Notas

[1] STOUFFER, Samuel A. et al. *The American Soldier*. Princeton, N.J.: Princeton University Press, 1949. v.1. p.122 et seq., esp. p.127.

[2] KENDALL, Patricia L., LAZARSFELD, Paul F. Problems of Social Analysis. In: MERTON, Robert K., LAZARSFELD, Paul F. (Ed.). *Continuities in Social Research:* Studies in the Scope and Method of "The American Soldier". New York: Free Press, 1950. p.133-196.

[3] Note, é claro, que os termos "variável independente" e "variável dependente" são, a rigor, usados incorretamente no diagrama. De fato, temos uma variável independente (a variável de teste) e duas variáveis dependentes. A terminologia incorreta foi usada apenas para dar continuidade ao exemplo anterior.

[4] N.T. Medicare: sistema de seguro-saúde. BABBIE, Earl R. *Science and Morality in Medicine*. Berkeley: University of California Press, 1970. Ver especialmente p.181.

[5] GLOCK, Charles Y., RINGER, Benjamin B., BABBIE, Earl R. *To Comfort and to Challenge*. Berkeley: University of California Press, 1967. p.92.

[6] ROSENBERG, Morris. *The Logic of Survey Analysis*. New York: Basic Books, 1968.

[7] Ibidem. p.88-89.

[8] Ibidem. p.94-95.

Leituras Adicionais

GLOCK, Charles Y. (Ed.). *Survey Research in the Social Sciences*. New York: Russel Sage Foundation, 1967. cap.I.

HIRSCHI, Travis, SELVIN, Hanan. *Principles of Survey Analysis*. New York: Free Press, 1973.

HYMAN, Herbert. *Survey Design and Analysis*. New York: Free Press, 1955.

LAZARSFELD, Paul F., PASANELLA, Ann K., ROSENBERG, Morris (Ed.). *Continuities in "The Language of Social Research"*. New York: Free Press, 1972. seção II.

ROSENBERG, Morris. *The Logic of Survey Analysis*. New York: Basic Books, 1968.

STOUFFER, Samuel A. *Social Research to Test Ideas*. New York: Free Press, 1962.

Capítulo 16

Estatística Social

Muitas pessoas se deixam intimidar pela pesquisa empírica por não ficarem confortáveis com a matemática e a estatística. Muitos relatórios de pesquisa são cheios de cálculos não especificados. O papel da estatística em pesquisa de *survey* é muito importante, mas é igualmente importante ver esse papel na perspectiva adequada. A pesquisa empírica é, antes de mais nada, uma operação lógica, e não uma operação matemática. A matemática é apenas uma linguagem conveniente e eficaz para descrever as operações lógicas inerentes à boa análise de dados. A estatística, um ramo da matemática aplicada, é especialmente adequada para várias análises de pesquisa.

Neste capítulo, examinaremos dois tipos de estatística: a *descritiva* e a *inferencial*. A *estatística descritiva* é um meio de descrever dados em formas manejáveis. A *estatística inferencial*, por seu lado, ajuda a tirar conclusões de observações; tipicamente, envolve tirar conclusões sobre uma população a partir de uma amostra daquela população.

Estatística Descritiva

A estatística descritiva é um método de apresentar descrições quantitativas de modo manejável. Às vezes, deseja-se descrever variáveis isoladamente, outras vezes quer-se descrever as associações que ligam uma variável a outra. Vejamos alguns dos modos de fazer essas descrições.

Redução de Dados

A pesquisa científica freqüentemente envolve coletar grandes massas de dados. Suponha que pesquisamos 2.000 pessoas, fazendo 100 perguntas a cada uma — um estudo de grandeza comum. Com isso, teríamos 200.000 respostas. Ninguém poderia ler todas e extrair delas alguma conclusão significativa. Portanto, muita análise científica implica a *redução* de dados, de detalhes imanejáveis para resumos mais fáceis de trabalhar.

Para começar, vejamos a matriz de dados brutos criados por um projeto de pesquisa quantitativa. A Tabela 16-1 apresenta uma matriz de dados parciais. Cada fileira na matriz representa uma pessoa (ou outra unidade de análise), cada coluna uma variável e cada célula na matriz o atributo ou valor codificado que uma pessoa tem numa variável. A primeira coluna da Tabela 16-1 representa sexo. Suponha que "1" represente homem e "2", mulher. Isso significa que as pessoas 1 e 2 são homens, a pessoa 3 é mulher, e assim por diante.

No caso de idade, o "3" da pessoa 1 pode significar de 30 a 39 anos de idade, e o "4" da pessoa 2, 40-49. Não importa como a idade tenha sido codificada (ver Capítulo 11), os números de código mostrados na coluna 2 da Tabela 16-1 descreveriam cada pessoa lá representada.

TABELA 16-1
Matriz de dados brutos parciais

	Sexo	Idade	Escolaridade	Renda	Profissão	Filiação Política	Orientação Política	Filiação Religiosa	Importância da religião
Pessoa 1	1	3	2	4	1	2	3	0	4
Pessoa 2	1	4	2	4	4	1	1	1	2
Pessoa 3	2	2	5	5	2	2	4	2	3
Pessoa 4	1	5	4	4	3	2	2	2	4
Pessoa 5	2	3	7	8	6	1	1	5	1
Pessoa 6	2	1	3	3	5	3	5	1	1

Note que os dados já foram reduzidos quando criamos uma matriz de dados como a da Tabela 16-1. Se a idade foi codificada como acima, a resposta específica "33 anos de idade" já foi reduzida à categoria 30-39. As pessoas entrevistadas podem ter dado 60 ou 70 idades diferentes, agora reduzidas a seis ou sete categorias.

O Capítulo 14 discutiu algumas formas de resumir dados univariados: medidas de tendência central, como a moda, a mediana e a média, e medidas de dispersão, como amplitude e desvio padrão. Também é possível resumir a associação entre variáveis.

Medidas de Associação

A associação entre duas variáveis quaisquer pode ser representada por uma matriz de dados produzida pelas distribuições de freqüência conjuntas das duas variáveis. A Tabela 16-2 apresenta uma matriz assim. Ela dá toda a informação necessária para determinar a natureza e a extensão da relação entre escolaridade e preconceito.

TABELA 16-2
Dados brutos hipotéticos sobre escolaridade e preconceito

	Nível de Escolaridade				
Preconceito	Nenhum	Primário	Secundário	Superior	Pós-Graduado
Alto	23	34	156	67	16
Médio	11	21	123	102	23
Baixo	6	12	95	164	77

Note, por exemplo, que 23 pessoas (a) não têm nenhuma escolaridade e (b) tiveram um escore alto de preconceito; 77 pessoas (a) tinham curso de pós-graduação e (b) tiveram baixo escore de preconceito. Como a matriz de dados brutos na Tabela 16-1, esta matriz dá mais informação do que o necessário à compreensão. Ao examiná-la com cuidado, vê-se que, à medida que a escolaridade aumenta de "nenhum" para "pós-graduado", há uma tendência geral de diminuição do preconceito, mas não é possível ter nada mais além dessa impressão geral. Várias estatísticas descritivas permitem resumir esta matriz de dados. A seleção da medida adequada depende, inicialmente, da natureza das duas variáveis.

Vejamos algumas alternativas para resumir a associação entre duas variáveis. Se quiser mais informação sobre esses tópicos, consulte um manual de estatística social.[1] As medidas de associação que vamos discutir baseiam-se no modelo de *redução proporcional de erro* (RPE). Para ver como esse modelo

funciona, suponha que eu lhe peça para adivinhar atributos de respondentes em alguma variável, por exemplo, se responderam sim ou não a algum item do questionário.

Primeiro, suponha que você conhece a distribuição geral de respostas na amostra total, digamos, 60% sim e 40% não. Você cometeria o menor número de erros se sempre usasse a resposta *modal* (a mais freqüente): o sim.

Segundo, suponha que você também conhece a relação empírica entre a primeira variável e alguma outra, digamos, sexo. Toda vez que eu perguntar se algum respondente respondeu sim ou não, lhe direi também se é homem ou mulher. Se as duas variáveis são relacionadas, você deve cometer menos erros da segunda vez. É possível, portanto, calcular o RPE sabendo a relação entre as duas variáveis: quanto maior a relação, maior a redução de erro.

O modelo básico de RPE é modificado ligeiramente para levar em conta níveis diferentes de medida: nominal, ordinal ou intervalo. As seções seguintes consideram cada nível de medição e apresentam uma medida de associação adequada a cada uma. As medidas discutidas são apenas três de muitas medidas apropriadas.

Variáveis Nominais. Se as duas variáveis consistem em dados nominais (por exemplo, sexo, filiação religiosa, raça), o *lambda* (l) seria uma medida adequada. O lambda se baseia na capacidade de prever valores de uma das variáveis, isto é, o RPE obtido do conhecimento dos valores de outra variável. Vejamos um exemplo hipotético simples da lógica e método do lambda. A Tabela 16-3 apresenta dados hipotéticos relacionando sexo a situação empregatícia. No geral, vemos que 1.100 pessoas estão empregadas e 900 desempregadas. Se você fosse predizer se as pessoas estão empregadas sabendo apenas a distribuição geral dessa variável, predizer sempre "empregado" levaria a menos erros do que sempre predizer "desempregado". Mas isso resultaria em 900 erros em 2.000 previsões.

TABELA 16-3
Dados hipotéticos relacionando sexo e emprego

	Homens	Mulheres	Total
Empregados	900	200	1.100
Desempregados	100	800	900
Total	1.000	1.000	2.000

Suponha que você teve acesso aos dados da Tabela 16-3 e ouvisse o sexo da pessoa antes de predizer a situação de emprego. A estratégia mudaria. Para cada homem a predição seria "empregado" e para cada mulher seria "desempregada". Neste caso, você faria 300 erros — os 100 homens desempregados e as 200 mulheres empregadas — ou 600 erros menos do que cometeria se não soubesse o sexo da pessoa.

Lambda, então, representa a redução de erros como proporção dos erros que teriam sido cometidos com base na distribuição geral. Neste exemplo, lambda seria igual a 0,67, ou seja, os 600 erros a menos divididos pelos 900 erros totais baseados apenas no emprego. Neste exemplo, lambda mede a associação estatística entre sexo e emprego.

Se sexo e emprego fossem estatisticamente independentes, encontraríamos a mesma distribuição de situação de emprego para homens e mulheres. Neste caso, saber o sexo dos respondentes não afetaria o número de erros cometidos ao se predizer situação de emprego, e o lambda resultante seria zero. Se, por outro lado, todos os homens estivessem empregados e todas as mulheres desempregadas, saber o sexo da pessoa evitaria erros na predição de situação de emprego. Especificamente, você cometeria 900 erros a menos (em 900), de forma que lambda seria 1,0, representando uma associação estatística perfeita.

Lambda é apenas uma de várias medidas de associação para a análise de duas variáveis nominais.

Variáveis Ordinais. Se as variáveis relacionadas são ordinais (por exemplo, classe social, religiosidade, alienação), *gamma* (γ) é uma medida de associação adequada. Como lambda, gamma se baseia na capacidade de adivinhar valores de uma variável sabendo valores de outra. Em vez de valores exatos, todavia, gamma se baseia no arranjo ordinal de valores. Para qualquer *par* de casos, você adivinha que a ordenação de uma variável vai corresponder, positiva ou negativamente, à ordenação da outra. Por exemplo, se você suspeita que religiosidade se relaciona positivamente com conservadorismo político e se a Pessoa A é mais religiosa que a Pessoa B, você adivinha que ela é também mais conservadora que B. Gamma se baseia no número de comparações emparelhadas que se ajustam a este padrão versus as que o contradizem.

TABELA 16-4
Dados hipotéticos relacionando classe social e preconceito

Preconceito	Classe Baixa	Classe Média	Classe Alta
Baixo	200	400	700
Médio	500	900	400
Alto	800	300	100

A Tabela 16-4 apresenta dados hipotéticos relacionando classe social e preconceito. A natureza geral da relação entre essas duas variáveis é que, à medida que aumenta a classe social, diminui o preconceito. Há uma associação negativa entre classe social e preconceito.

Gamma é calculado de duas quantidades: (1) o número de pares que têm a mesma posição de ordem nas duas variáveis e (2) o número de pares com posições de ordem opostas nas duas variáveis. O número de pares com a mesma posição é calculado multiplicando-se a freqüência de cada célula na tabela pela soma de todas as células que aparecem abaixo e à direita dela e depois somando-se todos os produtos. Na Tabela 16-4, o número de pares com a mesma posição seria calculado da seguinte forma: 200 (900 + 300 + 400 + 100) + 500 (300 + 100) + 400 (400 + 100) + 900 (100) ou 340.000 + 200.000 + 200.000 + 90.000 = 830.000.

O número de pares com posições opostas nas duas variáveis é calculado multiplicando-se a freqüência de cada célula na tabela pela soma de todas as células que aparecem abaixo e à esquerda dela e depois somando-se todos os produtos. Na Tabela 16-4, o número de pares com posições opostas seria calculado assim: 700 (500 + 800 + 900 + 300) + 400 (800 + 300) + 400 (500 + 800) + 900 (800) ou 1.750.000 + 440.000 + 520.000 + 720.000 = 3.430.000.

Gamma é calculado a partir dos números de pares com a mesma posição e de pares com posições opostas, da seguinte maneira:

$$\text{Gamma} = \frac{\text{igual - oposto}}{\text{igual + oposto}}$$

Neste exemplo, gamma é igual a (830.000 - 3.430.000) dividido por (830.000 + 3.430.000), ou - 0,61. O sinal negativo indica a associação negativa sugerida pela inspeção inicial da tabela. Classe social e preconceito, neste exemplo hipotético, são associados negativamente. A expressão numérica de gamma indica que 61% a mais dos pares examinados tinham posições opostas do que posições iguais.

Note que, enquanto os valores de lambda variam de 0 a 1, os valores de gamma variam de -1 a +1, representando a *direção* tanto quanto a magnitude da associação. Como as variáveis nominais não têm estrutura ordinal, não faz sentido falar da direção da relação. (Um lambda negativo indicaria que você cometeu mais erros na predição dos valores de uma variável sabendo os valores da segunda do que os que cometeu ignorando a segunda; isso não é logicamente possível.)

Considere o seguinte exemplo do uso de gamma na pesquisa de *survey* contemporânea. Para estudar até que ponto as viúvas santificavam seus maridos falecidos, Helena Znaniecki Lopata aplicou um questionário a uma amostra probabilística de 301 viúvas na área de Chicago.[2] Em parte, o questionário pedia a cada uma caracterizar seu marido falecido em termos da *escala de diferenciação semântica* mostrada na Tabela 16-5. Pediu-se a cada uma descrever o cônjuge falecido fazendo um círculo num número para cada par de características opostas. Note que a série de números conectando cada par de características é uma medida ordinal.

Depois, Lopata queria descobrir até que ponto as várias medidas se relacionavam uma com a outra e escolheu gamma como medida da associação. A Tabela 16-6 mostra como ela apresentou os resultados. O formato mostrado é chamado de *matriz de correlação*. Para cada par de medidas, Lopata calculou o gamma. "Bom" e "Útil", por exemplo, se relacionam um com o outro num gamma igual a 0,79. A matriz é uma forma conveniente de apresentar as intercorrelações entre várias variáveis e é encontrada com freqüência na literatura de pesquisa. Neste caso, vê-se que todas as variáveis se relacionam fortemente umas com as outras, embora para alguns pares a relação seja mais forte do que para outros.

Gamma é apenas uma de várias medidas de associação para variáveis ordinais.

TABELA 16-5
Escala de diferenciação semântica

Característica								
Extremo Positivo								Extremo Negativo
Bom	1	2	3	4	5	6	7	Mau
Útil	1	2	3	4	5	6	7	Inútil
Honesto	1	2	3	4	5	6	7	Desonesto
Superior	1	2	3	4	5	6	7	Inferior
Gentil	1	2	3	4	5	6	7	Cruel
Amigável	1	2	3	4	5	6	7	Hostil
Caloroso	1	2	3	4	5	6	7	Frio

TABELA 16-6
Associações gamma entre os itens de diferenciação semântica da escala de santificação

	Útil	Honesto	Superior	Gentil	Amigável	Caloroso
Bom	0,79	0,88	0,80	0,90	0,79	0,83
Útil	—	0,84	0,71	0,77	0,68	0,72
Honesto		—	0,83	0,89	0,79	0,82
Superior			—	0,78	0,60	0,73
Gentil				—	0,88	0,90
Amigável					—	0,90

FONTE – LOPATA, Helena Znaniecki. Widowhood and Husband Sanctification. *Journal of Marriage and the Family*, p.439-450, maio 1991.

Variáveis de Intervalo ou de Razão. Se estão sendo associadas variáveis de intervalo ou de razão (por exemplo, idade, renda, média de notas escolares), uma medida adequada de associação é a *correlação produto-momento* de Pearson (*r*). A derivação e a computação desta medida de associação são por demais complexos para os propósitos deste livro, portanto faremos apenas alguns comentários gerais.

Como gamma e lambda, *r* se baseia em adivinhar o valor de uma variável conhecendo-se a outra. Para variáveis de intervalo ou de razão contínuas, entretanto, é pouco provável que se possa adivinhar o valor *preciso* da variável. Por outro lado, apenas predizer o arranjo ordinal de valores para as duas variáveis deixaria de tirar vantagem do maior teor de informação transmitido por uma variável de intervalo ou razão. Em certo sentido, *r* reflete o *quão exata* foi a adivinhação do valor de uma variável, com base no conhecimento do valor da outra.

Para entender a lógica de *r*, considere como você pode, hipoteticamente, adivinhar os valores das unidades de análise numa variável qualquer. Com variáveis nominais, vimos que se pode sempre apostar no valor modal. Para dados de razão ou intervalo, você sempre minimiza os erros apostando no valor médio da variável. Embora esta prática produza nenhuma ou poucas estimativas perfeitas, a extensão dos erros será minimizada.

No cálculo de lambda, vimos o número de erros produzidos por apostar sempre no valor modal. No caso de *r*, os erros são medidos em termos da soma do quadrado das diferenças entre o valor real e a média. Esta soma é chamada de *variação total.* Para entender este conceito, devemos expandir o âmbito de nossa investigação. Retornaremos a esta discussão quando examinarmos a *análise de regressão*, no Capítulo 17. Por enquanto, mudaremos nosso foco da estatística descritiva para a estatística inferencial.

Estatística Inferencial

A maioria dos *surveys* examina dados coletados de amostras extraídas de populações maiores. Uma amostra de pessoas pode ser entrevistada num *survey*; uma amostra de registros de divórcio pode ser codificada e analisada; uma amostra de jornais pode ser examinada através de análise de conteúdo. Os pesquisadores raramente estudam amostras apenas para descrevê-las *per se*; na maioria dos casos, o objetivo final é fazer afirmações sobre a população maior da qual a amostra foi extraída. Freqüentemente, portanto, será necessário tratar os achados univariados e multivariados amostrais como a base para *inferências* a respeito de alguma população.

Esta seção examina as medidas estatísticas usadas para fazer tais inferências e suas bases lógicas. Começaremos com

dados univariados e passaremos para os multivariados na discussão de testes de significância estatística.

Inferências Univariadas

As primeiras seções do Capítulo 14 trataram dos métodos de apresentação de dados univariados. Cada medida de resumo pretendia oferecer um método para descrever a amostra estudada. Usaremos essas medidas para fazer afirmativas mais amplas sobre a população.

Se 50% de uma amostra de pessoas dizem que tiveram resfriados no ano anterior, 50% é também nossa melhor estimativa da proporção de resfriados na população total de onde foi retirada a amostra. (Esta estimativa supõe uma amostra aleatória simples, naturalmente.) É improvável, entretanto, que *precisamente* 50% da população tiveram resfriados durante o ano. Se foi seguido um desenho de amostragem rigoroso para seleção aleatória, seremos capazes de estimar a faixa de erro esperada quando aplicarmos o achado amostral à população.

O Capítulo 5, sobre a teoria da amostragem, tratou dos procedimentos para fazer tais estimativas, portanto faremos aqui apenas uma revisão. No caso de uma porcentagem, a quantidade

$$\sqrt{\frac{p \times q}{n}}$$

em que p é uma porcentagem, q é igual a 1 - p e n é o tamanho da amostra, chama-se *erro padrão*. Esta quantidade é muito importante na estimativa do erro amostral. Podemos ter 68% de confiança que o número para a população cairá dentro de mais ou menos um erro padrão do número para a amostra, 95% de confiança que cairá dentro de mais ou menos dois erros padrão ou 99,9% de confiança que cairá dentro de mais ou menos três erros padrão.

Qualquer afirmação sobre erro amostral, portanto, deve ter dois componentes essenciais: o *nível de confiança* (por exemplo, 95%) e o *intervalo de confiança* (por exemplo, ± 2,5%). Se 50% de uma amostra de 1.600 pessoas disseram ter tido resfriados durante o ano, podemos dizer que temos 95% de confiança que o número para a população vai estar entre 47,5% e 52,5%.

Neste exemplo, fomos além de apenas descrever a amostra, para fazer estimativas (inferências) sobre a população maior. Ao fazê-lo, devemos estar atentos às implicações de várias suposições.

Primeiro, a amostra deve ser tirada da população sobre a qual se está inferindo. Uma amostra tirada de uma lista telefônica não é base legítima para inferências estatísticas sobre a população de uma cidade, porque algumas pessoas não têm telefone e outras têm números não listados.

Segundo, a estatística inferencial supõe o uso de amostragem aleatória simples, o que virtualmente nunca é o caso em *surveys* por amostragem. A estatística supõe amostragem com reposição, que quase nunca é feita, mas talvez isso não seja um problema sério. Amostragem sistemática é usada mais que a aleatória; também isso não é problema sério, se feita corretamente. Amostragem estratificada, já que aumenta a representatividade, não é, claramente, um problema. Mas a amostragem por conglomerados pode ser problemática, porque as estimativas de erro amostral podem ser pequenas demais. É patente que amostragem de esquina não permite o uso de estatística inferencial. O cálculo do erro padrão também supõe que 100% da amostra será aproveitada; o problema aumenta de gravidade à medida que diminui esta taxa.

Terceiro, a estatística inferencial trata apenas do erro de amostragem. Não considera "erros não-amostrais". Assim, embora possa ser correto afirmar que entre 47,5% e 52,5% da população (com 95% de confiança) relataram ter tido resfriados no ano anterior, não podemos com tanta confiança estimar a porcentagem que realmente os teve. Alguns relatam resfriados quando tiveram outra doença; outros esquecem os resfriados que tiveram. Como os erros não-amostrais são provavelmente maiores que os erros de amostragem num desenho de amostra respeitável, é preciso muito cuidado ao generalizar de achados de amostra para a população.

Testes de Significância Estatística

Não há resposta científica à pergunta se uma associação entre duas variáveis é significativa, forte, importante, interessante ou digna de relato. Talvez o teste decisivo seja sua capacidade de persuadir sua audiência (presente e futura) da significância da associação. Mas os *testes de significância* podem ajudar nesse ponto. Como o nome sugere, a estatística paramétrica é a estatística que faz certas suposições sobre os parâmetros que descrevem a população da qual se extraiu a amostra.

Embora os *testes de significância estatística* sejam amplamente citados na literatura de ciências sociais, a lógica subjacente a eles é sutil e muitas vezes mal-entendida. Os testes de significância apóiam-se na mesma lógica de amostragem discutida neste livro. Para ajudar a entender essa lógica, retornemos ao conceito de erro amostral com relação a dados univariados.

Lembre que uma estatística amostral normalmente fornece a melhor estimativa do parâmetro de população correspondente, mas a estatística e o parâmetro raramente coincidem com exatidão. Desse modo, relatamos a probabilidade de que o parâmetro caia dentro de uma certa faixa (intervalo de confiança). O grau de incerteza dentro dessa faixa se deve a erro de amostragem normal. O corolário desta afirmação de probabilidade é o de que é *improvável* que o parâmetro caia fora da faixa especificada puramente como resultado de erro de amostragem. Assim, se estimamos que um parâmetro (com 99,9% de confiança) está entre 45% e 55%, dizemos, por implicação, que é *extremamente improvável* que o parâmetro seja na verdade, por exemplo, 90%, supondo que nosso único erro de estimativa se deva a amostragem normal. Essa é a lógica que fundamenta os testes de significância.

Uma Ilustração Gráfica. Vamos ilustrar esta lógica da significância estatística numa série de diagramas representando a seleção de amostras de uma população. Os elementos da lógica a ser ilustrada são:

1. Suposições com respeito à *independência* de duas variáveis no estudo da população.

2. Suposições com respeito à *representatividade* das amostras selecionadas através de procedimentos convencionais de amostragem probabilística.

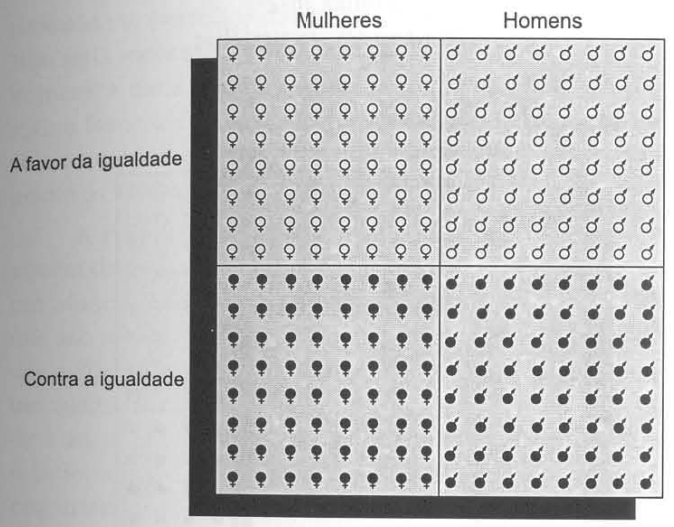
3. A *distribuição conjunta* observada dos elementos da amostra em termos das duas variáveis.

A Figura 16-1 representa uma população hipotética de 256 pessoas, metade homens e metade mulheres. O diagrama indica como cada pessoa se sente sobre as mulheres gozarem de igualdade com os homens. No diagrama, os que favorecem a igualdade são representados por círculos abertos, enquanto os que se opõem a ela são representados por círculos sombreados.

A questão investigada é se há relação entre gênero e sentimentos sobre igualdade entre homens e mulheres. Mais especificamente, se a tendência das mulheres favorecerem a igualdade é maior do que a dos homens, já que, presumivelmente, elas se beneficiariam mais dessa igualdade. Examine a Figura 16-1 e tente ver a resposta a essa pergunta.

FIGURA 16-1

Uma população hipotética de homens e mulheres que são favoráveis ou se opõem à igualdade sexual



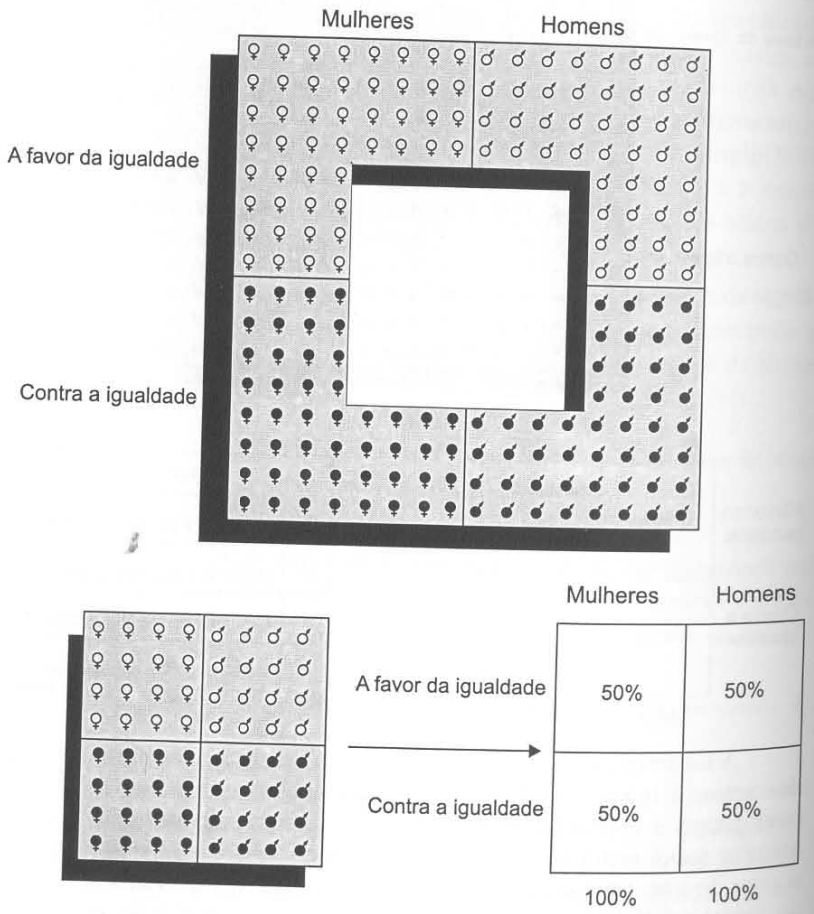
	Mulheres	Homens
A favor da igualdade	50%	50%
Contra a igualdade	50%	50%
	100%	100%

Legenda	
♀	Mulheres a favor da igualdade
♂	Homens a favor da igualdade
♀	Mulheres contra a igualdade
♂	Homens contra a igualdade

A ilustração indica que não há relação entre sexo e atitudes sobre a igualdade. Exatamente metade dos membros de cada grupo é a favor e metade é contra. Lembre a discussão anterior sobre redução proporcional de erro. Neste caso, saber o sexo de uma pessoa não reduziria os erros que cometeríamos ao tentar adivinhar a atitude dela com relação à igualdade. A tabela na parte inferior da Figura 16-1 fornece uma visão tabular do que se observa no diagrama.

A Figura 16-2 representa a seleção de uma amostra de um quarto da população hipotética. Na ilustração, um "quadrado" recortado no centro da população provê uma amostra representativa. Essa amostra tem dezesseis de cada tipo de pessoa. Metade são homens e metade são mulheres, e metade de cada grupo é a favor da igualdade, enquanto a outra metade se opõe a ela.

FIGURA 16-2
Uma amostra representativa



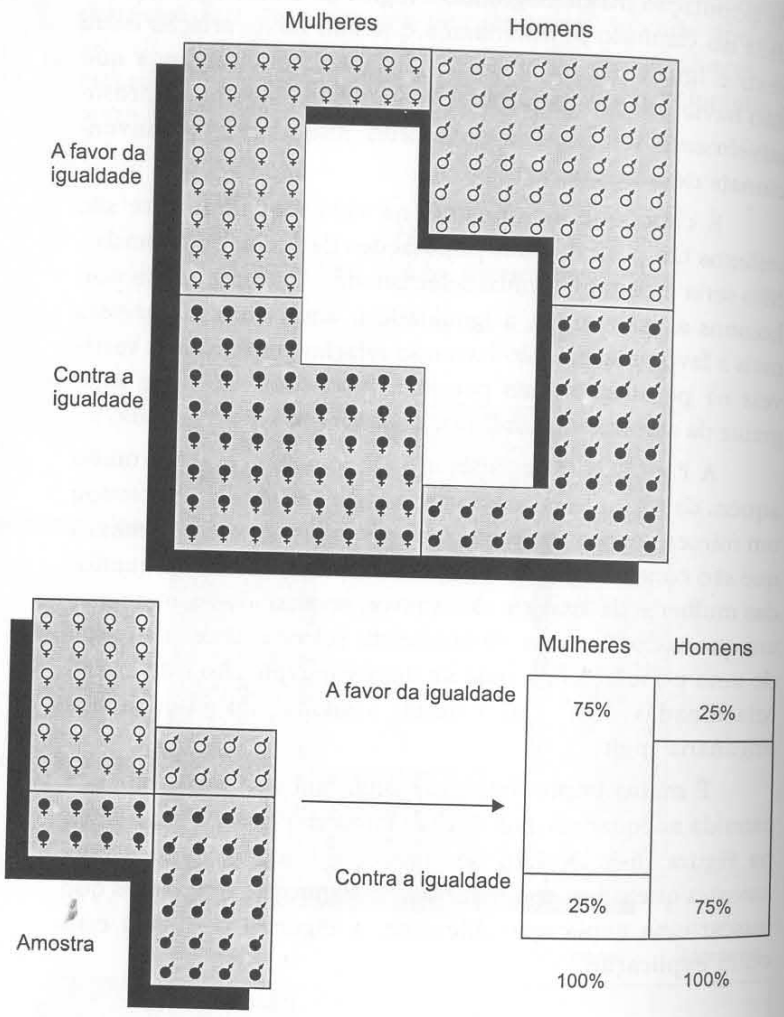
A amostra selecionada na Figura 16-2 permitiria tirar conclusões precisas sobre a relação entre sexo e igualdade na população maior. Seguindo a lógica de amostragem aprendida no Capítulo 5, notaríamos que não havia relação entre sexo e igualdade na amostra; portanto, concluiríamos que não havia relação na população maior (a amostra foi presumivelmente selecionada de acordo com as regras convencionais de amostragem).

É claro que as amostras na vida real raramente são reflexos tão perfeitos das populações de onde foram tiradas. Não seria incomum termos selecionado, digamos, um ou dois homens a mais contra a igualdade e umas duas mulheres a mais a favor, mesmo não havendo relação entre as duas variáveis na população. Tais pequenas variações são parte integrante da amostra probabilística, conforme se viu no Capítulo 5.

A Figura 16-3 representa uma amostra que fica muito aquém da representação da população maior. Ela selecionou um número excessivo de mulheres que apóiam e de homens que são contra a igualdade. Como mostra a tabela, três quartos das mulheres da amostra são a favor, comparados com apenas um quarto dos homens. Se tivéssemos selecionado esta amostra de uma população na qual as duas variáveis não estivessem relacionadas uma com a outra, a análise da amostra nos enganaria muito.

É muito improvável que uma amostra probabilística extraída adequadamente seja tão imprecisa quanto a mostrada na Figura 16-3. De fato, se realmente selecionássemos uma amostra que desse os resultados da Figura 16-3, teríamos que buscar uma explicação diferente. A Figura 16-4 ilustra essa outra explicação.

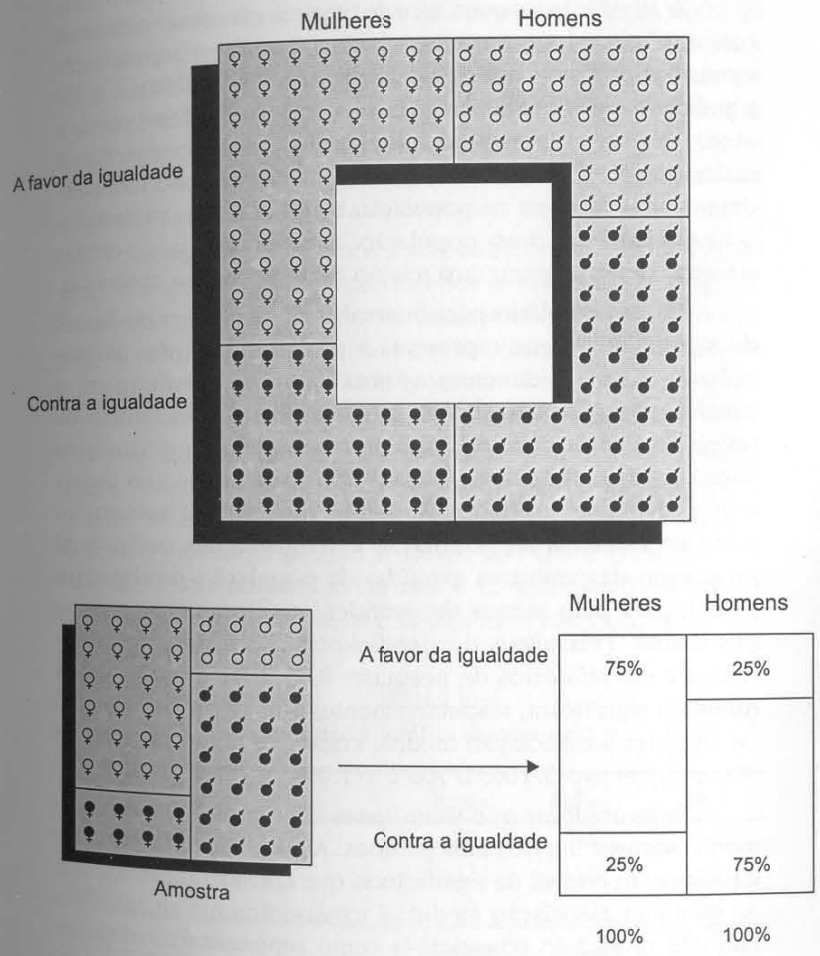
FIGURA 16-3
Uma amostra não representativa



Note que a amostra selecionada na Figura 16-4 também mostra uma forte relação entre gênero e igualdade, mas desta vez a razão é diferente. Selecionamos uma amostra perfeitamente representativa, mas há de fato uma forte relação entre as duas variáveis na população. Nesta figura, há, na população, maior probabilidade das mulheres apoiarem a igualdade do que os homens, e a amostra reflete isso.

Na prática, nunca sabemos o que é verdadeiro para a população total; é por isso que extraímos amostras. Se selecionássemos uma amostra e descobríssemos a forte relação mostrada nas Figuras 16-3 e 16-4, precisaríamos decidir se nosso achado refletiu a população de modo preciso ou foi apenas um resultado de erro de amostragem.

FIGURA 16-4
Uma amostra representativa de uma população na qual as variáveis se relacionam



A Lógica da Significância Estatística. A lógica fundamental dos testes de significação estatística, portanto, é a seguinte. Face a *qualquer* discrepância entre a suposta independência de variáveis numa população e a distribuição observada de elementos da amostra, podemos explicar a discrepância de um de dois modos: (1) podemos atribuí-la a uma amostra não representativa, ou

(2) podemos rejeitar a suposição de independência. A lógica e a estatística associadas com métodos de amostragem probabilística orientam sobre as probabilidades variadas de graus variados de não-representatividade (expressos como erro de amostragem). Ou seja, há uma *alta* probabilidade de um *pequeno* grau de não-representatividade e uma *baixa* probabilidade de um *grande* grau de não-representatividade.

A *significância estatística* de uma relação observada num conjunto de dados amostrais, portanto, é sempre expressa em termos de probabilidades. Significativo no nível 0,05 ($p < 0,05$) significa que a probabilidade de uma relação tão forte como a observada ser atribuível apenas a erro de amostragem não é maior que 5 em 100. Ou seja, se duas variáveis são independentes uma da outra na população e 100 amostras probabilísticas são extraídas desta população, não mais que cinco dessas amostras devem mostrar uma relação tão forte como a observada.

Há um corolário para intervalos de confiança em testes de significância, que representa a probabilidade das associações medidas se deverem *apenas* a erro de amostragem: o *nível de significância*. Como os intervalos de confiança, os níveis de significância derivam de um modelo lógico no qual várias amostras são extraídas de uma população. Ao determinar níveis de confiança, supomos que não há associação entre as variáveis na população e perguntamos então qual proporção das amostras extraídas da população produziriam associações pelo menos tão grandes quanto as encontradas nos dados. Três níveis de significância são usados freqüentemente em relatórios de pesquisa: 0,05, 0,01 e 0,001. Esses números significam, respectivamente, que as probabilidades de se obter a associação medida como resultado de erro de amostragem são 5/100, 1/100 e 1/1.000.

Pesquisadores que usam testes de significância normalmente seguem um dos dois padrões. Alguns especificam antecipadamente o nível de significância que consideram suficiente. Se qualquer associação medida é estatisticamente significativa naquele nível, vão considerá-la como representação de uma associação genuína entre as duas variáveis. Em outras palavras, estão dispostos a descontar a possibilidade de ela resultar de erro de amostragem.

Outros pesquisadores preferem informar o nível específico de significância para cada associação, desconsiderando as convenções de 0,05, 0,01 e 0,001. Em vez de relatar que uma certa associação é significativa no nível 0,05, podem relatar a significação no nível 0,023, indicando que as chances de ela ter resultado de erro de amostragem são 23 em 1000.

Qui-Quadrado. Qui-quadrado (χ^2) é um teste de significância muito usado nas ciências sociais. Baseia-se na *hipótese nula*, que é a suposição de não haver relação entre duas variáveis na população total. Dada a distribuição observada de valores nas duas variáveis separadas, computamos a distribuição conjunta que seria esperada se não houvesse relação entre as duas variáveis. O resultado desta computação é um conjunto de *freqüências esperadas* para todas as células na tabela de contingência. Comparamos esta distribuição esperada com a distribuição de casos realmente encontrada nos dados da amostra e determinamos a probabilidade de a discrepância descoberta ter resultado apenas de erro de amostragem. Vejamos um exemplo.

Suponha que estamos interessados na possível relação entre ir à igreja e gênero, no caso dos membros de uma igreja. Para testar a relação, selecionamos aleatoriamente uma amostra de 100 membros. Descobrimos que na amostra 40 são homens e 60, mulheres, e que 70% da amostra afirmam ter ido à igreja na semana anterior, os restantes 30% dizendo que não o fizeram.

Se não houver relação entre sexo e comparecimento à igreja, 70% dos homens na amostra deveriam ter ido à igreja na semana anterior e 30%, não. As mulheres deveriam ter comparecido na mesma proporção. A Tabela 16-7 (Parte I) mostra que, baseado neste modelo, 28 homens e 42 mulheres deveriam ter ido à igreja, com 12 homens e 18 mulheres não comparecendo.

TABELA 16-7
Ilustração hipotética de qui-quadrado

I. Freqüências Esperadas	Homens	Mulheres	Total
Compareceram	28	42	70
Não compareceram	12	18	30
Total	40	60	100
II. Freqüências Observadas	**Homens**	**Mulheres**	**Total**
Compareceram	20	50	70
Não compareceram	20	10	30
Total	40	60	100
III. (Observadas - Esperadas)^2 ÷ Esperadas	**Homens**	**Mulheres**	
Compareceram	2.29	1.52	χ^2=12.70
Não compareceram	5.33	3.56	p<.001

A Parte II da Tabela 16-7 mostra o comparecimento observado à igreja para a amostra de 100 membros. Observe que 20 homens afirmaram ter ido à igreja na semana anterior, enquanto outros 20 disseram que não foram. Das mulheres da amostra, 50 foram e 10 não. Comparando as freqüências esperadas e observadas (Partes I e II), observamos que menos homens foram à igreja do que o esperado, no caso das mulheres ocorrendo o contrário.

Qui-quadrado é calculado assim. Para cada célula na tabela, o pesquisador (1) subtrai a freqüência esperada para aquela célula da freqüência observada, (2) eleva esta quantidade ao quadrado e (3) divide a diferença ao quadrado pela freqüência esperada. Faz-se o mesmo para cada célula na tabela, e os resultados são agregados. (A Parte III da Tabela 16-7 apresenta as computações célula por célula.) O somatório final é o valor do qui-quadrado — 12,70 no exemplo atual.

Este valor é a discrepância geral entre a distribuição conjunta observada na amostra e a distribuição que devíamos esperar se as duas variáveis não se relacionassem entre si. Evidentemente, a simples descoberta de uma discrepância não prova que duas variáveis se relacionam. Um erro normal de amostragem pode produzir discrepâncias mesmo quando não há relação na população. Mas a grandeza do valor de qui-quadrado permite estimar a probabilidade disto haver acontecido.

Para determinar a significância estatística da relação observada, usamos um conjunto padrão de valores de qui-quadrado. Isto requer a computação dos *graus de liberdade*, que, para o qui-quadrado, são computados assim. O número de linhas na tabela de freqüências observadas, menos um, é multiplicado pelo número de colunas, menos um, que pode ser escrito como (l - 1) (c - 1). No exemplo, temos duas linhas e duas colunas (descontados os totais), portanto há um grau de liberdade.

Numa tabela de valores de qui-quadrado, vemos que, para um grau de liberdade e amostragem aleatória de uma população em que não há relação entre duas variáveis, 10% das vezes devemos esperar um qui-quadrado de pelo menos 2,7. Assim, se selecionarmos 100 amostras desta população, devemos esperar que cerca de 10 delas produzam um qui-quadrado igual ou maior que 2,7. Além disto, devemos esperar valores mínimos de qui-quadrado de pelo menos 6,6 em apenas 1% das amostras e valores de pelo menos 7,9 em apenas 0,5% das amostras. Quanto maior o valor do qui-quadrado, menor a probabilidade de este valor poder ser atribuído apenas a erro de amostragem.

No nosso exemplo, o valor do qui-quadrado é 12,70. Se não houvesse relação entre sexo e comparecimento à igreja na população de membros, e tivesse sido selecionado e estudado um grande número de amostras, esperaríamos um qui-quadrado desta magnitude em menos da metade de 1% das amostras. A probabilidade de obter um qui-quadrado desta magnitude é menos de 0,001%, supondo o uso de amostragem aleatória e não haver relação na população. Relatamos este achado afirmando que a relação é estatisticamente significativa no nível 0,001. Já que é tão improvável que a relação observada possa ter resultado apenas do erro de amostragem, rejeitamos a hipótese nula e supomos haver uma relação entre as duas variáveis na população de membros daquela Igreja.

A maioria das medidas de associação pode ser testada para significância estatística. Tabelas padronizadas de valores nos permitem determinar se uma associação tem significância estatística e em que nível. Qualquer manual de estatística traz instruções para o uso destas tabelas, de modo que não há necessidade de o fazermos aqui.

Revisão. Testes de significância fornecem uma medida objetiva em relação às quais podemos estimar a significância de associações entre variáveis, que nos ajudam a excluir associações que possam não representar associações genuínas na população estudada. Quem usar ou ler relatórios de testes de significância estatística deve, contudo, se precaver contra vários perigos na interpretação.

Primeiro, estamos discutindo testes de significância *estatística*; não há testes objetivos de significância substantiva. Assim, podemos estar legitimamente convencidos que uma certa associação não se deve a erro de amostragem, mas também podemos afirmar sem medo de contradição que duas variáveis têm uma relação mínima entre si. Erro de amostragem é uma função inversa ao tamanho da amostra: quanto maior a amostra, menor o erro esperado. Uma correlação de 0,1 pode ser significativa (em algum nível) se for descoberta numa amostra grande, mas a mesma correlação entre as mesmas variáveis não seria significativa numa amostra menor. Isto faz sentido se entendermos a lógica básica dos testes de significância. Na amostra maior, há menos chance de a correlação ser apenas o produto de erro de amostragem. Nas duas amostras,

porém, a correlação poderia representar uma correlação muito fraca, de essencialmente zero.

A distinção entre significância estatística e substantiva talvez seja melhor ilustrada em casos de certeza absoluta de que diferenças observadas não podem resultar de erros de amostragem. Este é o caso quando observamos uma população toda. Imagine que consigamos saber a idade de todas as autoridades governamentais nos EUA e na URSS. Só para argumentar, suponha que a média de idade das autoridades americanas fosse 45 e a das autoridades soviéticas 46. Não há erro de amostragem, pois sabemos a idade de todos. Sabemos com certeza que as autoridades soviéticas são mais velhas que as americanas. Obviamente diremos que a diferença não é substantivamente significativa. Concluiremos que, nos dois países, as autoridades têm essencialmente a mesma idade.

Segundo, a menos que sejamos enganados por este exemplo hipotético, *não* devemos calcular significância estatística de relações observadas em dados colhidos de populações inteiras. Testes de significância estatística medem a probabilidade de relações entre variáveis serem produtos apenas de erro de amostragem. Se não há amostragem, não pode haver erro de amostragem.

Terceiro, testes de significância se baseiam nas mesmas suposições de amostragem usados na computação dos intervalos de confiança. Se tais suposições não se enquadram no desenho da amostra, testes de significância não têm legitimidade.

Como acontece na maioria dos temas deste livro, também neste ponto tenho meu viés pessoal, neste caso, contra testes de significância. Não objeto à lógica estatística destes testes, que é sólida. Mas preocupa-me que estes testes parecem enganar mais que iluminar. Minhas reservas principais são:

1. Testes de significância fazem suposições de amostragem que virtualmente nunca são satisfeitas por desenhos de amostragens reais.

2. Dependem da ausência de erros não-amostrais, suposição questionável na maioria das medições empíricas.

3. Na prática, são aplicados com demasiada freqüência a medidas de associação cuja computação violou suposições destas medidas (por exemplo, correlações produto-momento computadas a partir de dados ordinais).

4. Significância estatística é facilmente mal entendida como "força de associação" ou significância substantiva.

Ao mesmo tempo — talvez paradoxalmente — acho que testes de significância podem ser valiosos para o pesquisador como instrumentos úteis para a compreensão de dados. Apesar de os comentários acima sugerirem uma abordagem conservadora dos testes de significância, isto é, que só devem ser usados quando todas as suposições são satisfeitas, minha perspectiva geral é justamente a inversa. Encorajo o uso de qualquer técnica estatística (qualquer medida de associação ou teste de significância) em qualquer conjunto de dados, se isto ajudar a compreender os dados. Se calcular correlações produto-momento entre variáveis nominais e testar significância estatística no contexto de amostragens não controladas satisfizerem este critério, então encorajo estas atividades. Digo isto no espírito do que Hanan Selvin chamou de técnicas para garimpar dados. Vale tudo, contanto que, ao cabo, haja compreensão dos dados e do mundo social sob estudo.

Porém, o preço desta liberdade radical é abdicar de interpretações estatísticas estritas. Você não poderá mais basear a importância definitiva de seus achados apenas numa correlação significativa no nível 0,05. Qualquer que seja a rota da descoberta, os dados empíricos têm de ser apresentados de forma legítima e sua importância deve ser argumentada logicamente.

Resumo

Neste capítulo, trabalhamos com uma visão ampla e abrangente do mundo das estatísticas sociais. Não pretendi fazer do leitor um especialista em estatísticas, mas entender a lógica básica de algumas técnicas estatísticas mais usadas. Começamos fazendo uma distinção ampla entre estatísticas *descritivas* e *inferenciais*.

Estatísticas descritivas são usadas para resumir dados. Algumas resumem a distribuição de atributos numa só variável; outras, chamadas *medidas de associação*, resumem a associação entre variáveis.

Estatísticas inferenciais são usadas para estimar a generalizabilidade de achados obtidos pela análise de amostras às populações das quais foram extraídas. Algumas estatísticas inferenciais estimam as características da população para uma variável (em termos de *níveis* e *intervalos de confiança*). Outras, chamadas *testes de significância estatística*, estimam as relações entre variáveis da população.

Muitas medidas de associação se baseiam num modelo de *redução proporcional de erro* (RPE), baseado na comparação de (a) o número de erros que cometeríamos se tentássemos adivinhar os atributos de uma variável para cada caso estudado, se tudo o que soubéssemos fosse a distribuição dos atributos desta variável, e (b) o número de erros que cometeríamos se soubéssemos a distribuição conjunta geral e fôssemos informados em cada caso sobre o atributo de uma variável, cada vez que procurássemos adivinhar o atributo da outra.

Lambda é uma medida de associação para a análise de duas varáveis *nominais*. Oferece uma ilustração límpida do modelo RPE. Gamma é uma medida de associação para a análise de duas varáveis *ordinais*. A *correlação produto-momento* de Pearson é uma medida de associação para a análise de duas variáveis *intervalo* ou *de razão*.

Inferências sobre alguma característica de uma população, tal como o percentual de eleitores a favor do candidato A, devem indicar o *intervalo de confiança* (âmbito dentro da qual o valor é esperado, por exemplo, 45% a 55% a favor do candidato A) e uma indicação do *nível de confiança* (a probabilidade de que o valor esteja dentro deste âmbito, por exemplo, 95% de confiança). Cálculos de níveis e intervalos de confiança se baseiam na teoria da probabilidade e supõem técnicas convencionais de amostragem probabilística.

Inferências sobre a generalizabilidade para a população das associações descobertas entre variáveis numa amostra envolvem *testes de significância estatística*, que estimam a probabilidade de uma associação tão forte quanto a observada resultar de erro de amostragem normal, caso não haja associação entre as variáveis na população maior. Portanto, testes de significância estatística também se baseiam na teoria da probabilidade e supõem técnicas convencionais de amostragem probabilística. Significância estatística não deve ser confundida com significância *substantiva*; esta quer dizer que uma associação observada é forte, importante, significativa.

O nível de significância de uma associação observada é relatado como a probabilidade de ela ter resultado apenas de erro de amostragem. Afirmar que uma associação é significativa no nível 0,05 é dizer que uma associação tão forte quanto a observada não poderia ser esperada como resultado de erro de amostragem mais do que 5 vezes em 100. Pesquisadores de *survey* tendem a usar um leque particular de níveis de significância em conexão com testes de significância estatística: 0,05, 0,01 e 0,001. Isto é apenas uma convenção.

Falando estritamente, testes de significância estatística fazem suposições sobre dados e métodos quase nunca completamente satisfeitos por *surveys* reais. Apesar disto, os testes podem ser úteis na análise e interpretação dos dados. Mas tome cuidado para não interpretar a "significância" do teste precisamente demais.

Notas

[1] Dois livros manuais de estatística social que podem ser úteis: JENDREK, Margaret Platt. *Through the Maze*: Statistics with Computer Applications. Belmont, CA: Wadsworth, 1985; BLALOCK, Hubert. *Social Statistics*. New York: McGraw-Hill, 1979.

[2] LOPATA, Helena Znaniecki. Widowhood and Husband Sanctification. *Journal of Marriage and the Family*, p.439-450, maio 1981.

Leituras Adicionais

BLALOCK, Hubert. *Social Statistics*. New York: McGraw-Hill, 1979.

HENKEL, Ramon E. *Tests of Significance*. Beverly Hills, CA: Sage, 1976.

JENDREK, Margaret Platt. *Through the Maze:* Statistics with Computer Applications. Belmont, CA: Wadsworth, 1985.

KISH, Leslie. Chance, Statistics and Statisticians. *Journal of the American Statistical Association*, v.73, n.361, p.1-6 mar. 1978.

MORRISON, Denton, HENKEL, Ramon (Ed.). *The Significance Test Controversy:* a Reader. Chicago: Aldine-Atherton, 1970.

SHARP, Vicki. *Statistics for the Social Sciences*. Boston: Little, Brown, 1979.