

Termos-chave

Amostragem aleatória	Corte transversal agrupado	Frequência de dados
Análise empírica	Dados em painel	Modelo econométrico
<i>Ceteris paribus</i>	Dados experimentais	Modelo econômico
Conjunto de dados de corte transversal	Dados não experimentais	
	Dados observacionais	
Conjunto de dados de séries temporais	Dados retrospectivos	
	Efeito causal	

Problemas

- Suponha que você seja solicitado a conduzir um estudo para determinar se salas de aula com menor número de alunos conduzem a um melhor desempenho do estudante de quarto ano.
 - Se você pudesse conduzir qualquer experimento que quisesse, o que você faria? Seja específico.
 - De modo mais realista, suponha que você possa reunir dados observacionais de milhares de estudantes de quarto ano de determinado Estado. Você pode obter o tamanho, em número de alunos, de suas salas de aula e um total de pontos do teste padronizado feito no final do quarto ano. Por que se pode esperar uma correlação negativa entre o tamanho da sala de aula e os pontos do teste padronizado?
 - Uma correlação negativa necessariamente mostrará que salas de aulas com menor número de alunos conduzem a um melhor desempenho. Explique.
- Uma das justificativas dos programas de treinamento de pessoal é que eles melhoram a produtividade dos trabalhadores. Suponha que você seja questionado se mais treinamentos tornam os profissionais mais produtivos. No entanto, em vez de ter dados de trabalhadores individuais, você tem acesso aos dados de empresas industriais de Ohio. Particularmente, de cada empresa, você tem informação sobre o número de horas de treinamento por trabalhador (*treinamento*) e o número de itens sem defeito produzido por trabalhador, por hora (*produção*).
 - Especifique claramente o conceito *ceteris paribus* do experimento inevidente nessa questão de política.
 - Parece provável que a decisão de uma empresa em treinar seus empregados será independente das características dos trabalhadores? Quais são algumas das características, mensuráveis e não mensuráveis, dos trabalhadores?
 - Cite um fator, exceto as características dos trabalhadores, que pode afetar a produtividade do trabalhador.
 - Se você encontrar uma correlação positiva entre *treinamento* e *produção*, você terá estabelecido de forma convincente que o treinamento pessoal torna os trabalhadores mais produtivos? Explique.
- Suponha que em sua universidade lhe seja pedido que encontre a relação entre horas semanais gastas estudando (*estudo*) e horas semanais gastas trabalhando (*trabalho*). Faz sentido caracterizar o problema como dedução, se o *estudo* “induz” ao *trabalho* ou o *trabalho* “induz” ao *estudo*? Explique.
- Estados (e províncias) que têm controle sobre as taxas às vezes reduzem os impostos para tentar estimular o crescimento econômico. Suponha que você tenha sido contratado por um estado para estimar os efeitos dos impostos corporativos sobre, digamos, o crescimento *per capita* do produto estadual bruto.

- (i) Que tipo de dados você precisaria coletar para realizar uma análise estatística?
- (ii) É viável fazer um experimento controlado? O que seria necessário?
- (iii) É possível que uma análise de correlação entre o crescimento do produto estadual bruto e os impostos seja convincente? Explique.

Exercícios em computador

- C1** Use os dados do arquivo WAGE1 para este exercício.
- (i) Encontre o nível de escolaridade médio da amostra. Quais são os menores e os maiores valores de anos de educação?
 - (ii) Encontre o salário-hora médio da amostra. Ele parece alto ou baixo?
 - (iii) Os dados salariais são reportados em dólares de 1976. Pesquisando na Internet ou em uma fonte impressa, encontre o Índice de Preços do Consumidor (IPC) para os anos de 1976 e 2013.
 - (iv) Use os valores do IPC da parte (iii) para encontrar o salário-hora médio em dólares de 2013. Agora o salário-hora médio parece razoável?
 - (v) Quantas mulheres existem na amostra? E quantos homens?
- C2** Use os dados do arquivo BWGHT para responder a essas questões.
- (i) Quantas mulheres existem na amostra e quantas relataram fumar durante a gravidez?
 - (ii) Qual é o número médio de cigarros consumidos por dia? A média é uma boa medida da mulher “típica” neste caso? Explique.
 - (iii) Entre mulheres que fumaram durante a gravidez, qual é o número médio de cigarros consumidos por dia? De que forma isso se compara com sua resposta ao item (ii) e por quê?
 - (iv) Encontre a média de *fatheduc* na amostra. Por que somente 1.192 observações são usadas para calcular essa média?
 - (v) Relate a renda média familiar e seu desvio padrão em dólares.
- C3** Os dados existentes no arquivo MEAP01 são do estado de Michigan no ano de 2001. Use estes dados para responder às seguintes questões.
- (i) Encontre os maiores e os menores valores de *math4*. Essa variação faz sentido? Explique.
 - (ii) Quantas escolas têm uma taxa de aprovação perfeita no teste de matemática? Que porcentagem da amostra total isso representa?
 - (iii) Quantas escolas têm taxas de aprovação em matemática de exatamente 50%?
 - (iv) Compare as taxas médias de aprovação em matemática e leitura. Qual teste tem a aprovação mais difícil?
 - (v) Encontre a correlação entre *math4* e *read4*. O que você conclui?
 - (vi) A variável *exppp* são os gastos por aluno. Encontre o *exppp* médio e seu desvio padrão. Você diria que há uma variação ampla nos gastos por aluno?
 - (vii) Suponha que a Escola A gaste US\$ 6.000 por estudante e a Escola B gaste US\$ 5.500 por aluno. Com que percentual os gastos da Escola A superam os da Escola B? Compare isso a $100 \cdot [\log(6.000) - \log(5.500)]$, que é a diferença percentual aproximada baseada na diferença dos logs naturais. Ver Seção A.4, no Apêndice A (Disponível no site da Cengage.)
- C4** Os dados contidos em JTRAIN2 são provenientes de um experimento de capacitação profissional direcionado para homens de baixa renda durante 1976-1977; ver Lalonde (1986).

- (i) Use a variável indicadora *train* para determinar a proporção de homens que recebeu treinamento profissional.
 - (ii) A variável *re78* são os ganhos de 1978, medidos em milhares de dólares de 1982. Encontre as médias de *re78* para a amostra de homens que recebeu capacitação profissional e para aquela que não recebeu. A diferença é economicamente grande?
 - (iii) A variável *unem78* é um indicador de um homem estar desempregado ou não em 1978. Que proporção dos homens que receberam treinamento profissional está desempregada? E entre aqueles que não receberam treinamento? Comente sobre a diferença.
 - (iv) A partir dos itens (ii) e (iii), o programa de treinamento profissional parece ter sido efetivo? O que tornaria suas conclusões mais convincentes?
- C5** Os dados em FERTIL2 foram coletados de mulheres que viviam na República de Botsuana em 1988. A variável *children* refere-se ao número de filhos vivos. A variável *electric* é um indicador binário igual a um se a residência da mulher tiver eletricidade, e zero se não tiver.
- (i) Encontre os menores e os maiores valores de *children* da amostra. Qual é a média de *children*?
 - (ii) Qual é a porcentagem de mulheres que têm eletricidade em casa?
 - (iii) Calcule a média de *children* para aquelas sem eletricidade e faça o mesmo para as que têm eletricidade. Comente o que descobriu.
 - (iv) A partir do item (iii), você pode deduzir que ter eletricidade “causa” mulheres com menos filhos? Explique.
- C6** Use os dados contidos no arquivo COUNTYMURDERS para responder a essas questões. Use somente o ano de 1996. A variável *murders* é o número de assassinatos relatados no condado. A variável *execs* é o número de execuções de pessoas sentenciadas à morte ocorridas naquele determinado condado. A maioria dos estados norte-americanos tem pena de morte, mas alguns deles não.
- (i) Quantos condados são listados no conjunto de dados? Destes, quantos tiveram zero assassinato? Qual é a porcentagem de condados que teve zero execução? (Lembre-se, use somente os dados de 1996.)
 - (ii) Qual é o maior número de assassinatos? Qual é o maior número de execuções? Por que o número médio de execuções é tão pequeno?
 - (iii) Calcule o coeficiente de correlação entre *murders* e *execs* e descreva o que encontrar.
 - (iv) Você deve ter encontrado uma correlação positiva no item (iii). Você acha que mais execuções *causam* mais assassinatos? O que poderia explicar a correlação positiva?
- C7** O conjunto de dados do arquivo ALCOHOL contém informações sobre uma amostra de homens dos Estados Unidos. Duas variáveis principais são o status de emprego autor-relatado e o abuso de álcool (ao lado de muitas outras variáveis). As variáveis *employ* e *abuse* são ambas binárias, ou indicadores: elas só recebem os valores zero e um.
- (i) Qual é a porcentagem de homens da amostra que relatou abuso de álcool? Qual é a taxa de emprego?
 - (ii) Considere o grupo de homens que abusa de álcool. Qual é a taxa de emprego desse grupo?
 - (iii) Qual é a taxa de emprego do grupo de homens que não abusam de álcool?
 - (iv) Discuta a diferença de suas respostas aos itens (ii) e (iii). Isso permite que você conclua que o abuso de álcool causa desemprego?