Prefácio

A disciplina econometria pode ser divertida tanto para o professor quanto para o aluno. O mundo real da economia, dos negócios e do governo é complicado e confuso, repleto de idéias que competem umas com as outras e de perguntas que exigem respostas. O que é mais eficaz para enfrentar motoristas bêbados: aprovar leis rígidas ou aumentar o imposto sobre bebidas alcoólicas? Você consegue ganhar dinheiro na bolsa de valores comprando quando os preços estão historicamente baixos em relação aos ganhos ou deve apenas permanecer na mesma posição como sugere a teoria do passeio aleatório dos preços das ações? Podemos melhorar o ensino fundamental reduzindo o tamanho das turmas ou devemos simplesmente fazer com que as crianças ouçam Mozart dez minutos por dia? A econometria nos ajuda a separar as idéias coerentes das absurdas e a encontrar respostas quantitativas para questões quantitativas importantes. A econometria abre uma janela em nosso mundo complexo que nos permite enxergar as relações em que as pessoas, as empresas e os governos baseiam suas decisões.

Este livro destina-se ao ensino de econometria em nível de graduação. A nossa experiência é de que, para tornar a econometria relevante, as aplicações interessantes devem motivar a teoria e a teoria deve ser apropriada para as aplicações. Esse princípio simples representa uma mudança significativa em relação à antiga geração de livros de econometria, em que os modelos teóricos e as hipóteses não eram apropriados para as aplicações. Não é de surpreender que alguns estudantes questionem a relevância da econometria, depois de ter gasto parte de seu tempo aprendendo hipóteses que posteriormente entendem como irrealistas, precisando então aprender "soluções" para "problemas" que surgem quando as aplicações não são apropriadas para as hipóteses. Consideramos que é muito melhor motivar a necessidade de ferramentas com uma aplicação concreta e, então, fornecer algumas hipóteses simples que sejam apropriadas para a aplicação. Como a teoria é relevante de imediato para as aplicações, este enfoque pode fazer com que a econometria tenha vida.

Características Deste Livro

Este livro difere dos demais em três aspectos principais. Primeiro, integramos questões e dados do mundo real ao desenvolvimento da teoria e tratamos com seriedade das descobertas importantes da análise empírica resultante. Segundo, nossa escolha de tópicos reflete a teoria e a prática modernas. Terceiro, fornecemos teorias e hipóteses que são apropriadas para as aplicações. Nosso objetivo é ensinar os estudantes a se tornarem consumidores sofisticados de econometria e a fazer isso em um nível de matemática apropriado para uma disciplina introdutória.

Dados e Questões do Mundo Real

Organizamos cada tópico metodológico em torno de uma questão importante do mundo real que exige uma resposta numérica específica. Por exemplo, ensinamos a regressão com uma variável, a regressão múltipla e a análise de forma funcional no contexto da estimação do efeito de insumos escolares sobre os produtos escolares. (Será que turmas menores no ensino fundamental geram notas mais altas nos exames?) Ensinamos métodos para dados de painel no contexto da análise do efeito das leis relativas a motoristas bêbados sobre os acidentes fatais de trânsito. Usamos a possível discriminação racial no mercado de empréstimos hipotecários como aplicação empírica para ensinar a regressão com uma variável dependente binária (logit e probit). Ensinamos a estimação de variáveis instrumentais no contexto da obtenção da elasticidade da demanda por cigarros. Embora esses exemplos envolvam raciocínio econômico, todos podem ser compreendidos tendo-se estudado apenas uma disciplina introdutória da área de economia, e muitos podem ser entendidos sem que se tenha estudado nenhuma disci-

plina da área de economia. Portanto, o professor pode ter o foco no ensino da econometria, não da microeconomia ou da macroeconomia.

Tratamos todas as nossas aplicações empíricas com seriedade e de tal modo que mostremos aos alunos como podem aprender com os dados, mas, ao mesmo tempo, com autocrítica e cientes das limitações das análises empíricas. Por meio de cada aplicação, ensinamos os alunos a explorar especificações alternativas e, assim, avaliar se suas descobertas principais são robustas. As questões propostas nas aplicações empíricas são importantes, e fornecemos respostas sérias e, imaginamos, verossímeis. Entretanto, encorajamos os professores e os alunos a discordarem, convidando-os a fazer uma nova análise dos dados disponíveis no site da Internet referente ao livro **(www.aw.com/stock_br).**

Escolha de Tópicos Contemporâneos

A econometria percorreu um longo caminho nas últimas duas décadas. Os tópicos que abordamos refletem o melhor da econometria aplicada contemporânea. Pode-se fazer isso apenas em uma disciplina introdutória, portanto enfocamos procedimentos e testes comumente usados na prática. Por exemplo:

- ***Regressão de variáveis instrumentais.*** Apresentamos a regressão de variáveis instrumentais como um método geral para lidar com a correlação entre o termo de erro e um regressor, que pode ocorrer por muitas razões, inclusive a causalidade simultânea. As duas hipóteses para um instrumento válido — exogeneidade e relevância — recebem a mesma importância. Segue-se a esta apresentação uma extensa discussão sobre a origem desses instrumentos.
- ***Avaliação de programas.*** Um número crescente de estudos econométricos analisa experimentos controlados aleatórios ou quase-experimentos, também conhecidos como experimentos naturais. Tratamos desses tópicos — freqüentemente chamados em conjunto de avaliação de programas — no Capítulo 11. Apresentamos essa estratégia de pesquisa como um enfoque alternativo aos problemas de variáveis omitidas, de causalidade simultânea e de seleção e avaliamos os pontos fortes e fracos de estudos que utilizam dados experimentais ou quase-experimentais.
- ***Previsão.*** O capítulo sobre previsão (Capítulo 12) considera previsões univariadas (auto-regressivas) ou multivariadas que utilizam a regressão de séries temporais, e não os grandes modelos estruturais com equações simultâneas. Enfocamos ferramentas simples e confiáveis, tais como auto-regressões e seleção de modelos por meio de um critério de informação, que funcionam bem na prática. Esse capítulo também apresenta um enfoque com sentido prático para tendências estocásticas (raízes unitárias), testes de raiz unitária, testes de quebras estruturais (com dados conhecidos e desconhecidos) e pseudoprevisão fora da amostra, todos no contexto do desenvolvimento de modelos de previsão de séries temporais estáveis e confiáveis.
- ***Regressão de séries temporais.*** Fazemos uma clara distinção entre duas aplicações muito diferentes da regressão de séries temporais: previsão e estimação de efeitos causais dinâmicos. O capítulo sobre inferências causais utilizando dados de séries temporais (Capítulo 13) examina de forma mais cuidadosa quando os métodos de estimação diferentes, incluindo mínimos quadrados generalizados, levarão ou não a inferências causais válidas e quando é aconselhável estimar regressões dinâmicas usando MQO com erros-padrão consistentes quanto à heteroscedasticidade e à autocorrelação.

Teoria Apropriada para as Aplicações

Embora as ferramentas econométricas sejam mais bem motivadas por aplicações empíricas, os alunos precisam aprender o suficiente de teoria econométrica para entender os pontos fortes e as limitações dessas ferramentas. Fornecemos um tratamento moderno por meio do qual o ajuste entre teoria e aplicações é o mais perfeito possível, mantendo a matemática em um nível que requer apenas álgebra.

As aplicações empíricas modernas compartilham algumas características comuns: as bases de dados são geralmente grandes (centenas de observações, freqüentemente mais); os regressores não são fixos em amostras repetidas, mas, em vez disso, são coletados por amostragem aleatória (ou algum outro mecanismo que os torne aleatórios); os dados não são normalmente distribuídos; e não há uma razão *a priori* para pensar que os erros são homoscedásticos (embora freqüentemente haja razões para se pensar que são heteroscedásticos).

Essas observações levam a diferenças importantes entre o desenvolvimento teórico neste livro e nos demais.

- ***Enfoque de amostras grandes.*** Como as bases de dados são grandes, utilizamos desde o princípio aproximações normais de amostras grandes para as distribuições amostrais nos testes de hipótese e nos intervalos de confiança. Nossa experiência é de que leva menos tempo para ensinar os rudimentos das aproximações de amostras grandes do que ensinar as distribuições F exatas e t de Student, correções para os graus de liberdade e assim por diante. Este enfoque de amostras grandes poupa os estudantes da frustração de descobrir que, devido aos erros não normais, a teoria da distribuição exata que acabaram de dominar é irrelevante. Uma vez que tenha sido ensinado sob o contexto de média amostral, o enfoque de amostras grandes para testes de hipótese e intervalos de confiança estende-se diretamente para a análise de regressão múltipla, logit e probit, estimação de variáveis instrumentais e métodos de séries temporais.
- ***Amostragem aleatória.*** Como os regressores raramente são fixos em aplicações econométricas, desde o princípio tratamos os dados sobre todas as variáveis (dependentes e independentes) como resultado da amostragem aleatória. Essa hipótese é apropriada para nossas aplicações iniciais e para dados de corte, estende-se facilmente para dados de painel e para dados de séries temporais e, devido ao nosso enfoque de amostras grandes, não impõe dificuldades conceituais ou matemáticas adicionais.
- ***Heteroscedasticidade.*** Os econometristas práticos usam rotineiramente erros padrão robustos quanto à heteroscedasticidade para eliminar dúvidas quanto à presença ou não de heteroscedasticidade. Neste livro, vamos além da consideração da heteroscedasticidade como uma exceção ou um "problema" a ser "resolvido"; em vez disso, permitimos a heteroscedasticidade desde o princípio e simplesmente utilizamos erros padrão robustos quanto à heteroscedasticidade. Apresentamos a homoscedasticidade como um caso especial que fornece uma motivação teórica para MQO.

Produtores Qualificados, Consumidores Sofisticados

Esperamos que os alunos que utilizam este livro tornem-se consumidores sofisticados de análise empírica. Para isso, precisam aprender não só como utilizar as ferramentas da análise de regressão, mas também como avaliar a validade das análises empíricas que lhes são apresentadas.

Nosso processo para ensinarmos como avaliar um estudo empírico ocorre em três etapas. Em primeiro lugar, imediatamente após apresentarmos as principais ferramentas da análise de regressão, dedicamos o Capítulo 7 às ameaças à validade interna e externa de um estudo empírico. Esse capítulo discute problemas com dados e questões referentes à generalização das descobertas para outros cenários. Também examina as principais ameaças à análise de regressão, incluindo variáveis omitidas, especificação incorreta da forma funcional, erros nas variáveis, seleção e simultaneidade — além de formas para reconhecer essas ameaças na prática.

Depois, aplicamos esses métodos para avaliação de estudos empíricos à análise empírica dos exemplos no decorrer do livro. Fazemos isso considerando especificações alternativas e tratando sistematicamente as várias ameaças à validade das análises apresentadas no livro.

Por fim, para se tornarem consumidores sofisticados, os alunos precisam de experiência direta como produtores. O aprendizado ativo supera o aprendizado passivo, e a econometria é uma disciplina ideal para um aprendizado ativo.

Enfoque Matemático e Nível de Rigor

Nosso objetivo é que os estudantes desenvolvam um entendimento sofisticado das ferramentas da análise de regressão moderna, disciplina seja ensinada utilizando-se um nível matemático "alto" ou "baixo". As partes I-IV do texto (que cobrem o material fundamental) são acessíveis a alunos que tenham estudado apenas a matemática

anterior ao curso de cálculo. As partes I-IV contêm menos equações e mais aplicações em relação a muitos livros introdutórios de econometria e um número muito menor de equações do que os livros voltados para matemática nas disciplinas de graduação. Entretanto, mais equações não implicam um tratamento mais sofisticado. Segundo nossa experiência, no caso da maioria dos estudantes, um tratamento mais matemático não leva a um entendimento mais profundo.

Tendo em vista isso, alunos diferentes aprendem de forma diferente, e, para aqueles com um bom conhecimento de matemática, o aprendizado pode ser aprimorado por um tratamento matemático mais explícito. Assim, a Parte V contém uma introdução à teoria econométrica que é apropriada para os estudantes com maior embasamento matemático. Acreditamos que, quando os capítulos sobre matemática na Parte V são usados em conjunto com o material das partes I-IV, este livro torna-se adequado para disciplinas de econometria, tanto avançadas da graduação quanto em nível de mestrado.

Conteúdo e Organização

O livro é composto de cinco partes, e consideramos que o aluno tenha cursado uma disciplina das áreas de probabilidade e estatística, embora revisemos o material na Parte I. Cobrimos o material fundamental da análise de regressão na Parte II. As partes III, IV e V apresentam tópicos adicionais que aprofundam o tratamento fundamental da Parte II.

Parte I

O Capítulo 1 introduz a econometria e enfatiza a importância de fornecer respostas quantitativas a perguntas quantitativas. Discute o conceito de causalidade nos estudos estatísticos e faz uma resenha dos tipos diferentes de dados encontrados na econometria. O material de probabilidade e estatística é revisado nos capítulos 2 e 3, respectivamente; dependendo da formação dos alunos, esses capítulos podem ser dados em uma determinada disciplina ou fornecidos apenas como referência.

Parte II

O Capítulo 4 introduz a regressão com um único regressor e os mínimos quadrados ordinários (MQO). No Capítulo 5, os alunos aprendem a lidar com o viés de variável omitida utilizando a regressão múltipla, estimando, desse modo, o efeito de uma variável independente e mantendo as demais variáveis independentes constantes. No Capítulo 6, os métodos de regressão múltipla são estendidos para modelos com funções de regressão da população não-lineares que são lineares nos parâmetros (de modo que podem ser estimados por MQO). No Capítulo 7, os alunos voltam um pouco atrás e aprendem a identificar os pontos fortes e as limitações dos estudos de regressão, vendo nesse processo como aplicar os conceitos de validade interna e externa.

Parte III

A Parte III apresenta extensões dos métodos de regressão. No Capítulo 8, os alunos aprendem a usar dados de painel para controlar variáveis não observadas que são constantes ao longo do tempo. O Capítulo 9 trata da regressão com uma variável dependente binária. O Capítulo 10 mostra como a regressão de variáveis instrumentais pode ser usada para tratar uma variedade de problemas que produzem correlação entre o termo de erro e o regressor e como se podem encontrar e avaliar instrumentos válidos. O Capítulo 11 apresenta aos alunos a análise de dados de experimentos e quase-experimentos (ou experimentos naturais), tópicos freqüentemente chamados de "avaliação de programas".

Parte IV

A Parte IV examina a regressão com dados de séries temporais. O Capítulo 12 enfoca a previsão e apresenta diversas ferramentas modernas para analisar regressões de séries temporais, tais como testes de raiz unitária e testes de

estabilidade. O Capítulo 13 discute o uso de dados de séries temporais para estimar relações causais. O Capítulo 14 apresenta algumas ferramentas mais avançadas para a análise de séries temporais, incluindo modelos de heteroscedasticidade condicional.

Parte V

A Parte V é uma introdução à teoria econométrica. Ela é mais do que um apêndice que completa os detalhes matemáticos omitidos do livro; é um tratamento em separado da teoria econométrica da estimação e da inferência no modelo de regressão linear. O Capítulo 15 desenvolve a teoria da análise de regressão para um único regressor; a exposição não utiliza álgebra matricial, embora exija um nível mais elevado de sofisticação matemática do que o restante do livro. O Capítulo 16 apresenta e estuda o modelo de regressão múltipla na forma matricial.

Pré-Requisitos dentro do Livro

Como professores diferentes gostam de enfatizar materiais diferentes, redigimos este livro tendo em mente diversas preferências didáticas. Na medida do possível, os capítulos das partes III, IV e V são "independentes" no sentido de que não requerem que todos os capítulos anteriores sejam estudados primeiro. Os pré-requisitos específicos para cada capítulo são descritos na Tabela 1. Embora tenhamos concluído que a seqüência de tópicos adotada no livro funciona bem em nossos próprios cursos, os capítulos foram escritos de tal modo que permitam que os professores apresentem os tópicos em uma ordem diferente se assim o desejarem.

TABELA 1 Guia de Pré-Requisitos dos Capítulos com Tópicos Especiais nas Partes III-V

	Partes ou Capítulos com Pré-Requisitos							
Capítulo	Parte I	Parte II	8.1, 8.2	10.1, 10.2	12.1-12.4	12.5-12.8	13	15
8	■	■						
9	■	■						
10.1, 10.2	■	■						
10.3-10.6	■	■	■	■				
11	■	■	■	■				
12	■	■						
13	■	■			■			
14	■	■			■	■	■	
15	■	■						
16	■	■						■

Esta tabela mostra os pré-requisitos mínimos necessários para cobrir o material de determinado capítulo. Por exemplo, a estimação de efeitos causais dinâmicos (Capítulo 13) requer que a Parte I (conforme necessário, dependendo do conhecimento do aluno), a Parte II e as seções 12.1-12.4 sejam vistas primeiro.

Amostras de Cursos

Pode-se utilizar este livro em diversos cursos.

Introdução à Econometria

Este curso apresenta a econometria (Capítulo 1) e faz uma revisão de probabilidade e estatística conforme necessário (capítulos 2 e 3). Em seguida, passa para a regressão com um único regressor, a regressão múltipla, o básico da análise de forma funcional, e a avaliação de estudos de regressão (toda a Parte II). O curso prossegue

cobrindo a regressão com dados de painel (Capítulo 8), a regressão com uma variável dependente limitada (Capítulo 9) e/ou a regressão de variáveis instrumentais (Capítulo 10), conforme o tempo permitir, e termina com experimentos e quase-experimentos no Capítulo 11, tópicos que fornecem uma oportunidade para retornar às questões de estimação de efeitos causais levantadas no início do semestre e para recapitular os métodos de regressão mais importantes. *Pré-requisitos: álgebra e introdução à estatística.*

Introdução à Econometria com Aplicações de Séries Temporais e Previsão

Assim como o curso introdutório padrão, este curso cobre toda a Parte I (conforme necessário) e toda a Parte II. Opcionalmente, o curso oferece a seguir uma breve introdução a dados de painel (seções 8.1 e 8.2) e examina a regressão de variáveis instrumentais (Capítulo 10 ou apenas as seções 10.1 e 10.2). O curso então segue para a Parte IV, cobrindo previsão (Capítulo 12) e a estimação de efeitos causais dinâmicos (Capítulo 13). Se o tempo permitir, o curso poderá incluir alguns tópicos avançados em análise de séries temporais, tais como heteroscedasticidade condicional (Seção 14.5). *Pré-requisitos: álgebra e introdução à estatística.*

Análise de Séries Temporais Aplicada e Previsão

Este livro também pode ser usado para um curso de curta duração sobre séries temporais aplicadas e previsão que tenha como pré-requisito um curso sobre análise de regressão. Gasta-se algum tempo com uma revisão das ferramentas da análise de regressão básica na Parte II, dependendo do conhecimento do aluno. O curso então prossegue diretamente para a Parte IV e trabalha a previsão (Capítulo 12), a estimação de efeitos causais dinâmicos (Capítulo 13) e tópicos avançados em análise de séries temporais (Capítulo 14), incluindo autoregressões vetoriais e heteroscedasticidade condicional. *Pré-requisitos: álgebra e introdução à econometria básica ou equivalente.*

Introdução à Teoria Econométrica

Este livro também é apropriado para um curso avançado de graduação com alunos que tenham um profundo conhecimento matemático ou para um curso de econometria em nível de mestrado. O curso faz uma breve revisão da teoria de estatística e probabilidade conforme necessário (Parte I) e apresenta a análise de regressão utilizando o tratamento não matemático baseado em aplicações da Parte II. Essa introdução é seguida pelo desenvolvimento teórico dos capítulos 15 e 16. O curso então examina a regressão com uma variável dependente limitada (Capítulo 9) e a estimação por máxima verossimilhança (Apêndice 9.2). A seguir, volta-se opcionalmente para regressão de variáveis instrumentais (Capítulo 10), métodos de séries temporais (Capítulo 12) e/ou estimação de efeitos causais utilizando dados de séries temporais e mínimos quadrados generalizados (Capítulo 13 e Seção 16.6). *Pré-requisitos: cálculo e introdução à estatística. O Capítulo 16 assume algum contato prévio com álgebra matricial.*

Recursos Pedagógicos

O livro inclui uma série de recursos pedagógicos com o objetivo de ajudar os alunos a entender, reter e aplicar as idéias essenciais. As *introduções aos capítulos* fornecem uma motivação e uma base do mundo real, bem como um breve guia que destaca a seqüência da discussão. Os *termos-chave* aparecem em negrito e são definidos no contexto ao longo de cada capítulo, e os *quadros de Conceito-Chave* apresentados em intervalos regulares recapitulam as idéias centrais. Os *quadros de interesse geral* fornecem menções interessantes a tópicos relacionados e destacam estudos do mundo real que utilizam os métodos ou conceitos em discussão no texto. Um *Resumo* numerado concluindo cada capítulo serve de quadro para a revisão dos principais pontos da cobertura. As questões da seção *Revisão dos Conceitos* verificam a compreensão do conteúdo principal por parte dos alunos, e os *Exercícios* servem para se colocarem em prática os conceitos e técnicas apresentados no capítulo. No final do livro, a seção *Referências Bibliográficas* lista fontes para leitura adicional, o *Apêndice* fornece tabelas estatísticas, e um *Glossário* define convenientemente todos os termos-chave do livro.

Recursos Adicionais

Professores e estudantes que adotam este livro têm à disposição uma gama de recursos adicionais no site do livro, em **www.aw.com/stock_br**. Nele os estudantes encontram arquivos de dados (*data sets*) usados no livro, bem como soluções para os exercícios selecionados (em inglês), e os professores encontram o manual de soluções (em inglês) e apresentações em PowerPoint.

Agradecimentos

Muitas pessoas contribuíram para este projeto. Nossos maiores agradecimentos são para nossos colegas de Harvard e Princeton que utilizaram manuscritos preliminares deste livro em sala de aula. Suzanne Cooper, da Kennedy School of Government em Harvard, fez sugestões inestimáveis e comentários detalhados sobre muitos manuscritos. Na qualidade de professora assistente de um dos autores (Stock), ela também ajudou a revisar boa parte do material deste livro durante seu desenvolvimento para uma disciplina de mestrado solicitada na Kennedy School. Também somos gratos a dois outros colegas da Kennedy School: Alberto Abadie e Sue Dynarski, por suas explicações pacientes sobre quase-experimentos e sobre o campo da avaliação de programas e por seus comentários detalhados sobre os manuscritos preliminares do livro. Em Princeton, Eli Tamer utilizou um manuscrito preliminar do livro em suas aulas e também forneceu comentários úteis sobre o penúltimo manuscrito.

Também somos gratos a muitos de nossos amigos e colegas econometristas que gastaram tempo discutindo conosco a essência deste livro e que no todo nos ofereceram tantas sugestões úteis. Bruce Hansen (Universidade de Wiscosin, Madison) e Bo Honore (Princeton) ofereceram um feedback útil sobre os esboços iniciais e as versões preliminares do material principal da parte II. Joshua Angrist (MIT) e Guido Imbens (Universidade de Berkeley, Califórnia) forneceram sugestões profundas sobre nossa abordagem dos materiais sobre avaliação de programas. Nossa apresentação do material sobre séries temporais foi beneficiada pelas discussões com Yacine Ait-Sahalia (Princeton), Graham Elliot (Universidade da Califórnia, San Diego), Andrew Harvey (Universidade de Cambridge) e Christopher Sims (Princeton). Por fim, muitas pessoas fizeram sugestões úteis sobre partes do manuscrito relacionadas a suas áreas de especialização: Don Andrews (Yale), John Bound (Universidade de Michigan), Gregory Chow (Princeton), Thomas Downes (Tufts), David Drukker (Stata, Corp.), Jean Baldwin Grossman (Princeton), Eric Hanushek (Hoover Institution), James Heckman (Universidade de Chicago), Han Hong (Princeton), Caroline Hoxby (Harvard), Alan Krueger (Princeton), Steven Levitt (Universidade de Chicago), Richard Light (Harvard), David Neumark (Michigan State University), Joseph Newhouse (Harvard), Pierre Perron (Universidade de Boston), Kenneth Warner (Universidade do Michigan) e Richard Zeckhauser (Harvard).

Muitas pessoas foram muito generosas em nos fornecer dados. Os dados sobre pontuação nos exames da Califórnia foram elaborados com a ajuda de Les Axelrod da Divisão de Padrões e Avaliações do Departamento de Educação da Califórnia. Somos gratos a Charlie DePascale, do Serviço de Avaliação de Alunos do Departamento de Educação de Massachusetts, por sua ajuda em aspectos da base de dados sobre pontuação nos exames de Massachusetts. Christopher Ruhm (Universidade da Carolina do Norte, Greensboro) nos forneceu gentilmente sua base de dados sobre leis relativas a motoristas bêbados e mortes no trânsito. O departamento de pesquisa do Federal Reserve Bank de Boston merece nossos agradecimentos por reunir seus dados sobre discriminação racial nos empréstimos hipotecários; agradecemos em particular a Geoffrey Tootell por nos fornecer a versão atualizada da base de dados que utilizamos no Capítulo 9 e a Lynn Browne por nos explicar o contexto de política a ele relacionada. Agradecemos a Jonathan Gruber (MIT) por compartilhar seus dados sobre vendas de cigarros, que analisamos no Capítulo 10, e a Alan Krueger (Princeton) por sua ajuda com os dados do Projeto STAR, do Tennessee, que analisamos no Capítulo 11.

Somos também muito gratos pelos muitos comentários construtivos, detalhados e profundos que recebemos daqueles que revisaram os diversos manuscritos para a Addison-Wesley:

Michael Abbott, Queen's University, Canadá
Richard J. Agnello, Universidade de Delaware

Clopper Almon, Universidade de Maryland
Joshua Angrist, Massachusetts Institute of Technology
Swarnjit S. Arora, Universidade de Wisconsin, Milwaukee
Christopher F. Baum, Boston College
McKinley L. Blackburn, Universidade da Carolina do Sul
Alok Bohara, Universidade do Novo México
Chi-Young Choi, Universidade de New Hampshire
Dennis Coates, University de Maryland, Baltimore
Tim Conley, Graduate School of Business, Universidade de Chicago
Douglas Dalenberg, Universidade de Montana
Antony Davies, Duquesne University
Joanne M. Doyle, James Madison University
David Eaton, Murray State University
Adrian R. Fleissig, California State University, Fullerton
Rae Jean B. Goodman, United States Naval Academy
Bruce E. Hansen, Universidade de Wisconsin, Madison
Peter Reinhard Hansen, Brown University
Ian T. Henry, Universidade de Melbourne, Austrália
Marc Henry, Universidade de Columbia
William Horrace, Universidade do Arizona
Òscar Jordà, Universidade da Califórnia, Davis
Frederick L. Joutz, The George Washington University
Elia Kacapyr, Ithaca College
Manfred W. Keil, Claremont McKenna College
Eugene Kroch, Villanova University
Gary Krueger, Macalester College
Kajal Lahiri, State University of New York, Albany
Daniel Lee, Shippensburg University
Tung Liu, Ball State University
Ken Matwiczak, LBJ School of Public Affairs, Universidade do Texas, Austin
KimMarie McGoldrick, Universidade de Richmond
Robert McNown, Universidade do Colorado, Boulder
H. Naci Mocan, Universidade do Colorado, Denver
Mototsugu Shintani, Vanderbilt University
Mico Mrkaic, Duke University
Serena Ng, Johns Hopkins University
Jan Ondrich, Universidade de Siracusa
Pierre Perron, Universidade de Boston
Robert Phillips, The George Washington University
Simran Sahi, Universidade de Minnesota
Sunil Sapra, California State University, Los Angeles
Frank Schorfheide, Universidade da Pennsylvania
Leslie S. Stratton, Virginia Commonwealth University
Jane Sung, Truman State University
Christopher Taber, Northwestern University
Petra Todd, Universidade da Pennsylvania
John Veitch, Universidade de San Francisco
Edward J. Vytlacil, Universidade de Stanford

M. Daniel Westbrook, Georgetown University
Tiemen Woutersen, University of Western Ontario
Phanindra V. Wunnava, Middlebury College
Zhenhui Xu, Georgia College and State University
Yong Yin, State University of New York, Buffalo
Jiangfeng Zhang, Universidade de Berkeley, Califórnia,
John Xu Zheng, Universidade do Texas, Austin

Agradecemos a várias pessoas por seu trabalho cuidadoso de revisão. Kerry Griffin e Yair Listokin leram todo o manuscrito e Andrew Fraker, Ori Heffetz, Amber Henry, Hong Li, Alessandro Tarozzi e Matt Watson examinaram cuidadosamente diversos capítulos.

Tivemos o benefício da ajuda de uma editora de desenvolvimento excepcional, Jane Tufts, cuja criatividade, trabalho duro e atenção aos detalhes aprimoraram o livro de várias formas. A Addison-Wesley nos proporcionou um suporte de primeira classe, começando por nossa excelente editora, Sylvia Mallory, e estendendo-se a toda a equipe de publicação. Jane e Sylvia nos ensinaram pacientemente muito sobre redação, organização e apresentação, e seus esforços são evidentes em cada página deste livro. Também estendemos nossos agradecimentos a todos os demais profissionais da excelente equipe da Addison-Wesley, que nos ajudaram em cada passo do processo complexo de publicação deste livro: Adrienne D'Ambrosio (gerente de marketing), Melissa Honig (produtora de mídia sênior), Regina Kolenda (designer sênior), Katherine Watson (supervisora de produção) e, especialmente, Denise Clinton (editora-chefe).

Acima de tudo, somos gratos a nossas famílias por sua tolerância durante este projeto. Escrever este livro consumiu um tempo enorme — para elas, o projeto deve ter parecido infindável. Elas, mais do que ninguém, suportaram o ônus deste compromisso, e somos profundamente gratos por sua ajuda e seu apoio.

PARTE UM

Introdução e Revisão

CAPÍTULO 1

Questões e Dados Econômicos

Pergunte a meia dúzia de econometristas a definição de econometria e você terá meia dúzia de respostas diferentes. Um deles poderia dizer que econometria é a ciência que testa teorias econômicas. Um segundo poderia dizer que é o conjunto de ferramentas utilizadas para prever valores futuros de variáveis econômicas, tais como as vendas de uma empresa, o crescimento global da economia ou os preços das ações. Outro poderia dizer que se trata do processo de ajuste dos modelos econômicos matematizados a dados do mundo real. Outro ainda diria que econometria é a arte e a ciência que utiliza dados históricos para fazer recomendações numéricas ou quantitativas de política no governo e nos negócios.

Na verdade, todas essas respostas estão certas. Sob um prisma mais amplo, econometria é a ciência e a arte que utiliza a teoria econômica e as técnicas estatísticas para analisar dados econômicos. Os métodos econométricos são usados em muitas áreas da economia, incluindo finanças, economia do trabalho, macroeconomia, microeconomia, marketing e política econômica. Esses métodos normalmente também são utilizados em outras ciências sociais, incluindo ciência política e sociologia.

Este livro apresenta o conjunto fundamental de métodos utilizados pelos econometristas. Usaremos esses métodos para responder a diversas questões quantitativas específicas, extraídas do mundo da política, dos negócios e do governo. Este capítulo apresenta quatro dessas questões e discute, em termos gerais, a abordagem econométrica para responder a elas. Termina com uma resenha dos principais tipos de dados disponíveis aos econometristas para responder a essas e a outras questões econômicas quantitativas.

1.1 Questões Econômicas que Examinamos

Muitas decisões tomadas na economia, nos negócios e no governo dependem da compreensão das relações entre variáveis no mundo que nos cerca. Essas decisões requerem respostas quantitativas para questões quantitativas.

Este livro examinará diversas questões quantitativas extraídas de tópicos atuais em economia. Quatro dessas questões referem-se à política educacional, ao viés racial na contratação de empréstimos hipotecários, ao consumo de cigarros e à previsão macroeconômica.

Questão nº 1: A Redução do Tamanho das Turmas Melhora o Aprendizado no Ensino Fundamental?

Propostas de reforma do sistema de ensino público dos Estados Unidos geram debates acalorados. Muitas propostas referem-se aos estudantes mais jovens, do ensino fundamental, o qual tem vários objetivos, tais como o desenvolvimento de habilidades sociais, mas para muitos pais e educadores o objetivo mais importante é o aprendizado acadêmico básico: ler, escrever e fazer as operações matemáticas elementares. Uma proposta importante para melhorar o aprendizado básico é a redução do tamanho das turmas do ensino fundamental. Com menos alunos nas turmas, segue o argumento, cada aluno recebe mais atenção do professor, há menos interrupções durante a aula, o aprendizado cresce e as notas melhoram.

Mas qual é, exatamente, o efeito da redução do tamanho das turmas sobre o ensino fundamental? Essa redução custa dinheiro: requer a contratação de mais professores e, se a escola já estiver operando em sua capacidade máxima, a construção de novas salas de aula. Um tomador de decisões que esteja considerando a contratação de mais professores deve pesar os custos e os benefícios. Para fazer isso, contudo, o tomador de decisões deve ter uma compreensão quantitativa exata dos prováveis benefícios. O efeito benéfico de turmas menores sobre o aprendizado básico é grande ou pequeno? É possível que turmas menores não tenham nenhum efeito sobre o aprendizado básico?

Embora o senso comum e a experiência cotidiana possam sugerir que um maior aprendizado ocorra quando há menos estudantes, o senso comum não pode fornecer uma resposta quantitativa à questão sobre qual é exatamente o efeito na redução do tamanho das turmas sobre o aprendizado básico. Para obter essa resposta, precisamos examinar a evidência empírica — isto é, a evidência baseada em dados — que relaciona o tamanho das turmas ao aprendizado básico no ensino fundamental.

Neste livro, examinaremos a relação entre o tamanho das turmas e o aprendizado básico utilizando dados coletados de 420 diretorias regionais de ensino da Califórnia em 1998. Segundo esses dados, alunos de diretorias com turmas pequenas tendem a ter desempenho melhor em exames padronizados do que alunos de diretorias regionais com turmas maiores. Embora esse fato seja coerente com a idéia de que turmas menores geram pontuações melhores nos exames, ele pode simplesmente refletir outras vantagens que alunos de diretorias de ensino com turmas pequenas têm sobre seus colegas de diretorias de ensino com turmas grandes. Por exemplo, diretorias de ensino com turmas pequenas tendem a ter moradores com melhor poder aquisitivo do que aquelas com turmas grandes; desse modo, os alunos de diretorias de ensino com turmas pequenas têm mais oportunidades para aprender fora da sala de aula. Pode ser que sejam essas oportunidades adicionais de aprendizado, e não as turmas menores, que levem a pontuações mais altas nos exames. Na Parte 2, utilizaremos análise de regressão múltipla para isolar o efeito de mudanças no tamanho das turmas em relação a mudanças em outros fatores, tais como a situação econômica dos alunos.

Questão nº 2: Há Discriminação Racial no Mercado de Empréstimos Hipotecários?

A maioria dos norte-americanos adquire sua residência com o auxílio de uma hipoteca, um grande empréstimo garantido pelo valor do imóvel. Por lei, as instituições financeiras norte-americanas não podem levar em consideração a raça do requerente ao decidir pela concessão ou recusa de um pedido de hipoteca: candidatos idênticos em todos os aspectos — com exceção de sua raça — deveriam ter seu pedido de hipoteca aprovado. Em teoria, portanto, não deveria haver viés racial na concessão de empréstimos hipotecários.

Em contraste com essa conclusão teórica, pesquisadores do Federal Reserve Bank (Banco da Reserva Federal), de Boston,* constataram (utilizando dados do início da década de 1990) que 28 por cento dos requerentes negros tinham seu pedido de hipoteca negado, enquanto apenas 9 por cento dos requerentes brancos tinham seu pedido negado. Será que esses dados indicam que, na prática, existe viés racial na concessão de empréstimos hipotecários? E, se for esse o caso, qual é a dimensão do viés?

De acordo com os dados do Fed de Boston, o fato de mais requerentes negros do que brancos terem seu pedido negado não fornece por si só evidência de discriminação por parte dos mutuantes, pois requerentes negros e brancos diferem sob muitos outros aspectos além da raça. Antes de concluir que existe viés racial no mercado hipotecário, esses dados devem ser examinados mais de perto para se constatar se existe uma diferença na probabilidade de requerentes *idênticos sob outros aspectos* terem o pedido negado e, se for esse o caso, se essa diferença é grande ou pequena. Para isso, no Capítulo 9 apresentaremos os métodos econométricos que permitem quantificar o efeito da raça sobre a probabilidade de se obter uma hipoteca, *mantendo constantes* outras características dos requerentes, particularmente sua capacidade de saldar o empréstimo.

Questão nº 3: Em Quanto a Tributação sobre Cigarros Reduz Seu Consumo?

O consumo de cigarros é um grande problema de saúde pública mundial. Muitos dos custos do hábito de fumar recaem sobre outros membros da sociedade, tais como as despesas com tratamento médico de doenças provocadas pelo cigarro e os custos menos quantificáveis para os não-fumantes, que preferem não respirar a fumaça de terceiros. Como esses custos recaem sobre outras pessoas além do fumante, há intervenção governamental na

* Agência Distrital do Banco Central dos Estados Unidos (Federal Reserve Bank), situada em Boston. O Banco Central norte-americano é normalmente chamado de Fed, nome que utilizaremos ocasionalmente no texto. Maiores detalhes sobre a organização do Banco Central dos Estados Unidos podem ser encontrados, *inter alia*, em Blanchard, Olivier J. *Macroeconomics*, 3. ed. Boston: Addison-Wesley, 2003 (N. do R.T.).

redução do consumo de cigarros. Um dos instrumentos mais flexíveis para a redução do consumo é o aumento dos impostos sobre cigarros.

De acordo com as noções de economia básica, se os preços dos cigarros subirem, o consumo cairá. Mas em quanto? Se o preço de venda subir 1 por cento, em qual percentual a quantidade de cigarros vendidos cairá? A mudança percentual da quantidade demandada resultante de um aumento de 1 por cento no preço é a *elasticidade-preço da demanda*. Se queremos reduzir o hábito de fumar em determinado montante, digamos 20 por cento, por meio do aumento de impostos, precisamos conhecer a elasticidade-preço para calcular o aumento de preços necessário para atingir essa redução do consumo. Mas qual é a elasticidade-preço da demanda por cigarros?

Embora a teoria econômica nos forneça conceitos que ajudam a responder a essa questão, ela não nos diz o valor numérico da elasticidade-preço da demanda. Para conhecer a elasticidade, precisamos examinar a evidência empírica sobre o comportamento dos fumantes e dos fumantes em potencial. Em outras palavras, precisamos analisar dados sobre o consumo e os preços dos cigarros.

Os dados que examinamos são relativos a vendas e preços de cigarros, impostos sobre cigarros e renda pessoal em nível estadual para os Estados Unidos nas décadas de 1980 e 1990. Segundo os dados, estados com baixos impostos sobre cigarros e, portanto, cigarros com preços baixos apresentam um alto índice de fumantes, enquanto estados com preços elevados apresentam um baixo índice de fumantes. Entretanto, a análise desses dados é complicada, uma vez que a causalidade se move nos dois sentidos: impostos baixos levam a uma demanda elevada, mas, se há muitos fumantes no Estado, os políticos locais podem tentar manter os impostos sobre cigarros em um patamar baixo para satisfazer a seus eleitores fumantes. No Capítulo 10 estudaremos os métodos que lidam com essa "causalidade simultânea" e utilizaremos esses métodos para estimar a elasticidade-preço da demanda por cigarros.

Questão nº 4: Qual Será a Taxa de Inflação no Próximo Ano?

Parece que as pessoas sempre desejam prever o futuro. Como serão as vendas no próximo ano em uma empresa que planeja um investimento em equipamentos novos? Será que a bolsa de valores vai subir no mês que vem, e, se for esse o caso, em quanto? Os impostos municipais arrecadados no ano que vem cobrirão os gastos planejados com serviços municipais? Sua prova de microeconomia na semana que vem se concentrará em externalidades ou monopólios? Sábado será um bom dia para ir à praia?

Um aspecto do futuro em que os macroeconomistas e economistas financeiros estão particularmente interessados é a taxa da inflação total dos preços durante o próximo ano. Um financista pode aconselhar um cliente a tomar um empréstimo ou a saldá-lo a uma dada taxa de juros, dependendo de uma melhor previsão para a taxa de inflação ao longo do próximo ano. Os economistas de bancos centrais como o Federal Reserve Board, em Washington, D.C., e o European Central Bank, em Frankfurt, Alemanha, são responsáveis por manter a taxa da inflação de preços sob controle e por isso suas decisões sobre a fixação das taxas de juros se apóiam na previsão da inflação ao longo do próximo ano. Se projetarem o aumento da taxa de inflação em um ponto percentual, eles devem aumentar as taxas de juros acima disso para desacelerar uma economia que, de seu ponto de vista, corre o risco de estar superaquecida. Se a sua projeção estiver errada, correm o risco de provocar uma recessão desnecessária ou um salto indesejável na taxa de inflação.

Os economistas que dependem de previsões numéricas precisas utilizam modelos econométricos para fazer essas previsões. O trabalho de um formulador de previsões consiste em prever o futuro usando o passado; os econometristas fazem isso utilizando-se da teoria econômica e de técnicas estatísticas para quantificar relações em dados históricos.

Os dados utilizados nos Estados Unidos para prever a inflação são as taxas de inflação e de desemprego. Uma relação empírica importante nos dados macroeconômicos é a "curva de Phillips", em que um valor atualmente baixo da taxa de desemprego está associado a um aumento da taxa de inflação ao longo do próximo ano. Uma das previsões de inflação que desenvolveremos e avaliaremos no Capítulo 12 baseia-se na curva de Phillips.

Questões Quantitativas, Respostas Quantitativas

Cada uma dessas quatro questões requer uma resposta numérica. A teoria econômica oferece pistas sobre a resposta — o consumo de cigarros deve diminuir se o preço aumenta —, mas o valor efetivo do número deve

ser obtido empiricamente, isto é, por meio da análise dos dados. Como utilizamos dados para responder a questões quantitativas, nossas respostas sempre possuem alguma incerteza: um conjunto diferente de dados produziria uma resposta numérica diferente. Desse modo, a estrutura conceitual para a análise precisa fornecer tanto uma resposta numérica para a questão quanto uma medida da precisão dessa resposta.

A estrutura conceitual utilizada neste livro é o modelo de regressão múltipla, o alicerce da econometria. Esse modelo, apresentado na Parte 2, proporciona uma forma matemática de quantificar como uma mudança em uma variável afeta outra variável, mantendo tudo o mais constante. Por exemplo, qual é o efeito de uma mudança no tamanho das turmas sobre as pontuações nos exames, *mantendo constantes* as características dos alunos (como renda familiar) que o diretor regional de ensino não pode controlar? Qual é o efeito de sua raça sobre suas oportunidades de ter um pedido de hipoteca aprovado, *mantendo constantes* fatores como sua capacidade de saldar o empréstimo? Que efeito tem um aumento de 1 por cento no preço dos cigarros sobre seu consumo, *mantendo constante* a renda dos fumantes e dos fumantes em potencial? O modelo de regressão múltipla e suas extensões fornecem uma estrutura para responder a essas questões utilizando dados e para quantificar a incerteza associada a essas respostas.

1.2 Efeitos Causais e Experimentos Idealizados

Assim como muitas questões encontradas na econometria, as primeiras três questões da Seção 1.1 dizem respeito a relações causais entre variáveis. No uso cotidiano, diz-se que uma ação causa um efeito se o efeito é o resultado direto, ou conseqüência, daquela ação. Encostar em um forno quente faz com que você se queime; beber água diminui sua sede; colocar ar nos pneus do carro faz com que eles fiquem cheios; aplicar fertilizante em plantações de tomates faz com que nasçam mais tomates. Causalidade significa que uma ação específica (aplicar fertilizante) leva a uma conseqüência específica, mensurável (mais tomates).

Estimação de Efeitos Causais

Qual é a melhor forma de medir o efeito causal sobre a produção de tomates (medida em quilos) da aplicação de determinado montante de fertilizante, digamos 100 gramas de fertilizante por metro quadrado?

Uma forma de medir esse efeito causal é conduzir um experimento. Nesse experimento, um pesquisador agrônomo planta diversos lotes de tomates. Cada lote é cuidado de forma idêntica, mas com uma exceção: alguns lotes recebem 100 gramas de fertilizante por metro quadrado, ao passo que os demais não recebem nada. Além disso, a atribuição quanto a um lote receber ou não fertilizante é feita aleatoriamente por um computador, o que assegura que quaisquer outras diferenças entre os lotes não estejam relacionadas com o recebimento ou não de fertilizante. Ao final da safra, o agrônomo pesa a colheita de cada lote. A diferença entre a produção média por metro quadrado dos lotes tratados e dos não tratados é o efeito do tratamento por fertilizante sobre a produção de tomates.

Esse é um exemplo de um **experimento controlado aleatório**. É controlado na medida em que existe tanto um **grupo de controle**, que não recebe nenhum tratamento (nenhum fertilizante), quanto um **grupo de tratamento**, que recebe o tratamento (100 g/m^2 de fertilizante). É aleatório porque o tratamento é atribuído aleatoriamente. Essa atribuição aleatória elimina a possibilidade de uma relação sistemática entre, por exemplo, a quantidade de luz solar no lote e o recebimento ou não de fertilizante, de modo que a única diferença sistemática entre os grupos de tratamento e de controle é o tratamento em si. Se o experimento for implementado de modo adequado em uma escala suficientemente grande, ele produzirá uma estimativa do efeito causal do tratamento (aplicar 100 g/m^2 de fertilizante) sobre o resultado de interesse (produção de tomates).

Neste livro, o **efeito causal** é definido como o efeito de dada ação ou tratamento sobre um resultado, conforme medido em um experimento controlado aleatório ideal. Nesse experimento, a única razão sistemática para as diferenças observadas nos resultados entre os grupos de tratamento e de controle é o tratamento em si.

É possível imaginar um experimento controlado aleatório ideal para responder a cada uma das três primeiras questões da Seção 1.1. Por exemplo, para estudar o tamanho das turmas, podemos imaginar a atribuição aleatória de "tratamentos" de diferentes tamanhos de turmas a diferentes grupos de alunos. Se o experimento for elaborado e implementado de forma que a única diferença sistemática entre os grupos de alunos seja o tamanho das turmas, teoricamente ele estimará o efeito da redução do tamanho das turmas sobre as pontuações dos exames, mantendo tudo o mais constante.

O conceito de experimento controlado aleatório ideal é útil porque fornece uma definição de efeito causal. Na prática, entretanto, não é possível conduzir experimentos ideais. Na verdade, experimentos são raros em econometria porque freqüentemente são antiéticos, impossíveis de executar de maneira satisfatória ou absurdamente caros. O conceito de experimento controlado aleatório ideal, porém, proporciona um ponto de referência teórico para uma análise econométrica de efeitos causais utilizando dados do mundo real.

Previsão e Causalidade

Embora as três primeiras questões da Seção 1.1 se refiram a efeitos causais, o mesmo não acontece com a quarta questão — previsão da inflação. Você não precisa conhecer uma relação causal para fazer uma boa previsão. Uma boa forma de "prever" se está chovendo é observar se os pedestres estão usando guarda-chuvas, mas o ato de utilizar um guarda-chuva não causa a chuva.

Mesmo que a previsão não envolva relações causais, a teoria macroeconômica sugere padrões e relações que podem ser úteis para a previsão da inflação. Como veremos no Capítulo 12, a análise de regressão múltipla permite quantificar relações históricas sugeridas pela teoria econômica, verificar se essas relações são estáveis ao longo do tempo, fazer projeções quantitativas sobre o futuro e avaliar a precisão dessas previsões.

1.3 Dados: Fontes e Tipos

Em econometria, os dados originam-se de experimentos ou de observações não experimentais do mundo. Este livro examina as bases de dados experimentais e não experimentais.

Dados Experimentais *versus* Dados Observacionais

Dados experimentais vêm de experimentos concebidos para a avaliação de um tratamento ou uma política ou para a investigação de um efeito causal. Por exemplo, o Estado do Tennessee, nos Estados Unidos, financiou um grande experimento controlado aleatório examinando o tamanho das turmas na década de 1980. Nesse experimento, que examinaremos no Capítulo 11, milhares de alunos foram atribuídos aleatoriamente para turmas de tamanhos diferentes durante vários anos e tiveram de fazer anualmente exames padronizados.

O experimento do tamanho das turmas do Tennessee custou milhões de dólares e exigiu a cooperação contínua de muitos diretores de ensino, pais e professores ao longo de muitos anos. Como os experimentos do mundo real com indivíduos são difíceis de administrar e controlar, eles apresentam falhas em relação aos experimentos controlados aleatórios ideais. Além disso, em algumas circunstâncias, os experimentos não apenas são caros e difíceis de controlar, mas também antiéticos. (Seria ético oferecer cigarros baratos a adolescentes selecionados aleatoriamente para ver o quanto eles compram?) Em virtude desses problemas financeiros, práticos e éticos, os experimentos em economia são raros. Em vez disso, a maioria dos dados econômicos é obtida por meio da observação do comportamento no mundo real.

Dados obtidos pela observação do comportamento efetivo fora de um ambiente experimental são chamados de **dados observacionais**. Esses dados são coletados por meio de pesquisas, como uma pesquisa por telefone com consumidores, e por registros administrativos, como registros históricos sobre pedidos de hipoteca mantidos pelas instituições financeiras.

Dados observacionais colocam grandes desafios para as tentativas econométricas de estimação de efeitos causais, e as ferramentas da econometria enfrentam esses desafios. No mundo real, os níveis de "tratamento" (a quantidade de fertilizante no exemplo do tomate, a razão aluno-professor no exemplo do tamanho das turmas) não são atribuídos aleatoriamente, de modo que é difícil separar o efeito do "tratamento" de outros fatores relevantes. Grande parte da econometria, e deste livro, dedica-se a métodos que fazem frente aos desafios encontrados quando os dados do mundo real são utilizados para a estimação de efeitos causais.

Sejam experimentais ou observacionais, as bases de dados podem ser de três tipos principais: dados de corte, dados de séries temporais e dados de painel. Neste livro você encontrará todos eles.

Dados de Corte

Dados sobre diferentes entidades — trabalhadores, consumidores, empresas, unidades governamentais e assim por diante — em um único período de tempo são chamados de **dados de corte**. Por exemplo, os dados sobre pontuações de exames nas diretorias de ensino da Califórnia são dados de corte: referem-se a 420 entidades (diretorias de ensino) para um único período de tempo (1998). Em geral, o número de entidades sobre as quais temos observações é representado por n, de modo que temos $n = 420$ para o exemplo da base de dados da Califórnia.

A base de dados sobre as pontuações nos exames da Califórnia contém medidas de variáveis diferentes para cada diretoria. Alguns desses dados estão na Tabela 1.1. Cada linha apresenta dados para uma diretoria diferente. Por exemplo, a pontuação média nos exames para a primeira diretoria ("diretoria nº 1") é 690,8; essa é a média das pontuações dos exames de matemática e ciências para todos os alunos da quinta série dessa diretoria em 1998 em um exame padronizado (SAT — Stanford Achievement Test, ou Exame de Aproveitamento de Stanford). A razão média aluno-professor naquela diretoria é de 17,89, isto é, o número de alunos na diretoria nº 1 dividido pelo número de professores é 17,89. O gasto médio por aluno nessa diretoria é de US$ 6.385. O percentual de alunos nessa diretoria que ainda estão aprendendo inglês — isto é, o percentual de alunos para os quais o inglês é o segundo idioma e que ainda não são proficientes nessa língua — é 0 por cento.

As linhas seguintes apresentam dados para outras diretorias. A ordem das linhas é arbitrária e o número da diretoria, chamado de **número da observação**, é um número atribuído arbitrariamente que organiza os dados. Como você pode ver na tabela, todas as variáveis listadas apresentam mudanças consideráveis.

Com dados de corte, podemos aprender sobre as relações entre variáveis estudando as diferenças entre pessoas, empresas ou outras entidades econômicas durante um único período de tempo.

Dados de Séries Temporais

Dados de séries temporais são dados para uma entidade única (pessoa, empresa, país) coletados em diversos períodos de tempo. Nossa base de dados sobre taxas de inflação e desemprego nos Estados Unidos é um

TABELA 1.1 Observações Selecionadas sobre Pontuações de Exames e Outras Variáveis para as Diretorias de Ensino da Califórnia em 1998

Número da observação (diretoria)	Pontuação média nos exames da diretoria (5ª série)	Razão aluno-professor	Gastos por aluno (US$)	Porcentagem de alunos aprendendo inglês
1	690,8	17,89	6.385	0,0
2	661,2	21,52	5.099	4,6
3	643,6	18,70	5.502	30,0
4	647,7	17,36	7.102	0,0
5	640,8	18,67	5.236	13,9
⋮	⋮	⋮	⋮	⋮
418	645,0	21,89	4.403	24,3
419	672,2	20,20	4.776	3,0
420	655,8	19,04	5.993	5,0

Nota: A base de dados sobre pontuação nos exames da Califórnia está no Apêndice 4.1.

exemplo de base de dados de séries temporais. A base de dados contém observações de duas variáveis (as taxas de inflação e desemprego) para uma única entidade (Estados Unidos) durante 167 períodos de tempo. Cada período de tempo dessa base de dados corresponde a um trimestre de um ano (o primeiro trimestre é janeiro, fevereiro e março; o segundo é abril, maio e junho, e assim por diante). As observações dessa base de dados se iniciaram no segundo trimestre de 1959, que passa a ser representado por 1959:II, e terminaram no quarto trimestre de 2000 (2000:IV). O número de observações (isto é, períodos de tempo) em uma base de dados de séries temporais é representado por T. Como há 167 trimestres de 1959:II a 2000:IV, essa base de dados contém $T = 167$ observações.

Algumas observações dessa base de dados são mostradas na Tabela 1.2. Os dados de cada linha correspondem a um período de tempo diferente (ano e trimestre). No segundo trimestre de 1959, por exemplo, a taxa de inflação de preços foi de 0,7 por cento ao ano a uma taxa anualizada. Em outras palavras, se a inflação tivesse continuado por 12 meses à taxa do segundo trimestre de 1959, o nível geral de preços (conforme medido pelo Índice de Preços ao Consumidor – IPC) teria aumentado 0,7 por cento. No segundo trimestre de 1959, a taxa de desemprego foi de 5,1 por cento, isto é, 5,1 por cento da força de trabalho declarou que estava sem trabalho e que procurava uma colocação. No terceiro trimestre de 1959, a taxa de inflação medida pelo IPC foi de 2,1 por cento e a taxa de desemprego foi de 5,3 por cento.

Ao monitorar uma única entidade ao longo do tempo, os dados de séries temporais podem ser utilizados para o estudo da evolução das variáveis ao longo do tempo e para a previsão dos valores futuros dessas variáveis.

Dados de Painel

Dados de painel, também chamados de **dados longitudinais**, são dados de diversas entidades em que cada uma delas é observada em dois ou mais períodos de tempo. Nossos dados sobre consumo de cigarros e preços de cigarros são um exemplo de base de dados de painel. Na Tabela 1.3, mostramos variáveis e observações selecionadas dessa base de dados. O número de entidades de uma base de dados de painel é representado por n, e o número de períodos de tempo, por T. Na base de dados de cigarros, temos observações para $n = 48$ estados norte-americanos (entidades) durante $T = 11$ anos (períodos de tempo), de 1985 a 1995. Portanto, existe um total de $n \times T = 48 \times 11 = 528$ observações.

TABELA 1.2 Observações Selecionadas para as Taxas de Inflação pelo Índice de Preços ao Consumidor (IPC) e de Desemprego nos Estados Unidos: Dados Trimestrais, 1959–2000.

Número da observação	Data (Ano: trimestre)	Taxa de inflação pelo IPC (Porcentagem ao ano a uma taxa anualizada)	Taxa de desemprego (%)
1	1959:II	0,7	5,1
2	1959:III	2,1	5,3
3	1959:IV	2,4	5,6
4	1960:I	0,4	5,1
5	1960:II	2,4	5,2
⋮	⋮	⋮	⋮
165	2000:II	3,0	4,0
166	2000:III	3,5	4,0
167	2000:IV	2,8	4,0

Nota: A base de dados sobre inflação e desemprego nos Estados Unidos está no Apêndice 12.1.

Dados de Corte, de Séries Temporais e de Painel

Conceito-Chave 1.1

- Dados de corte consistem em diversas entidades observadas em um único período.
- Dados de séries temporais consistem em uma única entidade observada em diversos períodos.
- Dados de painel (também conhecidos como dados longitudinais) consistem em diversas entidades em que cada uma delas é observada em dois ou mais períodos.

TABELA 1.3 Observações Selecionadas sobre Vendas e Preços de Cigarros e Impostos sobre Esse Produto por Estado e Ano para os Estados Unidos, 1985–1995

Número da observação	Estado	Ano	Vendas de cigarros (maços *per capita*)	Preço médio por maço (incluindo impostos)	Impostos totais (imposto sobre consumo + imposto sobre vendas)
1	Alabama	1985	116,5	US$ 1,022	US$ 0,333
2	Arkansas	1985	128,5	1,015	0,370
3	Arizona	1985	104,5	1,086	0,362
⋮	⋮	⋮	⋮	⋮	⋮
47	West Virginia	1985	112,8	1,089	0,382
48	Wyoming	1985	129,4	0,935	0,240
49	Alabama	1986	117,2	1,080	0,334
⋮	⋮	⋮	⋮	⋮	⋮
96	Wyoming	1986	127,8	1,007	0,240
97	Alabama	1987	115,8	1,135	0,335
⋮	⋮	⋮	⋮	⋮	⋮
528	Wyoming	1995	112,2	1,585	0,360

Nota: A base de dados sobre consumo de cigarros está no Apêndice 10.1.

Uma parte da base de dados de consumo de cigarros é mostrada na Tabela 1.3. O primeiro bloco de 48 observações mostra os dados de cada Estado dos Estados Unidos em 1985, organizados alfabeticamente do Alabama a Wyoming. O bloco seguinte de 48 observações lista os dados de 1986 e assim por diante, até 1995. Por exemplo, em 1985, as vendas de cigarros no Arkansas eram de 128,5 maços per capita (o número total de maços de cigarros vendidos no Arkansas em 1985 dividido pelo total da população do Estado em 1985 é igual a 128,5). O preço médio de um maço de cigarros no Arkansas em 1985, incluindo impostos, era US$ 1,015, dos quais US$ 0,37 eram impostos federais, estaduais e locais.*

* O contexto local dos Estados Unidos é próximo, no Brasil, ao contexto *municipal*. Na verdade, o caso norte-americano se refere a condados (N. do R.T.).

Os dados de painel podem ser usados para aprender sobre relações econômicas com base nas experiências de muitas entidades diferentes na base de dados e na evolução ao longo do tempo das variáveis de cada entidade.

As definições de dados de corte, dados de séries temporais e dados de painel estão resumidas no quadro Conceito-Chave 1.1.

Resumo

1. Muitas decisões em negócios e em economia requerem estimativas quantitativas de como uma mudança em uma variável afeta outra variável.
2. Conceitualmente, a forma de estimar um efeito causal é por meio de um experimento controlado aleatório, mas conduzir tais experimentos para aplicações econômicas é freqüentemente antiético, complexo e extremamente caro.
3. A econometria proporciona ferramentas para estimar efeitos causais utilizando tanto dados observacionais (não experimentais) quanto dados de experimentos imperfeitos do mundo real.
4. Dados de corte são coletados por meio da observação de diversas entidades em um único ponto no tempo; dados de séries temporais são coletados por meio da observação de uma única entidade em diversos pontos no tempo; e dados de painel são coletados por meio da observação de diversas entidades, cada qual observada em diversos pontos no tempo.

Termos-chave

experimento controlado aleatório (6)
grupo de controle (6)
grupo de tratamento (6)
efeito causal (6)
dados experimentais (7)
dados observacionais (7)
dados de corte (8)
número da observação (8)
dados de séries temporais (8)
dados de painel (9)
dados longitudinais (9)

Revisão dos Conceitos

1.1 Elabore um experimento controlado aleatório ideal hipotético para estudar o efeito do número de horas de estudo sobre o desempenho em provas de microeconomia. Sugira alguns obstáculos para a implementação desse experimento na prática.

1.2 Elabore um experimento controlado aleatório ideal hipotético para estudar o efeito do uso de cintos de segurança sobre o número de mortes em acidentes de trânsito. Sugira alguns obstáculos para a implementação desse experimento na prática.

1.3 Solicitaram a você um estudo sobre a relação entre horas gastas com treinamento de funcionários (medidas em horas por trabalhador por semana) em uma fábrica e a produtividade de seus trabalhadores (produção por trabalhador por hora). Descreva:

a. um experimento controlado aleatório ideal para medir esse efeito causal;

b. uma base de dados observacionais de corte com a qual você poderia estudar esse efeito;

c. uma base de dados observacionais de séries temporais com a qual você poderia estudar esse efeito e

d. uma base de dados observacionais de painel com a qual você poderia estudar esse efeito.

CAPÍTULO 2

Revisão de Probabilidade

Este capítulo traz uma revisão das idéias centrais da teoria da probabilidade necessárias para a compreensão da análise de regressão e da econometria. Nós presumimos que você tenha cursado uma disciplina de introdução à probabilidade e estatística. Se o seu conhecimento de probabilidade estiver defasado, você deve reforçá-lo lendo este capítulo. Caso se sinta seguro com o material, ainda assim deve percorrer rapidamente o capítulo e os termos e conceitos no final para se assegurar de que está familiarizado com as idéias e as notações.

A maioria dos aspectos do mundo que nos cerca tem um elemento de aleatoriedade. A teoria da probabilidade nos proporciona instrumentos matemáticos para quantificar e descrever essa aleatoriedade. Na Seção 2.1 fazemos uma revisão das distribuições de probabilidade para uma única variável aleatória e na Seção 2.2 falamos sobre a expectativa matemática,* a média e a variância de uma única variável aleatória. A maioria dos problemas interessantes em economia envolve mais de uma variável; assim, na Seção 2.3, apresentamos os elementos básicos da teoria da probabilidade para duas variáveis aleatórias. Na Seção 2.4, discutimos três distribuições de probabilidade especiais que desempenham um papel central em estatística e em econometria: as distribuições normal, qui-quadrado e $F_{m,\infty}$.

As duas últimas seções deste capítulo se concentram em uma fonte específica de aleatoriedade de importância fundamental na econometria: a aleatoriedade originada da seleção aleatória de uma amostra de dados com base em uma população maior. Por exemplo, suponha que você observe dez indivíduos recém-formados na universidade selecionados ao acaso, registre (ou "observe") seus ganhos e calcule o ganho médio utilizando esses dez dados (ou "observações"). Como você escolheu a amostra aleatoriamente, poderia ter selecionado dez outros recém-formados por mero acaso; se tivesse feito isso, teria observado dez ganhos diferentes e calculado uma média da amostra diferente. Como o ganho médio varia de uma amostra escolhida aleatoriamente para outra, a média da amostra é em si uma variável aleatória. Portanto, a média da amostra tem uma distribuição de probabilidade, que é chamada de distribuição amostral** da média da amostra porque descreve diferentes valores possíveis para a média da amostra que poderiam ter ocorrido se uma amostra diferente tivesse sido selecionada.

Na Seção 2.5, discutimos a amostragem aleatória e a distribuição amostral da média da amostra. A distribuição amostral é, geralmente, complicada. Entretanto, quando o tamanho da amostra é grande o bastante, a distribuição amostral da média da amostra é aproximadamente normal, um resultado conhecido como teorema central do limite, que será discutido na Seção 2.6.

2.1 Variáveis Aleatórias e Distribuições de Probabilidade

Probabilidades, Espaço Amostral e Variáveis Aleatórias

Probabilidades e resultados. O sexo da próxima pessoa que você vai conhecer, sua nota em uma prova e o número de vezes que seu computador vai travar enquanto você estiver digitando um trabalho têm um elemento de acaso ou aleatoriedade. Em cada um desses exemplos, há algo ainda desconhecido que será revelado oportunamente.

* A expressão tradicional "esperança matemática" normalmente é utilizada como tradução de *mathematical expectation* em cursos de estatística em virtude de seu uso difundido nessa área. No contexto da teoria macroeconômica, no entanto, a mesma expressão seria traduzida como "expectativa matemática", bastando lembrar, por exemplo, a literatura de modelos com expectativas racionais. Aqui optamos pela segunda tradução, lembrando o leitor de que ele pode utilizar a primeira se assim preferir (N. do R.T.).

** A tradução mais utilizada em estatística para *sampling distribution* é "distribuição amostral", que adotamos no texto. No entanto, uma tradução mais próxima do original seria "distribuição da amostragem" (N. do R.T.).

Ocorrências potenciais mutuamente exclusivas de um processo aleatório são chamadas de **resultados**. Por exemplo, pode ser que seu computador nunca trave, pode ser que trave uma vez, duas vezes e assim por diante. Apenas um desses resultados irá de fato ocorrer (os resultados são mutuamente exclusivos) e os resultados não precisam ser igualmente prováveis.

A **probabilidade** de um resultado é a proporção do tempo em que tal resultado ocorre no longo prazo. Se a probabilidade de seu computador não travar enquanto você digita um trabalho é de 80 por cento, na digitação de diversos trabalhos você concluirá 80 por cento deles sem que o computador trave.

Espaço amostral e eventos. O conjunto de todos os resultados possíveis é chamado de **espaço amostral**. Um **evento** é um subconjunto do espaço amostral, isto é, é um conjunto de um ou mais resultados. O evento "meu computador não vai travar mais de uma vez" é o conjunto que consiste em dois resultados: "não travar" e "travar uma vez".

Variáveis aleatórias. Uma variável aleatória é uma síntese numérica de um resultado aleatório. O número de vezes que seu computador trava enquanto você digita um trabalho é aleatório e assume um valor numérico, portanto é uma variável aleatória.

Algumas variáveis aleatórias são discretas, outras são contínuas. Conforme seus nomes sugerem, uma **variável aleatória discreta** assume somente um conjunto discreto de valores, como 0, 1, 2, ..., ao passo que uma **variável aleatória contínua** assume um conjunto contínuo de valores possíveis.

Distribuição de Probabilidade de uma Variável Aleatória Discreta

Distribuição de probabilidade. A **distribuição de probabilidade** de uma variável aleatória discreta é a lista de todos os valores possíveis para a variável e a probabilidade de que cada valor venha a ocorrer. A soma dessas probabilidades é um.

Por exemplo, seja M o número de vezes que seu computador trava enquanto você digita um trabalho. A distribuição de probabilidade da variável aleatória M é a lista das probabilidades de cada resultado possível: a probabilidade de $M = 0$, representada por $P(M = 0)$, é a probabilidade de o computador nunca travar; $P(M = 1)$ é a probabilidade de o computador travar somente uma vez; e assim por diante. Um exemplo de distribuição de probabilidade para M é dado na segunda linha da Tabela 2.1; nessa distribuição, se o seu computador travar quatro vezes, você desistirá e escreverá seu trabalho à mão. De acordo com essa distribuição, a probabilidade de o computador nunca travar é de 80 por cento, a probabilidade de travar uma vez é de 10 por cento e a probabilidade de travar duas, três ou quatro vezes é de 6, 3 e 1 por cento, respectivamente. Essas probabilidades somam 100 por cento. Essa distribuição de probabilidade é mostrada na Figura 2.1.

Probabilidades de eventos. A probabilidade de um evento pode ser calculada a partir da distribuição de probabilidade. Por exemplo, a probabilidade do evento de o computador travar uma ou duas vezes é a soma das probabilidades dos resultados individuais. Isto é, $P(M = 1 \text{ ou } M = 2) = P(M = 1) + P(M = 2) = 0{,}10 + 0{,}06 = 0{,}16$, ou 16 por cento.

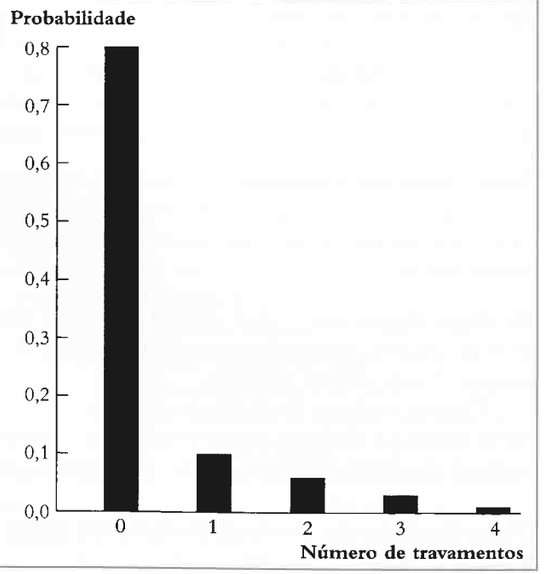
Distribuição de probabilidade acumulada. A **distribuição de probabilidade acumulada** é a probabilidade de que a variável aleatória seja menor ou igual a um determinado valor. A última linha da Tabela 2.1

TABELA 2.1 Probabilidade de Seu Computador Travar *M* Vezes

	Resultado (número de travamentos)				
	0	1	2	3	4
Distribuição de probabilidade	0,80	0,10	0,06	0,03	0,01
Distribuição de probabilidade acumulada	0,80	0,90	0,96	0,99	1,00

FIGURA 2.1 Distribuição de Probabilidade do Número de Travamentos do Computador

A altura de cada barra é a probabilidade de que o computador trave o número de vezes indicado. A altura da primeira barra é 0,80, portanto a probabilidade de 0 travamento é de 80 por cento. A altura da segunda barra é 0,1, portanto a probabilidade de 1 travamento é de 10 por cento e assim por diante para as outras barras.



mostra a distribuição de probabilidade acumulada da variável aleatória M. Por exemplo, a probabilidade de que ocorra no máximo um travamento, $P(M \leq 1)$, é de 90 por cento, que corresponde à soma da probabilidade de nenhum travamento (80 por cento) com a de um travamento (10 por cento).

Uma distribuição de probabilidade acumulada é também chamada de **função de distribuição acumulada**, **f.d.a.** ou **distribuição acumulada**.

Distribuição de Bernoulli. Um caso especial importante de variável aleatória discreta ocorre quando a variável aleatória é binária, isto é, os resultados são 0 ou 1. Uma variável aleatória binária é chamada de **variável aleatória de Bernoulli** (em homenagem a Jacob Bernoulli, matemático e cientista suíço do século XVII) e sua distribuição de probabilidade é chamada de **distribuição de Bernoulli**.

Por exemplo, seja G o sexo da próxima pessoa que você conhecer, em que $G = 0$ indica que a pessoa é do sexo masculino e $G = 1$ que é do sexo feminino. Os resultados de G e suas probabilidades são, portanto

$$G = \begin{cases} 1 \text{ com probabilidade } p \\ 0 \text{ com probabilidade } 1 - p, \end{cases} \tag{2.1}$$

em que p é a probabilidade de que a próxima pessoa que você conhecer seja uma mulher. A distribuição de probabilidade na Equação (2.1) é a distribuição de Bernoulli.

Distribuição de Probabilidade de uma Variável Aleatória Contínua

Distribuição de probabilidade acumulada. A distribuição de probabilidade acumulada para uma variável contínua é definida como no caso da variável aleatória discreta. Ou seja, a distribuição de probabilidade acumulada de uma variável aleatória contínua é a probabilidade de que a variável aleatória seja menor ou igual a dado valor.

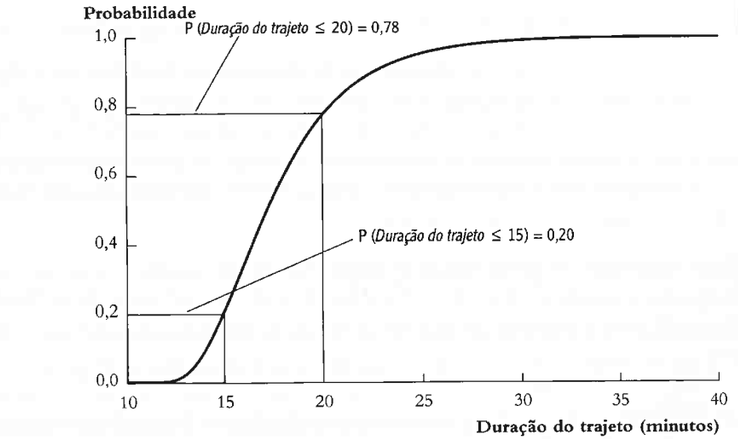
Por exemplo, considere uma aluna que dirija de sua casa até a escola. A duração do trajeto pode assumir um conjunto contínuo de valores e é natural que seja tratada como uma variável aleatória contínua, uma vez que depende de fatores aleatórios, como as condições meteorológicas e de trânsito. A Figura 2.2a mostra uma distribuição acumulada hipotética das durações do trajeto. Por exemplo, a probabilidade de que o trajeto leve menos do que 15 minutos é de 20 por cento e a probabilidade de que leve menos do que 20 minutos é de 78 por cento.

Função densidade de probabilidade. Como uma variável aleatória contínua pode assumir um conjunto contínuo de valores possíveis, a distribuição de probabilidade usada para variáveis discretas, que lista a probabilidade de cada valor possível da variável aleatória, não é apropriada para variáveis contínuas. Em vez disso, a probabilidade é resumida pela **função densidade de probabilidade**. A área sob a função densidade de probabilidade entre dois pontos quaisquer é a probabilidade de que a variável aleatória esteja entre esses dois pontos. Uma função densidade de probabilidade é também chamada de **f.d.p.**, **função densidade** ou simplesmente **densidade**.

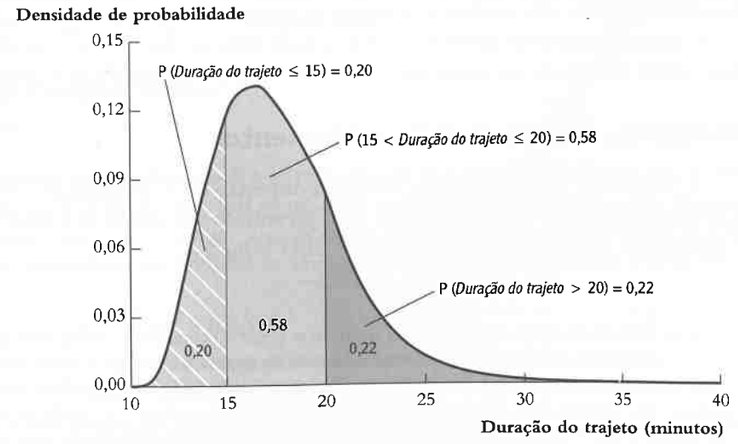
A Figura 2.2b mostra a função densidade de probabilidade para as durações do trajeto correspondentes à distribuição acumulada na Figura 2.2a. A probabilidade de que o trajeto leve entre 15 e 20 minutos é dada pela área sob a f.d.p. entre 15 e 20 minutos, que é 0,58 ou 58 por cento. De modo equivalente, essa probabilidade pode ser vista na distribuição acumulada da Figura 2.2a como a diferença entre a probabilidade de que o trajeto leve menos de 20 minutos (78 por cento) e a probabilidade de que leve menos de 15 minutos (20 por cento). Assim, a função densidade de probabilidade e a distribuição de probabilidade acumulada mostram a mesma informação em formatos diferentes.

FIGURA 2.2 Distribuição Acumulada e Funções Densidade de Probabilidade da Duração do Trajeto

A Figura 2.2a mostra a distribuição de probabilidade acumulada (ou f.d.a.) das durações do trajeto. A probabilidade de que a duração do trajeto seja menor do que 15 minutos é de 0,20 (ou 20 por cento) e a probabilidade de que seja menor do que 20 minutos é de 0,78 (78 por cento). A Figura 2.2b mostra a função densidade de probabilidade (ou f.d.p.) das durações do trajeto. As probabilidades são dadas pelas áreas sob a f.d.p. A probabilidade de que a duração do trajeto esteja entre 15 e 20 minutos é de 0,58 (58 por cento) e é dada pela área sob a curva entre 15 e 20 minutos.



(a) Função de distribuição acumulada da duração do trajeto



(b) Função densidade de probabilidade da duração do trajeto

2.2 Valores Esperados, Média e Variância

Valor Esperado de uma Variável Aleatória

Valor esperado. O **valor esperado** de uma variável aleatória Y, representado por $E(Y)$, é o valor médio de longo prazo da variável aleatória obtido por meio de muitos experimentos ou ocorrências repetidos.

O valor esperado de uma variável aleatória discreta é calculado como uma média ponderada dos resultados possíveis dessa variável aleatória, em que os pesos são as probabilidades desse resultado. O valor esperado de Y também é chamado de **expectativa** de Y ou **média** de Y e é representado por μ_Y.*

Por exemplo, suponha que você empreste US$ 100 a um amigo a juros de 10 por cento. Se o empréstimo for pago, você receberá US$ 110 (o principal de US$ 100 mais os juros de US$ 10), porém existe um risco de 1 por cento de seu amigo ficar inadimplente e você não receber nada. Desse modo, o montante a ser pago para você é uma variável aleatória igual a US$ 110 com probabilidade 0,99 e igual a US$ 0 com probabilidade 0,01. Ao longo de muitos empréstimos, em 99 por cento dos casos você receberia US$ 110, mas em 1 por cento dos casos não receberia nada; desse modo, na média, você receberia US$ $110 \times 0{,}99 +$ US$ $0 \times 0{,}01 =$ US$ 108,90. Assim, o valor esperado de seu pagamento (ou o "pagamento médio") é de US$ 108,90.

Como um segundo exemplo, considere o número de vezes M que seu computador trava, com a distribuição de probabilidade dada pela Tabela 2.1. O valor esperado de M é o número médio de travamentos ao longo de muitos trabalhos, ponderados pela freqüência com que um dado número de travamentos ocorre. Dessa forma,

$$E(M) = 0 \times 0{,}80 + 1 \times 0{,}10 + 2 \times 0{,}06 + 3 \times 0{,}03 + 4 \times 0{,}01 = 0{,}35 \tag{2.2}$$

isto é, o número esperado de travamentos do computador enquanto você digita um trabalho é 0,35. Obviamente, o número efetivo de travamentos sempre deve ser um número inteiro; não faz sentido dizer que seu computador travou 0,35 vezes durante a digitação de determinado trabalho! Em vez disso, o cálculo na Equação (2.2) significa que o número médio de travamentos ao longo de muitos trabalhos é 0,35.

A fórmula para o valor esperado de uma variável aleatória discreta Y que pode assumir k valores diferentes é dada no quadro Conceito-Chave 2.1.

Valor esperado de uma variável aleatória de Bernoulli. Um caso especial importante da fórmula geral do quadro Conceito-Chave 2.1 é a média de uma variável aleatória de Bernoulli. Seja G a variável aleatória de Bernoulli com a distribuição de probabilidade da Equação (2.1). O valor esperado de G é

$$E(G) = 1 \times p + 0 \times (1 - p) = p. \tag{2.3}$$

Portanto, o valor esperado de uma variável aleatória de Bernoulli é p, a probabilidade de que ela assuma o valor "1".

Valor esperado de uma variável aleatória contínua. O valor esperado de uma variável aleatória contínua é também a média dos resultados possíveis da variável aleatória ponderada pelas respectivas probabilidades.

Como uma variável aleatória contínua pode assumir um conjunto contínuo de valores possíveis, a definição matemática formal de sua expectativa envolve cálculo e sua definição é dada no Apêndice 15.1.

Variância, Desvio Padrão e Momentos

A variância e o desvio padrão medem a dispersão ou a "propagação" de uma distribuição de probabilidade. A **variância** de uma variável aleatória Y, representada por var(Y), é o valor esperado do quadrado do desvio de Y em relação à sua média, isto é, var$(Y) = E[(Y - \mu_Y)^2]$.

* Há uma distinção no texto original que na tradução se perde. O conceito de média nesse parágrafo é a tradução dos termos *mean* e *population mean*. Esses termos estão relacionados à média da população, ou seja, à média dos valores possíveis da variável aleatória ponderada pelas respectivas probabilidades (veja o Conceito-Chave 2.1). Vemos também no texto a palavra "média" como tradução de *average*. Um exemplo desse último uso, bastante freqüente neste capítulo, é a "média da amostra", tradução de *sample average*. O termo *average* refere-se neste caso à média aritmética obtida pelos valores coletados em uma amostra (veja a Seção 2.5). Os dois conceitos são diferentes e por isso se recomenda ao leitor cuidado durante a leitura (N. do R.T.).

Conceito-Chave 2.1

Valor Esperado e Média

Suponha que a variável aleatória Y assuma k valores possíveis, $y_1, \ldots, y_k$, onde y_1 representa o primeiro valor, y_2 o segundo etc., e que a probabilidade de Y assumir y_1 seja p_1, a probabilidade de Y assumir y_2 seja p_2, e assim por diante. O valor esperado de Y, representado por $E(Y)$, é

$$E(Y) = y_1 p_1 + y_2 p_2 + \cdots + y_k p_k = \sum_{i=1}^{k} y_i p_i, \tag{2.4}$$

onde a notação "$\sum_{i=1}^{k} y_i p_i$" significa "a soma de $y_i p_i$ para i variando de 1 a k". O valor esperado de Y é também chamado de média de Y ou expectativa de Y e é representado por μ_Y.

Como a variância envolve o quadrado de Y, a dimensão da variância é a dimensão do quadrado de Y, o que torna o problema de difícil interpretação. É, portanto, comum medir a dispersão pelo **desvio padrão**, que é a raiz quadrada da variância e é representado por σ_Y. O desvio padrão tem a mesma dimensão de Y. Essas definições estão resumidas no Conceito-Chave 2.2.

Por exemplo, a variância do número de travamentos do computador M é a média do quadrado da diferença entre M e sua média, 0,35, ponderada pelas respectivas probabilidades:

$$\begin{aligned}\text{var}(M) = {} & (0 - 0{,}35)^2 \times 0{,}80 + (1 - 0{,}35)^2 \times 0{,}10 + (2 - 0{,}35)^2 \times 0{,}06 \\ & + (3 - 0{,}35)^2 \times 0{,}03 + (4 - 0{,}35)^2 \times 0{,}01 = 0{,}6475.\end{aligned} \tag{2.5}$$

O desvio padrão de M é a raiz quadrada da variância, portanto $\sigma_M = \sqrt{0{,}6475} \cong 0{,}80$.

Variância de uma variável aleatória de Bernoulli. A média da variável aleatória de Bernoulli G com a distribuição de probabilidade dada pela Equação (2.1) é $\mu_G = p$ (Equação (2.3)) e, portanto, sua variância é

$$\text{var}(G) = \sigma_G^2 = (0 - p)^2 \times (1 - p) + (1 - p)^2 \times p = p(1 - p). \tag{2.6}$$

Dessa forma, o desvio padrão de uma variável aleatória de Bernoulli é $\sigma_G = \sqrt{p(1 - p)}$.

Momentos. A média de Y, $E(Y)$, é também chamada de primeiro momento de Y, e o valor esperado do quadrado de Y, $E(Y^2)$, também é chamado de segundo momento de Y. Em geral, o valor esperado de Y^r é chamado de ***momento r-ésimo*** da variável aleatória Y. Isto é, o momento r-ésimo de Y é $E(Y^r)$.

Assim como a média é uma medida do centro de uma distribuição e o desvio padrão é uma medida de sua dispersão, os momentos com $r > 2$ medem outros aspectos da forma de uma distribuição. Neste livro, momentos mais elevados de distribuições (momentos com $r > 2$) são usados principalmente nas hipóteses matemáticas e nas deduções que embasam os procedimentos estatísticos e econométricos.

Média e Variância de uma Função Linear de uma Variável Aleatória

Nesta seção, discutiremos variáveis aleatórias (digamos X e Y) relacionadas por uma função linear. Por exemplo, considere uma estrutura de alíquotas de imposto de renda na qual um trabalhador é tributado a uma alíquota de 20 por cento sobre seus rendimentos para então ganhar uma bolsa de estudos (isenta de impostos) de US$ 2.000. Sob essa estrutura de imposto, os rendimentos líquidos de impostos Y estão relacionados aos rendimentos brutos X pela equação

$$Y = 2.000 + 0{,}8X. \tag{2.7}$$

Isto é, os rendimentos líquidos Y são 80 por cento dos rendimentos brutos X, mais US$ 2.000.

Suponha que os rendimentos brutos de uma pessoa no próximo ano sejam uma variável aleatória com média μ_X e variância σ_X^2. Como os rendimentos brutos são aleatórios, os rendimentos líquidos também são. Sob esse imposto, qual será a média e o desvio padrão dos rendimentos líquidos? Depois dos impostos, seus rendimentos

Conceito-Chave 2.2

Variância e Desvio Padrão

A variância da variável aleatória discreta Y, representada por σ_Y^2, é

$$\sigma_Y^2 = \text{var}(Y) = E[(Y-\mu_Y)^2] = \sum_{i=1}^{k} (y_i - \mu_Y)^2 p_i. \quad (2.8)$$

O desvio padrão de Y é σ_Y, a raiz quadrada da variância. A dimensão do desvio padrão é a mesma dimensão de Y.

são 80 por cento dos rendimentos brutos originais, mais US$ 2.000. Desse modo, o valor esperado de seus rendimentos líquidos é

$$E(Y) = \mu_Y = 2.000 + 0{,}8\mu_X. \quad (2.9)$$

A variância dos rendimentos líquidos é o valor esperado de $(Y-\mu_Y)^2$. Como $Y = 2.000 + 0{,}8X$, temos que $Y - \mu_Y = 2.000 + 0{,}8X - (2.000 + 0{,}8\,\mu_X) = 0{,}8\,(X-\mu_X)$. Então, $E[(Y-\mu_Y)^2] = E\{[0{,}8(X-\mu_X)]^2\} = 0{,}64\,E[(X-\mu_X)^2]$. Segue-se que $\text{var}(Y) = 0{,}64\text{var}(X)$, logo, tirando-se a raiz quadrada da variância, obtém-se que o desvio padrão de Y é

$$\sigma_Y = 0{,}8\sigma_X. \quad (2.10)$$

Ou seja, o desvio padrão da distribuição de seus rendimentos líquidos é 80 por cento do desvio padrão da distribuição dos rendimentos brutos.

Essa análise pode ser generalizada de tal forma que Y dependa de X com um intercepto a (em vez de US$ 2.000) e uma declividade b (em vez de 0,8), de modo que

$$Y = a + bX. \quad (2.11)$$

Então, a média e a variância de Y são

$$\mu_Y = a + b\mu_X \text{ e} \quad (2.12)$$

$$\sigma_Y^2 = b^2\sigma_X^2, \quad (2.13)$$

e o desvio padrão de Y é $\sigma_Y = b\sigma_X$. As expressões nas equações (2.9) e (2.10) são aplicações das fórmulas mais gerais das equações (2.12) e (2.13) com $a = 2.000$ e $b = 0{,}8$.

2.3 Duas Variáveis Aleatórias

A maioria das questões interessantes em economia envolve duas ou mais variáveis. Será que pessoas com nível superior completo têm mais chance de conseguir um emprego do que aquelas que não possuem curso superior? Como a distribuição de renda entre as mulheres se compara com a distribuição de renda entre os homens? Essas questões dizem respeito à distribuição de duas variáveis aleatórias consideradas em conjunto (nível de instrução e situação empregatícia no primeiro exemplo, renda e sexo no segundo). Responder a essas questões requer uma compreensão dos conceitos de distribuição de probabilidade conjunta, marginal e condicional.

Distribuições Conjuntas e Distribuições Marginais

Distribuição de probabilidade conjunta. A **distribuição de probabilidade conjunta** de duas variáveis aleatórias discretas, digamos X e Y, é a probabilidade de que as variáveis aleatórias assumam simultaneamente determinados valores, digamos x e y. A soma das probabilidades de todas as combinações possíveis (x,y) é um. A distribuição de probabilidade conjunta pode ser escrita como a função $P(X = x,\ Y = y)$.

Por exemplo, as condições meteorológicas — se vai ou não chover — afetam a duração do trajeto da aluna da Seção 2.1. Seja Y uma variável aleatória binária que é igual a um se o trajeto for curto (menos de 20 minutos) e igual a zero se não for esse o caso, e seja X uma variável aleatória binária que é igual a zero se estiver chovendo e igual a um se não estiver. Para essas duas variáveis aleatórias há quatro resultados possíveis: chove e o trajeto é longo ($X = 0,\ Y = 0$); chove e o trajeto é curto ($X = 0,\ Y = 1$); não chove e o trajeto é longo ($X = 1,\ Y = 0$); não chove e o trajeto é curto ($X = 1,\ Y = 1$). A distribuição de probabilidade conjunta é a freqüência com que cada um desses quatro resultados ocorre ao longo de muitos trajetos repetidos.

Um exemplo de uma distribuição conjunta dessas duas variáveis é dado na Tabela 2.2. De acordo com essa distribuição, ao longo de muitos trajetos, há chuva e um trajeto longo em 15 por cento dos dias ($X = 0, Y = 0$), isto é, a probabilidade de um trajeto longo com chuva é de 15 por cento, ou $P(X = 0, Y = 0) = 0{,}15$. Também temos $P(X = 0, Y = 1) = 0{,}15$, $P(X = 1, Y = 0) = 0{,}07$ e $P(X = 1, Y = 1) = 0{,}63$. Esses quatro resultados possíveis são mutuamente exclusivos e constituem o espaço amostral de modo que a soma das quatro probabilidades seja igual a um.

Distribuição de probabilidade marginal. A **distribuição de probabilidade marginal** de uma variável aleatória Y é somente outro nome para sua distribuição de probabilidade. Esse termo é usado para distinguir a distribuição de Y sozinha (a distribuição marginal) da distribuição conjunta de Y e outra variável aleatória.

A distribuição marginal de Y pode ser calculada a partir da distribuição conjunta de X e Y somando-se as probabilidades de todos os resultados possíveis para os quais Y assume um valor específico. Se X pode assumir l valores diferentes, $x_1, \ldots, x_l$, então a probabilidade marginal de que Y assuma o valor y é

$$P(Y = y) = \sum_{i=1}^{l} P(X = x_i,\ Y = y). \quad (2.14)$$

Por exemplo, na Tabela 2.2 a probabilidade de um trajeto longo com chuva é de 15 por cento e a probabilidade de um trajeto longo sem chuva é de 7 por cento, logo a probabilidade de um trajeto longo (com ou sem chuva) é de 22 por cento. A distribuição marginal das durações do trajeto é dada na última coluna da Tabela 2.2. Da mesma forma, a probabilidade marginal de que irá chover é de 30 por cento, conforme mostrado na última linha da Tabela 2.2.

Distribuições Condicionais

Distribuição condicional. A distribuição de uma variável aleatória Y condicional a outra variável aleatória X assumindo um valor específico é chamada de **distribuição condicional de Y dado X**. A probabilidade condicional de que Y assuma o valor y quando X assume o valor x é escrita como $P(Y = y \mid X = x)$.

Por exemplo, qual é a probabilidade de que um trajeto seja longo ($Y = 0$) se você sabe que está chovendo ($X = 0$)? Da Tabela 2.2, a probabilidade conjunta de um trajeto curto com chuva é de 15 por cento e a probabilidade conjunta de um trajeto longo com chuva é de 15 por cento; logo, se está chovendo, um trajeto curto e um trajeto longo são igualmente prováveis. Portanto, a probabilidade de que um trajeto seja longo ($Y = 0$), condicional ao fato de estar chovendo ($X = 0$), é de 50 por cento, ou $P(Y = 0 \mid X = 0) = 0{,}50$. De forma equivalente, a probabilidade marginal de chuva é 30 por cento; isto é, ao longo de muitos trajetos chove 30 por cento das vezes. Desses 30 por cento de trajetos, em 50 por cento das vezes o trajeto é longo (0,15/0,30).

TABELA 2.2 Distribuição Conjunta das Condições Meteorológicas e das Durações do Trajeto

	Com chuva ($X = 0$)	Sem chuva ($X = 1$)	**Total**
Trajeto longo ($Y = 0$)	0,15	0,07	0,22
Trajeto curto ($Y = 1$)	0,15	0,63	0,78
Total	0,30	0,70	1,00

Em geral, a distribuição condicional de Y dado $X = x$ é

$$P(Y = y|X = x) = \frac{P(X = x,\ Y = y)}{P(X = x)}. \tag{2.15}$$

Por exemplo, a probabilidade condicional de um trajeto longo, dado que está chovendo, é $P(Y = 0|X = 0) = P(X = 0,\ Y = 0)/\ P(X = 0) = 0{,}15/0{,}30 = 0{,}50$.

Como segundo exemplo, considere uma modificação do caso do computador que trava. Suponha que você utilize um computador da biblioteca para digitar seu trabalho e que uma bibliotecária selecione aleatoriamente para você um dos computadores disponíveis, dos quais metade é nova e metade é antiga. Como a seleção é feita aleatoriamente, a idade do computador que você utiliza, A (= 1, se o computador for novo; = 0, se for antigo) é uma variável aleatória. Suponha que a distribuição conjunta das variáveis aleatórias M e A seja dada na Parte A da Tabela 2.3. Então, a distribuição condicional de travamentos do computador, dada a idade do computador, é dada na Parte B da tabela. Por exemplo, a probabilidade conjunta $M = 0$ e $A = 0$ é 0,35; como a metade dos computadores é antiga, a probabilidade condicional de nenhum travamento, dado que você está usando um computador antigo, é $P(M = 0|A = 0) = P(M = 0,\ A = 0)/P(A = 0) = 0{,}35/0{,}50 = 0{,}70$, ou 70 por cento. Em contraste, a probabilidade condicional de nenhum travamento, dado que foi selecionado um computador novo, é de 90 por cento. Segundo as distribuições condicionais na Parte B da Tabela 2.3, os computadores mais novos têm menos probabilidade de travar do que os antigos; por exemplo, a probabilidade de três travamentos é de 5 por cento para um computador antigo, mas de 1 por cento para um computador novo.

Expectativa condicional. A **expectativa condicional de Y dado X**, também chamada de **média condicional de Y dado X**, é a média da distribuição condicional de Y dado X. Isto é, a expectativa condicional é o valor esperado de Y, calculado utilizando a distribuição condicional de Y dado X. Se Y assumir k valores, y_1, ..., y_k, a média condicional de Y dado $X = x$ será

$$E(Y|X = x) = \sum_{i=1}^{k} y_i\, P(Y = y_i|X = x). \tag{2.16}$$

Por exemplo, com base nas distribuições condicionais da Tabela 2.3, o número esperado de travamentos do computador, dado um computador antigo, é $E(M|A = 0) = 0 \times 0{,}70 + 1 \times 0{,}13 + 2 \times 0{,}10 + 3 \times 0{,}05 + 4 \times 0{,}02 = 0{,}56$. O número esperado de travamentos do computador, dado um computador novo, é $E(M|A = 1) = 0{,}14$, menor do que para os computadores antigos.

A expectativa condicional de Y dado $X = x$ é apenas o valor médio de Y quando $X = x$. No exemplo da Tabela 2.3, o número médio de travamentos para computadores antigos é 0,56, portanto a expectativa condicional de Y, dado um computador antigo, é de 0,56 travamentos. Da mesma forma, entre os computadores novos, o número médio de travamentos é 0,14, isto é, a expectativa condicional de Y, dado um computador novo, é 0,14.

TABELA 2.3 Distribuições Conjunta e Condicional de Travamentos do Computador (M) e Idade do Computador (A)

A. Distribuição conjunta

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
Computador antigo ($A = 0$)	0,35	0,065	0,05	0,025	0,01	0,50
Computador novo ($A = 1$)	0,45	0,035	0,01	0,005	0,00	0,50
Total	0,8	0,1	0,06	0,03	0,01	1,00

B. Distribuições condicionais de M dado A

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
$P(M\|A = 0)$	0,70	0,13	0,10	0,05	0,02	1,00
$P(M\|A = 1)$	0,90	0,07	0,02	0,01	0,00	1,00

Lei das expectativas iteradas. A média de Y é a média ponderada da expectativa condicional de Y dado X, utilizando como peso a distribuição de probabilidade de X. Por exemplo, a altura média dos adultos é a média ponderada da altura média dos homens e da altura média das mulheres, utilizando como pesos as proporções de homens e de mulheres. Expressando matematicamente, se X assumir l valores, x_1, ... , x_l, então

$$E(Y) = \sum_{i=1}^{l} E(Y|X = x_i)P(X = x_i). \tag{2.17}$$

A Equação (2.17) resulta das equações (2.16) e (2.15) (veja o Exercício 2.9).

Expressa de outra forma, a expectativa de Y é a expectativa da expectativa condicional de Y dado X, isto é,

$$E(Y) = E[E(Y|X)], \tag{2.18}$$

onde a expectativa entre colchetes no lado direito da Equação (2.18) é calculada usando a distribuição condicional de Y dado X e a expectativa fora do colchete é calculada usando a distribuição marginal de X. A Equação (2.18) é conhecida como **lei das expectativas iteradas**.

Por exemplo, o número médio de travamentos M é a média ponderada da expectativa condicional de M, dado um computador antigo, e da expectativa condicional de M, dado um computador novo; portanto, $E(M) = E(M|A = 0) \times P(A = 0) + E(M|A = 1) \times P(A = 1) = 0{,}56 \times 0{,}50 + 0{,}14 \times 0{,}50 = 0{,}35$. Essa é a média da distribuição marginal de M, conforme calculada na Equação (2.2).

A lei das expectativas iteradas implica que, se a média condicional de Y dado X é zero, a média de Y é zero. Isso é uma conseqüência imediata da Equação (2.18): se $E(Y|X) = 0$, então $E(Y) = E[E(Y|X)] = E[0] = 0$. Dito de outra forma, se a média de Y dado X é zero, a média ponderada dessas médias condicionais utilizando como pesos as respectivas probabilidades é zero, isto é, a média de Y deve ser zero.

Variância condicional. A **variância de Y condicional a X** é a variância da distribuição condicional de Y dado X. Expressa matematicamente, a variância condicional de Y dado X é

$$\text{var}(Y|X = x) = \sum_{i=1}^{k} [y_i - E(Y|X = x)]^2\, P(Y = y_i|X = x). \tag{2.19}$$

Por exemplo, a variância condicional do número de travamentos, dado um computador antigo, é $\text{var}(M|A = 0) = (0 - 0{,}56)^2 \times 0{,}70 + (1 - 0{,}56)^2 \times 0{,}13 + (2 - 0{,}56)^2 \times 0\ ,10 + (3 - 0{,}56)^2 \times 0{,}05 + (4 - 0{,}56)^2 \times 0{,}02 \cong 0{,}99$. O desvio padrão da distribuição condicional de M dado $A = 0$ é, portanto, $\sqrt{0{,}99} = 0{,}99$. A variância condicional de M dado $A = 1$ é a variância da distribuição na segunda linha da Tabela 2.3, que é 0,22; assim, o desvio padrão de M para computadores novos é $\sqrt{0{,}22} = 0{,}47$. Para as distribuições condicionais da Tabela 2.3, o número esperado de travamentos para computadores novos (0,14) é menor do que para computadores antigos (0,56) e a dispersão da distribuição do número de travamentos, conforme medida pelo desvio padrão condicional, é menor para computadores novos (0,47) do que para computadores antigos (0,99).

Independência

Duas variáveis aleatórias X e Y são **independentemente distribuídas**, ou **independentes**, se o conhecimento do valor de uma das variáveis não fornece nenhuma informação sobre a outra. Especificamente, X e Y são independentes se a distribuição condicional de Y dado X é igual à distribuição marginal de Y. Isto é, X e Y são independentemente distribuídas se, para todos os valores de x e y,

$$P(Y = y|X = x) = P(Y = y) \quad \text{(independência de } X \text{ e } Y\text{)}. \tag{2.20}$$

Substituindo a Equação (2.20) na Equação (2.15), temos uma expressão alternativa para variáveis aleatórias independentes em termos de sua distribuição conjunta. Se X e Y são independentes, então

$$P(X = x, Y = y) = P(X = x)P(Y = y). \tag{2.21}$$

Isto é, a distribuição conjunta de duas variáveis aleatórias independentes é o produto de suas distribuições marginais.

Co-variância e Correlação

Co-variância. Uma medida da extensão com que duas variáveis aleatórias movem-se juntas é a sua co-variância. A **co-variância** entre X e Y é o valor esperado $E[(X-\mu_X)(Y-\mu_Y)]$, onde μ_X é a média de X e μ_Y é a média de Y. A co-variância é representada por $\text{cov}(X,Y)$ ou por σ_{XY}. Se X pode assumir l valores e Y pode assumir k valores, a co-variância é dada pela fórmula

$$\text{cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j-\mu_X)(y_i-\mu_Y)\text{P}(X=x_j, Y=y_i). \quad (2.22)$$

Para interpretar essa fórmula, suponha que, quando X é maior do que sua média (de modo que $X-\mu_X$ é positivo), Y tende a ser maior do que sua média (de modo que $Y-\mu_Y$ é positivo) e que, quando X é menor que sua média (de modo que $X-\mu_X<0$), Y tende a ser menor do que sua média (de modo que $Y-\mu_Y<0$). Em ambos os casos, o produto $(X-\mu_X)(Y-\mu_Y)$ tende a ser positivo, logo a co-variância é positiva. Em contraste, se X e Y tendem a se mover em direções opostas (de modo que X é grande quando Y é pequeno e vice-versa), a co-variância é negativa. Finalmente, se X e Y são independentes, a co-variância é zero (veja o Exercício 2.9).

Correlação. Como a co-variância é o produto de X por Y, expressos em termos de desvios com relação a suas médias, sua dimensão é, inconvenientemente, a dimensão de X multiplicada pela dimensão de Y. Esse problema de "dimensão" pode fazer com que valores numéricos da co-variância sejam de difícil interpretação.

A correlação é uma medida alternativa de dependência entre X e Y que soluciona o problema de "dimensão" da co-variância. Especificamente, a **correlação** entre X e Y é a co-variância entre X e Y dividida por seus desvios padrão:

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}. \quad (2.23)$$

Como as dimensões do numerador da Equação (2.23) são iguais às do denominador, as dimensões se cancelam e a correlação é um número puro. As variáveis aleatórias X e Y são chamadas **não-correlacionadas** se $\text{corr}(X, Y) = 0$.

A correlação está sempre entre -1 e 1, ou seja, como está provado no Apêndice 2.1,

$$-1 \le \text{corr}(X,Y) \le 1 \text{ (desigualdade da correlação).} \quad (2.24)$$

Correlação e média condicional. Se a média condicional de Y não depende de X, então Y e X são não-correlacionadas, isto é

$$\text{se } E(Y|X) = \mu_Y \text{ então } \text{cov}(Y, X) = 0 \text{ e } \text{corr}(Y, X) = 0. \quad (2.25)$$

Demonstraremos agora esse resultado. Primeiro, suponha que X e Y tenham média zero, de modo que $\text{cov}(Y, X) = E[(Y-\mu_Y)(X-\mu_X)] = E(YX)$. De acordo com a lei das expectativas iteradas (Equação (2.18)), $E(YX) = E[E(Y|X)X] = 0$, pois $E(Y|X) = 0$; logo, $\text{cov}(Y,X) = 0$. A Equação (2.25) segue-se ao se substituir $\text{cov}(Y, X) = 0$ na definição de correlação da Equação (2.23). Se Y e X não tiverem média zero, primeiro subtraia suas médias para que a prova anterior seja aplicada.

Não é necessariamente verdade, entretanto, que, se X e Y são não-correlacionadas, a média condicional de Y dado X não depende de X. Dito de outra forma, é possível que a média condicional de Y seja uma função de X, embora X e Y sejam não-correlacionadas. O Exercício 2.10 nos fornece um exemplo.

Média e Variância de Somas de Variáveis Aleatórias

A média da soma de duas variáveis aleatórias, X e Y, é a soma de suas médias:

$$E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y \quad (2.26)$$

Médias, Variâncias e Co-variâncias de Somas de Variáveis Aleatórias

Conceito-Chave 2.3

Sejam X, Y e V variáveis aleatórias, sejam μ_X e σ_X^2 a média e a variância de X, seja σ_{XY} a co-variância entre X e Y (e assim por diante para as outras variáveis) e sejam a, b e c constantes. Esses fatos seguem-se das definições de média, variância e co-variância:

$$E(a+bX+cY) = a + b\mu_X + c\mu_Y, \quad (2.27)$$

$$\text{var}(a+bY) = b^2\sigma_Y^2, \quad (2.28)$$

$$\text{var}(aX+bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \quad (2.29)$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \quad (2.30)$$

$$\text{cov}(a+bX+cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \text{ e} \quad (2.31)$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y. \quad (2.32)$$

$$|\text{corr}(X,Y)| \le 1 \text{ e } |\sigma_{XY}| \le \sqrt{\sigma_X^2\sigma_Y^2} \text{ (desigualdade da correlação).} \quad (2.33)$$

A variância da soma de X e Y é a soma de suas variâncias, mais duas vezes sua co-variância:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}. \quad (2.34)$$

Se X e Y são independentes, então a co-variância é zero e a variância de sua soma é a soma de suas variâncias:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) = \sigma_X^2 + \sigma_Y^2 \text{ (se } X \text{ e } Y \text{ são independentes).} \quad (2.35)$$

O quadro Conceito-Chave 2.3 reúne expressões úteis para médias, variâncias e co-variâncias que envolvem somas ponderadas de variáveis aleatórias. Os resultados do quadro Conceito-Chave 2.3 são deduzidos no Apêndice 2.1.

2.4 Distribuições Normal, Qui-Quadrado, $F_{m,\infty}$ e t de Student

As distribuições de probabilidade mais freqüentemente encontradas em econometria são as distribuições normal, qui-quadrado, $F_{m,\infty}$ e t de Student.

Distribuição Normal

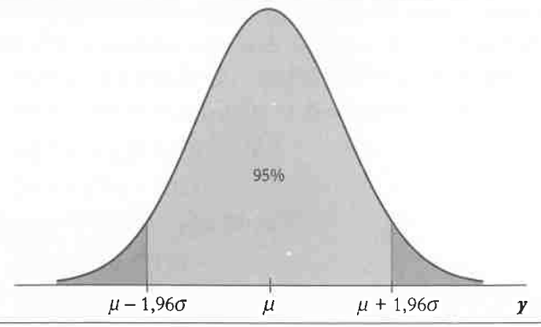
Uma variável aleatória contínua com uma **distribuição normal** tem a densidade de probabilidade com o formato familiar de sino mostrado na Figura 2.3. A função específica que define a densidade de probabilidade normal é dada no Apêndice 15.1. Como mostra a Figura 2.3, a densidade normal com média μ e variância σ^2 é simétrica em torno de sua média e tem 95 por cento de sua probabilidade entre $\mu - 1{,}96\sigma$ e $\mu + 1{,}96\sigma$.

Algumas notações e terminologias especiais foram desenvolvidas para a distribuição normal. A distribuição normal com média μ e variância σ^2 é chamada concisamente de "$N(\mu, \sigma^2)$". A **distribuição normal padrão** é a distribuição normal com média $\mu = 0$ e variância $\sigma^2 = 1$ e é representada por $N(0, 1)$. Variáveis aleatórias com distribuição $N(0, 1)$ são freqüentemente representadas por Z, e a função de distribuição acumulada normal padrão é representada pela letra grega Φ; dessa forma, $\text{P}(Z \le c) = \Phi(c)$, onde c é uma constante. Os valores da função de distribuição acumulada normal padrão estão na Tabela 1 do Apêndice.

Para o cálculo de probabilidades de uma variável normal com média e variância generalizadas, a variável deve ser **padronizada** subtraindo-se em primeiro lugar a média para em seguida dividir o resultado pelo desvio padrão. Por exemplo, suponha que Y seja distribuída como $N(1, 4)$, isto é, que Y seja uma variável normalmente distribuída com média 1 e variância 4. Qual é a probabilidade de que $Y \le 2$, isto é, qual é a área sombreada na Figura 2.4a? A versão padronizada de Y é Y menos sua média, dividido por seu desvio padrão, isto é, $(Y-1)/\sqrt{4} = \frac{1}{2}(Y-1)$. Assim, a variável aleatória $\frac{1}{2}(Y-1)$ é normalmente distribuída com média zero e variância um

FIGURA 2.3 Densidade de Probabilidade Normal

A função densidade de probabilidade normal com média μ e variância σ^2 é uma curva em forma de sino, centralizada em μ. A área sob a f.d.p. normal entre $\mu - 1{,}96\sigma$ e $\mu + 1{,}96\sigma$ é 0,95. A distribuição normal é representada por $N(\mu, \sigma^2)$.



(veja o Exercício 2.4); sua distribuição normal padrão é mostrada na Figura 2.4b. Agora, $Y \leq 2$ é equivalente a $\frac{1}{2}(Y-1) \leq \frac{1}{2}(2-1)$, isto é, $\frac{1}{2}(Y-1) \leq \frac{1}{2}$. Portanto,

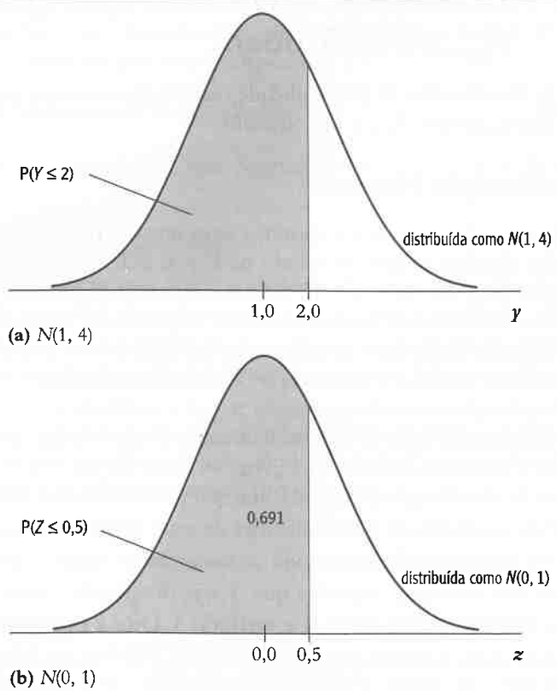
$$P(Y \leq 2) = P[\tfrac{1}{2}(Y-1) \leq \tfrac{1}{2}] = P(Z \leq \tfrac{1}{2}) = \Phi(0{,}5) = 0{,}691 \tag{2.36}$$

em que o valor 0,691 foi obtido da Tabela 1 do Apêndice.

A mesma abordagem pode ser aplicada para calcular a probabilidade de que uma variável aleatória normalmente distribuída exceda algum valor ou se encontre dentro de determinado intervalo. Os passos são resumidos no Conceito-Chave 2.4. O quadro "Um Dia Péssimo em Wall Street" apresenta uma aplicação incomum da distribuição normal acumulada.

FIGURA 2.4 Calculando a probabilidade de que $Y \leq 2$ quando Y é distribuída como $N(1, 4)$

Para calcular $P(Y \leq 2)$, padronize Y e então utilize a tabela de distribuição normal padrão. Y é padronizada pela subtração de sua média ($\mu = 1$) e pela divisão do resultado por seu desvio padrão ($\sigma_Y = 2$). A probabilidade de que $Y \leq 2$ é mostrada na Figura 2.4a e a probabilidade correspondente após a padronização de Y é mostrada na Figura 2.4b. Como a variável aleatória padronizada, $\frac{Y-1}{2}$, é uma variável aleatória (Z) normal padrão, $P(Y \leq 2) = P\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = P(Z \leq 0{,}5)$. Da Tabela 1 do Apêndice, $P(Z \leq 0{,}5) = 0{,}691$.



Conceito-Chave 2.4

Calculando Probabilidades Envolvendo Variáveis Aleatórias Normais

Suponha que Y seja normalmente distribuída, com média μ e variância σ^2, ou seja, Y é distribuída como $N(\mu, \sigma^2)$. Então, Y é padronizada subtraindo-se a sua média e dividindo o resultado por seu desvio padrão, isto é, calculando $Z = (Y - \mu)/\sigma$.

Sejam c_1 e c_2 dois números com $c_1 < c_2$, e sejam $d_1 = (c_1 - \mu)/\sigma$ e $d_2 = (c_2 - \mu)/\sigma$. Então,

$$P(Y \leq c_2) = P(Z \leq d_2) = \Phi(d_2), \tag{2.37}$$

$$P(Y \geq c_1) = P(Z \geq d_1) = 1 - \Phi(d_1), \text{ e} \tag{2.38}$$

$$P(c_1 \leq Y \leq c_2) = P(d_1 \leq Z \leq d_2) = \Phi(d_2) - \Phi(d_1). \tag{2.39}$$

A função distribuição acumulada normal Φ está na Tabela 1 do Apêndice.

Distribuição normal multivariada. A distribuição normal pode ser generalizada para descrever a distribuição conjunta de um conjunto de variáveis aleatórias. Nesse caso, a distribuição é chamada de **distribuição normal multivariada**, ou, se apenas duas variáveis estão sendo consideradas, de **distribuição normal bivariada**. A fórmula para a f.d.p. normal bivariada é dada no Apêndice 15.1, e a fórmula para a f.d.p. normal multivariada generalizada é dada no Apêndice 16.1.

A distribuição normal multivariada apresenta três propriedades importantes. Se X e Y possuem uma distribuição normal bivariada com co-variância σ_{XY} e se a e b são constantes, $aX + bY$ possui uma distribuição normal,

$$aX + bY \text{ é distribuída como } N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}) \quad (X, Y \text{ normal bivariada}). \tag{2.40}$$

De modo geral, se n variáveis aleatórias têm uma distribuição normal multivariada, qualquer combinação linear dessas variáveis (como sua soma) é normalmente distribuída.

Em segundo lugar, se um conjunto de variáveis tem uma distribuição normal multivariada, a distribuição marginal de cada variável é normal (isso resulta da Equação (2.40), quando colocamos $a = 1$ e $b = 0$).

Em terceiro lugar, se variáveis com distribuição normal multivariada têm co-variâncias iguais a zero, as variáveis são independentes. Portanto, se X e Y têm uma distribuição normal bivariada e $\sigma_{XY} = 0$, X e Y são independentes. Na Seção 2.3 dissemos que, se X e Y são independentes, então, seja qual for sua distribuição conjunta, $\sigma_{XY} = 0$. Se X e Y têm distribuição normal conjunta, então o inverso também é verdadeiro. Esse resultado — de que co-variância zero implica independência — é uma propriedade especial da distribuição normal multivariada que não é verdadeira para o caso geral.

Distribuições Qui-Quadrado e $F_{m,\infty}$

As distribuições qui-quadrado e $F_{m,\infty}$ são usadas para testar alguns tipos de hipóteses na estatística e na economia.

A **distribuição qui-quadrado** é a distribuição da soma de m variáveis aleatórias normais padrão independentes ao quadrado. Essa distribuição depende de m, que é chamado de graus de liberdade da distribuição qui-quadrado. Por exemplo, sejam Z_1, Z_2 e Z_3 variáveis aleatórias normais padrão independentes. Então, $Z_1^2 + Z_2^2 + Z_3^2$ possui uma distribuição qui-quadrado com três graus de liberdade. O nome dessa distribuição deriva da letra grega usada para representá-la: uma distribuição qui-quadrado com m graus de liberdade é representada por χ_m^2.

Percentis selecionados da distribuição χ_m^2 são fornecidos pela Tabela 3 do Apêndice. Por exemplo, essa tabela mostra que o 95º percentil da distribuição χ_3^2 é 7,81, de modo que $P(Z_1^2 + Z_2^2 + Z_3^2 \leq 7{,}81) = 0{,}95$.

Uma distribuição estreitamente relacionada é a distribuição $F_{m,\infty}$. A **distribuição $F_{m,\infty}$** é a distribuição de uma variável aleatória com uma distribuição qui-quadrado com m graus de liberdade, dividida por m. De forma equivalente, a distribuição $F_{m,\infty}$ é a distribuição da média de m variáveis aleatórias normais padrão ao quadrado. Por exemplo, se Z_1, Z_2 e Z_3 são variáveis aleatórias normais padrão independentes, então $(Z_1^2 + Z_2^2 + Z_3^2)/3$ possui uma distribuição $F_{3,\infty}$.

Um Dia Péssimo em Wall Street

Em um dia típico, o valor total de ações negociadas na bolsa de valores dos Estados Unidos pode subir ou cair 1 por cento ou mais. Isso é muito — mas não é nada se comparado ao que aconteceu no dia 19 de outubro de 1987, uma segunda-feira. Na "Segunda-feira Negra", o Dow Jones Industrial Average (uma média de 30 ações de indústrias grandes) despencou 25,6 por cento! De 1º de janeiro de 1980 a 16 de outubro de 1987, o desvio padrão dos retornos diários (isto é, a variação percentual diária dos preços) no Dow Jones foi de 1,16 por cento; desse modo, a queda de 25,6 por cento representou um retorno negativo de 22 (= 25,6/1,16) desvios padrão. A enormidade dessa queda pode ser vista na Figura 2.5, um gráfico dos retornos diários do Dow Jones durante a década de 1980.

Se os retornos de ações são normalmente distribuídos, a probabilidade de uma queda de no mínimo 22 desvios padrão é $P(Z \leq -22) = \Phi(-22)$. Você não encontrará esse valor na Tabela 1 do Apêndice, mas poderá calculá-lo usando um computador (tente!). Essa probabilidade é de $1{,}4 \times 10^{-107}$, isto é, 0,000 ... 00014, em que há um total de 106 zeros!

Em que medida $1{,}4 \times 10^{-107}$ é pequeno? Considere o seguinte:

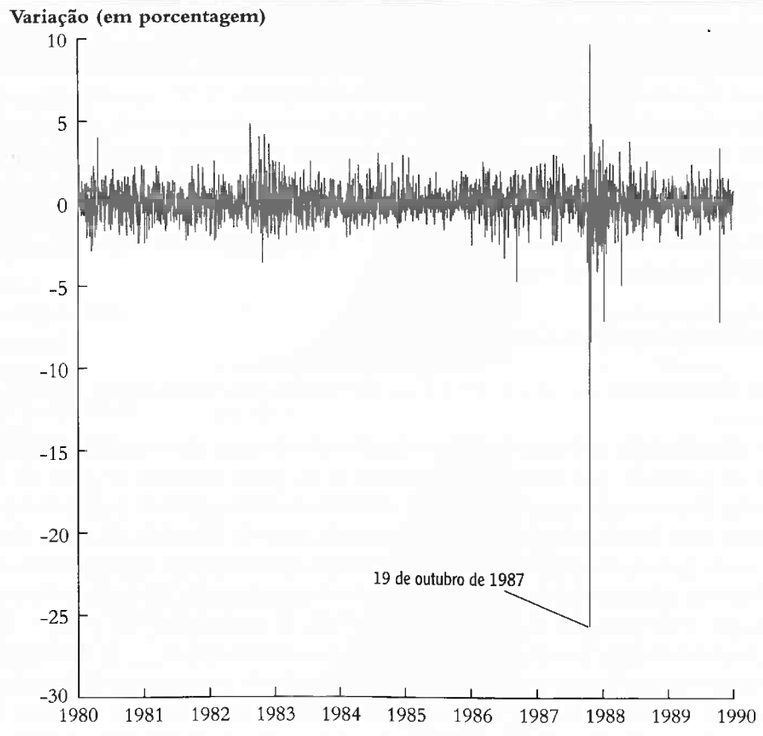
- A população mundial é de cerca de 6 bilhões, portanto a probabilidade de se ganhar em uma loteria entre todas as pessoas é de aproximadamente uma em 6 bilhões, ou 2×10^{-10}.
- Acredita-se que o universo exista há 15 bilhões de anos, ou cerca de 5×10^{17} segundos, portanto a probabilidade de se escolher ao acaso um segundo em particular dentre todos os segundos desde o surgimento do universo é de 2×10^{-18}.
- Existem aproximadamente 10^{43} moléculas de gás no primeiro quilômetro acima da superfície da Terra. A probabilidade de se escolher uma ao acaso é de 10^{-43}.

Embora realmente tenha sido um péssimo dia em Wall Street, o simples fato de sua ocorrência sugere que sua probabilidade foi maior do que $1{,}4 \times 10^{-107}$. Na verdade, o

(continua)

FIGURA 2.5 Variação Percentual Diária do Índice Dow Jones Industrial na Década de 1980

Durante a década de 1980, a variação percentual diária do índice Dow Jones foi de 0,05 por cento e seu desvio padrão foi de 1,16 por cento. Em 19 de outubro de 1987 — a "Segunda-feira Negra" —, o índice caiu 25,6 por cento, ou mais de 22 desvios padrão.



(continuação)

retorno sobre ações apresenta uma distribuição com caudas mais largas do que a distribuição normal; em outras palavras, existem mais dias com retornos grandes positivos ou negativos do que a distribuição normal sugeriria. No Capítulo 14 apresentaremos um modelo econométrico para retornos de ações utilizado por profissionais do mercado financeiro que é mais consistente com relação aos dias muito ruins — e muito bons — que vemos de fato em Wall Street.

Percentis selecionados da distribuição $F_{m,\infty}$ são fornecidos pela Tabela 4 do Apêndice. Por exemplo, o 95º percentil da distribuição $F_{3,\infty}$ é 2,60, de modo que $P[(Z_1^2 + Z_2^2 + Z_3^2)/3 \leq 2{,}60) = 0{,}95$. O 95º percentil da distribuição $F_{3,\infty}$ é o 95º percentil da distribuição χ_3^2 dividido por três (7,81/3 = 2,60).

A Distribuição *t* de Student

A **distribuição *t* de Student** com m graus de liberdade é definida como a distribuição da razão entre uma variável aleatória normal padrão e a raiz quadrada de uma variável aleatória qui-quadrado independentemente distribuída com m graus de liberdade, dividida por m. Isto é, seja Z uma variável aleatória normal padrão, seja W uma variável aleatória com uma distribuição qui-quadrado com m graus de liberdade e sejam Z e W independentemente distribuídas. Então, a variável aleatória $Z/\sqrt{W/m}$ possui uma distribuição t de Student (também chamada de **distribuição *t***) com m graus de liberdade. Essa distribuição é representada por t_m. Percentis selecionados da distribuição t_m são fornecidos na Tabela 2 do Apêndice.

A distribuição t de Student depende dos graus de liberdade m. Portanto, o 95º percentil da distribuição t_m depende dos graus de liberdade m. A distribuição t de Student tem um formato de sino semelhante ao da distribuição normal, mas quando m é pequeno (20 ou menos) ela apresenta mais massa nas caudas, ou seja, possui um formato de sino com caudas mais largas do que a normal. Quando m é maior do que 30, a distribuição t de Student tem uma boa aproximação pela distribuição normal padrão e a distribuição t_∞ se iguala à distribuição normal padrão.

2.5 Amostragem Aleatória e a Distribuição da Média da Amostra

Praticamente todos os procedimentos estatísticos e econométricos utilizados neste livro envolvem médias e médias ponderadas de uma amostra de dados. Caracterizar as distribuições das médias das amostras é, portanto, um passo essencial rumo à compreensão dos procedimentos econométricos.

Nesta seção, apresentamos alguns conceitos básicos sobre a amostragem aleatória e as distribuições de médias utilizadas ao longo do livro. Começamos discutindo a amostragem aleatória. O ato de amostragem aleatória, ou seja, a seleção aleatória de uma amostra de uma população maior tem o efeito de tornar a média da amostra em si uma variável aleatória. Como a média da amostra é uma variável aleatória, possui uma distribuição de probabilidade, chamada de distribuição amostral. Nesta seção, finalizamos com algumas propriedades da distribuição amostral da média da amostra.

Amostragem Aleatória

Amostragem aleatória simples. Suponha que nossa aluna da Seção 2.1 queira ser uma estatística e decida registrar a duração do trajeto que faz até a faculdade em diversos dias. Ela seleciona esses dias ao acaso durante o ano letivo e a duração do trajeto diário apresenta a função de distribuição acumulada da Figura 2.2a. Como os dias foram selecionados ao acaso, saber o valor da duração do trajeto em um desses dias selecionados aleatoriamente não fornece nenhuma informação sobre a duração do trajeto em outro dia, isto é, como os dias foram selecionados ao acaso, os valores da duração do trajeto em cada dia diferente são variáveis aleatórias distribuídas independentemente.

Amostragem Aleatória Simples e Variáveis Aleatórias i.i.d.

Conceito-Chave 2.5

Em uma amostra aleatória simples, n objetos são selecionados ao acaso de uma população e cada um tem a mesma probabilidade de ser selecionado. O valor da variável aleatória Y para o i-ésimo objeto selecionado aleatoriamente é representado por Y_i. Como cada objeto tem a mesma probabilidade de ser selecionado e a distribuição de Y_i é a mesma para todo i, as variáveis aleatórias Y_1, ..., Y_n são independente e identicamente distribuídas (i.i.d.), isto é, a distribuição de Y_i é a mesma para todo $i = 1$, ..., n e Y_1 é distribuído independentemente de Y_2, ..., Y_n e assim por diante.

A situação descrita no parágrafo anterior é um exemplo da estrutura de amostragem mais simples utilizada em estatística, conhecida como **amostragem aleatória simples**, em que n objetos são selecionados ao acaso de uma **população** (a população dos dias de aula) e cada membro da população (cada dia) tem a mesma probabilidade de ser incluído na amostra.

As n observações na amostra são representadas por Y_1, ..., Y_n, onde Y_1 é a primeira observação, Y_2 a segunda observação e assim por diante. No exemplo da aluna, Y_1 é a duração do trajeto no primeiro dos n dias selecionados aleatoriamente e Y_i é a duração do trajeto no i-ésimo dos dias selecionados aleatoriamente.

Como os membros da população incluídos na amostra são selecionados ao acaso, os valores das observações Y_1, ..., Y_n são aleatórios entre si. Se membros diferentes da população forem escolhidos, seus valores de Y serão diferentes. Portanto, o ato de amostragem aleatória significa que Y_1, ..., Y_n podem ser tratadas como variáveis aleatórias. Antes de serem selecionadas, Y_1, ..., Y_n podem assumir muitos valores possíveis; após sua seleção, um valor específico é registrado para cada observação.

Seleções i.i.d. Como Y_1, ..., Y_n são selecionadas aleatoriamente da mesma população, a distribuição marginal de Y_i é a mesma para cada $i = 1$, ..., n; essa distribuição marginal é a distribuição de Y na população selecionada. Quando Y_i possui a mesma distribuição marginal para $i = 1$, ..., n, diz-se que Y_1, ..., Y_n são **identicamente distribuídas**.

No contexto da amostragem aleatória simples, conhecer o valor de Y_1 não fornece nenhuma informação sobre Y_2, logo a distribuição condicional de Y_2 dado Y_1 é igual à distribuição marginal de Y_2. Em outras palavras, em uma amostragem aleatória simples, Y_1 é distribuída independentemente de Y_2, ..., Y_n.

Quando Y_1, ..., Y_n são selecionadas da mesma distribuição e são independentemente distribuídas, diz-se que são **independente e identicamente distribuídas**, ou **i.i.d.**

Amostragem aleatória simples e seleções i.i.d. são resumidas no Conceito-Chave 2.5.

Distribuição Amostral da Média da Amostra

A média da amostra, $\overline{Y}$, das n observações Y_1, ..., Y_n é

$$\overline{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i. \tag{2.41}$$

Um conceito essencial é o de que o ato de selecionar uma amostra aleatória torna a média da amostra $\overline{Y}$ uma variável aleatória. Como a amostra foi selecionada ao acaso, o valor de cada Y_i é aleatório. Como Y_1, ..., Y_n são aleatórias, sua média é aleatória. Caso uma amostra diferente tivesse sido selecionada, as observações e sua média da amostra seriam diferentes: o valor de $\overline{Y}$ difere de uma amostra selecionada aleatoriamente para a seguinte.

Por exemplo, suponha que nossa aluna tenha selecionado ao acaso cinco dias para registrar a duração de seu trajeto e então tenha calculado a média dessas cinco durações. Se ela tivesse escolhido cinco dias diferentes, teria registrado cinco durações diferentes — e dessa forma teria calculado um valor diferente para a média da amostra.

Como $\overline{Y}$ é aleatória, possui uma distribuição de probabilidade. A distribuição de $\overline{Y}$ é chamada de **distribuição amostral** de $\overline{Y}$, pois é a distribuição de probabilidade associada a valores possíveis de $\overline{Y}$ que poderiam ser calculados para diferentes amostras possíveis Y_1, ..., Y_n.

A distribuição amostral de médias e médias ponderadas desempenha um papel central em estatística e em econometria. Iniciamos nossa discussão da distribuição amostral de $\overline{Y}$ ao calcularmos sua média e sua variância sob condições gerais com relação à distribuição da população de Y.

Média e variância de $\overline{Y}$. Suponha que as observações Y_1, ..., Y_n sejam i.i.d. e sejam μ_Y e σ_Y^2 a média e a variância de Y_i (como as observações são i.i.d., a média e a variância são as mesmas para todo $i = 1$, ..., n). Quando $n = 2$, a média da soma $Y_1 + Y_2$ é obtida por meio da Equação (2.26), ou seja, $E(Y_1+Y_2) = \mu_Y + \mu_Y = 2\mu_Y$. Portanto, a média da média da amostra é $E[\frac{1}{2}(Y_1 + Y_2)] = \frac{1}{2} \times 2\mu_Y = \mu_Y$. Em geral,

$$E(\overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \mu_Y. \tag{2.42}$$

A variância de $\overline{Y}$ é obtida pela aplicação da Equação (2.28). Por exemplo, para $n = 2$, $\text{var}(Y_1 + Y_2) = 2\sigma_Y^2$, logo (aplicando a Equação (2.31) com $a = b = \frac{1}{2}$ e $\text{cov}(Y_1, Y_2) = 0$), $\text{var}(\overline{Y}) = \frac{1}{2}\sigma_Y^2$. Para um n geral, como Y_1, ..., Y_n são i.i.d., Y_i e Y_j são independentemente distribuídas para $i \neq j$, logo $\text{cov}(Y_i, Y_j) = 0$. Então,

$$\begin{aligned} \text{var}(\overline{Y}) &= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(Y_i) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\text{cov}(Y_i, Y_j) \\ &= \frac{\sigma_Y^2}{n}. \end{aligned} \tag{2.43}$$

O desvio padrão de $\overline{Y}$ é a raiz quadrada da variância, $\sigma_Y/\sqrt{n}$.

Em suma, a média, a variância e o desvio padrão de $\overline{Y}$ são

$$E(\overline{Y}) = \mu_Y, \tag{2.44}$$

$$\text{var}(\overline{Y}) = \sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n} \text{ e} \tag{2.45}$$

$$\text{DP}(\overline{Y}) = \sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}}. \tag{2.46}$$

Esses resultados são válidos para qualquer distribuição de Y_i, isto é, a distribuição de Y_i não precisa assumir uma forma específica, como a normal, para que as equações (2.44), (2.45) e (2.46) sejam válidas.

A notação $\sigma_{\overline{Y}}^2$ representa a variância da distribuição amostral da média da amostra $\overline{Y}$. Em contraste, σ_Y^2 é a variância de cada Y_i individual, ou seja, a variância da distribuição da população da qual a observação é selecionada. De modo semelhante, $\sigma_{\overline{Y}}$ representa o desvio padrão da distribuição amostral de $\overline{Y}$.

Distribuição amostral de $\overline{Y}$ quando Y é normalmente distribuída. Suponha que Y_1, ..., Y_n sejam seleções i.i.d. da distribuição $N(\mu_Y, \sigma_Y^2)$. Conforme foi dito após a Equação (2.37), a soma de n variáveis aleatórias normalmente distribuídas é em si normalmente distribuída. Como a média de $\overline{Y}$ é μ_Y e a variância de $\overline{Y}$ é σ_Y^2/n, isso significa que, se Y_1, ..., Y_n são seleções i.i.d. da distribuição $N(\mu_Y, \sigma_Y^2)$, $\overline{Y}$ é distribuída como $N(\mu_Y, \sigma_Y^2/n)$.

2.6 Aproximações de Distribuições Amostrais para Amostras Grandes

As distribuições amostrais desempenham um papel central no desenvolvimento de procedimentos estatísticos e econômicos e por isso é importante saber, em um sentido matemático, o que é a distribuição amostral

de $\overline{Y}$. Existem dois enfoques para caracterizar distribuições amostrais: um enfoque "exato" e um enfoque "aproximado".

O enfoque "exato" envolve a derivação de uma fórmula para a distribuição amostral que vale com exatidão para qualquer valor de n. A distribuição amostral que descreve exatamente a distribuição de $\overline{Y}$ para qualquer n é chamada de **distribuição exata** ou **distribuição de amostra finita** de $\overline{Y}$. Por exemplo, se Y é normalmente distribuído e $Y_1, \ldots, Y_n$ são i.i.d., então (conforme discutido na Seção 2.5), a distribuição exata de $\overline{Y}$ é normal com média μ_Y e variância σ_Y^2/n. Infelizmente, se a distribuição de Y não é normal, em geral a distribuição amostral exata de $\overline{Y}$ é muito complicada e depende da distribuição de Y.

O enfoque "aproximado" utiliza aproximações para a distribuição amostral que dependem do fato de o tamanho da amostra ser grande. A aproximação de distribuição amostral para amostras grandes freqüentemente é chamada de **distribuição assintótica** — "assintótica" porque as aproximações se tornam exatas no limite em que $n \longrightarrow \infty$. Como veremos nesta seção, essas aproximações podem ser muito precisas mesmo que o tamanho da amostra seja de apenas $n = 30$ observações. Como os tamanhos de amostra utilizados na prática em econometria são geralmente da ordem de centenas ou milhares, essas distribuições assintóticas fornecem aproximações muito boas para a distribuição amostral exata.

Nesta seção, apresentamos duas ferramentas importantes utilizadas para aproximar distribuições amostrais quando o tamanho da amostra é grande, a lei dos grandes números e o teorema central do limite. A lei dos grandes números afirma que, quando o tamanho da amostra for grande, $\overline{Y}$ estará próxima de μ_Y com uma probabilidade muito elevada. O teorema central do limite afirma que, quando o tamanho da amostra é grande, a distribuição amostral da média da amostra padronizada, $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$, é aproximadamente normal.

Embora as distribuições amostrais exatas sejam complicadas e dependam da distribuição de Y, as distribuições assintóticas são simples. Além disso — o que é notável —, a distribuição normal assintótica de $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$ *não* depende da distribuição de Y. Essa distribuição aproximada normal fornece simplificações enormes e forma a base da teoria da regressão utilizada ao longo deste livro.

Lei dos Grandes Números e Consistência

A **lei dos grandes números** afirma que, sob condições gerais, $\overline{Y}$ estará próxima de μ_Y com uma probabilidade muito elevada quando n for grande. Isso algumas vezes é chamado de "lei das médias". Quando se calcula a média de um número grande de variáveis aleatórias com a mesma média, os valores grandes contrabalançam os valores pequenos e a média da amostra fica próxima da média comum.

Por exemplo, considere uma versão simplificada do experimento da aluna, em que ela simplesmente registra se o seu trajeto foi curto (menos de 20 minutos) ou longo. Considere Y_i igual a um se o trajeto foi curto no i-ésimo dia selecionado aleatoriamente e igual a zero se foi longo. Como ela utilizou uma amostragem aleatória simples, $Y_1, \ldots, Y_n$ são i.i.d. Assim, Y_i, $i = 1, \ldots, n$ são seleções i.i.d. de uma variável aleatória de Bernoulli, em que (segundo a Tabela 2.2) a probabilidade de que $Y_i = 1$ é 0,78. Como a expectativa de uma variável aleatória de Bernoulli é sua probabilidade de sucesso, $E(Y_i) = \mu_Y = 0{,}78$. A média da amostra $\overline{Y}$ é a fração de dias de sua amostra em que o trajeto foi curto.

A Figura 2.6 mostra a distribuição amostral de $\overline{Y}$ para vários tamanhos de amostra n. Quando $n = 2$ (veja a Figura 2.6a), $\overline{Y}$ pode assumir apenas três valores: 0, $\frac{1}{2}$ e 1 (nenhum trajeto foi curto, um trajeto foi curto e ambos os trajetos foram curtos), nenhum dos quais está particularmente próximo da proporção verdadeira na população, 0,78. À medida que n aumenta, contudo (veja a Figura 2.6b-d), $\overline{Y}$ assume mais valores e a distribuição amostral torna-se bastante centrada em μ_Y.

A propriedade de $\overline{Y}$ estar próximo de μ_Y com probabilidade crescente à medida que n aumenta é chamada de **convergência na probabilidade** ou, de forma mais concisa, **consistência** (veja o Conceito-Chave 2.6). A lei dos grandes números afirma que, sob determinadas condições, $\overline{Y}$ converge em probabilidade para μ_Y ou, de forma equivalente, que $\overline{Y}$ é consistente para μ_Y.

As condições para a lei dos grandes números que utilizaremos neste livro são de que Y_i, $i = 1, \ldots, n$ seja i.i.d. e que a variância de Y_i, σ_Y^2, seja finita. O papel matemático dessas condições é esclarecido na Seção 15.2, em que a lei dos grandes números é provada. Se os dados são coletados por amostragem aleatória simples, a hipótese i.i.d. é válida. A hipótese de que a variância é finita afirma que valores de Y_i extremamente grandes são observados com pouca freqüência; caso contrário, a média da amostra não seria confiável. Essa hipótese é plausível para as

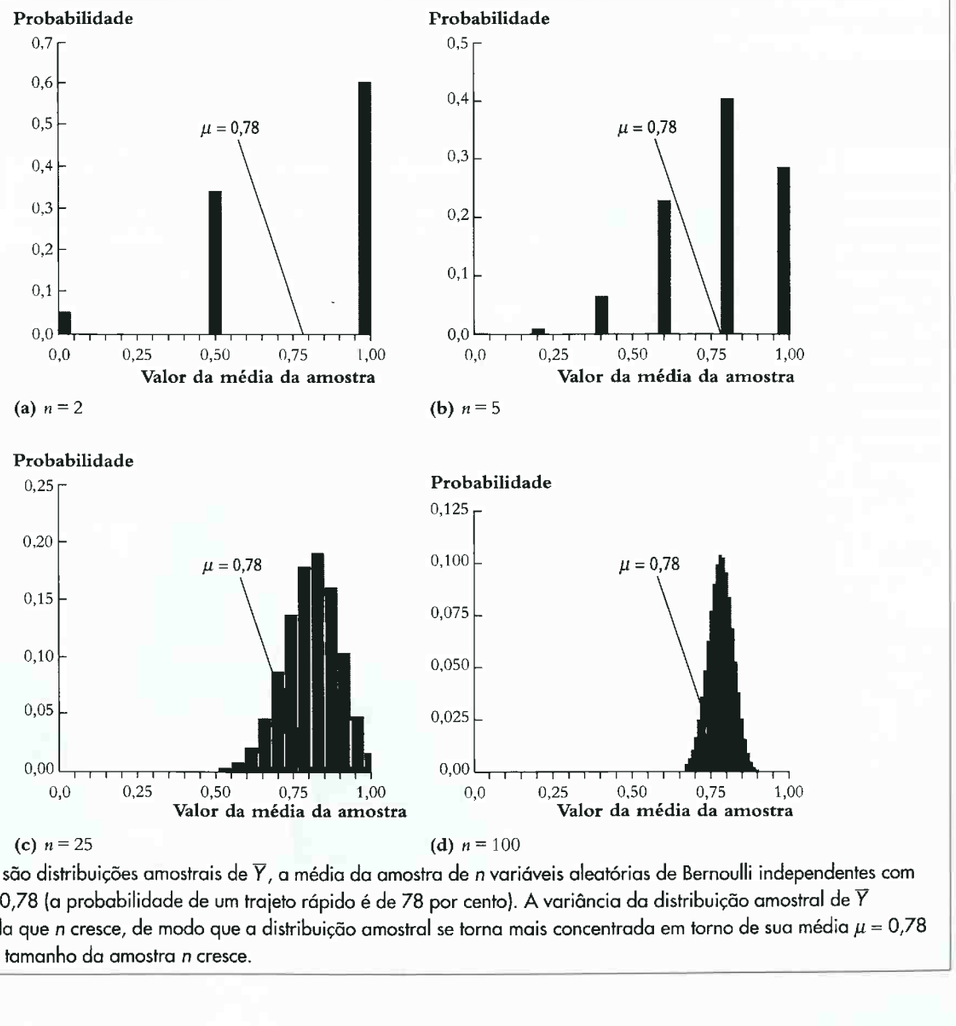
aplicações deste livro; por exemplo, como existe um limite superior para a duração do trajeto de nossa aluna (ela poderia estacionar e ir andando se o trânsito estivesse muito ruim), a variância da distribuição das durações do trajeto é finita.

Teorema Central do Limite

O **teorema central do limite** afirma que, sob condições gerais, a distribuição de $\overline{Y}$ aproxima-se bem de uma distribuição normal quando n é grande. Lembre-se de que a média de $\overline{Y}$ é μ_Y e sua variância é $\sigma_{\overline{Y}}^2 = \sigma_Y^2/n$. De acordo com o teorema central do limite, quando n é grande, a distribuição de $\overline{Y}$ é aproximadamente $N(\mu_Y, \sigma_{\overline{Y}}^2)$. Conforme discutimos no final da Seção 2.5, a distribuição de $\overline{Y}$ é *exatamente* $N(\mu_Y, \sigma_{\overline{Y}}^2)$ quando a amostra é selecionada de uma população com distribuição normal $N(\mu_Y, \sigma_Y^2)$. O teorema central do limite diz que esse mesmo resultado é *aproximadamente* verdadeiro quando n é grande mesmo que $Y_1, \ldots, Y_n$ em si não sejam normalmente distribuídas.

A convergência da distribuição de $\overline{Y}$ para a aproximação normal em formato de sino pode ser vista (um pouco) na Figura 2.6. Entretanto, como a distribuição fica bastante concentrada quando n é grande, é necessário forçar um pouco a vista. Seria mais fácil ver o formato da distribuição de $\overline{Y}$ se você utilizasse uma lupa ou tivesse outra forma de ampliar ou expandir o eixo horizontal do gráfico.

FIGURA 2.6 Distribuição Amostral da Média da Amostra de *n* Variáveis Aleatórias de Bernoulli



As distribuições são distribuições amostrais de $\overline{Y}$, a média da amostra de n variáveis aleatórias de Bernoulli independentes com $p = P(Y_i = 1) = 0{,}78$ (a probabilidade de um trajeto rápido é de 78 por cento). A variância da distribuição amostral de $\overline{Y}$ diminui à medida que n cresce, de modo que a distribuição amostral se torna mais concentrada em torno de sua média $\mu = 0{,}78$ à medida que o tamanho da amostra n cresce.

Conceito-Chave 2.6

Convergência na Probabilidade, Consistência e Lei dos Grandes Números

A média da amostra $\overline{Y}$ converge em probabilidade para μ_Y (ou, de forma equivalente, $\overline{Y}$ é consistente para μ_Y) se a probabilidade de que $\overline{Y}$ se encontra no intervalo de $\mu_Y - c$ para $\mu_Y + c$ torna-se arbitrariamente próxima de um à medida que n aumenta para qualquer constante $c > 0$. Isso é escrito como $\overline{Y} \xrightarrow{p} \mu_Y$.

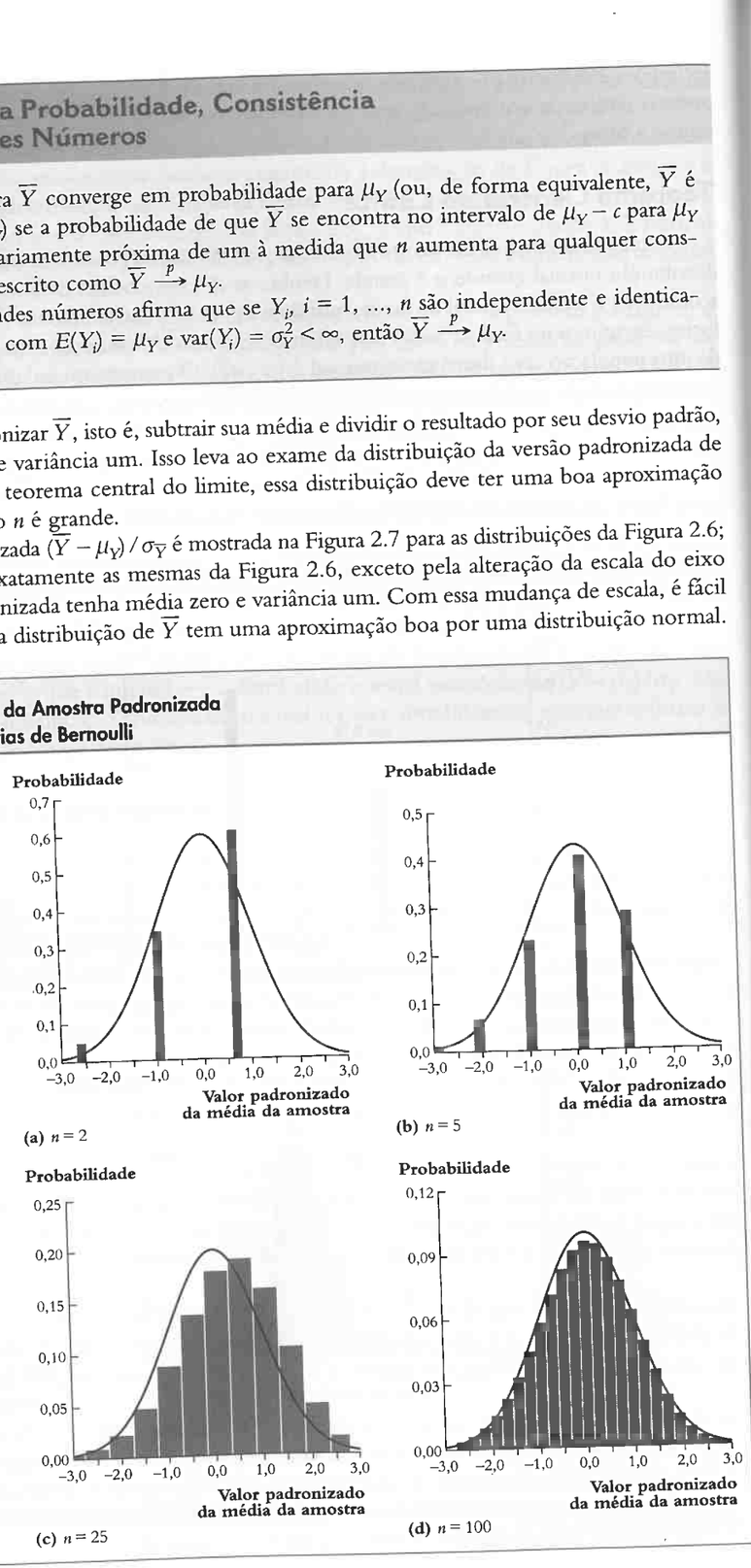
A lei dos grandes números afirma que se Y_i, $i = 1, \ldots, n$ são independente e identicamente distribuídas com $E(Y_i) = \mu_Y$ e $\text{var}(Y_i) = \sigma_Y^2 < \infty$, então $\overline{Y} \xrightarrow{p} \mu_Y$.

Uma forma de fazer isso é padronizar $\overline{Y}$, isto é, subtrair sua média e dividir o resultado por seu desvio padrão, de modo que ela tenha média zero e variância um. Isso leva ao exame da distribuição da versão padronizada de $\overline{Y}$, $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$. De acordo com o teorema central do limite, essa distribuição deve ter uma boa aproximação por uma distribuição $N(0,1)$ quando n é grande.

A distribuição da média padronizada $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$ é mostrada na Figura 2.7 para as distribuições da Figura 2.6; as distribuições da Figura 2.7 são exatamente as mesmas da Figura 2.6, exceto pela alteração da escala do eixo horizontal para que a variável padronizada tenha média zero e variância um. Com essa mudança de escala, é fácil ver que, se n é grande o suficiente, a distribuição de $\overline{Y}$ tem uma aproximação boa por uma distribuição normal.

FIGURA 2.7 Distribuição da Média da Amostra Padronizada de *n* Variáveis Aleatórias de Bernoulli

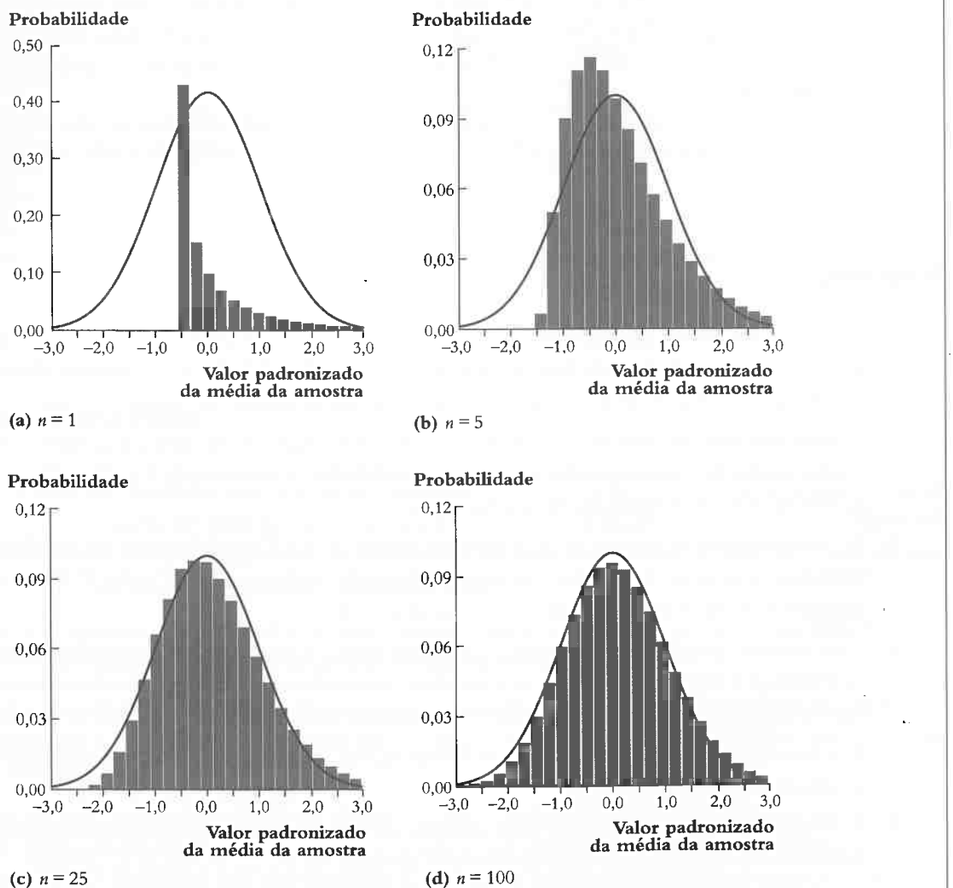
A distribuição amostral de $\overline{Y}$ na Figura 2.6 é mostrada aqui após a padronização de $\overline{Y}$. Isso centra as distribuições da Figura 2.6 e amplia a escala no eixo horizontal por um fator $\sqrt{n}$. Quando o tamanho da amostra é grande, as distribuições amostrais são cada vez mais bem aproximadas pela distribuição normal (a linha contínua), conforme previsto pelo teorema central do limite.



Alguém pode perguntar: o que significa "grande o suficiente"? Isto é, quão grande deve ser n para que a distribuição de $\overline{Y}$ seja aproximadamente normal? A resposta é: "depende". A qualidade da aproximação normal depende da distribuição dos Y_i subjacentes que compõem a média. Em um caso extremo, se os Y_i são normalmente distribuídos, então $\overline{Y}$ é normalmente distribuído de forma exata para todo n. No entanto, quando os Y_i subjacentes têm uma distribuição muito distante da normal, essa aproximação pode requerer $n = 30$ ou até mais.

Esse ponto é ilustrado pela Figura 2.8 para uma distribuição da população, mostrada na Figura 2.8a, que é muito diferente da distribuição de Bernoulli. Essa distribuição tem uma cauda direita longa (é assimétrica para a direita). A distribuição amostral de $\overline{Y}$, depois de centrada e ampliada, é mostrada nas figuras 2.8b, c e d para $n = 5$, 25 e 100. Embora a distribuição amostral esteja aproximando a forma de sino para $n = 25$, a aproximação normal ainda tem imperfeições evidentes. Entretanto, para $n = 100$, a aproximação normal é bastante boa. Na verdade, para $n \geq 100$, a aproximação normal para a distribuição de $\overline{Y}$ geralmente é muito boa para uma grande variedade de distribuições da população.

FIGURA 2.8 Distribuição da Média da Amostra Padronizada de *n* Seleções de uma Distribuição Assimétrica



As figuras mostram a distribuição amostral da média da amostra padronizada de n seleções da distribuição da população assimétrica mostrada na Figura 2.8a. Quando n é pequeno ($n = 5$), a distribuição amostral, como na distribuição da população, é assimétrica. Mas, quando n é grande ($n = 100$), as distribuições amostrais são bem aproximadas por uma distribuição normal padrão (linha contínua), conforme previsto pelo teorema central do limite.

Conceito-Chave 2.7

Teorema Central do Limite

Suponha que $Y_1, \ldots, Y_n$ sejam i.i.d. com $E(Y_i) = \mu_Y$ e $\text{var}(Y_i) = \sigma_Y^2$, onde $0 < \sigma_Y^2 < \infty$. À medida que $n \longrightarrow \infty$, a distribuição de $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$ (onde $\sigma_{\overline{Y}}^2 = \sigma_Y^2/n$) torna-se arbitrariamente bem aproximada pela distribuição normal padrão.

O teorema central do limite é um resultado notável. Enquanto as distribuições de $\overline{Y}$ para "n pequeno" nas partes b e c das figuras 2.7 e 2.8 são complicadas e completamente diferentes umas das outras, as distribuições para "n grande" nas figuras 2.7d e 2.8d são simples e, incrivelmente, têm um formato semelhante. Como a distribuição de $\overline{Y}$ se aproxima da normal à medida que n cresce bastante, diz-se que $\overline{Y}$ possui uma **distribuição normal assintótica**.

A conveniência da aproximação normal, associada a sua grande aplicabilidade devido ao teorema central do limite, faz dela um alicerce da estatística aplicada moderna. O teorema central do limite está resumido no Conceito-Chave 2.7.

Resumo

1. As probabilidades de que uma variável aleatória assuma valores diferentes são mostradas pela função de distribuição acumulada, pela função distribuição de probabilidade (para variáveis aleatórias discretas) e pela função densidade de probabilidade (para variáveis aleatórias contínuas).
2. O valor esperado de uma variável aleatória Y (também chamado de média μ_Y), representado por $E(Y)$, é o valor médio da variável ponderado pelas probabilidades. A variância de Y é $\sigma_Y^2 = E[(Y-\mu_Y)^2]$ e o desvio padrão de Y é a raiz quadrada de sua variância.
3. As probabilidades conjuntas de duas variáveis aleatórias X e Y são mostradas por sua distribuição de probabilidade conjunta. A distribuição de probabilidade condicional de Y dado $X = x$ é a distribuição de probabilidade de Y, condicional a que X assuma o valor x.
4. Uma variável aleatória normalmente distribuída possui a densidade de probabilidade em forma de sino da Figura 2.3. Para calcular uma probabilidade associada a uma variável aleatória normal, em primeiro lugar padronize a variável e então utilize a distribuição acumulada normal padrão da Tabela 1 do Apêndice.
5. A amostragem aleatória simples produz n observações aleatórias $Y_1, \ldots, Y_n$ que são independente e identicamente distribuídas (i.i.d.).
6. A média da amostra, $\overline{Y}$, varia de uma amostra selecionada aleatoriamente para a seguinte e, portanto, é uma variável aleatória com uma distribuição amostral. Se $Y_1, \ldots, Y_n$ são i.i.d., então:
   a. a distribuição amostral de $\overline{Y}$ tem média μ_Y e variância $\sigma_{\overline{Y}}^2 = \sigma_Y^2/n$;
   b. a lei dos grandes números afirma que $\overline{Y}$ converge em probabilidade para μ_Y; e
   c. o teorema central do limite afirma que a versão padronizada de $\overline{Y}$, $(\overline{Y} - \mu_Y)/\sigma_{\overline{Y}}$, possui uma distribuição normal padrão (distribuição $N(0,1)$) quando n é grande.

Termos-chave

resultados (13)
probabilidade (13)
espaço amostral (13)
evento (13)
variável aleatória discreta (13)
variável aleatória contínua (13)
distribuição de probabilidade (13)
distribuição de probabilidade acumulada (13)
função de distribuição acumulada (f.d.a.) (14)
variável aleatória de Bernoulli (14)
distribuição de Bernoulli (14)
função densidade de probabilidade (f.d.p.) (15)
função densidade (15)
densidade (15)
valor esperado (16)
média (16), variância (16) e desvio padrão (17)
momentos de uma distribuição (17)
distribuição de probabilidade conjunta (18)
distribuição de probabilidade marginal (19)
distribuição condicional (19)
expectativa condicional (20)
média condicional (20)
lei das expectativas iteradas (21)
variância condicional (21)
independência (21)
co-variância e correlação (22)
não-correlacionadas (22)
distribuição normal (23)
distribuição normal padrão (23)
padronizar uma variável aleatória (23)
distribuição normal multivariada (25)
distribuição normal bivariada (25)
distribuição qui-quadrado (25)
distribuição $F_{m,\infty}$ (25)
distribuição t de Student (27)
amostragem aleatória simples (27)
população (28)
identicamente distribuídas (28)
independente e identicamente distribuídas (i.i.d.) (28)
distribuição amostral (29)
distribuição exata (30)
distribuição assintótica (30)
lei dos grandes números (30)
convergência na probabilidade (30)
consistência (30)
teorema central do limite (31)
distribuição normal assintótica (34)

Revisão dos Conceitos

2.1 Exemplos de variáveis aleatórias utilizadas neste capítulo incluem: (a) o sexo da próxima pessoa que você vai conhecer, (b) o número de vezes que um computador trava, (c) a duração do trajeto até a faculdade, (d) se o computador que lhe disponibilizaram na biblioteca é novo ou antigo e (e) se está chovendo ou não. Explique por que cada uma delas pode ser considerada aleatória.

2.2 Suponha que as variáveis aleatórias X e Y sejam independentes e que você conheça suas distribuições. Explique por que conhecer o valor de X não lhe diz nada sobre o valor de Y.

2.3 Suponha que X represente o montante de precipitação atmosférica em sua cidade natal em um dado mês e que Y represente o número de nascimentos em Los Angeles durante o mesmo mês. Será que X e Y são independentes? Explique.

2.4 Uma turma de econometria possui oitenta alunos e o peso médio dos alunos é de 65,8 kg. Uma amostra aleatória de quatro alunos é selecionada e seu peso médio é calculado. O peso médio dos alunos da amostra será igual a 65,8 kg? Justifique. Use esse exemplo para explicar por que a média da amostra $\overline{Y}$ é uma variável aleatória.

2.5 Suponha que $Y_1, \ldots, Y_n$ sejam variáveis aleatórias i.i.d. com uma distribuição $N(1, 4)$. Esboce a densidade de probabilidade de $\overline{Y}$ para $n = 2$. Faça o mesmo para $n = 10$ e $n = 100$. Em palavras, descreva como as densidades diferem. Qual é a relação entre sua resposta e a lei dos grandes números?

2.6 Suponha que $Y_1, ..., Y_n$ sejam variáveis aleatórias i.i.d. com a distribuição de probabilidade dada pela Figura 2.8a. Você deseja calcular $P(\overline{Y} \leq 0{,}1)$. Seria razoável utilizar a aproximação normal se $n = 5$? O que dizer para $n = 25$ e $n = 100$? Explique.

Exercícios

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

***2.1** Utilize a distribuição de probabilidade fornecida na Tabela 2.2 para calcular (a) $E(Y)$ e $E(X)$; (b) σ_X^2 e σ_Y^2; e (c) σ_{XY} e corr(X, Y).

2.2 Utilizando as variáveis aleatórias X e Y da Tabela 2.2, considere duas variáveis aleatórias novas $W = 3 + 6X$ e $V = 20 - 7Y$. Calcule (a) $E(W)$ e $E(V)$, (b) σ_W^2 e σ_V^2 e (c) σ_{WV} e corr(W, V).

2.3 A tabela a seguir fornece a distribuição de probabilidade conjunta entre a situação empregatícia e o nível de instrução entre trabalhadores ou pessoas que estão procurando trabalho (desempregados) em idade ativa na população norte-americana, com base no censo dos Estados Unidos de 1990.

Distribuição Conjunta de Situação Empregatícia e Nível de Instrução da População dos Estados Unidos com Idade entre 25 e 64 Anos, 1990

	Desempregados ($Y = 0$)	Empregados ($Y = 1$)	**Total**
Sem curso superior ($X = 0$)	0,045	0,709	0,754
Com curso superior ($X = 1$)	0,005	0,241	0,246
Total	0,050	0,950	1,000

***a.** Calcule $E(Y)$.

b. A taxa de desemprego é a fração da força de trabalho que está desempregada. Demonstre que a taxa de desemprego é dada por $1 - E(Y)$.

***c.** Calcule $E(Y|X = 1)$ e $E(Y|X = 0)$.

d. Calcule a taxa de desemprego para (i) indivíduos com curso superior (ii) indivíduos sem curso superior.

***e.** Um membro da população selecionado aleatoriamente relata que está desempregado. Qual é a probabilidade de que esse trabalhador tenha curso superior? E de que não tenha curso superior?

f. As conquistas educacionais e a situação empregatícia são independentes? Explique.

2.4 A variável aleatória Y possui média 1 e variância 4. Seja $Z = \frac{1}{2}(Y - 1)$. Mostre que $\mu_Z = 0$ e $\sigma_Z^2 = 1$.

2.5 Calcule as seguintes probabilidades:

a. $P(Y \leq 3)$, se Y é distribuída como $N(1, 4)$.

b. $P(Y > 0)$, se Y é distribuída como $N(3, 9)$.

***c.** $P(40 \leq Y \leq 52)$, se Y é distribuída como $N(50, 25)$.

d. $P(6 \leq Y \leq 8)$, se Y é distribuída como $N(5, 2)$.

2.6 Calcule as seguintes probabilidades:

a. $P(Y \leq 6{,}63)$, se Y é distribuída como χ_1^2.

b. $P(Y \leq 7{,}78)$, se Y é distribuída como χ_4^2.

***c.** $P(Y > 2{,}32)$, se Y é distribuída como $F_{10,\infty}$.

2.7 Em uma população, $\mu_Y = 100$ e $\sigma_Y^2 = 43$. Utilize o teorema central do limite para resolver as seguintes questões:

a. Em uma amostra aleatória de tamanh4
o $n = 100$, calcule $P(\overline{Y} \leq 101)$.

b. Em uma amostra aleatória de tamanho $n = 165$, calcule $P(\overline{Y} > 98)$.

***c.** Em uma amostra aleatória de tamanho $n = 64$, calcule $P(101 \leq \overline{Y} \leq 103)$.

2.8 Em um ano qualquer, o tempo pode provocar danos em uma residência por meio de um temporal. De ano para ano, o dano é aleatório. Seja Y o valor monetário do dano em um dado ano qualquer. Suponha que em 95 por cento dos anos Y = US$ 0, mas que em 5 por cento dos anos Y = US$ 20.000.

a. Qual é a média e o desvio padrão do dano em qualquer ano?

b. Considere um "seguro coletivo" de cem pessoas cujas casas estão suficientemente dispersas de modo que, em um dado ano, o dano a casas diferentes possa ser visto como variáveis aleatórias independentemente distribuídas. Seja $\overline{Y}$ o dano médio a essas cem casas em um ano. (i) Qual é o valor esperado do dano médio $\overline{Y}$? (ii) Qual é a probabilidade de que $\overline{Y}$ exceda US$ 2.000?

2.9 Considere duas variáveis aleatórias X e Y. Suponha que Y assuma k valores $y_1, ..., y_k$, e que X assuma l valores $x_1, ..., x_l$.

a. Demonstre que $P(Y = y_j) = \sum_{i=1}^{l} P(Y = y_j|X = x_i)P(X = x_i)$. (*Dica*: Use a definição de $P(Y = y_j|X = x_i)$.)

b. Utilize sua resposta em (a) para verificar a Equação (2.17).

c. Suponha que X e Y sejam independentes. Mostre que $\sigma_{XY} = 0$ e corr(X, Y) = 0.

2.10 Este exercício fornece um exemplo de um par de variáveis aleatórias X e Y para as quais a média condicional de Y dado X depende de X, mas corr(X, Y) = 0. Sejam X e Z duas variáveis aleatórias normais padrão independentemente distribuídas e seja $Y = X^2 + Z$.

a. Demonstre que $E(Y|X) = X^2$.

b. Demonstre que $\mu_Y = 1$.

c. Demonstre que $E(XY) = 0$. (*Dica*: Utilize o fato de que momentos ímpares de uma variável aleatória normal padrão são todos iguais a zero.)

d. Demonstre que cov(X, Y) = 0 e, portanto, corr(X, Y) = 0.

APÊNDICE 2.1 Derivação dos Resultados do Conceito-Chave 2.3

Este apêndice deriva as equações do Conceito-Chave 2.3.

A Equação (2.29) vem da definição de expectativa.

Para derivar a Equação (2.30), utilize a definição de variância para escrever $\text{var}(a + bY) = E\{[a + bY - E(a + bY)]^2\} = E\{[b(Y - \mu_Y)]^2\} = b^2E[(Y - \mu_Y)^2] = b^2\sigma_Y^2$.

Para derivar a Equação (2.31), utilize a definição de variância para escrever

$$\begin{aligned} \text{var}(aX + bY) &= E\{[(aX + bY) - (a\mu_X + b\mu_Y)]^2\} \\ &= E\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\} \\ &= E[a^2(X - \mu_X)^2] + 2E[ab(X - \mu_X)(Y - \mu_Y)] + E[b^2(Y - \mu_Y)^2] \\ &= a^2\text{var}(X) + 2ab\text{cov}(X, Y) + b^2\text{var}(Y) \\ &= a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \end{aligned} \tag{2.47}$$

onde a segunda igualdade segue ao agrupar os termos, a terceira igualdade segue ao expandir o quadrado e a quarta igualdade segue da definição de variância e co-variância.

Para derivar a Equação (2.32), escreva $E(Y^2) = E\{[(Y - \mu_Y) + \mu_Y]^2\} = E[(Y - \mu_Y)^2] + 2\mu_Y E(Y - \mu_Y) + \mu_Y^2 = \sigma_Y^2 + \mu_Y^2$, pois $E(Y - \mu_Y) = 0$.

Para derivar a Equação (2.33), utilize a definição de co-variância para escrever

$$\begin{aligned} \text{cov}(a + bX + cV, Y) &= E\{[a + bX + cV - E(a + bX + cV)][Y - \mu_Y]\} \\ &= E\{[b(X - \mu_X) + c(V - \mu_V)][Y - \mu_Y]\} \\ &= E\{[b(X - \mu_X)][Y - \mu_Y]\} + E\{[c(V - \mu_V)][Y - \mu_Y]\} \\ &= b\sigma_{XY} + c\sigma_{VY}, \end{aligned} \tag{2.48}$$

que é a Equação (2.33).

Para derivar a Equação (2.34), escreva $E(XY) = E\{[(X - \mu_X) + \mu_X][(Y - \mu_Y) + \mu_Y]\} = E[(X - \mu_X)(Y - \mu_Y)] + \mu_X E(Y - \mu_Y) + \mu_Y E(X - \mu_X) + \mu_X\mu_Y = \sigma_{XY} + \mu_X\mu_Y$.

Agora provamos a desigualdade da correlação da Equação (2.35), a saber, $|\text{corr}(X,Y)| \leq 1$. Seja $a = -\sigma_{XY}/\sigma_X^2$ e $b = 1$. Aplicando a Equação (2.31), temos que

$$\begin{aligned} \text{var}(aX + Y) &= a^2\sigma_X^2 + \sigma_Y^2 + 2a\sigma_{XY} \\ &= (-\sigma_{XY}/\sigma_X^2)^2\,\sigma_X^2 + \sigma_Y^2 + 2(-\sigma_{XY}/\sigma_X^2)\sigma_{XY} \\ &= \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2. \end{aligned} \tag{2.49}$$

Como $\text{var}(aX + Y)$ é uma variância, ela não pode ser negativa; desse modo, da última linha da Equação (2.49), obtemos que $\sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2 \geq 0$. A reorganização dessa desigualdade produz

$$\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 \quad \text{(desigualdade da co-variância).} \tag{2.50}$$

A desigualdade de co-variância implica que $\sigma_{XY}^2/(\sigma_X^2\sigma_Y^2) \leq 1$ ou, de forma equivalente, $|\sigma_{XY}/(\sigma_X\sigma_Y)| \leq 1$, que (utilizando a definição de correlação) prova a desigualdade da correlação $|\text{corr}(X,Y)| \leq 1$.