

SCHUMAN, Howard, PRESSER, Stanley. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. New York: Academic Press, 1981.

SMITH, Tom W. The Art of Asking Questions. *Public Opinion Quarterly*, v.51, p.S95-S108, inverno 1987.

Capítulo 8

Construção de Índices e Escalas

Como notamos antes, muitas, se não a maioria, das pesquisas sociais visam determinar as associações entre variáveis. Geralmente, queremos estabelecer que X está relacionado a (ou causa) Y. Mas notamos também que a medição de variáveis, freqüentemente, é tarefa difícil. Normalmente, é impossível chegar a uma medida totalmente inequívoca e completamente aceitável de qualquer variável. Mesmo assim, os pesquisadores não desistem de tentar criar medidas cada vez melhores e mais úteis.

Este capítulo aborda a questão da medição. Especificamente, discute a construção de índices e escalas como medidas de variáveis. Estas medidas cumulativas são freqüentemente usadas na pesquisa social, por vários motivos. Primeiro, apesar do cuidado ao construir questionários, raramente é possível chegar a uma única pergunta que represente adequadamente uma variável complexa. Qualquer item único pode desprestatar alguns respondentes no estudo. Para medir religiosidade, talvez apenas freqüência à igreja não seja um indicador suficiente. Alguns respondentes que freqüentam regularmente a igreja podem ser considerados irreligiosos em outros termos, e alguns que nunca vão à igreja podem ser considerados religiosos. Índices e escalas combinam vários itens de questionário, evitando assim os vieses inerentes aos itens únicos.

Segundo, você pode querer usar uma medida ordinal de uma variável bem refinada — por exemplo, arranjando os respondentes em várias categorias ordinais, da mais baixa à mais alta da variável. Um único item pode não apresentar um número

suficiente de categorias de respostas para permitir esta amplitude de variação. Um índice ou escala, formados a partir de vários itens, pode.

Finalmente, índices e escalas são dispositivos *eficientes* para a análise de dados. Se um único item de questionário nos dá somente uma localização rudimentar de um respondente numa variável, vários itens podem dar uma medida mais exata e abrangente. Mas, normalmente, é impraticável considerar simultaneamente todas as respostas particulares dadas por um respondente. Índices e escalas (especialmente escalas) são dispositivos de *redução de dados*; as várias respostas de um respondente podem ser resumidas num único escore, e mesmo assim os detalhes específicos daquelas respostas serem mantidos quase que na totalidade.

Índices Versus Escalas

Os termos *índice* e *escala* são usados habitualmente de modo impreciso e intercambiável na literatura da pesquisa social. As definições técnicas originais destes termos perderam atualmente seus significados. Antes de considerarmos as distinções entre índices e escalas, vejamos o que eles têm em comum.

Ambos são, tipicamente, medidas *ordinais* de variáveis. São construídos de maneira a ordenar os respondentes do *survey* (ou outras unidades de análise) em termos de variáveis específicas, como religiosidade, alienação, situação socioeconômica, preconceito, sofisticação intelectual etc. O escore de um respondente numa escala ou índice de autoritarismo, por exemplo, dá uma indicação do seu autoritarismo relativo, quando comparado a outros respondentes.

No nosso uso dos termos, escalas e índices são medidas compostas de variáveis, ou seja, medições baseadas em respostas a mais de um item do questionário. Portanto, os escores dos respondentes num índice ou escala de religiosidade são determinadas pelas respostas específicas a vários itens do questionário, cada um dos quais fornecendo alguma indicação da religiosidade deles.

Neste livro, distinguiremos entre índices e escalas pela maneira como os escores são atribuídos aos respondentes. Um *índice* é construído pelo somatório simples de escores atribuídos a respostas específicas aos itens individuais que formam o índice. Uma *escala* é construída pela atribuição de escores a *padrões de respostas* entre os vários itens que formam a escala. Uma escala difere de um índice por tirar vantagem de qualquer possível

estrutura de intensidade entre os itens individuais. Um exemplo simples deve elucidar esta distinção.

Suponha que queiramos medir o apoio dos americanos aos direitos civis de comunistas declarados. Podemos perguntar se um comunista devia ter o direito de ter as seguintes ocupações: (1) advogado, (2) médico, (3) bibliotecário, (4) engenheiro, (5) repórter, (6) professor. Alguns respondentes estão dispostos a permitir que comunistas exerçam todas as ocupações listadas. Alguns não quererem permitir comunistas em nenhuma delas. Outros acham que algumas ocupações podem e outras não. Respondentes que dão respostas mistas presumivelmente indicam que consideram algumas ocupações mais importantes que outras. Contudo, as prioridades relativas das diferentes ocupações variam de respondente para respondente; nenhuma ordenação absoluta é inerente às ocupações em si.

Pelas respostas, você pode construir um *índice* dos compromissos relativos dos respondentes para com as liberdades civis dos comunistas, baseado no número de ocupações que eles lhes permitem exercer. Quem permitir aos comunistas exercer todas as ocupações apóia claramente um grau maior de liberdade civil que quem lhes fechar todas as portas. Além disso, você supõe que quem permitir a um comunista exercer três das ocupações apóia mais as liberdades civis dele do que quem permitir apenas uma ou duas, independente de quais forem as ocupações. Tal índice pode fornecer uma medida ordinal útil e precisa do libertarianismo civil. A parte A da Figura 8-1 ilustra a lógica simples da construção de índices.

Entretanto, suponha que as ocupações do exemplo anterior fossem: (1) escavador de valas, (2) professor de escola secundária, (3) presidente dos Estados Unidos. Neste caso, há toda razão para acreditar que esses três itens têm uma *estrutura de intensidade*. Quem permitir a um comunista ser presidente certamente o deixará ser escavador de valas ou professor secundário. Quem permitir a um comunista cavar valas pode ou não permitir-lhe exercer as outras duas ocupações. Provavelmente, sabendo *a quantidade* de ocupações que um respondente permite a comunistas, saberíamos *quais* profissões permitiria. Nessa situação, uma medida composta formada pelos três itens constituiria uma escala, tal como usamos o termo. A parte B da Figura 8-1 ilustra a construção de uma escala.

Deve estar claro que escalas geralmente são superiores a índices, se por nenhuma outra razão, porque escores de escalas dão mais informação do que escores de índices. Você deve ter cuidado com o uso errado do termo *escala*; chamar uma medida

qualquer de escala ao invés de índice não a torna melhor. Deve também estar alerta em relação a duas outras más interpretações da natureza do escalonamento. Primeiro, se a combinação de vários itens do questionário vai resultar ou não numa escala quase sempre depende da amostra específica de respondentes. Certos itens podem formar uma escala entre os respondentes de uma amostra, mas não de uma outra; não se deve presumir que um certo conjunto de itens formam uma escala só porque a formaram numa determinada amostra. Segundo, usar técnicas de escalonamento (discutidas depois neste capítulo) não assegura a criação de uma escala mais do que usar itens que previamente formaram escalas.

Um exame da literatura substantiva baseada em dados de *survey* mostra que índices são usados muito mais freqüentemente do que escalas. Mas, ironicamente, a literatura metodológica contém muito pouca, se alguma, discussão sobre construção de índices, apesar de abundarem as discussões sobre construção de escalas. Há duas razões aparentes para isto. Primeiro, índices são usados mais freqüentemente porque, em geral, é difícil ou impossível construir escalas a partir dos dados que se têm. Segundo, métodos de construção de índices não são discutidos porque parecem óbvios e diretos.

FIGURA 8-1
Índices versus escalas

Parte A. Lógica da Construção de Índices

Aqui estão várias ocupações que os americanos podem consentir ou não que um comunista declarado exerça. No geral, as diferentes ocupações provavelmente seriam consideradas como tendo mais ou menos o mesmo grau de "importância" pelos respondentes.

Para criar um índice de apoio às liberdades civis, podemos atribuir aos respondentes um ponto para cada ocupação que consentirem a um comunista exercer.

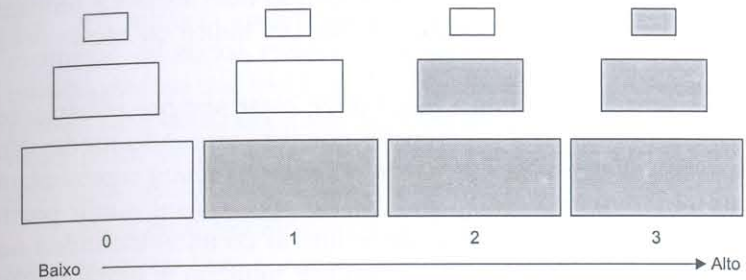
Advogado	Físico	Bibliotecário	Engenheiro	Professor	Repórter
----------	--------	---------------	------------	-----------	----------

Parte B. Lógica da Construção de Escalas

A seguir mostramos três ocupações que os respondentes provavelmente veriam como diferindo em grau de importância com relação ao seu exercício por comunistas. O tamanho das caixas representa os diferentes percentuais de aceitação com relação a um comunista exercendo tal cargo.



A seguir temos os quatro tipos de escalas que esperaríamos obter de uma análise das respostas. Uma caixa sombreada indica a aceitação da ocupação a ser exercida por comunistas.



Mas não é fácil construir índices. Acho que o fracasso generalizado em desenvolver técnicas de construção de índices resultou na criação de muitos índices ruins em pesquisa social. Pensando nisso, concentrei a maior parte deste capítulo nos métodos de construção de índices. Aprendendo a lógica desta construção, você estará melhor preparado para tentar construir escalas. Inclusive, o índice cuidadosamente construído pode acabar se tornando uma escala.

Construção de Índices

Seleção de Itens

Um índice composto é criado para medir uma variável. O primeiro critério para selecionar os itens a serem incluídos no índice é a *validade aparente* (validade lógica). Se, por exemplo, você quiser medir conservadorismo político, cada item deve aparecer para indicar o conservadorismo do respondente (ou seu oposto, o liberalismo). Filiação a partido político seria um desses itens. Se pedirmos aos respondentes para aprovarem ou desaprovarem as idéias de uma conhecida figura pública conservadora, as respostas podem logicamente fornecer outra indicação de conservadorismo. Um pesquisador construindo um índice de religiosidade pode considerar itens investigando a freqüência do respondente à igreja, aceitação de certas crenças religiosas, freqüência de orações, e assim por diante. Cada item listado parece oferecer alguma indicação da religiosidade dele.

Em geral, a literatura metodológica sobre conceitualização e medição enfatiza a necessidade de *unidimensionalidade* na construção de índices e escalas; isto é, uma medida composta

deve representar apenas uma dimensão. Assim, itens que refletem religiosidade não devem ser incluídos numa medida de conservadorismo político, mesmo que as duas variáveis tenham uma relação empírica. Neste sentido, um índice ou escala deve ser unidimensional.

Ao mesmo tempo, você deve estar sempre consciente das nuances sutis da dimensão que está tentando medir. No caso da religiosidade, os itens mencionados acima representam diferentes *tipos* de religiosidade. Se você quiser medir participação ritual em religião, deve limitar os itens incluídos no índice àqueles que especificamente indicam a participação ritual: frequência à igreja, comunhão, confissão, e semelhantes. Mas, se quiser medir religiosidade de modo mais geral, deve incluir um conjunto equilibrado de itens representando cada tipo diferente de religiosidade. Ao final, a natureza dos itens incluídos determinará quão específica ou genérica é a medida da sua variável.

Ao selecionar itens para incluir no índice, você deve considerar também a quantidade de *variância* fornecida pelos itens. Se um item fornece uma indicação de conservadorismo político, por exemplo, você deve anotar quantos conservadores são indicados pelo item. No limite, um item que indicar que ninguém é conservador ou que todos são conservadores não será muito útil para criar um índice. Se só 1% de uma amostra aprovar um político radical de direita, o item que medir isso provavelmente não será muito útil na construção de um índice.

Com relação à variância, temos duas opções. Primeiro, você pode selecionar vários itens que dividam os respondentes mais ou menos ao meio, em termos da variável. Assim, você pode selecionar vários itens, cada um indicando cerca de metade como conservadores e metade como liberais. Apesar de nenhum dos itens, por si só, justificar caracterizar um respondente como “muito conservador”, uma pessoa que parecer conservadora em todos os itens pode ser caracterizada assim. A segunda opção é selecionar itens que difiram em variância. Um item pode indicar que cerca de metade dos respondentes é conservadora, e outro pode indicar que poucos o são. (*Nota:* esta segunda opção é necessária para o escalonamento, mas é também um método razoável para construir índices).

Relações Bivariadas entre Itens

A segunda etapa na construção de índices é examinar as relações bivariadas entre os itens. Se cada item realmente der uma indicação da variável — em termos de validade aparente —,

então os itens devem estar empiricamente relacionados entre si. Por exemplo, se vários itens refletem o conservadorismo ou liberalismo de um respondente, quem aparecer como conservador num item deve também aparecer assim nos outros itens. Mas, tais itens raramente, ou nunca, estão perfeitamente relacionados entre si. (É essa disparidade que leva à necessidade de construir medidas compostas). De todo modo, quem aparecer conservador no item A tem maior probabilidade de aparecer como conservador no item B do que quem figurar como liberal no item A.

Você deve examinar todas as possíveis relações bivariadas entre os vários itens considerados para inclusão no índice, para determinar as forças relativas das relações entre os vários pares de itens. Para tal, pode usar tabelas de porcentagem, coeficientes de correlação, ou ambos. O critério primário para avaliar essas relações é a força delas. Contudo, deve-se usar este critério com sutileza, ficando alerta com itens não relacionados empiricamente entre si. É improvável que itens não relacionados entre si meçam a mesma variável. Um item não relacionado a vários outros provavelmente deve ser abandonado. Ao mesmo tempo, outro sinal de perigo é uma forte correlação entre dois itens. No limite, se dois itens têm uma correlação perfeita, só precisamos de um deles no índice, porque um carrega todas as indicações fornecidas pelo outro. (Esta questão ficará mais clara na próxima seção.)

Um exemplo da literatura substantiva sobre pesquisa de *survey* pode ajudar a ilustrar os passos na construção de índices.¹ Um *survey* com professores de uma escola de medicina estudou o efeito da “perspectiva científica” na qualidade dos cuidados que os médicos dispensam aos pacientes. O objetivo primário foi determinar se médicos com mais inclinação para a ciência eram mais impessoais no tratamento dos pacientes.

O questionário do *survey* ofereceu vários indicadores possíveis das perspectivas científicas dos respondentes. Destes, três itens pareceram — em termos de validade aparente — indicar especialmente se os médicos eram ou não cientificamente orientados. Os três itens foram:

1. Como professor da escola, em qual capacidade acha que pode dar maior contribuição acadêmica: como médico clínico ou como médico pesquisador?
2. Enquanto continua a aprimorar seus conhecimentos médicos, você diria que seus interesses médicos finais estão mais no tratamento completo do paciente ou na compreensão dos mecanismos básicos?

3. No campo da pesquisa terapêutica, em geral você está mais interessado em artigos relatando avaliações da efetividade de vários tratamentos, ou em artigos explorando os argumentos básicos subjacentes ao tratamento?

Em cada item, a segunda resposta indica maior orientação científica do que a primeira. Nas repostas a cada item, quem optar pela segunda alternativa é mais cientificamente orientado do que quem optar pela primeira. Esta conclusão *comparativa* é razoável, mas não devemos pensar que quem opta pela segunda alternativa é “cientista” em algum sentido absoluto. Ele é apenas *mais científico* do que quem opta pela primeira alternativa. Este ponto importante ficará mais claro quando examinarmos a distribuição de respostas que cada item produz.

Quanto ao primeiro item — o melhor papel acadêmico —, só um terço dos respondentes pareceu ser cientificamente orientado, dizendo que sua melhor contribuição acadêmica era como pesquisador. Mas este resultado não significa que só um terço da amostra é de “cientistas”, porque os outros dois itens sugeriram conclusões bem diferentes. Em resposta ao segundo item, o maior interesse médico, cerca de dois terços escolheram a resposta científica, indicando estarem mais interessados em aprender sobre mecanismos básicos do que sobre cuidado completo ao paciente. Em resposta ao terceiro item, preferências de leitura, aproximadamente 80% preferiram a resposta científica.

Estes três itens de questionário não podem nos dizer quantos “cientistas” há na amostra, porque nenhum deles está relacionado a um conjunto de critérios descrevendo o que constitui ser um cientista em sentido absoluto. Usar os itens para isso nos deixaria com o problema de três estimativas bem diferentes de quantos cientistas há na amostra.

Estes três itens de questionário nos dão três indicadores independentes das inclinações relativas dos respondentes quanto à ciência na medicina. Cada item separa os respondentes entre mais e menos científicos. Tendo em vista as diferentes distribuições de respostas produzidas pelos três itens, é claro que cada grupo resultante de respondentes mais ou menos científicos terá integrantes um pouco diferentes. Quem parece científico num item não o parece em outro. Mesmo assim, como cada item mede a mesma dimensão geral, devemos encontrar alguma correspondência entre os diferentes grupos. Quem parecer científico num item deve ter maior probabilidade de parecer científico na resposta a outro item do que quem parecer não-científico na resposta ao primeiro. Devemos encontrar associação ou correlação entre as respostas dadas a quaisquer dois itens.

As tabelas da Figura 8-2 permitem examinar as associações entre as respostas aos três itens.² Três tabelas bivariadas (duas variáveis) são apresentadas, mostrando a distribuição conjunta das respostas a cada par de itens. Apesar de cada item único produzir um agrupamento diferente de respondentes “científicos” e “não-científicos”, a Figura 8-2 mostra que as respostas dadas a cada item correspondem, em certo grau, às respostas dadas a cada outro item.

Um exame das três relações bivariadas da Figura 8-2 apóia a crença de que os três itens medem a mesma variável: orientação científica. Começamos observando a primeira relação bivariada da tabela. As escolhas dos professores de seus melhores papéis acadêmicos e de seus principais interesses médicos nos dão indicações de orientação científica. Quem escolheu “pesquisador”, no primeiro caso, indica maior inclinação científica do que quem respondeu “clínico”. Quem respondeu “mecanismos básicos” parece mais inclinado cientificamente do que quem respondeu “tratamento completo ao paciente”, na pergunta sobre o principal interesse. Se estes itens realmente medirem a mesma coisa, quem parece científico no primeiro item (“pesquisador”) deve parecer mais científico no segundo (“mecanismos básicos”) do que quem parece não científico no primeiro (“clínico”). Pelos dados, vemos que 87% dos “pesquisadores” parecem científicos no segundo item, em oposição a 51% dos “clínicos”. (*Nota:* O fato dos “clínicos” estarem divididos ao meio, em relação aos principais interesses médicos, é irrelevante. O relevante é que eles são menos científicos do que os “pesquisadores”, em seus interesses médicos). A força dessa relação pode ser resumida numa diferença de 36 pontos percentuais.

FIGURA 8-2
Relações bivariadas entre itens sobre orientação científica

		Melhor Papel Acadêmico	
		Clínico	Pesquisador
Principal Interesse Médico	Tratamento completo do paciente	49%	13%
	Mecanismos básicos	51%	87%
		100%	100%
		(268)	(159)

B.

		Preferências de Leitura	
		Efetividade	Argumentos
Principal Interesse Médico	Tratamento completo do paciente	68%	30%
	Mecanismos básicos	32%	70%
		100%	100%
		(78)	(349)

C.

		Preferências de Leitura	
		Efetividade	Argumentos
Melhor Papel Acadêmico	Clínico	85%	64%
	Pesquisador	15%	36%
		100%	100%
		(78)	(349)

Pode-se chegar à mesma conclusão geral em relação às outras relações bivariadas. A força da relação entre preferências de leitura e principal interesse médico pode ser resumida como diferença de 38 pontos percentuais; a força da relação entre preferências de leitura e os dois papéis acadêmicos pode ser resumida numa diferença de 21 pontos percentuais.

Inicialmente, os três itens foram selecionados com base na validade aparente; cada um pareceu dar alguma indicação das orientações científicas dos professores. Examinando as relações bivariadas entre os pares de itens, encontramos apoio para a crença inicial de que todos os itens medem basicamente a mesma coisa.

Relações Multivariadas entre Itens

Constatar uma relação bivariada, esperada, entre pares de itens sugere que eles podem ser incluídos num índice composto, mas esta justificativa não basta. O próximo passo na construção do índice é examinar as relações multivariadas entre os itens. É preciso examinar as relações simultâneas entre muitas variáveis antes de combiná-las num único índice.

A principal razão para construir um índice é desenvolver um método para classificar respondentes nos termos de alguma

variável, como conservadorismo político, religiosidade ou orientação científica. Por exemplo, um índice de conservadorismo político deve identificar respondentes que são muito conservadores, moderadamente conservadores, pouco conservadores e nada conservadores (ou, respectivamente, moderadamente liberais e muito liberais, ao invés das duas últimas categorias). As diferentes gradações em termos da variável resultam da combinação das respostas dadas aos itens do índice. Portanto, quem parece ser conservador em todos os itens é considerado muito conservador no total.

Um índice pode só fornecer gradações significativas se cada item acrescentar algo à avaliação de cada respondente. Falamos na seção anterior que não se deve incluir no mesmo índice dois itens perfeitamente relacionados entre si. Se um deles o for, incluir o outro não acrescentará nada à nossa avaliação dos respondentes. Examinar as relações multivariadas entre itens é um outro modo de eliminar material inútil, além de determinar a força combinada de um certo conjunto de itens quanto à medida da variável considerada.

Voltemos ao exemplo dos professores de medicina, para compreender melhor para que serve o exame multivariado. A Figura 8-3 mostra a relação trivariada entre os três itens. A tabela, agora, é um pouco diferente da Figura 8-2. Agora, a amostra foi dividida em quatro grupos, de acordo com (1) seu melhor papel acadêmico e (2) sua preferência de leitura. Os números entre parênteses são de respondentes em cada grupo. (Portanto, 66 professores disseram que poderiam ensinar melhor como clínicos e também disseram que preferiam artigos sobre a eficácia dos tratamentos). Para cada um dos quatro grupos, apresentamos a porcentagem de respondentes que disseram estar mais interessados nos mecanismos básicos do que no tratamento completo do paciente. (Dos 66 professores mencionados acima, 27% estão primariamente interessados nos mecanismos básicos).

FIGURA 8-3

Relações trivariadas entre os itens sobre orientação científica

	Melhor Papel Acadêmico		
	Clínico	Pesquisador	
Preferências de Leitura	Efetividade	27% (66)	58% (12)
	Argumentos	58% (219)	89% (130)

O arranjo dos quatro grupos se baseia na conclusão já tirada com relação à orientação científica. Respondentes do canto superior esquerdo da tabela são presumivelmente os menos cientificamente orientados dos quatro grupos, quanto a melhor papel acadêmico e preferências de leitura. Os no canto inferior direito da tabela são presumivelmente os mais cientificamente orientados, quanto aos dois itens.

Lembre-se de que expressar interesse primordial em “mecanismos básicos” também foi tomado como indicador de orientação científica. Como esperamos, os no canto inferior direito têm a maior probabilidade de dar esta resposta (89%), e os no canto superior esquerdo a menor probabilidade (27%). Os que deram respostas mistas em relação a papel acadêmico e preferências de leitura estão em posição intermediária no seu interesse por mecanismos básicos (58% em ambos os casos).

Esta tabela nos diz muitas coisas. Primeiro, vemos que as relações originais entre pares de itens não são afetadas significativamente pela presença de um terceiro item. Por exemplo, a relação entre papel acadêmico e principal interesse médico foi resumida como diferença de 36 pontos percentuais. Observando a Figura 8-3, vemos que, somente entre os respondentes mais interessados em artigos sobre a efetividade do tratamento, a relação entre papel acadêmico e principal interesse médico foi de 31 pontos percentuais (58% menos 27% na primeira linha), e o mesmo é verdadeiro para os mais interessados em artigos sobre argumentos básicos (89% menos 58% da segunda linha). A relação original entre papel acadêmico e principal interesse médico é essencialmente a mesma da Figura 8-2, mesmo entre os considerados científicos ou não-científicos quanto às preferências de leituras.

Podemos tirar a mesma conclusão do exame das colunas na Figura 8-3. Lembre que a relação original entre preferências de leitura e principal interesse médico foi resumida como diferença de 38 pontos percentuais. Observando apenas os “clínicos” na Figura 8-3, vemos que a relação entre os outros dois itens é agora 31 pontos percentuais. A mesma relação é encontrada entre os “pesquisadores” na segunda coluna.

A importância destas observações fica mais evidente quando consideramos o que pode ter acontecido. A tabela da Figura 8-4 apresenta dados hipotéticos para ilustrar este ponto.

FIGURA 8-4
Relação trivariada hipotética entre itens de orientação científica

	Melhor Papel Acadêmico	
	Clinico	Pesquisador
Porcentagem interessada nos mecanismos básicos	51% (66)	87% (12)
Preferências de Leitura	51% (219)	86% (130)

Estes dados hipotéticos contam uma história bem diferente da que contaram os dados verdadeiros da Figura 8-3. Neste caso, evidentemente persiste a relação original entre papel acadêmico e principal interesse médico, mesmo quando preferências de leitura são introduzidas no quadro. Em cada linha da tabela, os “pesquisadores” têm probabilidade maior do que os “clínicos” de expressar interesse pelos mecanismos básicos. Mas, observando as colunas, vemos que não há relação entre preferência de leitura e principal interesse médico. Se soubermos se os respondentes acham que podem ensinar melhor como clínicos ou como pesquisadores, conhecermos suas preferências de leitura nada acrescenta à nossa avaliação de suas orientações científicas. Se algo como a Figura 8-4 resultar de nossos dados, concluiremos que preferências de leitura não devem ser incluídas no mesmo índice de papel acadêmico, porque não contribuirá em nada para o índice composto.

Neste exemplo, estavam envolvidos apenas três itens de questionário. Se considerarmos mais itens, é necessário usar tabelas multivariadas mais complexas. Até aqui, limitamos a atenção à análise trivariada dos três itens. Lembre que o propósito desta etapa, na construção de índices, é examinar a interação simultânea dos itens, para determinar se eles podem ser incluídos no mesmo índice.

Escore de Índices

Após determinar os melhores itens para inclusão no índice, o próximo passo é atribuir escores a respostas específicas, assim criando um único índice composto a partir de vários itens. É preciso tomar duas decisões básicas sobre isso.

A primeira é decidir a amplitude desejada dos escores de índice. Uma das principais vantagens de um índice sobre

um item único é a amplitude de gradações oferecida na medição de uma variável. Como observamos antes, conservadorismo político pode ser medido desde “muito conservador” a “nada conservador” (ou “muito liberal”). Então, quão longe, nos extremos, deve se estender o índice? Como antes, esta decisão envolve a questão da variância. Quase sempre, quando se estende os possíveis extremos de um índice, menos casos são encontrados em cada extremidade. Se você tentar medir conservadorismo político no seu maior extremo, pode descobrir que quase ninguém cai nessa categoria.

Então, a primeira decisão está relacionada aos desejos conflitantes de (1) a maior amplitude de medição possível no índice e (2) um número adequado de casos em cada ponto do índice. Você tem que chegar a algum tipo de acordo entre estes desejos.

A segunda decisão concerne a atribuição de escores para cada resposta específica. Basicamente, você deve decidir se atribui peso igual ou diferentes pesos aos itens no índice. Como veremos mais tarde, a construção de escalas é bastante diferente neste aspecto, mas a atribuição de escores na construção de índices é uma questão aberta. Enquanto não há regras firmes a seguir, sugiro — e a prática tende a confirmar — que os itens devam ter pesos iguais, a menos que haja fortes razões para receberem pesos distintos. Ou seja, ponderar diferentemente deve ser justificado, enquanto pesos iguais deve ser a norma.

É claro que a decisão sobre a ponderação deve estar relacionada ao equilíbrio dos itens escolhidos. Se o índice representar a mistura de aspectos pouco diferentes de uma variável, então você deve ponderar igualmente cada um deles. Mas, em alguns casos, você pode achar que dois itens refletem essencialmente o mesmo aspecto, enquanto um terceiro reflete um aspecto diferente. Se quiser respeitar igualmente os dois aspectos no índice, você pode decidir dar ao item diferente um peso igual à combinação dos pesos dos dois itens similares. Nesse caso, você pode atribuir um escore máximo de 2 para o item diferente e escores máximos de 1 para cada um dos dois itens similares.

Embora as razões para ponderar respostas devam considerar questões como essas, normalmente você experimenta diferentes métodos de ponderação, examinando os pesos relativos atribuídos a diferentes aspectos e, ao mesmo tempo, considerando a amplitude e a distribuição dos casos. No fim, o método escolhido representa uma combinação entre essas

diferentes demandas. (Nota: Nesta atividade, como na maioria das atividades de *survey*, a decisão pode ser revista, com base num exame posterior. A validação de um índice, a ser discutida daqui a pouco, pode levá-lo a reciclar os esforços e construir um índice completamente diferente).

No exemplo do *survey* na escola de medicina, decidiu-se ponderar cada item igualmente, porque eles foram escolhidos, em parte, por representarem aspectos ligeiramente diferentes da variável orientação científica. Em cada item, os respondentes recebiam um escore de 1 quando davam resposta “científica” e 0 ao dar resposta “não-científica”. Assim, cada respondente podia receber um escore de 0, 1, 2, ou 3, dependendo do número de respostas “científicas” dadas. Este método de ponderação permitiu o que foi considerado uma amplitude útil de variação — quatro categorias de índices — e também um número suficiente de casos em cada categoria para as análises.

Trabalhando com Dados Faltantes

Em quase todo *survey*, alguns respondentes não respondem algumas perguntas (ou escolhem a opção “não sei”). Esses dados faltantes são um problema em todos os estágios de análise, mas especialmente na construção de índices. (As escalas são diferentes também neste ponto). Se alguns respondentes não responderem itens de um índice composto, você terá problemas para atribuir escores a estes entrevistados na construção do índice. Há vários métodos para lidar com este problema.

Primeiro, se relativamente poucos respondentes tiverem dados faltantes, você pode decidir excluí-los da construção e análise do índice. Os principais cuidados neste caso são se a quantidade restante para análise ainda é suficiente e se excluir estes respondentes resulta numa amostra tendenciosa ao usar o índice na análise. A possibilidade de viés pode ser examinada pela comparação, em outras variáveis relevantes, dos respondentes que seriam incluídos no índice e dos que seriam excluídos. (No exemplo da escola de medicina, esta foi a decisão tomada em relação aos dados faltantes).

Segundo, pode haver razões para tratar os dados faltantes do mesmo modo que as outras respostas dadas. Por exemplo, se o questionário pedir aos respondentes para indicar sua participação em várias atividades marcando “sim” ou “não” para cada uma, muitos entrevistados podem marcar “sim” para algumas atividades e deixar as outras em branco. Neste caso,

você pode decidir que uma resposta em branco significa um “não” e registrar os dados faltantes como se tivesse sido marcada a opção “não”.

Terceiro, uma análise cuidadosa dos dados faltantes pode permitir uma interpretação do seu significado. Ao construir uma medida de conservadorismo político, por exemplo, você pode descobrir que quem não respondeu uma certa pergunta era geralmente tão conservador, nos outros itens, quanto aqueles que deram uma resposta “conservadora” a esta pergunta. Outro exemplo: um *survey* recente, que mediu crença religiosa, descobriu que os respondentes que escolheram “não sei” para uma certa crença eram quase idênticos aos “descrentes” nas respostas relativas a outras crenças. (*Nota:* Você não deve tomar estes exemplos como guias empíricos para seus estudos, tome-os apenas como sugestões de formas para analisar os próprios dados). Sempre que a análise de dados faltantes permitir tais interpretações, você pode ponderar de acordo com estas soluções.

Quarto, você pode decidir atribuir um escore intermediário aos dados faltantes. Por exemplo, se um item for ponderado em 0, 1 e 2 para suas três possíveis respostas, você pode atribuir um escore intermediário (1) ao item para os respondentes que não responderam. (É a mesma lógica pela qual a resposta “indeciso” é muitas vezes registrada como estando entre “concordo” e “discordo”).

Quinto, você pode atribuir escores de índice proporcionalmente, baseado nas respostas dadas por um respondente. Por exemplo, suponha que seis itens são combinados num índice, com escores de 0 ou 1 para cada item. O escore máximo que um entrevistado pode receber é 6. Se um respondente responder apenas cinco itens, mas receber 5 nestes itens, poderá receber um escore proporcional de $5/5 \times 6 = 6$ no índice. Quem receber 2 nos quatro itens respondidos poderá receber um escore final de $2/4 \times 6 = 3$. Quando estes cálculos resultarem em frações, deve-se usar algum método de arredondamento, para simplificar os escores finais.

Finalmente, você pode não querer usar qualquer um destes métodos para lidar com dados faltantes, mas a análise posterior pode exigir que todos os respondentes na amostra sejam ponderados. Nesse caso, você pode atribuir escores aos dados faltantes de forma aleatória. Para um item que atribuir os escores de 0, 1 e 2, o primeiro que não respondê-lo poderá receber um escore de 1, o segundo 0 e o terceiro 2 etc. Este é o método mais conservador, do ponto de vista de análise

de pesquisas, porque você está “embaralhando as cartas” contra si mesmo. Se o índice resultante provar ser uma ferramenta poderosa na análise, você poderá concluir que seria ainda mais poderoso se todos houvessem respondido todas as questões. (Claro que, se sua meta for mostrar que o índice não está relacionado a outras variáveis, você terá embaralhado as cartas a seu favor).

A escolha de um determinado método para tratar dados faltantes depende tanto da situação da pesquisa que é impossível sugerir um único melhor método ou hierarquizá-los. Geralmente, sugiro uma análise dos que não responderam, tanto quanto a possíveis vieses na sua exclusão quanto às respostas deles a outros itens do índice. O objetivo final da análise é compreender seus dados.

Validação de Índices

Até agora, discutimos os passos da seleção e ponderação dos itens de um índice composto. Se cada passo for dado cuidadosamente, aumentamos a probabilidade do índice realmente medir a variável a que se refere. O sucesso na criação de um índice útil, contudo, ainda não foi provado; a *validação* do índice ajuda a alcançar este objetivo. A base lógica da validação é: supomos que um índice composto meça uma variável, ou seja, que os escores sucessivos de um índice ordenem os respondentes quanto àquela variável. Um índice de conservadorismo político ordena os grupos quanto ao seu conservadorismo relativo. Se um índice for bem sucedido nisso, então pessoas classificadas como relativamente conservadoras pelo índice devem aparecer como relativamente conservadoras em todos os itens do questionário (ou outros indicadores) que também reflitam orientação política. Há muitos métodos para validar um índice composto.

Análise de Item. O primeiro passo na validação de um índice é uma validação interna chamada *análise de item*. Examina-se o quanto o índice composto está relacionado (ou prediz as respostas) aos itens do questionário incluídos no próprio índice. Se o índice tiver sido construído cuidadosamente, examinando as relações bivariadas e multivariadas entre vários itens, este passo confirmará a validade do índice. Para um índice complexo com muitos itens, este passo provê um teste mais parcimonioso da contribuição independente de cada item para o índice. Se descobirmos que um item não está muito relacionado ao índice, podemos supor que outros itens do índice estão eliminando a contribuição daquele

item. Assim, o item em questão não contribuirá para a potência do índice e deverá ser excluído dele.

A análise de item é um importante primeiro teste para a validade do índice, mas não é suficiente. Se o índice medir adequadamente uma variável, deverá prever com sucesso outros indicadores dela. Para testar isso, precisamos lançar mão de itens não incluídos no índice.

Validação Externa. Pessoas classificadas como politicamente conservadoras no índice devem aparecer como conservadoras nas respostas a outros itens do questionário. Devemos compreender, é claro, que estamos falando de conservadorismo relativo; é impossível dar uma definição final absoluta do que vem a ser “conservadorismo” de maneira definitiva. Mas os respondentes classificados como os mais conservadores pelo índice devem ser os mais conservadores ao responder outras perguntas. Os classificados como menos conservadores no índice devem ser os menos conservadores nas respostas a outros itens. O ordenamento de grupos de respondentes no índice deve predizer o ordenamento deles nas respostas a outras perguntas que lidam com orientação política.

Em nosso exemplo do índice de orientação científica, várias perguntas do questionário ofereceram a possibilidade de validação adicional. A Tabela 8-1 apresenta alguns desses itens. Os itens listados permitem várias lições sobre validação de índices. Primeiro, notamos que o índice prediz fortemente as respostas aos itens de validação, no sentido em que o ordenamento das respostas científicas entre os quatro grupos é o mesmo do ordenamento fornecida pelo próprio índice.

TABELA 8-1
Validação do índice de orientação científica

	Índice da Orientação Científica			
	Baixo			Alto
	0	1	2	3
Porcentagem interessada em freqüentar palestras científicas na escola de medicina	34	42	46	65
Porcentagem que diz que os professores da faculdade devem ter experiência como médicos pesquisadores	43	60	65	89
Porcentagem que preferiria que as atividades dos professores fossem restritas apenas à pesquisa	0	8	32	66
Porcentagem envolvida em pesquisa no ano acadêmico anterior	61	76	94	99

Ao mesmo tempo, cada item dá uma descrição *diferente* da orientação científica como um todo. Por exemplo, o último item de validação indica que a grande maioria dos professores da faculdade pesquisou durante o ano anterior. Se esse fosse o único indicador de orientação científica, concluiríamos que quase todos os professores são cientificamente orientados. Mesmo assim, os professores considerados mais científicos pelo índice têm maior probabilidade de terem pesquisado do que os classificados como relativamente menos científicos. O terceiro item de validação desenha um quadro descritivo diferente. Somente uma minoria dos professores disse que preferia encargos exclusivamente de pesquisa. Ainda assim, as porcentagens desta resposta também correspondem aos escores atribuídos no índice.

Índices Ruins Versus Validadores Ruins. Um dilema que enfrentam quase todos construtores de índices é o fracasso dos itens externos em validarem o índice. Se a análise interna de item mostrar relações inconsistentes entre os itens incluídos no índice e o próprio índice, há alguma coisa errada com o índice. Entretanto, se o índice fracassar em predizer consistentemente os itens de validação externa, a conclusão a tirar é mais ambígua. Você tem que optar entre duas possibilidades: (1) o índice não mede adequadamente a variável em questão, ou (2) os itens de validação não medem adequadamente a variável e, portanto, não são um teste suficiente do índice.

Se você trabalhou longa e conscienciosamente na construção do índice, achará a segunda conclusão bastante atraente. Geralmente, você sente que incluiu os melhores indicadores da variável no índice e que os itens de validação, portanto, são indicadores de segunda classe. Mas, você deve reconhecer que o índice é, propositadamente, uma medida muito poderosa da variável; como tal, ele deve estar de alguma maneira relacionado a qualquer item que, mesmo de longe, se relacione com a variável.

Quando a validação externa fracassa, você deve reexaminar o índice antes de decidir que os itens de validação são insuficientes. Uma maneira de fazer isso é examinar as relações entre os itens de validação e os itens individuais do índice. Descobrir que alguns itens do índice se relacionam aos validadores mas outros não melhora nossa compreensão do índice construído inicialmente.

Não há “receita de bolo” para este dilema; ele é uma agonia com a qual o pesquisador sério deve aprender a conviver. Ao cabo, a sabedoria da decisão sobre o índice será determinada

pela utilidade dele nas suas análises posteriores. Talvez você decida que o índice é bom e os validadores defeituosos e, mais tarde, descubra que a variável em questão (como medida pelo índice) não está relacionada a outras variáveis das maneiras esperadas. Neste ponto, você poderá retornar à composição do índice.

Neste texto, demos considerável importância à construção de índices simples, por duas razões. Primeiro, uma revisão dos temas da literatura sobre pesquisa empírica mostra a popularidade destas medidas entre os pesquisadores de *survey*. Segundo, porque temos muito pouca, se é que alguma, discussão na literatura ou em manuais metodológicos sobre as técnicas para a construção de índices. A construção de índices "simples" talvez tenha sido vista como simples demais para merecer tal discussão, e as técnicas permaneceram parte da tradição oral da pesquisa de *survey*.

"Escalonamento" Likert

Definimos uma escala como uma medida composta construída com base numa estrutura de intensidade entre os itens da medida. Na construção de escalas, os padrões de resposta entre vários itens são ponderados, enquanto na construção de índices as respostas individuais são ponderadas e os escores independentes somados. Por esta definição, o método de medição desenvolvido por Rensis Likert, chamado *escalamento Likert*, representa uma maneira mais sistemática e refinada de construir índices; portanto, este método será discutido aqui e não nas seções sobre escalonamento.

O termo *escala Likert* é associado a um formato de pergunta freqüentemente usado nos questionários de *survey*. Basicamente, mostra-se aos respondentes uma declaração e se pergunta se eles "concordam fortemente", "concordam", "discordam" ou "discordam fortemente". Modificações na redação das categorias de resposta (por exemplo, "aprovam") podem ser feitas.

O valor particular deste formato é a ordinalidade não-ambígua das categorias de resposta. Se fosse permitido aos respondentes inventar ou escolher respostas como "concordo até certo ponto", "concordo mesmo", "concordo em quase tudo", e assim por diante, seria impossível julgar a força relativa da concordância dos vários respondentes. O formato Likert resolve facilmente este dilema.

O escalonamento Likert também se presta a um método bastante direto de construção de índices. Já que se usa categorias idênticas de resposta para os vários itens que medem

uma variável, cada um desses itens pode ser ponderado de maneira uniforme. Com cinco categorias de respostas, escores de 0 a 4 ou 1 a 5 podem ser atribuídos, levando em conta a "direção" do item (por exemplo, atribuir o escore 5 a "concordo fortemente" nos itens positivos e a "discordo fortemente" nos itens negativos). Cada respondente recebe então um escore geral, representando a soma dos escores recebidos pelas respostas aos itens individuais.

O método Likert baseia-se na suposição de que o escore geral, que resulta das respostas aos vários itens que aparentemente refletem a variável estudada, fornece uma medida razoavelmente boa da variável. Estes escores gerais não são o produto final da construção de índices; na verdade, são usados para se fazer uma *análise de itens* levando à escolha dos melhores itens. Essencialmente, cada item individual é correlacionado à grande medida composta. Presume-se que os itens que melhor se correlacionam com a medida composta fornecem os melhores indicadores da variável, e somente estes itens seriam incluídos no índice que será finalmente usado para as análises da variável.

Note que a atribuição uniforme de escores nas categorias de resposta de item-Likert presume que cada item tenha aproximadamente a mesma intensidade dos outros. Esta é o aspecto chave no qual o método Likert difere do escalonamento, como usamos o termo neste livro. Os itens tipo-Likert podem ser usados de várias maneiras e não estamos presos ao método descrito. Eles podem ser combinados com outros tipos de itens na construção de índices simples; igualmente, podem ser usados na construção de escalas. Mas, se todos os itens cuja inclusão for considerada para uma medida composta forem do formato Likert, deve-se considerar o método descrito antes.

Voltemos agora a atenção para as técnicas de escalonamento. Há muitos métodos à disposição do pesquisador de *survey*; consideraremos apenas as escalas Bogardus, Thurstone e Guttman.

Construção de Escalas

Bons índices fazem uma classificação ordinal dos respondentes numa variável. Todos os índices se baseiam na suposição de que uma pessoa com duas indicações de ter inclinação científica, por exemplo, deve ser mais científica do que uma pessoa com apenas uma indicação. No entanto, um índice

pode falhar em considerar que nem todas as indicações de uma variável são igualmente importantes. (É claro que se pode tentar resolver este problema pesando os indicadores diferentemente).

As escalas dão mais segurança com relação à ordinalidade, detectando *estruturas* entre os indicadores. Os itens de uma medida composta podem ter diferentes *intensidades* em termos da variável. As três descrições seguintes de procedimentos de escalonamento ilustram a variedade de técnicas disponíveis.

Escala de Distância Social de Bogardus

Um bom exemplo de escala é a *Escala de Distância Social de Bogardus*. Suponha que você queira saber o quanto os respondentes estão dispostos a se relacionar com albaneses. Pode fazer as seguintes perguntas a eles.

1. Você deixaria albaneses morarem no seu país?
2. Você deixaria albaneses morarem na sua comunidade?
3. Você deixaria albaneses morarem na sua vizinhança?
4. Você deixaria um albanês morar ao lado de sua casa?
5. Você deixaria seu filho (sua filha) se casar com um albanês?

As perguntas estão em ordem crescente de proximidade do contato que os respondentes podem querer ou não com albaneses. Começando com a intenção original de medir a disposição de se relacionar com albaneses, desenvolvemos várias questões indicando diferentes graus de intensidade nesta variável.

As claras diferenças de intensidade sugerem uma estrutura entre os itens. Presumivelmente, respondentes dispostos a aceitar uma espécie de associação estariam dispostos a aceitar todas as associações precedentes na lista, isto é, as com menores intensidades. Por exemplo, quem deixar albaneses morar na vizinhança certamente os aceitará na sua comunidade e na sua nação, mas pode ou não aceitá-los como vizinhos próximos ou parentes. Esta é a estrutura lógica da intensidade inerente aos itens.

Empiricamente, esperamos encontrar que a maioria dos respondentes aceita a co-cidadania e a minoria aceita o casamento. Falamos, nesse sentido, de “itens fáceis” (co-cidadania) e “itens difíceis” (casamento). Mais respondentes concordam

com os itens fáceis do que com os difíceis. Com algumas exceções inevitáveis, a lógica exige que, quando os respondentes recusam uma associação apresentada na escala, também recusam todas as mais difíceis que se seguem.

A Escala de Distância Social de Bogardus ilustra a economia de escalonamento como um dispositivo de redução de dados. Se soubermos *quantas* associações com albaneses um respondente aceitaria, saberemos *quais* associações aceitaria. Um único número pode resumir precisamente cinco ou seis respostas de *survey* sem perda de informação.

Escalas Thurstone

Muitas vezes, a estrutura inerente da Escala de Distância Social de Bogardus não é apropriada para a variável sendo medida. Na verdade, esta estrutura lógica entre vários indicadores raramente é aparente. O *escalonamento de Thurstone* procura desenvolver um formato para gerar grupos de indicadores de uma variável com pelo menos uma estrutura *empírica* entre eles. Um formato básico é o de “intervalos aparentemente iguais”.

Dá-se a um grupo de “juízes” talvez cem itens, que se acredita serem indicadores de alguma variável. Pede-se a cada um para estimar com que força cada item é indicador da variável, atribuindo escores de, por exemplo, 1 a 13. Se a variável for preconceito, pede-se aos juízes para marcar 1 para os indicadores mais fracos de preconceito, 13 para os mais fortes e números intermediários para os que estiverem entre eles.

Depois de os juízes terminarem a tarefa, examinam-se os escores que atribuíram a cada item, para determinar quais itens produziram mais acordo entre eles. Os itens sobre os quais eles mais discordaram são rejeitados como ambíguos. Seleciona-se, entre os itens que produziram acordo geral, um ou mais para representar cada escore da escala de 1 a 13.

Os itens assim selecionados são incluídos no questionário. Espera-se que quem parecer preconceituoso em itens com força 5 parece preconceituoso nos itens com menor força; também não se espera que quem não parecer preconceituoso nos itens com força 6 não o parece em itens com maiores intensidades.

Se os itens da Escala Thurstone forem adequadamente desenvolvidos e ponderados, aparecerá a economia e a eficácia da redução dos dados inerentes à Escala de Distância Social de Bogardus. Um único escore pode ser atribuído a cada respondente (aceita a força do item mais difícil), e este escore representaria adequadamente as respostas a vários itens do questionário.

Como acontece com a escala de Bogardus, um respondente com escore 6 pode ser considerado mais preconceituoso do que um que recebeu escore 5 ou menos.

O escalonamento de Thurstone é raramente usado em pesquisas de *survey* hoje, principalmente devido ao enorme gasto de energia requerido no “juízo” dos itens. Vários (talvez dez ou quinze) juízes têm que gastar um tempo considerável atribuindo escores aos muitos itens iniciais. Já que a qualidade do julgamento depende da experiência com e do conhecimento da variável considerada, a tarefa pode requerer pesquisadores profissionais. Além do mais, os significados dos vários itens que indicam uma certa variável tendem a mudar com o tempo. Portanto, um item que recebe um certo peso num momento pode ter um peso bem diferente mais tarde. Uma Escala Thurstone teria que ser periodicamente atualizada para ser efetiva.

Escalonamento de Guttman

Uma técnica de escalonamento muito popular hoje foi desenvolvida por Louis Guttman. Como as escalas de Bogardus e Thurstone, o *escalonamento de Guttman* baseia-se no fato de que alguns itens podem ser indicadores mais “difíceis” da variável do que outros. Quem aceita um item difícil também aceita os mais fáceis. Se tal estrutura aparecer nos dados examinados, podemos dizer que os itens formam um Escala de Guttman. Um exemplo deve ser suficiente.

No exemplo anterior descrevendo a medição da orientação científica entre professores de uma escola de medicina, foi construído um índice simples. Como veremos, os três itens do índice formam essencialmente uma Escala de Guttman. Esta possibilidade aparece quando procuramos indicadores relativamente “difíceis” e “fáceis” da orientação científica.

O item que perguntava aos respondentes se eles serviriam melhor como clínicos ou pesquisadores é o mais difícil dos três; se este fosse o único indicador da variável, só cerca de um terço deles seria considerado científico. Se o item sobre principal interesse médico (tratamento completo ao paciente versus mecanismos básicos) fosse usado como único indicador, quase dois terços seriam considerados científicos. A preferência de leitura (efetividade do tratamento versus argumentos) é o mais fácil; cerca de 80% dos entrevistados seriam considerados científicos neste item.

TABELA 8-2
Escalonamento da orientação científica

	Preferência de Leitura	Principal Interesse	Papel Acadêmico	Número de Casos
Tipos de escalas:	+	+	+	116
	+	+	-	127
Total = 383	+	-	-	92
	-	-	-	48
Tipos mistos:	-	+	-	18
	+	-	+	14
Total = 44	-	-	+	5
	-	+	+	7

Para determinar se há uma estrutura escalar entre as respostas aos três itens, devemos examinar os vários padrões de respostas possíveis aos três itens simultaneamente. Na Tabela 8-2, todos os padrões possíveis são apresentados de forma esquemática. Para cada um dos três itens, sinais positivos e negativos foram usados para indicar respostas científicas e não-científicas, respectivamente. (“Mais” indica resposta científica, “menos” resposta não-científica.)

Os primeiros quatro padrões de resposta da tabela formam o que chamamos *tipos de escala* — os padrões que formam uma estrutura escalar. Seguindo os respondentes que selecionaram todas as três respostas científicas (linha 1), vemos que os com apenas duas respostas científicas (linha 2) optaram pelas duas mais fáceis; aqueles com apenas uma resposta científica (linha 3) optaram pela mais fácil das três. Finalmente, os entrevistados que não optaram por respostas científicas estão na linha 4.

A segunda parte da tabela apresenta os padrões de respostas que violam a estrutura escalar dos itens. Os dois últimos padrões de resposta são os que saem mais da estrutura escalar: os que só aceitaram o item mais difícil e os que rejeitaram apenas o mais fácil.

A última coluna da tabela indica o número de respondentes no *survey*, em cada padrão de resposta. É imediatamente evidente que a grande maioria (90%) deles se encaixa num dos tipos de escala. Mas a existência de tipos mistos indica que os itens não formam uma Escala de Guttman perfeita.

Neste ponto, precisamos lembrar que uma das principais funções do escalonamento é a redução eficiente dos dados. As escalas provêm uma técnica de apresentação sinóptica dos

dados e, ao mesmo tempo, mantêm o máximo possível da informação original.

Quando os itens de orientação científica formaram um índice, em nossa discussão anterior, os respondentes receberam um ponto no índice para cada resposta científica que forneciam. Se os mesmos três itens forem ponderados como Escala de Guttman, alguns respondentes receberiam escores diferentes dos recebidos no índice. Eles receberiam escores de escala que permitiriam uma reprodução mais exata das respostas originais a todos os três itens.

Quem se encaixar nos tipos de escala recebe os mesmos escores que recebe na construção do índice. Por exemplo, as pessoas que dão todas as três respostas científicas recebem escore 3. Note que, se soubermos que alguém neste grupo recebeu um escore 3, podemos precisamente prever que ele escolheu todas as três respostas científicas. Para as pessoas na segunda fileira da tabela, a atribuição do escore de escala 2 nos levaria a prever exatamente respostas científicas aos dois itens mais fáceis e resposta não-científica ao mais difícil. Em cada um dos quatro tipos de escala, podemos prever exatamente as respostas reais dadas por todos os respondentes.

Entretanto, os tipos mistos na tabela são um problema. O primeiro tipo misto (- + -) recebeu escore 1 no índice para indicar apenas uma resposta científica. Se lhe for atribuído o escore de escala 1, porém, prediríamos que todo mundo neste grupo escolheu apenas o item mais fácil (+ - -), fazendo em consequência dois erros por respondente. Os escores de escala são atribuídos, portanto, para reduzir os erros na reconstrução das respostas originais. A Tabela 8-3 ilustra os escores de índice e de escala que seriam atribuídos a cada padrão de resposta no nosso exemplo.

Como mencionei antes, a ponderação original do índice para os quatro tipos de escala seria mantida na construção da Escala de Guttman e não seriam cometidos erros na reprodução das respostas dadas aos três itens. Os tipos mistos seriam ponderados diferentemente, na tentativa de reduzir os erros. Mas note que um erro é cometido para cada respondente nos tipos mistos. No primeiro tipo misto, prediríamos erroneamente uma resposta científica ao item mais fácil para cada um dos dezoito entrevistados deste grupo, com um total de dezoito erros.

TABELA 8-3
Escore de índice e escala

	Padrões de Respostas			Número de Cartões	Escore de Índice	Escore de Escala	Total de Erros de Escala
Tipos de Escalas:	+	+	+	116	3	3	0
	+	+	-	127	2	2	0
	+	-	-	92	1	1	0
	-	-	-	48	0	0	0
Tipos Mistos:	-	+	-	18	1	2	18
	+	-	+	14	2	3	14
	-	-	+	5	1	0	5
	-	+	+	7	2	3	7

* Nota- Esta tabela apresenta um método comum para se ponderar "tipos mistos", mas o leitor deve saber que outros métodos também são usados.

O grau com que um conjunto de respostas empíricas forma uma Escala de Guttman é determinado pela exatidão com que as respostas originais podem ser reconstruídas a partir dos escores de escala. Para cada um dos 427 respondentes do exemplo, prediremos três respostas, com um total de 1.281 previsões. A Tabela 8-3 indica que faremos quarenta e quatro erros usando os escores de escala atribuídos. A porcentagem de previsões corretas é o *coeficiente de reproducibilidade*, ou seja, a porcentagem de respostas "reproduzíveis". No exemplo, o coeficiente de reproducibilidade é de 1.237/ 1.281, ou seja, 96,6%.

Exceto para o caso de reproducibilidade perfeita (100%), não temos como afirmar, de forma absoluta, se um grupo de itens forma ou não uma Escala de Guttman. Virtualmente, todos os conjuntos de itens desse tipo *aproximam-se* de uma escala. Como regra geral, coeficientes de 90% ou 95% são os padrões comumente usados para determinar se há ou não uma Escala de Guttman. Se a reproducibilidade observada exceder o nível que você estabeleceu, provavelmente você decidirá ponderar e usar os itens como uma escala.³

Um comentário final sobre a Escala de Guttman. Ela é baseada na estrutura observada entre os *dados reais que estão sob exame*. Este ponto importante é freqüentemente mal compreendido pelos pesquisadores. Não faz sentido dizer que um conjunto de itens de questionário (talvez desenvolvido e usado anteriormente por outro pesquisador) constitua uma Escala de Guttman. Podemos apenas dizer que os itens formam uma escala dentro do corpo de dados que está sendo analisado.

Portanto, a escalabilidade é uma questão empírica, dependente da amostra. Se um grupo de itens formar uma Escala de Guttman numa amostra, não há garantia de que formarão tal escala entre os respondentes de outra amostra. Então, itens de questionário em si, e por si sós, nunca formam uma escala, mas sim um conjunto de observações empíricas.

Tipologias

Concluimos este capítulo com uma breve discussão da análise e construção de *tipologias*. Lembre que índices e escalas são construídos para fornecer medidas ordinais de variáveis. Tentamos atribuir escores de índice ou escala aos respondentes, de modo a indicar um grau crescente de preconceito, religiosidade, conservadorismo, e assim por diante. Nestes casos, estamos lidando com dimensões únicas. Em contraste, tipologias são *multidimensionais*.

Muitas vezes você quer resumir as interseções entre duas ou mais dimensões. Por exemplo, examinando orientações políticas separadamente, em termos de temas domésticos e de política externa.

TABELA 8-4
Uma tipologia política

Atitudes para com a Política Doméstica	Atitudes para com a Política Externa	
	Conservador	Liberal
Conservador	A	B
Liberal	C	D

A apresentação 2 x 2 da Tabela 8-4 descreve esta tipologia. As pessoas na célula A da tabela são conservadoras em política externa e política doméstica; as na célula D são liberais em ambas. Os respondentes nas células B e C são conservadores numa questão e liberais na outra.

Freqüentemente, ao tentar construir um índice ou uma escala, você chega a uma tipologia. Os itens que você achou que representavam uma única variável parecem representar duas. Nesse exemplo, você podia estar tentando construir um índice único das atitudes políticas, mas descobriu —

empiricamente — que as políticas externa e doméstica tinham de ser mantidas separadas.

Você deve estar consciente da dificuldade inerente à análise tipológica. Sempre que a tipologia for usada como a *variável independente*, provavelmente não haverá problemas. No exemplo acima, você pode computar e apresentar as porcentagens de pessoas em cada célula que dizem que normalmente votam no Partido Democrata. Assim, você pode facilmente examinar os efeitos das atitudes em relação às políticas doméstica e externa no comportamento eleitoral.

Mas é difícil analisar a tipologia como *variável dependente*. Se quiser descobrir por que os entrevistados caem nas diferentes células da tipologia, você terá problemas. Isto fica aparente ao considerarmos as maneiras pelas quais se pode construir e ler as tabelas. Por exemplo, suponha que você queira examinar os efeitos do sexo sobre atitudes políticas. Com uma dimensão única, você poderia facilmente determinar as porcentagens de homens e mulheres classificados como conservadores e liberais no seu índice ou escala. Entretanto, com uma tipologia você teria que apresentar a distribuição dos homens em sua amostra entre os tipos A, B, C e D. A seguir, teria que repetir o procedimento para as mulheres e comparar as duas distribuições. Suponha que 80% das mulheres sejam classificadas como tipo A (conservadoras nas duas dimensões), comparados a 30% dos homens. Além disso, suponha que apenas 5% das mulheres sejam classificadas como tipo B (conservadoras apenas em relação às questões domésticas), comparados a 40% dos homens. Seria incorreto concluir, de um exame do tipo B, que os homens são mais conservadores em assuntos domésticos do que as mulheres, já que 85% das mulheres, comparados a 70% dos homens, têm essa característica. A escassez relativa de mulheres classificadas como tipo B se deve à sua concentração no tipo A. Vê-se que é muito difícil uma interpretação de tais dados em áreas não descritivas. Você provavelmente acabará examinando as duas dimensões políticas separadamente, especialmente se as variáveis tiverem mais categorias de respostas do que no nosso exemplo.

TABELA 8-5
Nove estilos de vida americanos

Nome	Descrição Típica
Estilo de vida sobrevivente	Muito pobre; baixa educação; muitos idosos; muitos doentes.
Estilo de vida sustentador	À beira da pobreza; freqüentemente encontrado nas favelas urbanas.
Estilo de vida pertencente	Classe média tradicional
Estilo de vida emulador	Lutando para crescer financeiramente; ambicioso; geralmente jovem.
Estilo de vida conquistador	Bem-sucedido na profissão e nos negócios; a imagem do sucesso americano.
Estilo de vida eu-sou-eu	Filhos de conquistadores, renunciando ao sucesso tradicional em busca de um novo estilo de vida.
Estilo de vida experimental	Jovem, educado, bem-sucedido, pessoas mais ou menos místicas que dão grande importância a questões internas.
Estilo de vida socialmente Consciente	Pessoas bem-sucedidas, maduras, influentes, preocupadas e ativas nas áreas de conscientização e responsabilidade social.
Estilo de vida integrado	Pessoas psicologicamente maduras que possuem preocupações internas e externas integradas.

FONTE - Adaptado de MITCHELL, Arnold. *The Nine American Lifestyles*. New York: Warner Books, 1983. p.4-24.

Uma tipologia mais complexa, desenvolvida por Arnold Mitchell e colegas do Instituto de Pesquisas de Stanford, a *tipologia Valores e Estilos de Vida* (VAEV), tem atraído bastante atenção, especialmente nos círculos de *marketing*.⁴ Se os americanos podem ser agrupados ou separados em termos de muitas variáveis (sexo, idade, raça, educação etc.), a pesquisa de Mitchell sobre várias variáveis demográficas e atitudes sugeriu a possibilidade de identificar a maioria deles com um dos nove estilos de vida apresentados na Tabela 8-5.

Como Mitchell e outros demonstraram, as pessoas identificadas com os nove principais estilos de vida têm padrões de consumo bem diferentes, sugerindo que um determinado produto pode ser anunciado de maneira efetiva entre alguns estilos de vida e não entre os outros. Vendas de Cadillacs provavelmente não serão altas entre os sobreviventes, nem as vendas de rifles entre os socialmente conscientes. Além disso, os pesquisadores que usaram o modelo VAEV puderam identificar áreas geográficas onde se concentram certos estilos

de vida, tornando possível, por exemplo, usar mala direta, anunciando apenas nos CEPs com maior probabilidade de vendas para um determinado produto.

Apesar das tipologias apresentarem problemas especiais para a análise de *survey*, o modelo VAEV demonstra que elas não devem ser totalmente abandonadas. É claro que, em última instância, você deve projetar e construir medidas compostas — índices, escalas ou tipologias —, apropriadas às necessidades da sua pesquisa.

Resumo

Este capítulo abordou a lógica e a construção de índices e escalas. As técnicas descritas são comuns na análise de *survey* e extremamente valiosas. Boas medidas compostas como índices e escalas oferecem as seguintes vantagens:

1. Onde indicadores únicos (por exemplo, respostas a um item de questionário) podem produzir uma medida tendenciosa da variável em questão, uma medida composta a partir de vários indicadores diferentes pode eliminar o efeito de viés.
2. Medidas compostas permitem maior amplitude de variação da variável. Enquanto um único item dicotômico fornece apenas dois níveis de intensidade (por exemplo, alto e baixo), a combinação de cinco itens resulta na criação de um índice ou escala variando de um baixo 0 até um alto 5. Uma medida composta adequadamente construída dá maior poder explicativo à análise.
3. Escalas podem prover uma eficiente técnica para reduzir dados. Um conjunto extenso de respostas de questionário pode ser resumido na forma de um único escore de escala, sem perder muito da informação original.
4. Tipologias fornecem classificações multidimensionais, que freqüentemente são eficazes como variáveis independentes, mas problemáticas enquanto variáveis dependentes.

Notas

¹ O exemplo, inclusive as tabelas, foi tirado de BABBIE, Earl. *Science and Morality in Medicine*. Berkeley: University of California Press, 1970.

² Se você tiver dificuldade para entender as tabelas apresentadas neste capítulo, consulte o Capítulo 14, que trata da construção e interpretação de tabelas.

³ A decisão sobre os critérios a usar a este respeito é, claro, arbitrária. Um alto grau de reproducibilidade não assegura que a escala construída de fato meça o conceito em questão, embora aumente a confiança de que todos os itens medem a mesma coisa. Um alto coeficiente de reproducibilidade é mais provável quando menos itens estiverem envolvidos.

⁴ MITCHELL, Arnold. *The Nine American Lifestyles*. New York: Warner Books, 1983.

Leituras Adicionais

LAZARFELD Paul F., ROSENBERG, Morris (Ed.). *The Language of Social Research*. New York: Free Press, 1955. seção 1.

OPPENHEIM, A. N. *Questionnaire Design and Attitude Measurement*. New York: Basic Books, 1966.

SELLTIZ, Claire et al. *Research Methods in Social Relations*. New York: Holt, Rinehart & Winston, 1959. cap.10.

Coleta de Dados

3

CAPÍTULO 9

QUESTIONÁRIOS AUTO-ADMINISTRADOS

CAPÍTULO 10

SURVEYS POR ENTREVISTAS

CAPÍTULO 11

PROCESSAMENTO DOS DADOS

CAPÍTULO 12

PRÉ-TESTES E ESTUDOS-PILOTO

Os quatro capítulos da Parte 3 abordam os vários aspectos da coleta de dados de *survey*. Examinaremos as atividades que resultam num corpo de dados para análise e consideraremos as decisões que você precisa tomar em relação às opções disponíveis.

Os Capítulos 9 e 10 focalizam as técnicas de coleta de dados usadas em questionários auto-administrados e *surveys* por entrevistas, respectivamente. O Capítulo 10 considera as entrevistas cara a cara e as feitas por telefone.

O Capítulo 11 trata dos diferentes métodos usados na conversão dos questionários completos em dados quantitativos manipuláveis. Também estudaremos alguns métodos novos, computadorizados, que criam arquivos de dados à medida que vão sendo coletados.

O Capítulo 12 considera um componente da pesquisa de *survey* profissional freqüentemente esquecido: estudos-piloto e pré-testes. Discutiremos as diferentes técnicas que você pode usar para testar os vários aspectos do desenho de pesquisa, antes de comprometer muitos recursos, e investigaremos como você pode avaliar os resultados desses testes.