

Capítulo 13

Lógica da Medição e da Associação

O cerne da análise do *survey* se reparte entre a descrição e a explicação. O analista de *survey* mede variáveis e depois examina as associações entre elas. Mas, como vimos no Capítulo 1, há muita confusão sobre a natureza deste processo.



Como já apresentamos os vários aspectos do desenho do *survey*, especialmente a conceituação e o desenho do instrumento, convém retornarmos à imagem tradicional do método científico. Já criticamos essa imagem, usando um exemplo de *survey*; examinemos agora, esquematicamente, a imagem tradicional e, depois, uma imagem alternativa, mais correta, da ciência na prática. A seguir, consideraremos a noção de Paul Lazarsfeld sobre a “interpermutabilidade de índices”.

A Imagem Tradicional

A perspectiva tradicional do método científico baseia-se numa seqüência de passos, resumidos na seguinte lista, que se acredita que os cientistas sigam em seu trabalho.

1. Construção da teoria
2. Derivação de hipóteses teóricas
3. Operacionalização de conceitos
4. Coleta de dados empíricos
5. Teste empírico de hipóteses

Antes de refutar esta perspectiva tradicional, vamos examiná-la melhor.

Elementos do Modelo Tradicional

Construção da Teoria. Face a um aspecto interessante do mundo natural ou social, o cientista supostamente deduz uma teoria abstrata para descrevê-lo. Supõe-se que este exercício é puramente lógico. Suponha que você esteja interessado em comportamento desviante. Você constrói, baseado na teoria sociológica existente, uma teoria do comportamento desviante. Entre outras coisas, essa teoria inclui uma variedade de conceitos relevantes às causas do comportamento desviante.

Derivação das Hipóteses Teóricas. Com base na sua teoria geral do comportamento desviante, você, supostamente, deriva hipóteses relacionando os vários conceitos que formam a teoria. Este é, também, um procedimento puramente lógico. No exemplo acima, suponha que você derive, logicamente, a hipótese de que a delinqüência juvenil é uma função da supervisão. Especificamente: à medida que aumenta a supervisão, diminui a delinqüência juvenil.

Operacionalização de Conceitos. O próximo passo na visão tradicional é a especificação de indicadores empíricos que representarão seus conceitos teóricos. Enquanto conceitos teóricos devem ser um tanto abstratos e talvez um pouco vagos, seus indicadores empíricos precisam ser específicos e precisos. Portanto, em nosso exemplo, você pode operacionalizar o conceito "juvenil" como pessoas com menos de dezoito anos de idade, "delinqüência" como ter sido preso por um ato criminoso, e "supervisão" como a presença de um adulto em casa, não trabalhando.

O efeito da operacionalização é converter a hipótese teórica em empírica. No caso, a hipótese empírica seria a seguinte: entre as pessoas com menos de dezoito anos de idade, as que moram em casa com um adulto que não trabalha têm menor probabilidade de serem presas por atos criminosos do que as sem um adulto não trabalhando em casa.

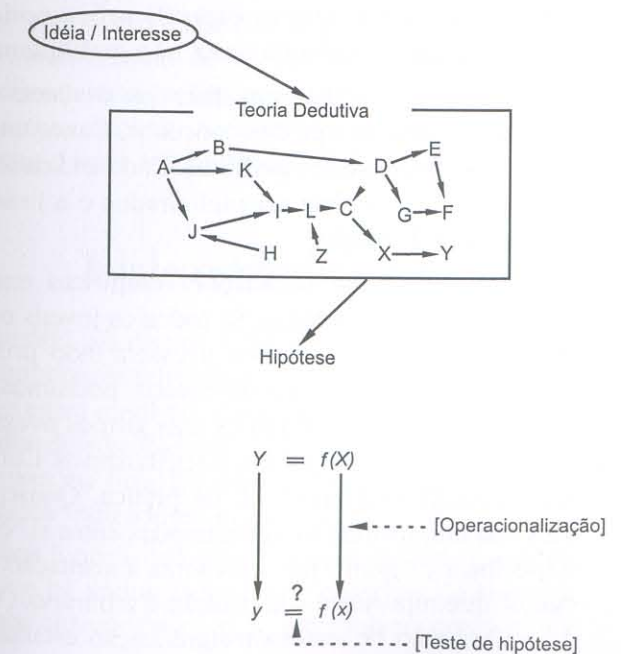
Coleta de Dados Empíricos. Baseado na operacionalização dos conceitos teóricos, você coleta dados relacionados aos indicadores empíricos. No nosso exemplo, você pode fazer um *survey* com pessoas com menos de dezoito anos de idade. Entre outras coisas, o questionário perguntaria a cada uma se ela mora em casa com um adulto que não trabalha e se já foi presa por cometer um ato criminoso.

Teste Empírico de Hipóteses. Após coletar os dados, o passo final é o teste estatístico da hipótese. Você determina empiricamente se os jovens com adultos que não trabalham em casa têm menor probabilidade de serem presos por atos criminosos do que os que não têm esses adultos. A confirmação ou desconfirmação da hipótese empírica é então usada para aceitar ou rejeitar a hipótese teórica.

Uma Apresentação Esquemática

A imagem tradicional da ciência é mostrada de modo esquemático na Figura 13-1. Você começa com um interesse particular sobre o mundo, cria uma teoria geral sobre ele e usa a teoria dedutiva para gerar uma hipótese sobre a associação entre duas variáveis. Esta hipótese é representada na forma $Y = f(X)$. Esta expressão é lida "Y é uma função de X", significando que os valores de Y são determinados ou causados por valores de X. No nosso exemplo, delinqüência (Y) é uma função de supervisão (X). A seguir, você operacionaliza as duas variáveis especificando medidas empíricas para representá-las no mundo real, coleta dados relevantes a tais medidas e, finalmente, testa empiricamente a relação esperada.

FIGURA 13-1
A imagem tradicional da ciência



Dois Problemas Básicos

Esta descrição da visão tradicional do método científico pode convencê-lo de que pesquisa científica é uma atividade rotineira. Tudo o que é preciso fazer é seguir religiosamente os passos 1 a 5 e se terá descoberto a verdade.

Infelizmente, a pesquisa científica e o mundo empírico que estudamos não são tão nítidos. A visão tradicional do método não representa o que realmente acontece na pesquisa científica — social ou outra. Dois problemas básicos impedem que o sonho se torne realidade.

Primeiro, conceitos teóricos raramente permitem operacionalizações não-ambíguas. Conceitos são abstratos e genéricos, enquanto a especificação de indicadores empíricos se faz por aproximação. No exemplo dado, é improvável que a variável “supervisão” seja representada adequadamente pela presença de um adulto que não trabalha em casa. A simples presença deste adulto não garante a supervisão de um jovem e, em algumas casas sem tal adulto, pode haver outros modos de fazê-la.

Igualmente, ser preso por um ato criminoso não é o mesmo que o conceito abstrato de “delinqüência”. Alguns jovens podem ter comportamento delinqüente sem serem presos. Outros podem ser presos erradamente. Além disso, a especificação de “juvenil” como uma pessoa abaixo de dezoito anos de idade é arbitrária. Outras especificações podem ser feitas e provavelmente nenhuma seria não-ambiguamente correta.

Além disso, não basta dizer que devíamos ter especificado “melhores” indicadores dos conceitos. Conceitos teóricos quase nunca têm indicadores perfeitos. Todo indicador empírico tem defeitos; todos podem ser melhorados e a busca de melhores indicadores é infinita.

Segundo, as associações empíricas entre as variáveis quase nunca são perfeitas. Se todos os jovens com adultos não trabalhando em casa nunca tivessem sido presos e todos os sem tal adulto tivessem sido presos, podíamos concluir que a hipótese foi confirmada. Se os dois grupos tivessem exatamente o mesmo número de presos, a rejeitaríamos. Contudo, nenhuma das alternativas é provável, na prática. Quase todas as variáveis são empiricamente relacionadas entre si “em algum grau”. Especificar o “grau” que representa a aceitação da hipótese e o “grau” que representa sua rejeição é arbitrário. (Veja no Capítulo 16 a discussão de testes de significação estatística.)

Na realidade, usamos indicadores aproximados dos conceitos teóricos para descobrir associações parciais. Estes problemas conspiram entre si contra você. Suponha que você especifique um grau de associação para a aceitação da hipótese e a análise empírica não o atinge. Naturalmente, você se perguntará se indicadores diferentes dos conceitos podiam ter resultado no grau especificado de associação.

Fiz estes comentários para indicar que a visão tradicional do método científico não é apropriada à pesquisa na prática. Não se faz pesquisa seguindo obstinadamente o caminho entre os passos 1 e 5. Isto não deve desanimá-lo, contudo, e sim desafiá-lo. Não deve ser considerado como negação da possibilidade da pesquisa científica, mas como base para a pesquisa esclarecida, realmente científica.

Medição e associação são conceitos inter-relacionados que devem ser trabalhados lógica e simultaneamente. Em vez de dar ordenadamente uma série de passos fixos, você deve dar estes passos para frente e para trás, incessantemente. Suas construções teóricas são, muitas vezes, feitas em torno de associações previamente observadas entre indicadores empíricos, construções teóricas parciais podem sugerir o exame de novos dados empíricos, e assim por diante. Espera-se que após cada atividade você entenda um pouco melhor o tema do seu interesse. É raro o “experimento crítico” que determina de uma vez por todas o destino de toda uma teoria.

Portanto, a pesquisa científica é uma empreitada sem fim, visando o entendimento de algum fenômeno. Para isso, você mede e examina associações, sempre consciente da natureza das suas inter-relações. As seções seguintes explicam melhor a natureza dessas inter-relações.

A Intercambialidade de Índices

Paul Lazarsfeld, ao discutir a “intercambiabilidade de índices”, criou uma importante ferramenta conceitual para a compreensão da relação entre medição e associação, assim como uma solução parcial dos dois problemas discutidos na seção anterior.¹ Essa discussão nasceu do reconhecimento de que qualquer conceito tem vários indicadores possíveis.

Retornemos à noção de uma hipótese teórica $Y = f(X)$. Lazarsfeld viu que há vários indicadores possíveis de supervisão, que podemos chamar de x_1, x_2, x_3 etc. Embora possa haver

razões para acreditar que alguns possíveis indicadores são melhores do que outros, eles são essencialmente intercambiáveis. Portanto, enfrentamos o dilema de quais indicadores usar no teste da hipótese $Y = f(X)$.

A solução para o dilema é usar todos os indicadores. Assim, testamos as hipóteses empíricas: $y = f(x_1)$, $y = f(x_2)$, $y = f(x_3)$ etc. Em vez de ter um teste para a hipótese, temos vários, como indicado esquematicamente na Figura 13-2.

Surge um novo dilema. Se o cientista que segue a visão tradicional do método enfrenta o problema da única associação empírica talvez não ser perfeita, o cientista que usa este método está diante de várias associações empíricas, nenhuma das quais perfeita, algumas delas podendo entrar em conflito com outras. Portanto, mesmo havendo especificado um grau de associação como suficiente para confirmar a hipótese, talvez você descubra que os testes usando x_1 , x_3 e x_5 satisfazem o critério especificado, mas os testes usando x_2 e x_4 não. Como resolver este dilema?

Nos termos da noção da "intercambiabilidade de índices", a hipótese teórica é aceita como uma proposição *geral* se for confirmada por todos os testes empíricos específicos. Se, por exemplo, a delinqüência juvenil for uma função da supervisão, no sentido geral deste conceito, ela deve estar relacionada empiricamente a todo indicador empírico de supervisão.

Se, no entanto, o cientista descobrir que somente certos indicadores da supervisão se relacionam à delinqüência juvenil, então os tipos de supervisão aos quais a proposição se aplica foram especificados. Na prática, isto pode ajudar a reconceituar "supervisão" em termos mais gerais. Talvez, por exemplo, a delinqüência juvenil seja uma função de restrições estruturais, e alguns tipos de supervisão sejam indicadores destas restrições e outros não.

É importante entender o que se alcança através deste processo. Em vez de testar rotineiramente uma hipótese fixa, relacionando supervisão a delinqüência, você ganha uma compreensão mais abrangente da natureza da associação. Mas este feito só é significativo se virmos o objetivo da ciência como compreensão crescente, em vez de somente construir teorias e testar hipóteses.

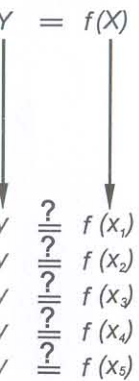
A implicação principal dos comentários anteriores é que a medição e a associação estão fortemente interligadas. A medição de uma variável faz pouco sentido fora dos contextos empíricos e teóricos das associações a serem testadas. Perguntado "Como devo medir classe social?", o cientista experiente responderá "Qual é seu propósito ao medi-la?" A maneira adequada de medir uma variável depende muito das variáveis que serão associadas a ela. Outro exemplo deve esclarecer isto.

Recentemente, surgiu uma controvérsia na sociologia religiosa sobre a relação entre religiosidade e preconceito. Um livro de Charles Y. Glock e Rodney Stark, *Christian Beliefs and Anti-Semitism*,² relatou dados empíricos indicando que membros de igrejas cristãs com crenças ortodoxas tinham maior probabilidade de serem anti-semitas do que membros menos ortodoxos. Estas descobertas acirraram consideráveis discussões nas igrejas e resultaram em outras pesquisas sobre o mesmo tópico por outros pesquisadores.

Um outro projeto de pesquisa chegou à conclusão oposta à de Glock e Stark, relatando que, quando a ortodoxia crescia, diminuía o preconceito. Mas análise mais minuciosa mostrou que as medidas de ortodoxia, nesta pesquisa, se basearam na aceitação de declarações de questionário que refletiam as doutrinas cristãs tradicionais de que "Todos os homens são irmãos" e "Ame o seu próximo". Não foi surpreendente que respondentes que aceitaram as declarações baseadas nessas doutrinas pareceram menos preconceituosos do que os que as rejeitaram. Normalmente, estas descobertas seriam (e foram) questionadas por causa de "contaminação". As duas variáveis examinadas (ortodoxia religiosa e tolerância) na verdade mediam qualidades idênticas ou similares. Chamar um conjunto de indicadores de "ortodoxia" e outro de "tolerância" não mostra que preconceito diminui com o aumento da ortodoxia em qualquer sentido geral.

As discussões deste capítulo sugerem uma reação um pouco diferente às descobertas da pesquisa. Perguntando *como* ortodoxia e preconceito estão associados entre si, em vez de perguntar se eles estão associados, concluímos que a ortodoxia medida pelos indicadores de Glock e Stark (crença em Deus, Jesus, milagres, e coisas assim) está positivamente associada

FIGURA 13-2
Intercambiabilidade
de índices



ao preconceito, enquanto a ortodoxia medida como compromisso com as normas de amor fraterno e igualdade está negativamente associada a preconceito. Ambas conclusões estão empiricamente corretas, apesar de nenhuma delas responder à pergunta geral, se religião e preconceito estão relacionados entre si.

É claro que o passo final é avaliar a utilidade relativa das conclusões. O achado de que ortodoxia e preconceito estão negativamente associados provavelmente seria descartado como tautológico ou trivial. (É claro que a medida da ortodoxia em termos de amor fraterno e de igualdade pode ser muito útil em algum outro contexto.)

Resumo

O objetivo deste capítulo foi prover uma perspectiva saudável sobre os objetivos da medição e da associação na ciência. Tentei proporcionar-lhe, a este respeito, melhor compreensão da empresa científica. O capítulo começou revendo o discurso tradicional de manual sobre o método científico, indicando que esta visão não reflete com exatidão a pesquisa científica na prática. No lugar da perspectiva tradicional, tentei oferecer um modelo alternativo, que acredito ser mais útil nas atividades de pesquisa reais.

Minha motivação para abordar essa questão vem da observação das dificuldades de pesquisadores inexperientes, que aceitam a perspectiva tradicional como um quadro verdadeiro de como cientistas "reais" trabalham, resultando em profundo desapontamento com a própria pesquisa. Não quero dizer que a pesquisa científica é descuidada ou "não-científica", apenas que ela não é rotineira. O cientista é bem diferente do técnico.

Notas

¹ LAZARFELD, Paul. Problems in Methodology. In: MERTON, Robert K. (Ed.). *Sociology Today*. New York: Basic Books, 1959. p.375.

² GLOCK, Charles Y., STARK, Rodney. *Christian Beliefs and Anti-Semitism*. New York: Harper & Row, 1967.

Leituras Adicionais

BABBIE, Earl. *Observing Ourselves: Essays in Social Research*. Belmont, CA: Wadsworth, 1986.

LAZARFELD, Paul. Problems in Methodology. In: MERTON, Robert K. (Ed.). *Sociology Today*. New York: Basic Books, 1959.

LAZARFELD, Paul F., PASANELLA, Ann K., ROSENBERG, Morris (Ed.). *Continuities in the Language of Social Research*. New York: Free Press, 1972. seção I.

Capítulo 14

Construindo e Compreendendo Tabelas

A maioria das análises de *survey* caem na rubrica geral da análise multivariada; a Parte 4 concentra-se nas variedades da análise multivariada. O termo se refere ao exame simultâneo de diversas variáveis. A análise das associações simultâneas entre idade, escolaridade e preconceito é um exemplo de análise multivariada.

A análise multivariada não é uma forma específica de análise; técnicas específicas para fazer análise multivariada incluem análise fatorial, análise do menor espaço, correlação múltipla, regressão múltipla e análise de trajetória [*path analysis*], entre outras. A lógica básica da análise multivariada pode ser mais bem-estudada através de tabelas simples, chamadas tabelas de contingência ou tabulações cruzadas. Portanto, este capítulo debruça-se sobre a construção e a compreensão dessas tabelas. Além disso, a análise multivariada não pode ser totalmente compreendida sem a compreensão firme de dois modos analíticos ainda mais fundamentais, análise univariada e análise bivariada. Portanto, o grosso deste capítulo discute estes modos analíticos.

Análise Univariada

Análise univariada é o exame da distribuição de casos de apenas uma variável de cada vez. Começaremos nossa discussão com a lógica e os formatos para a análise dos dados univariados.

Distribuições

O formato mais básico para apresentar dados univariados é relatar todos os casos individuais, isto é, listar o atributo de cada caso estudado na variável em questão. Suponha que você esteja interessado nas idades de executivos pesquisados num estudo sobre práticas empresariais. A maneira mais direta de relatar as idades dos executivos seria listá-las: 63, 57, 49, 62, 80, 72, 55 etc. Apesar desta lista dar ao leitor os dados mais detalhados, ela seria complicada demais para a maioria dos propósitos. Poderíamos distribuir os dados de maneira mais administrável sem perder qualquer detalhe, relatando que cinco executivos têm 38 anos de idade, sete têm 39, dezoito têm 40, e assim por diante. Tal formato impediria a duplicação de dados nesta variável.

Um formato ainda mais administrável, com alguma perda de detalhe, seria relatar as idades dos executivos como *marginais*, que são *distribuições de frequências* de dados agrupados: 246 executivos com menos de 45 anos, 517 entre 45 e 50 anos, e assim por diante. Neste caso, o leitor teria menos dados para examinar e interpretar mas não seria capaz de reproduzir completamente as idades originais dos executivos. Por exemplo, o leitor não teria como saber quantos deles tinham 41 anos.

O exemplo anterior apresentou os marginais na forma de números brutos. Uma forma alternativa seria o uso de *porcentagens*. Por exemplo, você pode relatar que x% dos executivos têm menos de 45 anos, y% estão entre 45 e 50 etc. (veja a Tabela 14-1). Frequentemente, você precisa decidir qual será a *base* a partir da qual computar a porcentagem, ou seja, o número que representa 100%. Nos exemplos mais diretos, a base é o número total de casos estudados. Entretanto, temos problemas sempre que faltarem dados para alguns casos. Por exemplo, suponha um *survey* no qual os respondentes devem informar a idade. Se alguns não responderem essa pergunta, há duas alternativas. Primeira, basear as porcentagens sobre o número total de respondentes, relatando os que não respondem como porcentagens do total. Segunda, usar o número de pessoas que responderam a pergunta como a base sobre a qual computar as porcentagens. Deve-se relatar o número dos que não responderam, mas estes não figuram nas porcentagens.

TABELA 14-1
Uma ilustração de análise univariada

Idades de Executivos de Empresas (Hipotéticas)	
Abaixo de 35	9%
36-45	21
46-55	45
56-65	19
66 ou mais	6
	100 % = (433)
Sem dados =	(18)

A escolha da base depende inteiramente do propósito da análise. Se este for comparar a distribuição etária da amostra com dados comparáveis da população da qual foi tirada a amostra, provavelmente deve-se omitir da computação os "sem resposta". A melhor estimativa da distribuição etária de todos os respondentes é a distribuição dos que responderam a pergunta. Como "sem resposta" não é uma categoria etária significativa, sua presença entre as categorias de base confunde a comparação dos números da amostra e da população. (Veja o exemplo da Tabela 14-1.)

Tendência Central

Além de informar os marginais, você pode apresentar os dados na forma de *médias* resumidas ou medidas de *tendência central*. Suas opções são a *moda* (o atributo mais freqüente, agrupado ou não agrupado), a *média aritmética* e a *mediana* (o atributo do meio na distribuição ordenada dos atributos observados). Vejamos como as três medidas são calculadas.

Suponha que você está fazendo um estudo-piloto que entreviste adolescentes, cujas idades, na sua amostra, variam de 13 a 19 anos.

Idade	Número
13	3
14	4
15	6
16	8
17	4
18	3
19	3

Tendo listado as idades dos 31 respondentes, qual você diria é a idade deles em geral, ou “na média”? Há três diferentes maneiras de responder esta pergunta.

A medida mais fácil de calcular é a *moda*, o valor mais freqüente. Como se vê na lista, há mais pessoas com 16 anos (oito) do que com qualquer outra idade, de forma que a idade modal é 16, como indicado na Figura 14-1.

A Figura 14-1 também mostra o cálculo da média, que envolve três passos: (1) Multiplique cada idade pelo número de entrevistados com esta idade, (2) some os resultados destas multiplicações e (3) divida o total pelo número de entrevistados. Como se vê, a idade média neste exemplo é 15,87.

A *mediana* representa o valor “do meio”; metade está acima dele e metade abaixo. Se tivéssemos a *exata* idade de cada entrevistado (por exemplo, 17 anos e 124 dias), poderíamos pôr os 31 entrevistados em ordem de idade, e a mediana para todo o grupo seria a idade do entrevistado do meio. Mas não sabemos as idades exatas; nossos dados são “dados agrupados”, no sentido em que três pessoas que não têm precisamente a mesma idade foram agrupadas na categoria “13 anos de idade”, por exemplo.

FIGURA 14-1
Três “Médias”

Idade	Número
13	☺☺☺
14	☺☺☺☺
15	☺☺☺☺☺☺
16	☺☺☺☺☺☺☺☺
17	☺☺☺☺
18	☺☺☺
19	☺☺☺

Modo = 16
Mais freqüente

Idade	Número
13	☺☺☺
14	☺☺☺☺
15	☺☺☺☺☺☺
16	☺☺☺☺☺☺☺☺
17	☺☺☺☺
18	☺☺☺
19	☺☺☺

13 x 3 = 39
14 x 4 = 56
15 x 6 = 90
16 x 8 = 128
17 x 4 = 68
18 x 3 = 54
19 x 3 = 54

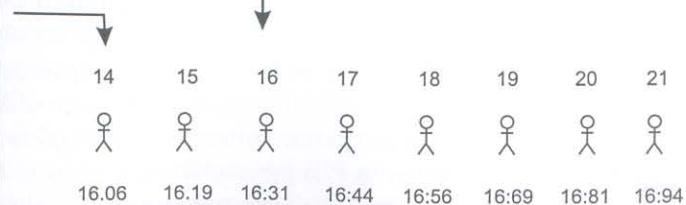
Média = 15.87
Média Aritmética

$$\frac{492}{31} = 15.87$$

(Total) (Casos)

Idade	Número
13	☺☺☺ 1-3
14	☺☺☺☺ 4-7
15	☺☺☺☺☺☺ 8-13
16	☺☺☺☺☺☺☺☺
17	☺☺☺☺ 22-25
18	☺☺☺☺ 26-28
19	☺☺☺☺ 29-31

Mediana = 16.31
Ponto Médio



A Figura 14-1 exemplifica a lógica do cálculo da mediana para dados agrupados. Como há 31 entrevistados, o entrevistado “do meio” seria o número 16 se eles forem ordenados por idade; quinze seriam mais jovens e quinze mais velhos. Na parte de baixo da Figura 14-1, você vê que a pessoa do meio é uma das oito com 16 anos de idade. Na ampliação do grupo, vemos que o número 16 é o terceiro da esquerda para a direita.

Não sabemos as idades exatas dos entrevistados deste grupo; a convenção estatística em tais casos é presumir que eles estejam distribuídos igualmente ao longo da extensão do grupo. Neste caso, as *possíveis* idades dos entrevistados variam de 16 anos e zero dias a 16 anos e 364 dias. Falando de maneira mais rigorosa, a amplitude é de 364/365 anos. Na prática, basta chamar isso de um ano.

Se os oito entrevistados deste grupo estivessem igualmente distribuídos de um extremo ao outro, estariam com um oitavo de ano de diferença — um intervalo de 0,125-ano. O exemplo mostra que, se colocarmos o primeiro entrevistado na metade do intervalo a partir do limite inferior e acrescentarmos um intervalo completo à idade de cada entrevistado seguinte, o último entrevistado estará à metade de um intervalo de distância do limite superior.

Portanto, o que fizemos foi calcular hipoteticamente as idades exatas dos oito entrevistados, supondo que suas idades estejam distribuídas igualmente. Feito isso, anotamos a idade do entrevistado do meio (16,31) e a usamos como a idade mediana para o grupo.

Sempre que o número total de entrevistados for par, não haverá caso no meio, é claro. Nesta situação, você calcula a média dos dois valores que o cercam. Suponha, por exemplo, que houvesse mais um adolescente de 19 anos. O ponto do meio então cairia entre os números 16 e 17. A média, neste caso, seria calculada como $(16,31 + 16,44)/2 = 16,38$.

Na literatura de pesquisas, se mostram as médias e as medianas. Sempre que as médias forem mostradas, lembre que elas são susceptíveis ao efeito de valores extremos — poucos números muito grandes ou muito pequenos. Por exemplo, o Produto Interno Bruto (PIB) *per capita* de 1988 para os Estados Unidos foi de US\$17.500, em contraste com US\$400 para o Sri Lanka, US\$5.080 para a Irlanda, e US\$11.910 para a Austrália. Já o pequeno reino petrolífero dos Emirados Árabes Unidos teve um PIB *per capita* de US\$14.410, apesar de a maioria dos residentes daquele país viverem na pobreza, como se pode inferir

do fato de que sua taxa de mortalidade infantil foi quase quatro vezes maior que a dos Estados Unidos em 1988. O alto PIB *per capita* dos Emirados Árabes reflete a enorme petro-riqueza de poucos.¹

Dispersão

As medidas de tendência central oferecem a vantagem de reduzir os dados brutos a uma forma mais manejável. Um único número (ou atributo) pode representar todos os dados detalhados coletados sobre uma variável. Esta vantagem tem um custo, porque o leitor não poderá reconstruir os dados originais a partir da média, ou mediana. Esta desvantagem pode ser mais ou menos compensada pela informação sobre a *dispersão* das respostas. A medida mais simples da dispersão é a *amplitude* — a distância que separa o valor mais alto do mais baixo. Assim, além de informar que nossos entrevistados têm uma idade média de 15,87 anos, também podemos informar que suas idades variam de 13 a 19. Uma medida um pouco mais sofisticada de dispersão é o *desvio padrão*. A lógica dessa medida foi discutida no Capítulo 5, ao descrever o erro padrão da distribuição amostral.

Há várias outras medidas da dispersão, como por exemplo o *desvio quartílico* (ou *amplitude semi-interquartilica*). Um quartil representa um quarto dos casos estudados. Por exemplo, se estivermos estudando QI, o primeiro quartil seria composto pelos indivíduos com os QIs mais altos, o quarto quartil com aqueles com os QIs mais baixos. O desvio quartílico é a metade da distância entre o fim do primeiro quartil e o começo do quarto quartil. Se o quartil mais alto tiver QIs variando de 120 a 150, e o mais baixo tiver valores indo de 60 a 90, o desvio quartílico será $(120-90)/2 = 15$.

Veja que, já que os quatro quartis são de mesmo tamanho, o primeiro e segundo quartis têm metade das pessoas com os valores mais altos, e o terceiro e quarto, a metade das pessoas com os valores mais baixos. Assim, o ponto médio entre o segundo e o terceiro quartis é o valor mediano.

Variáveis Contínuas e Discretas

Os cálculos descritos acima não são apropriados para todas as variáveis. Para entender por que, precisamos distinguir dois tipos de variáveis: *contínuas* e *discretas*. Idade é uma variável contínua, de razão; ela aumenta firmemente em frações minúsculas em vez de pular de categoria para categoria como

acontece com variáveis discretas, como sexo ou posto militar. Se estivermos analisando variáveis discretas — uma variável nominal ou ordinal, por exemplo —, não se aplicam algumas das técnicas descritas acima. Falando estritamente, medianas e médias devem ser calculadas apenas para dados de intervalo e de razão, respectivamente. Se, por exemplo, a variável em questão for sexo, números brutos ou marginais de porcentagens são análises adequadas e úteis. Calcular a moda é uma análise legítima, apesar de não muito reveladora, mas médias, medianas e resumos de dispersão não seriam apropriados. Algumas vezes, os pesquisadores podem aprender algo de valor violando regras como estas, mas é preciso cautela ao fazê-lo.

Descrições de Subgrupos

Análises univariadas servem para *descrever* a amostra do *survey* e, por extensão, a população da qual foi extraída. Análises bivariadas e multivariadas objetivam primariamente temas explicativos. Mas, antes de entrarmos na explicação, consideremos o caso intermediário da descrição de subgrupos.

Freqüentemente, você quer descrever subconjuntos da sua amostra. Numa análise univariada direta, você pode querer apresentar a distribuição das respostas a uma questão relacionada a direitos iguais para homens e mulheres. Na exploração detalhada das respostas, faz sentido examinar separadamente as respostas de homens e mulheres da amostra. Ao examinar as atitudes com relação à Ku Klux Klan, faz sentido descrever separadamente os respondentes brancos e negros, assim como os respondentes de diferentes regiões do país.

Ao computar e apresentar descrições estratificadas, siga os mesmos passos descritos na seção sobre análise univariada; a diferença é que os passos são dados independentemente para cada subgrupo relevante. Por exemplo, todos os homens na amostra são tratados como uma amostra total, representando 100%, e as distribuições de respostas ou as medidas de tendência central são computadas para eles. O mesmo é feito para as mulheres. Assim, você poderia informar, por exemplo, que 75% das mulheres aprovam a igualdade entre os sexos e que 63% dos homens também aprovam. Cada grupo é submetido a uma análise univariada simples. Distribuições de frequência para subgrupos são geralmente referidas como marginais *estratificados*.

Em algumas situações, você apresenta os marginais estratificados ou outras análises de subgrupo apenas por razões descritivas. O relatório de dados do censo geralmente tem esse propósito. O valor médio de unidades de moradia em diferentes bairros pode ser apresentado por razões descritivas. Assim, você pode anotar o valor médio de uma casa para qualquer bairro.

Mas a razão mais freqüente da descrição de subgrupos é comparar. No caso da igualdade sexual, você estaria interessado em determinar se as mulheres têm *probabilidade maior* que os homens de aprovar a proposta. Além disso, na maioria dos casos, comparações não são motivadas por curiosidade ociosa. Tipicamente, a comparação se baseia na expectativa de que a variável de estratificação terá alguma forma de efeito causal sobre a variável de descrição. Ser homem ou mulher afetará a atitude com relação à igualdade entre os sexos. Igualmente, ser branco ou negro afetará atitudes com relação à Ku Klux Klan. Sempre que a análise for motivada por tais expectativas, entramos no domínio da explicação em vez da descrição.

Antes de entrarmos na lógica da análise causal, bivariada, vamos considerar outro exemplo de comparações de subgrupos que nos permitirá abordar alguns assuntos de formatação de tabelas.

Fundindo Categorias de Respostas

“Exemplos de manual” de tabelas são freqüentemente mais simples do que as que em geral encontramos publicadas em relatórios de pesquisas ou em nossas próprias análises de dados; esta seção e a próxima abordam dois problemas comuns e sugerem soluções.

Começamos voltando à Tabela 14-2, que mostra os dados coletados numa pesquisa multinacional de opinião pública feita pelo *New York Times*, *CBS News* e o *Herald Tribune* em 1985, sobre atitudes com relação às Nações Unidas. A pergunta da Tabela 14-2 trata de atitudes gerais sobre como as Nações Unidas davam conta de seu trabalho.

Comparemos como as pessoas das cinco nações na Tabela 14-2 apóiam o trabalho feito pela ONU. Ao inspecionarmos a tabela, descobrimos que há números demais para permitir ver qualquer padrão significativo.

Parte do problema é a relativamente pequena porcentagem de entrevistados que selecionaram as duas categorias de

respostas extremas, trabalho *muito bom* e trabalho *muito ruim*. Embora seja tentador ler apenas a segunda linha da tabela, que diz “bom trabalho”, seria incorreto fazê-lo. Observando apenas a segunda linha, concluiríamos que a Alemanha Ocidental e os Estados Unidos foram os mais positivos (46%) sobre o desempenho das Nações Unidas, seguidos de perto pela França (45%), com a Inglaterra menos positiva (39%) do que os outros três e o Japão (11%) o menos positivo de todos.

Este procedimento não é correto, porque ignora os respondentes que deram a resposta mais positiva, “muito bom trabalho”. Numa situação como essa, devemos combinar, ou fundir, os dois extremos da amplitude de variação. Isto é, combinar o “muito bom” com o “bom” e o “muito ruim” com o “ruim”. Se você fizer isso na análise dos seus próprios dados, é aconselhável somar as frequências brutas e recomputar as porcentagens para as categorias combinadas; ao analisar uma tabela publicada como a Tabela 14-2, você pode simplesmente somar as porcentagens, como mostrado na Tabela 14-3.

TABELA 14-2

Atitude com relação às Nações Unidas: “Como a ONU está resolvendo os problemas que ela tem que enfrentar?”

	Alemanha Ocidental	Inglaterra	França	Japão	Estados Unidos
Muito bom	2 %	7 %	2 %	1 %	5 %
Bom	46	39	45	11	46
Ruim	21	28	22	43	27
Muito ruim	6	9	3	5	13
Não sei	26	17	28	41	10

FONTE: 5-Nation Survey Finds Hope for U.N. *New York Times*, p.6, 26 June 1985.

TABELA 14-3
Fundindo categorias extremas

	Alemanha Ocidental	Inglaterra	França	Japão	Estados Unidos
Trabalho bom ou melhor	48 %	46 %	47 %	12 %	51 %
Trabalho ruim ou pior	27	37	25	48	40
Não sei	26	17	28	41	10

As categorias fundidas na Tabela 14-3 nos permitem ler facilmente as porcentagens nacionais, mostrando os que dizem que a ONU está fazendo um trabalho pelo menos bom. Os Estados Unidos agora são os mais positivos; Alemanha, Inglaterra e França são pouco menos positivas e são quase indistinguíveis uma da outra; o Japão fica sozinho em sua avaliação bastante negativa. Apesar das conclusões a serem tiradas agora não diferirem radicalmente das que poderíamos ter tirado a partir da leitura da segunda linha da Tabela 14-2, vemos que a Inglaterra agora parece relativamente mais positiva.

Quero poupá-lo de um erro comum. Suponha que você tenha lido apressadamente a segunda coluna da Tabela 14-2 e notado que a Inglaterra tem uma avaliação um pouco pior do trabalho das ONU do que os Estados Unidos, a Alemanha e a França. Você pode se sentir obrigado a elaborar uma explicação para isso, talvez criando uma engenhosa teoria psico-histórica sobre o doloroso declínio da outrora digna e poderosa Inglaterra. Então, tendo alardeado sua “teoria”, alguém pode mostrar-lhe que uma leitura apropriada dos dados mostra que a Inglaterra não é menos positiva em sua avaliação do que as outras nações européias. Este *não é* um risco hipotético. Ele ocorre com frequência, mas é possível evitá-lo fundindo categorias de respostas onde for o caso.

Lidando com os “Não Sei”

As Tabelas 14-2 e 14-3 ilustram outro problema comum na análise dos dados de *survey*. Geralmente, é bom dar às pessoas a opção de dizer “não sei” ou “sem opinião”, quando se pede suas opiniões sobre certos assuntos, mas o que fazer com estas respostas ao analisar os dados?

Repare a grande amplitude de variação nas porcentagens nacionais dizendo “não sei” na Tabela 14-3 — de apenas 10% nos Estados Unidos a 41% no Japão. Porcentagens substanciais respondendo “não sei” podem confundir os resultados de uma tabela. Por exemplo, será que os japoneses têm uma probabilidade tão menor de dizer que a ONU está fazendo um bom trabalho somente porque um grande número de japoneses não expressou sua opinião?

Há um modo fácil de calcular as porcentagens excluindo os “não sei”. Observe a primeira coluna de porcentagens na Tabela 14-3 — as respostas dos alemães ocidentais. Note que 26% dos respondentes disseram que não sabiam. Isso significa que os que disseram “trabalho bom” ou “ruim”, juntos, representam 74% (100 menos 26) do total. Se dividirmos os 48% que disseram “trabalho bom ou melhor” por 0,74 (a proporção que expressou uma opinião), podemos dizer que 63% “dos com opinião” disseram que as Nações Unidas estavam fazendo um trabalho bom ou muito bom ($48\%/0,74 = 63\%$). A Tabela 14-4 apresenta toda a tabela, excluídos os “não sei”.

Repare que estes novos dados oferecem uma interpretação um pouco diferente da das tabelas anteriores. Especificamente, agora parece que a França e a Alemanha Ocidental são os mais positivos nas avaliações das Nações Unidas, com os Estados Unidos e a Inglaterra um pouco abaixo. Apesar de o Japão continuar o mais negativo, ele passou de 12% positivos para 20%.

Neste ponto, tendo visto três versões dos dados, você pode estar se perguntando qual é a versão *certa*. A resposta depende do seu propósito na análise e interpretação dos dados. Por exemplo, se não for essencial para você distinguir entre “muito bom” e “bom”, faz sentido combiná-los porque assim será mais fácil ler a tabela.

É mais difícil decidir, no abstrato, se vamos excluir ou incluir os “não sei”. O fato de uma grande porcentagem de japoneses não ter opinião pode ser uma descoberta muito importante, se você quiser descobrir se as pessoas conhecem o trabalho das Nações Unidas. Por outro lado, se você quiser saber como elas votariam numa questão, é melhor excluir os “não sei”, baseado no suposto de que elas não votariam ou dividiriam seus votos entre os dois lados da questão.

De qualquer modo, a *verdade* contida nos dados é que uma certa porcentagem disse que não sabia e que os demais dividem suas opiniões na proporção relatada. Frequentemente, é certo informar seus dados em ambas as formas — com e sem os “não sei” —, para que os leitores possam tirar as próprias conclusões.

Agora vamos ajustar ligeiramente a lógica do nosso exame das comparações de subgrupos e virar nossa discussão na direção das análises bivariadas.

TABELA 14-4
Omitindo os “não sei”

	Alemanha Ocidental	Inglaterra	França	Japão	Estados Unidos
Trabalho bom ou melhor	63 %	55 %	65 %	20 %	57 %
Trabalho ruim ou pior	36	45	35	81	44

Análise Bivariada

A análise bivariada explicativa é, basicamente, a mesma coisa que descrição de subgrupos, com certas restrições especiais. Nas descrições de subgrupos, você tem total liberdade para escolher a variável de estratificação que quiser e descrever cada subgrupo nos termos de qualquer outra variável. No exemplo da igualdade sexual, você pode descrever homens e mulheres separadamente nos termos das porcentagens de homens e mulheres que aprovam ou desaprovam (veja Tabela 14-5), ou pode descrever separadamente os que aprovam e os que desaprovam em termos das porcentagens de homens e mulheres (veja Tabela 14-6).

Qualquer uma das tabelas é uma representação legítima de descrições de subgrupos. Entretanto, os dados apresentados nas duas tabelas são lidos diferentemente. Na Tabela 14-5, vemos que 63% dos homens da amostra aprovam a igualdade sexual, comparados a 75% das mulheres. Na Tabela 14-6, vemos que, dos que concordam, 46% são homens, enquanto, dos que discordam, 60% são homens. Ou podemos notar que 37% dos

400 homens discordam, comparados a 25% de 400 mulheres (Tabela 14-5), e, dos que concordam, 54% são mulheres, enquanto, dos que discordam, 40% são mulheres (Tabela 14-6).

TABELA 14-5

"Você concorda ou discorda da proposição de que homens e mulheres devem ser tratados igualmente em todos os aspectos?"

	Homens	Mulheres
Concordam	63%	75%
Discordam	37	25
	100%	100%
	(400)*	(400)

* Os números mostrados em parênteses representam a base para a porcentagem. Neste exemplo, há 400 homens ao todo, 63% (252 homens) dos quais "concordam". 37% (148 homens) "discordam".

TABELA 14-6

Direção oposta da porcentagem

	Concordam	Discordam
Homens	46%	60%
Mulheres	54	40
	100%	100%
	100% =(552)	(248)

Mas, numa análise bivariada explicativa, só a Tabela 14-5 faria sentido. As razões são:

1. Geralmente, mulheres têm status inferior na sociedade americana; portanto, elas apoiariam mais a proposta de igualdade entre os sexos.

2. O sexo do respondente afeta sua resposta ao item do questionário; mulheres têm maior probabilidade de aprovar do que os homens.

3. Se os respondentes homens e mulheres forem descritos separadamente em termos de suas respostas, uma porcentagem maior de mulheres do que de homens aprovaria.

Seguindo esta lógica, a Tabela 14-5 divide a amostra em dois grupos, homens e mulheres, depois descreve as atitudes dos dois grupos separadamente. As porcentagens expressando aprovação nos dois grupos são comparadas, e vemos que mulheres realmente têm maior probabilidade de aprovar do que homens.

Se a Tabela 14-6 fosse apresentada como análise bivariada explicativa, a lógica da tabela seria a seguinte. As atitudes relacionadas à igualdade sexual afetam o sexo da pessoa que tem estas atitudes. Aprovar a igualdade sexual tende a tornar a pessoa mais mulher do que homem. É claro que este raciocínio é absurdo. O sexo dos respondentes é predeterminado bem antes se formarem atitudes sobre igualdade sexual. Diferentes atitudes sobre igualdade sexual não podem ter efeito no fato de a pessoa com esta atitude ser homem ou mulher.

Mas a Tabela 14-6 seria legítima do ponto de vista da descrição de subgrupos e mesmo para fins de *predição*. Se soubermos a atitude de um respondente sobre igualdade sexual e quisermos predizer se ele é homem ou mulher, a Tabela 14-6 seria uma boa base sobre a qual fundamentar tal predição. Se soubermos que ele aprova a igualdade sexual, prediremos que o respondente é mulher. (*Nota:* Se fizermos várias predições independentes como esta, erraremos 46% das vezes.) Se soubermos que o respondente desaprova a igualdade sexual, adivinharemos que ele é homem (e erraremos 40% das vezes em testes repetidos).

Porém, para propósito de explicação, só a Tabela 14-5 é legítima. Para isso, é preciso conhecer a lógica das variáveis *independentes* e *dependentes*. Basicamente, você tenta explicar valores da variável dependente baseado nos valores da variável independente. Neste sentido, você raciocina que a variável independente causa a variável dependente (tipicamente, de forma probabilística). No exemplo anterior, as atitudes sobre igualdade sexual formam a variável dependente, e os sexos dos respondentes, a variável independente. Logo, sexo causa atitudes sobre igualdade sexual.

Algumas vezes, é difícil e até arbitrário determinar qual variável é dependente e qual é independente, mas algumas diretrizes ajudam. Para começar, sempre que houver uma clara

ordem temporal relacionando as duas variáveis, a variável cujos valores são determinados antes é sempre a variável independente; aquela cujos valores são determinados depois é sempre a variável dependente. A noção de causalidade retroativa no tempo é ilógica. Como os sexos dos respondentes são determinados antes de suas atitudes sobre igualdade, sexo deve ser a variável independente.

Uma implicação disso é que duas variáveis ocorrendo simultaneamente não podem ser ligadas causalmente. O sexo e a raça de uma pessoa não podem ser analisados, *per se*, de forma explicativa. É claro que a raça dos pais e o sexo da criança poderiam ser logicamente analisados desta maneira, se você acreditar, por exemplo, que pais negros têm maior probabilidade de ter filhos homens do que pais brancos. Neste caso, poderíamos determinar uma ordem temporal das variáveis. (Contudo, parece não haver evidência lógica ou empírica apoiando esta especulação.)

Em muitos casos, nenhuma ordem temporal clara parece relacionar as duas variáveis. Por exemplo, se você quiser examinar a relação causal entre escolaridade e preconceito, a ordem temporal das duas variáveis é um pouco mais ambígua do que no exemplo do sexo e atitudes. Você pode dizer que, aumentando a escolaridade, tornamos a pessoa menos preconceituosa; então, escolaridade seria a variável independente, e preconceito, a variável dependente. Por outro lado, pode ser que o preconceito afete a quantidade de educação que uma pessoa procura ou recebe. Você pode dizer que a educação seria anátema para uma pessoa muito preconceituosa e/ou uma pessoa muito preconceituosa teria maior probabilidade de ser expulsa da escola. Portanto, pode-se justificar o argumento que faz do preconceito a variável independente e da escolaridade a variável dependente.

Em situações onde a ordem temporal das variáveis não é clara, a designação de variáveis independentes e dependentes precisa ser feita e apresentada em bases lógicas. Frequentemente, o argumento não é bastante forte para satisfazer todos os leitores. Em outras situações, podemos achar que as duas variáveis se afetam de forma cíclica. Por exemplo, para alguns respondentes, escolaridade afeta o preconceito, enquanto para outros é o contrário; para outros, pode haver uma dinâmica ainda mais complexa: uma pessoa sem preconceito pode ser levada a buscar mais educação e esta reduzir ainda mais seu preconceito.

Seja como for, toda tabela bivariada explicativa implicitamente designa uma variável independente e uma dependente. Se você achar que nenhuma ordem temporal conecta as variáveis, precisa designar arbitrariamente uma ordem quase-temporal ao construir a tabela. A seguinte discussão presume que se tenha designado uma variável como dependente, desconsiderando os fundamentos desta designação.

Construção de Tabelas

A construção de tabelas bivariadas explicativas segue estes passos:

1. A amostra é dividida em valores ou categorias da variável dependente.
2. Cada subgrupo é descrito em termos dos valores ou categorias das variáveis dependentes.
3. Finalmente, a tabela é lida comparando-se os subgrupos da variável independente em termos de um valor da variável dependente.

Lembre nossa análise anterior sobre sexo e atitudes com relação à igualdade sexual. Seguindo os passos acima, sexo sendo designado como variável independente e atitudes sobre igualdade sexual, como variável dependente, procedemos assim:

1. A amostra é dividida em homens e mulheres.
2. Cada subgrupo sexual é descrito em termos de aprovação ou desaprovação da igualdade sexual.
3. Homens e mulheres são comparados em termos das porcentagens que aprovam a igualdade sexual.

A Figura 14-2 nos dá uma representação gráfica dos passos a dar na criação de tabelas de porcentagem. O número de casos foi reduzido para simplificar.

Um problema que costuma confundir pesquisadores experientes merece um comentário. A tabela deve ser percentuada "para baixo" ou "através"? Deve uma coluna de porcentagens

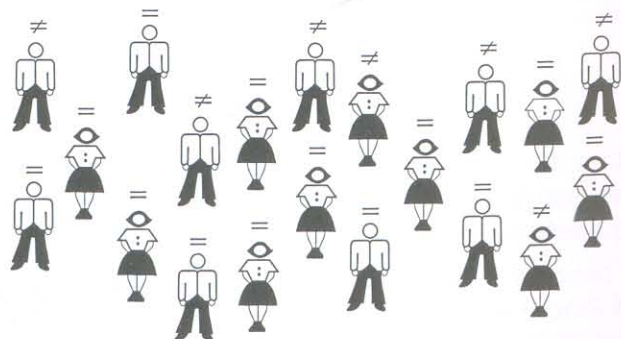
totalizar 100% ou o deve a linha de porcentagens? A resposta é completamente arbitrária. Neste livro, tendi a padronizar o procedimento calculando a porcentagem para baixo, de modo que as colunas de porcentagens totalizassem 100%, mas esta convenção é só uma questão de gosto e hábito pessoais.

Entretanto, podemos derivar dessa questão geral uma diretriz útil. Se uma tabela for percentuada para baixo, deve ser lida horizontalmente. Se é percentuada através, deve ser lida verticalmente. Usando a Tabela 14-5 como exemplo, vemos que foi percentuada para baixo, de modo que as porcentagens em cada coluna somem 100%. A tabela é interpretada lendo através: 63% dos homens aprovam, comparado a 75% das mulheres.

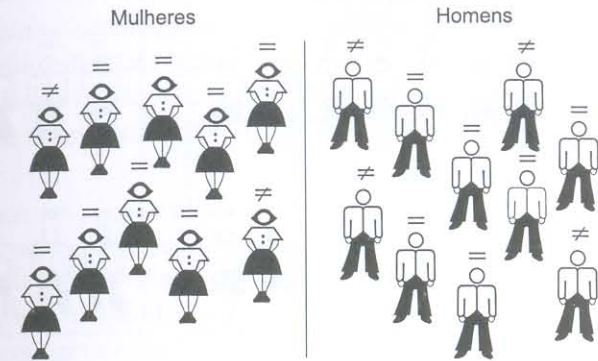
A obediência completa a esta regra de construção e interpretação de tabelas ajuda a evitar um erro comum. Muitos pesquisadores inexperientes lêem a Tabela 14-5 assim: "63% dos homens concordam com igualdade sexual enquanto 37% discordam. Portanto, os homens têm maior probabilidade de concordarem." Essa interpretação é incorreta. Embora seja verdade que os homens têm maior probabilidade de aprovar a igualdade sexual do que *desaprovar*, esse achado só tem significado como descrição das atitudes dos homens. A observação mais importante é que os homens têm uma probabilidade menor de aprovar *do que as mulheres*. Como a tabela é percentuada na vertical, deve ser lida horizontalmente.

FIGURA 14-2
Percentuando uma tabela

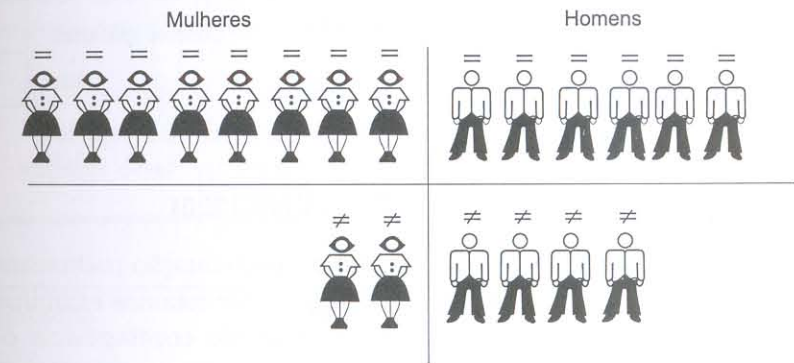
A. Alguns homens e mulheres que ou são a favor (sinal de igual) da igualdade sexual, ou são contra (≠)



B. Separar homens e mulheres (a variável independente)



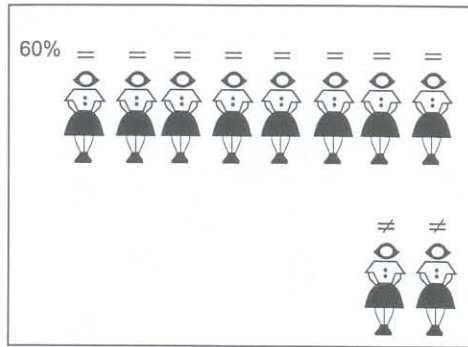
C. Em cada grupo do gênero, separar os que são a favor da igualdade sexual dos que são contra (variável dependente)



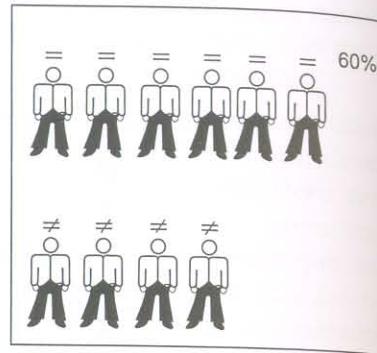
D. Conte os números em cada célula da tabela

	Mulheres	Homens
8		
2		

E. Qual a porcentagem das mulheres é a favor da igualdade?



F. Qual a porcentagem dos homens é a favor da igualdade?



G. Conclusões

Enquanto a maioria dos homens e mulheres são a favor da igualdade sexual, mulheres tem maior chances de o ser.

Portanto sexo parece ser uma das causas da atitude em relação a igualdade sexual.

	Mulheres	Homens
A favor da igualdade	80 %	60 %
Contra a igualdade	20 %	40 %
Total	100 %	100 %

Formatos de Tabelas Bivariadas

Não há formato de apresentação padronizado para o tipo de tabelas de porcentagens que estamos examinando (frequentemente chamadas tabelas de contingência ou tabulações cruzadas). Por isso, encontramos uma grande variedade de formatos na literatura de pesquisa. Se as tabelas puderem ser lidas e interpretadas facilmente, provavelmente não há razão para forçar a padronização. Contudo, certas diretrizes devem ser seguidas na apresentação da maioria dos dados tabulares. (Veja a Tabela 14-7 como exemplo de uma boa tabela.)

1. Tabelas devem ter cabeçalhos ou títulos descrevendo seu conteúdo.

2. O conteúdo original das variáveis deve ser apresentado claramente — na própria tabela, se for possível, ou no texto com uma paráfrase na tabela. Isso é especialmente crítico quando uma variável é descrita a partir de respostas a uma pergunta atitudinal, porque o significado das

respostas depende muito da construção da frase usada no item de questionário.

3. Os valores ou categorias de cada variável devem ser claramente indicados. Categorias complexas de resposta devem ser abreviadas, mas o significado deve estar claro na tabela, e a resposta completa deve ser relatada no texto.

4. Quando são apresentadas porcentagens na tabela, deve-se indicar a base sobre a qual foram computadas. É redundante apresentar todos os números brutos para cada categoria de porcentagem, já que eles podem ser reconstruídos a partir das porcentagens e das bases. Além disso, a apresentação de números e de porcentagens frequentemente torna a tabela confusa e mais difícil de ser lida.

5. Se alguns respondentes forem omitidos da tabela, devido a dados faltantes (por exemplo, "sem resposta"), seus números devem ser indicados na tabela.

TABELA 14-7

"Você aprova ou desaprova a proposição geral de que homens e mulheres devem ser tratados igualmente em todos os aspectos?"

	Homens	Mulheres
Aprovam	63%	75%
Desaprovam	37	25
	100%	100%
	(400)	(400)
Sem resposta =	(12)	(5)

Análise Multivariada

A lógica da análise multivariada é o assunto de capítulos posteriores deste livro, especialmente o Capítulo 15. Entretanto, neste ponto será útil discutir rapidamente a construção de tabelas multivariadas, isto é, tabelas construídas a partir de várias variáveis.

Tabelas multivariadas podem ser construídas com base numa descrição mais complicada de subgrupo, seguindo essencialmente os mesmos passos descritos para as tabelas bivariadas. Porém, em vez de uma variável independente e uma dependente, teremos mais de uma variável independente. Em vez de explicar a variável dependente com base numa única variável independente, buscaremos uma explicação baseada em mais de uma variável independente.

Voltando ao exemplo das atitudes sobre igualdade sexual, suponha que você acredite que idade também afeta tais atitudes, com os jovens tendo maior probabilidade de aprovar a igualdade sexual do que os mais velhos. O primeiro passo na construção de tabelas é dividir a amostra total em subgrupos baseados nos vários valores das duas variáveis independentes simultaneamente: homem jovem, homem velho, mulher jovem, mulher velha. Depois, os vários subgrupos são descritos nos termos da variável dependente, e se fazem as comparações. A Tabela 14-8 exemplifica uma tabela hipotética multivariada.

Seguindo a convenção, a tabela foi percentuada para baixo e, por isso, deve ser lida na horizontal. A interpretação leva a várias conclusões.

1. Entre homens e mulheres, os mais jovens dão mais apoio à igualdade sexual do que os mais velhos. Entre as mulheres, 90% das com menos de 30 e 60% das com 30 ou mais aprovam.
2. Dentro de cada grupo de idade, as mulheres apóiam a igualdade sexual mais que os homens. Entre os respondentes abaixo de 30 anos, 90% das mulheres apóiam, comparadas a 78% dos homens. Entre os de 30 ou mais, 60% das mulheres apóiam, contra 48% dos homens.
3. Como medido na tabela, idade parece ter maior efeito sobre atitudes do que sexo. Para homens e mulheres, o efeito da idade pode ser resumido como uma diferença de 30 pontos percentuais. Dentro de cada grupo de idade, a diferença em pontos percentuais entre homens e mulheres é de 12.
4. Idade e sexo têm efeitos independentes sobre as atitudes. Dentro de cada valor de uma variável independente, diferentes valores da segunda afetarão as atitudes.

5. Igualmente, as duas variáveis independentes têm um efeito cumulativo sobre atitude. Mulheres jovens são as que mais apóiam, e homens velhos os que apóiam menos.

TABELA 14-8
"Você aprova ou desaprova a proposição geral de que homens e mulheres devem ser tratados igualmente em todos os aspectos?"

	Mulheres		Homens	
	Abaixo de 30	30 e acima	Abaixo de 30	30 e acima
Aprovam	90%	60%	78%	48%
Desaprovam	10	40	22	52
	100%	100%	100%	100%
	(200)	(200)	(200)	(200)
Sem resposta =	(2)	(3)	(10)	(2)

O Capítulo 15 sobre o modelo de elaboração examina com mais detalhes a lógica da análise multivariada. Mas, antes de concluir esta seção, vejamos um formato alternativo para apresentar dados.

Cada tabela apresentada neste capítulo é ineficiente de algum modo. Como a variável dependente — atitude sobre igualdade sexual — é dicotômica (tem dois valores), conhecer um valor permite reconstruir o outro facilmente. Assim, se sabemos que 90% das mulheres com menos de 30 anos aprovam a igualdade sexual, sabemos automaticamente que 10% desaprovam (supondo que removemos os "não sei" e outras respostas desse tipo). Portanto, é desnecessário relatar a porcentagem dos que desaprovam. Por isso, a Tabela 14-8 poderia ser apresentada no formato alternativo da Tabela 14-9.

Nesta, as porcentagens aprovando a igualdade sexual são relatadas nas células representando as interseções das duas variáveis independentes. Os números entre parênteses abaixo de cada porcentagem representam os números de casos sobre os quais as porcentagens são baseadas. Por exemplo, sabemos que havia 200 mulheres abaixo de 30 anos na amostra e que 90%

delas aprovavam a igualdade sexual. Além disso, que 180 das 200 mulheres aprovavam e que as outras 20 (10%) desaprovavam. É mais fácil ler a Tabela 14-9 do que a 14-8, sem sacrificar qualquer detalhe.

TABELA 14-9

"Você concorda ou discorda da proposição geral de que homens e mulheres devem ser tratados igualmente em todos os aspectos?"

Porcentagem que Concorda	Mulheres	Homens
Abaixo de 30	90 (200)	78 (200)
30 e acima	60 (200)	48 (200)

Resumo

Este capítulo introduziu a lógica e a mecânica da construção de tabelas. À primeira vista, tabelas parecem muito simples para merecer discussão extensa. Na verdade, elas são bastante complexas. Por isso, são freqüentemente malconstruídas e mal-interpretadas.

Tentei apresentar a lógica da construção e interpretação de tabelas, começando com a análise univariada, passando para a descrição de subgrupos e depois discutindo a análise explicativa bivariada e a análise multivariada. Os capítulos seguintes dependem muito da compreensão das tabelas de contingência, portanto é essencial que você as entenda bem.

Vale a pena repetir os passos essenciais da construção e interpretação das tabelas explicativas.

1. Divida a amostra em subgrupos baseados nos valores da(s) variável(eis) independente(s).

2. Descreva cada grupo com base nos valores da variável dependente.

3. Compare os subgrupos da variável independente em relação ao valor da variável dependente.

Finalmente, memorize esta regra básica: percentue para baixo e leia através, ou percentue através e leia para baixo.

Nota

¹ Dados extraídos de HAUB, Carl, KENT, Mary Mederios. *1988 World Population Data Sheet* (a poster). Washington, DC: Population Reference Bureau.

Leituras Adicionais

COLE, Stephen. *The Sociological Method*. Chicago: Markham, 1972.

DAVIS, James. *Elementary Survey Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

LABOVITZ, Sanford, HAGEDORN, Robert. *Introduction to Social Research*. New York: McGraw-Hill, 1971.

ZEISEL, Hans. *Say It with Figures*. New York: Harper & Row, 1957.