

# Capítulo 7

## Conceituação e Desenho de Instrumentos

A pesquisa científica tem duas metas principais: *descrição* e *explicação*. Pesquisadores medem as distribuições empíricas de valores nas variáveis (descrição) e as associações entre as variáveis para explicar essas distribuições de valores. O Capítulo 13 discute as interconexões lógicas entre descrição e explicação. A finalidade deste capítulo é fundamentar estes processos, examinando a lógica e as possibilidades da medição descritiva.

Seja trabalhando a partir de uma teoria rigorosamente deduzida, seja a partir de um conjunto de suspeitas e palpites, em algum ponto você se depara com conceitos abstratos, não especificados, que acha que o ajudarão a compreender o mundo que o cerca. Na pesquisa de *survey*, estes conceitos devem ser convertidos em perguntas num questionário, de forma a permitir a coleta de dados empíricos relevantes para a análise.

O capítulo começa com notas gerais relativas à lógica da conceituação e da operacionalização. A seguir, considera os diferentes tipos de dados que se pode levantar num *survey*. Finalmente, se debruça sobre algumas técnicas para construir boas perguntas.

### Lógica da Conceituação

Muitas vezes se quer estudar conceitos como classe social, alienação, preconceito, sofisticação intelectual, e assim por diante. Talvez você suspeite que a alienação decresce com a mobilidade social ascendente, ou que a sofisticação intelectual

reduz o preconceito. No entanto, antes de embarcar na pesquisa empírica, estes conceitos podem ser apenas idéias gerais na sua cabeça, sem definição precisa do que significam para você.

Por exemplo, o conceito "status social" é freqüentemente usado na pesquisa social, mas de forma alguma está claro seu significado último. Definições diferentes de classe social incluem: renda, prestígio ocupacional, educação, riqueza, poder, status familiar tradicional e avaliação moral. Provavelmente, nenhuma combinação específica destes elementos produz uma definição de status social plenamente satisfatória para algum pesquisador, muito menos para todos os pesquisadores.

Muitas vezes se diz que tais conceitos não especificados têm "riqueza de significado", no sentido de combinar uma variedade de elementos, sumariando um fenômeno complexo. Embora o termo "status social" evoque diferentes imagens para pesquisadores diferentes, estas imagens provavelmente parecem a todos eles importantes e significativas. Tais conceitos gerais devem ser especificados para a pesquisa empírica rigorosa, ou seja, reduzidos a indicadores empíricos específicos. A operacionalização de conceitos é inevitavelmente insatisfatória, tanto para os pesquisadores quanto para suas audiências. Em última análise, conceitos ricos em significado devem ser reduzidos a indicadores empíricos simplificados e superficiais.

Penso que a maioria dos conceitos interessantes para os pesquisadores sociais *não tem significado real*, nem definição última. Status social é um bom exemplo. O conceito geral de status social seguramente é antigo. Embora não o tenham verbalizado em termos abstratos, os primeiros humanos reconheceram que alguns deles tinham status social maior do que outros. Alguns eram mais poderosos, alguns eram mais respeitados, e outros recebiam maior autoridade sobre seus pares. O reconhecimento destas diferenças persistiu no decorrer da história humana. Max Weber extraiu muitas implicações dessas diferenças. Karl Marx derivou uma teoria da sociedade com base nelas. W. Lloyd Warner e outros sociólogos americanos dos anos 30 e 40 observaram as mesmas diferenças entre moradores de pequenas cidades, e descobriram que membros de uma comunidade podiam identificar os vizinhos com maior ou menor status social.

A existência de diferenças de status entre membros de uma sociedade é clara. Acemais, tais diferenças parecem importantes para compreender outros aspectos da sociedade. Em vista disso, o termo "status social" tem sido aceito como uma anotação resumida do fenômeno. Mas o que realmente significa o termo "status social"? Uma vez que é apenas um termo para denotar um conceito geral, não tem significado

último. Por si, status social só existe como uma assinalação conveniente para uma diversidade de observações empíricas. Deste ponto de vista, nenhum pesquisador pode pesquisar status social correta ou incorretamente; pode apenas fazer medições mais ou menos úteis.

Neste sentido, os cientistas nunca coletam dados, eles *criam dados*. Esta idéia é essencialmente o mesmo sentimento expresso pelo filósofo da ciência Alfred North Whitehead, quando escreveu o seguinte:

A natureza recebe crédito pelo que deveria estar reservado para nós; a rosa pelo seu perfume; o rouxinol pelo seu canto; e o Sol pela sua radiação. Os poetas estão totalmente enganados. Deviam endereçar sua lírica para si próprios, fazendo dela odes de autocongratulação pela excelência da mente humana. A natureza é assunto sem graça, sem som, sem cor; meramente o apressar da matéria, de forma interminável e sem significado.<sup>1</sup>

Quando faz várias perguntas e combina as respostas num índice que chama de classe social, você *criou* uma medida de classe social; você realmente *criou* um agrupamento e uma hierarquia de classe social entre seus respondentes. Você não descobriu, apenas, um agrupamento e uma hierarquia que já existia, num sentido real.

Talvez as medidas criadas tenham utilidade, ajudando-o a entender os dados que tem, ou ajudando-o a desenvolver teorias sobre classes sociais. Entretanto, não tem sentido perguntar se você *realmente* mediu classe social, num sentido válido em última análise, porque o conceito só existe em nossas mentes.

A noção de que pesquisadores criam dados e não os coletam é até mais fundamental do que sugere o exemplo do status social. Nos primeiros dias da II Guerra Mundial, Hadley Cantril fez dois *surveys* nacionais do povo americano.<sup>2</sup> O desenho das amostras e a checagem cruzada das suas características indicavam que ambos forneciam boas estimativas da população americana. Num dos *surveys*, foi perguntado aos respondentes: "Você acha que os Estados Unidos conseguirão ficar fora da guerra?" A maioria (55% dos que opinaram) disse sim. No outro *survey*, a pergunta foi: "Você acha que os Estados Unidos entrarão na guerra antes que ela termine?" A maioria (59% dos que opinaram) disse sim.

Estes itens são muitas vezes citados como exemplo dos efeitos de perguntas "tendenciosas". Entretanto, sempre que é mostrado a estudantes, eles perguntam: "Qual pergunta forneceu a resposta *correta* sobre as expectativas americanas

quanto a entrar na guerra?" Esta pergunta não pode ser respondida. Podemos apenas concluir que, em 1939, não havia algo como "a atitude americana sobre a probabilidade de entrarmos na guerra". Nenhuma porcentagem representava a proporção que *realmente* pensava que entraríamos na guerra. Portanto, não se podia *coletar* tais dados; podia-se apenas criá-los fazendo perguntas, e a forma pela qual as perguntas eram feitas tinha um impacto sobre as respostas recebidas.

Nunca conseguimos fazer medidas *precisas*, apenas medidas *úteis*. Mas esta suposição não deve ser tomada como justificativa para a anarquia científica. Pesquisa rigorosa ainda é possível; simplesmente é mais difícil do que se imagina. O resto deste capítulo ocupa-se da lógica e das habilidades requeridas na feitura de medidas úteis.

## Um Quadro de Referências de Operacionalização

Como sugere a discussão anterior, conceitos são codificações gerais da experiência e das observações. Observamos pessoas vivendo em tipos diferentes de estruturas residenciais e desenvolvemos o conceito de unidade residencial. Observamos diferenças em posição social e desenvolvemos o conceito de status social. Notamos diferenças no grau de compromisso religioso das pessoas e desenvolvemos o conceito de religiosidade. Entretanto, é imperativo reconhecer que todos estes conceitos são anotações resumidas da experiência e das observações.

Na ciência, tais conceitos muitas vezes assumem a forma de *variáveis* reunindo uma coleção de *atributos* relacionados. Assim, o conceito de "gênero" resume dois atributos distintos: masculino e feminino. "Período escolar" é uma variável composta pelos atributos calouro e veterano. "Filiação religiosa" pode incluir atributos como protestante, católico, judeu, budista e assim por diante.

Muitos dos conceitos mais interessantes na ciência social representam *variáveis ordinais* feitas de valores dispostos ao longo de uma *dimensão*. Por exemplo, o conceito de status social implica uma hierarquia de valores como "status elevado", "status médio", "status inferior" etc. Religiosidade, preconceito, alienação, sofisticação intelectual, liberalismo e conceitos similares implicam também um conjunto de valores hierarquizados.

A *operacionalização* é o processo pelo qual pesquisadores especificam observações empíricas que podem ser tomadas como indicadores dos atributos contidos em algum conceito. Se o conceito for religiosidade, a operacionalização é o processo

para especificar medições empíricas que indicarão se os respondentes são altamente religiosos, moderadamente religiosos, não religiosos etc. Em geral, diversos destes indicadores são especificados e combinados durante a análise de dados para fornecer uma medida composta (índice ou escala) representando o conceito.

Uma vez que virtualmente todos os conceitos são resumos *ad hoc* da experiência e das observações, não têm significados reais, últimos. Portanto, você não faz medidas corretas ou incorretas, apenas pode determinar até que ponto as medições contribuem para compreender os dados empíricos e desenvolver teorias do comportamento social. Desta perspectiva, é possível dar diretrizes para o processo de operacionalização, para ampliar sua utilidade nas atividades de pesquisa.

Suponha que você queira estudar religiosidade. Talvez queira aprender por que algumas pessoas são mais religiosas do que outras, ou saber as conseqüências de ser mais ou menos religioso. Embora possa iniciar com uma noção geral do que quer dizer com religiosidade, é igualmente provável que você não tenha qualquer indicador em mente. Para iniciar o processo da operacionalização, você deve enumerar todas as diversas subdimensões da variável. Ao fazer isto, deve prestar atenção a pesquisas prévias sobre o tema, bem como às concepções de senso comum sobre religiosidade.

Charles Y. Glock estudou, teórica e empiricamente, as diversas subdimensões da religiosidade; seus estudos propiciam um começo excelente para a abordagem do tema.<sup>3</sup> Ele discute *envolvimento ritual* enquanto participação em atividades como serviços religiosos semanais, comunhão (para os cristãos), orações antes das refeições etc. *Envolvimento ideológico* concerne a aceitação de crenças religiosas tradicionais. *Envolvimento intelectual* refere-se à extensão do conhecimento sobre religião. *Envolvimento experiencial* se refere à extensão das experiências religiosas dos respondentes, como ouvir a voz de Deus, possessões religiosas etc. Finalmente, Glock examina o *envolvimento consuetudinário*, isto é, o quanto o comportamento social é motivado por preocupações religiosas, de acordo com ensinamentos religiosos.

Embora estas cinco subdimensões da religiosidade ajudem a organizar muitos indicadores possíveis, outros indicadores talvez não se encaixem facilmente nelas. Dar dinheiro para a igreja, ou para atividades relacionadas à igreja, pode ser considerado uma indicação de religiosidade, bem como a participação nas atividades sociais da igreja. A lista poderia prosseguir. Portanto, a primeira tarefa é compilar uma lista, tão exaustiva quanto possível, de todos os diferentes indicadores que poderiam ser incluídos no conceito geral de religiosidade.

Ao decidir o que a religiosidade pode ser, é preciso considerar também o que ela *não é*. Deve-se tomar muito cuidado para não construir itens de questionário que indiquem ao mesmo tempo a religiosidade e as variáveis que se irá relacionar a ela, na análise. Se você quiser, por exemplo, determinar a relação entre religiosidade e atitudes em relação à guerra, itens medindo o compromisso com a doutrina cristã da "paz na Terra" não seriam boas medidas de religiosidade. Como as respostas refletiriam as atitudes dos respondentes frente à guerra — além de preocupações religiosas —, tais itens seriam "contaminados". Mesmo que as respostas certamente se relacionem com atitudes relativas à participação na guerra, os achados não contribuiriam em nada para entender a relação geral entre religiosidade e atitudes em relação à guerra.

Deve-se prestar atenção especial ao *oposto* da variável que se busca medir. Se a meta é medir religiosidade, devemos estar sensíveis à variável anti-religiosidade. Algumas pessoas são muito religiosas e outras não religiosas; outras são anti-religiosas. Como anti-religiosidade pode variar em intensidade, você deve determinar a amplitude de variação conceitual da variável. Talvez optando por uma medida de religiosidade variando de baixa a alta, agrupando os anti-religiosos e os que apenas não são positivamente religiosos. Ou talvez estendendo a variável de muito religioso a muito anti-religioso. Em qualquer caso, os itens usados na medida devem basear-se na sua decisão sobre a amplitude. Considere o seguinte exemplo de um item de questionário:

Por favor, diga se você concorda ou discorda da seguinte declaração: "*A religião organizada causa mais mal do que bem.*"

Concordância indicaria uma orientação anti-religiosa. Por outro lado, a discordância não indicaria necessariamente religiosidade. Embora as respostas certamente estejam fortemente correlacionadas com medidas de religiosidade, o item propriamente dito não mede religiosidade, apenas anti-religiosidade.

Este exemplo ilustra um problema comum na medida de dimensões. Ao invés de medir uma variável de baixo a alto, muitas vezes medimos dois pólos opostos. Por exemplo, raramente medimos graus de conservadorismo político, medindo ao invés variações entre liberalismo e conservadorismo. Num estudo sobre ciência e medicina, tentei medir graus diversos de compromisso com perspectivas científicas entre médicos, mas me vi constantemente envolvido em questões sobre perspectivas anticientíficas.<sup>4</sup>

Este problema não tem solução simples. Os interesses de pesquisa variam demais para permitir a elucidação de regras adequadas de procedimentos. O único conselho, agora, é realizar o exercício demorado de listar todas as possíveis subdimensões da variável, anotando as dimensões a serem excluídas do conceito e especificando os pontos conceituais extremos da dimensão que descreve o conceito. Cada item do questionário examinado deve ser avaliado segundo estas decisões. Você deve se perguntar quais seriam as implicações de cada possível resposta: como esta resposta reflete o conceito básico? Apenas através deste processo você pode gerar dados relevantes para uma análise posterior significativa.

## Tipos de Dados

O formato de pesquisa de *survey* gera vários tipos de dados úteis para a pesquisa social, dos quais esta seção oferece uma visão geral.

Embora o capítulo tenha começado afirmando que os dados só existem através do processo científico de gerá-los, mesmo assim tem sentido ver alguns tipos de dados como "fatos". Por "fatos", queremos dizer itens de informação que o respondente acredita representarem a *verdade* e que você, em geral, aceita como sendo verdade. As características demográficas dos respondentes se encaixam nesta categoria. Quando pede aos entrevistados para indicarem seu sexo como masculino ou feminino, eles acreditam que as respostas representam fatos indiscutíveis, e você os aceita como tal. O mesmo pode ser dito das informações sobre idade, raça, região de origem etc.

Às vezes, pede-se aos respondentes para darem informações que eles aceitam como declarações verdadeiras, mas que você não aceita necessariamente como tal. Por exemplo, pode-se perguntar se Deus existe. Respondendo sim ou não, eles dizem o que acreditam ser a verdade sobre o tema. Já você vê estas respostas apenas como descrições dos respondentes, e não como respostas à questão da existência de Deus. Solicitados a concordar ou discordar da declaração "As mulheres são mais emotivas do que os homens", os respondentes dirão o que acreditam ser verdade e você toma as respostas apenas como descrições dos respondentes.

Em outros casos, você pede ao respondente para dar informações reconhecidas por você dois como atitudes subjetivas. Você pode perguntar: "Você acha que o presidente está realizando um bom ou um mau trabalho?" Tanto ele quanto você entendem

que se trata de uma opinião e não de um fato. A distinção entre crenças e atitudes, neste sentido, muitas vezes não é clara, mas, em certos problemas de pesquisa, pode vir a ser importante.

Muita pesquisa social interessante envolve medir orientações, como o preconceito, muitas vezes não reconhecidas pelos próprios respondentes. Num *survey*, você pode fazer diversas perguntas que, combinadas, permitem descrever os respondentes como mais ou menos preconceituosos. Em muitos casos, eles não entendem a finalidade latente das perguntas e podem muito bem discordar das suas caracterizações na análise.

A pesquisa de *survey* não permite medida direta do comportamento, embora frequentemente o comportamento social seja a referência última da pesquisa social. Entretanto, a pesquisa de *survey* permite medida *indireta* do comportamento, muitas vezes de maneiras úteis.

Pode-se pedir aos respondentes para informar seu comportamento passado. Qual sabonete compraram recentemente? Foram à igreja na semana passada? Em quem votaram na última eleição? É claro que estas perguntas estão sujeitas aos problemas de lembrar e ser honesto. Por exemplo, eles podem não lembrar em quem votaram, digamos, na eleição presidencial de 1960. Podem permitir que a memória os engane, particularmente se o comportamento em questão é visto hoje como bom ou mau. (Pesquisas de opinião pós-eleitorais muitas vezes mostram maior porcentagem do eleitorado votando no vencedor do que o que realmente aconteceu.) No entanto, apesar destas desvantagens, relatos de comportamentos passados podem ser muito úteis.

A pesquisa de *survey* também pode examinar comportamento prospectivo, real ou hipotético. Assim, pesquisas de opinião perguntam "Em quem você votará?" Às vezes, é útil criar situações hipotéticas e perguntar aos respondentes como agiriam: "Se seu partido apresentasse uma mulher para presidente, você acha que votaria nela?"

Na maioria das vezes, medidas de comportamento prospectivo são menos confiáveis do que as de comportamento passado. De qualquer forma, tenha em mente essas limitações, quando estiver gerando tais dados. Ao perguntar aos respondentes se votariam numa mulher indicada para presidente pelo seu partido, provavelmente você aprenderá pouca coisa de valor para uma futura candidata feminina. Os dados descritivos produzidos por respostas a uma pergunta destas são provavelmente pouco confiáveis. Entretanto, ao mesmo tempo, os dados gerados pelo item podem ser úteis para descrever os respondentes como mais ou menos antifeministas.

## Níveis de Medição

Vimos na seção precedente que o pesquisador de *survey* pode medir diversas variáveis sociais. Nesta seção, olharemos para estas variáveis de outra perspectiva, considerando quatro níveis de medição: nominal, ordinal, intervalo e razão.<sup>5</sup>

### Medição Nominal

A medição *nominal* apenas distingue as categorias que compõem uma variável. Sexo, por exemplo, é uma variável composta das categorias masculino e feminino. Outros exemplos são filiação religiosa ou política, região do país e tipo de curso superior.

As categorias compõem uma variável nominal, são mutuamente excluídas, não tendo qualquer outra relação entre si. Os níveis restantes de medição, discutidos abaixo, refletem relações adicionais entre as categorias.

### Medição Ordinal

A medição *ordinal* reflete uma ordenação entre as categorias que compõem uma variável. Classe social é um exemplo de variável ordinal, composta, talvez, das categorias classe baixa, classe média e classe alta. Outros exemplos são religiosidade, alienação e anti-semitismo.

A medição ordinal é muito usada na pesquisa social científica. Embora tais medidas sejam muitas vezes representadas por números num índice ou escala, estes números não têm outro significado além da indicação da ordem. Assim, uma pessoa com nota 5 numa escala de alienação é *mais alienada* do que uma pessoa com nota 4 neste índice, embora esta medida ordinal não dê qualquer indicação sobre *quão mais alienada* ela é.

### Medição de Intervalo

A medição de *intervalo* também usa números para descrever condições, mas eles têm mais sentido do que na medição ordinal, porque as distâncias entre os pontos têm significado real. O exemplo mais comum é a escala de temperatura Fahrenheit. A diferença entre 80 e 90 graus é a mesma que a diferença entre 60 e 70 graus. (Veja que a diferença entre as notas 4 e 5 num índice de alienação não é necessariamente a mesma que a diferença entre 3 e 4.)

## Medição de Razão

A medição de *razão* tem as mesmas características da de intervalo, com a característica adicional de ter um *zero real*. Em comparação com a escala Fahrenheit (uma medida de intervalo), a escala de temperatura Kelvin (baseada em graus de temperatura de zero absoluto) é uma medida de razão: enquanto 40 graus Fahrenheit não são necessariamente duas vezes mais quentes do que 20 graus Fahrenheit, 200 Kelvins são duas vezes mais quentes do que 100 Kelvins.

No contexto da pesquisa social científica, a idade é um exemplo de medida de razão. Uma pessoa com vinte anos tem o dobro da idade de alguém com dez anos. Altura, peso e tempo de moradia numa cidade são outros exemplos.

## Implicações dos Níveis de Medição

Na discussão sobre a análise de *survey* na Parte 3 do livro, veremos que diferentes técnicas analíticas podem requerer níveis específicos de medição. Por exemplo, se você está analisando a relação entre duas variáveis nominais, algumas técnicas analíticas são inadequadas.

Ao mesmo tempo, é importante compreender que uma variável pode ser tratada de forma diferente em termos dos níveis de medição. Lembre, por exemplo, que idade é uma medida de razão. Portanto, você pode calcular uma equação de regressão associando idade e altura (outra medida de razão). Mas, suponha que você está estudando uma amostra de estudantes variando de quatorze a vinte anos. Dentro desta faixa limitada, você pode aproveitar os intervalos iguais entre as idades, mas não o caráter de razão da variável. Em outro tipo de estudo, você pode escolher juntar idades nas seguintes categorias: menos de quarenta anos = *jovem*; quarenta a sessenta anos = *meia idade*; acima de sessenta = *idoso*. Finalmente, idade pode ser convertida em medida nominal para certos fins. Respondentes podem ser categorizados como membros do “*baby boom*” pós II Guerra Mundial ou não; podem ser categorizados como tendo nascido durante a Depressão dos anos 30 ou não. (Outra medida nominal — baseada na data do nascimento e não na idade — é a classificação por signos astrológicos.)

Ao desenhar o questionário, considere o tipo de análise que fará após a coleta dos dados. Você deve determinar se e como pode medir cada variável, de forma a permitir a análise

requerida. Se a análise requer dados na forma de medidas de razão, você não deve construir seus questionários de forma a criar apenas variáveis nominais.

## Guias para a Elaboração de Questões

Ao construir itens do questionário, há várias opções. A experiência passada dos pesquisadores de *survey* fornece muitas diretrizes que podem ajudá-lo a gerar dados úteis para análise. Esta seção lida com os dois temas.

### Questões e Declarações

A pesquisa de *survey* é habitualmente vista como fazer perguntas, mas o exame de um *survey* típico provavelmente mostrará tantas declarações quanto perguntas. Isto tem suas razões. Muitas vezes você quer determinar o quanto os respondentes apóiam determinada atitude ou perspectiva. Se for capaz de resumir a atitude numa declaração curta, você pode apresentá-la aos respondentes e perguntar-lhes se concordam ou discordam dela. Rensis Likert formalizou este procedimento com a *escala Likert*, um formato no qual solicita-se aos entrevistados “concordar fortemente”, “concordar”, “discordar”, “discordar fortemente” ou “aprovar fortemente”, “aprovar” etc.

Questões e declarações podem ser usadas vantajosamente numa pesquisa de *survey*. O uso de ambas dá mais flexibilidade ao desenho dos itens e pode tornar o questionário mais interessante.

### Perguntas Abertas e Fechadas

Há duas opções para fazer perguntas. Você pode fazer perguntas *abertas*, solicitando aos respondentes darem suas próprias respostas. Por exemplo, pode perguntar “Qual tema você considera mais importante para o país hoje?”, dando espaço para escrever a resposta (ou solicitando resposta verbal ao entrevistador).

Nas perguntas *fechadas*, pede-se para respondentes escolherem uma alternativa numa lista apresentada. Perguntas fechadas são muito populares, porque dão maior uniformidade de respostas e são mais facilmente processadas. Respostas abertas devem ser codificadas antes de entrar com os dados, e há o perigo de alguns respondentes darem respostas irrelevantes para a intenção do pesquisador. Por outro lado, muitas vezes pode-se entrar com as

respostas fechadas diretamente a partir do questionário e, em alguns casos, podem ser marcadas diretamente em folhas para sensores óticos pelos próprios respondentes, para entrada automática dos dados.

A principal desvantagem das perguntas fechadas está na estruturação das respostas. Quando as respostas relevantes a uma questão são relativamente claras, estruturar as respostas pode não apresentar problema. Entretanto, em outros casos, a estruturação das respostas pode deixar passar algumas que são importantes. Por exemplo, ao perguntar sobre “as questões mais importantes com as quais o país se depara”, você pode apresentar uma lista delas, mas deixar passar questões que os respondentes consideram importantes. (Lembre nossa afirmação anterior de que os dados são criados e não coletados.)

Duas diretrizes devem sempre ser seguidas ao construir questões fechadas. Primeiro, as categorias de respostas propostas devem ser *exaustivas*, ou seja, devem incluir todas as respostas possíveis que se pode esperar. Por isso, os pesquisadores muitas vezes acrescentam a categoria “Outros (especifique por favor)”. Ao fazê-lo, lembre que os respondentes tentarão encaixar suas respostas pessoais numa das categorias fornecidas, mesmo que o encaixe não seja perfeito.

Segundo, as categorias de respostas devem ser *mutuamente excludentes*, ou seja, os respondentes não devem se sentir forçados a escolher mais de uma resposta. (Às vezes, o pesquisador pode querer respostas múltiplas, mas estas respostas criarão dificuldades no processamento.) Você pode garantir respostas mutuamente excludentes considerando cuidadosamente cada combinação de respostas e indagando se uma pessoa poderia, razoavelmente, escolher ambas. Muitas vezes, você pede ao respondente para “selecionar a *melhor* resposta”, mas esta técnica não deve ser usada para compensar um conjunto de respostas malpensadas.

## Tornando os Itens Claros

Os itens de questionários devem ser claros e não ambíguos, e a proliferação de perguntas pouco claras e ambíguas encontradas em muitos *surveys* torna necessário enfatizar este ponto. Muitas vezes, você fica tão envolvido com o tema estudado que opiniões e perspectivas ficam claras para você, mas não para seus respondentes, muitos dos quais terão dado pouca ou nenhuma atenção ao tema.

Por outro lado, você pode ter uma compreensão superficial do tema, deixando de especificar suficientemente a intenção da pergunta. A questão “O que você pensa sobre o sistema proposto de mísseis antibalísticos?” pode evocar no respondente a contra-pergunta: “Qual sistema proposto de mísseis antibalísticos?” Itens de questionário devem ser precisos, de forma que o respondente saiba exatamente qual pergunta espera-se que ele responda.

## Evitando Questões Duplas

Com frequência, os pesquisadores pedem uma resposta única para uma combinação de perguntas. Esta situação parece ocorrer com mais frequência quando o pesquisador se identifica pessoalmente com uma posição complexa. Por exemplo, você pode pedir aos respondentes para concordarem ou discordarem da declaração “Os Estados Unidos devem abandonar seu programa espacial e gastar dinheiro com programas domésticos”. Embora muitas pessoas concordem e outras discordem de forma inequívoca desta declaração, outras não conseguem responder. Alguns podem querer abandonar o programa espacial e devolver o dinheiro aos contribuintes. Outros podem querer continuar com o programa espacial e também aplicar mais dinheiro em programas domésticos. Estes últimos não podem concordar nem discordar da declaração.

Como regra geral, sempre que a palavra *e* aparece numa pergunta ou declaração, você deve checar se não está fazendo uma pergunta dupla.

## Garantindo a Competência dos Entrevistados em Responder

Ao pedir informações, você deve sempre se perguntar se os respondentes são capazes de fazê-lo de forma confiável. Num estudo sobre a criação de crianças, você pode pedir aos respondentes para informarem a idade em que as crianças primeiro “responderam” aos pais. Além do problema de definir o que é “responder aos pais”, é duvidoso que a maioria se lembre disso com precisão.

Em outro exemplo, líderes estudantis governamentais ocasionalmente pedem aos seus constituintes para indicar como as mensalidades dos alunos devem ser gastas. Tipicamente, pede-se aos respondentes para indicar qual porcentagem dos fundos disponíveis deve ser gasta numa longa lista de atividades. Sem um bom conhecimento da natureza destas atividades e dos custos de cada uma, eles não têm como dar respostas com

sentido. ("Custos administrativos" receberão pouco apoio, embora esta categoria seja essencial para o programa como um todo.)

Um grupo de pesquisadores que pesquisava a experiência ao volante de adolescentes insistia em fazer uma pergunta aberta quanto ao número de quilômetros dirigidos desde o momento da habilitação. Embora os consultores advertissem que poucos motoristas estimariam esta informação com alguma precisão, a pergunta foi feita assim mesmo. Alguns adolescentes afirmaram haver dirigido centenas de milhares de quilômetros.

## Fazendo Perguntas Relevantes

Da mesma forma, as perguntas devem ser relevantes para a maioria dos respondentes. Quando se pede atitudes sobre um tema sobre o qual poucos pensam ou se importam, os resultados têm pouca possibilidade de serem úteis. Ademais, os respondentes podem expressar atitudes mesmo nunca tendo pensado sobre o assunto, e você corre o risco de se equivocar.

Este ponto é evidenciado quando pesquisadores pedem respostas sobre pessoas e temas fictícios. Numa pesquisa de opinião que fiz para um candidato, perguntei aos respondentes se conheciam quinze figuras políticas da comunidade. Como exercício metodológico, inventei um nome: Tom Sakumoto. Nove por cento dos respondentes disseram que o conheciam. Destes, metade relatou tê-lo visto na televisão e lido sobre ele nos jornais.

Quando se obtêm respostas a temas fictícios, você pode desconsiderá-las. Mas quando o tema é real, você pode não ter como diferenciar respostas que refletem atitudes genuínas de respostas sem sentido para uma questão irrelevante.

O grau em que os respondentes inventam respostas é até certo ponto controlável pelo pesquisador. Um estudo por Bishop, Tuchfarber e Oldendick o demonstra.<sup>6</sup> Perguntou-se aos respondentes uma de três questões sobre se o "Ato sobre Questões Públicas de 1975" devia ser revogado. Ao primeiro grupo, a pergunta foi "Você concorda ou discorda desta idéia, ou você não pensou muito sobre este assunto?" Apenas 3,7% deram uma opinião sobre a revogação desta lei fictícia. O segundo grupo recebeu a mesma pergunta, mas não lhes foi dada a opção de dizer que não tinham tido chance para pensar sobre isso; 24,3% expressaram uma opinião. No grupo final, os respondentes não tiveram a possibilidade de fugir da resposta, e os que disseram não ter opinião foram pressionados pelos entrevistadores a dar uma resposta; 31,5% expressaram uma opinião.

## Usando Itens Curtos

Querendo ser não ambíguo, preciso e sublinhar a relevância de um tema, muitas vezes você é levado a formular itens longos e complicados. Isto deve ser evitado. É irrelevante, para respondentes que não estudaram o item profundamente, que a intenção de um item seja clara, se for estudada com cuidado. O respondente deve poder ler um item rapidamente, entender sua intenção, e escolher ou dar uma resposta sem dificuldade. Em geral, você deve supor que os entrevistados *irão* ler os itens rapidamente e dar respostas rápidas; portanto, você deve fornecer itens claros e curtos que não serão mal interpretados sob tais condições.

## Evitando Itens Negativos

Uma negação num item de questionário abre caminho para uma fácil má interpretação. Solicitados a concordarem ou discordarem da declaração "Os Estados Unidos não devem reduzir seus arsenais de armas nucleares", grande parte dos respondentes passará por cima da palavra *não*. Assim, alguns concordarão com a declaração, mesmo sendo a favor da redução de armas nucleares, enquanto outros concordarão por opor-se a ela. E você talvez jamais saiba quais respostas refletem esta confusão.

Num estudo sobre liberdades civis, perguntou-se aos respondentes se sentiam que "os seguintes tipos de pessoas devem ser proibidos de lecionar em escolas públicas", de uma lista incluindo comunistas, membros da Ku Klux Klan, e assim por diante. As categorias de resposta "sim" e "não" foram colocadas ao lado de cada item. Uma comparação das respostas a este item com respostas a outros itens refletindo apoio às liberdades civis sugeriu enfaticamente que muitos respondentes davam a resposta *sim* para indicar que tal pessoa podia dar aulas em vez de indicar que ela devia ser "proibida". (Um estudo subsequente que dava como categorias de respostas "permitir" e "proibir" produziu resultados muito mais claros.)

## Evitando Itens e Termos "Tendenciosos"

Tenho repetido que dados de *survey* são criados e não apenas coletados. Portanto, a maneira como os dados são procurados determina a natureza dos dados recebidos. Você deve estar sempre atento ao efeito da redação das perguntas sobre os resultados que obterá.



A maioria dos pesquisadores reconhecera o efeito provável de uma pergunta começando com "Você não concorda com o presidente dos Estados Unidos na crença de que..." e nenhum pesquisador de reputação usaria tal item. Infelizmente, o efeito tendencioso de itens e termos é muito mais sutil do que sugere este exemplo.

A mera identificação de uma atitude ou posição com uma pessoa ou agência de prestígio pode enviesar as respostas. O item "Você concorda ou discorda da proposta do presidente de..." teria este efeito. "Você concorda ou discorda com a recente decisão da Suprema Corte de que..." teria um efeito similar. Isso não quer dizer que esta redação produziria necessariamente um consenso ou mesmo uma maioria em apoio à posição identificada com a pessoa ou agência de prestígio, mas apenas que o apoio provavelmente seria maior do que o que seria obtido sem esta identificação.

Itens de questionários podem ser negativa ou positivamente tendenciosos. "Você concorda ou discorda da posição de Adolf Hitler quando declarou que..." é um exemplo de viés negativo. Muitas vezes o uso de termos como liberal, conservador, comunista, ateu e outros introduz um viés não intencional, embora tais termos às vezes sejam essenciais e apropriados para a intenção da pergunta (por exemplo, "Como você se descreveria politicamente: liberal, conservador, de centro ou alguma outra coisa?") Da mesma forma, Tom Smith<sup>7</sup> demonstrou que as pesquisas de opinião pública informam apoio muito maior para assistência governamental a "pessoas pobres" do que para "pessoas recebendo benefícios". O termo benefícios [*welfare*] tem uma carga muito negativa na opinião pública americana.

Como em outros exemplos, você deve examinar cuidadosamente o objetivo da sua investigação e construir itens que serão mais úteis para ele. Você jamais deve ser levado a pensar que existem, em última análise, formas "certas" ou "erradas" de fazer as perguntas.

## Qualidade das Medições

Antes de nos voltarmos para as técnicas específicas de medição, consideremos alguns critérios gerais sobre a qualidade das medições. Para começar, pode-se fazer medições com graus variados de *precisão*, um termo que representa a qualidade das distinções feitas entre os atributos de uma variável. A descrição de uma mulher como "tendo quarenta e três anos" é mais

precisa do que "nos seus quarenta". Dizer "11,75 polegadas de comprimento" é uma descrição mais precisa do que dizer "mais ou menos um pé de comprimento".

Como regra geral, medidas precisas são superiores a medidas imprecisas, como dita o senso comum. Não há condições em que medições imprecisas sejam superiores a medições precisas. Mas a precisão nem sempre é necessária ou desejável. Se para o objetivo de pesquisa basta saber que uma mulher está nos seus quarenta, qualquer esforço adicional para descobrir a idade precisa é inútil. Portanto, a operacionalização dos conceitos deve ser guiada parcialmente pela compreensão do grau de precisão requerido. Se suas necessidades não são claras, seja mais e não menos preciso.

Entretanto, não confunda a precisão com a *exatidão*. Descrever alguém como "nascido em Stowe, Vermont" é mais preciso do que "nascido na Nova Inglaterra", mas suponha que a pessoa em questão de fato tenha nascido em Dover, New Hampshire. Neste caso, a descrição menos precisa seria mais exata, ou seja, refletiria melhor o mundo real.

Precisão e exatidão são qualidades obviamente importantes nas medições de pesquisa e provavelmente não precisam de mais explicações. Mas, quando cientistas sociais constroem e avaliam medições, prestam atenção especial a duas considerações técnicas: confiabilidade e validade.

## Confiabilidade

*Confiabilidade* é o problema de uma determinada técnica, ao ser aplicada repetidamente a um mesmo objeto, produzir, a cada vez, os mesmos resultados. Suponha que eu lhe pedisse para avaliar meu peso. Você me olharia com cuidado e adivinharia que tenho 80 quilos. Agora, suponha que eu lhe peça para avaliar o peso de trinta ou quarenta outras pessoas, e enquanto você estivesse fazendo isso, eu entrasse na fila usando um disfarce e, quando chegasse a minha vez, você adivinhasse 90 quilos. Te peguei! Este pequeno exercício teria demonstrado que fazer estimar o peso das pessoas não é uma técnica confiável.

Entretanto, suponha que eu lhe emprestasse minha balança de banheiro. Não importa o meu disfarce, você presumivelmente anunciaria o mesmo peso para mim nas duas vezes, mostrando que o uso da balança fornece uma medida mais confiável de peso do que a adivinhação. Mas, tal como a precisão, a confiabilidade não garante a exatidão. Suponha que eu calibrasse

a balança para tirar 3 quilos do meu peso apenas para me sentir melhor. Embora você (confiavelmente) informasse o mesmo peso cada vez, sempre estaria errado.

Suponha que estejamos interessados em saber se as pessoas preferem a cerveja A ou a cerveja B. Podemos, é claro, simplesmente perguntar-lhes qual delas preferem — provavelmente a melhor solução. Mas suponha que desejássemos dados mais “objetivos”. Podemos pedir aos respondentes para dizer quantas As e quantas Bs beberam até hoje. Se você pensar bem, verá que esta é uma técnica pouco confiável. Dificilmente alguém saberá dar esta informação. Da mesma forma, não é provável que as pessoas saibam quantas vezes já foram à igreja, quantos quilômetros já dirigiram etc.

Fazer perguntas que as pessoas acham também irrelevantes provavelmente levará a respostas não confiáveis. Por exemplo, a maioria das pessoas desconhece qual a política externa da China em relação à Albânia, e nem se importam.

Problemas de confiabilidade surgem em vários pontos da pesquisa de *survey*. Como veremos no Capítulo 10, entrevistadores diferentes às vezes obtêm respostas diferentes dos entrevistados como resultado das suas próprias atitudes e aparência. Igualmente, codificadores diferentes podem codificar as mesmas respostas abertas de forma diferente. Por exemplo, se formos codificar algumas centenas de ocupações específicas em termos de algum esquema padronizado de codificação, digamos, um conjunto de categorias criado pelo Ministério do Trabalho ou pelo Bureau de Recenseamento, você e eu não codificaríamos todas aquelas ocupações nas mesmas categorias.

Como se criam medições confiáveis? Há diversas técnicas para medir a confiabilidade de itens de questionários, mas os métodos para *maximizar* a confiabilidade são bastante diretos. Faça apenas as perguntas cujas respostas as pessoas provavelmente sabem, pergunte coisas relevantes para elas, e seja claro no que está perguntando. O perigo é que as pessoas *darão* respostas — confiáveis ou não. As pessoas lhe contarão o que acham da relação entre a China e a Albânia mesmo não tendo a menor idéia de qual seja ela.

## Validade

No uso convencional, a *validade* se refere ao grau com que uma medida empírica reflete adequadamente o *significado real* do conceito considerado. A primeira consideração

é sobre a *validade aparente*. Certas medidas empíricas podem ou não coincidir com nossas convenções e imagens mentais associadas a um conceito. Podemos discutir quanto à adequação de medir religiosidade através de comparecimento à igreja, mas concordaríamos que ir à igreja tem pelo menos alguma relevância para a religiosidade.

Segundo, os pesquisadores chegaram a acordos concretos sobre a melhor forma de medir alguns conceitos. O Bureau do Recenseamento, por exemplo, criou definições operacionais de conceitos como família, residência e status empregatício que parecem ter validade operacional na maioria dos estudos que usam aqueles conceitos.

Edward Carmines e Richard Zeller<sup>8</sup> discutem três tipos adicionais de validade: validade relacionada a critério, validade de conteúdo e validade de construção. A *validade relacionada a critério* às vezes é chamada de *validade preditiva* e baseia-se em algum critério externo. Por exemplo, a validade do conselho de faculdade é evidenciada por sua capacidade de prever o sucesso dos alunos. A validade de um teste escrito para motoristas é evidenciada pela relação entre as notas que as pessoas obtêm no teste e sua competência para dirigir. Nestes exemplos, sucesso escolar e competência ao volante são os *critérios*. Em geral, o comportamento pode servir como padrão da validade de critério para muitas medidas de atitude em pesquisa social (por exemplo, pessoas “preconceituosas” de fato discriminam as minorias?), embora a relação entre atitudes e comportamento seja também assunto importante de estudo por si própria.

A *validade de conteúdo* refere-se ao grau com que uma medição cobre a amplitude de significados incluídos no conceito. Por exemplo, um teste de capacidade matemática não pode se limitar apenas à adição, também precisando cobrir a subtração, a multiplicação, a divisão etc. Em outro exemplo, se dizemos que estamos medindo preconceito em geral, nossas medidas refletem o preconceito contra grupos raciais e étnicos, minorias religiosas, mulheres, idosos etc.?

Finalmente, a *validade de construção* baseia-se no modo como uma medida se relaciona a outras variáveis num sistema de relações teóricas. Por exemplo, suponha que você quer estudar as fontes e consequências da “satisfação conjugal”, desenvolva uma medida de satisfação conjugal e deseje avaliar sua validade. Além da medida, você terá desenvolvido também certas expectativas teóricas sobre como a variável satisfação conjugal se relaciona a outras. Por exemplo, você pode ter concluído

que maridos satisfeitos têm menos probabilidade de espancarem suas mulheres do que maridos não satisfeitos. Se sua medida de satisfação conjugal se relaciona ao espancamento de mulheres da forma esperada, este resultado constitui evidência da validade de construção da sua medida. Mas, se maridos "satisfeitos" e "insatisfeitos" tiveram igual probabilidade de espancar suas mulheres, este resultado desafiará a validade da sua medida.

## Tensão entre Confiabilidade e Validade

Com rodapé destas discussões, deve ser assinalado que muitas vezes há uma certa tensão entre os critérios de confiabilidade e validade. Muitas vezes parece que fazemos uma troca entre as duas.

A maioria dos conceitos realmente interessantes que queremos estudar tem muitas nuances sutis, e é difícil especificar precisamente o que queremos dizer com eles. Os pesquisadores muitas vezes falam desses conceitos como tendo "riqueza de significado". Se você duvida, tente criar definições satisfatórias e medidas para conceitos como preconceito, satisfação, alienação, religiosidade e liberalismo.

Entretanto, a ciência precisa ser específica para poder gerar medições confiáveis. Portanto, com muita frequência, a especificação de definições operacionais e medições confiáveis parece roubar de tais conceitos a riqueza de significado. Religiosidade é mais do que frequentar a igreja; preconceito é mais do que a expressão de uma atitude preconceituosa num questionário. Esta situação é um dilema persistente e inevitável para o pesquisador, e você estará bem preparado para isso se estiver de sobreaviso. Esteja preparado e lide com isso. Se você não chegar a um acordo claro sobre como medir um conceito, meça-o de várias formas diferentes. Se o conceito tiver várias dimensões diferentes, meça todas. Acima de tudo, saiba que o conceito não tem qualquer significado além daquele que você lhe atribuiu, e que a única justificativa para dar um significado particular a um conceito é a utilidade. A meta é medir conceitos de forma a nos ajudar a entender o mundo que nos cerca.

## Formato Geral dos Questionários

O formato de um questionário pode ser tão importante quanto a natureza e a redação das perguntas. Um questionário malformatado pode levar os respondentes a saltar perguntas,

confundi-las quanto à natureza dos dados pedidos e, no limite, levá-los a jogá-lo fora. Vamos sugerir diretrizes tanto gerais quanto específicas.

Como regra geral, o questionário deve estar bem distribuído e não amontoado. Você deve maximizar o "espaço em branco". Pesquisadores inexperientes tendem a temer que seus questionários possam parecer muito longos e, por isso, apertam várias perguntas na mesma linha, abreviam perguntas, e tentam usar o menor número de páginas possível. Tudo isso é desaconselhável e mesmo perigoso. Mais de uma pergunta por linha levará as pessoas a saltarem a segunda. Abreviar perguntas resulta em más interpretações. Em geral, se os respondentes observarem que terminaram rapidamente as primeiras páginas de um questionário que, a princípio, parecia longo, se sentirão menos desmoralizados do que se tiverem gasto um longo tempo na primeira página do que parecia um questionário curto. Ademais, cometerão menos erros e não serão forçados a reler perguntas confusas e abreviadas, nem terão sido forçados a escrever uma resposta longa em espaço pequeno.

Não há como exagerar a valia de espalhar as perguntas num questionário. Questionários espremidos são desastrosos, seja em *surveys* por correio ou por entrevistas.

## Formatos para Respostas

Há vários métodos para apresentar uma série de categorias para o respondente marcar em resposta a uma pergunta. Minha experiência diz que *caixas* espaçadas adequadamente são a melhor escolha. Se o questionário for composto em tipografia ou formatado num bom sistema de processamento de textos, as caixas podem ser geradas de forma fácil e limpa. Também é possível fazer caixas numa máquina de escrever.

Se o questionário for datilografado em máquina de escrever com chaves, caixas excelentes podem ser produzidas com um colchete esquerdo, um espaço e um colchete direito: [ ]. Se não houver colchetes, parênteses funcionam razoavelmente bem no mesmo modo: ( ). Mas você não deve usar barras e sublinhamentos. Primeiro, porque requer mais esforço de datilografia; segundo, porque o resultado não é muito limpo, especialmente se as categorias de resposta tiverem de ser em espaçamento único. A Figura 7-1 mostra uma comparação dos diversos métodos.

O pior método é dar espaços em branco para se assinalar, porque os respondentes muitas vezes farão marcas muito