. Conforme discutido na Seção 2.5 (veja o Conceito-Chave 2.5), essa é uma declaração sobre a seleção de uma amostra. Se as observações são selecionadas por amostragem aleatória simples de uma única população grande, então (X_i, Y_i), $i = 1, \ldots, n$ são i.i.d. Por exemplo, seja X a idade de um trabalhador e seja Y seu salário; imagine a seleção aleatória de uma pessoa da população de trabalhadores. A pessoa selecionada aleatoriamente terá uma idade e um salário dados (isto é, X e Y aceitarão alguns valores). Se uma amostra de n trabalhadores for selecionada dessa população, então (X_i, Y_i), $i = 1, \ldots, n$, terão necessariamente a mesma distribuição e, se forem selecionadas ao acaso, elas também serão independentemente distribuídas de uma observação para a próxima, isto é, serão i.i.d.

A hipótese i.i.d. é razoável para muitos esquemas de coleta de dados. Por exemplo, em geral, dados de pesquisa de um subconjunto da população escolhido aleatoriamente podem ser tratados como i.i.d.

Entretanto, nem todos os esquemas de amostragem geram observações i.i.d. sobre (X_i, Y_i). Um exemplo disso ocorre quando os valores de X não são selecionados de uma amostra aleatória da população, mas sim definidos pelo pesquisador como parte de um experimento. Por exemplo, suponha que uma horticultora queira estudar os efeitos de diferentes métodos de eliminação orgânica de ervas daninhas (X) sobre a produção de tomates (Y); para isso, ela planta diferentes lotes de tomates utilizando técnicas diferentes de eliminação orgânica de ervas daninhas. Se ela seleciona a técnica (o nível de X) a ser utilizada no i-ésimo lote e aplica a mesma técnica ao i-ésimo lote em todas as repetições do experimento, o valor de X_i não muda de uma amostra para a seguinte. Portanto, X_i é não aleatória (embora o resultado Y_i seja aleatório), de modo que o esquema de amostragem não

é i.i.d. Os resultados apresentados neste capítulo desenvolvidos para regressores i.i.d. também são válidos se os regressores são não aleatórios (isso será discutido com mais profundidade no Capítulo 15). O caso de um regressor não aleatório é, contudo, bastante especial. Por exemplo, protocolos experimentais modernos fariam com que a horticultora atribuísse o nível de X a diferentes lotes utilizando um gerador computadorizado de números aleatórios, eliminando desse modo qualquer viés possível pela horticultora (ela poderia aplicar seu método favorito de eliminação de ervas daninhas dos tomates no lote mais ensolarado). Quando esse protocolo experimental moderno é utilizado, o nível de X é aleatório e (X_i, Y_i) é i.i.d.

Outro exemplo de amostragem não i.i.d. são observações que se referem à mesma unidade de observação ao longo do tempo. Por exemplo, podemos ter dados sobre níveis de estoque (Y) de uma empresa e sobre a taxa de juros pela qual a empresa pode tomar um empréstimo (X), em que os dados são coletados ao longo do tempo de uma empresa específica; por exemplo, os dados podem ser registrados quatro vezes por ano (trimestralmente) durante 30 anos. Esse é um exemplo de séries temporais de dados e uma característica importante dos dados de séries temporais é que observações muito próximas no tempo não são independentes, mas tendem a ser correlacionadas; se as taxas de juros estiverem baixas agora, provavelmente continuarão baixas no próximo trimestre. Esse padrão de correlação viola a "independência" da hipótese i.i.d. Dados de séries temporais introduzem um conjunto de complicações que são melhores de lidar depois do desenvolvimento das ferramentas básicas da análise de regressão, de modo que adiamos as discussões adicionais sobre análise de séries temporais para a Parte 4.

Hipótese nº 3: X_i e u_i Têm Quatro Momentos

A terceira hipótese dos mínimos quadrados é que os quartos momentos de X_i e u_i são diferentes de zero e finitos ($0 < E(X_i^4) < \infty$ e $0 < E(u_i^4) < \infty$) ou, de forma equivalente, que os quatro momentos de X_i e Y_i são diferentes de zero e finitos. Essa hipótese limita a probabilidade de que seja selecionada uma observação com valores extremamente grandes de X_i ou u_i. Se fôssemos selecionar uma observação com X_i ou Y_i extremamente grande — isto é, com X_i ou Y_i bastante fora do intervalo normal dos dados —, essa observação receberia grande importância em uma regressão de MQO e tornaria os resultados da regressão enganosos.

A hipótese dos quartos momentos finitos é utilizada na matemática que justifica as aproximações de amostras grandes para as distribuições da estatística do teste de MQO. Encontramos essa hipótese no Capítulo 3 quando discutimos a consistência da variância da amostra. Especificamente, a Equação (3.8) declara que a variância da amostra s_Y^2 é um estimador consistente da variância da população σ_Y^2 (isto é, que $s_Y^2 \xrightarrow{p} \sigma_Y^2$). Se $Y_1, \ldots, Y_n$ são i.i.d. e o quarto momento de Y_i é finito, a lei dos grandes números no Conceito-Chave 2.6 se aplica à média $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)^2$, um passo importante na prova do Apêndice 3.3 ao mostrar que s_Y^2 é consistente. O papel da hipótese dos quartos momentos na teoria matemática da regressão de MQO é discutido mais adiante na Seção 15.3.

É possível argumentar que essa hipótese é uma particularidade técnica que normalmente vale na prática. O tamanho da turma está limitado pela capacidade física de uma sala de aula; o melhor que você pode fazer em um exame nacional é acertar todas as questões e o pior é errar todas. Como o tamanho da turma e a pontuação nos exames têm um intervalo finito, eles necessariamente têm quartos momentos finitos. De forma mais geral, as distribuições mais comumente utilizadas, como a normal, têm quatro momentos. Ainda assim, sob o aspecto matemático, algumas distribuições possuem quartos momentos infinitos; essa hipótese descarta essas distribuições. Se essa hipótese é válida, é improvável que inferências estatísticas que utilizam MQO sejam controladas por algumas poucas observações.

Uso das Hipóteses de Mínimos Quadrados

As três hipóteses de mínimos quadrados para o modelo de regressão linear estão resumidas no Conceito-Chave 4.3. As hipóteses de mínimos quadrados desempenham papéis duplos; voltaremos a eles várias vezes ao longo deste livro.

O primeiro papel é matemático: se essas hipóteses são válidas, então, como é mostrado na próxima seção, para amostras grandes, os estimadores de MQO possuem distribuições amostrais que são normais. Por sua vez, essa distribuição normal de amostra grande permite desenvolver métodos para teste de hipótese e construção de intervalos de confiança utilizando os estimadores de MQO.

Hipóteses de Mínimos Quadrados

Conceito-Chave 4.3

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, ..., n, \text{ onde:}$$

1. o termo de erro tem média condicional zero dado X_i, ou seja, $E(u_i | X_i) = 0$;
2. (X_i, Y_i), $i = 1, ..., n$ são seleções independente e identicamente distribuídas (i.i.d.) de sua distribuição conjunta; e
3. (X_i, u_i) têm quartos momentos finitos diferentes de zero.

O segundo papel é organizar as circunstâncias que colocam dificuldades para a regressão de MQO. Como veremos, a primeira hipótese dos mínimos quadrados é a mais importante para se considerar na prática. Uma razão pela qual a primeira hipótese dos mínimos quadrados pode não ser válida na prática é discutida na Seção 4.10 e no Capítulo 5; razões adicionais serão discutidas na Seção 7.2.

É importante também considerar se a segunda hipótese é válida em uma aplicação. Embora seja plausível sua aplicação em muitos conjuntos de dados de corte, a segunda hipótese não é apropriada para dados de séries temporais. Por isso, a hipótese i.i.d. será substituída por uma hipótese mais apropriada quando discutirmos regressão com dados de séries temporais na Parte 4.

Consideramos a terceira hipótese uma condição técnica que é comumente válida na prática, desse modo não nos estenderemos mais sobre ela.

4.4 Distribuição Amostral dos Estimadores de MQO

Como os estimadores de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ são calculados a partir de uma amostra selecionada aleatoriamente, os próprios estimadores são variáveis aleatórias com uma distribuição de probabilidade — a distribuição amostral — que descreve os valores que eles poderiam assumir em amostras aleatórias possíveis. Essas distribuições amostrais serão apresentadas nesta seção. Em amostras pequenas, essas distribuições são complexas, mas, em amostras grandes, são aproximadamente normais em razão do teorema central do limite.

Distribuição Amostral dos Estimadores de MQO

Revisão da distribuição amostral de $\overline{Y}$. Lembre-se da discussão nas seções 2.5 e 2.6 sobre a distribuição amostral da média de amostra $\overline{Y}$, um estimador da média da população desconhecida de Y, μ_Y. Como $\overline{Y}$ é calculado utilizando uma amostra selecionada aleatoriamente, $\overline{Y}$ é uma variável aleatória que assume valores diferentes de uma amostra para a seguinte; a probabilidade de que esses valores sejam diferentes está resumida em sua distribuição amostral. Embora a distribuição amostral de $\overline{Y}$ possa ser complexa quando o tamanho da amostra for pequeno, determinadas afirmações sobre ela são válidas para todo n. Em particular, a média da distribuição amostral é μ_Y, isto é, $E(\overline{Y}) = \mu_Y$, então $\overline{Y}$ é um estimador não viesado de μ_Y. Se n é grande, é possível dizer mais sobre a distribuição amostral. Em particular, o teorema central do limite (veja a Seção 2.6) afirma que essa distribuição é aproximadamente normal.

Distribuição amostral de $\hat{\beta}_0$ e $\hat{\beta}_1$. Essas idéias podem ser estendidas aos estimadores de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ do intercepto desconhecido β_0 e da declividade desconhecida β_1 da reta de regressão da população. Como os estimadores de MQO são calculados utilizando uma amostra aleatória, $\hat{\beta}_0$ e $\hat{\beta}_1$ são variáveis aleatórias que assumem valores diferentes de uma amostra para a seguinte; a probabilidade de que esses valores sejam diferentes está resumida em suas distribuições amostrais.

Embora a distribuição amostral de $\hat{\beta}_0$ e $\hat{\beta}_1$ possa ser complexa quando o tamanho da amostra for pequeno, determinadas afirmações sobre ela são válidas para todo n. Em particular, a média das distribuições amostrais

de $\hat{\beta}_0$ e $\hat{\beta}_1$ são β_0 e β_1. Em outras palavras, sob as hipóteses de mínimos quadrados do Conceito-Chave 4.3, temos

$$E(\hat{\beta}_0) = \beta_0 \text{ e } E(\hat{\beta}_1) = \beta_1, \tag{4.13}$$

isto é, $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores não viesados de β_0 e β_1. A prova de que $\hat{\beta}_1$ é não viesado é dada no Apêndice 4.3 e a prova de que $\hat{\beta}_0$ é não viesado é deixada como o Exercício 4.4.

Se a amostra for suficientemente grande, pelo teorema central do limite, a distribuição amostral de $\hat{\beta}_0$ e $\hat{\beta}_1$ tem uma boa aproximação pela distribuição normal bivariada (veja a Seção 2.4). Isso implica que as distribuições marginais de $\hat{\beta}_0$ e $\hat{\beta}_1$ são normais para amostras grandes.

Esse argumento invoca o teorema central do limite. Tecnicamente, o teorema central do limite diz respeito à distribuição de médias (como $\overline{Y}$). Se você examinar o numerador da Equação (4.8) para $\hat{\beta}_1$, verá que ele também é um tipo de média — não uma média simples, como $\overline{Y}$, mas uma média do produto, $(Y_i - \overline{Y})(X_i - \overline{X})$. Conforme discutido no Apêndice 4.3, o teorema central do limite se aplica a essa média de modo que, assim como a média mais simples $\overline{Y}$, ela possua distribuição normal para amostras grandes.

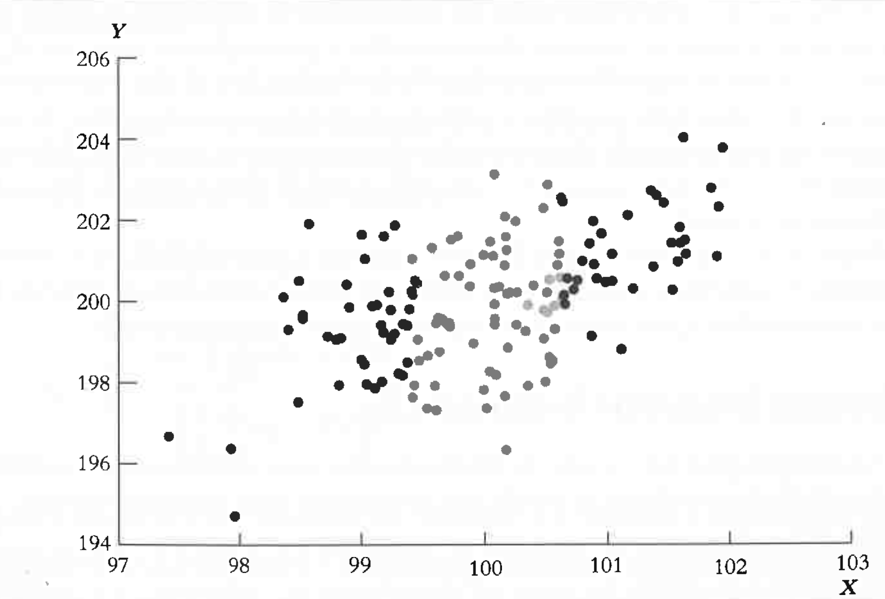
A aproximação normal para a distribuição dos estimadores de MQO para amostras grandes está resumida no Conceito-Chave 4.4. (O Apêndice 4.3 resume a derivação dessas fórmulas.) Uma questão relevante na prática é: quão grande n deve ser para que essas aproximações sejam confiáveis? Na Seção 2.6, sugerimos que $n = 100$ é suficientemente grande para que a distribuição amostral de $\overline{Y}$ tenha uma boa aproximação por uma distribuição normal; algumas vezes, um n menor já é suficiente. Esse critério se estende para as médias mais complicadas que aparecem na análise de regressão. Em praticamente todas as aplicações modernas da econometria, $n > 100$, de modo que trataremos as aproximações normais para as distribuições dos estimadores de MQO como confiáveis, a menos que haja bons motivos para pensar o contrário.

Os resultados no Conceito-Chave 4.4 indicam que os estimadores de MQO são consistentes, isto é, quando o tamanho da amostra é grande, $\hat{\beta}_0$ e $\hat{\beta}_1$ são próximos dos verdadeiros coeficientes da população β_0 e β_1 com probabilidade elevada. Isso porque as variâncias $\sigma^2_{\hat{\beta}_0}$ e $\sigma^2_{\hat{\beta}_1}$ dos estimadores tendem a zero à medida que n aumenta (n aparece no denominador das fórmulas para as variâncias), de modo que a distribuição dos estimadores de MQO estará fortemente concentrada em torno de suas médias, β_0 e β_1, quando n for grande.

Outra implicação das distribuições do Conceito-Chave 4.4 é que, em geral, quanto maior é a variância de X_i, menor é a variância $\sigma^2_{\hat{\beta}_1}$ de $\hat{\beta}_1$. Em termos matemáticos, isso ocorre porque a variância de $\hat{\beta}_1$ na Equação (4.14) é inversamente proporcional ao quadrado da variância de X_i: quanto maior é a $\text{var}(X_i)$, maior é o denominador da Equação (4.14), de modo que menor é $\sigma^2_{\hat{\beta}_1}$. Para entender melhor por que isso ocorre, veja a Figura 4.5, que

FIGURA 4.5 Variância de $\hat{\beta}_1$ e Variância de X

Os pontos acinzentados representam um conjunto de X_i's com uma variância pequena. Os pontos pretos representam um conjunto de X_i's com uma variância grande. A reta de regressão pode ser estimada de maneira mais precisa com os pontos pretos do que com os pontos acinzentados.



Distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$ para Amostras Grandes

Conceito-Chave 4.4

Se as hipóteses de mínimos quadrados do Conceito-Chave 4.3 são válidas, então, para amostras grandes, $\hat{\beta}_0$ e $\hat{\beta}_1$ possuem uma distribuição amostral normal conjunta. A distribuição normal de $\hat{\beta}_1$ para amostras grandes é $N(\beta_1, \sigma^2_{\hat{\beta}_1})$, onde a variância distribuição, $\sigma^2_{\hat{\beta}_1}$, é

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\text{var}[(X_i - \mu_X)u_i]}{[\text{var}(X_i)]^2}. \tag{4.14}$$

A distribuição normal de $\hat{\beta}_0$ para amostras grandes é $N(\beta_0, \sigma^2_{\hat{\beta}_0})$, onde

$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{\text{var}(H_i u_i)}{[E(H_i^2)]^2}, \quad \text{onde } H_i = 1 - \left(\frac{\mu_X}{E(X_i^2)}\right)X_i. \tag{4.15}$$

apresenta um gráfico de dispersão com 150 observações artificiais sobre X e Y. As observações indicadas por pontos acinzentados são as 75 observações mais próximas de $\overline{X}$. Suponha que lhe peçam para desenhar a reta mais precisa possível com pontos coloridos *ou* pretos – qual você escolheria? Seria mais fácil desenhar uma reta precisa com pontos pretos, que possuem uma variância maior do que os pontos coloridos. Da mesma forma, quanto maior a variância de X, mais preciso é $\hat{\beta}_1$.

A aproximação normal para a distribuição amostral de $\hat{\beta}_0$ e $\hat{\beta}_1$ é uma ferramenta poderosa. Com essa aproximação em mãos, seremos capazes de desenvolver métodos para fazer inferências sobre os verdadeiros valores da população dos coeficientes de regressão utilizando apenas uma amostra de dados.

4.5 Testando Hipóteses sobre um dos Coeficientes de Regressão

Sua cliente, a superintendente, liga para você com um problema. Um contribuinte irritado que está em sua sala afirma que a redução do tamanho das turmas não melhorará a pontuação nos exames, de modo que uma redução adicional seria um desperdício de dinheiro. O tamanho das turmas, alega o contribuinte, não tem nenhum efeito sobre a pontuação nos exames.

A alegação do contribuinte pode ser reformulada na linguagem da análise de regressão. Como o efeito de uma variação unitária no tamanho da turma sobre a pontuação nos exames é $\beta_{TamTurma}$, o contribuinte está afirmando que a reta de regressão da população é horizontal, isto é, que a declividade $\beta_{TamTurma}$ da reta de regressão da população é zero. Será que existe evidência, pergunta a superintendente, de que essa declividade seja diferente de zero em sua amostra de 420 observações das secretarias de ensino da Califórnia? Você pode rejeitar a hipótese do contribuinte de que $\beta_{TamTurma} = 0$ ou deve aceitá-la pelo menos de forma experimental na falta de novas evidências adicionais?

Nesta seção, discutimos testes de hipóteses sobre a declividade β_1 ou o intercepto β_0 da reta de regressão da população. Começamos discutindo em detalhe os testes bicaudais da declividade β_1 para então passarmos para os testes monocaudais e para os testes de hipóteses relativas ao intercepto β_0.

Hipóteses Bicaudais Relativas a β_1

O enfoque geral para o teste de hipóteses sobre esses coeficientes é o mesmo que para o teste de hipóteses sobre a média da população, de modo que começamos com uma breve revisão.

Teste de hipóteses sobre a média da população. Na Seção 3.2, você viu que a hipótese nula de que a média de Y é um valor específico $\mu_{Y,0}$ pode ser escrita como H_0: $E(Y) = \mu_{Y,0}$ e a alternativa bicaudal é H_1: $E(Y) \neq \mu_{Y,0}$.

Fórmula Geral da Estatística t

Conceito-Chave 4.5

Em geral, a estatística t possui a forma

$$t = \frac{\text{estimador} - \text{valor hipotético}}{\text{erro padrão do estimador}}. \tag{4.16}$$

O teste da hipótese nula H_0 contra a alternativa bicaudal segue os três passos resumidos no Conceito-Chave 3.6. O primeiro passo consiste em calcular o erro padrão de $\overline{Y}$, $EP(\overline{Y})$, que é um estimador do desvio padrão da distribuição amostral de $\overline{Y}$. O segundo passo consiste em calcular a estatística t, que possui a forma geral dada no Conceito-Chave 4.5; aplicada aqui, a estatística t é $t = (\overline{Y} - \mu_{Y,0})/EP(\overline{Y})$.

O terceiro passo consiste em calcular o valor p, que é o menor nível de significância em que a hipótese nula pode ser rejeitada, baseada na estatística de teste efetivamente observada; de modo equivalente, o valor p é a probabilidade de que seja obtida uma estatística por variação de amostragem aleatória no mínimo tão diferente do valor da hipótese nula quanto a estatística efetivamente observada, supondo que a hipótese nula esteja correta (veja o Conceito-Chave 3.5). Como a estatística t possui uma distribuição normal padrão para amostras grandes sob a hipótese nula, o valor p para um teste de hipótese bicaudal é $2\Phi(-|t^{ef}|)$, onde t^{ef} é o valor da estatística t efetivamente calculado e Φ é a distribuição acumulada normal padrão tabulada na Tabela 1 do Apêndice. De maneira alternativa, o terceiro passo pode ser substituído pela simples comparação da estatística t com o valor crítico apropriado para o teste com o nível de significância desejado; por exemplo, um teste bicaudal com nível de significância de 5 por cento rejeitaria a hipótese nula se $|t^{ef}| > 1{,}96$. Nesse caso, diz-se que a média da população é, em termos estatísticos, significativamente diferente do valor hipotético ao nível de significância de 5 por cento.

Testando hipóteses sobre a declividade β_1. Teoricamente, a característica crítica que justifica o procedimento de teste a seguir para a média de população é que, para amostras grandes, a distribuição amostral de $\overline{Y}$ é aproximadamente normal. Como $\hat{\beta}_1$ também possui distribuição amostral normal para amostras grandes, hipóteses sobre o valor verdadeiro da declividade β_1 podem ser testadas utilizando o mesmo enfoque geral.

As hipóteses nula e alternativa devem ser expressas com precisão antes de serem testadas. A hipótese do contribuinte irritado é de que $\beta_{TamTurma} = 0$. De modo mais geral, sob a hipótese nula, a verdadeira declividade da população β_1 assume algum valor específico, $\beta_{1,0}$. Sob a alternativa bicaudal, β_1 não é igual a $\beta_{1,0}$. Isto é, a **hipótese nula** e a **hipótese alternativa bicaudal** são

$$H_0: \beta_1 = \beta_{1,0} \textit{ versus } H_1: \beta_1 \neq \beta_{1,0} \text{ (hipótese alternativa bicaudal).} \tag{4.17}$$

Para testar a hipótese nula H_0, seguimos os três passos utilizados para a média da população.

O primeiro passo consiste em calcular o **erro padrão de $\hat{\beta}_1$**, $EP(\hat{\beta}_1)$. O erro padrão de $\hat{\beta}_1$ é um estimador de $\sigma_{\hat{\beta}_1}$, o desvio padrão da distribuição amostral de $\hat{\beta}_1$. Especificamente,

$$EP(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}, \tag{4.18}$$

onde

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \overline{X})^2\hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^2}. \tag{4.19}$$

O estimador da variância na Equação (4.19) é discutido no Apêndice 4.4. Embora a fórmula para $\hat{\sigma}^2_{\hat{\beta}_1}$ seja complicada, em aplicações o erro padrão é calculado por programas de regressão, de modo que é fácil utilizá-lo na prática.

O segundo passo consiste em calcular a **estatística** t:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{EP(\hat{\beta}_1)}. \quad (4.20)$$

O terceiro passo consiste em calcular o **valor p**, isto é, a probabilidade de se observar um valor de $\hat{\beta}_1$ no mínimo tão diferente de $\beta_{1,0}$ quanto a estimativa efetivamente calculada ($\hat{\beta}_1^{ef}$), supondo que a hipótese nula esteja correta. Expressando matematicamente,

$$\text{valor } p = P_{H_0}[|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{ef} - \beta_{1,0}|]$$

$$= P_{H_0}\left[\left|\frac{\hat{\beta}_1 - \beta_{1,0}}{EP(\hat{\beta}_1)}\right| > \left|\frac{\hat{\beta}_1^{ef} - \beta_{1,0}}{EP(\hat{\beta}_1)}\right|\right] = P_{H_0}(|t| > |t^{ef}|), \quad (4.21)$$

onde P_{H_0} representa a probabilidade calculada sob a hipótese nula, a segunda igualdade segue-se pela divisão por $EP(\hat{\beta}_1)$ e t^{ef} é o valor da estatística t efetivamente calculado. Como $\hat{\beta}_1$ tem distribuição aproximadamente normal para amostras grandes, sob a hipótese nula, a estatística t é distribuída aproximadamente como uma variável aleatória normal padrão, de modo que, para amostras grandes, temos

$$\text{valor } p = P(|Z| > |t^{ef}|) = 2\Phi(-|t^{ef}|). \quad (4.22)$$

Um valor pequeno do valor p, digamos menos de 5 por cento, fornece evidência contra a hipótese nula no sentido de que a probabilidade de se obter um valor de $\hat{\beta}_1$ por variação aleatória pura de uma amostra para a seguinte é menor do que 5 por cento se a hipótese nula estiver de fato correta. Se for esse o caso, a hipótese nula é rejeitada ao nível de significância de 5 por cento.

Alternativamente, a hipótese pode ser testada ao nível de significância de 5 por cento simplesmente pela comparação do valor da estatística t com $\pm 1{,}96$, o valor crítico para um teste bicaudal, e pela rejeição da hipótese nula ao nível de 5 por cento se $|t^{ef}| > 1{,}96$.

Esses passos estão resumidos no Conceito-Chave 4.6.

Aplicação à pontuação nos exames. O estimador de MQO do coeficiente da declividade, estimado utilizando-se as 420 observações da Figura 4.2 e expresso na Equação (4.7), é $-2{,}28$. Seu erro padrão é 0,52, isto é, $EP(\hat{\beta}_1) = 0{,}52$. Assim, para testar a hipótese nula de que $\beta_{TamTurma} = 0$, construímos a estatística t utilizando a Equação (4.20); portanto, $t^{ef} = (-2{,}28 - 0)/0{,}52 = -4{,}38$.

A estatística t excede 2,58, o valor crítico bicaudal de 1 por cento, de modo que a hipótese nula é rejeitada em favor da alternativa bicaudal ao nível de significância de 1 por cento. Alternativamente, podemos calcular o valor p associado a $t = -4{,}38$. Essa probabilidade é a área nas caudas da distribuição normal padrão, como mostrado na Figura 4.6, e é bastante pequena, aproximadamente 0,00001 ou 0,001 por cento. Isto é, se a hipótese nula $\beta_{TamTurma} = 0$ é verdadeira, a probabilidade de se obter um valor de $\hat{\beta}_1$ tão distante da hipótese nula quanto o valor efetivamente obtido é muito pequena, menor do que 0,001 por cento. Como esse acontecimento é bastante improvável, é razoável concluir que a hipótese nula é falsa.

Hipóteses Monocaudais Relativas a β_1

A discussão até o momento concentrou-se em testar a hipótese de que $\beta_1 = \beta_{1,0}$ contra a hipótese de que $\beta_1 \neq \beta_{1,0}$. Esse é um teste de hipótese bicaudal, uma vez que, sob a alternativa, β_1 pode ser ou maior ou menor do que $\beta_{1,0}$. Algumas vezes, entretanto, é apropriado utilizar um teste de hipótese monocaudal. Por exemplo, no problema da pontuação nos exames e da razão aluno-professor, muitas pessoas acreditam que turmas menores proporcionam um ambiente de aprendizado melhor. Sob essa hipótese, β_1 é negativo: turmas menores levam a notas maiores. Pode fazer sentido, portanto, testar a hipótese nula de que $\beta_1 = 0$ (nenhum efeito) contra a alternativa monocaudal de que $\beta_1 < 0$.

Para um teste monocaudal, a hipótese nula e a hipótese alternativa monocaudal são:

$$H_0: \beta_1 = \beta_{1,0} \textit{ versus } H_1: \beta_1 < \beta_{1,0} \text{ (alternativa monocaudal)} \quad (4.23)$$

Testando a Hipótese $\beta_1 = \beta_{1,0}$ contra a Alternativa $\beta_1 \neq \beta_{1,0}$

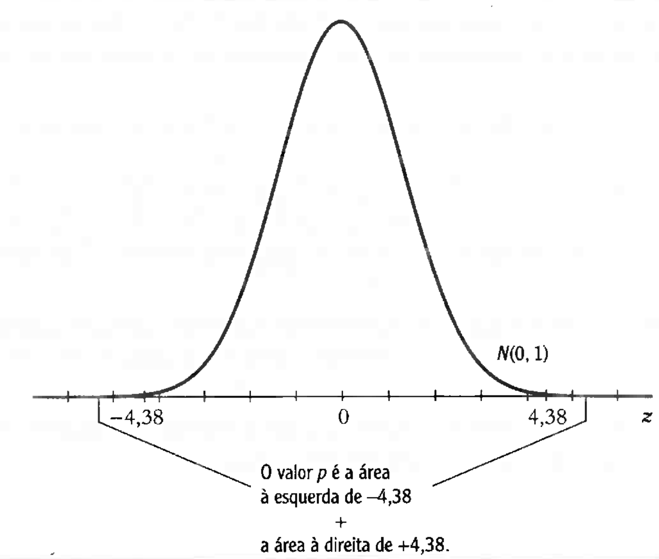
Conceito-Chave 4.6

1. Calcule o erro padrão de $\hat{\beta}_1$, $EP(\hat{\beta}_1)$ (veja a Equação (4.18)).
2. Calcule a estatística t (veja a Equação (4.20)).
3. Calcule o valor p (veja a Equação (4.22)). Rejeite a hipótese ao nível de significância de 5 se o valor p for menor do que 0,05 ou, de forma equivalente, se $|t^{ef}| > 1{,}96$.

O erro padrão e (em geral) a estatística t e o valor p que testam se $\beta_1 = 0$ são calculados automaticamente pelo programa de regressão.

FIGURA 4.6 Calculando o Valor p de um Teste Bicaudal Quando $t^{ef} = -4{,}38$

O valor p de um teste bicaudal é a probabilidade de que $|Z| \geq |t^{ef}|$, onde Z é uma variável aleatória normal padrão e t^{ef} é o valor da estatística t calculada a partir da amostra. Quando $t^{ef} = -4{,}38$, o valor p é de apenas 0,00001.



onde $\beta_{1,0}$ é o valor de β_1 sob a hipótese nula (0 no exemplo da razão aluno-professor) e β_1 é menor do que $\beta_{1,0}$ sob a hipótese alternativa. Se a alternativa é que β_1 é maior que $\beta_{1,0}$, a desigualdade na Equação (4.23) é invertida.

Como a hipótese nula é a mesma para um teste de hipótese monocaudal e um teste de hipótese bicaudal, a construção da estatística t é a mesma. A única diferença entre um teste de hipótese monocaudal e um bicaudal é o modo como você interpreta a estatística t. Para a alternativa monocaudal em (4.23), a hipótese nula é rejeitada contra a alternativa monocaudal para valores da estatística t negativos e grandes, mas não para positivos e grandes: em vez de ser rejeitada se $|t^{ef}| > 1{,}96$, a hipótese é rejeitada ao nível de significância de 5 por cento se $t^{ef} < -1{,}645$.

O valor p para um teste monocaudal é obtido de uma distribuição acumulada normal padrão como

$$\text{valor } p = P(Z < t^{ef}) = \Phi(t^{ef}) \text{ (valor } p\text{, teste monocaudal à esquerda)} \quad (4.24)$$

Se a hipótese alternativa é de que β_1 é maior do que $\beta_{1,0}$, as desigualdades nas equações (4.23) e (4.24) são invertidas, de modo que o valor p é a probabilidade correspondente à cauda direita, $P(Z > t^{ef})$.

Quando um teste monocaudal deve ser utilizado? Na prática, as hipóteses alternativas monocaudais devem ser utilizadas quando existe um motivo claro para β_1 estar em determinado lado do valor nulo $\beta_{1,0}$ sob a alternativa. Esse motivo pode vir da teoria econômica, da evidência empírica anterior ou de ambas. Entretanto, mesmo que a princípio pareça que a alternativa relevante é monocaudal, isso pode não ser necessariamente verdade após alguma reflexão. Um medicamento de nova geração submetido a testes clínicos pode mostrar-se na realidade perigoso em virtude dos efeitos colaterais previamente desconhecidos. No exemplo do tamanho das turmas, lembramo-nos da piada segundo a qual o segredo do sucesso de uma universidade é admitir alunos talentosos e depois assegurar que os professores os deixem em paz e provoquem o mínimo de perturbação possível. Na prática, essa ambigüidade freqüentemente leva os econometristas a utilizar testes de hipótese bicaudais.

Aplicação à pontuação nos exames. A estatística t para o teste de hipótese de que o tamanho da turma não tem efeito sobre a pontuação nos exames (de modo que $\beta_{1,0} = 0$ na Equação (4.23)) é $t^{ef} = -4{,}38$. Isso é menos de $-2{,}33$ (o valor crítico para um teste monocaudal a um nível de significância de 1 por cento), de modo que a hipótese nula é rejeitada contra a alternativa monocaudal ao nível de 1 por cento. Na verdade, o valor p é menor do que 0,0006 por cento. Com base nesses dados, você pode rejeitar a alegação do contribuinte irritado de que a estimativa negativa da declividade teve origem simplesmente na variação de amostragem aleatória ao nível de significância de 1 por cento.

Testando Hipóteses sobre o Intercepto β_0

Essa discussão se concentrou no teste de hipóteses sobre a declividade β_1. Ocasionalmente, contudo, a hipótese diz respeito ao intercepto, β_0. As hipóteses nulas relativas ao intercepto e à alternativa bicaudal são

$$H_0: \beta_0 = \beta_{0,0} \text{ } versus \text{ } H_1: \beta_0 \neq \beta_{0,0} \text{ (alternativa bicaudal).} \tag{4.25}$$

O enfoque geral para o teste dessa hipótese nula consiste nos três passos do Conceito-Chave 4.6 aplicados a β_0 (a fórmula para o erro padrão de $\hat{\beta}_0$ é dada no Apêndice 4.4). Se a alternativa é monocaudal, o enfoque é modificado conforme discutido na subseção anterior relativa a hipóteses sobre a declividade.

Testes de hipótese são úteis quando você tem uma hipótese nula específica em mente (assim como o contribuinte irritado). Poder aceitar ou rejeitar essa hipótese nula com base na evidência estatística fornece uma ferramenta poderosa para lidar com a incerteza inerente de utilizar uma amostra para aprender sobre uma população. Entretanto, muitas vezes não existe somente uma hipótese sobre um coeficiente de regressão que seja dominante e, em vez disso, o que se deseja conhecer é um intervalo de valores do coeficiente consistentes com os dados. Isso requer a construção de um intervalo de confiança.

4.6 Intervalos de Confiança para um Coeficiente de Regressão

Como qualquer estimativa estatística da declividade β_1 possui necessariamente incerteza com relação à amostragem, não podemos determinar exatamente o valor verdadeiro de β_1 a partir dos dados da amostra. Mas podemos utilizar o estimador de MQO e seu erro padrão para construir um intervalo de confiança para a declividade β_1 ou para o intercepto β_0.

Intervalo de confiança para β_1. Lembre-se de que um **intervalo de confiança de 95 por cento** para β_1 possui duas definições equivalentes: (1) é o conjunto de valores que não podem ser rejeitados utilizando-se um teste de hipótese bicaudal a um nível de significância de 5 por cento; (2) é um intervalo que possui 95 por cento de probabilidade de conter o valor verdadeiro de β_1; isto é, em 95 por cento das amostras possíveis que possam ser selecionadas, o intervalo de confiança irá conter o valor verdadeiro de β_1. Como esse intervalo contém o valor verdadeiro em 95 por cento de todas as amostras, diz-se que ele possui um **nível de confiança** de 95 por cento.

O motivo pelo qual essas duas definições são equivalentes é descrito a seguir. Um teste de hipótese a um nível de significância de 5 por cento irá, por definição, rejeitar o valor verdadeiro de β_1 em apenas 5 por cento de todas as amostras possíveis, isto é, em 95 por cento de todas as amostras possíveis o valor verdadeiro de β_1 *não* será rejeitado. Como o intervalo de confiança de 95 por cento (conforme a primeira definição) é o conjunto de todos os valores de β_1 que *não* são rejeitados ao nível de significância de 5 por cento, segue-se que o valor verdadeiro de β_1 estará contido no intervalo de confiança em 95 por cento de todas as amostras possíveis.

Intervalos de Confiança para β_1

Conceito-Chave 4.7

Um intervalo de confiança bicaudal de 95 por cento para β_1 é um intervalo que contém o valor verdadeiro de β_1 com uma probabilidade de 95 por cento, isto é, contém o valor verdadeiro de β_1 em 95 por cento de todas as amostras possíveis selecionadas aleatoriamente. De modo equivalente, ele é também o conjunto de todos os valores de β_1 que não podem ser rejeitados por um teste de hipótese bicaudal ao nível de significância de 5 por cento. Quando o tamanho da amostra é grande, o intervalo de confiança é construído como:

$$\text{intervalo de confiança de 95 por cento para } \beta_1 = (\hat{\beta}_1 - 1{,}96EP(\hat{\beta}_1), \hat{\beta}_1 + 1{,}96EP(\hat{\beta}_1)). \tag{4.26}$$

Como no caso de um intervalo de confiança para a média da população (veja a Seção 3.3), em princípio um intervalo de confiança de 95 por cento pode ser calculado testando-se todos os valores possíveis de β_1 (isto é, testando a hipótese nula $\beta_1 = \beta_{1,0}$ para todos os valores de $\beta_{1,0}$) ao nível de significância de 5 por cento utilizando a estatística t. O intervalo de confiança de 95 por cento é então a coleção de todos os valores de β_1 que não são rejeitados. Porém, construir a estatística t para todos os valores de β_1 levaria uma eternidade.

Uma forma mais fácil de construir o intervalo de confiança é observar que a estatística t irá rejeitar o valor hipotético $\beta_{1,0}$ quando $\beta_{1,0}$ estiver fora do intervalo $\hat{\beta}_1 \pm 1{,}96\ EP(\hat{\beta}_1)$. Isto é, o intervalo de confiança de 95 por cento para β_1 é o intervalo $(\hat{\beta}_1 - 1{,}96\ EP(\hat{\beta}_1), \hat{\beta}_1 + 1{,}96\ EP(\hat{\beta}_1))$. Esse argumento é equivalente ao argumento utilizado para desenvolver um intervalo de confiança para a média da população.

A construção de um intervalo de confiança para β_1 é resumida no Conceito-Chave 4.7.

Intervalo de confiança para β_0. Um intervalo de confiança de 95 por cento para β_0 é construído conforme descrito no Conceito-Chave 4.7, com $\hat{\beta}_0$ e $EP(\hat{\beta}_0)$ substituindo $\hat{\beta}_1$ e $EP(\hat{\beta}_1)$.

Aplicação à pontuação nos exames. A regressão de MQO da pontuação nos exames contra a razão aluno-professor, descrita na Equação (4.7), resultou em $\hat{\beta}_0 = 698{,}7$ e $\hat{\beta}_1 = -2{,}28$. Os erros padrão para essas estimativas são $EP(\hat{\beta}_0) = 10{,}4$ e $EP(\hat{\beta}_1) = 0{,}52$.

Em virtude da importância dos erros padrão, de agora em diante iremos incluí-los entre parênteses abaixo dos coeficientes estimados quando descrevermos as retas de regressão de MQO:

$$\widehat{PontExame} = \underset{(10,4)}{698{,}9} - \underset{(0,52)}{2{,}28} \times RAP. \tag{4.27}$$

O intervalo de confiança bicaudal de 95 por cento para β_1 é $\{-2{,}28 \pm 1{,}96 \times 0{,}52\}$, isto é, $-3{,}30 \leq \beta_1 \leq -1{,}26$. O valor $\beta_1 = 0$ não está nesse intervalo de confiança; desse modo (como já sabíamos da Seção 4.5) a hipótese $\beta_1 = 0$ pode ser rejeitada ao nível de significância de 5 por cento.

Intervalos de confiança para efeitos previstos de variações em X. O **intervalo de confiança de 95 por cento** para β_1 pode ser utilizado para construir um intervalo de confiança de 95 por cento para o efeito previsto de uma variação global em X.

Considere uma variação em X de determinado montante, Δx. A variação prevista em Y associada a essa variação em X é $\beta_1 \Delta x$. A declividade da população β_1 é desconhecida, mas, como podemos construir um intervalo de confiança para β_1, podemos construir um intervalo de confiança para o efeito previsto $\beta_1 \Delta x$. Como um dos extremos de um intervalo de confiança de 95 por cento para β_1 é $\hat{\beta}_1 - 1{,}96EP(\hat{\beta}_1)$, o efeito previsto da variação Δx utilizando essa estimativa de β_1 é $(\hat{\beta}_1 - 1{,}96EP(\hat{\beta}_1)) \times \Delta x$. O outro extremo do intervalo de confiança é $\hat{\beta}_1 + 1{,}96EP(\hat{\beta}_1)$; o efeito previsto da variação que utiliza essa estimativa é $(\hat{\beta}_1 + 1{,}96EP(\hat{\beta}_1)) \times \Delta x$. Portanto, um intervalo de confiança de 95 por cento para o efeito de mudar x no montante Δx pode ser expresso como

intervalo de confiança de 95 por cento $\beta_1\Delta x =$
$$(\hat{\beta}_1\Delta x - 1{,}96EP(\hat{\beta}_1) \times \Delta x, \hat{\beta}_1\Delta x + 1{,}96EP(\hat{\beta}_1) \times \Delta x). \tag{4.28}$$

Por exemplo, nossa superintendente hipotética está considerando a redução da razão aluno-professor em 2. Como o intervalo de confiança de 95 por cento para β_1 é (–3,30, –1,26), o efeito da redução da razão aluno-professor em 2 pode ser tão grande quanto –3,30 × (–2) = 6,60, ou tão pequeno quanto –1,26 × (–2) = 2,52. Portanto, pode-se prever que a diminuição da razão aluno-professor em 2 deve aumentar a pontuação nos exames entre 2,52 e 6,60 pontos, com um nível de confiança de 95 por cento.

4.7 Regressão Quando X É uma Variável Binária

A discussão até o momento concentrou-se no caso em que o regressor é uma variável contínua. A análise de regressão também pode ser utilizada quando o regressor é binário, isto é, quando assume apenas dois valores, 0 ou 1. Por exemplo, X pode ser o sexo de um trabalhador (= 1 se feminino, = 0 se masculino), se uma diretoria regional de ensino é urbana ou rural (= 1 se urbana, = 0 se rural) ou se o tamanho da turma da diretoria é pequeno ou grande (= 1 se pequeno, = 0 se grande). Uma variável binária também é chamada de **variável indicador** ou algumas vezes de **variável *dummy***.

Interpretação dos Coeficientes de Regressão

O mecanismo da regressão com um regressor binário é o mesmo que seria se a variável fosse contínua. A interpretação de β_1, entretanto, é diferente; disso conclui-se que a regressão com uma variável binária é equivalente a fazer uma análise da diferença entre médias, conforme descrito na Seção 3.4.

Para visualizar isso, suponha que você tenha uma variável D_i que seja igual a 0 ou a 1, dependendo de a razão aluno-professor ser menor do que 20:

$$D_i = \begin{cases} 1 \text{ se a razão aluno-professor na } i\text{-ésima diretoria} < 20 \\ 0 \text{ se a razão aluno-professor na } i\text{-ésima diretoria} \geq 20. \end{cases} \tag{4.29}$$

O modelo de regressão da população com D_i como regressor é

$$Y_i = \beta_0 + \beta_1 D_i + u_i, i = 1, \ldots, n. \tag{4.30}$$

Esse modelo é igual ao modelo de regressão com o regressor contínuo X_i, exceto pelo fato de que agora o regressor é a variável binária D_i. Como D_i não é contínua, não é útil pensar em β_1 como uma declividade; de fato, como D_i pode assumir somente dois valores, não existe uma "reta", de modo que não faz sentido falar em declividade. Portanto, não iremos nos referir a β_1 como a declividade na Equação (4.30); em vez disso, nos referiremos a β_1 simplesmente como o **coeficiente que multiplica D_i** nessa regressão ou, de forma mais sucinta, o **coeficiente de D_i**.

Se β_1 da Equação (4.30) não é uma declividade, então o que é? A melhor forma de interpretar β_0 e β_1 em uma regressão com um regressor binário é considerar, um de cada vez, os dois casos possíveis, $D_i = 0$ e $D_i = 1$. Se a razão aluno-professor é alta, então $D_i = 0$ e a Equação (4.30) torna-se

$$Y_i = \beta_0 + u_i \quad (D_i = 0). \tag{4.31}$$

Como $E(u_i | D_i) = 0$, a expectativa condicional de Y_i quando $D_i = 0$ é $E(Y_i | D_i = 0) = \beta_0$, isto é, β_0 é o valor da média da população no que tange à pontuação nos exames quando a razão aluno-professor é alta. Do mesmo modo, quando $D_i = 1$,

$$Y_i = \beta_0 + \beta_1 + u_i \quad (D_i = 1). \tag{4.32}$$

Assim, quando $D_i = 1$, $E(Y_i | D_i = 1) = \beta_0 + \beta_1$; isto é, $\beta_0 + \beta_1$ é o valor da média da população no que tange à pontuação nos exames quando a razão aluno-professor é baixa.

Como $\beta_0 + \beta_1$ é a média da população de Y_i quando $D_i = 1$ e β_0 é a média da população de Y_i quando $D_i = 0$, a diferença $(\beta_0 + \beta_1) - \beta_0 = \beta_1$ é a diferença entre essas duas médias. Em outras palavras, β_1 é a diferença entre a expectativa condicional de Y_i quando $D_i = 1$ e quando $D_i = 0$, ou $\beta_1 = E(Y_i | D_i = 1) - E(Y_i | D_i = 0)$. No exemplo da pontuação nos exames, β_1 é a diferença entre a pontuação média nos exames nas diretorias com razão aluno-professor baixa e a pontuação média nos exames nas diretorias com razão aluno-professor alta.

Como β_1 é a diferença entre as médias da população, faz sentido pensar que o estimador de MQO β_1 seja a diferença entre as médias da amostra de Y_i nos dois grupos, e de fato esse é o caso.

Testes de hipótese e intervalos de confiança. Se as duas médias da população são iguais, β_1 da Equação (4.30) é zero. Portanto, a hipótese nula de que as duas médias da população são iguais pode ser testada contra a hipótese alternativa de que elas diferem por meio do teste da hipótese nula $\beta_1 = 0$ contra a alternativa $\beta_1 \neq 0$. Essa hipótese pode ser testada com o procedimento esboçado na Seção 4.5. Especificamente, a hipótese nula pode ser rejeitada a um nível de 5 por cento contra a alternativa bicaudal quando a estatística t de MQO $t = \hat{\beta}_1/EP(\hat{\beta}_1)$ excede 1,96 em valor absoluto. Do mesmo modo, um intervalo de confiança de 95 por cento para β_1, construído como $\hat{\beta}_1 \pm 1{,}96EP(\hat{\beta}_1)$ conforme descrito na Seção 4.6, fornece um intervalo de confiança de 95 por cento para a diferença entre as duas médias da população.

Aplicação à pontuação nos exames. Como exemplo, uma regressão da pontuação nos exames contra a variável binária razão aluno-professor D definida na Equação (4.29), estimada por MQO utilizando as 420 observações da Figura 4.2, gera

$$\widehat{PontExame} = \underset{(1,3)}{650{,}0} + \underset{(1,8)}{7{,}4}D \tag{4.33}$$

onde os erros padrão das estimativas de MQO dos coeficientes β_0 e β_1 estão entre parênteses abaixo das estimativas de MQO. Desse modo, a pontuação média nos exames para a subamostra com razões aluno-professor maiores ou iguais a 20 (isto é, para $D = 0$) é 650,0 e a pontuação média nos exames para a subamostra com razões aluno-professor menores do que 20 (de modo que $D = 1$) é 650,0 + 7,4 = 657,4. Assim, a diferença entre a pontuação média nos exames da amostra para os dois grupos é 7,4. Essa é a estimativa de MQO de β_1, o coeficiente da variável binária razão aluno-professor D.

Será que a diferença entre as pontuações médias nos exames da população dos dois grupos é, em termos estatísticos, significativamente diferente de zero ao nível de 5 por cento? Para descobrir essa resposta, construa a estatística t para β_1 : $t = 7{,}4/1{,}8 = 4{,}04$. Isso excede 1,96 em valor absoluto, de modo que a hipótese de que a pontuação média nos exames da população nas diretorias com razões aluno-professor alta e baixa é a mesma pode ser rejeitada ao nível de significância de 5 por cento.

O estimador de MQO e seu erro padrão podem ser utilizados para construir um intervalo de confiança de 95 por cento para a diferença verdadeira entre as médias. Este é 7,4 ± 1,96 × 1,8 = (3,9, 10,9). Esse intervalo de confiança exclui $\beta_1 = 0$, de modo que (como sabemos do parágrafo anterior) a hipótese $\beta_1 = 0$ pode ser rejeitada ao nível de significância de 5 por cento.

4.8 R^2 e o Erro Padrão da Regressão

O R^2 e o erro padrão da regressão são duas medidas do quanto a reta de regressão de MQO se ajusta aos dados. O R^2 varia entre zero e um e mede a fração da variância de Y_i explicada pela variação em X_i. O erro padrão da regressão mede quão distante Y_i está, em geral, de seu valor previsto.

R^2

O **R^2 da regressão** é a fração da variância da amostra de Y_i explicada por (ou prevista por) X_i. As definições do valor previsto e do resíduo (veja o Conceito-Chave 4.2) nos permitem escrever a variável dependente Y_i como a soma do valor previsto $\hat{Y}_i$, mais o resíduo $\hat{u}_i$:

$$Y_i = \hat{Y}_i + \hat{u}_i. \tag{4.34}$$

Nessa notação, o R^2 é a razão entre a variância da amostra de $\hat{Y}_i$ e a variância da amostra de Y_i.

Matematicamente, o R^2 pode ser escrito como a razão entre a soma dos quadrados explicada e a soma dos quadrados total. A **soma dos quadrados explicada**, ou ***SQE***, é a soma dos quadrados dos desvios dos valores previstos de Y_i, $\hat{Y}_i$, em relação a sua média, e a **soma dos quadrados total**, ou ***SQT***, é a soma dos quadrados dos desvios de Y_i em relação a sua média:

$$SQE = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 \text{ e} \tag{4.35}$$

$$SQT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2, \tag{4.36}$$

onde a Equação (4.35) utiliza o fato de que $\overline{Y}$ é igual ao valor previsto da média da amostra de MQO (provado no Apêndice 4.3).

O R^2 é a razão entre a soma dos quadrados explicada e a soma dos quadrados total:

$$R^2 = \frac{SQE}{SQT}. \tag{4.37}$$

Alternativamente, o R^2 pode ser escrito em termos da fração da variância de Y_i *não* explicada por X_i. A **soma dos quadrados dos resíduos**, ou ***SQR***, é a soma dos quadrados dos resíduos de MQO:

$$SQR = \sum_{i=1}^{n} \hat{u}_i^2. \tag{4.38}$$

O Apêndice 4.3 mostra que $SQT = SQE + SQR$. Portanto, o R^2 também pode ser expresso como um menos a razão entre a soma dos quadrados dos resíduos e a soma dos quadrados total:

$$R^2 = 1 - \frac{SQR}{SQT}. \tag{4.39}$$

Finalmente, o R^2 da regressão de Y sobre o único regressor X é o quadrado do coeficiente de correlação entre Y e X.

O R^2 situa-se entre zero e um. Se $\hat{\beta}_1 = 0$, então X_i não explica nenhuma variação de Y_i e o valor previsto de Y_i baseado na regressão é apenas a média da amostra de Y_i. Nesse caso, a soma dos quadrados explicada é zero e a soma dos quadrados dos resíduos é igual à soma dos quadrados total; assim, R^2 é zero. Se, em vez disso, X_i explica toda variação de Y_i, então $Y_i = \hat{Y}_i$ para todo i e cada resíduo é igual a zero (isto é, $\hat{u}_i = 0$), de modo que $SQE = SQT$ e $R^2 = 1$. Em geral, o R^2 não assume os valores extremos zero ou um, mas situa-se entre eles. Um R^2 próximo de um indica que o regressor é bom na previsão de Y_i, ao passo que um R^2 próximo de zero indica que o regressor não é muito bom na previsão de Y_i.

Erro Padrão da Regressão

O **erro padrão da regressão**, ou ***EPR***, é um estimador do desvio padrão do erro da regressão u_i. Como os erros da regressão $u_1, ..., u_n$ não são observados, o *EPR* é calculado utilizando seus correspondentes da amostra, os resíduos de MQO $\hat{u}_1, ..., \hat{u}_n$. A fórmula para o *EPR* é

$$EPR = s_{\hat{u}}, \text{ onde } s_{\hat{u}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SQR}{n-2}, \tag{4.40}$$

onde a fórmula para $s_{\hat{u}}^2$ utiliza o fato (provado no Apêndice 4.3) de que a média da amostra dos resíduos de MQO é zero.

A fórmula para o *EPR* na Equação (4.40) é igual à fórmula para o desvio padrão da amostra de Y dado na Equação (3.7) da Seção 3.2, exceto pelo fato de que nessa equação $Y_i - \overline{Y}$ é substituído por $\hat{u}_i$, e o divisor na Equação (3.7) é $n - 1$, ao passo que aqui é $n - 2$. Os motivos para a utilização do divisor $n - 2$ aqui (em vez de n) e do divisor $n - 1$ na Equação (3.7) são os mesmos: corrigir um pequeno viés para baixo introduzido pela estimação de dois coeficientes de regressão. Isso se chama correção de "graus de liberdade"; como foram estimados dois coeficientes (β_0 e β_1), dois "graus de liberdade" dos dados foram perdidos, de modo que o divisor nesse fator é $n - 2$. (A matemática por trás disso é discutida na Seção 15.4.) Quando n é grande, a diferença entre a divisão por n, $n - 1$ ou $n - 2$ é desprezível.

4.9 Heteroscedasticidade e Homoscedasticidade

Nossa única hipótese sobre a distribuição de u_i condicional a X_i é que ela possui uma média igual a zero (a primeira hipótese dos mínimos quadrados). Se, além disso, a *variância* dessa distribuição condicional não depende de X_i, os erros são chamados de homoscedásticos. Nesta seção, discutimos a homoscedasticidade, suas implicações teóricas, as fórmulas simplificadas para os erros padrão dos estimadores de MQO que surgem quando os erros são homoscedásticos e os riscos que você corre ao utilizar essas fórmulas simplificadas na prática.

O Que São Heteroscedasticidade e Homoscedasticidade?

Definições de Heteroscedasticidade e Homoscedasticidade. O termo de erro u_i é **homoscedástico** se a variância da distribuição condicional de u_i dado X_i é constante para $i = 1, ..., n$ e em particular não depende de X_i. Caso contrário, o termo de erro é **heteroscedástico**.

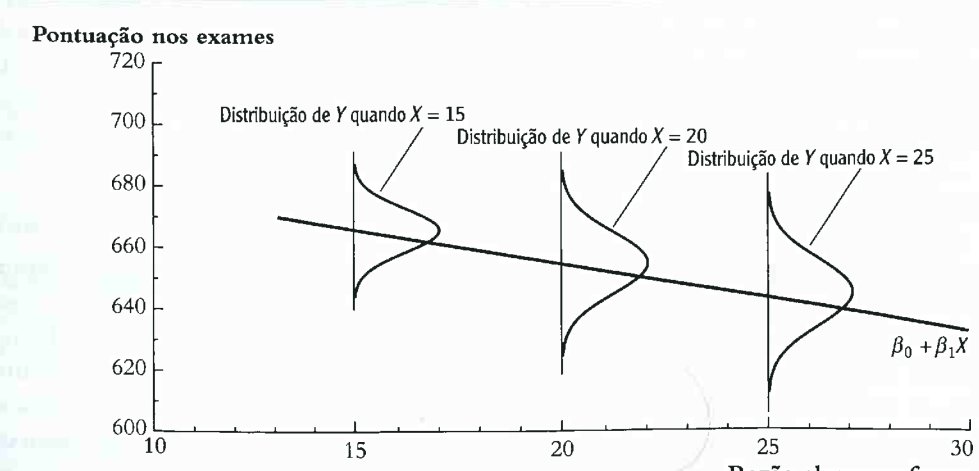
Para fins ilustrativos, volte à Figura 4.4. A distribuição dos erros u_i é mostrada para vários valores de x. Como essa distribuição se aplica especificamente ao valor indicado de x, essa é a distribuição condicional de u_i dado $X_i = x$. Conforme desenhado na figura, todas essas distribuições condicionais possuem a mesma dispersão; de modo mais preciso, a variância dessas distribuições é a mesma para os diversos valores de x. Isto é, na Figura 4.4, a variação condicional de u_i dado $X_i = x$ não depende de x, de modo que os erros ilustrados nessa figura são homoscedásticos.

A Figura 4.7, por sua vez, ilustra um caso em que a distribuição condicional de u_i tem dispersão maior à medida que x aumenta. Para valores pequenos de x, a distribuição possui uma dispersão pequena, mas, para valores grandes de x, a dispersão é maior. Assim, na Figura 4.7, a variância de u_i dado $X_i = x$ aumenta com x, de modo que os erros nessa figura são heteroscedásticos.

O Conceito-Chave 4.8 apresenta as definições de heteroscedasticidade e homoscedasticidade.

FIGURA 4.7 Um Exemplo de Heteroscedasticidade

Assim como a Figura 4.4, esta mostra a distribuição condicional da pontuação nos exames para três tamanhos de turma diferentes. Ao contrário da Figura 4.4, essas distribuições tornam-se mais dispersas (possuem uma variância maior) para tamanhos de turmas maiores. Como a variância da distribuição de u dado X, var(u|X) depende de X, u é heteroscedástico.



Heteroscedasticidade e Homoscedasticidade

O termo de erro u_i é homoscedástico se a variância da distribuição condicional de u_i dado X_i, $\text{var}(u_i | X_i = x)$, é constante para $i = 1, ..., n$ e em particular não depende de x; caso contrário, o termo de erro é heteroscedástico.

Conceito-Chave 4.8

Exemplo. Esses termos são verdadeiros palavrões e suas definições podem parecer abstratas. Para esclarecê-los com um exemplo, deixemos de lado o problema da pontuação nos exames e da razão aluno-professor e voltemos ao exemplo dos salários dos homens com curso superior *versus* mulheres com curso superior, discutido na Seção 3.5. Seja $HOMEM_i$ uma variável binária igual a 1 para homens com curso superior e igual a 0 para mulheres com curso superior. O modelo de regressão com variável binária que relaciona o salário de uma pessoa a seu sexo é

$$Salário_i = \beta_0 + \beta_1 HOMEM_i + u_i \tag{4.41}$$

para $i = 1, ..., n$. Como o regressor é binário, β_1 é a diferença entre as médias da população dos dois grupos; nesse caso, a diferença entre o salário médio dos homens e das mulheres com curso superior.

A definição de homoscedasticidade afirma que a variância de u_i não depende do regressor. Aqui o regressor é $HOMEM_i$, de modo que o que está em questão é se a variância do termo de erro depende de $HOMEM_i$. Em outras palavras, será que a variância do termo de erro é igual para homens e mulheres? Se for esse o caso, o erro é homoscedástico; caso contrário, é heteroscedástico.

Decidir se a variância de u_i depende de $HOMEM_i$ requer bastante reflexão sobre o que realmente é o termo de erro. Para isso, é interessante escrever a Equação (4.41) como duas equações separadas, uma para homens e outra para mulheres:

$$Salário_i = \beta_0 + u_i \text{ (mulheres) e} \tag{4.42}$$

$$Salário_i = \beta_0 + \beta_1 + u_i \text{ (homens).} \tag{4.43}$$

Portanto, para mulheres, u_i é o desvio do salário da i-ésima mulher em relação ao salário médio da população de mulheres (β_0) e, para os homens, u_i é o desvio do salário do i-ésimo homem em relação ao salário médio da população de homens ($\beta_0 + \beta_1$). Segue-se que a afirmação "a variância de u_i não depende de $HOMEM$" é equivalente à afirmação "a variância do salário é a mesma para homens e mulheres". Em outras palavras, nesse exemplo, o termo de erro é homoscedástico quando a variância da distribuição da população de salários é a mesma para homens e mulheres; quando essas variâncias são diferentes, o termo de erro é heteroscedástico.

Implicações Matemáticas da Homoscedasticidade

Os estimadores de MQO permanecem não viesados e assintoticamente normais. Como as hipóteses de mínimos quadrados no Conceito-Chave 4.3 não impõem restrições à variância condicional, elas se aplicam tanto ao caso geral de heteroscedasticidade quanto ao caso especial de homoscedasticidade. Portanto, os estimadores de MQO permanecem não viesados e consistentes mesmo que os erros sejam homoscedásticos. Além disso, os estimadores de MQO têm distribuições amostrais que são normais para amostras grandes mesmo que os erros sejam homoscedásticos. Independentemente de os erros serem homoscedásticos ou heteroscedásticos, o estimador de MQO é não viesado, consistente e assintoticamente normal.

Eficiência do estimador de MQO. Se as hipóteses de mínimos quadrados do Conceito-Chave 4.3 são válidas e, além disso, os erros são homoscedásticos, os estimadores de MQO $\hat\beta_0$ e $\hat\beta_1$ são eficientes entre todos os estimadores que são lineares em $Y_1, ..., Y_n$ e são não viesados, condicionais a $X_1, ..., X_n$. Isto é, os estimadores de MQO têm a menor variância de todos os estimadores não viesados que são médias ponderadas de $Y_1, ..., Y_n$. Em outras palavras, se, além das hipóteses de mínimos quadrados, os erros são homoscedásticos, os estimadores de MQO $\hat\beta_0$ e $\hat\beta_1$ são os **melhores estimadores lineares não viesados**, ou **MELNV**.* Esse resultado foi expresso para a média da amostra $\overline{Y}$ no Conceito-Chave 3.3 e se estende para o MQO na presença de homoscedasticidade. Esse resultado, conhecido como teorema de Gauss-Markov, é provado no Capítulo 15.

Se os erros são heteroscedásticos, o MQO não é mais MELNV. Em teoria, se os erros são heteroscedásticos, é possível construir um estimador que tenha uma variância menor do que o estimador de MQO. Esse método é chamado de **mínimos quadrados ponderados**, em que as observações são ponderadas pelo inverso da raiz quadrada da variância condicional de u_i dado X_i. Em razão dessa ponderação, os erros dessa regressão ponderada são homoscedásticos, de modo que o MQO, aplicado a essa regressão ponderada, é MELNV. Embora teoricamente excelente, o problema desse método na prática é que você deve saber como a variância condicional de u_i efetivamente depende de X_i, o que raramente se sabe nas aplicações. Como os mínimos quadrados ponderados têm principalmente interesse teórico, adiamos discussões adicionais para o Capítulo 15.

Fórmula de variância somente homoscedástica. Se o termo de erro é homoscedástico, as fórmulas para as variâncias de $\hat\beta_0$ e $\hat\beta_1$ no Conceito-Chave 4.4 são simplificadas. Conseqüentemente, se os erros são homoscedásticos, uma fórmula especial pode ser utilizada para os erros padrão de $\hat\beta_0$ e $\hat\beta_1$. Essas fórmulas são dadas no Apêndice 4.4. No caso especial em que X é uma variável binária, o estimador da variância de $\hat\beta_1$ na presença de homoscedasticidade (isto é, o quadrado do erro padrão de $\hat\beta_1$ na presença da homoscedasticidade) é a chamada fórmula da variância agrupada para a diferença entre as médias, discutida na nota de rodapé 1 da Seção 3.4.

Como essas fórmulas alternativas são derivadas para o caso especial em que os erros são homoscedásticos e não se aplicam quando os erros são heteroscedásticos, elas serão identificadas como fórmulas "somente homoscedásticas" para a variância e para o erro padrão dos estimadores de MQO. Como o nome sugere, se os erros são heteroscedásticos, os **erros padrão somente homoscedásticos** não são apropriados. Especificamente, se os erros são heteroscedásticos, a estatística t calculada utilizando o erro padrão somente homoscedástico não possui uma distribuição normal padrão nem mesmo para amostras grandes. Na verdade, os valores críticos corretos que são utilizados para essa estatística t somente homoscedástica dependem da natureza precisa da heteroscedasticidade, de modo que esses valores críticos não podem ser tabulados. Do mesmo modo, se os erros são heteroscedásticos, mas um intervalo de confiança é construído como $\pm 1{,}96$ erros padrão somente homoscedásticos, temos que em geral a probabilidade de que esse intervalo contenha o valor verdadeiro do coeficiente não é de 95 por cento, mesmo em amostras grandes.

Em vez disso, como a homoscedasticidade é um caso especial da heteroscedasticidade, os estimadores $\hat\sigma^2_{\hat\beta_1}$ e $\hat\sigma^2_{\hat\beta_0}$ das variâncias de $\hat\beta_1$ e $\hat\beta_0$ dados pelas equações (4.19) e (4.59) geram inferências estatísticas válidas, sejam os erros heteroscedásticos ou homoscedásticos. Desse modo, testes de hipótese e intervalos de confiança baseados nesses erros padrão serão válidos se os erros forem heteroscedásticos ou não. Como os erros padrão que utilizamos até o momento (isto é, aqueles baseados nas equações (4.19) e (4.59)) levam a inferências estatísticas válidas para erros heteroscedásticos ou não, eles são chamados de **erros padrão robustos quanto à heteroscedasticidade**. Como essas fórmulas foram propostas por Eicker (1967), Huber (1967) e White (1980), também são chamadas de erros padrão de Eicker-Huber-White.

O que Isso Significa na Prática?

Qual é mais realista, a heteroscedasticidade ou a homoscedasticidade? A resposta para essa questão depende da aplicação. Contudo, as questões podem ser esclarecidas retornando-se ao exemplo da diferença relativa ao sexo entre salários de indivíduos com curso superior. A familiaridade com a forma pela qual as pessoas

* No original, *best linear unbiased estimators*, ou *Blue*. Em virtude do nome em inglês, um econometrista pode se referir a esse tipo de estimador como "estimador *blue*" ou "estimador azul" (N. do R.T.).

são pagas no mundo ao nosso redor nos dá algumas pistas sobre a hipótese que é mais sensata. Durante muitos anos — e atualmente em uma extensão menor — as mulheres não ocuparam os cargos mais bem remunerados: sempre houve homens com salários baixos, mas raramente houve mulheres com salários altos. Esse fato sugere que a distribuição de salários entre mulheres possui uma dispersão menor do que entre homens. Em outras palavras, é plausível que a variância do termo de erro na Equação (4.42) para as mulheres seja menor do que a variância do termo de erro na Equação (4.43) para homens. Portanto, a presença de um "telhado de vidro" para os cargos e os salários das mulheres sugere que o termo de erro no modelo de regressão com variável binária na Equação (4.41) é heteroscedástico. A menos que existam motivos convincentes para o contrário — e não conseguimos pensar em nenhum —, faz sentido tratar o termo de erro desse exemplo como heteroscedástico.

Como esse exemplo de modelagem de salários ilustra, a heteroscedasticidade aparece em muitas aplicações econométricas. De modo geral, a teoria econômica raramente oferece qualquer motivo para se acreditar que os erros sejam homoscedásticos. É, portanto, prudente supor que sejam heteroscedásticos, a menos que você tenha motivos convincentes para pensar o contrário.

Implicações práticas. A principal questão de relevância prática nesta discussão é: devemos utilizar erros padrão robustos quanto à heteroscedasticidade ou somente homoscedásticos? Para responder a ela, é interessante imaginar o cálculo de ambos para então escolher um entre eles. Se o erro padrão somente homoscedástico e o erro padrão robusto quanto à heteroscedasticidade são iguais, nada se perde com o uso de erros padrão robustos quanto à heteroscedasticidade; se, entretanto, eles são diferentes, você deve utilizar os mais confiáveis que permitam a ocorrência de heteroscedasticidade. O mais simples, então, é sempre utilizar os erros padrão robustos quanto à heteroscedasticidade.

Por razões históricas, muitos programas econométricos utilizam os erros padrão somente homoscedásticos em sua configuração padrão, de modo que fica a critério do usuário especificar a opção erros padrão robustos quanto à heteroscedasticidade. Os detalhes da implementação dos erros padrão robustos quanto à heteroscedasticidade dependem do programa econométrico que você utiliza.

Todos os exemplos empíricos neste livro empregam erros padrão robustos quanto à heteroscedasticidade, exceto quando declarado explicitamente de outra forma.[2]

4.10 Conclusão

Volte por um momento ao problema da superintendente que está considerando a contratação de mais professores para diminuir a razão aluno-professor. De acordo com o que aprendemos, o que ela pode considerar útil?

Nossa análise de regressão — com base em 420 observações da base de dados sobre pontuações nos exames da Califórnia de 1998 — mostrou que há uma relação negativa entre a razão aluno-professor e as pontuações nos exames: diretorias com turmas menores apresentam pontuação maior nos exames. O coeficiente é relativamente grande, em termos práticos: diretorias com dois alunos a menos por professor têm, em média, pontuações nos exames maiores em 4,6 pontos. Isso é o mesmo que dizer que uma diretoria passou do 50º percentil da distribuição de pontuação nos exames para o 60º percentil.

O coeficiente da razão aluno-professor é, em termos estatísticos, significativamente diferente de 0 ao nível de significância de 5 por cento. O coeficiente da população poderia ser 0, e poderíamos simplesmente ter estimado nosso coeficiente negativo por variação de amostragem aleatória. Entretanto, a probabilidade de isso ser feito (e de se obter uma estatística t em β_1 tão grande quanto possível) puramente por variação aleatória ao longo de amostras potenciais é extremamente pequena, de aproximadamente 0,001 por cento. Um intervalo de confiança de 95 por cento para β_1 é $-3{,}30 \leq \beta_1 \leq -1{,}26$.

Fizemos um progresso considerável rumo à resposta da questão da superintendente. Mesmo assim, ainda resta uma preocupação persistente. Estimamos uma relação negativa entre a razão aluno-professor e a pontuação

[2] Caso este livro seja utilizado com outros textos, é interessante observar que alguns livros acrescentam a homoscedasticidade à relação de hipóteses de mínimos quadrados. Como acabamos de discutir, entretanto, essa hipótese adicional não é necessária para a validade da análise de regressão de MQO, desde que erros padrão robustos quanto à heteroscedasticidade sejam utilizados.

nos exames, mas esta é necessariamente a relação *causal* que a superintendente precisa para tomar sua decisão? Constatamos que diretorias com razões aluno-professor menores têm, em média, pontuações maiores nos exames. Mas isso significa que a redução da razão aluno-professor irá aumentar a pontuação?

Há, na verdade, motivos para duvidar de que isso irá acontecer. Contratar mais professores, afinal, custa dinheiro, de modo que diretorias regionais de ensino mais ricas podem custear turmas menores. E os alunos de escolas mais ricas têm outras vantagens em relação aos seus vizinhos mais pobres, incluindo instalações melhores, livros mais novos e professores mais bem pagos. Além disso, alunos de escolas mais ricas geralmente são de famílias abastadas e, portanto, dispõem de outras vantagens que não estão diretamente associadas a sua escola. Por exemplo, a Califórnia tem uma grande comunidade de imigrantes; esses imigrantes tendem a ser mais pobres do que a população em geral e, em muitos casos, seus filhos não falam inglês fluentemente. Assim, pode ser que a relação negativa que estimamos entre a pontuação nos exames e a razão aluno-professor sejam conseqüência da existência de turmas menores aliada a muitos outros fatores que são, na verdade, a verdadeira causa de pontuações menores nos exames.

Esses outros fatores, ou "variáveis omitidas", poderiam indicar que a análise de MQO feita até o momento na verdade tem pouco valor para a superintendente. Ela poderia mesmo ser enganadora: mudar apenas a razão aluno-professor não alteraria esses outros fatores que determinam o desempenho de uma criança na escola. Para tratar desse problema, precisamos de um método que nos permita isolar o efeito da variação na razão aluno-professor sobre a pontuação nos exames, *mantendo esses outros fatores constantes*. Esse método é a análise de regressão múltipla, tema do Capítulo 5.

Resumo

1. A reta de regressão da população, $\beta_0 + \beta_1 X$, é a média de Y como uma função do valor de X. A declividade, β_1, é a variação esperada em Y associada a uma variação unitária em X. O intercepto, β_0, determina o nível (ou altura) da reta de regressão. O Conceito-Chave 4.1 resume a terminologia do modelo de regressão linear da população.
2. A reta de regressão da população pode ser estimada utilizando observações da amostra (Y_i, X_i), $i = 1, ..., n$ por mínimos quadrados ordinários (MQO). Os estimadores de MQO do intercepto e da declividade são representados por $\hat{\beta}_0$ e $\hat{\beta}_1$.
3. Existem três hipóteses principais para o modelo de regressão linear: (1) os erros da regressão, u_i, possuem uma média igual a zero condicional aos regressores X_i; (2) as observações da amostra são seleções aleatórias i.i.d. da população; e (3) as variáveis aleatórias possuem quatro momentos. Se essas hipóteses forem válidas, os estimadores de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ serão (1) não viesados, (2) consistentes e (3) normalmente distribuídos quando a amostra é grande.
4. O teste de hipótese para os coeficientes de regressão é análogo ao teste de hipótese para a média da população: utilize a estatística t para calcular os valores p e aceitar ou rejeitar a hipótese nula. Assim como um intervalo de confiança para a média da população, um intervalo de confiança de 95 por cento para um coeficiente de regressão é calculado como o estimador $\pm 1{,}96$ erros padrão.
5. Quando X é binário, o modelo de regressão pode ser utilizado para estimar e testar hipóteses sobre a diferença entre as médias da população do grupo "$X = 0$" e do grupo "$X = 1$".
6. O R^2 e o erro padrão da regressão (EPR) são medidas de quanto os valores de Y_i estão próximos da reta de regressão estimada. O R^2 situa-se entre zero e um; um valor maior indica que os Y_i estão mais próximos da reta. O erro padrão da regressão é um estimador do desvio padrão do erro da regressão.
7. Em geral, o erro u_i é heteroscedástico, isto é, a variância de u_i para um dado valor de X_i, $\text{var}(u_i | X_i = x)$ depende de x. Um caso especial ocorre quando o erro é homoscedástico, isto é, $\text{var}(u_i | X_i = x)$ é constante. Erros padrão somente homoscedásticos não geram inferências estatísticas válidas quando os erros são heteroscedásticos, mas erros padrão robustos quanto à heteroscedasticidade geram-nas.

Termos-chave

Revisão dos Conceitos

4.1 Explique a diferença entre $\hat{\beta}_1$ e β_1; entre o resíduo $\hat{u}_i$ e o erro da regressão u_i; e entre o valor previsto de MQO $\hat{Y}_i$ e $E(Y_i|X_i)$.

4.2 Exponha os procedimentos para o cálculo do valor p de um teste bicaudal de H_0: $\mu_Y = 0$ utilizando um conjunto de observações i.i.d. Y_i, $i = 1, ..., n$. Exponha os procedimentos para o cálculo do valor p de um teste bicaudal de H_0: $\beta_1 = 0$ em um modelo de regressão utilizando um conjunto de observações i.i.d. (Y_i, X_i), $i = 1, ..., n$.

4.3 Explique como você poderia utilizar um modelo de regressão para estimar a diferença relativa ao sexo entre salários por meio do uso de dados da Seção 3.5. Quais são as variáveis dependentes e as independentes?

4.4 Esboce um gráfico de dispersão com dados hipotéticos para uma regressão estimada com $R^2 = 0{,}9$. Esboce um gráfico de dispersão com dados hipotéticos para uma regressão com $R^2 = 0{,}5$.

Exercícios

As soluções para os exercícios marcados com * podem ser encontradas, em inglês, no site relativo ao livro em **www.aw.com/stock_br**.

***4.1** Suponha que um pesquisador, utilizando dados sobre o tamanho da turma (*TT*) e a pontuação média nos exames de 100 turmas do terceiro ano, estime a regressão de MQO,

$$\widehat{PontExame} = 520{,}4 - 5{,}82 \times TT,\ R^2 = 0{,}08,\ EPR = 11{,}5.$$
$$(20{,}4) \quad (2{,}21)$$

a. Uma turma tem 22 alunos. Qual é a previsão da regressão para a pontuação média nos exames dessa turma?

b. No ano passado uma turma tinha 19 alunos e neste ano tem 23. Qual é a previsão da regressão para a variação na pontuação média nos exames da turma?

c. Construa um intervalo de confiança de 95 por cento para β_1, o coeficiente de declividade da regressão.

d. Calcule o valor p para o teste bicaudal da hipótese nula H_0: $\beta_1 = 0$. Você rejeita a hipótese nula ao nível de 5 por cento? E ao nível de 1 por cento?

e. A média da amostra do tamanho da turma entre as 100 turmas é 21,4. Qual é a média da amostra da pontuação nos exames entre as 100 turmas? (*Dica*: Faça uma revisão das fórmulas para os estimadores de MQO.)

f. Qual é o desvio padrão da amostra da pontuação nos exames para as 100 turmas? (*Dica*: Faça uma revisão das fórmulas para R^2 e *EPR*.)

4.2 Suponha que um pesquisador, utilizando dados sobre salários de 250 trabalhadores e 280 trabalhadoras selecionados aleatoriamente, estime a regressão de MQO,

$$\widehat{Salário} = 12{,}68 + 2{,}79\ Homem,\ R^2 = 0{,}06,\ EPR = 3{,}10$$
$$(0{,}18) \quad (0{,}84)$$

onde *Salário* é medido em US$/hora e *Homem* é uma variável binária igual a um se a pessoa for do sexo masculino e 0 se a pessoa for do sexo feminino. Defina a diferença relativa ao sexo entre salários como a diferença entre o salário médio de homens e de mulheres.

a. Qual é a diferença relativa ao sexo estimada?

b. Será que a diferença relativa ao sexo estimada é significativamente diferente de zero? (Calcule o valor p para o teste de hipótese nula de que não há diferença relativa ao sexo.)

c. Construa um intervalo de confiança de 95 por cento para a diferença relativa ao sexo.

d. Na amostra, qual é o salário médio das mulheres? E dos homens?

e. Outro pesquisador utiliza esses mesmos dados, mas regride *Salário* sobre *Mulher*, uma variável igual a 1 se a pessoa for do sexo feminino e 0 se a pessoa for do sexo masculino. Quais são as estimativas da regressão calculadas a partir dessa regressão?

$$\widehat{Salário} = \_\_\_\_ + \_\_\_\_\ Mulher,\ R^2 = \_\_\_\_,\ EPR = \_\_\_\_.$$

***4.3** Mostre que a primeira hipótese dos mínimos quadrados, $E(u_i|X_i) = 0$, implica que $E(Y_i|X_i) = \beta_0 + \beta_1 X_i$.

4.4 Mostre que $\hat{\beta}_0$ é um estimador não viesado de β_0. (*Dica*: Use o fato de que $\hat{\beta}_1$ é não viesado, conforme o Apêndice 4.3.)

4.5 Suponha que uma amostra aleatória de 200 homens com 20 anos de idade seja selecionada de uma população e sua altura e seu peso sejam registrados. Uma regressão de peso sobre altura resulta em:

$$\widehat{Peso} = -99{,}41 + 3{,}94\ Altura,\ R^2 = 0{,}81,\ EPR = 10{,}2,$$
$$(2{,}15) \quad (0{,}31)$$

onde *Peso* é medido em libras e *Altura* é medida em polegadas.

a. Qual é a previsão de peso da regressão para uma pessoa com 70 polegadas de altura? E com 65 polegadas? E com 74 polegadas?

b. Uma pessoa tem um surto tardio de crescimento e cresce 1,5 polegada ao longo de um ano. Qual é a previsão da regressão para o aumento de peso da pessoa?

c. Construa um intervalo de confiança de 99 por cento para o aumento de peso em (b).

d. Suponha que, em vez de medir o peso em libras e a altura em polegadas, eles sejam medidos em quilogramas e centímetros. Quais são as estimativas de regressão da nova regressão, quilograma-centímetro? (Dê todos os resultados, coeficientes estimados, erros padrão, R^2 e EPR.)

4.6 A partir da Equação (4.15), derive a variância de $\hat{\beta}_0$ na presença da homoscedasticidade dada na Equação (4.61) do Apêndice 4.4.

APÊNDICE 4.1 Base de Dados de Pontuação nos Exames da Califórnia

A base de dados California Standardized Testing and Reporting contém dados sobre o desempenho dos alunos nos exames, as características da escola e o histórico demográfico dos alunos. Os dados utilizados aqui são de todas as 420 diretorias K-6 e K-8 na Califórnia com dados disponíveis de 1998 e 1999. A pontuação nos exames é a média da pontuação nos exames de leitura e matemática do Stanford 9 Achievement Test, um exame nacional aplicado a alunos da quinta série. As características da escola (medidas como uma média na diretoria) incluem matrícula, número de professores (medido como "equivalentes em dedicação integral"), número de computadores por sala de aula e gastos por aluno. A razão aluno-professor utilizada aqui é o número equivalente de professores com dedicação integral na diretoria dividido pelo número de alunos. Variáveis demográficas relativas aos alunos também são medidas como uma média na diretoria. As variáveis demográficas incluem a porcentagem de alunos que estão incluídos no programa público de assistência CalWorks (anteriormente AFDC), a porcentagem de alunos que tem direito a almoço subsidiado, a porcentagem de alunos que está aprendendo inglês (isto é, alunos para os quais o inglês é o segundo idioma). Todos esses dados foram obtidos do Departamento de Educação da Califórnia (California Department of Education) (www.cde.ca.gov).

APÊNDICE 4.2 Derivação dos Estimadores de MQO

Este apêndice utiliza cálculo para derivar as fórmulas para os estimadores de MQO dados no Conceito-Chave 4.2. Para minimizar a soma dos quadrados dos erros de previsão $\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2$ (veja a Equação (4.6)), em primeiro lugar tome as derivadas parciais com relação a b_0 e b_1:

$$\frac{\partial}{\partial b_0}\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2 = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i) \text{ e} \tag{4.44}$$

$$\frac{\partial}{\partial b_1}\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2 = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)X_i. \tag{4.45}$$

Os estimadores de MQO, $\hat{\beta}_0$ e $\hat{\beta}_1$, são os valores de b_0 e b_1 que minimizam $\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2$ ou, de forma equivalente, os valores de b_0 e b_1 para os quais as derivadas das equações (4.44) e (4.45) são iguais a zero. Desse modo, fazer essas derivadas iguais a zero, agrupar os termos e dividi-los por n mostra que os estimadores de MQO, $\hat{\beta}_0$ e $\hat{\beta}_1$, devem satisfazer as duas equações,

$$\overline{Y} - \hat{\beta}_0 - \hat{\beta}_1\overline{X} = 0 \text{ e} \tag{4.46}$$

$$\frac{1}{n}\sum_{i=1}^{n}X_iY_i - \hat{\beta}_0\overline{X} - \hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n}X_i^2 = 0. \tag{4.47}$$

Resolvendo esse par de equações para $\hat{\beta}_0$ e $\hat{\beta}_1$, temos

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}X_iY_i - \overline{X}\,\overline{Y}}{\frac{1}{n}\sum_{i=1}^{n}X_i^2 - (\overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{4.48}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.49}$$

As equações (4.48) e (4.49) são fórmulas para $\hat{\beta}_0$ e $\hat{\beta}_1$ dadas no Conceito-Chave 4.2; a fórmula $\hat{\beta}_1 = s_{XY}/s_X^2$ é obtida ao se dividir o numerador e o denominador na Equação (4.48) por $n - 1$.

APÊNDICE 4.3 Distribuição Amostral do Estimador de MQO

Neste apêndice, mostramos que o estimador de MQO $\hat{\beta}_1$ é não viesado e que, para amostras grandes, tem uma distribuição amostral normal dada no Conceito-Chave 4.4.

Representação de $\hat{\beta}_1$ em Termos de Regressores e Erros

Começamos fornecendo uma expressão para $\hat{\beta}_1$ em termos de regressores e erros. Como $Y_i = \beta_0 + \beta_1X_i + u_i$, $Y_i - \overline{Y} = \beta_1(X_i - \overline{X}) + u_i - \overline{u}$, de modo que o numerador da fórmula para $\hat{\beta}_1$ na Equação (4.48) é

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) &= \sum_{i=1}^{n}(X_i - \overline{X})[\beta_1(X_i - \overline{X}) + (u_i - \overline{u})] \\ &= \beta_1\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(X_i - \overline{X})(u_i - \overline{u}).\end{aligned} \tag{4.50}$$

Agora $\sum_{i=1}^{n}(X_i - \overline{X})(u_i - \overline{u}) = \sum_{i=1}^{n}(X_i - \overline{X})u_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{u} = \sum_{i=1}^{n}(X_i - \overline{X})u_i$, onde a igualdade final resulta da definição de $\overline{X}$, que implica que $\sum_{i=1}^{n}(X_i - \overline{X})\overline{u} = \left[\sum_{i=1}^{n}X_i - n\overline{X}\right]\overline{u} = 0$. Substituindo $\sum_{i=1}^{n}(X_i - \overline{X})(u_i - \overline{u}) = \sum_{i=1}^{n}(X_i - \overline{X})u_i$ na expressão final da Equação (4.50), temos $\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \beta_1\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(X_i - \overline{X})u_i$. A substituição dessa expressão, por sua vez, na fórmula para $\hat{\beta}_1$ na Equação (4.48) gera

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{4.51}$$

Prova de que $\hat{\beta}_1$ É Não Viesado

A expectativa de $\hat{\beta}_1$ é obtida tomando-se a expectativa dos dois lados da Equação (4.51). Portanto,

$$\begin{aligned}E(\hat{\beta}_1) &= \beta_1 + E\left[\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}\right] \\ &= \beta_1 + E\left[\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})E(u_i|X_1, \ldots, X_n)}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}\right] = \beta_1,\end{aligned} \tag{4.52}$$

onde a segunda igualdade da Equação (4.52) resulta do uso da lei de expectativas iteradas (veja a Seção 2.3). Pela segunda hipótese dos mínimos quadrados, u_i é distribuído independentemente de X para todas as observações diferentes de i, de modo que $E(u_i|X_1, \ldots, X_n) = E(u_i|X_i)$. Pela primeira hipótese dos mínimos quadrados, no entanto, $E(u_i|X_i) = 0$. Assim, o numerador do último termo na Equação (4.52) é zero, de modo que $E(\hat{\beta}_1) = \beta_1$, isto é, o estimador de MQO é não viesado.

Distribuição Normal para Amostras Grandes do Estimador de MQO

A aproximação normal para amostras grandes da distribuição limite de $\hat{\beta}_1$ (veja o Conceito-Chave 4.4) é obtida considerando-se o comportamento do último termo na Equação (4.51).

Em primeiro lugar, considere o numerador desse termo. Como $\overline{X}$ é consistente, se o tamanho da amostra for grande, $\overline{X}$ será aproximadamente igual a μ_X. Portanto, para uma boa aproximação, o termo no numerador da Equação (4.51) é a média de amostra $\overline{v}$, em que $v_i = (X_i - \mu_X)u_i$. Pela primeira hipótese dos mínimos quadrados, v_i tem uma média igual a zero. Pela segunda hipótese dos mínimos quadrados, v_i é i.i.d. A variância de v_i é $\sigma_v^2 = \text{var}[(X_i - \mu_X)u_i]$, que, pela terceira hipótese dos mínimos quadrados, é diferente de zero e finita. Assim, $\overline{v}$ satisfaz todos os requisitos do teorema central do limite (veja o Conceito-Chave 2.7). Portanto, $\overline{v}/\sigma_{\overline{v}}$ é, para amostras grandes, distribuído como $N(0, 1)$, onde $\sigma_{\overline{v}}^2 = \sigma_v^2/n$. Desse modo, a distribuição de $\overline{v}$ tem uma boa aproximação pela distribuição $N(0, \sigma_v^2/n)$.

A seguir, considere a expressão no denominador da Equação (4.51); essa é a variância da amostra de X (exceto pela divisão por n, e não por $n-1$, que não trará conseqüências se n for grande). Conforme discutido na Seção 3.2 (veja a Equação (3.8)), a variância da amostra é um estimador consistente da variância da população, de modo que, para amostras grandes, ela está arbitrariamente próxima da variância da população de X.

Combinando esses dois resultados, temos que, para amostras grandes, $\hat{\beta}_1 - \beta_1 \cong \overline{v}/\text{var}(X_i)$, de modo que a distribuição amostral de $\hat{\beta}_1$ é, para amostras grandes, $N(\beta_1, \sigma^2_{\hat{\beta}_1})$, onde $\sigma^2_{\hat{\beta}_1} = \text{var}(\overline{v})/[\text{var}(X_i)]^2 = \text{var}[(X_i - \mu_X)u_i]/\{n[\text{var}(X_i)]^2\}$, que é a expressão da Equação (4.14).

Alguns Fatos Algébricos Adicionais sobre MQO

Os resíduos e valores previstos de MQO satisfazem:

$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0, \tag{4.53}$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i = \overline{Y}, \tag{4.54}$$

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0 \text{ e } s_{\hat{u}X} = 0 \text{ e} \tag{4.55}$$

$$SQT = SQR + SQE. \tag{4.56}$$

As equações (4.53) a (4.56) dizem que a média da amostra dos resíduos de MQO é zero; a média da amostra dos valores previstos de MQO é igual a $\overline{Y}$; a co-variância da amostra $s_{\hat{u}X}$ entre os resíduos de MQO e os regressores é zero; e a soma dos quadrados total é a soma da soma dos quadrados dos resíduos com a soma dos quadrados explicada (SQE, SQT e SQR são definidas nas equações (4.35), (4.36) e (4.38)).

Para verificar a Equação (4.53), observe que a definição de $\hat{\beta}_0$ nos permite escrever os resíduos de MQO como $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = (Y_i - \overline{Y}) - \hat{\beta}_1(X_i - \overline{X})$; portanto,

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} (Y_i - \overline{Y}) - \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \overline{X}).$$

Porém, as definições de $\overline{Y}$ e $\overline{X}$ implicam que $\sum_{i=1}^{n}(Y_i - \overline{Y}) = 0$ e $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$, de modo que $\sum_{i=1}^{n}\hat{u}_i = 0$.

Para verificar a Equação (4.54), observe que $Y_i = \hat{Y}_i + \hat{u}_i$, de modo que $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i + \sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} \hat{Y}_i$, onde a segunda igualdade é uma conseqüência da Equação (4.53).

Para verificar a Equação (4.55), observe que $\sum_{i=1}^{n} \hat{u}_i = 0$ implica que $\sum_{i=1}^{n} \hat{u}_i X_i = \sum_{i=1}^{n} \hat{u}_i (X_i - \overline{X})$, de modo que

$$\begin{aligned}\sum_{i=1}^{n} \hat{u}_i X_i &= \sum_{i=1}^{n} [(Y_i - \overline{Y}) - \hat{\beta}_1(X_i - \overline{X})]\,(X_i - \overline{X}) \\ &= \sum_{i=1}^{n} (Y_i - \overline{Y})(X_i - \overline{X}) - \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \overline{X})^2 = 0,\end{aligned} \tag{4.57}$$

onde a última igualdade da Equação (4.57) é obtida pelo uso da fórmula para $\hat{\beta}_1$ da Equação (4.48). Esse resultado, combinado com os resultados anteriores, indica que $s_{\hat{u}X} = 0$.

A Equação (4.56) é deduzida dos resultados anteriores e de um pouco de álgebra:

$$\begin{aligned} SQT &= \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i + \hat{Y}_i - \overline{Y})^2 \\ &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) \\ &= SQR + SQE + 2\sum_{i=1}^{n}\hat{u}_i\hat{Y}_i = SQR + SQE, \end{aligned} \tag{4.58}$$

onde a última igualdade resulta de $\sum_{i=1}^{n}\hat{u}_i\hat{Y}_i = \sum_{i=1}^{n}\hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \hat{\beta}_0\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i X_i = 0$ pelos resultados anteriores.

APÊNDICE 4.4 Fórmulas para os Erros Padrão de MQO

Este apêndice discute as fórmulas para os erros padrão de MQO. Estes são apresentados primeiramente sob as hipóteses de mínimos quadrados do Conceito-Chave 4.3, que permite a heteroscedasticidade: são os erros padrão robustos quanto à heteroscedasticidade. As fórmulas para a variância dos estimadores de MQO e para os erros padrão associados são, então, fornecidas para o caso especial de homoscedasticidade.

Erros Padrão Robustos Quanto à Heteroscedasticidade

O estimador $\hat{\sigma}^2_{\hat{\beta}_1}$ definido na Equação (4.19) é obtido substituindo-se as variâncias da população na Equação (4.14) pelas variâncias da amostra correspondentes, com uma modificação. A variância no numerador da Equação (4.14) é estimada por $\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \overline{X})^2\hat{u}_i^2$, onde o divisor $n-2$ (em vez de n) incorpora um ajuste de graus de liberdade para o viés para baixo, de forma análoga ao ajuste de graus de liberdade utilizado na definição do EPR na Seção 4.8. A variância no denominador é estimada por $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. A substituição de $\text{var}[(X_i - \mu_X)u_i]$ e $\text{var}(X_i)$ na Equação (4.14) por esses dois estimadores produz $\hat{\sigma}^2_{\hat{\beta}_1}$ na Equação (4.19). A consistência dos erros padrão robustos quanto à heteroscedasticidade é discutida na Seção 15.3.

O estimador da variância de $\hat{\beta}_0$ é

$$\hat{\sigma}^2_{\hat{\beta}_0} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{H}_i^2\hat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{H}_i^2\right)^2}, \tag{4.59}$$

onde $\hat{H}_i = 1 - [\overline{X}/\frac{1}{n}\sum_{i=1}^{n}X_i^2]X_i$. O erro padrão de $\hat{\beta}_0$ é $EP(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_0}}$. O raciocínio por trás do estimador $\hat{\sigma}^2_{\hat{\beta}_0}$ é o mesmo que está por trás de $\hat{\sigma}^2_{\hat{\beta}_1}$ e resulta da substituição das expectativas da população por médias da amostra.

Variâncias Somente Homoscedásticas

Na presença de homoscedasticidade, a variância condicional de u_i dado X_i é uma constante, isto é, $\text{var}(u_i|X_i) = \sigma_u^2$. Se os erros são homoscedásticos, as fórmulas do Conceito-Chave 4.4 são simplificadas para

$$\sigma^2_{\hat{\beta}_1} = \frac{\sigma_u^2}{n\sigma_X^2} \text{ e} \tag{4.60}$$

$$\sigma^2_{\hat{\beta}_0} = \frac{E(X_i^2)}{n\sigma_X^2}\sigma_u^2. \tag{4.61}$$

Para derivar a Equação (4.60), escreva o numerador da Equação (4.14) como $\text{var}[(X_i - \mu_X)u_i] = E(\{(X_i - \mu_X)u_i - E[(X_i - \mu_X)u_i]\}^2) = E\{[(X_i - \mu_X)u_i]^2\} = E[(X_i - \mu_X)^2 u_i^2] = E[(X_i - \mu_X)^2 \text{var}(u_i \mid X_i)]$, onde a segunda igualdade segue — uma vez que $E[(X_i - \mu_X)u_i] = 0$ (pela primeira hipótese dos mínimos quadrados) — e onde a última igualdade resulta da lei de expectativas iteradas (veja a Seção 2.3). Se u_i for homoscedástico, então $\text{var}(u_i \mid X_i) = \sigma_u^2$, de modo que $E[(X_i - \mu_X)^2 \text{var}(u_i \mid X_i)] = \sigma_u^2 E[(X_i - \mu_X)^2] = \sigma_u^2 \sigma_X^2$. O resultado da Equação (4.60) segue pela substituição dessa expressão no numerador da Equação (4.14) e da simplificação. Um cálculo semelhante gera a Equação (4.61).

Erros Padrão Somente Homoscedásticos

Os erros padrão somente homoscedásticos são obtidos pela substituição de médias e variâncias da população por médias e variâncias da amostra nas equações (4.60) e (4.61) e pela estimação da variância de u_i pelo quadrado do *EPR*. Os estimadores somente homoscedásticos dessas variâncias são

$$\tilde{\sigma}_{\hat{\beta}_1}^2 = \frac{s_{\hat{u}}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ (somente homoscedásticos) e} \tag{4.62}$$

$$\tilde{\sigma}_{\hat{\beta}_0}^2 = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)s_{\hat{u}}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ (somente homoscedásticos),} \tag{4.63}$$

onde $s_{\hat{u}}^2$ é dado na Equação (4.40). Os erros padrão somente homoscedásticos são a raiz quadrada de $\tilde{\sigma}_{\hat{\beta}_0}^2$ e $\tilde{\sigma}_{\hat{\beta}_1}^2$.