

Sumário

Capítulo 1	A Natureza da Econometria e dos Dados Econômicos	1
1.1	O que é Econometria?	1
1.2	Passos na Análise Econômica Empírica	2
1.3	A Estrutura dos Dados Econômicos	5
1.4	A Causalidade e a Noção de <i>Ceteris Paribus</i> na Análise Econométrica	12
PARTE 1		
ANÁLISE DE REGRESSÃO COM DADOS DE CORTE TRANSVERSAL		19
<hr/>		
Capítulo 2	O Modelo de Regressão Simples	20
2.1	Definição do Modelo de Regressão Simples	20
2.2	Derivação das Estimativas de Mínimos Quadrados Ordinários	25
2.3	Mecânica do Método MQO	34
2.4	Unidades de Medida e Forma Funcional	39
2.5	Valores Esperados e Variâncias dos Estimadores de MQO	45
2.6	Regressão através da Origem	58
Capítulo 3	Análise de Regressão Múltipla: Estimação	64
3.1	Funcionabilidade da Regressão Múltipla	64
3.2	Mecânica e Interpretação dos Mínimos Quadrados Ordinários	69
3.3	O Valor Esperado dos Estimadores de MQO	80
3.4	A Variância dos Estimadores de MQO	91
3.5	Eficiência de MQO: O Teorema de Gauss-Markov	99
Capítulo 4	Análise de Regressão Múltipla: Inferência	110
4.1	Distribuições Amostrais dos Estimadores de MQO	110
4.2	Testes de Hipóteses sobre um único Parâmetro Populacional: O Teste t	113
4.3	Intervalos de Confiança	131
4.4	Testes de Hipóteses sobre uma Combinação Linear dos Parâmetros	134
4.5	Testes de Restrições Lineares Múltiplas: O Teste F	137
4.6	Descrição dos Resultados da Regressão	149

Capítulo 5	Análise de Regressão Múltipla: MQO Assimptótico	158
5.1	Consistência	158
5.2	Normalidade Assimptótica e Inferência de Amostras Grandes	163
5.3	Eficiência Assimptótica de MQO	169
Capítulo 6	Análise de Regressão Múltipla: Problemas Adicionais	174
6.1	Efeitos da Dimensão dos Dados nas Estatísticas MQO	174
6.2	Um pouco mais sobre a Forma Funcional	179
6.3	Um pouco mais sobre o Grau de Ajuste e a Seleção de Regressores	189
6.4	Previsão e Análise de Resíduos	195
	<i>Intervalos de Confiança de Previsões</i>	195
	<i>Análise de Resíduos</i>	199
	<i>Previsão de y quando a Variável Dependente é log(y)</i>	201
	Resumo	203
	Problemas	204
Capítulo 7	Análise de Regressão Múltipla com Informações Qualitativas: Variáveis Binárias (ou <i>Dummy</i>)	207
7.1	A Descrição das Informações Qualitativas	207
7.2	Uma Única Variável <i>Dummy</i> Independente	209
7.3	O Uso de Variáveis <i>Dummy</i> para Categorias Múltiplas	216
7.4	Interações Envolvendo Variáveis <i>Dummy</i>	221
7.5	Uma Variável Dependente Binária: O Modelo de Probabilidade Linear	230
7.6	Um Pouco mais sobre Análise e Avaliação de Políticas e	
Capítulo 8	Heteroscedasticidade	243
8.1	Conseqüências da Heteroscedasticidade para o Método MQO	243
8.2	Inferência Robusta em Relação à Heteroscedasticidade após e Estimacão MQO	244
8.3	O Teste da Existência de Heteroscedasticidade	251
8.4	Estimacão de Mínimos Quadrados Ponderados	256
8.5	O Modelo de Probabilidade Linear Revisitado	266
Capítulo 9	Problemas Adicionais de Especificacão e de Dados	272
9.1	Má Especificacão da Forma Funcional	272
9.2	Utilizando Variáveis Proxy para Variáveis Explicativas Não-Observadas	278
9.3	Propriedades do Método MQO quando há Erros de Medida	285
9.4	Ausência de Dados, Amostras Não-Aleatórias e Observacões Extremas	292
PARTE 2		
ANÁLISE DE REGRESSÃO COM DADOS DE SÉRIES TEMPORAIS		305
Capítulo 10	O Básico da Análise de Regressão com Dados de Séries Temporais	306
10.1	A Natureza dos Dados das Séries Temporais	306

10.2	Exemplos de Modelos de Regressão de Séries Temporais	307
10.3	Propriedades de Amostra Finita do MQO sob as Hipóteses Clássicas	311
10.4	Forma Funcional, Variáveis <i>Dummy</i> e Números-Índices	320
10.5	Tendência e Sazonalidade	327
Capítulo 11	Questões Adicionais quanto ao Uso do MQO com Dados de Séries Temporais	340
11.1	Séries Temporais Estacionárias e Fracamente Dependentes	340
11.2	Propriedades Assintóticas do MQO	345
11.3	O Uso de Séries Temporais Altamente Persistentes na Análise de Regressão	352
11.4	Modelos Dinamicamente Completos e a Ausência de Correlação Serial	360
11.5	A Hipótese de Homoscedasticidade para Modelos de Séries Temporais	363
Capítulo 12	Correlação Serial e Heteroscedasticidade em Regressões de Séries Temporais	368
12.1	As Propriedades do MQO com Erros Serialmente Correlacionados	368
12.2	O Teste da Correlação Serial	372
12.3	A Correção da Correlação Serial com Regressores Estritamente Exógenos	380
12.4	Diferenciação e Correlação Serial	387
12.5	Inferência Robusta em Relação à Correlação Serial após o MQO	388
12.6	Heteroscedasticidade em Regressões de Séries Temporais	392
PARTE 3		
TÓPICOS AVANÇADOS		401
Capítulo 13	O Agrupamento de Cortes Transversais ao Longo do Tempo. Métodos Simples de Dados de Painei	402
13.1	O Agrupamento Independente de Cortes Transversais ao Longo do Tempo	403
13.2	Análise de Decisões Governamentais com Agrupamentos	
13.3	Análise de Dados de Painei de dois Períodos	414
13.4	Análise de Decisões Governamentais com Dados de Painei de dois Períodos	421
13.5	A Diferenciação com mais de dois Períodos de Tempo	424
Capítulo 14	Métodos Avançados de Dados de Painei	433
14.1	Estimação de Efeitos Fixos	433
14.2	Modelos de Efeitos Aleatórios	441
14.3	A Aplicação de Métodos de Dados de Painei a outras Estruturas de Dados	445
Capítulo 15	Estimação de Variáveis Instrumentais e Mínimos Quadrados de dois Estágios	453
15.1	Motivação: Variáveis Omitidas em um Modelo de Regressão Simples	454
15.2	Estimação de VI do Modelo de Regressão Múltipla	464
15.3	Mínimos Quadrados de dois Estágios	468
15.4	Soluções de VI de Problemas de Erros nas Variáveis	473
15.5	O Teste de Endogeneidade e o Teste de Restrições	

Sobreidentificadoras	475
15.6 O MQ2E com Heteroscedasticidade	478
15.7 A Aplicação do MQ2E a Equações de Séries Temporais	479
15.8 A Aplicação do MQ2E em Cortes Transversais Agrupados e em Dados de Painel	481
Capítulo 16 Modelos de Equações Simultâneas	491
16.1 A Natureza dos Modelos de Equações Simultâneas	491
16.2 Viés de Simultaneidade no MQO	496
16.3 A Identificação e a Estimação de uma Equação Estrutural	498
16.4 Sistemas com mais de duas Equações	505
16.5 Modelos de Equações Simultâneas com Séries Temporais	506
16.6 Modelos de Equações Simultâneas com Dados de Painel	510
Capítulo 17 Modelos com Variáveis Dependentes Limitadas e Correções da Seleção Amostral	517
17.1 Modelos Logit e Probit de Resposta Binária	518
17.2 O Modelo Tobit para Resposta de Solução de Canto	529
17.3 O Modelo de Regressão de Poisson	537
17.4 Modelos de Regressão Censurada e Truncada	542
17.5 Correções da Seleção Amostral	549
Capítulo 18 Tópicos Avançados sobre Séries Temporais	559
18.1 Modelos de Defasagem Distribuída Infinita	560
18.2 O Teste de Raízes Unitárias	567
18.3 Regressão Espúria	572
18.4 Co-Integração e Modelos de Correção de Erro	574
18.5 Previsão	581
Capítulo 19 A Montagem de um Projeto na Prática	602
19.1 A Formulação de uma Pergunta	602
19.2 A Revisão da Literatura	604
19.3 A Compilação dos Dados	605
19.4 A Análise Econométrica	609
19.5 A Redação de um Ensaio Empírico	613
Apêndice G Tabelas Estatísticas	629
Referências Bibliográficas	637
Glossário	645
Índice Remissivo	667

1

A Natureza da Econometria e dos Dados Econômicos

O Capítulo 1 examina o escopo da econometria e propõe questões gerais que resultam da aplicação dos métodos econométricos. A Seção 1.3 examina os tipos de dados usados em negócios, economia e outras ciências sociais. A Seção 1.4 faz uma discussão intuitiva das dificuldades associadas com a inferência da causalidade nas ciências sociais.

1.1 O QUE É ECONOMETRIA?

Imagine que você seja contratado pelo governo de seu Estado para avaliar a eficácia de um programa de treinamento financiado com recursos públicos. Suponha que esse programa ensine aos trabalhadores várias maneiras de como usar computadores no processo produtivo. O programa, com duração de 20 semanas, oferece cursos fora do horário do expediente. Qualquer trabalhador horista da produção pode participar, e a matrícula em todo o programa, ou em parte dele, é voluntária. Você deve determinar qual o efeito, se houver, do programa de treinamento sobre o salário-hora de cada trabalhador.

Suponha, agora, que você trabalhe para um banco de investimentos. Você deve estudar os retornos de diferentes estratégias de investimento que envolvem títulos do Tesouro dos Estados Unidos para decidir se elas estão de acordo com as teorias econômicas a elas associadas.

A tarefa de responder a tais questões pode parecer desanimadora à primeira vista. Nesse ponto, você deve ter somente uma vaga idéia de qual tipo de dados coletar. Até o fim deste curso de princípios de econometria, você provavelmente saberá como usar os métodos econométricos para avaliar, formalmente, um programa de treinamento ou testar uma simples teoria econômica.

A econometria é baseada no desenvolvimento de métodos estatísticos para estimar relações econômicas, testar teorias, avaliar e implementar políticas de governo e de negócios. A aplicação mais comum da econometria é a previsão de importantes variáveis macroeconômicas, tais como taxas de juros, taxas de inflação e produto interno bruto (PIB). Ainda que as previsões de indicadores econômicos sejam bastante visíveis e, muitas vezes, extensamente publicadas, os métodos econométricos podem ser usados em áreas econômicas que não têm nada a ver com previsões macroeconômicas. Por exemplo, estudaremos os efeitos de gastos em campanhas políticas sobre os resultados de eleições. No campo da educação, consideraremos o efeito de gastos públicos com escolas sobre o desempenho de estudantes. Além disso, aprenderemos como usar métodos econométricos para prever séries de tempo econômicas.

A econometria evoluiu como uma disciplina separada da estatística matemática, porque enfoca problemas inerentes à coleta e à análise de dados econômicos não-experimentais. **Dados não-experimentais** não são acumulados por meio de experimentos controlados de indivíduos, firmas ou seg-

mentos da economia. (Dados não-experimentais são, às vezes, chamados de **dados observacionais** para enfatizar o fato de que o pesquisador é um coletor passivo de dados.) **Dados experimentais** são freqüentemente coletados em ambientes de laboratório nas ciências naturais, mas são muito mais difíceis de serem obtidos nas ciências sociais. Embora seja possível realizar alguns experimentos sociais, geralmente é impossível conduzir os tipos de experimentos controlados necessários para avaliar questões econômicas, seja porque eles são proibitivamente dispendiosos ou moralmente repugnantes. Na Seção 1.4, apresentaremos alguns exemplos específicos das diferenças entre dados experimentais e não-experimentais.

Naturalmente, os econométricos, sempre que possível, valem-se dos estatísticos matemáticos. O método de análise da regressão múltipla é o esteio de ambos os campos, mas seu foco e sua interpretação podem diferir de forma marcante. Além disso, os economistas criaram novas técnicas para lidar com as complexidades dos dados econômicos e para testar as previsões das teorias econômicas.

1.2 PASSOS NA ANÁLISE ECONÔMICA EMPÍRICA

Os métodos econométricos são relevantes em, virtualmente, todos os ramos da economia aplicada. Eles entram em cena quando temos uma teoria econômica para testar ou quando temos em mente uma relação que apresenta alguma importância para decisões de negócios ou análises de políticas públicas. Uma análise empírica usa dados para testar uma teoria ou estimar uma relação.

Como se estrutura uma análise econômica empírica? Pode parecer óbvio, mas é importante enfatizar que o primeiro passo em qualquer análise empírica é a formulação cuidadosa da questão de interesse. Essa questão pode ser a de testar certo aspecto de uma teoria ou os efeitos de uma política governamental. Em princípio, métodos econométricos podem ser usados para responder a uma gama de questões.

Em alguns casos, especialmente aqueles que envolvem o teste de teorias econômicas, constrói-se um **modelo econômico** formal. Um modelo econômico consiste em equações matemáticas que descrevem várias relações. Os economistas são conhecidos por suas construções de modelos os quais descrevem um amplo leque de comportamentos. Por exemplo, em microeconomia intermediária, as decisões de consumo individual, sujeitas a uma restrição orçamentária, são descritas por modelos matemáticos. A premissa básica que fundamenta esses modelos é a *maximização da utilidade*. A hipótese de que os indivíduos fazem escolhas para maximizar seu bem-estar, sujeitas às restrições de recursos, oferece-nos um arcabouço muito poderoso para criar modelos econômicos tratáveis e fazer previsões bem definidas. No contexto das decisões de consumo, a maximização da utilidade leva a um conjunto de *equações de demanda*. Em uma equação de demanda, a quantidade demandada de cada produto depende do seu próprio preço, do preço dos bens substitutos e complementares, da renda do consumidor e das características individuais que influem no gosto. Essas equações podem formar a base de uma análise econométrica da demanda do consumidor.

Os economistas têm usado ferramentas econômicas básicas, tais como o arcabouço da maximização da utilidade, para explicar comportamentos que, à primeira vista, podem parecer de natureza não econômica. Um exemplo clássico é o modelo econômico de Becker (1968) sobre o comportamento criminoso.

EXEMPLO 1.1**(Modelo Econômico do Crime)**

Em um artigo inspirador, o prêmio Nobel Gary Becker postulou um arcabouço da maximização da utilidade para descrever a participação de um indivíduo no crime. Certos crimes têm recompensas econômicas evidentes, mas muitos comportamentos criminosos têm custos. O custo de oportunidade do crime impede o criminoso de participar de outras atividades, como um emprego legal. Além disso, há custos associados com a possibilidade de ser capturado, e, se condenado, há os custos associados com o cumprimento de pena. Da perspectiva de Becker, a decisão de empreender a atividade ilegal é uma decisão de alocação de recursos com os benefícios e custos das atividades concorrentes sendo considerados.

Sob hipóteses gerais podemos derivar uma equação que descreve a quantidade de tempo gasto na atividade criminosa como uma função de vários fatores. Podemos representar tal função como

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7), \quad (1.1)$$

em que

- y = horas gastas em atividades criminosas,
- x_1 = “salário” por hora ocupada em atividade criminosa,
- x_2 = salário-hora em emprego legal,
- x_3 = renda de outras atividades que não o crime ou um emprego legal,
- x_4 = probabilidade de ser capturado,
- x_5 = probabilidade de ser condenado se capturado,
- x_6 = sentença esperada se condenado, e
- x_7 = idade.

Outros fatores geralmente afetam a decisão de uma pessoa de participar de atividades criminosas, mas a lista acima representa o que poderia resultar de uma análise econômica formal. Como é comum na teoria econômica, não fomos específicos sobre a função $f(\cdot)$ em (1.1). Essa função depende de uma função utilidade subjacente, raramente conhecida. Entretanto, podemos usar a teoria econômica – ou a introspecção – para prever o efeito que cada variável teria sobre as atividades criminosas. Essa é a base para uma análise econométrica das atividades criminosas individuais.

A modelagem econômica formal é, às vezes, o ponto de partida da análise empírica, porém é mais comum usar a teoria econômica de modo menos formal, ou mesmo contar inteiramente com a intuição. Você pode concordar quanto aos determinantes do comportamento criminoso que aparecem na equação (1.1) serem razoavelmente baseados no senso comum; poderíamos chegar a tal equação diretamente, sem partir da maximização da utilidade. Essa visão tem algum mérito, embora haja casos em que derivações formais geram idéias que a intuição pode ignorar.

Vejam os exemplos de uma equação que foi derivada por meio de um raciocínio um tanto informal.

EXEMPLO 1.2**(Treinamento e Produtividade do Trabalhador)**

Considere o problema proposto no início da Seção 1.1. Um economista especializado em trabalho gostaria de examinar os efeitos do treinamento sobre a produtividade do trabalhador. Nesse caso, há pouca necessidade de teoria econômica formal. Um entendimento econômico básico é suficiente para perceber que fatores tais como educação, experiência e treinamento influenciam a produtividade do trabalhador. Os economistas também estão bem cientes de que os trabalhadores são pagos de acordo com sua produtividade. Esse raciocínio simples leva a um modelo tal que

$$\text{saláριο}_h = f(\text{educ}, \text{exper}, \text{treina}), \quad (1.2)$$

em que saláριο_h é o salário-hora, educ representa os anos de educação formal, exper refere-se aos anos de experiência no mercado de trabalho e treina corresponde a semanas ocupadas em treinamento. Novamente, outros fatores geralmente influenciam a taxa de salário, mas (1.2) captura a essência do problema.

Após especificarmos um modelo econômico, precisamos voltar ao que chamamos de modelo econométrico. Visto que trabalharemos com modelos econométricos ao longo deste texto, é importante saber como eles se relacionam com os modelos econômicos. Considere a equação (1.1) como exemplo. A forma da função $f(\cdot)$ deve ser especificada antes de podermos empreender uma análise econométrica. Uma segunda questão concernente a (1.1) é como lidar com variáveis que não podem ser razoavelmente observadas. Por exemplo, considere o “salário” que uma pessoa pode receber na atividade criminosa. Em princípio, tal quantidade é bem-definida, mas poderia ser difícil, se não impossível, observar o “salário” para um determinado indivíduo. Mesmo variáveis como a probabilidade de ser preso não podem ser obtidas de modo realista para um determinado indivíduo, mas pelo menos podemos observar estatísticas de detenção relevantes e derivar uma variável que se aproxime da probabilidade de prisão. Muitos outros fatores que não podemos listar, nem mesmo observar, afetam o comportamento criminoso, mas devemos de algum modo considerá-los.

As ambigüidades inerentes ao modelo econômico do crime são resolvidas ao se especificar um modelo econométrico particular, tal como:

$$\text{crime} = \beta_0 + \beta_1 \text{saláριο}_m + \beta_2 \text{outrenda} + \beta_3 \text{freqpris} + \beta_4 \text{freqcond} + \beta_5 \text{sentmed} + \beta_6 \text{idade} + u, \quad (1.3)$$

em que crime é alguma medida de freqüência da atividade criminosa, saláριο_m é o salário que poderia ser ganho em um emprego legal, outrenda é a renda de outras fontes (ativos, herança etc.), freqpris é a freqüência de prisões por infrações anteriores (para aproximar a probabilidade de detenção), freqcond é a freqüência de condenações e sentmed é a duração média da sentença após as condenações. A escolha dessas variáveis é determinada pela teoria econômica e por considerações sobre os dados. O termo u contém fatores não observados, tais como o “salário” da atividade criminosa, o caráter moral, a formação da família e erros na mensuração de coisas como a atividade criminosa e a probabilidade de de tenção. Podemos adicionar variáveis de formação da família ao modelo, tais como o número de irmãos, a educação dos pais, e assim por diante, mas nunca poderemos eliminar u inteiramente. De fato, lidar com esse termo de erro ou termo de distúrbância é, talvez, o componente mais importante de qualquer análise econométrica.

As constantes $\beta_0, \beta_1, \dots, \beta_6$ são os *parâmetros* do modelo econométrico e descrevem as direções e as influências da relação entre *crime* e os fatores usados para determinar *crime* no modelo.

Um modelo econométrico completo para o Exemplo 1.2 poderia ser

$$\text{salário} = \beta_0 + \beta_1 \text{edu} + \beta_2 \text{exper} + \beta_3 \text{treina} + u, \quad (1.4)$$

em que o termo u contém fatores tais como “aptidão inata”, qualidade da educação, formação da família e uma miríade de outros fatores que podem influenciar o salário de uma pessoa. Se estivermos especialmente interessados nos efeitos do treinamento de trabalho, então β_3 é o parâmetro de interesse.

Na maioria dos casos, a análise econométrica começa pela especificação de um modelo econométrico, sem consideração de detalhes da criação do modelo. Geralmente seguimos essa abordagem, pois, em grande parte, a derivação cuidadosa de algo como o modelo econômico do crime toma muito tempo e pode nos levar para algumas áreas especializadas e freqüentemente difíceis da teoria econômica. O raciocínio econômico desempenhará um papel importante em nossos exemplos, e incorporaremos toda teoria econômica subjacente na especificação do modelo econométrico. No modelo econômico do exemplo do crime, começaríamos com um modelo econométrico tal como (1.3) e usaríamos o raciocínio econômico e o senso comum como guias para escolher as variáveis. Embora essa abordagem perca algumas das profusões da análise econômica, ela é comum e efetivamente aplicada por pesquisadores cautelosos.

Visto que um modelo econométrico tal como (1.3) ou (1.4) tenha sido especificado, várias *hipóteses* de interesse podem ser formuladas em termos dos parâmetros desconhecidos. Por exemplo, na equação (1.3), poderíamos levantar a hipótese de que salário_m , o salário que poderia ser ganho no emprego legal, não tem efeito sobre o comportamento criminoso. No contexto desse modelo econométrico específico, a hipótese é equivalente a $\beta_1 = 0$.

Uma análise empírica, por definição, requer dados. Após os dados sobre as variáveis relevantes terem sido coletados, os métodos econométricos são usados para estimar os parâmetros do modelo econométrico e para, formalmente, testar as hipóteses de interesse. Em alguns casos, o modelo econométrico é usado para fazer previsões com a finalidade de testar de uma teoria a estudo do impacto de uma política.

Como a coleta de dados é muito importante em trabalhos empíricos, a Seção 1.3 descreverá os tipos de dados com os quais, provavelmente, nos defrontaremos.

1.3 A ESTRUTURA DOS DADOS ECONÔMICOS

Os dados econômicos apresentam-se em uma variedade de tipos. Embora alguns métodos econométricos possam ser aplicados com pouca ou nenhuma modificação para muitos tipos diferentes de informações, as características especiais de alguns dados devem ser consideradas ou deveriam ser exploradas. Descreveremos a seguir as estruturas de dados mais importantes encontradas nos trabalhos aplicados.

Dados de Corte Transversal

Um **conjunto de dados de corte transversal** consiste em uma amostra de indivíduos, consumidores, empresas, cidades, estados, países ou uma variedade de outras unidades, tomada em um determinado ponto no tempo. Às vezes, os dados de todas as unidades não correspondem precisamente ao mesmo período. Por exemplo, muitas famílias podem ser pesquisadas durante diferentes semanas de um ano. Em uma aná-

lise pura de dados de corte transversal, ignoraríamos, na coleta de dados, quaisquer diferenças de tempo não importantes. Se o conjunto de famílias fosse pesquisado durante diferentes semanas do mesmo ano, ainda veríamos isso como um conjunto de dados de corte transversal.

Uma importante característica dos dados de corte transversal é que não podemos, freqüentemente, assumir que eles foram obtidos por amostragem aleatória da população subjacente. Por exemplo, se obtemos informações sobre salários, educação, experiência e outras características ao extrair aleatoriamente 500 pessoas de uma população de trabalhadores, teremos uma amostra aleatória da população de todas as pessoas que trabalham. A amostragem aleatória, matéria aprendida nos cursos introdutórios de estatística, simplifica a análise de dados de corte transversal. Uma revisão sobre amostragem aleatória aparece no Apêndice C disponível em www.thomsonlearning.com.br, na página deste livro.

Algumas vezes, a amostragem aleatória não é apropriada como uma hipótese para analisar dados de corte transversal. Por exemplo, suponha que estejamos interessados em estudar fatores que influenciam na acumulação de riqueza das famílias. Podemos estudar uma amostra aleatória de famílias, mas algumas talvez se recusem a relatar suas riquezas. Se, por exemplo, for menos provável que famílias mais ricas revelem sua riqueza, a amostra resultante sobre a riqueza não é uma amostra aleatória extraída da população de todas as famílias. Este é um exemplo de um problema de seleção amostral, um tópico avançado que discutiremos no Capítulo 17.

Outra violação da amostragem aleatória ocorre quando construímos uma amostra a partir de unidades grandes relacionadas à população, em especial a unidades geográficas. O problema provável em tais casos é que a população não é suficientemente grande para se supor, de maneira razoável, que as observações são extrações independentes. Por exemplo, se queremos explicar novas atividades de negócios entre estados, como função de taxas de salários, preços de energia, alíquotas de impostos, serviços prestados, qualidade da força de trabalho e outras características estaduais, é improvável que as atividades de negócios em um Estado próximo a outro sejam independentes. Isso revela que os métodos econométricos que discutimos funcionam, de fato, em tais situações, mas algumas vezes necessitam ser refinados. Na maioria dos casos, ignoraremos as complexidades que surgem ao analisar tais situações e trataremos esses problemas dentro do arcabouço da amostragem aleatória, mesmo quando não for tecnicamente correto fazê-lo.

Os dados de corte transversal são amplamente usados em economia e em outras ciências sociais. Em economia, a análise de dados de corte transversal está intimamente alinhada com campos da microeconomia aplicada, tais como economia do trabalho, finanças públicas estaduais e locais, organização industrial, economia urbana, demografia e economia da saúde. Dados sobre indivíduos, famílias, empresas e cidades em um determinado ponto do tempo são importantes para testar hipóteses microeconômicas e avaliar políticas governamentais.

Para a análise econométrica, os dados de corte transversal usados podem ser representados e armazenados em computadores. A Tabela 1.1 contém, de forma abreviada, um conjunto de dados de corte transversal para o ano de 1976, de 526 trabalhadores. (Esse é um subconjunto dos dados do arquivo WAGE1.RAW*.) As variáveis incluem *salárioh* (salário por hora), *educ* (anos de educação formal), *exper* (anos de experiência no mercado de trabalho), *feminino* (indicador de gênero) e *casado* (estado civil). Estas duas últimas variáveis são binárias (zero-um) por natureza, e servem para indicar características qualitativas dos indivíduos. (A pessoa é do sexo feminino ou não; a pessoa é casada ou não.) Falaremos mais sobre variáveis binárias no Capítulo 7 e seguintes.

* NRT: Todos os arquivos mencionados no texto têm a designação “*.RAW”, mas no banco de dados os arquivos são planilhas em Excel (“.XLS”), e os arquivos de trabalho de programas econométricos (como “*.WFI”, do Eviews[®]). Portanto, a designação “*.RAW” é genérica e dá um significado de “*.matéria-prima” para aplicações e exercícios.

Tabela 1.1

Conjunto de Dados de Corte Transversal sobre Salários e outras Características Individuais

<i>nobsi</i>	<i>salárioh</i>	<i>educ</i>	<i>exper</i>	<i>feminino</i>	<i>casado</i>
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
.
.
.
525	11,56	16	5	0	1
526	3,50	14	5	1	0

A variável *nobsi* na Tabela 1.1 é o número da observação atribuído a cada indivíduo na amostra. Diferentemente das outras variáveis, ela não é uma característica do indivíduo. Todos os programas econométricos e estatísticos atribuem a cada unidade um número de observação. A intuição deveria dizer-lhe que, para dados como os da Tabela 1.1, não importa qual pessoa é classificada como observação um, qual pessoa é designada pela observação dois, e assim por diante. O fato de que a ordenação dos dados não importa para a análise econométrica é uma característica fundamental dos conjuntos de dados de corte transversal obtidos a partir da amostragem aleatória.

Às vezes, variáveis diferentes correspondem a diferentes períodos nos conjuntos de dados de corte transversal. Por exemplo, a fim de determinar os efeitos de políticas governamentais sobre o crescimento econômico de longo prazo, os economistas têm estudado a relação entre crescimento do PIB *per capita* real ao longo de certo período (digamos, 1960 a 1985) e variáveis determinadas, em parte, pela política governamental em 1960 (consumo do governo como percentagem do PIB e taxas de ensino médio de adultos). Tais conjuntos de dados poderiam ser representados como na Tabela 1.2, a qual constitui parte do conjunto de dados usados no estudo de De Long e Summers (1991) sobre as taxas de crescimento entre países.

A variável *cpibpcr* representa o crescimento médio do PIB *per capita* real ao longo do período 1960 a 1985. O fato de que *consgov60* (consumo do governo como percentagem do PIB) e *second60* (percentagem da população adulta com ensino médio) correspondem ao ano de 1960, enquanto *cpibpcr* é o crescimento médio ao longo do período 1960 a 1985, não leva a quaisquer problemas especiais ao tratar essas informações como um conjunto de dados de corte transversal. As observações estão ordenadas alfabeticamente por país, mas essa ordenação não afeta em nada qualquer análise subsequente.

Tabela 1.2

Conjunto de Dados sobre Taxas de Crescimento Econômico e Características de Países

<i>nobs</i>	<i>país</i>	<i>cpibpcr</i>	<i>consgov60</i>	<i>second60</i>
1	Argentina	0,89	9	32
2	Áustria	3,32	16	50
3	Bélgica	2,56	13	69
4	Bolívia	1,24	18	12
.
.
.
61	Zimbábue	2,30	17	6

Dados de Séries de Tempo

Um **conjunto de dados de séries de tempo** consiste em observações sobre uma variável ou muitas variáveis ao longo do tempo. Exemplos de dados de séries temporais incluem preços de ações, oferta de moeda, índice de preços ao consumidor, produto interno bruto, taxas anuais de homicídios e números de vendas de automóveis. Como eventos passados podem influenciar eventos futuros, e como, nas ciências sociais, as defasagens do comportamento são prevalentes, o tempo é uma dimensão importante em um conjunto de dados de séries de tempo. Diferentemente do arranjo dos dados de corte transversal, a ordenação cronológica das observações em uma série de tempo transmite informações potencialmente importantes.

Uma característica essencial dos dados de séries de tempo que torna mais difícil a análise do que os dados de corte transversal é o fato de que raramente é possível assumir (se é que é possível) que as observações econômicas são independentes ao longo do tempo. A maioria das séries de tempo econômicas, bem como de outras séries de tempo, está relacionada – muitas vezes fortemente relacionada – com seus históricos recentes. Por exemplo, saber algo sobre o produto interno bruto do último trimestre nos diz muito sobre a provável variação do PIB durante este trimestre, visto que o PIB tende a permanecer razoavelmente estável de um trimestre para o próximo. Embora muitos procedimentos econométricos possam ser usados tanto com dados de corte transversal como com dados de séries de tempo, outros pontos podem ser considerados para especificar, apropriadamente, os modelos econométricos que usam dados de séries de tempo. Além disso, as modificações e embelezamentos das técnicas econométricas comuns foram desenvolvidas com a finalidade de considerar e explorar a natureza dependente das séries de tempo e para tratar de outras questões, tal como o fato de que algumas variáveis econômicas tendem a exibir claras tendências ao longo do tempo.

Outra característica dos dados de séries de tempo que pode requerer atenção especial é a **frequência dos dados**, na qual eles são coletados. Em economia, as frequências mais comuns são: diária, semanal, mensal, trimestral e anual. Os preços de ações são registrados em intervalos diários (excluindo sábados e domingos). A oferta de moeda na economia dos Estados Unidos é informada semanalmente. Muitas séries macroeconômicas são tabuladas mensalmente, incluindo as taxas de inflação e desemprego. Outras séries macroeconômicas são registradas menos frequentemente, como a cada três meses (todo trimestre). O produto interno bruto é um exemplo importante de uma série trimestral. Outras séries de tempo, como as taxas de mortalidade infantil dos estados norte-americanos, estão disponíveis somente em bases anuais.

Muitas séries de tempo econômicas, sejam semanais, mensais ou trimestrais, exibem um forte padrão sazonal, o qual pode ser um importante fator na análise de séries de tempo. Por exemplo, dados mensais sobre o início da construção de moradias se diferenciam entre os meses simplesmente devido a mudanças das condições climáticas. Aprenderemos como trabalhar com séries de tempo no Capítulo 10.

A Tabela 1.3 contém um conjunto de dados de séries de tempo, obtido de um artigo de Castillo-Freeman e Freeman (1992), sobre os efeitos do salário mínimo em Porto Rico. O ano mais antigo no conjunto de dados é a primeira observação, e o ano mais recente disponível é a última observação. Quando os métodos econométricos são utilizados para analisar dados de séries de tempo, os dados devem ser armazenados em ordem cronológica.

Tabela 1.3

Salário Mínimo, Desemprego e Dados Relacionados para Porto Rico

<i>nobsa</i>	<i>ano</i>	<i>minmed</i>	<i>cobmed</i>	<i>desemp</i>	<i>pnb</i>
1	1950	0,20	20,1	15,4	878,7
2	1951	0,21	20,7	16,0	925,0
3	1952	0,23	22,6	14,8	1.015,9
.
.
.
37	1986	3,35	58,1	18,9	4.281,6
38	1987	3,35	58,2	16,8	4.496,7

A variável *minmed* se refere ao salário mínimo médio no ano, *cobmed* é a taxa de cobertura média (o percentual de trabalhadores cobertos pela lei de salário mínimo), *desemp* é a taxa de desemprego e *pnb* é o produto nacional bruto. Usaremos esses dados mais adiante em uma análise de séries de tempo do efeito do salário mínimo sobre o emprego.

Cortes Transversais Agrupados

Alguns conjuntos de dados têm tanto características de corte transversal quanto de séries de tempo. Por exemplo, suponha que dois estudos sobre famílias sejam realizados nos Estados Unidos com dados de corte transversal, um em 1985 e outro em 1990. Em 1985, uma amostra aleatória de famílias é pesquisada para variáveis tais como renda, poupança, tamanho da família, e assim por diante. Em 1990, uma *nova* amostra aleatória de famílias é extraída usando as mesmas questões da pesquisa. A fim de aumentar nosso tamanho de amostra, podemos formar um **corte transversal agrupado** ao combinar os dois anos.

Agrupar cortes transversais de diferentes anos é, freqüentemente, um modo eficaz de analisar os efeitos de uma nova política de governo. A idéia é coletar dados de anos anteriores e posteriores a uma importante mudança de política governamental. Como exemplo, considere o seguinte conjunto de dados sobre os preços da moradia coletados em 1993 e 1995 nos Estados Unidos, quando houve uma redução nos impostos sobre a propriedade em 1994. Suponha que tenhamos dados sobre 250 residências para 1993 e sobre 270 para 1995. Um modo de armazenar tais dados é apresentado na Tabela 1.4.

Tabela 1.4

Cortes-Transversais Agrupados: Dois Anos de Preços de Moradias

<i>nobsm</i>	<i>ano</i>	<i>preçoc</i>	<i>impro</i>	<i>arquad</i>	<i>qtdorm</i>	<i>banhos</i>
1	1993	85.500	42	1.600	3	2,0
2	1993	67.300	36	1.440	3	2,5
3	1993	134.000	38	2.000	4	2,5
.
.
.
250	1993	243.600	41	2.600	4	3,0
251	1995	65.000	16	1.250	2	1,0
252	1995	182.400	20	2.200	4	2,0
253	1995	97.500	15	1.540	3	2,0
.
.
.
520	1995	57.200	16	1.100	2	1,5

As observações 1 a 250 correspondem às residências vendidas em 1993, e as observações 251 a 520 correspondem às 270 residências vendidas em 1995. Embora a ordem na qual armazenamos os dados não se revele crucial, não se esqueça de que o ano de cada observação é, geralmente, muito importante. Essa é a razão de introduzirmos *ano* como uma variável separada.

A análise de um corte transversal agrupado é muito parecida com a de um corte transversal padrão, exceto pelo fato de que precisamos, freqüentemente, considerar diferenças periódicas das variáveis ao longo do tempo. De fato, além de aumentar o tamanho da amostra, a característica de uma análise de corte transversal agrupada é, freqüentemente, ver como uma relação fundamental mudou ao longo do tempo.

Dados de Painel ou Longitudinais

Um conjunto de **dados de painel** (ou dados longitudinais) consiste em uma série de tempo para *cada* membro do corte transversal do conjunto de dados. Como exemplo, suponha que tenhamos o histórico de salário, educação e emprego para um conjunto de indivíduos ao longo de um período de dez anos, ou que possamos coletar informações, tais como dados de investimento e financeiros, sobre o mesmo conjunto de empresas ao longo de um período de cinco anos. Dados de painel também podem ser coletados para unidades geográficas. Por exemplo, podemos coletar dados para o mesmo conjunto de municípios dos Estados Unidos sobre fluxos de imigração, impostos, taxas de salários, gastos governamentais etc., para os anos de 1980, 1985 e 1990.

A característica essencial dos dados de painel que os distingue dos dados de corte transversal agrupado é o fato de que as *mesmas* unidades do corte transversal (indivíduos, empresas ou municípios nos exemplos anteriores) são acompanhadas ao longo de um determinado período. Os dados na Tabela 1.4 não são considerados um conjunto de dados de painel porque as residências vendidas são provavelmente diferentes em 1993 e 1995; se houver quaisquer repetições, o número é provavelmente bem pequeno para ser significativo. Em contraste, a Tabela 1.5 contém um conjunto de dados de painel de dois anos sobre o crime e as estatísticas relacionadas para 150 cidades nos Estados Unidos.

Há várias características interessantes na Tabela 1.5. Primeiro, a cada cidade foi dado um número de 1 a 150. Qual cidade decidimos chamar de cidade 1, cidade 2, e assim por diante, é irrelevante. Assim como em um corte transversal puro, a ordenação no corte transversal de um conjunto de dados de painel não é importante. Poderíamos usar o nome da cidade em lugar de um número, mas é frequentemente útil ter ambos.

Tabela 1.5

Conjunto de Dados de Painel sobre Estatísticas de Crime nas Cidades para Dois Anos

<i>nobs</i>	<i>cidade</i>	<i>ano</i>	<i>homicds</i>	<i>população</i>	<i>desemp</i>	<i>polícia</i>
1	1	1986	5	350.000	8,7	440
2	1	1990	8	359.200	7,2	471
3	2	1986	2	64.300	5,4	75
4	2	1990	1	65.100	5,5	75
.
.
.
297	149	1986	10	260.700	9,6	286
298	149	1990	6	245.000	9,8	334
299	150	1986	25	543.000	4,3	520
300	150	1990	32	546.200	5,2	493

Um segundo ponto é que os dois anos dos dados para a cidade 1 preenchem as duas primeiras linhas ou observações. As observações 3 e 4 correspondem à cidade 2, e assim por diante. Como cada uma das 150 cidades tem duas linhas de dados, qualquer pacote econométrico verá isso como 300 observações. Esse conjunto de dados pode ser tratado como um corte transversal agrupado, em que as mesmas cidades aparecem em cada ano. Porém, como veremos nos Capítulos 13 e 14, podemos também usar a estrutura de painel para responder a questões que não podem ser respondidas simplesmente vendo isso como um corte transversal agrupado.

Ao organizar as observações na Tabela 1.5, colocamos os dois anos dos dados de cada cidade um ao lado do outro, com o primeiro ano antecedendo o segundo em todos os casos. Apenas por questões práticas, esse é o modo preferido de se ordenar conjuntos de dados de painel. Essa organização contrasta com o modo pelo qual os cortes transversais agrupados são armazenados na Tabela 1.4. Em resu-

mo, a razão para ordenar os dados de painel como na Tabela 1.5 é que precisaremos fazer transformações dos dados para cada cidade nos dois anos.

Como os dados de painel requerem a repetição das mesmas unidades ao longo do tempo, os conjuntos de dados de painel, especialmente aqueles sobre indivíduos, famílias e empresas, são mais difíceis de se obter que os cortes transversais agrupados. Não surpreendentemente, observar as mesmas unidades ao longo do tempo traz várias vantagens sobre os dados de corte transversal ou mesmo sobre os de dados de cortes transversais agrupados. O benefício que salientaremos neste livro é que ter múltiplas observações sobre as mesmas unidades nos permite controlar certas características não observáveis dos indivíduos, firmas etc. Como veremos, o uso de mais de uma observação pode facilitar a inferência causal em situações em que inferir causalidade seria muito difícil se somente um único corte transversal estivesse disponível. Uma segunda vantagem dos dados de painel é que eles, frequentemente, nos permitem estudar a importância das defasagens do comportamento ou o resultado de tomar decisões. Essa informação pode ser importante, visto que se pode esperar o impacto em muitas políticas públicas somente após algum tempo.

A maior parte dos livros para cursos de nível superior não contém uma discussão de métodos econométricos para dados de painel. Entretanto, os economistas agora reconhecem que algumas questões são difíceis, se não impossíveis, de serem respondidas satisfatoriamente sem dados de painel. Como você verá, podemos fazer consideráveis progressos com análises simples de dados de painel, um método que não é muito mais difícil do que trabalhar com um conjunto de dados de corte transversal padrão.

Um Comentário sobre Estruturas de Dados

A Parte 1 deste livro cobre a análise de dados de corte transversal, já que ela propõe menos conceitos e dificuldades técnicas. Ao mesmo tempo, ela ilustra muitos dos temas essenciais da análise econométrica. Usaremos os métodos e as idéias da análise de corte transversal no restante do texto.

Embora a análise econométrica de séries de tempo use muitas das mesmas ferramentas que a análise de corte transversal, ela é mais complicada devido à existência de tendência que traduz a natureza altamente persistente de muitas séries de tempo econômicas. Acredita-se que agora são considerados falhos muitos exemplos que têm sido tradicionalmente usados para ilustrar a maneira pela qual os métodos econométricos podem ser aplicados a dados de séries de tempo. Faz pouco sentido usar tais exemplos inicialmente, visto que esse hábito somente reforça uma prática econométrica insatisfatória. Portanto, postergaremos o tratamento da econometria de séries de tempo até a Parte 2, quando questões importantes concernentes a tendência, persistência, dinâmica e sazonalidade serão introduzidas.

Na Parte 3, trataremos explicitamente de cortes transversais agrupados e dados de painel. A análise de cortes transversais independentemente agrupados e a análise simples de dados de painel são ambas independentemente, extensões claras e diretas da análise pura de corte transversal. Entretanto, vamos esperar até o Capítulo 13 para tratar desses tópicos.

1.4 A CAUSALIDADE E A NOÇÃO DE *CETERIS PARIBUS* NA ANÁLISE ECONOMÉTRICA

Em muitos testes de teoria econômica, e certamente para avaliar políticas públicas, o objetivo do economista é inferir que uma variável (tal como a educação) tem um **efeito causal** sobre outra variável (tal como a produtividade do trabalhador). Encontrar simplesmente uma associação entre duas ou mais variáveis pode ser sugestivo, mas, a não ser que se possa estabelecer uma causalidade, raramente ela é convincente.