

Introdução

I.1 O que é econometria?

Em uma interpretação literal, *econometria* significa “medição econômica”. Embora a medição seja uma parte importante da econometria, seu escopo é muito mais amplo, como mostram as seguintes citações:

A econometria, resultado de determinada perspectiva sobre o papel da economia, consiste na aplicação da estatística matemática a dados econômicos para dar suporte empírico aos modelos formulados pela economia matemática e obter resultados numéricos.¹

[...] a econometria pode ser definida como a análise quantitativa dos fenômenos econômicos ocorridos com base no desenvolvimento paralelo da teoria e das observações e com o uso de métodos de inferência adequados.²

A econometria pode ser definida como a ciência social em que as ferramentas da teoria econômica, da matemática e da inferência estatística são aplicadas à análise dos fenômenos econômicos.³

A econometria diz respeito à determinação empírica das leis econômicas.⁴

A arte do econometrista está em encontrar o conjunto de hipóteses suficientemente específicas e realistas que lhe permitam tirar o melhor proveito dos dados de que dispõe.⁵

Os econometristas [...] são um auxílio positivo na tentativa de dissipar a imagem pública negativa da economia (seja ela quantitativa ou não) como assunto em que caixas vazias são abertas supondo-se a existência de abridores de lata para revelar conteúdos que dez economistas interpretarão de 11 maneiras distintas.⁶

O método da pesquisa econométrica visa, essencialmente, a conjugação da teoria econômica com medições concretas, usando a teoria e a técnica da inferência estatística como uma ponte.⁷

¹ TINTNER, Gerhard. *Methodology of mathematical economics and econometrics*. Chicago: The University of Chicago Press, 1968. p. 74.

² SAMUELSON, P. A.; KOOPMANS, T. C.; STONE, J. R. N. *Report of the evaluative committee for econometrica*. *Econometrica*. Abr. 1954, v. 22, n. 2, p. 141-146.

³ GOLDBERGER, Arthur S. *Econometric theory*. Nova York: John Wiley & Sons, 1964. p. 1.

⁴ THEIL, H. *Principles of econometrics*. Nova York: John Wiley & Sons, 1971. p. 1.

⁵ MALINVAUD, E. *Statistical methods of econometrics*. Chicago: Rand McNally, 1966. p. 514.

⁶ DARNELL, Adrian C.; EVANS, J. Lynne. *The limits of econometrics*. Rants, Inglaterra: Edward Elgar Publishing, 1990. p. 54.

⁷ HAAVELMO, T. “The probability approach in econometrics”. Suplemento da *Econometrica*. 1944. v. 12, prefácio p. iii.

I.2 Por que uma disciplina separada?

Como as definições apresentadas sugerem, a econometria é um amálgama de teoria econômica, economia matemática, estatística econômica e estatística matemática. Contudo, o assunto merece ser estudado de modo independente pelas seguintes razões.

A teoria econômica faz declarações ou hipóteses principalmente de natureza qualitativa. Por exemplo, a teoria microeconômica afirma que, tudo o mais permanecendo igual, uma redução no preço de uma mercadoria deve resultar no aumento da quantidade demandada por esta mercadoria. Portanto, a teoria econômica postula uma relação negativa ou inversa entre o preço e a quantidade demandada de uma mercadoria. Mas a teoria em si não oferece nenhuma medida quantitativa da relação entre as duas variáveis; ela não nos informa quanto a quantidade aumentará ou diminuirá em consequência de determinada variação no preço da mercadoria. Cabe ao econometrista oferecer essas estimativas numéricas. Em outras palavras, o econometrista proporciona conteúdo prático à maior parte da teoria econômica.

A principal preocupação da economia matemática é expressar a teoria econômica de forma matemática (equações) sem levar em conta se a teoria pode ser medida ou verificada empiricamente. A econometria, como já mencionado, está principalmente interessada na verificação da teoria econômica. Conforme veremos, o econometrista frequentemente usa as equações matemáticas formuladas pelo economista matemático, mas as aplica de forma que possam ser testadas na prática. E essa conversão de equações matemáticas em equações econométricas requer bastante engenhosidade e habilidade.

A estatística econômica busca principalmente a coleta, processamento e apresentação dos dados econômicos na forma de gráficos e tabelas. Essa é a tarefa do estatístico econômico. É ele o principal responsável por coletar dados sobre o produto nacional bruto (PNB), o emprego, o desemprego, os preços etc. As informações coletadas constituem os dados brutos do trabalho econométrico. Mas o trabalho do estatístico econômico não vai além disso; seu foco não é usar os dados para testar as teorias econômicas. É claro, se fosse, ele se tornaria um econometrista.

Embora a estatística matemática proporcione muitas das ferramentas usadas em sua atividade, os econometristas em geral precisam de métodos especiais em vista da natureza específica da maioria dos dados econômicos, isto é, por serem dados que não foram gerados por meio de experimentos controlados. O econometrista, como o meteorologista, depende em geral de dados que não podem ser controlados diretamente. Como Spanos observa corretamente:

Na econometria, quem modela muitas vezes se depara com dados provenientes de **observações** em oposição aos dados **experimentais**. Isso tem duas implicações importantes para a modelagem empírica na econometria. Primeiro, quem modela deve dominar habilidades muito diferentes das necessárias à análise de dados experimentais [...]. Segundo, a diferença entre quem coleta dados e quem os analisa exige que quem modela esteja profundamente familiarizado com a natureza e a estrutura dos dados em questão.⁸

I.3 A metodologia econométrica

Como fazem os econometristas para analisar um problema econômico? Qual metodologia utilizam? Embora existam várias escolas de pensamento sobre metodologia econométrica, aqui apresentamos a **tradicional** ou **clássica**, que ainda domina a pesquisa na economia e em outras ciências sociais e comportamentais.⁹

⁸ SPANOS, Aris. *Probability. Theory and statistical inference: econometric modelling with observational data*. Reino Unido: Cambridge University Press, 1999. p. 21.

⁹ Para uma discussão esclarecedora, embora avançada, do método econométrico, veja HENDRY, David F. *Dynamic econometrics*. Nova York: Oxford University Press, 1995. Veja também SPANOS, Aris op. cit.

Em termos gerais, a metodologia econométrica tradicional segue os seguintes passos:

1. Exposição da teoria ou hipótese.
2. Especificação do modelo matemático da teoria.
3. Especificação do modelo estatístico ou econométrico.
4. Obtenção dos dados.
5. Estimação dos parâmetros do modelo econométrico.
6. Teste de hipóteses.
7. Projeção ou previsão.
8. Uso do modelo para fins de controle ou de política.

Para ilustrarmos esses passos, vejamos a conhecida teoria do consumo keynesiana.

1. Exposição da teoria ou hipótese

Keynes afirmou:

A lei psicológica fundamental [...] é que os homens [as mulheres] estão dispostos, como regra e em média, a aumentar seu consumo conforme sua renda aumenta, mas não na mesma proporção que o aumento na renda.¹⁰

Em resumo, Keynes postulava que a **propensão marginal a consumir (PMC)**, a taxa de variação do consumo por variação de uma unidade (digamos, um dólar) de renda, é maior que zero, mas menor que 1.

2. Especificação do modelo matemático da teoria

Embora Keynes postulasse uma relação positiva entre consumo e renda, ele não especificou a forma exata da relação funcional entre as duas variáveis. Para simplificar, um economista matemático poderia sugerir a seguinte forma para a função de consumo keynesiana:

$$Y = \beta_1 + \beta_2 X \quad 0 < \beta_2 < 1 \quad (\text{I.3.1})$$

em que Y = despesas de consumo e X = renda, e β_1 e β_2 , conhecidos como os **parâmetros** do modelo, são, respectivamente, o **intercepto** e o coeficiente **angular**.

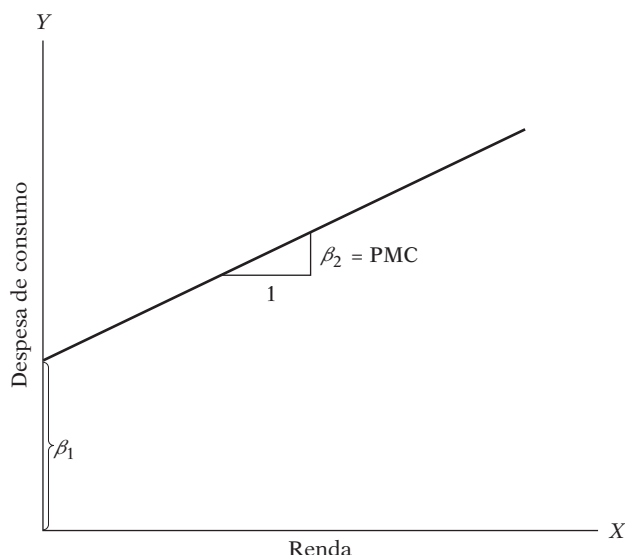
O coeficiente angular, β_2 , mede a PMC. A Figura 1.1 mostra a representação geométrica da Equação (I.3.1). Essa equação, que especifica que o consumo se relaciona linearmente à renda, é um exemplo de modelo matemático da relação entre consumo e renda e é conhecida como **função consumo** em economia. O modelo é apenas um conjunto de equações matemáticas. Se o modelo tem apenas uma equação, como no apresentado, denomina-se **modelo uniequacional**, enquanto se tiver mais de uma equação será denominado **modelo de múltiplas equações** (que veremos mais adiante).

Na Equação (I.3.1), a variável que aparece do lado esquerdo do sinal de igualdade é chamada de *variável dependente* e a(s) variável(eis) do lado direito é(são) chamada(s) de variável(eis) **independente(s)** ou **explanatória(s)**. Assim, na função consumo keynesiana, o consumo (despesa) é a variável dependente e a renda é a variável explanatória.

¹⁰ KEYNES, John Maynard. *The general theory of employment, interest and money*. Nova York: Harcourt Brace Jovanovich, 1936. p. 96.

FIGURA I.1

Função consumo
keynesiana.



3. Especificação do modelo estatístico ou econométrico

O modelo puramente matemático da função consumo apresentado na Equação (I.3.1) é de interesse limitado para o econometrista, pois supõe que existe uma relação *exata* ou *determinística* entre o consumo e a renda. Mas as relações entre variáveis econômicas são, em geral, *inexatas*. Portanto, se coletarmos dados sobre despesas de consumo e renda disponível (a renda depois de descontados os impostos) de uma amostragem de, digamos, 500 famílias americanas e traçarmos um gráfico em que o eixo vertical representa as despesas de consumo e o eixo horizontal, a renda disponível, não devemos esperar que as 500 observações se situem exatamente na reta dada pela Equação (I.3.1). Isso porque, além da renda, outras variáveis afetam as despesas de consumo. O tamanho da família, a idade de seus integrantes, a religião etc., por exemplo, provavelmente exercem certa influência sobre o consumo.

Para levar em conta as relações *inexatas* entre as variáveis econômicas, o econometrista deve modificar a função consumo determinística da Equação (I.3.1) do seguinte modo:

$$Y = \beta_1 + \beta_2 X + u \quad (\text{I.3.2})$$

em que u , conhecido como **distúrbio**, ou **termo de erro**, é uma **variável aleatória (estocástica)** que tem propriedades probabilísticas conhecidas. O termo de erro u pode representar bem todos esses fatores que afetam o consumo, mas que não são levados em conta explicitamente.

A Equação (I.3.2) é um exemplo de modelo econométrico. Mais tecnicamente, é um exemplo de **modelo de regressão linear**, que é o principal tema deste livro. A função consumo econométrica baseia-se na hipótese de que a variável dependente Y (o consumo) se relaciona linearmente com a variável explanatória X (a renda), mas que a relação entre ambas não é exata: está sujeita a variações individuais.

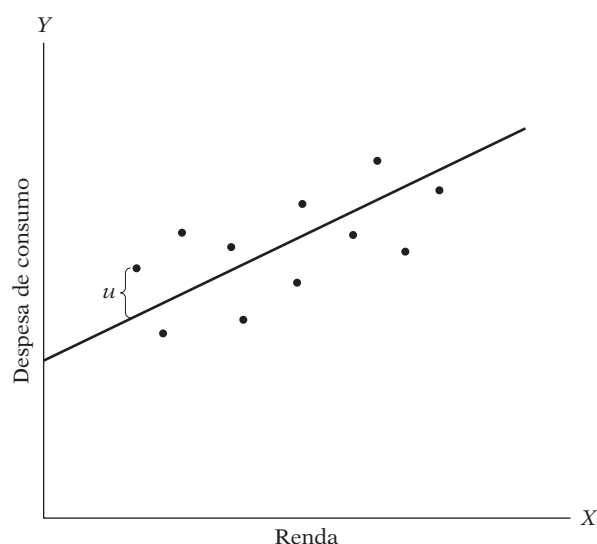
O modelo econométrico da função consumo pode ser representado pelo gráfico da Figura 1.2.

4. Obtenção dos dados

Para estimarmos o modelo econométrico da Equação (I.3.2), isto é, para obtermos os valores numéricos de β_1 e β_2 , precisamos de dados. Embora no próximo capítulo falaremos mais sobre a importância crítica dos dados para a análise econômica, por enquanto vamos examinar os dados apresentados na Tabela I.1, que se refere à economia dos Estados Unidos

FIGURA I.2

Modelo econométrico
da função consumo
keynesiana.



no período 1960-2005. Na tabela, a variável Y corresponde às despesas de consumo pessoal (DCP) *agregada* (isto é, para a economia como um todo) e a variável X ao produto interno bruto (PIB), um indicador de renda agregada, ambas medidas em termos de bilhões de dólares de 2000. Portanto, os dados são apresentados em termos “reais”, isto é, foram medidos a preços constantes (de 2000). Os dados estão representados graficamente na Figura I.3 (compare com a Figura I.2). Por enquanto, deixemos de lado a linha traçada no gráfico.

5. Estimação dos parâmetros do modelo econométrico

Agora que temos os dados, nossa próxima tarefa é estimar os parâmetros da função consumo. A estimativa numérica dos parâmetros fornece conteúdo empírico à função consumo. O mecanismo para estimar os parâmetros será examinado no Capítulo 3. Por enquanto, note que a técnica estatística da **análise de regressão** é a principal ferramenta para obter as estimativas. Aplicando essa técnica aos dados da Tabela I.1, obtemos as seguintes estimativas de β_1 e β_2 , especificamente, — 299,5913 e 0,7218. Portanto, a função consumo estimada é:

$$\hat{Y}_t = -299,5913 + 0,7218X_t \quad (I.3.3)$$

O acento circunflexo em cima do Y indica que se trata de uma estimativa.¹¹ A Figura I.3 mostra a função consumo estimada (isto é, a linha de regressão).

Como indica a Figura I.3, a linha de regressão ajusta-se bem aos dados, no sentido de que os pontos no gráfico que representam os dados ficam muito próximos da linha de regressão. A figura nos mostra que, para o período 1960-2005, o coeficiente angular (a **PMC**) era de quase 0,72, indicando que, no período amostrado, um aumento de um dólar na renda real levava, *em média*, a um aumento de cerca de 72 centavos nas despesas reais de consumo.¹² Dizemos *em média* porque a relação entre consumo e renda é inexata; como fica claro na Figura I.3, nem todos os pontos dos dados estão exatamente sobre a linha de regressão. Em termos simples, podemos dizer que, de acordo com nossos dados, as despesas médias de consumo aumentam cerca de 70 centavos a cada aumento real de um dólar na renda real.

¹¹ O uso de um acento circunflexo sobre uma variável ou parâmetro indica, por convenção, que se trata de um valor estimado.

¹² Por enquanto, não se preocupe sobre como esses valores foram obtidos. Como mostraremos no Capítulo 3, essas estimativas foram obtidas através do método estatístico dos **mínimos quadrados**. Também, por enquanto, não se preocupe com o valor negativo do intercepto.

TABELA I.1

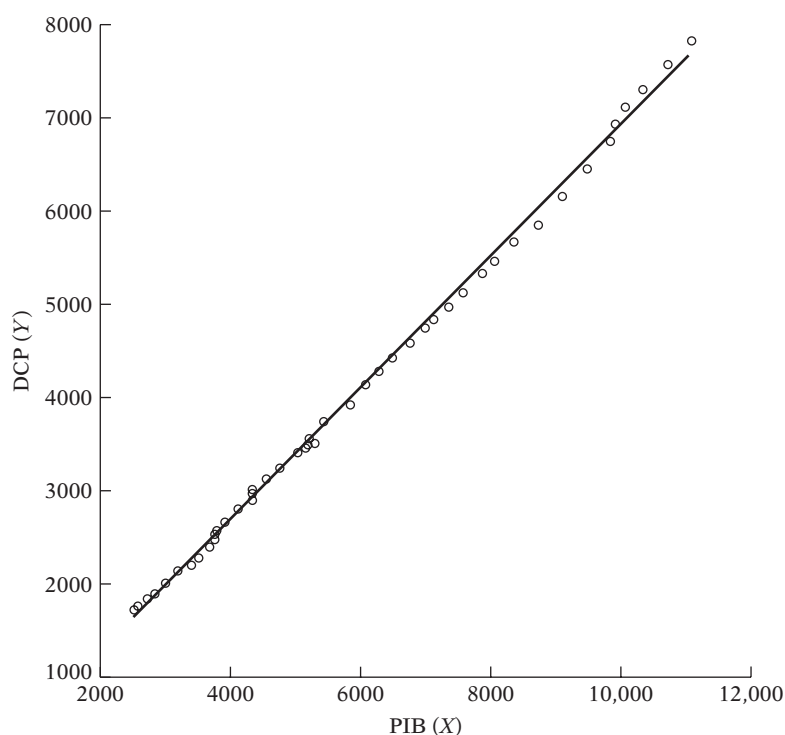
Despesas de consumo pessoal (Y) e produto interno bruto (X), 1960-2005 em bilhões de dólares de 2000

Fonte: *Economics Report of the President*, 2007, Tabela B-2, p, 230.

Ano	DCP(Y)	PIB(X)
1960	1597,4	2501,8
1961	1630,3	2560,0
1962	1711,1	2715,2
1963	1781,6	2834,0
1964	1888,4	2998,6
1965	2007,7	3191,1
1966	2121,8	3399,1
1967	2185,0	3484,6
1968	2310,5	3652,7
1969	2396,4	3765,4
1970	2451,9	3771,9
1971	2545,5	3898,6
1972	2701,3	4105,0
1973	2833,8	4341,5
1974	2812,3	4319,6
1975	2876,9	4311,2
1976	3035,5	4540,9
1977	3164,1	4750,5
1978	3303,1	5015,0
1979	3383,4	5173,4
1980	3374,1	5161,7
1981	3422,2	5291,7
1982	3470,3	5189,3
1983	3668,6	5423,8
1984	3863,3	5813,6
1985	4064,0	6053,7
1986	4228,9	6263,6
1987	4369,8	6475,1
1988	4546,9	6742,7
1989	4675,0	6981,4
1990	4770,3	7112,5
1991	4778,4	7100,5
1992	4934,8	7336,6
1993	5099,8	7532,7
1994	5290,7	7835,5
1995	5433,5	8031,7
1996	5619,4	8328,9
1997	5831,8	8703,5
1998	6125,8	9066,9
1999	6438,6	9470,3
2000	6739,4	9817,0
2001	6910,4	9890,7
2002	7099,3	10048,8
2003	7295,3	10301,0
2004	7577,1	10703,5
2005	7841,2	11048,6

FIGURA I.3

Despesas de consumo pessoal (Y) em relação ao PIB (X), 1960-2005 em bilhões de dólares de 2000.



6. Teste de hipóteses

Considerando que o modelo ajustado seja uma aproximação razoavelmente boa da realidade, é preciso desenvolver critérios adequados para verificar se as estimativas obtidas, digamos, na Equação (I.3.3) estão de acordo com as expectativas da teoria que está sendo testada. Segundo economistas “positivos” como Milton Friedman, uma teoria ou hipótese que não for verificável com evidências empíricas pode não ser admissível como parte de uma pesquisa científica.¹³

Conforme observado anteriormente, Keynes esperava que a PMC fosse positiva, mas menor que 1. Em nosso exemplo, a PMC é de cerca de 0,72. Entretanto, antes de aceitarmos esse valor como uma confirmação da teoria do consumo keynesiana, precisamos nos perguntar se essa estimativa está suficientemente abaixo da unidade para nos convencer de que não é um resultado devido ao acaso ou uma peculiaridade dos dados que utilizamos. Em outras palavras, 0,72 é *estatisticamente menor que 1*? Se for, será um respaldo para a teoria de Keynes.

Tal confirmação ou refutação de teorias econômicas com base em evidências amostrais se alicerça em um ramo da teoria estatística conhecido como **inferência estatística (teste de hipóteses)**. Ao longo do livro, veremos como esse processo é conduzido na prática.

7. Projeção ou previsão

Se o modelo escolhido não refutar a hipótese ou teoria considerada, podemos utilizá-lo para prever o(s) valor(es) futuro(s) da **variável previsão** Y , ou variável dependente, com base no(s) valor(es) futuro(s) conhecidos ou esperados da **variável previsora** X , ou variável explanatória.

¹³ Veja FRIEDMAN, Milton. “The methodology of positive economics.” *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953.

Para fins de ilustração, suponha que desejemos prever as despesas médias de consumo para 2006. O valor do PIB nesse ano foi de \$ 11319,4 bilhão.¹⁴ Colocando o valor do PIB no lado direito da Equação (I.3.3), obtemos:

$$\begin{aligned}\hat{Y}_{2006} &= -299,5913 + 0,7218(11319,4) \\ &= 7870,7516\end{aligned}\tag{I.3.4}$$

ou cerca de \$7.870 bilhões. Assim, dado o valor do PIB, as despesas de consumo médias previstas são de cerca de \$7.870 bilhões. O valor dessas despesas efetivamente registrado em 2006 foi de \$ 8.044 bilhões. Portanto, o modelo estimado (I.3.3) **subestimou** as despesas de consumo reais em cerca de \$ 174 bilhões. Podemos dizer que o **erro de previsão** é de cerca \$174 bilhões, que é aproximadamente de 1,5% do valor do PIB registrado em 2006. Quando examinarmos mais a fundo o modelo de regressão linear nos próximos capítulos, tentaremos verificar se um erro desse tipo é “pequeno” ou “grande”. Mas o importante agora é observar que esses erros de previsão são inevitáveis, dada a natureza estatística de nossa análise.

Há outro uso para o modelo estimado na Equação (I.3.3). Suponha que o presidente decida propor uma redução na alíquota do imposto de renda. Qual seria o efeito dessa política sobre a renda e, por conseguinte, sobre as despesas de consumo e, por fim, sobre o emprego?

Suponha que, como resultado da mudança proposta, as despesas com investimento aumentem. Qual seria o efeito sobre a economia? Como mostra a teoria macroeconômica, a mudança na renda que se segue, digamos, à variação de um dólar nas despesas com investimento é dada pelo **multiplicador da renda** M , que é definido como:

$$M = \frac{1}{1 - PMC}\tag{I.3.5}$$

Se utilizarmos a PMC de 0,72 obtida na Equação (I.3.3), esse multiplicador será de cerca de $M = 3,57$. Isto é, um aumento (redução) de um dólar no investimento levará *por fim* a um aumento (redução) de mais de três vezes na renda; observe que o multiplicador demora a produzir seu efeito.

Um valor crítico nesses cálculos é a PMC, pois o multiplicador depende dela. E essa estimativa da PMC pode ser obtida por meio de modelos de regressão como o da Equação (I.3.3). As estimativas quantitativas da PMC proporcionam informações valiosas para a formulação da política econômica. Conhecendo a PMC, podemos prever o curso futuro da renda, das despesas de consumo e do emprego após uma alteração da política fiscal do governo.

8. Uso do modelo para fins de controle ou de política

Suponha que tenhamos estimada a função consumo da Equação (I.3.3). Suponha, ainda, que o governo acredite que as despesas de consumo de cerca de \$ 8.750 bilhões (em dólares de 2000) manterão a taxa de desemprego em seu nível atual de cerca de 4,2% (no início de 2006). Que nível de renda garantirá o montante almejado (meta) de despesas de consumo?

Se os resultados da regressão da Equação (I.3.3) parecem razoáveis, um cálculo aritmético simples mostrará que:

$$8.750 = -299,5913 + 0,7218 (PIB_{2006})\tag{I.3.6}$$

o que dá aproximadamente $X = 1.2537$. Ou seja, um nível de renda de cerca de \$ 1.2537 bilhões, dada uma PMC de cerca de 0,72, gerará uma despesa de cerca de \$ 8.750 bilhões.

¹⁴ Os dados do DCP e do PIB para 2006 estavam disponíveis, mas os deixamos de lado propositalmente para ilustrar o tópico examinado nesta seção. Como veremos em capítulos subsequentes, é uma boa ideia guardar uma parte dos dados para verificar como o modelo ajustado prevê as observações que estão fora da amostra.

Como esses cálculos sugerem, um modelo estimado pode ser usado para fins de controle ou de formulação de políticas. Com uma combinação apropriada de políticas fiscais e monetárias, o governo pode manejar a **variável de controle X** para gerar o nível desejado da **variável meta Y** .

A Figura I.4 resume a anatomia da modelagem econométrica clássica.

Escolha do modelo

Quando um órgão governamental (por exemplo, o Departamento de Comércio dos Estados Unidos) coleta dados como os apresentados na Tabela I.1, não tem necessariamente uma teoria econômica qualquer em mente. Como sabemos então que os dados realmente confirmam a teoria do consumo keynesiana? Seria porque a função consumo keynesiana (a linha de regressão) da Figura I.3 está extremamente próxima dos dados disponíveis? É possível que outro modelo (teoria) do consumo se ajuste igualmente bem aos dados? Por exemplo, Milton Friedman desenvolveu um modelo de consumo, chamado *hipótese da renda permanente*.¹⁵ Robert Hall também formulou um modelo de consumo, conhecido como *hipótese da renda permanente no ciclo de vida*.¹⁶ Algum desses modelos, ou ambos, também poderiam servir para os da Tabela I.1?

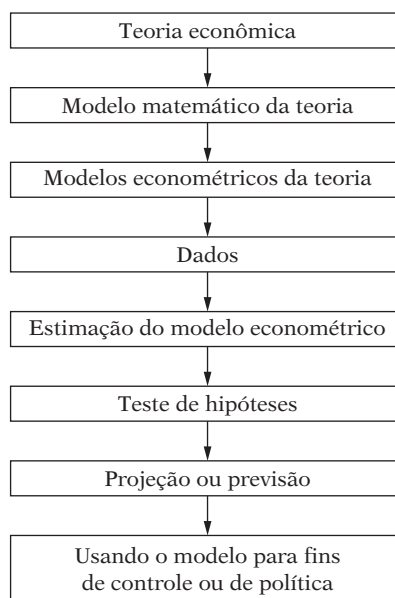
Em resumo, a dúvida com que o pesquisador se depara na prática é como escolher entre as diferentes hipóteses ou modelos para um dado fenômeno, como a relação consumo-renda. Como Miller argumenta:

Nenhum encontro com os dados é uma etapa no sentido de confirmação autêntica a menos que a hipótese lide melhor com os dados que algum rival natural [...]. O que fortalece uma hipótese, nesse caso, é a vitória que, ao mesmo tempo, é a derrota para outra hipótese plausível.¹⁷

Como, então, escolher entre os diversos modelos ou hipóteses concorrentes? Vale a pena ter em mente o conselho de Clive Granger:¹⁸

FIGURA I.4

Anatomia da modelagem econométrica.



¹⁵ FRIEDMAN, Milton. *A theory of consumption function*. Princeton: N. J.: Princeton University Press, 1957.

¹⁶ HALL, R. "Stochastic implications of the life cycle permanent income hypothesis: theory and evidence." *Journal of Political Economy*, 1978. v. 86, p. 971-987.

¹⁷ MILLER, R. W. *Fact and method: Explanation, confirmation, and reality in the natural and social sciences*. Princeton, N.J.: Princeton University Press, 1978. p. 176.

¹⁸ GRANGER, Clive W. J. *Empirical modeling in economics*. Reino Unido: Cambridge University Press, 1999. p. 58.

Gostaria de sugerir que, no futuro, você faça as seguintes perguntas quando lhe apresentarem uma nova teoria ou modelo empírico:

- (i) Qual o propósito disto? Para que decisão econômica contribuirá? e;
- (ii) Existe alguma prova que me permita avaliar sua qualidade em comparação com teorias ou modelos alternativos?

Penso que, se for dada a devida atenção a essas perguntas, a pesquisa e a discussão econômicas serão fortalecidas.

À medida que avançarmos no livro, encontraremos várias hipóteses que concorrem para explicar os vários fenômenos econômicos. Por exemplo, os estudantes de economia conhecem bem o conceito da função de produção, que é basicamente uma relação entre produto e insumos (capital e trabalho). Na literatura, duas das mais conhecidas são as funções *Cobb-Douglas* e a da *elasticidade de substituição constante*. Precisaremos descobrir, em virtude dos dados de produção e de insumos, se alguma delas, reflete melhor os dados.

O método econométrico clássico de oito etapas apresentado anteriormente é neutro no sentido de que pode ser usado para testar qualquer uma dessas hipóteses rivais.

É possível formular uma metodologia que seja suficientemente abrangente para incluir hipóteses concorrentes? Esse é um tópico complexo e controverso que será discutido no Capítulo 13, depois que tivermos adquirido conhecimentos teóricos suficientes.

I.4 Tipos de econometria

Como sugere a estrutura classificatória da Figura I.5, a econometria pode ser dividida em duas categorias amplas: **econometria teórica** e **econometria aplicada**. Em cada categoria podemos abordar a disciplina segundo as tradições **clássica** e **bayesiana**. Neste livro, damos ênfase à clássica. Quanto à abordagem bayesiana, o leitor pode consultar as referências indicadas no final do capítulo.

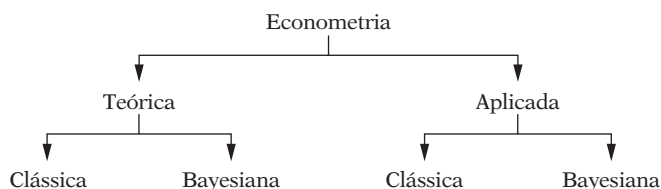
A econometria teórica trata do desenvolvimento de métodos adequados para medir as relações econômicas especificadas nos modelos econométricos. Sob esse aspecto, a econometria depende fortemente da estatística matemática. Por exemplo, um dos métodos mais usado neste livro é o dos **mínimos quadrados**. A econometria teórica deve deixar claras as hipóteses deste método, suas propriedades e o que acontece com elas quando uma ou mais hipóteses do método não são atendidas.

Na econometria aplicada, utilizamos as ferramentas da econometria teórica para estudar um ou mais campos especiais da economia e dos negócios, como a função de produção, a função investimento, as funções de oferta e de demanda, a teoria do portfólio etc.

Este livro trata, em grande parte, a formulação dos métodos econométricos, suas hipóteses, usos, limitações. Esses métodos são ilustrados com exemplos de várias áreas da economia e dos negócios. No entanto, este *não* é um livro de econometria aplicada, no sentido de se aprofundar particularmente em quaisquer dos campos da aplicação econômica. Essa tarefa cabe às obras escritas especificamente com esse fim. Ao final do livro, ofereceremos algumas referências bibliográficas.

FIGURA I.5

Categorias da econometria.



I.5 Pré-requisitos matemáticos e estatísticos

Embora este livro seja escrito em um nível elementar, o autor pressupõe que o leitor esteja familiarizado com os conceitos básicos da estimação estatística e do teste de hipóteses. O Apêndice A oferece uma visão geral e não técnica dos conceitos estatísticos básicos empregados aqui para aqueles que desejam reciclar seus conhecimentos. No que se refere à matemática, é desejável ter uma noção sobre cálculo diferencial, embora não seja essencial. Apesar de a maioria dos livros de econometria voltados à pós-graduação utilizar amplamente a álgebra matricial, quero deixar claro que ela não é necessária para tirar proveito deste livro. Estou completamente convencido de que as ideias fundamentais da econometria podem ser transmitidas sem recorrer à álgebra matricial. No entanto, para os estudantes mais inclinados à matemática, o Apêndice C apresenta um resumo da teoria básica da regressão em notação matricial e o Apêndice B oferece um resumo dos principais resultados da álgebra matricial.

I.6 O papel do computador

A análise de regressão, a principal ferramenta da econometria, é impensável nos dias de hoje sem o computador e o acesso a alguns softwares estatísticos ou pacotes estatísticos. (Acreditem em mim, sou da geração criada com a régua de cálculo!) Felizmente, vários pacotes excelentes para regressões estão disponíveis no mercado, tanto para computadores de grande porte quanto para microcomputadores, e a lista cresce a cada dia. Softwares como **ET**, **LIMDEP**, **SHAZAM**, **MICRO TSP**, **MINITAB**, **EViews**, **SAS**, **SPSS**, **STATA**, **Microfit**, **PcGive** e **BMD** atendem à maioria das técnicas e testes econométricos examinados neste livro.

Ao longo do livro, o leitor será convidado vez por outra a conduzir experimentos **Monte Carlo** com auxílio de um ou mais pacotes estatísticos. Os experimentos **Monte Carlo** são “divertidos” e permitirão ao leitor apreciar as propriedades de vários métodos estatísticos utilizados aqui. Os detalhes dos experimentos Monte Carlo serão discutidos no momento apropriado.

I.7 Sugestões para leituras complementares

O tema metodologia econométrica é vasto e controverso. Para os interessados, sugiro os livros a seguir:

MARCHI, Neil de; GILBERT, Christopher. (Eds.). *History and methodology of econometrics*. Nova York: Oxford University Press, 1989. Esta coletânea de textos discute alguns trabalhos pioneiros em metodologia econométrica e examina extensamente a abordagem britânica da econometria e sua relação com séries temporais, ou seja, dados coletados ao longo do tempo.

CHAREMZA, Wojciech W.; DEADMAN, Derek F. *New directions in econometric practice: general to specific modelling, cointegration and vector autoregression*. 2. ed. Hants, Inglaterra: Edward Elgar Publishing Ltd., 1997. Os autores criticam a abordagem tradicional da econometria e apresentam uma exposição detalhada das novas abordagens ao método econométrico.

DARNELL, Adrian C.; EVANS, J. Lynne. *The limits of econometrics*. Hants, Inglaterra: Edward Elgar Publishers Ltd., 1990. Este livro oferece um exame bastante equilibrado das várias abordagens metodológicas econométricas, com uma fidelidade renovada ao método tradicional.

MORGAN, Mary S. *The history of econometric ideas*. Nova York: Cambridge University Press, 1990. A autora oferece uma excelente perspectiva histórica da teoria e da prática econométricas, com um exame profundo das contribuições iniciais de Haavelmo (ganhador do Nobel de Economia de 1990) à econometria. No mesmo espírito, o livro de David F. Hendry e Mary S. Morgan, *The foundation of econometric analysis*, Reino Unido: Cambridge University Press, 1995, reúne uma seleção de textos seminais para mostrar a evolução das ideias econométricas ao longo do tempo.

COLANDER, David; BRENNER, Reuven. (Eds.). *Educating economists*. Ann Arbor, Michigan: University of Michigan Press, 1992. O livro apresenta uma visão crítica, às vezes agnóstica, do ensino e da prática econômicos.

Para os interessados em estatística e econometria bayesianas, os seguintes livros são muito úteis: DEY, John H. *Data in doubt*. Inglaterra: Basil Blackwell Ltd., Oxford University Press, 1985; Peter, M. LEE. *Bayesian statistics: an introduction*. Inglaterra: Oxford University Press, 1989; e PORIER, Dale J. *Intermediate statistics and econometrics: a comparative approach*. Cambridge, Massachusetts: MIT Press, 1995. ZELLER, Arnold. *An introduction to bayesian inference in econometrics*. Nova York: John Wiley & Sons, 1971, este é um livro de referência avançado. Outro livro de referência avançado é *Palgrave handbook of econometrics: volume 1: econometric theory*, editado por Terence C. Mills e Kerry Patterson, Nova York: Palgrave Macmillan, 2007.

Modelos de regressão com equação única

A Parte I do livro apresenta modelos de regressão com equação única. Nesses modelos, uma variável, chamada *variável dependente*, é expressa como função linear de uma ou mais variáveis, denominadas *variáveis explanatórias*. Em tais modelos, supõe-se implicitamente que as relações causais, se existirem, entre a variável dependente e as explanatórias ocorrem apenas em uma direção, especificamente, das variáveis explanatórias para a dependente.

No Capítulo 1, discutimos a interpretação histórica, assim como a moderna, do termo regressão e ilustramos a diferença entre elas com vários exemplos da economia e de outros campos.

No Capítulo 2, apresentamos alguns conceitos fundamentais da análise de regressão, com auxílio de um modelo de regressão linear de duas variáveis, um modelo em que a variável dependente é expressa como função linear de uma única variável explanatória.

No Capítulo 3, continuamos abordando o modelo de duas variáveis e apresentamos o que é conhecido como o *modelo clássico de regressão linear*, um modelo que adota várias hipóteses simplificadoras. Com essas premissas, expomos o *método dos mínimos quadrados ordinários* (MQO) para estimar os parâmetros do modelo de regressão com duas variáveis. A aplicação do método dos MQO é simples e tem algumas propriedades estatísticas muito desejáveis.

No Capítulo 4, apresentamos o modelo de regressão linear clássico *normal* (com duas variáveis), um modelo que pressupõe que a variável dependente estocástica siga a distribuição de probabilidade normal. Desse modo, os estimadores de mínimos quadrados ordinários obtidos no Capítulo 3 adquirem algumas propriedades estatísticas mais fortes do que o modelo de regressão linear clássico não normal — propriedades que nos permitem realizar a inferência estatística, isto é, testar hipóteses.

O Capítulo 5 é dedicado ao tópico do teste de hipóteses. Tentamos verificar se os coeficientes de regressão estimados são compatíveis com as hipóteses feitas em relação ao valor desses coeficientes, sendo que os valores hipotéticos são sugeridos pela teoria e/ou por trabalhos empíricos anteriores.

O Capítulo 6 aborda algumas extensões do modelo de regressão com duas variáveis. Em particular, discute tópicos como: (1) regressão que passa pela origem; (2) escalas e unidades de medidas; e (3) formas funcionais dos modelos de regressão tais como log-log, semilog e modelos recíprocos.

No Capítulo 7, consideramos o modelo de regressão múltipla em que há mais de uma variável explanatória e mostramos como o método dos mínimos quadrados ordinários pode ser ampliado para estimar os parâmetros de tais modelos.

No Capítulo 8, estendemos os conceitos apresentados no Capítulo 5 ao modelo de regressão múltipla e destacamos algumas das complicações que surgem com a introdução de diversas variáveis explanatórias.

O Capítulo 9, sobre variáveis explanatórias binárias (*dummy*) ou qualitativas, conclui a primeira parte do livro. O capítulo ressalta que nem todas as variáveis explanatórias precisam ser quantitativas (isto é, ter escalas proporcionais). Variáveis como gênero, raça, religião, nacionalidade e região de residência não podem ser prontamente quantificadas, contudo desempenham um papel importante para explicar vários fenômenos econômicos.

A natureza da análise de regressão

Conforme mencionado na Introdução, a regressão é a principal ferramenta da econometria e neste capítulo examinaremos sucintamente sua natureza.

1.1 Origem histórica do termo regressão

O termo regressão foi criado por Francis Galton. Em um artigo famoso, Galton verificou que, embora existisse uma tendência de que pais altos tivessem filhos altos e pais baixos tivessem filhos baixos, a estatura média das crianças nascidas de pais com uma dada altura tendia a mover-se ou "regredir" à altura média da população como um todo.¹ Em outras palavras, a altura dos filhos de pais mais altos ou mais baixos que o padrão tende a mover-se no sentido da altura média da população. A lei da regressão universal de Galton foi confirmada por seu amigo Karl Pearson, que coletou mais de mil registros de altura de membros de grupos familiares.² Ele constatou que a altura média dos filhos de um grupo de pais altos era menor do que a de seus pais e que a altura média de um grupo de filhos de pais baixos era maior do que a de seus pais; portanto, filhos de pais altos e baixos "regrediam" igualmente à altura média de todos os homens. Nas palavras de Galton, isso era uma "regressão à mediocridade".

1.2 A interpretação moderna da regressão

Contudo, a interpretação moderna da regressão é bastante diferente. De modo geral, podemos dizer:

A análise de regressão diz respeito ao estudo da dependência de uma variável, a *variável dependente*, em relação a uma ou mais variáveis, as *variáveis explanatórias*, visando estimar e/ou prever o valor médio (da população) da primeira em termos dos valores conhecidos ou fixados (em amostras repetidas) das segundas.

A importância desta visão da análise de regressão ficará mais clara à medida que avançarmos, mas alguns exemplos simples podem esclarecer o conceito básico.

Exemplos

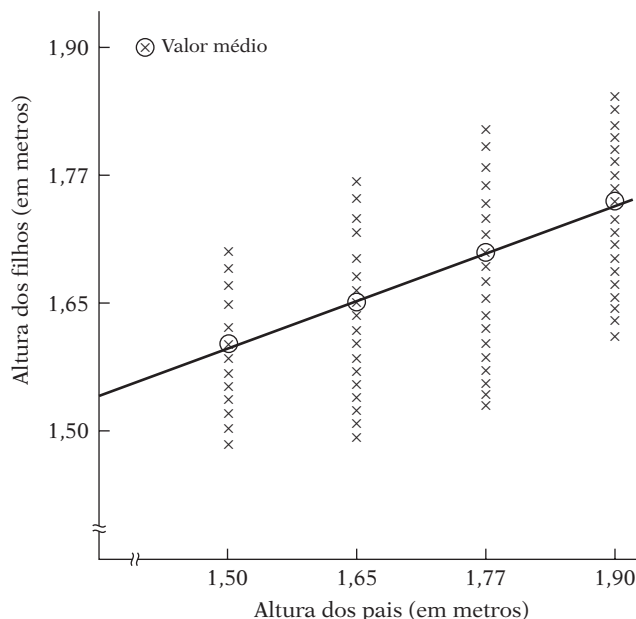
1. Voltemos à lei da regressão universal de Galton. Ele estava interessado em descobrir por que havia estabilidade na distribuição de alturas em uma população. Mas, na visão moderna, nossa preocupação não é essa, mas sim descobrir como a altura *média* dos filhos varia, dada a altura dos pais. Estamos interessados em prever a altura média dos filhos com base no conhecimento da altura dos pais. Para melhor entender o conceito, observe a Figura 1.1, um **diagrama de dispersão**.

¹ GALTON, Francis. "Family likeness in stature". *Proceedings of Royal Society*, Londres, 1886. v. 40, p. 42-72.

² PEARSON, K.; LEE, A. "On the laws of inheritance". *Biometrika*, nov. 1903. v. 2, p. 357-462.

FIGURA 1.1

Distribuição hipotética das alturas dos filhos em relação à altura dos pais.



A figura mostra a distribuição das alturas dos filhos, em uma população hipotética, correspondentes aos valores dados, ou *fixos*, da altura do pai. Note que, para cada altura do pai dada, há uma *faixa* ou distribuição de alturas dos filhos. Observe que apesar da variação da altura dos filhos para um dado valor da altura dos pais, a altura média dos filhos, em geral, aumenta com o aumento da altura do pai. As cruzes assinaladas com um círculo indicam a altura *média* dos filhos correspondente a uma dada altura dos pais. Conectando essas médias, obtemos a linha mostrada na figura. Essa linha, como veremos, é conhecida como **linha de regressão**. Ela mostra como a altura *média* dos filhos aumenta com a altura dos pais.³

- Consideremos o diagrama de dispersão da Figura 1.2, que apresenta a distribuição, em uma população hipotética, das alturas de meninos em idades *fixas*. Para cada idade, temos uma faixa, ou distribuição, de alturas. Obviamente, nem todos os meninos de uma mesma idade terão uma altura idêntica.

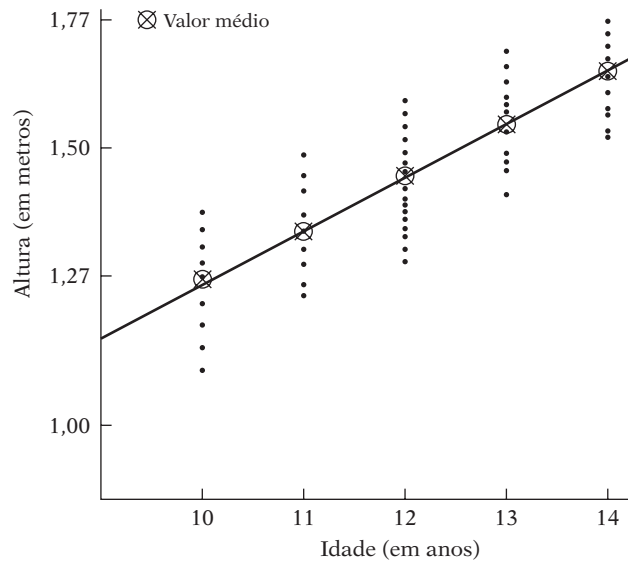
Mas a altura, *em média*, aumenta com a idade (até certa idade, é claro), o que pode ser visto claramente se traçarmos uma linha (a de regressão) que passe pelos pontos circulados que representam a altura média em cada idade. Conhecendo a idade, podemos prever, por meio da linha de regressão, a altura média correspondente a essa idade.

- Passando a exemplos econômicos, um economista poderia estar interessado em estudar a relação de dependência das despesas de consumo pessoal e a renda pessoal disponível, após o pagamento de impostos. Essa análise é útil para estimar a propensão marginal a consumir (PMC), isto é, a variação média nas despesas de consumo, para uma variação de um dólar na renda real.
- Um monopolista com possibilidade de fixar o preço ou a produção (mas não ambos) pode querer descobrir a resposta da demanda por um produto perante variações nos preços. Isso nos permitiria estimar a **elasticidade-preço** (isto é, a resposta dos preços) da demanda pelo produto e contribuiria para determinar o preço mais lucrativo.
- Um economista do trabalho pode querer estudar a relação entre a variação dos salários nominais e a taxa de desemprego. O diagrama de dispersão da Figura 1.3 mostra os dados históricos. A curva traçada é um exemplo da famosa *curva de Phillips*, que relaciona as variações nos salários

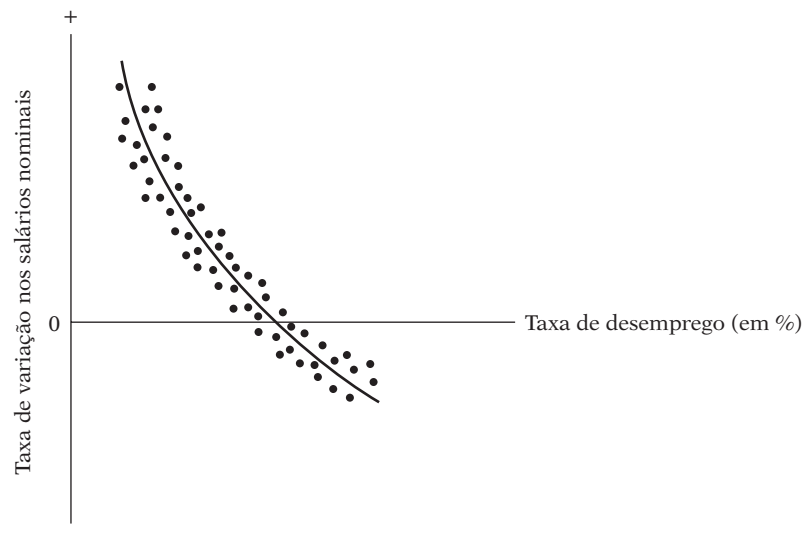
³ Neste estágio do desenvolvimento de nosso assunto, chamaremos a linha de regressão de *linha que conecta o valor médio da variável dependente (altura dos filhos) correspondente a um dado valor da variável explanatória (altura dos pais)*. Note que essa linha tem uma inclinação positiva, mas menor que 1, o que está de acordo com a regressão à mediocridade de Galton. (Por quê?)

FIGURA 1.2

Distribuição hipotética de alturas em relação a idades selecionadas.

**FIGURA 1.3**

Curva de Phillips hipotética.

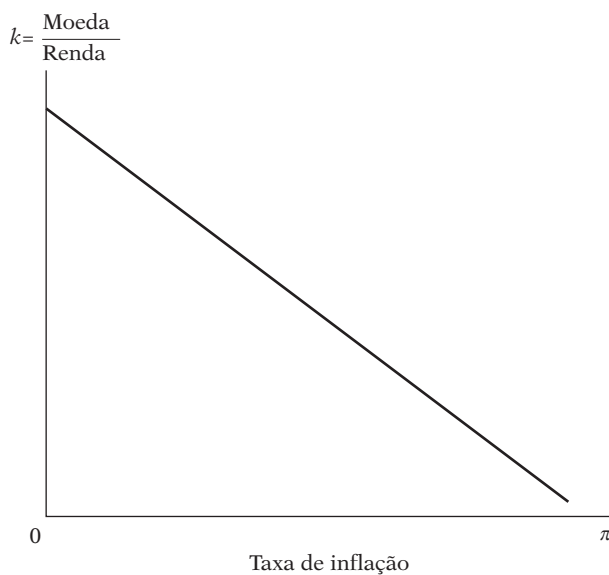


nominais com a taxa de desemprego. Esse diagrama de dispersão permitiria ao economista prever a variação média dos salários para uma dada taxa de desemprego. Tal conhecimento poderia contribuir para esclarecer o processo inflacionário de uma economia, visto que o aumento dos salários nominais tende a refletir-se em aumento de preços.

6. A economia monetária diz que, sendo tudo o mais constante, quanto mais alta a taxa de inflação, π , menor a proporção k de sua renda que as pessoas desejam manter sob forma monetária, conforme mostra a Figura 1.4. A curva da linha representa a mudança em k em relação ao índice de inflação. Uma análise quantitativa dessa relação permitiria ao economista prever a quantidade de moeda, como proporção de sua renda, que as pessoas desejariam manter sob diferentes taxas de inflação.
7. O diretor de *marketing* de uma empresa pode querer saber como a demanda dos produtos de sua empresa relaciona-se com as despesas com publicidade. Esse estudo seria de grande utilidade para determinar a **elasticidade da demanda** em relação às despesas com publicidade, isto é, a variação percentual da demanda em resposta a uma variação de 1% nas despesas com publicidade. Essa informação pode ser útil para determinar o orçamento de publicidade "ideal".

FIGURA 1.4

Saldos monetários em relação à taxa de inflação, π .



8. Por fim, um agrônomo pode estar interessado em estudar a dependência do rendimento de determinada plantação em relação à temperatura, à quantidade de chuva e de sol e à aplicação de fertilizantes. A análise de dependência permitiria a ele prever ou prognosticar o rendimento médio da lavoura, dadas as informações sobre as variáveis explanatórias.

O leitor pode imaginar muitos outros exemplos da dependência de uma variável em relação a uma ou mais variáveis. As técnicas de análise de regressão examinadas neste livro destinam-se especialmente a estudar essa dependência entre variáveis.

1.3 Relações estatísticas *versus* determinísticas

Nos exemplos citados na Seção 1.2, o leitor notou que, na análise de regressão, estamos preocupados com o que é conhecido como dependência *estatística*, e não *funcional* ou *determinística*, entre as variáveis, como aquelas da física clássica. Nas relações estatísticas entre variáveis, lidamos essencialmente com variáveis **aleatórias** ou **estocásticas**,⁴ isto é, variáveis que têm distribuições probabilísticas. Por outro lado, na dependência funcional ou determinística, também lidamos com variáveis, mas estas não são aleatórias ou estocásticas.

A dependência do rendimento das lavouras em relação à temperatura, pluviosidade, luz solar ou fertilizante, por exemplo, é de natureza estatística no sentido de que as variáveis explanatórias, embora importantes, não permitirão ao agrônomo prever exatamente o rendimento devido aos erros envolvidos na medição dessas variáveis, assim como a diversos outros fatores (variáveis) que coletivamente afetam o rendimento, mas cuja identificação individual pode ser difícil. Uma variabilidade "intrínseca" ou aleatória tende a existir na variável dependente "rendimento da lavoura" que não pode ser totalmente explicada independentemente do número de variáveis explanatórias consideradas.

Por outro lado, nos fenômenos determinísticos, lidamos com relações do tipo exibido pela lei da gravidade de Newton, que diz: cada partícula do universo atrai todas as outras partículas com uma força diretamente proporcional ao produto de suas massas e inversamente proporcional ao quadrado da distância entre elas. Simbolicamente, $F = k(m_1 m_2 / r^2)$, em que F = força, m_1 e m_2 são as massas das duas partículas, r é a distância entre elas e k = constante de proporcionalidade. Outro exemplo é a lei de Ohm,

⁴A palavra *estocástica* vem do grego *stokhos*, que significa "olho de boi" [ou "centro do alvo" em inglês]. O resultado do lançamento de um dardo ao alvo é um processo estocástico, isto é, um processo em que nem sempre o centro do alvo é atingido.

que diz: no caso de condutores metálicos dentro de certa faixa de temperatura, a corrente C é proporcional à voltagem V ; isto é $C = (\frac{1}{k})V$, em que $\frac{1}{k}$ é a constante de proporcionalidade. Outros exemplos dessas relações determinísticas são a lei de Boyle para o gás, a lei de Kirchhoff para a eletricidade e a de Newton para o movimento.

Neste livro, não nos preocupamos com essas relações determinísticas. Obviamente, se houver erros de medição no k da lei da gravidade de Newton, a relação determinística torna-se uma relação estatística. Nessa situação, a força só pode ser prevista aproximadamente com base no valor de k (e m_1 , m_2 e r), que contém erros. A variável F , nesse caso, torna-se uma variável aleatória.

1.4 Regressão versus causação

Embora a análise de regressão lide com a dependência de uma variável em relação a outras, isso não implica necessariamente uma causação. Nas palavras de Kendall e Stuart, "uma relação estatística, por mais forte e sugestiva que seja, nunca pode estabelecer uma conexão causal: nossas ideias de causação devem vir de fora da estatística, em última análise, de alguma teoria".⁵

No exemplo do rendimento da lavoura citado anteriormente, não há *razão estatística* para supor que a chuva não dependa do rendimento da lavoura. O fato de tratarmos o rendimento da lavoura como dependente (dentre outras coisas) da chuva decorre de considerações não estatísticas: o senso comum sugere que a relação não pode ser invertida, pois não podemos controlar a pluviosidade por meio de uma variação no rendimento da lavoura.

Em todos os exemplos citados na Seção 1.2, é importante notar que **uma relação estatística por si própria não implica logicamente uma causação**. Para atribuir causação, devemos recorrer a considerações *a priori* ou teóricas. Portanto, no terceiro exemplo citado, podemos invocar a teoria econômica para dizer que as despesas de consumo dependem da renda real.⁶

1.5 Regressão versus correlação

A **análise de correlação**, cujo principal objetivo é medir a *força* ou o *grau* de *associação linear* entre duas variáveis está estreitamente relacionada à análise de regressão, mas conceitualmente é muito diferente. O **coeficiente de correlação**, que estudaremos em detalhe no Capítulo 3, mede a força dessa associação (linear). Por exemplo, podemos estar interessados em determinar o coeficiente de correlação entre: fumar e câncer de pulmão; entre as notas obtidas nas provas de estatística e de matemática; entre as notas obtidas no ensino médio e na faculdade e assim por diante. Na análise de regressão, como já mencionamos, não estamos interessados prioritariamente nessa medida. Em vez disso, buscamos estimar ou prever o valor médio de uma variável com base nos valores fixos de outras variáveis. Portanto, talvez queiramos saber se é possível prever a nota média de uma prova de estatística conhecendo as notas do estudante na prova de matemática.

A regressão e a correlação têm algumas diferenças fundamentais que vale a pena mencionar. Na análise de regressão, existe uma assimetria na maneira como as variáveis dependente e explanatória são tratadas. Supomos que a variável dependente seja estatística, aleatória ou estocástica, isto é, que tenha distribuição probabilística. Por outro lado, consideramos que as variáveis explanatórias têm valores fixos (em amostras repetidas),⁷ que foram explicitados na definição de regressão dada na Seção 1.2. Portanto, na Figura 1.2, supusemos que a variável "idade" foi fixada em dados níveis e

⁵ KENDALL, M. G.; STUART, A. *The advanced theory of statistics*. Nova York: Charles Griffin Publishers, 1961. v. 2, cap. 26, p. 279.

⁶ Mas, como veremos no Capítulo 3, a análise de regressão clássica se alicerça no pressuposto de que o modelo utilizado na análise é o correto. Portanto, a direção da causalidade pode estar implícita no modelo postulado.

⁷ É fundamental observar que as variáveis explanatórias podem ser intrinsecamente estocásticas, mas, para fins de análise de regressão, pressupomos que seus valores são fixados em amostras repetidas (isto é, que X assume os mesmos valores em várias amostras), tornando-as, assim, não aleatórias ou não estocásticas. Veremos esse assunto com mais detalhes no Capítulo 3, Seção 3.2.

que as medições de altura foram obtidas nesses níveis. Na análise de correlação, por outro lado, tratamos quaisquer (duas) variáveis simetricamente; não há distinção entre as variáveis dependente e explanatória. Afinal, a correlação entre as notas nas provas de matemática e de estatística é a mesma que aquela entre as notas das provas de estatística e de matemática. Além disso, supõe-se que as duas variáveis sejam aleatórias. Como veremos, a maior parte da teoria da correlação baseia-se na premissa da aleatoriedade das variáveis, enquanto boa parte da teoria da regressão a ser exposta neste livro está condicionada à premissa de que a variável dependente é estocástica, mas as variáveis explanatórias são fixas ou não estocásticas.⁸

1.6 Terminologia e notação

Antes de prosseguirmos para a análise formal da teoria da regressão, vejamos brevemente a terminologia e a notação. Na literatura, os termos *variável dependente* e *variável explanatória* são descritos de vários modos. Uma lista representativa é:

Variável dependente	Variável explicativa
↕	↕
Variável explicada	Variável independente
↕	↕
Variável prevista	Previsor
↕	↕
Regressando	Regressor
↕	↕
Resposta	Estímulo
↕	↕
Variável endógena	Variável exógena
↕	↕
Saída	Entrada
↕	↕
Variável controlada	Variável de controle

Embora seja uma questão de gosto pessoal e de tradição, neste texto adotaremos a terminologia variável dependente/variável explanatória ou os termos mais neutros regressando e regressor.

Se estudamos a dependência de uma variável em relação a uma única variável explanatória, como é o caso das despesas de consumo em relação à renda real, esse estudo é conhecido como **análise de regressão simples** ou **de duas variáveis**. No entanto, se estudarmos a dependência de uma variável a mais de uma variável explanatória, como no caso da relação entre rendimento da lavoura e chuva, temperatura, luz do sol e fertilizantes, será uma **análise de regressão múltipla**. Em outras palavras, em uma regressão de duas variáveis há somente uma única variável explanatória, enquanto na regressão múltipla há mais de uma variável explanatória.

O termo **aleatório** é sinônimo de **estocástico**. Como já mencionado, uma variável aleatória, ou estocástica, é aquela que pode assumir qualquer valor, positivo ou negativo, dentro de um conjunto de valores com uma dada probabilidade.⁹

⁸No tratamento avançado da econometria, pode-se relaxar a premissa do caráter não estocástico das variáveis explanatórias (veja introdução à Parte 2).

⁹Veja o **Apêndice A** para uma definição formal e detalhes adicionais.

A menos que seja definido de outra forma, a letra Y denotará a variável dependente e os X ($X_1, X_2, \dots, X_k$) representarão as variáveis explanatórias, sendo X_k a k -ésima variável explanatória; os subscritos i ou t denotarão a i -ésima ou t -ésima observação ou valor; X_{ki} (ou X_{kt}) denotará a i -ésima (ou t -ésima) observação da variável X_k ; N (ou T) denotará o número total de observações ou valores da população; e n (ou t), o número total de observações de uma amostra. Por convenção, o subscrito i da observação será usado para dados de **corte transversal** (isto é, dados coletados em um ponto no tempo) e o subscrito t para os dados de **séries temporais** (isto é, dados coletados ao longo de um intervalo de tempo). A natureza dos dados de corte transversal e das séries temporais, bem como o importante tópico da natureza e das fontes de dados para a análise empírica, é examinada a seguir.

1.7 Natureza e fonte dos dados para a análise econômica¹⁰

O sucesso de qualquer análise econométrica depende, em última instância, da disponibilidade de dados adequados. Portanto é essencial dedicarmos algum tempo examinando a natureza, as fontes e as limitações dos dados que podem aparecer na análise empírica.

Tipos de dados

Três tipos de dados podem estar disponíveis para a análise empírica: dados de **séries temporais**, de **corte transversal** e **combinados** (isto é, a combinação de séries temporais com os dados de corte transversal).

Séries temporais

Os dados apresentados na Tabela 1.1 da Introdução são um exemplo de dados de séries temporais. Uma *série temporal* é um conjunto de observações dos valores que uma variável assume em diferentes momentos do tempo. Esses dados podem ser coletados a intervalos regulares, como **diariamente** (preços das ações, relatórios meteorológicos), **semanalmente** (informações sobre oferta de moeda), **mensalmente** (taxa de desemprego, índice de preços ao consumidor [IPC]), **trimestralmente** (PIB), **anualmente** (orçamento do governo), **quinquenalmente**, isto é, a cada cinco anos (censo industrial dos Estados Unidos), ou **decenalmente** (censo demográfico). Às vezes, os dados estão disponíveis em séries trimestrais e anuais, como no caso do PIB e das despesas de consumo. Com o advento dos computadores de alta velocidade, os dados agora podem ser coletados a intervalos extremamente curtos, como os relativos a preços das ações, obtidos de forma praticamente contínua (as chamadas *cotações em tempo real*).

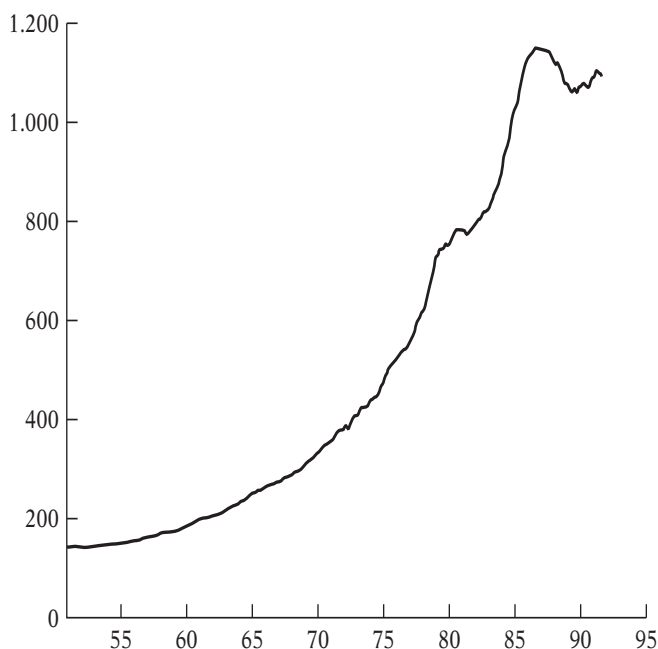
Embora as séries temporais sejam muito usadas nos estudos econométricos, apresentam problemas especiais para o econometrista. Como veremos mais adiante nos capítulos sobre **econometria das séries temporais**, a maioria dos estudos empíricos embasados nesse tipo de dado pressupõe que a série temporal subjacente seja **estacionária**. Embora ainda seja muito cedo para apresentar o significado técnico preciso dessa característica, *de maneira geral, uma série é estacionária se sua média e variância não variam sistematicamente ao longo do tempo*. Para entender o que isso significa, considere a Figura 1.5, que ilustra o comportamento da oferta de moeda, no seu conceito de M1, nos Estados Unidos de 1º de janeiro de 1959 a setembro de 1999. (Os dados numéricos são apresentados no Exercício 1.4.) Como podemos ver na figura, a oferta de moeda M1 mostra uma firme **tendência** ascendente, assim como uma variabilidade ao longo dos anos, sugerindo que a série temporal de M1 não é estacionária.¹¹ Exploraremos este tópico por completo no Capítulo 21.

¹⁰ Para um relato informativo, veja o livro de INTRILIGATOR, Michael D. *Econometric models, techniques and applications*. Englewood Cliffs, N. J.: Prentice Hall, 1978.

¹¹ Para melhor entendermos, dividimos os dados em quatro períodos: de janeiro de 1951 a dezembro de 1962, de janeiro de 1963 a dezembro de 1974; de janeiro de 1975 a dezembro de 1986; e de janeiro de 1987 a setembro de 1999. Os valores médios da oferta de moeda (e seus desvios padrão entre parênteses) foram de, respectivamente, 165,88 (23,27); 323,20 (72,66); 788,12 (195,43); e 1.099 (27,84), todos os dados em bilhões de dólares. Isso é uma indicação grosseira do fato de que a oferta de moeda não foi estacionária durante todo o período.

FIGURA 1.5

Oferta de moeda M1,
nos Estados Unidos,
janeiro de 1951 a
setembro de 1999.



Dados em corte transversal

Estes são dados em que uma ou mais variáveis foram coletadas *no mesmo ponto do tempo*, como o censo demográfico que é feito a cada dez anos (o mais recente é de 2000), as pesquisas de despesas do consumidor conduzidas pela Universidade de Michigan e, naturalmente, as pesquisas de opinião feitas pelo Gallup e inúmeras outras organizações. A Tabela 1.1 apresenta um exemplo concreto de dados em corte transversal. Nela estão dados da produção e dos preços dos ovos nos 50 Estados americanos nos anos de 1990 e 1991.

Para cada ano, os dados dos 50 Estados são um corte transversal. Portanto, na Tabela 1.1 temos duas amostras em corte transversal.

Assim como as séries temporais têm problemas especiais (devido à questão do caráter estacionário), os dados em corte transversal também têm seus problemas, especificamente o da *heterogeneidade*. Nos dados da Tabela 1.1 podemos ver que alguns estados produzem uma imensa quantidade de ovos (por exemplo, a Pensilvânia) e outros, muito pouco (como o Alasca). Quando incluímos dados tão heterogêneos em uma análise estatística, o **efeito magnitude** ou **escala** deve ser levado em conta a fim de não misturarmos maçãs com laranjas. Para melhor entendermos, na Figura 1.6, plotamos os dados sobre ovos produzidos e seus preços nos 50 Estados no ano de 1990. Esta figura mostra como as observações estão dispersas. No Capítulo 11, veremos como o efeito escala pode ser um fator importante ao estimarmos relações entre variáveis econômicas.

Dados combinados

Nos dados combinados, há elementos tanto de séries temporais quanto de corte transversal. Nos dados da Tabela 1.1, há um exemplo de dados combinados. Para cada um dos anos, temos 50 observações de corte transversal e, para cada estado, duas observações de séries temporais de preços e quantidade de ovos, em um total de cem observações combinadas.

Do mesmo modo, os dados no Exercício 1.1 são dados combinados, pois o índice de Preços ao Consumidor (IPC) de cada país para o período 1980-2005 é uma série temporal, enquanto os dados do IPC para os sete países em um único ano são de corte transversal. Nos dados combinados, temos um total de 182 observações — 26 observações anuais para cada um dos sete países.

TABELA 1.1 Produção de ovos nos Estados Unidos

Estado	Y_1	Y_2	X_1	X_2	Estado	Y_1	Y_2	X_1	X_2
AL	2.206	2.186	92,7	91,4	MT	172	164	68,0	66,0
AK	0,7	0,7	151,0	149,0	NE	1.202	1.400	50,3	48,9
AZ	73	74	61,0	56,0	NV	2,2	1,8	53,9	52,7
AR	3.620	3.737	86,3	91,8	NH	43	49	109,0	104,0
CA	7.472	7.444	63,4	58,4	NJ	442	491	85,0	83,0
CO	788	873	77,8	73,0	NM	283	302	74,0	70,0
CT	1.029	948	106,0	104,0	NY	975	987	68,1	64,0
DE	168	164	117,0	113,0	NC	3.033	3.045	82,8	78,7
FL	2.586	2.537	62,0	57,2	ND	51	45	55,2	48,0
GA	4.302	4.301	80,6	80,8	OH	4.667	4.637	59,1	54,7
HI	227,5	224,5	85,0	85,5	OK	869	830	101,0	100,0
ID	187	203	79,1	72,9	OR	652	686	77,0	74,6
IL	793	809	65,0	70,5	PA	4.976	5.130	61,0	52,0
IN	5.445	5.290	62,7	60,1	RI	53	50	102,0	99,0
IA	2.151	2.247	56,5	53,0	SC	1.422	1.420	70,1	65,9
KS	404	389	54,5	47,8	SD	435	602	48,0	45,8
KY	412	483	67,7	73,5	TN	277	279	71,0	80,7
LA	273	254	115,0	115,0	TX	3.317	3.356	76,7	72,6
ME	1.069	1.070	101,0	97,0	UT	456	486	64,0	59,0
MD	885	898	76,6	75,4	VT	31	30	106,0	102,0
MA	235	237	105,0	102,0	VA	943	988	86,3	81,2
MI	1.406	1.396	58,0	53,8	WA	1.287	1.313	74,1	71,5
MN	2.499	2.697	57,7	54,0	WV	136	174	104,0	109,0
MS	1.434	1.468	87,8	86,7	WI	910	873	60,1	54,0
MO	1.580	1.622	55,4	51,5	WY	1,7	1,7	83,0	83,0

Nota: Y_1 = milhões de ovos produzidos em 1990.

Y_2 = milhões de ovos produzidos em 1991.

X_1 = preço dos ovos em 1990 (centavos de dólar por dúzia).

X_2 = preço dos ovos em 1991 (centavos de dólar por dúzia).

Fonte: *World Almanac*, 1993, p. 119. Os dados são do *Economic Research Service. U.S. Department of Agriculture*.

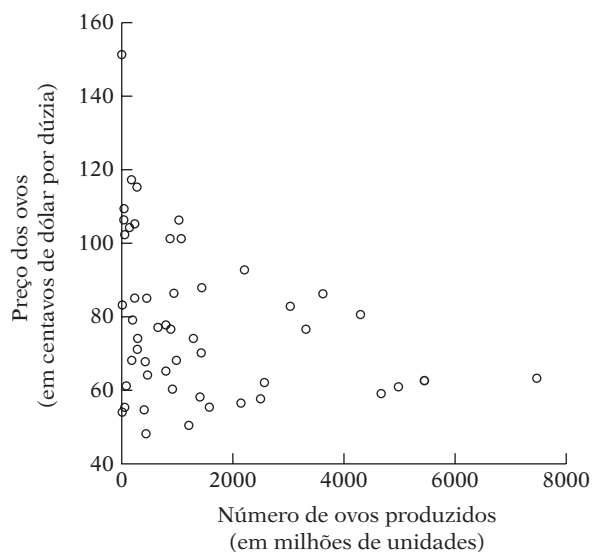
Dados em painel, longitudinais ou de micropainel

São um tipo especial de dados combinados nos quais a *mesma* unidade em corte transversal (por exemplo, uma família ou uma empresa) é pesquisada ao longo do tempo. Por exemplo, o Departamento de Comércio dos Estados Unidos realiza periodicamente um censo habitacional. Em cada levantamento, o mesmo domicílio (ou as pessoas que moram no mesmo endereço) é entrevistado para verificar se houve alguma alteração nas condições da residência e das finanças domiciliares desde o último levantamento. Ao entrevistarmos os mesmos domicílios periodicamente, os dados em painel proporcionam informações muito úteis sobre a dinâmica do seu comportamento, como veremos no Capítulo 16.

Como um exemplo concreto, considere os dados fornecidos na Tabela 1.2. Os dados da tabela, coletados originalmente por Y. Grunfeld, referem-se ao investimento, ao valor da empresa e ao estoque de capital reais de quatro empresas americanas – a saber: General Electric (GE), U.S. Steel (US),

FIGURA 1.6

Relação entre
quantidade e preços
de ovos produzidos,
1990.



General Motors (GM) e Westinghouse (WEST) – para o período de 1935-1954.¹² Como os dados foram coletados para diversas empresas ao longo de um número de anos, este é um exemplo clássico de dados em painel. Nessa tabela, o número de observações para cada empresa é o mesmo, mas nem sempre é esse o caso. Se o número de observações for o mesmo para todas as empresas, teremos um **painel balanceado**; se o número de observações não for o mesmo para todas as empresas, teremos um **painel desbalanceado**. No Capítulo 16, “Modelos de regressão de dados em painel”, examinaremos tais dados e mostraremos como estimar os modelos.

O propósito de Grunfeld ao coletar esses dados era verificar quanto o investimento bruto real (I) depende do valor real da empresa (F) no ano anterior e do estoque de capital real (C) no ano anterior. Como as empresas incluídas no exemplo operam no mesmo mercado de capital, ao estudá-las juntas, Grunfeld queria verificar se tinham as mesmas funções de investimento.

As fontes de dados¹³

Os dados utilizados para as análises empíricas podem ser coletados por órgãos do governo (por exemplo, Departamento de Comércio dos Estados Unidos), organismos internacionais (por exemplo, Fundo Monetário Internacional, FMI, ou o Banco Mundial), por organizações privadas (por exemplo, Standard & Poor’s Corporation) ou por pessoas físicas. Há literalmente milhares dessas instituições coletando dados com diversas finalidades.

A Internet

A Internet causou uma verdadeira revolução na coleta de dados. Basta navegar na rede a partir de uma palavra-chave (como taxa de câmbio) que você será soterrado por todo tipo de fonte de dados. No **Apêndice E**, apresentamos uma seleção de sites frequentemente visitados que proporcionam dados econômicos e financeiros de todos os tipos. A maior parte dos dados pode ser baixada sem grande custo. Você pode incluir alguns desses sites que oferecem dados econômicos úteis em sua lista de favoritos.

Os dados coletados pelas várias instituições podem ser **experimentais** e **não experimentais**. Os experimentais referem-se, em geral, às ciências naturais e o pesquisador costuma coletá-los mantendo

¹² GRUNFELD, Y. The determinants of corporate investment. 1958. Tese (Doutorado) – Departamento de Economia da Universidade de Chicago, 1958. Trabalho não publicado. Esses dados se tornaram referência para ilustrar modelo de regressão de dados em painel.

¹³ Para informações esclarecedoras, veja SOMERS, Albert T. *The U.S. economy demystified: What the major economic statistics mean and their significance for business*. Lexington, Mass.: D. C. Heath, 1985.

TABELA 1.2 Dados de investimentos para quatro empresas, 1935-1954

Observação	I	F_{-1}	C_{-1}	Observação	I	F_{-1}	C_{-1}
GE				US			
1935	33,1	1170,6	97,8	1935	209,9	1362,4	53,8
1936	45,0	2015,8	104,4	1936	355,3	1807,1	50,5
1937	77,2	2803,3	118,0	1937	469,9	2673,3	118,1
1938	44,6	2039,7	156,2	1938	262,3	2039,7	260,2
1939	48,1	2256,2	172,6	1939	230,4	1957,3	312,7
1940	74,4	2132,2	186,6	1940	361,6	2202,9	254,2
1941	113,0	1834,1	220,9	1941	472,8	2380,5	261,4
1942	91,9	1588,0	287,8	1942	445,6	2168,6	298,7
1943	61,3	1749,4	319,9	1943	361,6	1985,1	301,8
1944	56,8	1687,2	321,3	1944	288,2	1813,9	279,1
1945	93,6	2007,7	319,6	1945	258,7	1850,2	213,8
1946	159,9	2208,3	346,0	1946	420,3	2067,7	232,6
1947	147,2	1656,7	456,4	1947	420,5	1796,7	264,8
1948	146,3	1604,4	543,4	1948	494,5	1625,8	306,9
1949	98,3	1431,8	618,3	1949	405,1	1667,0	351,1
1950	93,5	1610,5	647,4	1950	418,8	1677,4	357,8
1951	135,2	1819,4	671,3	1951	588,2	2289,5	341,1
1952	157,3	2079,7	726,1	1952	645,2	2159,4	444,2
1953	179,5	2371,6	800,3	1953	641,0	2031,3	623,6
1954	189,6	2759,9	888,9	1954	459,3	2115,5	669,7
GM				WEST			
1935	317,6	3078,5	2,8	1935	12,93	191,5	1,8
1936	391,8	4661,7	52,6	1936	25,90	516,0	0,8
1937	410,6	5387,1	156,9	1937	35,05	729,0	7,4
1938	257,7	2792,2	209,2	1938	22,89	560,4	18,1
1939	330,8	4313,2	203,4	1939	18,84	519,9	23,5
1940	461,2	4643,9	207,2	1940	28,57	628,5	26,5
1941	512,0	4551,2	255,2	1941	48,51	537,1	36,2
1942	448,0	3244,1	303,7	1942	43,34	561,2	60,8
1943	499,6	4053,7	264,1	1943	37,02	617,2	84,4
1944	547,5	4379,3	201,6	1944	37,81	626,7	91,2
1945	561,2	4840,9	265,0	1945	39,27	737,2	92,4
1946	688,1	4900,0	402,2	1946	53,46	760,5	86,0
1947	568,9	3526,5	761,5	1947	55,56	581,4	111,1
1948	529,2	3245,7	922,4	1948	49,56	662,3	130,6
1949	555,1	3700,2	1020,1	1949	32,04	583,8	141,8
1950	642,9	3755,6	1099,0	1950	32,24	635,2	136,7
1951	755,9	4833,0	1207,7	1951	54,38	732,8	129,7
1952	891,2	4924,9	1430,5	1952	71,78	864,1	145,5
1953	1304,4	6241,7	1777,3	1953	90,08	1193,5	174,8
1954	1486,7	5593,6	2226,3	1954	68,60	1188,9	213,5

Notas: $Y = I$ = investimento bruto = adições à planta e equipamentos mais manutenção e reparos, em milhões de dólares deflacionados por P_1 .

$X_2 = F$ = valor da empresa = preço das ações ordinárias e preferenciais em 31 de dezembro (ou preço médio em 31 de dezembro e 31 de janeiro do ano seguinte) multiplicado pelo número de ações preferenciais em circulação mais o valor presente total da dívida em 31 de dezembro, em milhões de dólares deflacionados por P_2 .

$X_3 = C$ = estoque na planta e de equipamentos = soma acumulada das adições líquidas à planta e aos equipamentos deflacionadas por P_1 subtraída da provisão de depreciação deflacionada por P_3 nessas definições.

P_1 = deflator de preço implícito dos equipamentos duráveis do produtor (1947 = 100).

P_2 = deflator de preço implícito PIB (1947 = 100).

P_3 = deflator de despesas de depreciação = variação média em 10 anos do índice de preços no atacado dos metais e produtos metálicos (1947 = 100).

Fonte: reproduzido de VINOD, H. D.; ULLAH, Aman. *Recent advances in regression methods*. Nova York: Marcel Dekker, 1981. p. 259–261.

certos fatores constantes para avaliar o impacto de outros aspectos sobre o fenômeno. Por exemplo, para avaliar o impacto da obesidade sobre a pressão sanguínea, o pesquisador coletará dados para manter constantes os hábitos de alimentação, fumo e bebida das pessoas a fim de minimizar a influência dessas variáveis sobre a pressão sanguínea.

Nas ciências sociais, os dados encontrados em geral são de natureza não experimental, isto é, não são controlados pelo pesquisador.¹⁴ Por exemplo, os dados relativos ao PIB, ao desemprego, aos preços das ações etc. não estão sob o controle direto do pesquisador. Como veremos, a falta de controle geralmente cria problemas para o pesquisador que busca a(s) causa(s) exata(s) que afeta(m) uma situação em particular. Por exemplo, é a oferta de moeda que determina o PIB (nominal) ou é o inverso?

A precisão dos dados¹⁵

Embora haja fartura de dados disponíveis para a pesquisa econômica, sua qualidade muitas vezes deixa a desejar. Há várias razões para tanto:

1. Como já foi mencionado, a maioria dos dados das ciências sociais são não experimentais por natureza. Portanto, há a possibilidade de ocorrerem erros de observação, sejam intencionais ou não.
2. Mesmo no caso dos dados experimentais, erros de medição surgem em decorrência de aproximações e arredondamentos.
3. Nos levantamentos feitos por meio de questionários, o problema da falta de respostas pode ser grave; o pesquisador terá sorte se conseguir que 40% dos questionários sejam respondidos. Análises embasadas nessas respostas parciais podem não refletir verdadeiramente o comportamento dos 60% que não responderam, levando assim ao que é conhecido como viés de seleção (da amostra). E há, ainda, o problema de que aqueles que respondem ao questionário não o fazem de forma completa, especialmente perguntas sobre tópicos financeiros delicados, levando a novo viés de seleção.
4. Os métodos de amostragem usados para a obtenção dos dados variam tanto que muitas vezes é difícil comparar os resultados obtidos em diversas amostras.
5. Os dados econômicos em geral são apresentados em um nível muito agregado. Por exemplo, a maior parte dos macrodados (por exemplo, PIB, emprego, inflação, desemprego) só estão disponíveis para a economia como um todo ou para algumas grandes regiões geográficas. Esses dados muito agregados pouco nos dizem sobre as microunidades individuais que podem ser o objetivo final do estudo.
6. Devido à confidencialidade, certos dados só são publicados em forma muito agregada. A Receita Federal, por exemplo, não pode por lei revelar dados sobre declarações individuais; só pode liberar alguns valores totais muito agregados. Quem deseja saber quanto as pessoas com dado nível de renda gastam com saúde, só pode verificar em nível muito agregado, mas essa macroanálise muitas vezes deixa de revelar a dinâmica do comportamento das microunidades. Do mesmo modo, o Departamento de Comércio, que conduz um censo das empresas a cada cinco anos, não pode revelar informações sobre produção, emprego, consumo de energia, gastos com pesquisa e desenvolvimento, etc. em nível de empresa. Portanto, é difícil estudar as diferenças entre empresas em relação a essas variáveis.

Como consequência desses e de muitos outros problemas, **o pesquisador sempre deve ter em mente que os resultados de sua pesquisa terão a mesma qualidade dos dados coletados.** Em certas situações, quando os pesquisadores concluem que os resultados de seu trabalho são "insatisfatórios", a causa pode não estar no uso do modelo errado, mas na má qualidade dos dados. Infelizmente, devido à natureza

¹⁴ Nas ciências sociais às vezes também é possível fazer experimentos controlados. No Exercício 1.6, daremos um exemplo.

¹⁵ Para uma revisão crítica, veja MORGENSTERN, O. *The accuracy of economic observations*. 2. ed. Princeton, N. J.: Princeton University Press, 1963.

não experimental dos dados usados na maior parte dos estudos das ciências sociais, os pesquisadores não têm outra opção senão depender daqueles disponíveis. Mas devem ter sempre em mente que os dados empregados podem não ser os melhores e devem procurar não ser muito dogmáticos quanto aos resultados obtidos em determinado estudo, especialmente quando a qualidade dos dados não é confiável.

Uma nota sobre as escalas de medição das variáveis¹⁶

As variáveis que geralmente encontramos enquadram-se em quatro categorias amplas: *escalas de razão, de intervalo, ordinal e nominal*. É importante entender o que cada uma delas representa.

Escala de razão

Dada uma variável X , que assume dois valores, X_1 e X_2 , a razão X_1/X_2 e a distância $(X_2 - X_1)$ são quantidades significativas. Também há um ordenamento natural (ascendente ou descendente) dos valores ao longo da escala. Portanto, comparações como $X_2 \leq X_1$ ou $X_2 \geq X_1$ fazem sentido. A maioria das variáveis econômicas pertence a esta categoria. Faz sentido perguntar qual a diferença quantitativa entre o PIB deste ano e o do ano anterior. A renda pessoal, medida em dólares, é uma escala de razão. Alguém com um salário de \$ 100 mil ganha duas vezes mais que alguém com um salário de \$ 50 mil (sem o desconto dos impostos, é claro).

Escala de intervalo

A escala de intervalo atende às duas últimas propriedades da escala de razão, mas não à primeira. A distância entre dois períodos de tempo, por exemplo (2000-1995), é significativa, mas não a razão entre eles (2000/1995). Às 11 horas (horário local) de 11 de agosto de 2007, a cidade de Portland, em Oregon, registrava uma temperatura de 60°F enquanto que Tallahassee, na Flórida, chegou a 90°F. A temperatura não é medida em uma escala de razão, pois não faz sentido afirmar que Tallahassee estava 50% mais quente que Portland. Isso se deve principalmente ao fato de que a escala Fahrenheit não usa 0 grau como uma base natural.

Escala ordinal

Uma variável se enquadra nesta categoria apenas se satisfaz à terceira propriedade da escala de razão (isto é, o ordenamento natural). Como exemplos, podemos citar os sistemas de avaliação de alunos (conceitos A, B, C) ou as classes de renda (alta, média, baixa). No caso dessas variáveis, há um ordenamento, mas a distância entre as categorias não pode ser quantificada. Os estudantes de economia irão recordar-se das *curvas de indiferença* entre dois bens, em que cada curva mais elevada indica um nível mais alto de utilidade, mas não se pode quantificar quanto uma curva de indiferença é superior a outras.

Escala nominal

As variáveis desta categoria não têm nenhuma das características das variáveis da escala de razão. Variáveis como gênero (feminino, masculino) e estado civil (solteiro, casado, divorciado, separado) apenas denotam categorias. Pergunta: por que essas variáveis não podem ser expressas em nenhuma das escalas anteriores?

Como veremos, as técnicas econométricas que podem ser adequadas a variáveis com escala de razão podem não ser para as variáveis com escala nominal. É importante ter em mente os quatro tipos de escalas de medida descritas acima.

Resumo e conclusões

1. A ideia principal por trás da análise de regressão é a dependência estatística de uma variável, a dependente, a uma ou mais variáveis, as explanatórias.
2. O objetivo dessa análise é estimar e/ou prever o valor médio da variável dependente com base no valor conhecido ou fixado das variáveis explanatórias.

¹⁶ As considerações a seguir baseiam-se em SPANOS, Aris. *Probability theory and statistical inference: econometric modeling with observational data*. Nova York: Cambridge University Press, 1999. p. 24.

3. Na prática, o sucesso da análise de regressão depende da disponibilidade de dados adequados. Este capítulo discutiu a natureza, as fontes e as limitações dos dados que em geral estão disponíveis para as pesquisas, especialmente no campo das ciências sociais.
4. Em qualquer pesquisa, o pesquisador deve explicitar claramente as fontes dos dados usados na análise, suas definições, seus métodos de coleta e quaisquer lacunas ou omissões nos dados, bem como quaisquer revisões realizadas. Não devemos esquecer que os dados macroeconômicos publicados pelo governo são revistos frequentemente.
5. Como o leitor pode não ter tempo, energia ou recursos para confirmar os dados, ele tem o direito de acreditar que os dados usados pelo pesquisador foram coletados de forma adequada e que os cálculos e análises estão corretos.

EXERCÍCIOS

- 1.1. A Tabela 1.3 apresenta dados relativos ao Índice de Preços ao Consumidor (IPC) de sete países industrializados. A base do índice é 1982–1984 = 100.

TABELA 1.3

IPC em sete países industrializados, 1980 – 2005 (1982 – 1984 = 100)

Fonte: *Economic Report of the President*, 2007, Tabela 108, p. 354.

Ano	EUA	Canadá	Japão	França	Alemanha	Itália	Reino Unido
1980	82,4	76,1	91,0	72,2	86,7	63,9	78,5
1981	90,9	85,6	95,3	81,8	92,2	75,5	87,9
1982	96,5	94,9	98,1	91,7	97,0	87,8	95,4
1983	99,6	100,4	99,8	100,3	100,3	100,8	99,8
1984	103,9	104,7	102,1	108,0	102,7	111,4	104,8
1985	107,6	109,0	104,2	114,3	104,8	121,7	111,1
1986	109,6	113,5	104,9	117,2	104,6	128,9	114,9
1987	113,6	118,4	104,9	121,1	104,9	135,1	119,7
1988	118,3	123,2	105,6	124,3	106,3	141,9	125,6
1989	124,0	129,3	108,0	128,7	109,2	150,7	135,4
1990	130,7	135,5	111,4	132,9	112,2	160,4	148,2
1991	136,2	143,1	115,0	137,2	116,3	170,5	156,9
1992	140,3	145,3	117,0	140,4	122,2	179,5	162,7
1993	144,5	147,9	118,5	143,4	127,6	187,7	165,3
1994	148,2	148,2	119,3	145,8	131,1	195,3	169,3
1995	152,4	151,4	119,2	148,4	133,3	205,6	175,2
1996	156,9	153,8	119,3	151,4	135,3	213,8	179,4
1997	160,5	156,3	121,5	153,2	137,8	218,2	185,1
1998	163,0	157,8	122,2	154,2	139,1	222,5	191,4
1999	166,6	160,5	121,8	155,0	140,0	226,2	194,3
2000	172,2	164,9	121,0	157,6	142,0	231,9	200,1
2001	177,1	169,1	120,1	160,2	144,8	238,3	203,6
2002	179,9	172,9	119,0	163,3	146,7	244,3	207,0
2003	184,0	177,7	118,7	166,7	148,3	250,8	213,0
2004	188,9	181,0	118,7	170,3	150,8	256,3	219,4
2005	195,3	184,9	118,3	173,2	153,7	261,3	225,6

- a. Com base nos dados fornecidos, calcule a taxa de inflação de cada país.¹⁷
 - b. Represente graficamente a taxa de inflação de cada país em relação ao tempo (isto é, use o eixo horizontal para o tempo e o eixo vertical para a taxa de inflação).
 - c. Que conclusões gerais é possível tirar sobre a evolução da inflação nos sete países?
 - d. Em que país a taxa de inflação parece ser a mais flutuante? Há alguma explicação para isso?
- 1.2. a. Usando a Tabela 1.3, represente as taxas de inflação do Canadá, França, Alemanha, Itália, Japão e Reino Unido em relação à taxa de inflação dos Estados Unidos.
 - b. Faça um comentário geral sobre o comportamento das taxas de inflação dos seis países em relação à inflação dos Estados Unidos.
 - c. Se você constatar que as taxas de inflação dos seis países evoluem no mesmo sentido que a dos Estados Unidos, isso sugere que a inflação dos Estados Unidos "causa" inflação nos outros países? Justifique.
- 1.3. A Tabela 1.4 apresenta as taxas de câmbio em sete países industrializados, no período 1985–2006. Exceto no caso do Reino Unido, as taxas de câmbio estão definidas como unidades de moeda estrangeira por um dólar; no caso do Reino Unido, a taxa de câmbio é dada como o número de dólares por uma libra esterlina.

TABELA 1.4 Taxa de câmbio de nove países: 1985–2006

Ano	Austrália	Canadá	China	Japão	México	Coreia do Sul	Suécia	Suíça	Reino Unido
1985	0,7003	1,3659	2,9434	238,47	0,257	872,45	8,6032	2,4552	1,2974
1986	0,6709	1,3896	3,4616	168,35	0,612	884,60	7,1273	1,7979	1,4677
1987	0,7014	1,3259	3,7314	144,60	1,378	826,16	6,3469	1,4918	1,6398
1988	0,7841	1,2306	3,7314	128,17	2,273	734,52	6,1370	1,4643	1,7813
1989	0,7919	1,1842	3,7673	138,07	2,461	674,13	6,4559	1,6369	1,6382
1990	0,7807	1,1668	4,7921	145,00	2,813	710,64	5,9231	1,3901	1,7841
1991	0,7787	1,1460	5,3337	134,59	3,018	736,73	6,0521	1,4356	1,7674
1992	0,7352	1,2085	5,5206	126,78	3,095	784,66	5,8258	1,4064	1,7663
1993	0,6799	1,2902	5,7795	111,08	3,116	805,75	7,7956	1,4781	1,5016
1994	0,7316	1,3664	8,6397	102,18	3,385	806,93	7,7161	1,3667	1,5319
1995	0,7407	1,3725	8,3700	93,96	6,447	772,69	7,1406	1,1812	1,5785
1996	0,7828	1,3638	8,3389	108,78	7,600	805,00	6,7082	1,2361	1,5607
1997	0,7437	1,3849	8,3193	121,06	7,918	953,19	7,6446	1,4514	1,6376
1998	0,6291	1,4836	8,3008	130,99	9,152	1.400,40	7,9522	1,4506	1,6573
1999	0,6454	1,4858	8,2783	113,73	9,553	1.189,84	8,2740	1,5045	1,6172
2000	0,5815	1,4855	8,2784	107,80	9,459	1.130,90	9,1735	1,6904	1,5156
2001	0,5169	1,5487	8,2770	121,57	9,337	1.292,02	10,3425	1,6891	1,4396
2002	0,5437	1,5704	8,2771	125,22	9,663	1.250,31	9,7233	1,5567	1,5025
2003	0,6524	1,4008	8,2772	115,94	10,793	1.192,08	8,0787	1,3450	1,6347
2004	0,7365	1,3017	8,2768	108,15	11,290	1.145,24	7,3480	1,2428	1,8330
2005	0,7627	1,2115	8,1936	110,11	10,894	1.023,75	7,4710	1,2459	1,8204
2006	0,7535	1,1340	7,9723	116,31	10,906	954,32	7,3718	1,2532	1,8434

Fonte: *Economic Report of The President*, 2007, Table B-110, p. 356.

¹⁷ Subtraia do IPC do ano corrente o IPC do ano anterior, divida a diferença pelo IPC do ano anterior e multiplique o resultado por 100. Assim, a taxa de inflação do Canadá em 1981 foi de $[(85,6 - 76,1)/76,1] \times 100 = 12,48\%$, aproximadamente.

- a. Represente graficamente a evolução das taxas de câmbio ao longo do tempo e comente sobre o comportamento geral dessa evolução.
 - b. Diz-se que o dólar *apreciou-se* quando pode comprar mais unidades de moeda estrangeira. Opostamente, diz-se que se *depreciou* quando compra menos unidades da moeda estrangeira. No período 1985–2006, qual foi o comportamento geral do dólar dos Estados Unidos? Aproveite para pesquisar em algum livro de macroeconomia ou de economia internacional os fatores que determinam a apreciação ou depreciação de uma moeda.
- 1.4. A Tabela 1.5 apresenta os dados relativos à oferta monetária, no conceito de M1, que aparecem na Figura 1.5. Você poderia apresentar razões para o aumento da oferta de moeda no período considerado?

TABELA 1.5

Oferta monetária ajustada, no conceito de M1: janeiro de 1959 a julho de 1999 (em bilhões de dólares)

Fonte: Board of Governors, Federal Reserve Bank, Estados Unidos.

1959:01	138,8900	139,3900	139,7400	139,6900	140,6800	141,1700
1959:07	141,7000	141,9000	141,0100	140,4700	140,3800	139,9500
1960:01	139,9800	139,8700	139,7500	139,5600	139,6100	139,5800
1960:07	140,1800	141,3100	141,1800	140,9200	140,8600	140,6900
1961:01	141,0600	141,6000	141,8700	142,1300	142,6600	142,8800
1961:07	142,9200	143,4900	143,7800	144,1400	144,7600	145,2000
1962:01	145,2400	145,6600	145,9600	146,4000	146,8400	146,5800
1962:07	146,4600	146,5700	146,3000	146,7100	147,2900	147,8200
1963:01	148,2600	148,9000	149,1700	149,7000	150,3900	150,4300
1963:07	151,3400	151,7800	151,9800	152,5500	153,6500	153,2900
1964:01	153,7400	154,3100	154,4800	154,7700	155,3300	155,6200
1964:07	156,8000	157,8200	158,7500	159,2400	159,9600	160,3000
1965:01	160,7100	160,9400	161,4700	162,0300	161,7000	162,1900
1965:07	163,0500	163,6800	164,8500	165,9700	166,7100	167,8500
1966:01	169,0800	169,6200	170,5100	171,8100	171,3300	171,5700
1966:07	170,3100	170,8100	171,9700	171,1600	171,3800	172,0300
1967:01	171,8600	172,9900	174,8100	174,1700	175,6800	177,0200
1967:07	178,1300	179,7100	180,6800	181,6400	182,3800	183,2600
1968:01	184,3300	184,7100	185,4700	186,6000	187,9900	189,4200
1968:07	190,4900	191,8400	192,7400	194,0200	196,0200	197,4100
1969:01	198,6900	199,3500	200,0200	200,7100	200,8100	201,2700
1969:07	201,6600	201,7300	202,1000	202,9000	203,5700	203,8800
1970:01	206,2200	205,0000	205,7500	206,7200	207,2200	207,5400
1970:07	207,9800	209,9300	211,8000	212,8800	213,6600	214,4100
1971:01	215,5400	217,4200	218,7700	220,0000	222,0200	223,4500
1971:07	224,8500	225,5800	226,4700	227,1600	227,7600	228,3200
1972:01	230,0900	232,3200	234,3000	235,5800	235,8900	236,6200
1972:07	238,7900	240,9300	243,1800	245,0200	246,4100	249,2500
1973:01	251,4700	252,1500	251,6700	252,7400	254,8900	256,6900
1973:07	257,5400	257,7600	257,8600	259,0400	260,9800	262,8800
1974:01	263,7600	265,3100	266,6800	267,2000	267,5600	268,4400
1974:07	269,2700	270,1200	271,0500	272,3500	273,7100	274,2000
1975:01	273,9000	275,0000	276,4200	276,1700	279,2000	282,4300
1975:07	283,6800	284,1500	285,6900	285,3900	286,8300	287,0700
1976:01	288,4200	290,7600	292,7000	294,6600	295,9300	296,1600
1976:07	297,2000	299,0500	299,6700	302,0400	303,5900	306,2500

(Continua)

TABELA 1.5
(Continuação)

1977:01	308,2600	311,5400	313,9400	316,0200	317,1900	318,7100
1977:07	320,1900	322,2700	324,4800	326,4000	328,6400	330,8700
1978:01	334,4000	335,3000	336,9600	339,9200	344,8600	346,8000
1978:07	347,6300	349,6600	352,2600	353,3500	355,4100	357,2800
1979:01	358,6000	359,9100	362,4500	368,0500	369,5900	373,3400
1959:01	138,8900	139,3900	139,7400	139,6900	140,6800	141,1700
1959:07	141,7000	141,9000	141,0100	140,4700	140,3800	139,9500
1960:01	139,9800	139,8700	139,7500	139,5600	139,6100	139,5800
1960:07	140,1800	141,3100	141,1800	140,9200	140,8600	140,6900
1961:01	141,0600	141,6000	141,8700	142,1300	142,6600	142,8800
1961:07	142,9200	143,4900	143,7800	144,1400	144,7600	145,2000
1962:01	145,2400	145,6600	145,9600	146,4000	146,8400	146,5800
1962:07	146,4600	146,5700	146,3000	146,7100	147,2900	147,8200
1963:01	148,2600	148,9000	149,1700	149,7000	150,3900	150,4300
1963:07	151,3400	151,7800	151,9800	152,5500	153,6500	153,2900
1964:01	153,7400	154,3100	154,4800	154,7700	155,3300	155,6200
1964:07	156,8000	157,8200	158,7500	159,2400	159,9600	160,3000
1965:01	160,7100	160,9400	161,4700	162,0300	161,7000	162,1900
1965:07	163,0500	163,6800	164,8500	165,9700	166,7100	167,8500
1966:01	169,0800	169,6200	170,5100	171,8100	171,3300	171,5700
1966:07	170,3100	170,8100	171,9700	171,1600	171,3800	172,0300
1967:01	171,8600	172,9900	174,8100	174,1700	175,6800	177,0200
1967:07	178,1300	179,7100	180,6800	181,6400	182,3800	183,2600
1968:01	184,3300	184,7100	185,4700	186,6000	187,9900	189,4200
1968:07	190,4900	191,8400	192,7400	194,0200	196,0200	197,4100
1969:01	198,6900	199,3500	200,0200	200,7100	200,8100	201,2700
1969:07	201,6600	201,7300	202,1000	202,9000	203,5700	203,8800
1970:01	206,2200	205,0000	205,7500	206,7200	207,2200	207,5400
1970:07	207,9800	209,9300	211,8000	212,8800	213,6600	214,4100
1971:01	215,5400	217,4200	218,7700	220,0000	222,0200	223,4500
1971:07	224,8500	225,5800	226,4700	227,1600	227,7600	228,3200
1972:01	230,0900	232,3200	234,3000	235,5800	235,8900	236,6200
1972:07	238,7900	240,9300	243,1800	245,0200	246,4100	249,2500
1973:01	251,4700	252,1500	251,6700	252,7400	254,8900	256,6900
1973:07	257,5400	257,7600	257,8600	259,0400	260,9800	262,8800
1974:01	263,7600	265,3100	266,6800	267,2000	267,5600	268,4400
1974:07	269,2700	270,1200	271,0500	272,3500	273,7100	274,2000
1975:01	273,9000	275,0000	276,4200	276,1700	279,2000	282,4300
1975:07	283,6800	284,1500	285,6900	285,3900	286,8300	287,0700
1976:01	288,4200	290,7600	292,7000	294,6600	295,9300	296,1600
1976:07	297,2000	299,0500	299,6700	302,0400	303,5900	306,2500

(Continua)

TABELA 1.5
(Continuação)

1977:01	308,2600	311,5400	313,9400	316,0200	317,1900	318,7100
1977:07	320,1900	322,2700	324,4800	326,4000	328,6400	330,8700
1978:01	334,4000	335,3000	336,9600	339,9200	344,8600	346,8000
1978:07	347,6300	349,6600	352,2600	353,3500	355,4100	357,2800
1979:01	358,6000	359,9100	362,4500	368,0500	369,5900	373,3400
1979:07	377,2100	378,8200	379,2800	380,8700	380,8100	381,7700
1980:01	385,8500	389,7000	388,1300	383,4400	384,6000	389,4600
1980:07	394,9100	400,0600	405,3600	409,0600	410,3700	408,0600
1981:01	410,8300	414,3800	418,6900	427,0600	424,4300	425,5000
1981:07	427,9000	427,8500	427,4600	428,4500	430,8800	436,1700
1982:01	442,1300	441,4900	442,3700	446,7800	446,5300	447,8900
1982:07	449,0900	452,4900	457,5000	464,5700	471,1200	474,3000
1983:01	476,6800	483,8500	490,1800	492,7700	499,7800	504,3500
1983:07	508,9600	511,6000	513,4100	517,2100	518,5300	520,7900
1984:01	524,4000	526,9900	530,7800	534,0300	536,5900	540,5400
1984:07	542,1300	542,3900	543,8600	543,8700	547,3200	551,1900
1985:01	555,6600	562,4800	565,7400	569,5500	575,0700	583,1700
1985:07	590,8200	598,0600	604,4700	607,9100	611,8300	619,3600
1986:01	620,4000	624,1400	632,8100	640,3500	652,0100	661,5200
1986:07	672,2000	680,7700	688,5100	695,2600	705,2400	724,2800
1987:01	729,3400	729,8400	733,0100	743,3900	746,0000	743,7200
1987:07	744,9600	746,9600	748,6600	756,5000	752,8300	749,6800
1988:01	755,5500	757,0700	761,1800	767,5700	771,6800	779,1000
1988:07	783,4000	785,0800	784,8200	783,6300	784,4600	786,2600
1989:01	784,9200	783,4000	782,7400	778,8200	774,7900	774,2200
1989:07	779,7100	781,1400	782,2000	787,0500	787,9500	792,5700
1990:01	794,9300	797,6500	801,2500	806,2400	804,3600	810,3300
1990:07	811,8000	817,8500	821,8300	820,3000	822,0600	824,5600
1991:01	826,7300	832,4000	838,6200	842,7300	848,9600	858,3300
1991:07	862,9500	868,6500	871,5600	878,4000	887,9500	896,7000
1992:01	910,4900	925,1300	936,0000	943,8900	950,7800	954,7100
1992:07	964,6000	975,7100	988,8400	1004,340	1016,040	1024,450
1993:01	1030,900	1033,150	1037,990	1047,470	1066,220	1075,610
1993:07	1085,880	1095,560	1105,430	1113,800	1123,900	1129,310
1994:01	1132,200	1136,130	1139,910	1141,420	1142,850	1145,650
1994:07	1151,490	1151,390	1152,440	1150,410	1150,440	1149,750
1995:01	1150,640	1146,740	1146,520	1149,480	1144,650	1144,240
1995:07	1146,500	1146,100	1142,270	1136,430	1133,550	1126,730
1996:01	1122,580	1117,530	1122,590	1124,520	1116,300	1115,470
1996:07	1112,340	1102,180	1095,610	1082,560	1080,490	1081,340
1997:01	1080,520	1076,200	1072,420	1067,450	1063,370	1065,990

(Continua)

TABELA 1.5
(Continuação)

1997:07	1067,570	1072,080	1064,820	1062,060	1067,530	1074,870
1998:01	1073,810	1076,020	1080,650	1082,090	1078,170	1077,780
1998:07	1075,370	1072,210	1074,650	1080,400	1088,960	1093,350
1999:01	1091,000	1092,650	1102,010	1108,400	1104,750	1101,110
1999:07	1099,530	1102,400	1093,460			

- 1.5. Suponha que você quisesse desenvolver um modelo econômico de atividades criminosas, como as horas gastas nessas atividades (por exemplo, a venda de drogas ilegais). Que variáveis consideraria? Verifique se seu modelo combina com o desenvolvido pelo economista ganhador do Nobel Gary Becker.¹⁸
- 1.6. *Experimentos controlados de economia*: em 7 de abril de 2000, o presidente Clinton sancionou uma lei aprovada pelo Congresso que eliminava as restrições aos ganhos dos beneficiários da Previdência Social. Até então, os beneficiários com idade entre 65 e 69 anos que ganhassem mais de \$ 17 mil ao ano perderiam o equivalente a \$ 1 do benefício para cada \$ 3 ganhos além daqueles \$ 17 mil. Como você conceberia um estudo visando avaliar o impacto dessa mudança legal? *Nota*: na lei antiga, não havia nenhuma limitação de renda para os beneficiários com mais de 70 anos.
- 1.7. Os dados apresentados na Tabela 1.6 foram divulgados na edição do *The Wall Street Journal* de 1º de março de 1984. Relacionam o orçamento de publicidade (em milhões de dólares) de

TABELA 1.6

Impacto das
despesas com
publicidade

Fonte: Disponível em:
<http://lib.start.cmu.edu/DASL/Datafiles/tvadsdat.html>.

Empresa	Impressão (em milhões)	Despesas (em milhões de dólares de 1983)
1. Miller Lite	32,1	50,1
2. Pepsi	99,6	74,1
3. Stroh's	11,7	19,3
4. Fed'l Express	21,9	22,9
5. Burger King	60,8	82,4
6. Coca Cola	78,6	40,1
7. McDonald's	92,4	185,9
8. MCI	50,7	26,9
9. Diet Cola	21,4	20,4
10. Ford	40,1	166,2
11. Levi's	40,8	27,0
12. Bud Lite	10,4	45,6
13. ATT/Bell	88,9	154,9
14. Calvin Klein	12,0	5,0
15. Wendy's	29,2	49,7
16. Polaroid	38,0	26,9
17. Shasta	10,0	5,7
18. Meow Mix	12,3	7,6
19. Oscar Meyer	23,4	9,2
20. Crest	71,1	32,4
21. Kibbles 'N Bits	4,4	6,1

¹⁸ BECKER, G. S. "Crime and punishment: an economic approach". *Journal of Political Economy*, 1968. v. 76, p. 169-217.

21 empresas em 1983 com as impressões retidas, semanalmente, pelos que viram os produtos anunciados por essas empresas. Os dados foram obtidos em uma pesquisa realizada com 4 mil adultos, em que foi pedido aos usuários dos produtos que citassem um comercial da categoria do produto que tivessem assistido na semana anterior.

- a. Trace um gráfico com as impressões no eixo vertical e os gastos com publicidade no eixo horizontal.
- b. O que você poderia dizer sobre a natureza da relação entre as duas variáveis?
- c. Examinando o gráfico, você acha que vale a pena anunciar? Pense em todos os comerciais veiculados em finais de campeonatos de esportes ou no horário nobre.

Nota: nos próximos capítulos exploraremos mais os dados da Tabela 1.6.