

foram usados testes adequados de significância, a fim de verificar se diferenças quanto à frequência de discriminação em restaurantes de diferentes níveis de preços, diferente nacionalidade, com ou sem "maîtres", etc., representavam, provavelmente, diferenças reais ou apenas variações casuais. Das variáveis examinadas, verificou-se que apenas o preço estava significativamente relacionado com discriminação. Verificou-se que, quando os restaurantes foram divididos em três níveis de preços, apenas uma proporção muito menor dos restaurantes de nível inferior apresentava discriminação em seu tratamento para com negros.

Resumo

Neste capítulo, mostramos que a função do planejamento de pesquisa é permitir a coleta de provas significativas, com mínimo de esforço, tempo e dinheiro. Tais considerações são importantes em qualquer estudo, qualquer que seja o seu objetivo. Mas a *maneira* de atingi-las depende, em grande parte, do objetivo da pesquisa. Quando o objetivo é a exploração, é adequado um planejamento flexível de pesquisa, que dá oportunidade para considerar muitos aspectos diferentes de um problema. Quando o objetivo de um estudo é descrição precisa de uma situação ou de uma ligação entre variáveis, a exatidão se torna uma consideração básica; há necessidade de um planejamento que reduza o viés e aumente a precisão das provas coligidas. Os planejamentos adequados para estudos exploratórios e descritivos foram discutidos neste capítulo.

Quando o objetivo de um estudo é verificar uma hipótese de uma relação de causa e efeito entre variáveis, apresentam-se outras exigências. No capítulo seguinte são discutidos planejamentos de pesquisa adequados para tais estudos.

mativa de que, nessa época, provavelmente entre 12 e 20 por cento de todos os restaurantes, nessa área e nesse nível de preço, apresentavam discriminação contra os negros.

4

PLANEJAMENTO DE PESQUISA

II. Estudos que Verificam Hipóteses Causais

A Lógica da Verificação de Hipóteses sobre Relações Causais

Inferência Causal a Partir de Experimentos

Inferência Causal a Partir de Outros Planejamentos de Estudo

Resumo

As observações experimentais são apenas experiência cuidadosamente planejada com antecipação, destinada a constituir uma base segura de novo conhecimento.

R. A. FISHER

A Lógica da Verificação de Hipóteses sobre Relações Causais¹

UMA HIPÓTESE de relação causal afirma que determinada característica ou ocorrência (X) é um dos fatores que determinam outra característica ou ocorrência (Y). Os estudos planejados para verificar tais hipóteses precisam fornecer dados a partir dos quais se possa inferir, legitimamente, se X entra ou não na determinação de Y . Antes de considerarmos os tipos de processos de pesquisa que podem dar fundamentos para inferências desse tipo, é necessário discutir um pouco o conceito de "causalidade."

O CONCEITO DE "CAUSALIDADE"

O conceito de causalidade é complexo, e sua análise completa ultrapassaria de muito o objetivo deste livro. Limitaremos nossa discussão aos aspectos que parecem essenciais para a compreensão das exigências para os processos de pesquisa, em estudos planejados para a verificação de hipóteses causais.

A idéia de "senso comum" a respeito de causalidade tende a admitir que um único acontecimento (a "causa") sempre provoca outro acontecimento único (o "efeito"). Na ciência moderna, ao contrário, tende-se a acentuar a multiplicidade de "condições determinantes", que, reunidas,

¹ Para discussão mais minuciosa desse tópico, ver Churchman (1948) e Cohen e Nagel (1934, Capítulo 13).

tornam provável a ocorrência de determinado acontecimento. Tanto o pensamento científico quanto o de senso comum procuram descobrir condições *necessárias* e *suficientes* para um acontecimento. (Tais termos serão definidos nos parágrafos seguintes.) Todavia, enquanto o senso comum leva uma pessoa a esperar que um fator possa dar uma explicação completa, o cientista raramente — e talvez nunca — espera encontrar um único fator ou condição que seja necessário e suficiente para provocar um acontecimento. Ao contrário, está interessado em condições *contribuintes*, condições *contingentes*, condições *alternativas* — todas as quais espera ver como atuantes, a fim de tornar provável, mas não certa, a ocorrência do acontecimento. (Maior explicação desses termos será apresentada a seguir.)

Uma condição *necessária*, como o supõe o termo, é aquela que *precisa* ocorrer, para que ocorra o fenômeno de que é a "causa." Se X é uma condição necessária de Y, então Y nunca ocorrerá a não ser que ocorra a condição X. *Exemplo*: a experiência anterior com entorpecentes é uma condição necessária de vício com entorpecentes, pois seria impossível o vício se o indivíduo nunca tivesse experimentado entorpecentes.

Uma condição *suficiente* é aquela que é *sempre* seguida pelo fenômeno de que é uma "causa." Se X é uma condição suficiente de Y, então sempre que ocorra X, ocorrerá Y. *Exemplo*: a destruição do nervo ótico é uma condição suficiente de cegueira, pois nenhuma pessoa cujo nervo ótico tenha sido destruído pode ver.

Uma condição pode ser *necessária e suficiente* para a ocorrência de um fenômeno. Nesse caso, Y nunca ocorreria, a não ser que X ocorresse, e sempre que X ocorresse, Y também ocorreria. Em outras palavras, não haveria caso em que X ou Y aparecessem sós. Evidentemente, nenhum de nossos exemplos é adequado a esse modelo. Embora o vício em entorpecentes (Y) nunca possa ocorrer a não ser que a pessoa tenha experimentado o uso de entorpecentes (X), é verdade que um indivíduo pode experimentar entorpecentes sem se tornar viciado; assim, a experiência de entorpecentes é uma condição *necessária mas não suficiente* do vício. A fim de compreender o vício em entorpecentes, precisamos encontrar outras condições contribuintes.

De outro lado, embora a destruição do nervo ótico (X) sempre provoque a cegueira (Y), esta pode ocorrer de outras formas, ainda que o nervo ótico não esteja defeituoso; a destruição do nervo é uma condição *suficiente mas não necessária* da cegueira. A fim de compreender todas as "causas" de cegueira, precisamos buscar condições alternativas que possam provocá-la.

Usaremos o exemplo de vício em entorpecentes a fim de ilustrar a busca de condições *contribuintes*, *contingentes* e *alternativas*. Uma condição *contribuinte* é aquela que aumenta a probabilidade da ocorrência de determinado fenômeno, mas não a torna certa; isso ocorre porque é apenas um dentre vários fatores que, juntamente, determinam a ocorrência do fenômeno. Por isso, a pesquisa sobre vício em entorpecentes não se satisfaz com o reconhecimento de que a experiência com entorpecentes é uma condição necessária do vício, mas considera quais os fatores pessoais, de família ou de vizinhança que provocam essa experiência, bem como quais os fatores que tornam mais provável que um indivíduo com a experiência se torne viciado (Chein, 1956). Os estudos comparativos entre rapazes viciados e não-viciados podem verificar, por exemplo, que uma proporção consideravelmente mais elevada dos que se tornaram viciados vem de lares em que o pai não estava presente; em outras palavras, tais estudos sugerem que, nos rapazes, a ausência da figura paterna, durante a infância, é uma influência contribuinte no desenvolvimento de vício em entorpecentes.

No entanto, o comportamento com que lida a ciência social é extremamente complexo; é preciso considerar a interação de fatores. Um fator que atua como uma condição contribuinte de um fenômeno, sob um conjunto de condições, pode não fazê-lo sob outras condições. As condições em que determinada variável é uma causa contribuinte de determinado fenômeno são denominadas condições *contingentes*. Muitas pesquisas em ciência social se referem à identificação de tais condições. Em nosso exemplo de vício com entorpecentes, outros estudos podem procurar verificar se, sob diferentes condições, ocorre a relação entre ausência do pai e vício em entorpecentes. Podem verificar que, nos bairros em que o uso de entorpecentes por adolescentes é raro ou não-existente, os rapazes que cresceram sem um pai

em casa não se voltam para o vício. Depois, a hipótese pode ser refinada da seguinte maneira: nos bairros em que o uso de entorpecentes é comum (condição contingente), a ausência do pai de um rapaz contribui para a probabilidade de que o rapaz se torne um viciado.

A consciência da multiplicidade de causas contribuintes conduz também a um interesse em condições *alternativas* que tornam mais provável a ocorrência de um fenômeno. Assim, pode-se verificar que, em bairros com elevado índice de uso de entorpecentes, o índice de vício é muito elevado, não apenas entre rapazes que cresceram sem um pai, mas também entre rapazes cujos pais estavam nos lares durante a sua infância, mas que os trataram com hostilidade ou indiferença. Depois, a hipótese seria reformulada, a fim de considerar essas condições contribuintes alternativas: a ausência de uma figura paterna *ou* a ocorrência de tratamento hostil ou indiferente pelo pai contribui para a probabilidade de vício, em bairros em que é comum o uso de entorpecentes. Depois, o cientista social pode procurar uma hipótese que inclua um fator comum a *ambas* as condições contribuintes alternativas — por exemplo, que em bairros em que os entorpecentes são facilmente acessíveis, a falta de oportunidade para identificações com uma figura paterna, durante a infância, torna o vício em entorpecentes mais provável na adolescência.

BASES PARA INFERIR A EXISTÊNCIA DE UMA RELAÇÃO CAUSAL ENTRE DUAS VARIÁVEIS

Quase todas as hipóteses em ciência social se referem a condições contribuintes ou alternativas, e às contingências em que atuam. No entanto, é impossível *demonstrar diretamente* que certa característica ou ocorrência (*X*), isoladamente ou em combinação com outras características ou ocorrências (*A*, *B*, *C*, etc.), determina outra característica ou ocorrência (*Y*). Ao contrário, estamos sempre na posição de *inferir*, dos dados observados, que a hipótese segundo a qual *X* é uma condição para a ocorrência de *Y* é ou não sustentável com certo grau especificado de confiança. Qual a prova necessária para justificar tais inferências?

VARIÇÃO CONCOMITANTE. Um tipo de prova significativa refere-se à *variação concomitante* — isto é, até que ponto *X* e *Y* ocorrem juntamente ou juntamente variam de acordo com a forma prevista pela hipótese. Por exemplo, desejamos verificar a hipótese de que *X* é uma condição contribuinte de *Y*. A menos que verifiquemos que a proporção de casos com a característica *Y* é maior entre os casos que têm a característica *X* do que entre os que não têm essa característica, de modo geral concluiremos que a hipótese não é sustentável. Além disso, se nossa hipótese também especifica que a proporção de *Y* é determinada pela proporção de *X*, devemos também verificar que, no conjunto, os casos com maior proporção de *X* mostram também maior proporção de *Y*.

Um exemplo pode ajudar a esclarecer isso. Suponha-se que os médicos de determinada cidade do Norte dos Estados Unidos notem um súbito aumento no número de pacientes com graves perturbações digestivas. Suponha-se que uma fruta tropical — por exemplo, manga — tivesse recentemente aparecido nos mercados da cidade. Um dos médicos pensa que os dois conjuntos de acontecimentos podem estar relacionados; apresenta a hipótese de que a ingestão de mangas (*X*) pode provocar sérias perturbações digestivas (*Y*) em pessoas não acostumadas com a fruta (condição contingente). Durante certo período de tempo, muitos médicos perguntam a todos os pacientes com sérias perturbações digestivas se, recentemente, comeram mangas. Se nenhum deles tiver comido, a hipótese de que comer mangas é uma causa² de perturbações digestivas será rejeitada.

Suponha-se, no entanto, que alguns *tenham comido* mangas. Nesse caso, perguntaríamos a certo de número de pessoas *sem* perturbações digestivas se comeram mangas. Se se verificar que, nesse grupo, a proporção dos que comeram a fruta é tão grande quanto a proporção no grupo dos que têm perturbações digestivas — por exemplo, 30 por cento em cada grupo — inferir-se-á que a ingestão de mangas não

² Como, segundo já foi indicado, a ciência moderna considera a causalidade em função de múltiplas condições determinantes, e não com a esperança de descobrir um único fator que sempre conduza a determinado acontecimento, a palavra "causa" deve ser entendida, em todo o livro, com o sentido de "uma dentre várias condições determinantes que, em conjunto, tornam provável a ocorrência de determinado acontecimento".

tem muita relação com perturbações digestivas; em outras palavras, rejeitaremos a hipótese de que era causa das perturbações. É importante notar, no entanto, que podemos estar errados ao rejeitar a hipótese. Pode ter ocorrido que os que comeram e ficaram doentes tinham pequena perturbação do fígado, que foi agravada pela ingestão de mangas e isso provocou as perturbações digestivas; os que comeram mangas e não ficaram doentes não tinham essa dificuldade no fígado. Em outras palavras, comer mangas pode ser uma condição *contribuinte* de perturbação digestiva, sob a condição *contingente* de mau funcionamento do fígado. Ao desprezar a possibilidade de uma condição contingente sob a qual a ingestão de mangas provoca perturbações digestivas, podemos ter feito uma inferência incorreta, a partir do fato de que proporções iguais de pessoas, com e sem perturbações digestivas, tinham comido a fruta.

Suponha-se, de outro lado, que 90 por cento dos pacientes com perturbações digestivas *tivessem comido* mangas, e que apenas 30 por cento das pessoas sem perturbações digestivas o tivessem feito. Nesse caso, concluiríamos ser sustentável a hipótese de que comer mangas provoca perturbações digestivas.

A prova de que X e Y variam juntamente, de acordo com a forma prevista, pode ser procurada a partir da outra direção. Podemos localizar um grupo de pessoas que tenha comido mangas e outro grupo que não o tenha feito, e comparar a incidência de perturbações digestivas nos dois grupos. Se proporções semelhantes de pessoas nos dois grupos mostrarem sintomas da doença, concluiremos que a ingestão da fruta não provocou o desenvolvimento da doença. Suponha-se, no entanto, que 70 por cento dos que comeram mangas tenham tido perturbações digestivas, enquanto que apenas 10 por cento dos que não a comeram mostrem tais sintomas; nesse caso, concluiremos que a hipótese de relação causal entre mangas e perturbações digestivas é sustentável.

Todavia, seria apenas *sustentável*, e não *provada*. Nesse ponto, seriam sustentáveis três outras possíveis explicações para a relação: (1) as perturbações digestivas de certa forma levaram as pessoas a desejar mangas, de modo que a direção da relação seria oposta à suposta pela hipótese; (2) alguma outra condição provocou *tanto* a ingestão de mangas *quanto*

as perturbações digestivas; (3) alguma outra condição, simplesmente associada com a ingestão de mangas, era responsável pelas perturbações.

ORDEM TEMPORAL DE OCORRÊNCIA DE VARIÁVEIS. A hipótese alternativa de que as perturbações digestivas provocaram um desejo de mangas e, assim, a sua ingestão, conduz a um segundo tipo de provas, significativo para inferências a respeito de causalidade — a ordem temporal dos dois acontecimentos. Um acontecimento não pode ser considerado como a “causa” de outro, se ocorre *depois* desse outro. A ocorrência de um fator causal pode preceder ou pode ser simultânea com a ocorrência de um efeito; por definição, um efeito não pode ser provocado por um acontecimento que só ocorre depois do efeito. [No entanto, é possível que cada termo na relação seja tanto a “causa” quanto o “efeito” do outro. Muitas das relações básicas na ciência física são desse tipo. Por exemplo, a lei de Boyle afirma que, sob condições de temperatura constante, a pressão de um gás está inversamente relacionada com seu volume. Isso significa que qualquer mudança que possamos provocar no volume do gás resultará, simultaneamente, numa mudança em sua pressão, e também que qualquer mudança que possamos provocar na pressão resultará, simultaneamente, numa mudança em seu volume. As duas variáveis na relação funcionam de maneira intermutável como fator causal e como efeito, de forma que se diz que a relação é *simétrica*.]

Um exemplo de uma relação causal simétrica nas ciências sociais é a hipótese de Homans, segundo a qual “quanto mais elevado o posto de uma pessoa num grupo, mais nitidamente suas atividades se conformam às normas do grupo (...). A relação é rigorosamente mútua: quanto mais próximas da norma estiverem as atividades de uma pessoa, mais elevada tende a ser sua posição, mas é também verdade que, se a posição for considerada como a variável independente, quanto mais elevada a posição da pessoa, mais próximas da norma estarão suas atividades” (Homans, 1950).

Embora freqüentemente se encontrem relações causais simétricas nos fenômenos sociais, é muitas vezes útil e conveniente focalizar a influência de um fator sobre o outro. Ao verificar a hipótese de Homans, poder-se-ia desejar ver se é possível fazer com que um membro divergente do grupo

se conforme mais estritamente às normas do grupo, se elevarmos sua posição nesse grupo. Não se concluiria que a elevação na posição do indivíduo no grupo foi a causa de maior conformismo, a não ser que a elevação na posição fosse seguida por um aumento de comportamento de conformismo. Ao distinguir entre causa e efeito, é útil verificar qual dos dois acontecimentos ocorreu em primeiro lugar, supondo-se que não tenha ocorrido simultaneamente. Sabendo-se que uma elevação na posição, em determinado caso, precedeu ao aumento de conformismo, sabemos que o aumento de conformismo não foi o fator causal. (A verificação de que, nesse caso, uma elevação na posição provocou maior conformismo não elimina, naturalmente, a possibilidade de que o inverso também seria verdade.) No entanto, o conhecimento da prioridade temporal não é, em si mesmo, base suficiente para a inferência de causalidade; outros fatores, diferentes da causa presumível, podem ter provocado o efeito.

Em nosso exemplo de perturbação digestiva, se as pessoas comeram mangas *depois* de ter tido perturbações digestivas, e não antes, a ingestão de mangas não seria considerada como uma causa de tais perturbações. Digamos, no entanto, que o interrogatório mostrou que todos os pacientes que sofreram perturbação digestiva, e que comeram mangas, fizeram-no antes do aparecimento dos sintomas. A hipótese de que a ingestão de mangas é uma condição contribuinte continuaria a ser sustentável, e teríamos eliminado uma das hipóteses alternativas — isto é, que as perturbações digestivas levaram as pessoas a desejar mangas. Todavia, precisamos ainda considerar a outra hipótese alternativa: que alguma outra condição provocou a ingestão de mangas e as perturbações digestivas; e que alguma outra condição, apenas associada à ingestão de mangas, era responsável pelas perturbações.

A ELIMINAÇÃO DE OUTROS POSSÍVEIS FATORES CAUSAIS. Poderíamos considerar a possibilidade de que algum desequilíbrio químico provocasse um desejo de mangas e as perturbações digestivas. Tal hipótese poderia ser verificada pelo exame de uma amostra de pessoas, identificando-se aquelas que têm o tipo específico de desequilíbrio químico, e estudando-se os dois subgrupos durante um certo período de tempo, a fim de descobrir se maior proporção dos que têm

desequilíbrio químico comeram mangas e apresentaram perturbações digestivas.

Seria possível procurar fatores acidentalmente ligados com a ingestão de mangas e que poderiam ser responsáveis pelas perturbações digestivas. Suponhamos que as mangas tenham sido vendidas em apenas quatro lojas da cidade. Suponhamos que em três dessas lojas as mangas tenham sido pulverizadas com certo preparado a fim de melhorar sua aparência, enquanto que na quarta loja isso não ocorreu. Suponhamos, além disso, a verificação de que todos os que comeram mangas e tiveram perturbações digestivas tenham comprado a fruta em uma das lojas em que houve pulverização, enquanto que todos os que comeram mangas e não tiveram tais sintomas tenham comprado a fruta na loja em que não houve pulverização. A hipótese de que a ingestão de mangas era, em si mesma, uma causa de perturbação digestiva, seria afastada, e a atenção se voltaria para os efeitos da pulverização; por exemplo, será que a pulverização provocou as perturbações digestivas quando foi usada em outra fruta, ou será que alguma interação entre a pulverização e a composição química das mangas produziu os efeitos tóxicos?

RESUMO DOS TIPOS DE PROVAS SIGNIFICATIVAS PARA A INFERÊNCIA CAUSAL. Sumariemos, rapidamente, os três tipos principais de provas que são significativas para a verificação de hipóteses sobre relações causais:

1. A prova de variação concomitante — isto é, que X (a suposta variável *causal* ou *independente*) e Y (o suposto *efeito*, ou variável *dependente*, ou *critério*) estão ligados segundo a maneira prevista na hipótese. No caso de hipótese de que X é uma condição *contribuinte* de Y , isso significaria que Y apareceria em mais casos em que X estivesse presente que em casos em que X estivesse ausente. Outros tipos de hipóteses causais (que X é uma causa necessária ou suficiente de Y , ou uma causa contingente ligada a A , ou uma causa-alternada com B) exigiriam outros modelos específicos de ligação entre X e Y .

2. Prova de que Y não ocorre antes de X .

3. Prova que elimine outros fatores como possíveis condições determinantes de Y .

Deve-se acentuar que tal prova apenas dá uma base razoável para inferir que *X* é ou não uma causa de *Y*; não dá certeza absoluta. De outro lado, se a prova não confirma a hipótese de uma relação causal, pode ocorrer que tenhamos deixado de lado alguma condição sob a qual *X* é um determinante de *Y*, é, assim, que estejamos errados ao rejeitar a possibilidade de que *X* seja uma condição determinante de *Y*. De outro lado, se a prova confirma a hipótese, ainda pode ocorrer que tenhamos deixado de lado algum outro fator ligado a *X*, e que, na realidade, é a condição determinante de *Y*. Em resumo, podemos concluir que é razoável acreditar que *X* é ou não uma causa de *Y*, mas nunca podemos estar certos de que a relação tenha sido conclusivamente demonstrada. O acúmulo de estudos que indicam uma ou outra conclusão ajuda a aumentar nossa confiança em sua provável correção, mas ainda não constitui prova absoluta.

ADEQUAÇÃO DE DIFERENTES PLANEJAMENTOS DE PESQUISA COMO FONTES DE PROVA

Alguns tipos de planejamento de estudo dão bases mais convincentes para inferências causais. Em nosso exemplo, todos os processos até aqui sugeridos para a verificação da hipótese de que a ingestão de mangas é uma causa de perturbações digestivas utilizaram situações do curso normal de acontecimentos: comparação de indivíduos que ficaram doentes com aqueles que não apresentaram tais sintomas; comparação dos que, por qualquer razão, comeram mangas, com os que não o fizeram; etc. Essa abordagem exigiu certo número de diferentes estudos, qualquer um dos quais apresentou apenas bases tênues para a verificação da hipótese, pois não testava hipóteses alternativas. Por exemplo, uma pesquisa focalizada no fato de a ingestão de mangas ter ou não precedido a perturbação digestiva não daria provas quanto à possível atuação de outros fatores.

Um planejamento *experimental* dá mais certeza e maior eficiência, pois torna possível a reunião simultânea de várias linhas de prova. Numa verificação experimental da hipótese sobre a relação entre mangas e perturbação digestiva, o pesquisador faria com que algumas pessoas comessem mangas e

com que outras pessoas comparáveis não as comessem, durante o período do experimento. Em outras palavras, selecionaria as pessoas destinadas a diferentes "tratamentos", e num tratamento exporia as pessoas à suposta variável causal, enquanto no outro não as exporia (o que, tecnicamente, se denomina "manipulação da variável independente"). Destinaria as pessoas a diferentes tratamentos, de forma a estar certo de que os dois grupos não fossem diferentes, a não ser por acaso, antes de comer as mangas. A comparação da incidência de perturbações digestivas nos dois grupos, depois de um ter comido mangas, enquanto que o outro não o fez, daria provas quanto ao fato de a ingestão de mangas e as perturbações digestivas terem variado conjuntamente. Através de registros cuidadosos do momento de ingestão de mangas e do momento do início das perturbações digestivas, obteria provas quanto à variável que ocorreu antes. A igualdade dos dois grupos antes da ingestão de mangas daria segurança, não apenas quanto à inexistência de diferenças de incidência de perturbações digestivas, antes da ingestão de mangas (o que contribui para maior prova quanto à relação temporal entre as duas variáveis), mas também quanto à inexistência de diferenças quanto a algum outro aspecto ou aspectos que poderiam conduzir a uma diferença subsequente na incidência de perturbações digestivas. Além disso, o experimentador poderia introduzir "controles", a fim de eliminar a possibilidade de que experiências diferentes durante o experimento, além da ingestão ou não de mangas, pudessem ser responsáveis por uma diferença na ocorrência de perturbações digestivas; por exemplo, poderia fazer com que fossem iguais todos os outros alimentos ingeridos por todas as pessoas. Além disso, poderia inserir em seu experimento recursos para verificar hipóteses sobre fatores causais alternativos e específicos. Por exemplo, poderia verificar, simultaneamente, a hipótese a respeito dos efeitos da pulverização usada nas mangas, ao fazer com que algumas pessoas comessem mangas não pulverizadas, algumas comessem mangas pulverizadas, outras comessem outras frutas pulverizadas, etc. Dessa forma, num único estudo poderia obter provas dos três tipos necessários para uma base para inferência de relação causal: variação concomitante, ordem temporal e a possível influência de outros fatores.

Quando um experimento é possível, é o método mais eficiente para verificar uma hipótese de que uma variável, X , tem influência causal em outra variável, Y . Muitas questões de relação causal são facilmente tratadas em estudo experimental. Por exemplo, o pesquisador que está interessado na relativa eficiência de discussão e decisão de grupo, opostas à leitura de um folheto ou à frequência a uma aula — como um método para modificar o comportamento — pode estabelecer uma situação em que um ou mais grupos de indivíduos discutem um determinado problema e chegam a uma decisão a respeito, enquanto indivíduos comparáveis lêem um folheto ou ouvem uma aula sobre o assunto. De forma semelhante, a influência de estímulos subliminares na percepção de estímulos supraliminares subsequentes pode ser pesquisada pela exposição de indivíduos a tais estímulos. Os efeitos da “embalagem” na avaliação de um produto podem ser verificados, dando-se a uma amostra de pessoas um produto em embalagem de determinado estilo e, a uma amostra comparável, o mesmo produto em outra embalagem. Em tais casos, o pesquisador manipula a variável independente.

Em outros problemas, todavia, não é possível a manipulação da variável independente pelo experimentador, nem é possível submeter as pessoas a diferentes tratamentos. Suponha-se que alguém deseje estudar os efeitos de diferentes métodos de educação de crianças na estrutura de personalidade das crianças. É pouco provável que possa fazer com que certas crianças sejam criadas de uma forma, enquanto outras são criadas de outra maneira. (Poderia ser capaz de fazer algo parecido, no caso de crianças de asilo, mas seus resultados não poderiam ser adequadamente generalizados para crianças em situação familiar.) Precisa localizar crianças educadas de forma diferente e avaliar suas personalidades. Se verifica uma correlação, conseguiu prova de variação concomitante. A fim de conseguir uma base para a inferência de que os métodos de educação de crianças (X) são uma causa da estrutura de personalidade (Y), precisa conseguir provas de que Y não precede X , e de que outros possíveis fatores não são os determinantes. De modo geral, as provas sobre tais aspectos serão menos convincentes que as apresentadas por um experimento.

As hipóteses sobre os efeitos de atributos de indivíduos (e não de situações em que estão colocados) com frequência não são tratáveis experimentalmente, no sentido de manipulação da variável “independente” pelo pesquisador. Certamente é possível verificar, de maneira experimental, que pessoas com fome têm maior tendência para interpretar, como representação de alimento, figuras ambíguas, do que pessoas que não sentem fome; o grau de fome pode ser razoavelmente controlado através da especificação do período de tempo em que as pessoas devem ficar sem alimento, antes de ver as figuras. No entanto, muitas características dos indivíduos não podem ser manipuladas dessa forma. Características não-manipuláveis existem, por exemplo, em algumas hipóteses, tais como as seguintes: a lesão cerebral prejudica a capacidade para pensar abstratamente; as pessoas tendem a lembrar as partes de uma mensagem que são coerentes com suas opiniões pessoais e a esquecer as que são contrárias a tais opiniões. O pesquisador que trabalha com seres humanos não manipula a variável de lesão cerebral e, portanto, não destruirá partes do cérebro; precisa procurar os casos existentes de lesão cerebral. E não pode atribuir determinadas opiniões a determinados indivíduos; estes já trazem tais opiniões. O pesquisador consegue a variação que deseja, não por manipulação direta da variável, mas pela seleção de indivíduos em que a variável está presente ou ausente, é forte ou fraca, etc. Dá a indivíduos com e sem lesão cerebral a mesma tarefa; pede que indivíduos com opiniões diferentes leiam o mesmo trecho, etc.

A lógica da verificação de hipóteses sobre o suposto efeito de uma característica da pessoa — por exemplo, a lesão cerebral — que não é criada experimentalmente, é fundamentalmente a mesma da empregada ao verificar as hipóteses, em qualquer outro estudo não-experimental. O estudo não-experimental, em seu planejamento, não permite a exclusão antecipada, com qualquer grau de confiança, da possibilidade de que o efeito tenha sido criado por algum outro fator, correlacionado com o suposto fator causal. Por isso, enfrenta-se a necessidade de eliminar, *ex post facto* (isto é, depois de a suposta variável causal ter ocorrido), a possibilidade de que outros fatores correlacionados com o suposto fator causal possam ter produzido o efeito observado. Por

exemplo, se pacientes com e sem lesão cerebral respondem a um teste, as diferenças em sua realização poderiam refletir os efeitos de lesão cerebral ou poderiam refletir alguns outros fatores — por exemplo, diferenças em angústia que estão ligadas a diferentes tipos de doença, diferenças ligadas a variáveis sócio-econômicas (por ex., a lesão cerebral ocorre mais freqüentemente em certos tipos de profissão), etc. Hipoteticamente, se nos sentíssemos no direito de *provocar* lesão cerebral (tal como se faz com animais inferiores), poderíamos colocar algumas pessoas no tratamento de “sofrer lesão cerebral”, ou no tratamento de “não sofrer lesão cerebral”, de forma a eliminar, antecipadamente, a possibilidade de que as diferenças no teste fossem explicáveis por outros fatores. Depois, seria possível eliminar possíveis explicações alternativas e conhecidas, tais como angústia e variáveis sócio-econômicas, ao colocar as pessoas nas condições de “lesão cerebral” e “sem lesão cerebral” de forma que os dois grupos fossem semelhantes quanto aos tipos de doença representados, o grau de angústia, a diversidade de profissões, etc. Além disso, no experimento *ideal*, o pesquisador é capaz de eliminar explicações alternativas que não foram consideradas preliminarmente. Por exemplo, seria possível não considerar, antecipadamente, a possibilidade de que as supostas diferenças devidas a “lesão cerebral” pudessem refletir diferenças de sexo ou idade. Mesmo assim, se as condições de “lesão cerebral” e “sem lesão cerebral” fossem criadas com a distribuição *casual* dos sujeitos aos dois tratamentos, as pessoas colocadas em ambos não deveriam ser diferentes quanto à distribuição por sexo e idade, a não ser por acaso. Por isso, a distribuição casual das pessoas pelas condições experimentalmente criadas eliminaria, automaticamente, tais explicações alternativas, tivessem sido ou não consideradas antecipadamente. Todavia, é evidente que os pesquisadores *não* provocam lesão cerebral em seres humanos, a fim de obter conhecimento mais exato sobre seus efeitos. Em vez disso, procuram descobrir métodos substitutos para eliminar os possíveis efeitos de outras variáveis.

Ocasionalmente, as situações naturais podem apresentar tanto as condições contrastantes desejadas, quanto a oportunidade para processos suficientemente rigorosos, que tornem possível uma base razoavelmente segura para inferências. Por

exemplo, suponha-se que um cientista social tenha informação sobre as decisões de um sindicato ligado a determinado serviço, importante para os consumidores. Sabe que os líderes do sindicato consideram que, proximamente, uma greve será inevitável. O pesquisador está interessado em várias hipóteses referentes aos efeitos de incômodo pessoal nas atitudes com relação à pessoa ou ao grupo vistos como responsáveis por tal incômodo. Rapidamente o pesquisador se vale da situação iminente para organizar um estudo em duas cidades diferentes, numa das quais ocorrerá a greve, enquanto que na outra isso não irá acontecer. Com sua rapidez, consegue medidas de atitudes com relação a sindicatos, processos para decidir disputas industriais, etc., em dois grupos comparáveis (isto é, os moradores das duas cidades), um dos quais passa pela experiência crítica, enquanto que isso não ocorre com o outro. Importará saber que não foi ele quem provocou a greve? Evidentemente não, desde que possa supor que as duas cidades são realmente comparáveis, antes da greve, e que a decisão de fazer a greve numa cidade, e não na outra, não foi influenciada por características das cidades que poderiam estar ligadas às suas potenciais reações à greve.

Geralmente, no entanto, as situações naturais são complexas e não permitem que o pesquisador suponha, com qualquer grau de confiança, que os dois grupos a ser comparados apenas casualmente sejam diversos. Segundo a afirmação de Greenwood (1945):

(...) a situação criada nos dá maior controle de nossos fenômenos. Podemos determinar, à vontade, as circunstâncias atuantes e, assim, chegar a uma prova mais conclusiva de causalidade. (...) A capacidade para criar as mudanças necessárias permite a verificação de hipóteses que, de outro modo, não seriam acessíveis à verificação.

Por mais cuidadosamente controlado que seja o estudo, todavia, não existe forma para estar *inteiramente* certo da validade de inferências. Por exemplo, sempre existe a possibilidade de que pesquisas posteriores revelem a influência de algum fator não considerado no estudo; ou de que a prova coligida não era decisiva para a hipótese examinada (em outras palavras, as definições de trabalho não dão indicadores

adequados dos conceitos da hipótese); pode ocorrer que os processos de pesquisa não satisfaçam às exigências para fazer inferências. Certamente nas ciências sociais, onde pouco se sabe quanto aos fatores que devem ser controlados, e onde é difícil ou impossível controlar muitos dos fatores significativos (por exemplo, características individuais ou acontecimentos externos à situação experimental), tais possibilidades estão sempre presentes; todavia o princípio se aplica também às ciências físicas, sobretudo nos limites mais adiantados do conhecimento.

Essa possibilidade de inferência errada torna necessária a avaliação dos resultados experimentais no contexto de outros conhecimentos. A confiança no resultado de uma pesquisa exige não apenas a prova estatística de sua precisão (por exemplo, que os efeitos X e X' são, em nível significativo, estatisticamente diversos)³, mas, além disso, a prova de que a interpretação está de acordo com as interpretações de outros "fatos" a respeito dos quais se tem grande confiança. É por isso que a confiança na atribuição de qualquer relação causal entre acontecimentos exige a repetição da pesquisa e a relação entre os resultados e outras pesquisas.

Inferência Causal a Partir de Experimentos

O esquema básico de um experimento é simples: um grupo "experimental" é exposto à suposta variável causal (ou independente), enquanto um grupo "de controle" (ou "testemunha") não o é; depois, os dois grupos são comparados através do suposto efeito (ou variável dependente). Esse modelo permite a coleta de três principais tipos de provas, significativos para a verificação de hipóteses sobre relações causais: (1) prova de variação concomitante — isto é, que a variável causal e a variável dependente estão ligadas; (2) prova de que a variável dependente não ocorreu antes da variável causal; (3) prova que elimina outros fatores como possíveis condições determinantes da variável dependente.

³ Para uma discussão do conceito de significância estatística, ver o Capítulo 11, págs. 466-475.

A prova do primeiro tipo — variação concomitante — é dada de maneira muito simples num experimento. O pesquisador sabe quais as pessoas expostas à suposta variável causal (X); mede todas as pessoas através da suposta variável dependente (Y). Depois, verifica se Y ocorre mais frequentemente entre as pessoas expostas a X que entre as não expostas, ou verifica se as expostas a X mostram maior proporção de Y que as não expostas, ou qualquer outra relação específica entre X e Y , prevista pela hipótese. Por exemplo, suponha-se que um pesquisador verifique a hipótese de que o emprego de programas de instrução por televisão, em classes de oitavo ano que estudam ciência geral, provocará maior obtenção de informação que os métodos mais convencionais de ensino. Para isso, utilizará televisão em algumas classes (o grupo experimental, exposto a X , programas televisionados), mas não em outras (o grupo de controle, não exposto a X). No fim do semestre, irá comparar as notas dos grupos experimental e de controle num exame de informação sobre ciência geral. Se, em média, as crianças das classes onde se usou televisão têm nota mais alta que as de classes comuns, terá prova de que existe relação entre os programas de ensino através de televisão e a informação sobre o assunto.

A prova do segundo tipo — que o efeito suposto não ocorreu antes da suposta causa — é obtida por uma dentre as duas seguintes maneiras: estabelecimento de grupos experimental e de controle de forma que seja razoável supor que não diferem, quanto à variável dependente, antes da exposição à variável independente, ou através da medida de sua posição com relação à variável dependente, antes da exposição à variável independente. Em nosso exemplo de televisão, o pesquisador poderia distribuir as classes para "tratamento" com televisão e sem televisão, de forma que fosse razoável supor que eram comparáveis quanto à informação de ciência geral, antes de fazer o curso; ou poderia aplicar um teste de informação no início do curso; ou poderia fazer as duas coisas. (Vários processos que poderiam ser usados, bem como algumas de suas vantagens e desvantagens, serão discutidos mais adiante neste capítulo). Note-se que a exigência de que X não pode ser considerado como causa de Y , se Y ocorreu antes de X , não significa, no caso de uma hipótese como a da

televisão, que os alunos não devam ter conhecimento anterior de ciência geral. Como a hipótese afirma que o emprego de televisão provocará aquisição de mais conhecimento que os métodos convencionais, exige-se que, antes do início do curso, o grupo experimental não tenha *mais* informação que o grupo de controle.

A prova do terceiro tipo — eliminar outros fatores como possíveis condições determinantes — pode ser obtida de várias formas. Diversos fatores precisam ser considerados como possíveis condições determinantes. Entre os principais estão os seguintes: (1) fatores que ocorreram anteriormente ou que são características relativamente duradouras dos sujeitos; (2) acontecimentos contemporâneos, além da exposição à variável experimental; (3) mudanças de maturação ou desenvolvimento; (4) a influência do processo de mensuração. Processos um pouco diferentes são adequados para a eliminação de cada tipo de fator como possível condição determinante.

1. A fim de eliminar fatores que ocorreram anteriormente ou que são características relativamente duradouras dos sujeitos, o pesquisador pode estabelecer grupos experimental e de controle de forma que seja razoável supor que não diferem sistematicamente, quanto às características ou experiências passadas, que poderiam ser significativas; pode medi-las antes do experimento, através de tais fatores; pode fazer as duas coisas. Por exemplo, no caso da hipótese da televisão, outro fator que surge imediatamente, como possível condição determinante para a aquisição de informação de ciência geral, é a inteligência. A fim de eliminar diferenças de inteligência como a causa de uma diferença de informação, entre os grupos experimental e de controle, depois do experimento, o pesquisador organizará os dois grupos de forma que sejam comparáveis quanto ao QI, ou aplicará testes de inteligência antes do experimento; ou poderá fazer as duas coisas. (Também para este caso, os processos serão discutidos posteriormente).

2. Ao tentar eliminar acontecimentos contemporâneos — além da exposição à variável experimental — como causas de diferenças na posição quanto à variável dependente, o pesquisador pode estabelecer que determinadas condições sejam iguais nos grupos experimental e de controle, ou pode

deliberadamente variá-las durante o experimento. Pode, por exemplo, certificar-se de que o mesmo assunto é apresentado nas classes com e sem televisão, que os professores que usam os dois métodos são aproximadamente iguais quanto à capacidade para ensinar, que os cursos demoram aproximadamente o mesmo tempo, etc. De outro lado, pode intencionalmente escolher professores com diferentes graus de competência. Neste caso, distribuirá alguns dos mais capazes e alguns dos menos capazes pelas classes com televisão e pelas outras. Dessa forma, pode avaliar não só os efeitos da instrução por televisão, mas também os efeitos da capacidade do professor, bem como verificar se os programas televisionados são mais úteis para professores relativamente muito capazes ou para os relativamente pouco capazes.

Alguns acontecimentos contemporâneos que podem influir no resultado dos experimentos em ciências sociais não podem ser controlados, seja no sentido de mantê-los constantes, seja no sentido de intencionalmente manipulá-los. Por exemplo, um acontecimento científico de grande repercussão, e que ocorra durante o experimento, pode provocar um aumento de informação, qualquer que seja o curso de ciência geral. Se esse acontecimento influi da mesma forma no grupo experimental e no de controle, não se cria qualquer problema, pois um efeito comum aos dois grupos não poderia ser, quanto à variável dependente, uma causa de diferenças entre eles. No entanto, pode não ser correto supor que o acontecimento influi da mesma maneira nos dois grupos; seu impacto pode ser diferente no grupo exposto à variável experimental. Foram criados planejamentos complexos — discutidos posteriormente neste capítulo — a fim de considerar essa possibilidade.

3. As mudanças normais, ligadas à maturação e ao desenvolvimento, podem também confundir-se com os resultados do tratamento experimental. Por exemplo, freqüentemente ou não cursos de ciências, as crianças que crescem numa sociedade voltada para a ciência, como é a nossa, podem mostrar, durante o período de um ano, um aumento de conhecimento de ciência geral. Sempre que um experimento se prolonga por um longo período de tempo, os processos de crescimento e desenvolvimento devem ser considerados como possíveis condições determinantes de mudanças. Tam-

bém aqui, se se pode supor que os processos de maturação são iguais no grupo experimental e no de controle, e se se pode supor que o efeito da variável experimental não é específico a determinado estágio de desenvolvimento, os efeitos da maturação podem ser eliminados através da comparação entre os dois grupos. É possível usar planejamentos mais complexos, se existe razão para acreditar que tais suposições não são justificadas.

4. O processo de mensuração empregado no experimento pode influir no resultado. Se as pessoas acham que são "cobaias" de um experimento, ou se acham que estão sendo "testadas" e devem dar uma boa impressão, ou se o método de coleta de dados sugere respostas ou estimula um interesse que o sujeito não sentia antes, o processo de mensuração pode deformar os resultados experimentais. Foram criadas variações no planejamento experimental a fim de considerar os efeitos do processo de mensuração; serão discutidos mais adiante neste capítulo.

Todo o planejamento de um experimento tem a função de permitir que a coleta de provas se realize de tal forma que seja possível fazer, legitimamente, inferências de uma relação causal entre a variável dependente e a independente. No entanto, alguns aspectos são mais importantes: o método de seleção do grupo experimental e do de controle; os momentos em que a variável dependente é medida, bem como o modelo empregado de grupos de controle; o número de possíveis variáveis causais que foram sistematicamente incluídas no estudo. Cada um desses aspectos será discutido nas seções seguintes.⁴

A ESCOLHA DO GRUPO EXPERIMENTAL E DE CONTROLE

Em qualquer planejamento que exija a comparação de dois ou mais grupos de sujeitos que foram expostos a diferentes tratamentos experimentais, existe a suposição de que os grupos comparados sejam equivalentes, antes da introdução dos tratamentos. No entanto, o pesquisador não pode

⁴ Para uma discussão mais completa de fatores que podem influir no resultado de um experimento e das maneiras pelas quais diferentes planejamentos experimentais lidam com tais fatores, ver Campbell (1957).

apenas *fazer* essa suposição; precisa tomar providências a fim de que seja satisfeita.

Evidentemente, é absurda a tarefa de criar ou descobrir grupos equivalentes sob todos os aspectos. Antes de considerar como se resolve esse problema, é necessário distinguir duas diferentes razões para desejar que o grupo experimental e o de controle sejam equivalentes. A primeira é dar uma base para inferir que as diferenças que possam ser encontradas quanto à variável dependente não resultam de diferenças iniciais entre os dois grupos, através da posição na variável dependente ou de outros fatores. Aqui, o objetivo é assegurar, na medida do possível, a *validade* das inferências feitas a partir do experimento. Mas existe um segundo objetivo — o de aumentar a *sensibilidade* do experimento, isto é, aumentar sua capacidade para registrar pequenos efeitos do tratamento experimental, e que poderiam ser obscurecidos pelos efeitos de outros fatores.

Esses dois objetivos exigem processos um pouco diferentes no estabelecimento da equivalência entre os grupos. O objetivo da garantia de validade do experimento — através da segurança de que o grupo experimental e o de controle diferem, inicialmente, apenas por acaso — é atingido através de processos denominados *casualização*. O objetivo do aumento da sensibilidade do experimento, de forma que os efeitos da variável causal se tornem aparentes, mesmo que sejam relativamente pequenos ou que existam poucos sujeitos, é atingido através de processos de *emparelhamento*. Os dois tipos de processo são discutidos mais amplamente nos parágrafos seguintes. É importante compreender que todo experimento bem realizado inclui a casualização, pois esta é suposta pela inferência experimental; por isso, sempre que os processos de emparelhamento são usados, idealmente devem ser usados também os processos de casualização. No entanto, não é necessário complementar a casualização com o emparelhamento, embora isso seja frequentemente desejável.

CASUALIZAÇÃO. A casualização dá a garantia básica para diferenças, entre o grupo experimental e o de controle, que poderiam reduzir a validade de inferências sobre os efeitos do tratamento experimental. A distribuição casual de membros de um conjunto de sujeitos para o grupo experimental e o de controle inclui os mesmos princípios e processos incluídos

na seleção de uma amostra casual simples para um estudo descritivo (ver Apêndice B). Os processos de distribuição devem dar a todos os sujeitos a mesma probabilidade de serem destinados a determinada condição. Os processos são de molde a eliminar qualquer seleção baseada, consciente ou inconscientemente, no julgamento do pesquisador. Por exemplo, é possível atirar uma moeda para cada sujeito, destinando-o para o grupo experimental se cai "coroa", e para o grupo de controle se cai "cara". É possível também numerar todas as pessoas e depois, através do emprego de uma tabela de números casuais (ver Apêndice B), selecionar tantos sujeitos quantos são os desejados para o grupo experimental, e destinar os outros para o grupo de controle.⁵ A distribuição casual de indivíduos para as diferentes condições impede a possibilidade de diferenças *sistemáticas* ou não-casuais entre os grupos selecionados. Isso não significa que o grupo experimental e o de controle são exatamente iguais, mas, ao contrário, que quaisquer diferenças existentes antes da introdução da variável experimental resultam apenas do acaso. As regras de probabilidade permitem especificar a extensão de diferenças que, com maior número de casos (isto é, se a seleção fosse repetida um grande número de vezes), poderiam ser esperadas por acaso. Se, depois de um grupo ter sido submetido ao tratamento experimental, se verifica que os dois grupos são mais diferentes do que seria esperado por acaso, pode-se inferir que a variável experimental provocou a diferença. Evidentemente, essa inferência sempre deve ser feita de maneira provisória, sujeita à possibilidade de que algum outro fator tenha provocado a diferença.

R. A. Fisher (1951), uma das mais notáveis figuras na criação de planejamento experimental, indicou que:

(...) as causas não-controladas que podem influir no resultado (de um experimento) são sempre positivamente inumeráveis. Quando qualquer uma dessas causas é indicada, geralmente se percebe que, com maior trabalho e maior despesa, em grande parte

5 Na maioria dos experimentos, são usados grupos do mesmo tamanho. No entanto, quando é muito elevado o custo da introdução da variável experimental, pode ser mais econômico empregar mais casos no grupo de controle que no grupo experimental. Para discussão mais minuciosa, ver McNemar (1940).

poderia ter sido eliminada. Muito freqüentemente se supõe que tais refinamentos constituam aperfeiçoamentos do experimento. (...) qualquer que seja o grau de cuidado e habilidade experimental para igualar as condições, diferentes daquela que está sendo verificada, e que podem influir no resultado, essa uniformização será sempre mais ou menos incompleta, e em muitos e importantes casos práticos certamente será grosseiramente deficiente. (...) a simples precaução de casualização será suficiente para garantir a verificação de significância, pela qual o resultado do experimento deve ser julgado.

Para voltar ao nosso exemplo de ensino com televisão: digamos que todos os nossos sujeitos devem ser crianças de oitavo ano de determinada escola, que a metade (o grupo experimental) seja destinada a uma classe em que se usará a televisão; a outra metade (o grupo de controle), a uma classe que empregará os métodos convencionais. Mas certamente nem todas as crianças terão o mesmo QI; algumas podem já ter mais informação de ciência que outras; algumas serão mais interessadas pelo assunto que outras; algumas terão melhor visão que outras, etc. Do ponto de vista da validade das inferências que serão feitas, é necessário que o grupo experimental e o de controle não sejam diferentes, em qualquer dessas variáveis, a ponto de tais diferenças provocarem uma diferença em informação sobre ciência — medida na conclusão do experimento — e que será interpretada, incorretamente, como resultante da diferença nos métodos de ensino. Como todas as verificações de significância estatística se baseiam na suposição de que os casos foram casualmente distribuídos pelos grupos comparados, são planejados especificamente para considerar diferenças casuais nas características iniciais dos dois grupos. Por isso, o teste estatístico de significância⁶ oferece proteção contra a possibilidade de que as diferenças na variável dependente — resultantes de diferenças casuais e iniciais entre o grupo

6 Para uma rápida discussão do sentido de testes estatísticos de significância, ver o Capítulo 11, págs. 467-468. Para discussão mais completa, consultar qualquer bom manual de estatística.

experimental e o de controle — sejam incorretamente interpretadas como efeitos do tratamento experimental.

Nem sempre, porém, o cientista social pode distribuir, casualmente, os sujeitos para condições diferentes. Frequentemente, as circunstâncias práticas exigem compromissos com o ideal de seleção casual. Por exemplo, no caso do ensino por televisão, pode não ser possível escolher casualmente, entre todas as crianças de oitavo ano, as que devem ser colocadas na classe experimental. A fim de não perturbar a rotina escolar, pode ser necessário destinar as classes existentes para um ou outro tratamento. Neste caso, as classes podem ser casualmente destinadas a um ou outro tratamento, mas isso não permite tanta proteção quanto a distribuição casual de indivíduos. Às vezes, é possível fazer tais compromissos sem invalidar as bases para inferência no estudo (embora, nos termos de nossa definição, um estudo em que os casos não são casualmente selecionados não constitua um experimento). No entanto, uma forma extrema de distribuição não-casual prejudica seriamente os fundamentos da inferência. Isto ocorre na distribuição de acordo com auto-seleção. Por exemplo, se um pesquisador deseja verificar a hipótese de que a assistência social individual a famílias de crianças delinquentes reduz o comportamento delinquente, seria errado retirar sua amostra experimental de famílias que vieram voluntariamente às instituições sociais, e retirar sua amostra de controle entre famílias com crianças delinquentes semelhantes e que não procuraram esse auxílio. A razão é óbvia: as famílias que procuram esse tipo de auxílio podem ter determinadas características que, direta ou indiretamente, influem na probabilidade de que o comportamento delinquente fosse reduzido, mesmo sem a assistência individual, ou que tornem eficiente a assistência com elas realizada, mesmo que isso não ocorra com outras famílias. Podemos supor, por exemplo, que a mãe que procura o auxílio de uma instituição social, para lidar com a delinquência de seu filho, esteja mais preocupada com o comportamento delinquente e mais ciente dos recursos da comunidade para lidar com tal comportamento, que a mãe que não o faz. Qualquer dessas características poderia fazer com que tivesse mais probabilidade de tomar medidas para mudar o comportamento de seu filho, ainda que não tivesse o auxílio da

assistente social. E o fato de ter procurado o auxílio da assistência social poderia significar que fosse mais receptiva a esta, e, assim, que a assistência tivesse mais probabilidade de influência no comportamento delinquente do que se recebesse, involuntariamente, tal auxílio. O mesmo princípio se aplica sempre que os sujeitos se colocam no grupo "experimental" ou no de "controle".⁷

EMPARELHAMENTO. Embora a distribuição casual, quando possível, seja geralmente considerada como capaz de dar proteção adequada contra a interpretação de diferenças na variável dependente como se fossem resultantes da variável independente — quando, na realidade, decorrem de diferenças anteriores entre os dois grupos — não é o processo mais eficiente para aumentar a sensibilidade do experimento. Para a eficiência da pesquisa, é desejável que o experimento revele diferenças reais, provocadas pelo tratamento experimental, mesmo que sejam pequenas com relação às diferenças provocadas por outras variáveis. Em nosso exemplo de ensino por televisão, o método de ensino pode ter menos influência, na informação sobre ciência, do que o QI. A distribuição casual das crianças pelos grupos ensinados por um ou outro método não conduziria, provavelmente, a um emparelhamento (igualização) exato dos dois grupos, em função do QI. Essa diferença no QI poderia provocar, no fim do estudo, uma diferença na informação. Como já foi observado, os testes estatísticos de significância, baseados na suposição de amostragem casual, dariam proteção contra a atribuição dessa diferença à diferença nos métodos de ensino. No entanto, poderia haver uma pequena diferença entre os dois métodos de ensino, e que seria obscurecida pela diferença em informação, ligada ao QI. Quanto maior a redução de tais diferenças "exteriores", maior a probabilidade de que apareçam os efeitos do tratamento experimental.

Um exemplo hipotético supersimplificado pode auxiliar a esclarecer este aspecto. Suponha-se que, em nosso estudo dos efeitos do ensino de ciência com o auxílio da televisão, consideremos, como unidades de amostragem, as classes

⁷ Este é um problema freqüente em estudos planejados para verificar hipóteses causais que não seguem o modelo de experimentos controlados. Os métodos para enfrentar tal problema são considerados na secção deste Capítulo que discute tais estudos.

existentes, e não os indivíduos. Suponha-se, além disso, que oito classes sejam usadas no experimento — quatro das quais recebem instrução por televisão, e quatro servem como controles. Digamos que tais classes sejam diferentes quanto ao QI: quatro tenham um QI médio acima de 100 (serão denominadas “superiores” na tabela abaixo) e quatro estejam abaixo de 100. Se as classes fossem casualmente destinadas aos tratamentos de televisão e sem televisão, poderíamos ter um padrão como o apresentado abaixo; as notas no exame final são apresentadas na coluna da direita.

Classe	Televisão	QI médio	Nota média no exame
1	Sim	Superior	90
2	Não	Superior	80
3	Sim	Inferior	70
4	Não	Superior	80
5	Sim	Inferior	70
6	Não	Inferior	60
7	Sim	Inferior	70
8	Não	Superior	80

Neste exemplo, o resultado final médio de todas as classes, no exame final, é 75. O resultado médio das classes com QI superior é 82,5; o das classes com QI inferior, 67,5. Contudo, tanto as classes com televisão quanto as que não a tiveram apresentam resultados médios de 75, embora seja evidente, quando se olha a tabela, que as classes com QI superior, com instrução com televisão, apresentam resultados mais elevados que as classes com QI superior que não tiveram televisão, e que as classes com QI inferior, que tiveram instrução com televisão, apresentam resultados mais elevados que as classes com QI inferior que não a tiveram. No entanto, o fato de a distribuição casual ter provocado uma situação em que três das quatro classes que tiveram instrução com televisão têm QI inferior, enquanto que três, das quatro que não tiveram televisão, têm QI superior, obscurece a influência da televisão quando se compara o resultado médio de todas as classes que recebem o tratamento experimental com o de todas as classes de controle.

Para exemplificar o efeito do emparelhamento, suponhamos que os tratamentos com televisão e sem televisão tenham tido números iguais de classes com QI superior e inferior. Fazendo-se as mesmas suposições na tabela anterior, quanto à relativa contribuição de inteligência e método de ensino, para o exame final, os resultados seriam os seguintes:

Televisão	QI médio	Nota média no exame
Sim	Inferior	70
Sim	Superior	90
Sim	Inferior	70
Sim	Superior	90
Não	Inferior	60
Não	Superior	80
Não	Inferior	60
Não	Superior	80

Agora, o resultado médio das classes com televisão é 80; o das classes sem televisão é 70. Pelo emparelhamento dos grupos em função da inteligência, permitiu-se o aparecimento dos efeitos do método de ensino.

Deve-se notar que, no processo de emparelhamento, é importante não sacrificar a casualização. Em nosso exemplo, a casualização poderia ter sido incluída de diversas formas no processo; por exemplo, lançando-se moedas a fim de determinar quais as duas classes, dentre as quatro de QI superior, e quais as duas classes, dentre as quatro de QI inferior, deveriam receber o tratamento com televisão. Ou, se houvesse um grande número de classes para a seleção, poderiam ser divididas em dois grupos — QI superior e QI inferior; depois, através de uma tabela de números casuais, duas classes de cada grupo poderiam ser selecionadas para o tratamento com televisão, e duas de cada grupo serviriam como controles.

O fato de o emparelhamento poder tornar um experimento mais sensível — pelo controle dos efeitos de outras variáveis, que poderiam obscurecer o da variável em que o pesquisador está interessado — freqüentemente leva os experimentadores a complementar a casualização com processos de emparelhamento. Comumente são usados dois métodos:

o controle de exatidão e o controle de frequência de distribuição.⁸ Ambos, quando combinados — como realmente devem ser — com processos de casualização, são métodos de amostragem casual estratificada (ver o Apêndice B).

A igualização dos grupos, através de controle de exatidão, exige o emparelhamento dos indivíduos nos grupos, caso por caso. Para considerar um problema complexo — suponha-se que desejemos verificar o efeito de terapia psicanalítica de certo tipo nas atitudes de pessoas com preconceito. Tentaríamos estabelecer dois grupos de pessoas emparelhadas, indivíduo por indivíduo, quanto a atitudes e fatores que poderiam ser significativos para sua predisposição a mudança de atitude. Vale dizer, para a pessoa A, que tem muito preconceito, que está submetida a pressões do seu grupo social para ter preconceito, que é inteligente, que não tem intensas necessidades inconscientes que motivem seus preconceitos, etc., tentaríamos encontrar uma contraparte exata, A'. A seria destinada a um grupo, A' ao outro. Para B, que tem pouco preconceito, que está submetida a pressões sociais para não ter preconceito, que tem inteligência média, que tem inseguranças subjacentes que se exprimem nos seus preconceitos, etc., tentaríamos encontrar uma pessoa B'. E assim por diante, até que, para cada indivíduo no grupo experimental, tivéssemos um indivíduo emparelhado no grupo de controle.

Evidentemente, por várias razões o emparelhamento de indivíduos é uma tarefa muito difícil. Em primeiro lugar, para que o emparelhamento seja exato e para que os indivíduos sejam emparelhados quanto a vários fatores, precisa haver um grande número de casos para a seleção, a fim de se chegar a um emparelhamento adequado. Todos esses casos devem ser medidos quanto aos fatores significativos, mas apenas alguns serão usados. Quanto mais exato o emparelhamento, e quanto maior o número de fatores que devem ser emparelhados, maior o número de casos para os quais não existe emparelhamento. Em segundo lugar, frequentemente é difícil saber que fatores, dentre os muitos que são importantes e possíveis, são os mais importantes para a

8 Para discussão mais minuciosa de tais métodos de emparelhamento, bem como do método de distribuição casual, ver Greenwood (1945).

obtenção de controle de exatidão. Raramente é possível o emparelhamento, com qualquer grau de exatidão, em mais que dois ou três fatores. Felizmente, no entanto, os fatores importantes estão, com frequência, tão interrelacionados que o emparelhamento em um fator provoca o emparelhamento parcial em outros fatores; existe um "rendimento decrescente", à medida que se controlam outros fatores. Em terceiro lugar, é frequentemente difícil conseguir medidas adequadas dos fatores que talvez seja importante emparelhar; considere-se, por exemplo, o suposto experimento sobre os efeitos da psicanálise. Se não existem medidas adequadas dos fatores supostamente importantes, é muito difícil, evidentemente, que o emparelhamento seja muito exato.

O bom emparelhamento pode aumentar muito a eficiência de um experimento, pois diminui o tamanho das diferenças — quanto à variável dependente — que devem ocorrer, apenas por acaso, entre o grupo de controle e o grupo experimental. Quando as diferenças casuais são pequenas, é mais fácil demonstrar uma diferença devida ao efeito da variável experimental. Quando é possível emparelhar os membros de cada par — em função da posição inicial na variável dependente ou alguma característica muito correlacionada com ela — é necessário um número muito menor de casos do que seria necessário com grupos não-emparelhados para se estar certo de que uma diferença obtida não é explicável pelo acaso. No entanto, se o emparelhamento é feito em função de alguma outra característica — que, na realidade, não tem relação com a variável dependente — o processo de emparelhamento, por mais complexo que seja, não apresenta qualquer vantagem estatística ou experimental. Suponha-se, em nosso exemplo de efeito da terapia nas atitudes, em vez da possibilidade de emparelhar casos em todas as características sugeridas, a possibilidade de emparelhar apenas através da inteligência. Se esta não estivesse relacionada com preconceito, não teríamos conseguido o emparelhamento de casos quanto ao seu grau inicial de preconceito, e assim não teríamos eliminado tais diferenças iniciais como uma das causas das diferenças observadas depois da terapia. Em tal caso, o emparelhamento não apenas deixa de representar uma vantagem, mas, por razões excessivamente técnicas para serem

aqui discutidas, o mau emparelhamento pode até representar uma desvantagem estatística.⁹

Como já foi dito, o emparelhamento é, para a distribuição casual, não um substituto, mas um complemento. Os processos de emparelhamento podem considerar apenas algumas variáveis; as que não são consideradas devem ser casualmente distribuídas entre o grupo experimental e o de controle. Assim, se o nosso processo de emparelhamento, no experimento de terapia, nos deu um número de pares de indivíduos emparelhados (*A* e *A'*, *B* e *B'*, *C* e *C'*, etc.), precisamos empregar algum processo casual para determinar se *A* ou *A'* devem ir para o grupo experimental, se *B* ou *B'* devem ir para esse grupo, etc.

Deve-se acentuar um outro aspecto: quando se faz experimento com grupos sociais que têm certa unidade psicológica, se existe interesse pela coletividade (um clube, uma fábrica, uma classe), e não por indivíduos separados, é adequado emparelhar grupo com grupo, e não indivíduo com indivíduo. As interações dos indivíduos, suas atitudes com relação um ao outro, seus interesses e suas capacidades complementares — tudo isso exige que, quando os grupos constituem o objeto de estudo, os grupos, e não os indivíduos, sejam emparelhados (ver Deutsch, 1949).

O controle de distribuição de frequência é uma tentativa de colher algumas das vantagens do emparelhamento, sem pagar o preço da perda de tantos casos, que é o que geralmente ocorre no controle de exatidão. De acordo com o nome, o controle de distribuição de frequência tenta emparelhar um grupo experimental e um grupo de controle — através da distribuição geral de determinado fator ou determinados fatores nos dois grupos, e não de indivíduo por indivíduo. Por exemplo, se a idade for considerada importante para os efeitos estudados, tentar-se-ia fazer com que a média das idades nos dois grupos e as distribuições de idade nos dois grupos fossem semelhantes. Para cada um dos outros fatores considerados importantes, tentar-se-ia igualar as distribuições nos dois grupos. Pode ser necessário eliminar alguns casos para igualar as distribuições, mas, em comparação com o método de emparelhar indivíduo por

9 Para discussão minuciosa, ver Edwards (1950, Capítulo 14).

indivíduo, perde-se um número relativamente pequeno de indivíduos.

O emparelhamento por distribuição, e não por controle de exatidão, tem uma desvantagem potencial de certa importância. Embora sejam iguais as distribuições em fatores isolados, os grupos podem ser, na realidade, muito mal emparelhados em combinações de tais fatores. Suponha-se, por exemplo, que estivéssemos emparelhando dois grupos, em função de renda e idade. Apesar de, separadamente, as distribuições de renda e idade estarem iguais nos dois grupos, seria possível ter, num grupo, pessoas jovens com renda elevada e pessoas mais velhas com baixa renda, e ter, no outro grupo, a combinação exatamente oposta.

Dois outros aspectos devem ser notados ao considerar se o emparelhamento por distribuição de frequência conseguiu, na realidade, grupos iguais. Em primeiro lugar, não se deve supor que as distribuições nos dois grupos sejam semelhantes, apenas porque suas médias são semelhantes. Isso pode não ocorrer; um grupo pode ter grande amplitude com relação à média, o outro pode ter pequena amplitude de variação. Evidentemente, tais diferenças influirão nos resultados da pesquisa. Além da média, para a igualização das distribuições são importantes outras medidas estatísticas (como o desvio padrão, a assimetria, o achatamento — ver Capítulo 11, págs. 461-475) que dão uma indicação da natureza da distribuição de casos. Em segundo lugar, ainda que um teste estatístico indique que duas distribuições não são significativamente diferentes (isto é, que as duas amostras poderiam ter sido retiradas da mesma população), não há uma justificação necessária para sua equivalência. É um erro lógico supor que, se você não demonstrou conclusivamente que as coisas são *desiguais*, demonstrou que são *iguais*. As diferenças entre os grupos devem ser tão pequenas quanto possível, considerando-se os recursos disponíveis. Deve-se lembrar que, idealmente, a casualização deve sempre ser usada para igualar os fatores não-emparelhados, sempre que se emprega o controle de distribuição de frequência ou de exatidão.

TIPOS DE PLANEJAMENTO EXPERIMENTAL

Se desejamos verificar a hipótese de que X é uma causa de Y , através da comparação de um grupo que foi submetido a X com um grupo que não o foi, é evidentemente essencial medir os dois grupos com relação a Y , seja durante o período em que estão submetidos a X , seja depois disso. Às vezes é desejável ou até essencial ter, além disso, medidas de sua posição com relação a Y , antes de terem sido submetidos a X . O momento em que a variável dependente é medida dá uma base para a classificação dos experimentos sob dois agrupamentos principais: "apenas depois" e "antes-de-isto". O modelo empregado de grupos de controle apresenta uma base para outras subdivisões.

EXPERIMENTOS "APENAS DEPOIS". Em estudos deste tipo, como o nome indica, o grupo experimental e o de controle são observados ou medidos com relação à variável dependente (Y) apenas durante ou após o período em que o grupo experimental está submetido à suposta variável causal (X).¹⁰ Tal como ocorre em todos os planejamentos experimentais que exigem grupos de controle, o grupo de controle e o grupo experimental são selecionados antes da introdução da variável experimental (X), e a variável é introduzida, seja especificamente para o experimento, seja de um modo e em um mo-

10 Geralmente, o "grupo de controle" é submetido apenas à "ausência de X ", e não a alguma variável contrária, X' , cuja influência deva ser comparada a X . No entanto, é possível submeter o "grupo de controle" a X' e, assim, fazer com que constitua um segundo grupo experimental. Nesse caso, é às vezes possível usar apenas um grupo de sujeitos, que "servem como seus próprios controles". Por exemplo, ao verificar a hipótese de que as pessoas tendem a lembrar melhor as tarefas inacabadas que as completadas, um grupo de sujeitos pode ter permissão para terminar todas as suas tarefas, enquanto outro grupo pode ser impedido de completar qualquer das tarefas. Neste caso, haveria o emprego de dois grupos de sujeitos — um dos quais é submetido a X ("tarefas completadas"), outro a X' ("tarefas inacabadas"). É possível verificar a mesma hipótese com o emprego de um grupo de sujeitos; neste caso, todos seriam submetidos a X e a X' , isto é, teriam permissão para completar algumas tarefas, mas não outras. Ao fazer isso, é necessário considerar cuidadosamente a seqüência de tarefas terminadas e inacabadas, a fim de evitar complicações que poderiam surgir de alguns fatores, tais como a tendência para lembrar melhor a primeira tarefa, a última, etc.

mento especificados, previamente conhecidos pelo experimentador. Como se mostrou na seção anterior, os sujeitos são casualmente destinados aos dois grupos, com ou sem processos complementares de emparelhamento.¹¹ Se se usa uma das técnicas de emparelhamento, este se faz em função de outras características, diversas da posição inicial na suposta variável dependente (Y), pois esta não é medida antes do tratamento experimental. Os dois grupos são observados ou medidos, com relação a Y , apenas durante ou após o período em que o grupo é submetido à suposta variável causal (X). A coluna I na tabela da pág. 126 mostra uma representação esquemática desse tipo de estudo.

Neste planejamento, conclui-se que a hipótese é ou não sustentável apenas pela comparação da ocorrência de Y (ou sua extensão, ou sua natureza, etc.) no grupo experimental, depois de este ter sido submetido à variável X , com a ocorrência (etc.) de Y no grupo de controle, que não foi submetido a X . Assim, na Coluna I, comparamos Y_2 com Y_2 . Evidentemente, isso prova se X e Y variam ou não conjuntamente.

O que dizer a respeito da prova de que Y não precedeu a X no tempo? Isso se infere do método empregado no estabelecimento dos dois grupos. Se foram casualmente escolhidos, a teoria da probabilidade nos diz até que ponto se pode esperar que sejam diferentes, por acaso, em função de sua posição com relação a Y , antes da introdução do tratamento experimental, e o teste de significância leva em conta tais diferenças casuais. Se, além disso, foram emparelhados em função de alguma variável ou algumas variáveis ligadas à posição quanto a Y , as prováveis diferenças iniciais quanto a Y serão até menores do que seria esperado por acaso. Os processos de casualização, usados no estabelecimento dos grupos (tenham sido ou não complementados por emparelhamento), também dão uma base para concluir que as diferenças entre os grupos — quanto a outros fatores passados ou características relativamente duradouras que poderiam influir em sua posição quanto a Y , depois do tratamento experi-

11 Os estudos em que os grupos que não foram escolhidos casualmente são submetidos a diferentes condições, e depois comparados, não constituem, a rigor, experimentos. Tais estudos são discutidos mais adiante neste capítulo.

TIPOS DE PLANEJAMENTO EXPERIMENTAL

Condição	1 "Apenas depois"		2 "Antes-depois" com um único grupo		3 "Antes-depois" com grupos intercambiáveis		4 "Antes-depois" com um grupo de controle	
	Grupo experimental	Grupo de controle	Grupo experimental	Grupo de controle	Grupo experimental	Grupo de controle	Grupo experimental	Grupo de controle
Seleção anterior de grupos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Mensuração "antes"	Não	Não	Sim	Não	Sim	Sim	Sim	Sim
Submete a variáveis experimentais	Sim	Não	Sim	Sim	Sim	Talvez	Sim	Não
Submete a acontecimentos não-controlados	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Mensuração "depois"	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim
Mudança	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$
5 "Antes-depois", com dois grupos de controle								
Condição	Grupo experimental	Grupo de controle I	Grupo de controle II	Grupo experimental	Grupo de controle I	Grupo de controle II	Grupo de controle III	Grupo de controle III
Seleção prévia de grupos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Mensuração "antes"	Sim	Sim	Não	Sim	Sim	Sim	Não	Não
Submete a variáveis experimentais	Sim	Não	$(Y''_1 = \frac{Y_1 + Y'_1}{2})$	Sim	Sim	Sim	$(Y''_1 = \frac{Y_1 + Y'_1}{2})$	$(Y''_1 = \frac{Y_1 + Y'_1}{2})$
Submete a acontecimentos não-controlados	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Mensuração "depois"	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$
Mudança	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$	$d = Y_2 - Y_1$
Interação		$I = d - (d' + d'')$	$I = d - (d' + d'')$	$I = d - (d' + d'')$	$I = d - (d' + d'')$	$I = d - (d' + d'')$	$I = d - (d' + d'')$	$I = d - (d' + d'')$

mental — não são maiores que as diferenças casuais, consideradas pelo teste de significância.

Embora este planejamento tenha um problema comum a toda a pesquisa social — isto é, que os processos de mensuração usados possam alterar a característica que pretendem medir —, o problema é menos sério aqui que nos estudos de "antes-depois." (As dificuldades criadas por mensurações feitas antes da "exposição" à variável experimental serão discutidas mais adiante, com relação aos planejamentos "antes-depois.")

O que dizer sobre os efeitos de outros acontecimentos contemporâneos ou da maturação? Supõe-se que, entre o momento da seleção e o da medida de Y , ambos os grupos estejam submetidos aos mesmos acontecimentos externos e que passem por processos semelhantes de maturação. Se esta suposição é justificada, a posição do grupo de controle na variável dependente (Y_2), no fim do experimento, inclui a influência dos acontecimentos externos e dos processos de maturação que influíram nos dois grupos. Portanto, a diferença (d) entre Y_2 e Y'_2 pode ser considerada como uma indicação do efeito do tratamento experimental, desde que os acontecimentos externos e os processos de maturação não tenham interagido com a variável experimental, de modo a mudar seus efeitos. (A possibilidade de interação entre a variável experimental e outros fatores será discutida mais adiante.)

O planejamento "apenas depois" pode ser exemplificado por um estudo dos efeitos de um filme, *A Batalha da Inglaterra*, realizado pela Experimental Section of the Research Branch in the War Department's Information and Education Division (Hovland, Lumsdaine e Sheffield, 1949). Neste estudo, o grupo experimental viu o filme, enquanto o grupo de controle não o viu. Ao distribuir as pessoas pelos dois grupos, não parecia possível a seleção casual de indivíduos, pois isso exigiria a retirada de soldados de suas unidades regulares. Esse processo diferente não apenas teria criado dificuldades administrativas, mas presumivelmente também despertaria, no pensamento dos soldados, algumas questões sobre o objetivo da operação. Por isso, a seleção de casos para a amostra foi feita a partir de companhias, e não de indivíduos. Dadas as provas de que as companhias

eram, sob certos aspectos, diferentes, os pesquisadores estabeleceram dois grupos de companhias comparáveis quanto a algumas características, tais como resultado médio no Teste de Classificação Geral do Exército, educação, idade, região de nascimento, estágio de treinamento, etc. A casualização foi feita através do lançamento de moedas, a fim de decidir quais dos grupos emparelhados veriam o filme, quais não o veriam.¹²

O grupo experimental viu o filme durante sua hora semanal de orientação, como parte do processo normal de treinamento. O grupo de controle não viu o filme. Aproximadamente uma semana depois, os soldados de ambos os grupos foram solicitados a preencher um questionário, como parte de um levantamento do Ministério da Guerra destinado a "verificar como um grupo típico de soldados se sente a respeito de vários assuntos ligados à guerra." Juntamente com itens de "camuflagem", havia certo número de itens de fatos e opiniões que, segundo se poderia esperar, deviam ter sido influenciados pelo filme, mas não estavam tão especificamente ligados a ele que sugerissem uma ligação entre o "levantamento" e o filme. A medida dos efeitos do filme era a diferença (acima de diferenças casuais) entre a proporção do grupo de controle e a do experimental que respondia de certo modo aos itens significativos.

EXPERIMENTOS "ANTES-DEPOIS." Por várias razões, um experimentador pode desejar ter, além das medidas de *Y* depois da exposição à variável experimental, medidas de *Y* antes de tal exposição.

1. Como já se discutiu na seção sobre a seleção de grupos experimentais e grupos de controle, pode desejar

¹² Este processo de estabelecer dois grupos emparelhados e utilizar a casualização apenas para verificar qual deve ser submetido ao tratamento experimental é, naturalmente, um nítido desvio com relação ao processo recomendado, que consiste em decidir, casualmente, que indivíduo ou que unidade de cada par emparelhado deve ser indicado para o grupo experimental. A rigor, uma estimativa das prováveis diferenças casuais entre dois grupos, obtida através da suposição de que os indivíduos foram casualmente distribuídos, não apresenta uma estimativa adequada das prováveis diferenças casuais entre dois grupos estabelecidos dessa forma. Esse fato foi reconhecido pelos pesquisadores. O estudo exemplifica o fato de que as exigências realistas às vezes obrigam a concessões aos processos ideais.

aumentar a sensibilidade de seu experimento, através do emparelhamento de casos, em função de sua posição inicial na variável dependente. Como se indicou nessa seção, esse emparelhamento deve ser acompanhado por processos de distribuição casual.

2. Pode desejar verificar se existe "lugar" para que a variável experimental tenha um efeito. Por exemplo, suponha-se que alguém estivesse estudando a eficácia de uma campanha publicitária para fazer as mulheres usarem perfume. Se, por acaso, fossem selecionados um grupo experimental e um grupo de controle em que 100 por cento das mulheres já usassem perfume antes do início da campanha de anúncios, seria impossível verificar a eficácia da propaganda, pois haveria a atuação do "efeito de teto". Uma medida "antes" permite verificar a possibilidade de mudança na variável dependente e considerá-la na avaliação dos efeitos da variável experimental.¹³

3. A hipótese a que se refere o estudo pode especificar a posição inicial na variável dependente, como uma das condições determinantes, ou como uma das condições em que é contingente o efeito de alguma outra variável. Por exemplo, a hipótese pode afirmar que um filme com uma mensagem de amizade interracial terá maior influência sobre pessoas que já são neutras ou favoráveis, com relação a negros, do que sobre aquelas que têm preconceito. Nesse caso, uma medida inicial do preconceito, juntamente com outra medida depois do filme, é evidentemente exigida pela hipótese.

4. Embora a hipótese principal do estudo não se refira à posição inicial na variável dependente, o pesquisador pode desejar analisar os dados de forma a verificar se o tratamento experimental tem efeitos diferentes em pessoas que, inicialmente, estavam em posições diferentes na variável dependente. Para fazer isso, evidentemente precisa ter uma medida da posição na variável dependente, antes de os sujeitos serem submetidos ao tratamento experimental.

5. Se não conseguiu distribuir as pessoas, com base puramente casual, pelas condições diferentes, pode desejar verificar sua comparabilidade inicial. Como já foi indicado,

¹³ Ver Hovland, Lumsdaine e Sheffield (1949) para uma discussão do problema de "efeitos de teto" e soluções propostas.

a impossibilidade de distribuir as pessoas com base puramente casual, pelas condições experimentais e de controle, é um afastamento com relação às exigências de um experimento ideal, mas às vezes são necessários compromissos nesse aspecto, sobretudo em estudos realizados em condições da vida real. Em tais casos, a prova — a partir de uma medida “antes” — de que o grupo experimental e o de controle estavam inicialmente iguais na variável dependente ajuda a aumentar a confiança de que uma diferença na medida “depois” se deve aos efeitos da variável experimental. No entanto, na ausência de casualização permanece a possibilidade de que os dois grupos fossem diferentes em outras características ou experiências passadas, que podem interagir com a variável experimental, de forma a provocar diferenças finais na variável dependente. Por isso, a prova de que os dois grupos eram inicialmente comparáveis em sua posição na variável dependente é apenas um substituto parcial para a distribuição casual.

Os estudos que empregam tanto medidas “antes” quanto “depois”, na variável dependente, podem obedecer a várias disposições com relação aos grupos de controle. (1) Apenas um grupo pode ser usado no estudo, e a medida “antes” serve como “controle”, na proporção em que se supõe que represente o nível da variável dependente, na ausência do tratamento experimental. (2) A medida “antes” pode ser obtida num grupo, e a medida “depois” obtida em outro, presumivelmente equivalente. (3) As medidas “antes” e “depois” podem ser obtidas no grupo experimental e no de controle. (4) Pode haver dois ou mais grupos de controle. Esses quatro modelos serão discutidos abaixo.

Qualquer que seja o modelo dos grupos de controle, o experimento “antes-depois”, como o “apenas depois”, dá provas de variação concomitante entre a variável independente e a dependente, através da comparação da ocorrência (ou extensão, ou aumento) de Y, no grupo que foi submetido a X, com a ocorrência (etc.) de Y no grupo que não foi submetido a X. Infere-se que Y não ocorreu antes de X, a partir da garantia, dada pela casualização, de que era pouco provável que os grupos fossem inicialmente diferentes em sua posição com relação a Y, além da proporção casual considerada no teste de significância. Essa equivalência inicial com relação a Y pode

ser verificada através da comparação das medidas “antes” dos dois grupos. Se não foi possível a distribuição casual, as medidas “antes” ainda dão provas quanto a diferenças em Y, anteriores às diferenças em X. Mas, se a distribuição casual não foi possível, não existe base para afastar a possibilidade da existência de diferenças, superiores às do acaso, em outros fatores que poderiam explicar uma diferença na posição com relação a Y, depois da exposição ao tratamento experimental.

As variações nas disposições dos grupos de controle referem-se a tentativas para considerar acontecimentos contemporâneos, processos de maturação e efeitos da mensuração inicial. Embora, em qualquer tipo de pesquisa social, o processo de mensuração possa influir na característica que está sendo medida, o planejamento “antes-depois” está especialmente sujeito a essa dificuldade. Por exemplo, a tentativa de medir as atitudes dos sujeitos antes do início do experimento pode cristalizar as atitudes; pode esgotar a boa vontade das pessoas, etc. A segunda mensuração, isto é, a mensuração “depois”, pode apresentar outros problemas: o sujeito pode estar aborrecido e, por isso, não ter disposição para responder; pode tentar dar respostas coerentes com suas respostas anteriores (e, dessa forma, reduzir a mudança aparente); ou pode tentar fazer que suas respostas fiquem “interessantes”, modificando-as de uma entrevista para outra (e, dessa forma, aumentar a mudança aparente). O processo de mensuração repetida pode também influir no “instrumento de medida”; por exemplo, durante as medidas repetidas, um observador pode ficar fatigado, aborrecido, mais ou menos sensível ao fenômeno que deve registrar.

As diferentes disposições de grupos de controle diferem quanto à proteção que oferecem contra a errada atribuição, à variável independente, de diferenças na variável dependente que podem, na realidade, ser devidas a outros acontecimentos contemporâneos, à maturação ou aos efeitos da medida inicial.

O estudo “antes-depois” com um único grupo. Barker, Dembo e Lewin (1941), em seu estudo sobre os efeitos da frustração no brincar de crianças pequenas, usaram um planejamento “antes-depois”, sem um grupo de controle. Cada criança era levada para uma sala, onde havia brinquedos simples, com os quais poderia brincar durante meia hora;

durante esse tempo, seu brinquedo era avaliado por um observador, de acordo com uma escala de "produtividade." A seguir, levantava-se uma separação; na parte da sala que a criança podia usar a partir desse momento, havia um conjunto de brinquedos atraentes e complexos. Quando a criança ficava totalmente integrada nesses novos brinquedos, o experimentador a segurava pela mão, levando-a de volta para a parte da sala em que havia brincado antes, e trancava os novos brinquedos com uma tela de arame, através da qual a criança continuava a vê-los. A atividade da criança com os brinquedos mais simples era novamente avaliada quanto à sua produtividade, durante um período de meia hora. A diferença nas avaliações de produtividade do brinquedo, durante o período de "pré-frustração" e o de "pós-frustração", era considerada como prova da quantidade de regressão provocada pela experiência frustradora. (No fim do experimento, a criança podia brincar quanto tempo quisesse com os brinquedos mais atraentes, a fim de desfazer os efeitos frustradores do experimento.)

Nesse planejamento, cada sujeito "serve como seu controle." A diferença entre sua posição na variável dependente, antes e depois da exposição à variável independente, é considerada como uma medida do efeito da variável independente. (Ver a Coluna 2 da tabela na pág. 126). No entanto, outras influências podem ter atuado entre as medidas "antes" e "depois." Os acontecimentos externos, não ligados ao tratamento experimental, podem provocar uma mudança na posição quanto à variável dependente; o mesmo pode ocorrer com os processos de crescimento e desenvolvimento. A mensuração inicial (nesse caso, o período de brinquedo em que se fez a avaliação inicial) pode provocar mudanças. Este planejamento não permite separar tais efeitos dos resultantes do tratamento experimental. Por isso, o seu uso é justificado apenas quando existem boas razões para acreditar (como o fizeram Barker, Dembo e Lewin): (1) que a medida "antes" não influirá de alguma forma, seja na resposta ao tratamento experimental, seja na medida "depois"; (2) que não há probabilidade de quaisquer outras influências, além do tratamento experimental, que durante o estudo possam influir na resposta do sujeito no momento da segunda mensuração. Para ter uma razoável segurança de que tais suposições sejam

justificadas, é preciso ter considerável conhecimento dos prováveis efeitos da mensuração e de outras condições, diferentes do tratamento experimental, e que tendem a influir na variável dependente. Isso pode ser verdade em muitos problemas de alguns campos — como aprendizagem e percepção sensorial — onde se fez muito trabalho experimental; é muito menos provável que isso ocorra, atualmente, na psicologia social e na sociologia.

*O estudo "antes-depois" com grupos intercambiáveis.*¹⁴ Uma forma de eliminar os efeitos da mensuração inicial é medir um grupo antes da introdução do fator experimental, e medir outros grupos depois da exposição ao fator experimental. Os dois grupos são antecipadamente selecionados na população que deve ser submetida à variável experimental; tal como ocorre em outros planejamentos, a seleção casual assegura que, antes da introdução da variável experimental, os grupos provavelmente não diferem mais que uma quantidade especificável e que, portanto, podem ser tratados como intercambiáveis. Também aqui, o emparelhamento pode ser usado para complementar a casualização. Supõe-se que a diferença entre a medida "antes", obtida no primeiro grupo (Y_1), e a medida "depois", obtida no segundo grupo (Y_2), seja uma medida do efeito do fator experimental. (Ver Coluna 3 da tabela na pág. 126).

Esse planejamento foi usado num estudo de campanha de publicidade sobre as Nações Unidas, na cidade de Cincinnati (Star e Hughes, 1950). Duas amostras equivalentes, cada uma com mil pessoas, foram retiradas da população. Uma foi entrevistada antes do início da campanha de publicidade; a outra, dois meses depois. Para verificar a eficácia da campanha, foram comparadas as respostas dos dois grupos. Segundo se verificou, havia uma diferença muito pequena entre elas.

O estudo "antes-depois" com grupos intercambiáveis elimina a possibilidade de confundir um efeito da mensuração inicial com o da variável experimental. Suponha-se que

¹⁴ Na edição anterior deste livro, isso foi denominado planejamento "antes e depois simulado". D. T. Campbell, em artigo ainda inédito ("Quasi Experimental Designs for Use in Social Science Settings"), refere-se a ele como planejamento "antes e depois contra-balançado".

o mesmo grupo tivesse sido entrevistado antes e depois da campanha. A entrevista inicial poderia ter despertado seu interesse pelas Nações Unidas e, dessa forma, ter feito com que se tornassem muito sensíveis à campanha de publicidade. Se isso acontecesse, teria sido enganador um simples estudo "antes-depois", no qual a diferença entre as respostas "antes" e "depois" de um único grupo entrevistado fosse considerada como a medida do efeito da variável experimental (a campanha de publicidade).

De outro lado, o estudo "antes-depois" com grupos intercambiáveis não apresenta proteção contra o efeito de outros acontecimentos que se dêem entre as duas mensurações. Suponha-se que se encontrasse uma diferença entre a resposta "antes" de um grupo e a resposta "depois" do outro grupo; suponha-se, também, a existência de acontecimentos internacionais — em que a ONU desempenhasse um papel muito notável — entre as duas medidas. Este planejamento não teria maneira de verificar se a mudança na resposta se devia à campanha de publicidade, aos acontecimentos internacionais, ou a uma combinação de ambas as coisas.

O estudo "antes-depois" com um grupo de controle. A fim de tentar considerar os efeitos da mensuração inicial e dos fatores contemporâneos, freqüentemente se inclui um grupo de controle no planejamento "antes-depois." Esse planejamento é apresentado na Coluna 4 da tabela da pág. 126. Nesse estudo, o grupo de controle e o grupo experimental são medidos no início e no fim do período experimental. Evidentemente, a variável experimental é introduzida apenas no grupo experimental. Como o grupo de controle e o experimental são submetidos à mensuração inicial e às influências contemporâneas, a diferença entre os resultados dos dois grupos deve constituir uma medida da eficácia da variável experimental isolada. Para isso, é possível comparar os resultados finais dos dois grupos (Y_2 e Y_2') ou seus resultados de mudança (d e d').

Outro estudo sobre os efeitos do filme *A Batalha da Inglaterra*, paralelo ao estudo de "apenas depois", descrito nas págs. 127-128 apresenta um exemplo do estudo "antes-depois" com um grupo de controle. Num acampamento diferente, os efeitos do filme foram verificados com esse planejamento. O processo foi idêntico ao já descrito, mas, duas

semanas antes da exibição do filme ao grupo experimental, membros dos dois grupos foram solicitados a preencher o questionário como parte de "um levantamento." Depois, o filme foi exibido para o grupo experimental, durante os seus processos normais de treinamento; evidentemente, não foi exibido para o grupo de controle. Aproximadamente uma semana depois da exibição do filme, os soldados de ambos os grupos foram novamente solicitados a preencher um questionário, que repetia os itens fundamentais do instrumento "antes", mas que era apresentado como versão "revista" do instrumento de levantamento, modificado a partir da primeira aplicação. Para confirmar essa explicação, seu formato foi modificado, e alguns itens de "camuflagem" foram omitidos, enquanto que houve o acréscimo de outros. As diferenças entre as respostas "depois" do grupo experimental e do de controle foram consideradas como a medida dos efeitos do filme.

Se a medida "antes" e os acontecimentos diferentes do tratamento experimental influem da mesma maneira no grupo de controle e no grupo experimental, um planejamento "antes-depois", com um grupo de controle, apresenta garantias adequadas contra a atribuição, à variável experimental, de uma diferença na medida "depois" que é, na realidade, devida aos efeitos da mensuração inicial ou aos acontecimentos contemporâneos. No entanto, pode ocorrer que a mensuração inicial ou os acontecimentos externos *interajam* com a variável experimental, de forma que se mude o seu efeito. Por exemplo, a comparação dos dois estudos sobre o filme *A Batalha da Inglaterra* mostrou uma interação entre a mensuração inicial e a variável experimental; no estudo "antes-depois" houve menos diferença entre as respostas "depois" do grupo experimental e do grupo de controle que no estudo "apenas depois". Os pesquisadores sugeriram que um esforço para ser coerente nas respostas às duas versões do questionário do levantamento podem ter reduzido as mudanças de opinião, que, de outra forma, poderiam ter sido provocadas pelo filme, ou que o fato de, explicitamente, assumir uma posição ou outra, no primeiro questionário, pode ter feito aparecer uma tendência para cristalizar opiniões que, de outra forma, poderiam ser mais fluidas.

Quando existe essa interação entre a medida inicial ou os acontecimentos contemporâneos e a variável experimental, um estudo "antes-depois", com um grupo de controle, não apresenta uma base firme para inferências a respeito do efeito da variável experimental, pois esse planejamento não apresenta qualquer forma para descobrir a interação e, assim, separar o efeito isolado da variável experimental. Solomon (1949) criou planejamentos mais complexos para possibilitar a consideração de tais interações. Tais planejamentos exigem o emprego de outros grupos de controle.

O estudo "antes-depois" com dois grupos de controle. Este planejamento, apresentado na Coluna 5 da tabela, permite separar a influência da variável experimental e a influência da mensuração inicial, ainda que exista interação entre elas. Exige a adição de um segundo grupo de controle, que não é pré-medido, mas que é submetido à variável experimental e a uma mensuração "depois". Tal como no planejamento "antes-depois" com um grupo de controle, o primeiro grupo de controle é submetido à medida "antes" e à medida "depois", mas não submetido à variável experimental. Os três grupos (o experimental e os dois de controle) devem, naturalmente, ser escolhidos de tal forma que, no início do experimento, sejam diferentes apenas por acaso. Tal como ocorre nos outros planejamentos experimentais, isso significa que devem ser constituídos por distribuição casual ou por algum processo de emparelhamento, suplementado por casualização. Se isso foi feito, podemos supor que a pré-medida para o grupo de controle II teria sido semelhante aos outros resultados de pré-teste, isto é, o resultado inicial médio para o grupo experimental e o grupo de controle I. Este resultado inicial inferido para o grupo de controle II é indicado por Y''_1 na Coluna 5. Isso permite um grupo com uma medida "antes" (suposta) e uma medida "depois", que recebeu o tratamento experimental, mas para a qual não existe possibilidade de interação entre o tratamento experimental e os efeitos da medida "antes".

Pode-se notar que o tratamento do segundo grupo de controle neste planejamento (sem pré-medida, mas exposição à variável experimental e uma medida posterior) é igual ao tratamento do grupo experimental no planejamento "apenas depois" e do grupo que é medido apenas depois da exposição

à variável experimental, no planejamento "antes-depois" com grupos intercambiáveis. No entanto, o planejamento de dois grupos de controle não apenas apresenta um grupo para o qual não existe possibilidade de interação entre a pré-medida e o tratamento experimental, mas também apresenta uma medida dessa interação, através da comparação do grupo experimental com os dois grupos de controle.

Se existe razão para acreditar que os acontecimentos contemporâneos e os processos de desenvolvimento provavelmente não representam influências importantes em determinado estudo, o planejamento "antes-depois" com dois grupos de controle pode ser interpretado da seguinte maneira: (1) a mudança no grupo de controle II (d'') é devida apenas à variável experimental; (2) a mudança do grupo de controle I (d') é devida apenas aos efeitos da pré-mensuração; (3) se o resultado de mudança do grupo experimental (d) é diferente da soma dos resultados de mudança dos dois grupos de controle ($d' + d''$), isso é um reflexo da interação entre a pré-medida e a variável experimental. Essa interação pode acentuar ou reduzir os efeitos da variável experimental.

Embora geralmente se reconheça que a solução de Solomon é correta, seu planejamento não tem sido muito empregado. No entanto, alguns poucos estudos que o empregaram sugerem que os efeitos da interação são, frequentemente, tão grandes que têm efeitos marcantes nos resultados experimentais. Por exemplo, Canter (1951) usou este planejamento de três grupos num estudo sobre a eficiência de um curso de treinamento em relações humanas para supervisores. Antes do curso de treinamento, os membros do grupo experimental responderam a vários testes de informação e opinião, fizeram o curso e depois novamente responderam aos questionários. O primeiro grupo de controle respondeu aos testes nos mesmos momentos que o grupo experimental (isto é, antes e depois do treinamento), mas não recebeu o programa de treinamento. O segundo grupo de controle não foi pré-medido, mas recebeu o programa de treinamento e foi submetido à mensuração "depois". Ao comparar as mudanças nos resultados dos três grupos, Canter verificou que cada um dos testes da pré-medida interagiu de alguma forma com a variável experimental; alguns dos testes iniciais pareciam tornar os supervisores mais receptivos aos aspectos

significativos do curso de treinamento; outros pareciam torná-los menos receptivos.

O estudo "antes-depois" com três grupos de controle. Solomon sugeriu complexidade ainda maior para esse planejamento, a fim de conseguir garantia, quando se pode esperar que os acontecimentos contemporâneos ou as mudanças de desenvolvimento influam nos resultados experimentais. Isso exige a adição de um terceiro grupo de controle, como se vê na Coluna 6 da tabela. Ainda aqui, todos os grupos devem ser selecionados de forma que sejam diferentes apenas por acaso. Neste planejamento, o grupo experimental e o grupo de controle I são pré-medidos. Os grupos de controle II e III não são pré-medidos; supõe-se que seus resultados de pré-mensuração seriam iguais à média dos resultados combinados do grupo experimental e do grupo de controle I. A variável experimental é introduzida no grupo experimental e no grupo de controle II; está ausente nos grupos de controle I e III. Supõe-se que os quatro grupos estejam igualmente submetidos aos acontecimentos contemporâneos, e os quatro são medidos depois do experimento.

Nesse planejamento, a mudança no grupo de controle III (isto é, a diferença entre a medida posterior e a pré-medida inferida) representa o efeito de acontecimentos contemporâneos ou de processos de desenvolvimento, pois nesse grupo não havia a pré-medida, nem a variável experimental. A mudança no grupo de controle II representa os efeitos da variável experimental e dos acontecimentos contemporâneos ou de processos de desenvolvimento; a mudança no grupo de controle I, os efeitos da pré-mensuração e de acontecimentos contemporâneos ou de processos de desenvolvimento. O efeito isolado do fator experimental pode ser verificado subtraindo-se a mudança no grupo de controle III (o resultado de acontecimentos contemporâneos e de processo de maturação) da mudança no grupo de controle II (o resultado de acontecimentos contemporâneos, processos de maturação e variável experimental). Pode-se verificar até que ponto a mudança no grupo experimental reflete os efeitos de pré-medida, da variável experimental, de acontecimentos incontrolados ou da interação dos fatores, através da comparação da mudança nesse grupo com as dos outros.

Pode-se observar que esse planejamento de quatro grupos corresponde a fazer duas vezes o experimento: uma vez com o planejamento "antes-depois" com um grupo de controle (grupo experimental e grupo de controle I), e uma vez com um planejamento "apenas depois" (grupos de controle II e III). Se os resultados desses dois experimentos são coerentes, temos maior segurança — de que o que foi observado não é artifício — que com uma das versões isolada, pois repetimos a verificação dentro do estudo.

VERIFICAÇÃO DE DOIS OU MAIS TRATAMENTOS EXPERIMENTAIS

Todos os exemplos até aqui considerados exigem a comparação de um grupo experimental, submetido a determinado tratamento, com um grupo de controle, não submetido a tal tratamento. Alguns deles, como foi indicado, permitem separar os efeitos do tratamento dos efeitos da mensuração inicial, da maturação ou de acontecimentos contemporâneos. No entanto, os planejamentos até aqui descritos não são adequados para a verificação de hipóteses a respeito da contribuição conjunta de duas ou mais variáveis independentes que influem numa variável dependente, ou para comparar os efeitos de dois ou mais tratamentos experimentais, com a ausência de tratamento experimental. Por exemplo, podem dar uma resposta a uma pergunta como a seguinte: Será que uma campanha de publicidade sobre as Nações Unidas muda as atitudes com relação a essa organização? No entanto, em sua forma simples, não dão respostas eficientes a algumas perguntas, tais como as seguintes: o que é mais eficiente para mudar atitudes: aulas, fitas de cinema, programas de televisão ou discussões de grupo? Ou, então: Será que uma campanha de publicidade tem maior probabilidade de mudar atitudes de pessoas que inicialmente eram desfavoráveis, neutras ou favoráveis? Todavia, todos os planejamentos básicos podem ser adaptados para responder eficientemente a tais perguntas, se o experimento for planejado de forma a permitir uma *análise da variância*.

Suponhamos que estejamos interessados em verificar qual, dentre três técnicas para responder a observações de preconceito, é mais eficiente para influir em espectadores:

(1) um apelo à inteligência, dizendo-se que o preconceito é uma exibição de ignorância e estupidez; (2) um apelo à "justiça" e à tradição democrática; (3) um apelo à noção de diferenças individuais — isto é, que existem indivíduos "bons e maus" em todos os grupos. Podemos decidir, também, incluir uma situação em que não se dá resposta; isso equivale a acrescentar um grupo de controle. Embora seja possível dispensar os grupos de controle em alguns experimentos comparativos — pois os grupos contrastados podem ser considerados como "controles" uns dos outros —, esse processo não é recomendado. O uso de grupos de controle em experimentos comparativos estabelece uma linha de base com a qual é possível avaliar os diversos métodos que estão sendo estudados. Por exemplo, uma comparação entre duas técnicas de anúncios pode revelar que uma provoca mais vendas do que outra. Este "fato" pode adquirir significação inteiramente diversa se se verifica que um grupo de controle, não submetido aos anúncios, compra mais do que qualquer um dos grupos experimentais.

Voltando às respostas diante de observações de preconceito, suponhamos também a suspeita de que cada uma das técnicas de resposta possa ter diferentes resultados em diferentes circunstâncias; por exemplo, uma resposta pode ser melhor quando a pessoa reage de maneira indignada e emocional; outra, quando suas maneiras são calmas e altivas.

Oito grupos de pessoas são necessários para o experimento. Os membros do conjunto de pessoas são casualmente distribuídos pelos oito grupos. Cada grupo vê uma cena na qual se faz uma observação com preconceito, mas a resposta apresentada a cada grupo inclui uma diferente combinação de "conteúdo" e "maneira", como se vê na tabela (pág. 141). Suponhamos o emprego de uma escala de atitudes para medir a resposta dos espectadores. Se o estudo é estabelecido como um experimento "apenas depois", as diferenças, entre os grupos, nos resultados nessa única aplicação da escala, serão usadas como medida dos efeitos das variáveis experimentais. Se o estudo é estabelecido como um experimento "antes-depois", o efeito das variáveis experimentais pode ser avaliado através da comparação dos resultados finais dos diferentes grupos ou das diferenças na extensão da mudança entre as medidas "antes" e "depois".

Grupo	Conteúdo da resposta	Maneira de responder
1	Inteligência	Indignação
2	Inteligência	Calma, altivez
3	Justiça	Indignação
4	Justiça	Calma, altivez
5	Diferenças individuais	Indignação
6	Diferenças individuais	Calma, altivez
7	Sem resposta	Afastamento encolerizado, com expressão de aversão
8	Sem resposta	Agir como se nada tivesse sido dito

Com tal planejamento, é possível responder, num único estudo, a diferentes questões de pesquisa, através da comparação de diferentes agrupamentos de sujeitos. A fim de medir a relativa eficiência do apelo à inteligência, à justiça ou ao conceito de diferenças individuais, ou do silêncio, calcula-se e compara-se o resultado médio de sujeitos nos grupos 1 e 2 (inteligência), nos grupos 3 e 4 (justiça), nos grupos 5 e 6 (diferenças individuais), etc. De forma semelhante, para medir a relativa eficiência de respostas dadas de maneiras diferentes, compara-se o resultado médio de indivíduos nos grupos 1, 3, 5 e 7 (indignação) com o de indivíduos nos grupos 2, 4, 6 e 8 (calma). A fim de verificar se a eficiência dos vários conteúdos é influenciada pela sua maneira de apresentação, comparam-se os resultados dos oito grupos separados. Se as diferenças entre eles são maiores que as explicadas por diferenças de conteúdo e de maneira, consideradas separadamente, conclui-se que a eficiência dos vários conteúdos é influenciada pela maneira de apresentação. Por exemplo, pode-se verificar que o argumento de "justiça" é mais eficiente quando apresentado com calma, e o argumento de "inteligência" quando apresentado com indignação. A técnica estatística que permite a pesquisa dessas três questões, num único estudo, é conhecida como *análise de variância*.¹⁵

¹⁵ Os planejamentos deste tipo e as técnicas estatísticas para sua análise foram criados por R. A. Fisher e são discutidos em seu livro, *The Design of Experiments* (1951). Discussões mais simples podem ser encontradas em Cochran e Cox (1957), Edwards (1950), Lindquist (1953) e outros manuais de estatística.

Evidentemente, um planejamento que permite uma análise de variância torna possível um estudo de interrelações complexas. Permite também um planejamento eficiente; isto é, permite conclusões mais precisas sobre maior número de hipóteses, com menor número de casos, que se as hipóteses fossem verificadas em estudos separados.

PLANEJAMENTO REPRESENTATIVO

Brunswik (1956) apresentou uma crítica radical aos tipos de planejamento experimental que descrevemos neste capítulo. Apresenta a opinião de que os planejamentos clássicos de experimentos têm a consequência involuntária de artificialmente "ligar" ou "desligar" variáveis, de uma forma que não é representativa da sua maneira de existir na realidade. Assim, segundo Brunswik (1955), os planejamentos experimentais clássicos na medicina conduziram, às vezes, a práticas prejudiciais. Por exemplo, a pesquisa que demonstrou que a ebulição de líquidos destruiria as bactérias levou à prática de ferver o leite, a fim de eliminar os germes. A fervura do leite, no entanto, não apenas destrói bactérias, mas também provoca a perda de vitaminas. A tese de Brunswik é que o usual planejamento experimental sistemático, que tenta ligar uma variável independente a uma variável dependente, desequilibra o quadro da interação de fatores: a ligação entre a fervura do leite e sua perda de vitaminas é ignorada, se o estudo focaliza apenas a ligação entre ebulição e antissepsia.

De forma semelhante, nas ciências sociais seria possível citar pesquisas em que diferentes tipos de estilo de liderança, ou diferentes tipos de apresentação visual, etc., têm sido estudados em experimentos de laboratório muito controlados, com grande precaução para manter constantes outras variáveis. A crítica de Brunswik a esses experimentos seria que o estilo de liderança, ou determinado tipo de apresentação visual, ou qualquer outra variável a ser estudada, ocorrem em diferentes contextos; para compreender, por exemplo, a relação entre "liderança liberal" e "participação dos membros nas questões do grupo", é preciso estudar uma amostra representativa de situações em que ocorre a liderança. Pode haver muitas e diferentes maneiras para atingir elevado grau de participação

dos membros do grupo. A "liberalidade" pode provocá-la em um contexto (por exemplo, com membros muito motivados e muito inteligentes) e não conseguir provocá-la em outro contexto; a liderança autoritária pode ser eficiente ou ineficiente em diferentes circunstâncias.

Na realidade, Brunswik deseja um planejamento de pesquisa de tipo de levantamento, em que se faça uma amostragem representativa de situações. Este tipo de planejamento significa que as variáveis estudadas podem ser pesquisadas no contexto da variação concomitante e natural de outros fatores. O uso desse planejamento exige que se abandonem as noções usuais de controle experimental; precisamos considerar a interação entre variáveis, encontrada nas situações incluídas na amostra. Os instrumentos básicos de análise são os processos de correlação parcial e múltipla. A produtividade da metodologia proposta por Brunswik foi demonstrada em experimentos de percepção (Brunswik, 1956) e algumas investigações de psicologia clínica (Hammond, 1955).

As críticas de Brunswik ao planejamento experimental sistemático tiveram o efeito salutar de acentuar algumas interpretações erradas e algumas generalizações excessivas e errôneas, comumente feitas a partir de experimentos tradicionais. Além disso, acentuaram a necessidade de uma amostragem adequada de condições e situações experimentais, bem como de amostragem adequada de pessoas, para que seja possível fazer interpretações e generalizações apropriadas. No entanto, existem limitações em sua metodologia, o que se torna evidente quando tentamos responder a algumas das questões necessárias para a realização de um "planejamento representativo". Por exemplo, se se deseja estudar as condições que conduzem a "ativa participação dos membros nas questões do grupo", seria necessário compilar uma lista completa de situações de grupo em que pode ocorrer a participação dos membros e estudar uma amostra representativa desse universo, realizando-se mensurações de todas as variáveis importantes de cada situação que apareceu na amostra. As dificuldades teóricas para a compilação dessa lista e para a identificação das variáveis importantes para a participação dos membros, além das dificuldades práticas para o estudo de uma amostra representativa de tais situações

de grupo, são, naturalmente, insuperáveis em qualquer pesquisa.

Do ponto de vista prático, o pesquisador sempre precisa desejar menos que uma investigação de todas as condições importantes dos fenômenos que deseja compreender. Um componente da arte de pesquisa experimental está em ter "sensibilidade" para o assunto, ser sensível para a natureza dos fenômenos pesquisados, de forma a considerar as dimensões significativas. A pesquisa científica, como qualquer outra atividade humana significativa, exige julgamento. Evidentemente, não existem garantias de que mesmo os julgamentos bem disciplinados sejam corretos ou produtivos.

*Inferência Causal a partir de Outros Planejamentos de Estudo*¹⁶

O pesquisador pode não ter possibilidade de verificar uma hipótese através da distribuição dos sujeitos por diferentes condições, na qual controla diretamente a suposta variável causal. Considere-se, por exemplo, a pesquisa sobre a relação entre o cigarro e o câncer de pulmão em seres humanos. A quantidade de cigarros não tem sido controlada, como o seria num estudo experimental, através da distribuição de diferentes indivíduos por condições de diferentes números de cigarros; por muitas razões, é pouco provável que se faça experimento desse tipo com seres humanos. Ao contrário, obtém-se um registro (ou uma estimativa retrospectiva) de quanto um indivíduo fumou, e se tem câncer de pulmão; calcula-se depois a relação entre a quantidade de cigarros e a ocorrência de câncer no pulmão.

No entanto, como já foi indicado, a existência de uma correlação entre duas variáveis não demonstra, necessariamente, que uma é "causa" da outra. Por isso, precisamos, pelo menos, considerar a possibilidade de que os tipos de pessoas que se tornam fumantes são também, por uma razão

¹⁶ Para discussão minuciosa e esclarecedora da verificação de hipóteses causais através de estudos não-experimentais, ver Hyman (1955, Parte III). Nossa discussão do assunto deve muito à apresentação de Hyman.

ainda desconhecida, os tipos de pessoas que apresentam câncer no pulmão. Podemos lembrar que, antes de o fumo tornar-se comum entre estudantes universitários, verificou-se que os fumantes obtinham notas piores que os não-fumantes. Finalmente, tornou-se claro que, pelo menos nessa época, os estudantes que fumavam eram também os que tinham tendência para negligenciar os estudos.

Qualquer que seja a forma de um estudo, para que apresente uma verificação de uma hipótese causal, precisa apresentar bases para inferências a respeito da causalidade e garantias contra inferências não-justificadas. Os estudos não-experimentais não podem apresentar garantias tão adequadas quanto as apresentadas por distribuição casual de sujeitos por grupos experimentais e grupos de controle, manipulação direta da variável experimental e controle de algumas das variáveis externas que poderiam atuar durante a realização do experimento. Quais as garantias substitutas de que dispomos? O pesquisador pode substituir a manipulação direta da variável experimental por uma ou várias direções de provas: comparação entre pessoas submetidas a experiências contrárias, tentativas de verificar a ordem temporal de variáveis associadas, exame da relação entre variáveis, através do modelo de relações que poderia ser esperado se uma ou outra fosse o fato causal. A distribuição de sujeitos por grupo de controle e grupo experimental pode ser substituída por provas que apresentem uma base para a inferência de que os grupos submetidos a experiências contrárias eram ou não diferentes, antes de tais experiências; ou o pesquisador pode selecionar, de seu grupo total, sub-amostras emparelhadas, através de algumas características, mas com experiências contrárias; ou pode limitar sua amostra a pessoas com determinadas características. O pesquisador pode substituir o controle direto de variáveis externas, passadas ou contemporâneas, pela reunião de dados sobre outras características ou experiências de seus sujeitos que, segundo acredita, podem ser importantes para a posição na variável dependente, e usa tais dados em sua análise. Esta última operação será discutida no Capítulo 11. No resto deste capítulo serão discutidos vários processos destinados a substituir a manipulação direta da suposta va-

riável causal e a distribuição casual de sujeitos por grupos de controle e grupos experimentais.

SUBSTITUTOS PARA A MANIPULAÇÃO DIRETA DA SUPOSTA VARIÁVEL CAUSAL

COMPARAÇÃO DE GRUPOS SUBMETIDOS A EXPERIÊNCIAS CONTRÁRIAS. Se um pesquisador não tem possibilidade de distribuir os sujeitos por diferentes grupos, um dos quais será submetido a determinado tratamento, enquanto outro não o será, uma óbvia solução substituta é localizar grupos de pessoas que serão ou foram submetidas a experiências diferentes com relação à suposta variável causal em que o pesquisador está interessado. Por exemplo, Cook, Havel e Christ (1957), ao planejar um estudo sobre os efeitos de programa de orientação de verão para estudantes estrangeiros nos Estados Unidos, sabiam antecipadamente quais os estudantes destinados a freqüentar os centros de orientação, embora não tivessem interferido em sua escolha. A partir de todas as fontes disponíveis, compilaram uma lista de estudantes estrangeiros que iriam começar seus estudos formais nos Estados Unidos — ao mesmo tempo que os estudantes destinados aos centros de orientação — mas que não receberiam tal orientação. Os estudantes de ambos os grupos eram semelhantes pelo fato de irem aos Estados Unidos com alguma forma de bolsa de estudos oferecida pelo governo norte-americano. A partir da lista de estudantes que deveriam freqüentar os centros de orientação, os pesquisadores escolheram estudantes de lista comparável de não-orientação, a partir de controle de distribuição de freqüência, e emparelharam, grosseiramente, as duas amostras, através de algumas características, tais como nacionalidade, idade e campo de estudo. Como os estudantes foram antecipadamente identificados, foi possível obter medidas iniciais, logo que chegaram aos Estados Unidos, e assim verificar se os dois grupos eram inicialmente semelhantes em sua posição com relação às variáveis dependentes. Tal estudo se aproximava de um experimento "antes-depois", com um grupo de controle: os estudantes que freqüentavam os centros de orientação representavam o grupo "experimental"; os que não os freqüentavam representavam o grupo de "controle". No

entanto, havia uma limitação importante: o grupo "experimental" e o conjunto de sujeitos disponíveis para o grupo de "controle" foram constituídos a partir de considerações administrativas, referentes ao tipo de bolsa de estudos que receberam, e, por isso, podiam diferir sistematicamente. Num experimento "puro", o grupo total de estudantes teria constituído um conjunto a partir do qual os indivíduos seriam casualmente distribuídos (com ou sem processos suplementares de emparelhamento) a fim de freqüentar ou não um centro de orientação.

Ao contrário do que ocorreu com Cook, Havel e Christ, freqüentemente um pesquisador não tem possibilidade de saber, antecipadamente, quais os indivíduos que serão e quais não serão submetidos a determinada experiência. Em vez disso, pode localizar um grupo de pessoas submetidas ao tipo de experiência em que está interessado, e um grupo, sob outros aspectos semelhante, que não foi submetido a tal experiência; compara os dois grupos numa base *ex post facto*. O estudo de Deutsch e Collins (1951), sobre atitudes de locatários brancos com relação a negros, em conjuntos residenciais públicos (descrito no Capítulo 1), seguiu esse processo. Os pesquisadores selecionaram conjuntos residenciais diferentes quanto à variável independente em que estavam interessados — padrão de residência. Nesses conjuntos, escolheram, casualmente, donas de casa brancas. Dessa forma, aproximaram-se de um grupo "experimental", submetido à experiência de viver com negros, como vizinhos num padrão não-segregado de moradia, e um grupo de "controle", que não tinha tido essa experiência. Ao selecionar, como seu grupo de "controle", residentes de conjuntos residenciais bi-raciais, nos quais brancos e negros eram segregados — e não outros moradores da cidade —, controlaram certas variáveis externas, tais como o nível sócio-econômico das pessoas e a experiência de viver num conjunto residencial. Os estudos que empregam esse modelo *ex post facto* têm uma limitação ainda mais séria que os estudos semelhantes à pesquisa sobre estudantes estrangeiros, há pouco descrito; não apenas os sujeitos não são casualmente distribuídos pelas condições diferentes, mas não existe possibilidade de mensuração preliminar, a fim de verificar se os dois grupos eram inicialmente semelhantes em sua posição com relação à variável dependente ou com

relação a determinadas características que, segundo se acredita, são importantes para ela. Por tal razão, esses estudos devem permitir a coleta de dados a partir dos quais se possa inferir se os dois grupos eram ou não inicialmente semelhantes. Depois, neste capítulo, serão discutidos os tipos de provas que podem apresentar uma base para tal inferência.

PROVA DA ORDEM TEMPORAL DE VARIÁVEIS. Como já foi dito, um tipo de informação que pode auxiliar a dar fundamento para uma inferência de que uma variável especificada (X) é a causa de outra (Y), é a prova de que Y não ocorreu antes de X . Em alguns casos, a relação temporal é tão clara que não existe necessidade de provas complementares. Por exemplo, se se verifica que os indivíduos nascidos em diferentes países são diferentes quanto a alguma questão política, não cabe discutir qual das variáveis — nacionalidade ou opiniões políticas — é anterior no tempo. Frequentemente, no entanto, a relação temporal entre as duas variáveis não é tão clara. Embora uma pareça anterior à outra, isso pode não ser verdade. Por exemplo, num estudo sobre o efeito de experiências infantis no ajustamento adulto, um pesquisador pode confiar nas descrições que os adultos dão de sua infância. Evidentemente, as experiências da infância são anteriores, mas a seleção e interpretação de acontecimentos podem ser coloridas pela disposição atual, e disso resulta que a variável aparentemente anterior, sob a forma em que é medida, é na realidade um reflexo da variável posterior. Em outros casos, pode não haver qualquer base — sequer uma base enganadora — para julgar as relações temporais; isso tende a ocorrer, por exemplo, quando lidamos com duas atitudes relacionadas entre si.

O pesquisador pode introduzir vários processos, numa tentativa de conseguir provas sobre a relação temporal entre a ocorrência de variáveis. Nos parágrafos seguintes são discutidos dois desses processos — interrogar as pessoas quanto às relações temporais e obter provas durante um certo período de tempo.

Perguntas às pessoas, a respeito de relações temporais. O pesquisador pode incluir, em seu questionário ou em suas perguntas de entrevista, questões a respeito do período em que certas coisas aconteceram, ou perguntar como a pessoa

se sentia a respeito de alguma coisa antes de determinado acontecimento, ou se houve quaisquer mudanças em seus sentimentos. Por exemplo, Deutsch e Collins apresentaram algumas perguntas, tais como as seguintes: "Antes de você se mudar (para o conjunto residencial), o que é que você pensava desta residência? (...) Desde que você se mudou (para o conjunto residencial), você gostou mais, a mesma coisa ou menos da idéia de morar num conjunto residencial em que há famílias de cor e famílias brancas? (...) Houve muitas mudanças em suas idéias a respeito de pessoas de cor, desde que você mora no conjunto residencial? (...) (Se houve alguma mudança) De que maneira mudaram suas idéias?"

Naturalmente, sempre existe o perigo de que as respostas a tais perguntas não sejam corretas. As pessoas podem não lembrar; suas atitudes atuais podem deformar suas recordações, etc. Embora não seja possível verificar definitivamente a correção de tais descrições retrospectivas, às vezes o pesquisador pode criar verificações indiretas da probabilidade de deformação. Por exemplo, Deutsch e Collins compararam as respostas — das pessoas nos quatro conjuntos residenciais estudados — à pergunta sobre as atitudes anteriores com relação aos negros. Outras perguntas da entrevista mostraram notáveis diferenças, entre locatários de conjuntos integrados e segregados, no momento do estudo. Os pesquisadores pensaram que, se as atitudes atuais estivessem deformando a recordação, isso deveria conduzir a diferenças coerentes entre os locatários dos dois tipos de conjuntos residenciais, em suas descrições das atitudes anteriores à mudança. Segundo se verificou, embora os dois conjuntos residenciais segregados fossem muito semelhantes na proporção de locatários que descreviam atitudes favoráveis, neutras e desfavoráveis com relação aos negros — antes de se mudarem —, não havia essa semelhança entre os dois conjuntos residenciais integrados. Em um deles, os que descreviam atitudes inicialmente favoráveis aos negros eram mais numerosos que nos conjuntos residenciais segregados; todavia, o outro conjunto tinha a maior proporção dos que descreviam atitudes inicialmente desfavoráveis. Do fato de as descrições dos moradores nos conjuntos residenciais integrados e segregados — a respeito de suas atitudes iniciais — não apresentarem

diferenças *coerentes*, enquanto que isso ocorria em suas expressões atuais de atitudes, os pesquisadores concluíram que a recordação não era sistematicamente deformada por atitudes no momento do estudo. Além disso, verificaram que as mulheres semelhantes em educação, religião e atitudes políticas davam descrições semelhantes de sua atitude inicial com relação aos negros, qualquer que fosse o conjunto residencial em que viviam; essa coerência foi considerada como outra prova para desmentir a deformação da recordação por atitudes atuais.

Obtenção de provas através de estudos em certo período de tempo. Nos estudos limitados a uma única entrevista ou observação, ou outra medida de cada pessoa estudada, e nos quais o pesquisador não dispõe de informação complementar sobre as experiências do indivíduo, existe pouca possibilidade de obtenção de provas de seqüências temporais, a não ser através da recordação das pessoas quanto ao momento em que as coisas aconteceram. No entanto, em estudos que focalizam as mesmas pessoas durante certo período de tempo, o pesquisador pode conseguir provas de relações temporais entre variáveis. Esses estudos longitudinais podem apresentar a forma de observações repetidas das mesmas pessoas, ou de repetidas entrevistas com elas,¹⁷ ou ainda de diferentes processos de medida em épocas diferentes.

Stouffer *et al.* (1949a) apresentam um exemplo de estudo que emprega diferentes tipos de dados a respeito das mesmas pessoas, em diferentes momentos. Os pesquisadores estavam interessados na relação entre a aceitação do sistema oficial de valorização do Exército e a promoção. Se tivessem apenas entrevistado uma amostra de pessoal do Exército e verificado que os de posto mais elevado exprimem atitudes e opiniões mais de acordo com os valores oficiais do Exército, não teriam bases para inferir se a aceitação do sistema oficial de valores facilita a promoção ou se o fato de ser promovido aumenta a aceitação do sistema. Para evitar esse problema, entrevistaram um grupo de soldados recentemente convocados, através de questões com as quais poderia ser construído um índice de "aceitação do sistema de valores do

17 Para discussão minuciosa de estudos que empregam entrevistas repetidas com as mesmas pessoas ("estudos de painel"), ver Rosenberg, Thielens e Lazarsfeld (1951).

Exército." Quatro meses depois, examinaram os registros de Exército desses mesmos homens, e verificaram que maior proporção dos que tinham apresentado opiniões de acordo com os valores do Exército tinham-se tornado soldados de primeira classe. Assim, estava claro que o conformismo com o sistema de valores do Exército facilitava a promoção. (Evidentemente, existe muita probabilidade de que a relação entre essas duas variáveis seja mutuamente reforçadora; pesquisas posteriores poderiam mostrar que, depois da promoção, as opiniões estariam ainda mais de acordo com a orientação oficial do Exército).

BUSCA DE PADRÕES DE RELAÇÕES, INFERIDOS DE SUPOSIÇÕES CAUSAIS CONCORRENTES As vezes, é possível inferir qual, de dois fatores que variam juntamente, é o "causal", supondo-se que as duas variáveis mostrariam certo padrão de associação se X fosse a "causa", e um padrão diferente se Y fosse a "causa." Por exemplo, às vezes é razoável esperar que, se X fosse a causa, influiria cumulativamente em Y — isto é, que os indivíduos submetidos a X por longo tempo mostrariam maior grau de Y —, mas que isso não ocorreria se Y fosse o fator causal.

Uma inferência desse tipo era fundamental no plano de um estudo de Newcomb (1943, 1947), que procurou saber quais os tipos de pessoas que aceitam determinados tipos de mudança social. Uma das hipóteses do estudo era que, "em grande parte, os valores se tornam valores através da mediação de grupos com os quais o indivíduo tem contato direto." Ao estudar as alunas do *Bennington College*, Newcomb considerou a comunidade escolar como um grupo com o qual os estudantes tinham contato direto, e a atitude com relação às questões públicas como um valor importante. Essa atitude foi escolhida porque a escola se caracterizava por elevado grau de interesse por questões públicas e por uma atitude "liberal" em questões discutíveis. O pesquisador pensou que se a participação no grupo fosse realmente a variável causal, os que estivessem submetidos por mais tempo à atmosfera do grupo mostrariam atitudes mais de acordo com as características do grupo (neste caso, atitudes mais liberais). De outro lado, se as atitudes liberais conduzissem à escola, haveria menos razão para esperar um aumento de liberalismo, com maior número de anos de fre-

quência.¹⁸ Com o emprego de diferentes medidas, Newcomb verificou que o fato de estar na comunidade de Bennington, o que era indicado pela série escolar, era acompanhado por um aumento na informação quanto a questões públicas e maior liberalismo de opiniões. Em todas as suas medidas, as calouras eram mais conservadoras, as alunas do último ano eram menos conservadoras.

De maneira semelhante, em estudo anterior, Klineberg (1935a) pensou que os QI relativamente baixos dos negros, no Sul dos Estados Unidos e nas zonas rurais, poderiam ser explicados por maus ambientes. Isso o levou a esperar que os QI de crianças negras aumentariam com o tempo de residência numa cidade como Nova Iorque. As pesquisas com crianças negras de 12 anos de idade, na cidade de Nova Iorque, confirmaram essa expectativa.

No entanto, o fato de que os resultados em Y se tornem diferentes com diferentes períodos de submissão a X não apresenta uma base indiscutível para uma inferência de causalidade. Não apenas X e Y podem ser mutuamente reforçadores, como já notamos. Outros fatores podem estar associados a diferenças no tempo de submissão à variável independente, e talvez esses outros fatores expliquem as diferenças na variável dependente. Por exemplo, os processos seletivos podem ser diferentes em diferentes épocas. No exemplo de Bennington, poderia ter ocorrido que um prestígio crescente da escola (que tinha sido fundada apenas poucos anos antes do estudo) ou uma mudança na política de admissões conduzissem a uma mudança no tipo de estudantes que ingressavam em diferentes anos, e essa mudança, em vez da influência da atmosfera da escola, é que explicaria as diferenças entre calouras e alunas do último ano. De forma semelhante, é concebível que os negros mais inteligentes tivessem emigrado antes para Nova Iorque, e que o QI mais elevado de crianças que estavam há mais tempo em Nova Iorque refletisse essa característica de seus pais.

18 Também aqui, naturalmente, existe a possibilidade de que a frequência à escola pudesse ter acentuado mais as atitudes liberais iniciais, mas presumivelmente isso não conduziria a diferenças tão marcantes entre os estudantes submetidos por muito tempo e os submetidos por pouco tempo a essa experiência, como se deveria esperar se a participação na escola fosse o principal fator causal.

Têm sido usadas diferentes verificações para essas possibilidades. Uma é a repetição do estudo em outra época. Newcomb repetiu suas medidas de atitudes de estudantes de Bennington em quatro anos consecutivos (1935-1939). Em cada ano, verificou que as alunas de quarto ano e terceiro ano eram mais liberais que as de classes mais atrasadas; assim, fortaleceu-se a inferência de que a experiência na escola, e não algum outro fator, explicava a mudança de atitudes e informação. Klineberg também usou o processo de repetição, e realizou pesquisas paralelas em dois anos consecutivos (1931 e 1932). Segundo seu argumento:

Se as verificações [do estudo de 1931 — segundo o qual as crianças negras que estão há mais tempo no Norte têm maior QI] são devidas a progressiva deteriorização na qualidade dos migrantes e não a um efeito do ambiente, os resultados obtidos (...) em 1932 deveriam estar consistentemente *abaixo* dos obtidos (...) em 1931. Um exemplo específico tornará mais claro esse raciocínio. Os meninos de 12 anos de idade, que no estudo de 1931 estavam em Nova Iorque há quatro anos, por exemplo, chegaram em 1927; os do estudo de 1932, que estavam em Nova Iorque durante um período igual, chegaram em 1928. Se os migrantes se tornam inferiores com a passagem do tempo, o grupo de quatro anos do último estudo deveria ser inferior ao grupo correspondente no estudo anterior.

No entanto, a comparação dos resultados de 1931 e 1932 mostrou, para cada grupo de determinado período de residência, uma leve mas consistente superioridade no estudo de 1932. Por exemplo, o QI médio de alunos que, em 1931, estavam em Nova Iorque há três ou quatro anos (isto é, que tinham emigrado em 1927 ou 1928) era 66,86, enquanto que o QI médio dos que, em 1932, estavam em Nova Iorque há três ou quatro anos (isto é, que tinham emigrado em 1928 ou 1929) era 79,06. Em outras palavras, as diferenças, embora pequenas, estavam consistentemente a favor dos que tinham chegado há menos tempo. Klineberg comentou:

Essa diferença pode ser devida a aperfeiçoamento nas escolas do Sul; de qualquer forma, não existe prova de que os que chegaram há menos tempo sejam piores. Portanto, justifica-se a conclusão de que os resultados superiores dos que têm maior período de residência são devidos a essa residência, e não a qualquer mudança na qualidade dos migrantes.

Evidentemente, existem outros métodos, além da repetição do estudo, para verificar se outros fatores podem ser responsáveis pelas diferenças encontradas na variável dependente. Por exemplo, Klineberg apresenta duas outras verificações da hipótese de que as diferenças na inteligência de negros que deixaram o Sul em épocas diferentes poderiam explicar as diferenças no QI, ligadas ao período de residência em Nova Iorque. Pesquisou as razões para a emigração do Sul, e nessas razões nada encontrou para confirmar a hipótese de que se pudesse esperar que os fatores que conduzem à emigração tivessem correlação com a inteligência. Também estudou os registros de escolas do Sul freqüentadas por crianças negras, e não encontrou diferença sistemática na posição relativa na escola, entre crianças que depois emigraram e aquelas que não o fizeram. Essas duas direções de provas sugeriram que, pelo menos durante o período abrangido por essas pesquisas interligadas, a inteligência não era um fator seletivo na emigração negra do Sul.

A *idade* é outro fator que pode ser confundido com tempo de exposição a uma variável. É provável que os indivíduos que foram submetidos a determinada experiência por maior período de tempo sejam mais velhos, e talvez a diferença de idade, e não de exposição à suposta variável causal, explique as diferenças na variável dependente. Aqui, a solução é "controlar" a idade — isto é, comparar indivíduos da mesma idade e que diferem no tempo de exposição à variável. Newcomb não usou essa abordagem, presumivelmente por causa da pequena diferença de idade na população universitária. No estudo de Klineberg, em que todas as crianças tinham doze anos de idade, a idade não era um fator. No entanto, num estudo que abranja grande amplitude de idade, esse é um aspecto importante a ser verificado, e é relativamente fácil fazê-lo.

SUBSTITUTOS PARA DISTRIBUIÇÃO CASUAL DE SUJEITOS PELO GRUPO DE CONTROLE E GRUPO EXPERIMENTAL

Como já foi indicado, uma das maneiras pelas quais o *experimenter* se protege contra inferências errôneas é a escolha casual de seu grupo de controle e grupo experimental, com ou sem emparelhamento complementar. A teoria da probabilidade permite verificar até que ponto os grupos casualmente escolhidos na mesma população tendem a diferir por acaso, seja na posição inicial com relação à variável dependente, seja nas características ou na história passada, que, segundo se pode esperar, influem na posição subsequente na variável dependente. Os testes de significância estatística, que consideram a probabilidade de diferenças casuais, permitem dizer, com grau especificado de certeza, se uma diferença entre os dois grupos na variável dependente, depois da exposição de um deles ao tratamento experimental, é maior do que seria esperado por acaso, e assim pode ser legitimamente atribuída aos efeitos da variável experimental. O pesquisador que usa um planejamento não-experimental tem várias maneiras substitutas para tentar atingir esse objetivo.

PROVA DE COMPARABILIDADE INICIAL DOS GRUPOS. O pesquisador pode obter dados a partir dos quais pode inferir se os indivíduos que passaram por diferentes experiências, e agora mostram diferenças na variável dependente, eram comparáveis *antes* de terem sido submetidos a tais experiências. No estudo a que já se fez referência, Newcomb usou, entre outros, esse processo. Comparou as alunas de Bennington com alunas de duas outras escolas que não tinham tido a mesma atmosfera de liberalismo e interesse por questões públicas. Verificou que as alunas de último ano de Bennington eram nitidamente menos conservadoras que as das duas outras escolas. No entanto, antes de poder inferir que essa diferença resultava de diferenças na atmosfera da escola, precisava verificar a possibilidade de que as estudantes de Bennington fossem menos conservadoras, mesmo antes de irem para essa escola. A comparação entre as atitudes de calouras das três escolas serviu para isso, e indicou que as estudantes de Bennington, no início de sua vida universitária, eram apenas um pouco menos conservadoras que as calouras de outras

escolas; assim, pôde concluir que os grupos de estudantes eram, inicialmente, muito semelhantes.

Freqüentemente, é impossível obter medidas de indivíduos, nos estágios iniciais da experiências cujos efeitos desejamos verificar. Todavia, se se sabe o suficiente a respeito de outros fatores que tendem a influir na variável dependente pela qual se interessa o pesquisador, ou pelo menos a estar associados com tal variável, talvez seja possível fazer estimativas provisórias da posição inicial na variável dependente, obtendo-se informação a respeito de tais características com ela relacionadas. Essa foi uma das verificações usadas por Wilner, Walkley e Cook (1955), em seu estudo das atitudes de locatários brancos com relação a negros, em conjuntos residenciais públicos com diferentes padrões de moradia. Estudos anteriores tinham indicado que algumas características — tais como educação, ideologia política geral, religião e experiência anterior com negros — estão relacionadas com atitudes com relação a negros. Por isso, esses pesquisadores incluíram, em sua entrevista, perguntas referentes a tais aspectos. A partir das respostas, construíram, para cada entrevistado, uma estimativa de seu provável grau de atitude favorável ou desfavorável aos negros, no momento da mudança para o conjunto residencial; depois, compararam os locatários de conjuntos residenciais integrados e segregados, através de sua estimativa de atitude inicial. Evidentemente, as provas desse tipo estão longe de ser decisivas, a não ser que se saiba que, para esse caso, é suficientemente elevada a usual correlação entre a variável dependente e as características sobre as quais o pesquisador tem informação.

Geralmente, existe uma possibilidade de que as pessoas que passaram por experiências diferentes tenham escolhido tais experiências. Essa escolha significa, quase certamente, que os grupos comparados não eram inicialmente equivalentes; podem ter sido diferentes sob aspectos que exerceriam grande influência em sua posição na variável dependente. Por isso, sempre que possível, é importante verificar a probabilidade de escolha pessoal. Wilner e seus colaboradores incluíram, entre suas verificações, provas a esse respeito: algumas foram obtidas com os locatários, algumas em outras fontes. Perguntaram aos locatários brancos se, quando se inscreveram para admissão nos conjuntos residenciais em que estavam vivendo, sabiam que tais conjuntos também tinham moradores negros. Se um indivíduo não sabia que um con-

junto era bi-racial, havia pouca probabilidade de que sua atitude com relação aos negros interferisse em sua decisão para essa mudança. Com os diretores dos conjuntos residenciais, obtiveram a seguinte informação: saber se todos os conjuntos residenciais na comunidade tinham o mesmo padrão de residência racial, o que reduzia a oportunidade de escolha; registros de recusas, por candidatos brancos, de apartamentos vizinhos de negros, ou de transferência de tais apartamentos; registros de recusas e proporção de transferências em apartamentos distantes de uma família negra, etc.

Às vezes, um pesquisador tem a sorte de encontrar uma situação em que não existe possibilidade de escolha pessoal. Por exemplo, a Information and Education Division of the U. S. War Dept. (1947) verificou que os soldados brancos, em unidades que tinham companhias constituídas de batalhões brancos e negros, eram muito mais favoráveis a batalhões brancos e negros na mesma companhia que os soldados brancos de unidades exclusivamente brancas. Dada a natureza das designações militares, não havia possibilidade de que os soldados brancos, inicialmente mais favoráveis aos negros, se tivessem colocado em unidades mistas.

COMPARAÇÃO DE SUBGRUPOS EMPARELHADOS. Suponha-se que as provas coligidas indiquem que os grupos que passaram por experiências diferentes não eram inicialmente comparáveis. Isso ocorreu no estudo de Deutsch e Collins (1951) sobre atitudes de locatários de conjuntos residenciais públicos. Esses autores verificaram que os moradores de seus conjuntos segregados e integrados eram diferentes quanto a religião, educação e atitudes políticas — características que outros estudos verificaram estar ligadas com atitudes com relação a negros. Além disso, a natureza das diferenças era de tal ordem que se podia supor que os locatários de conjuntos residenciais integrados tivessem atitudes de menos preconceito, antes de mudar para os conjuntos residenciais, que os locatários de conjuntos segregados. Os autores dizem:

(. . .) está claro que na análise e na apresentação de nossos resultados será necessário eliminar ou controlar os efeitos de tais diferenças de população, a fim de atribuir significação causal aos efeitos do padrão de residência.

Para isso, compararam subgrupos emparelhados, em cada conjunto. Por exemplo, considerando separadamente os locatários politicamente liberais entrevistados, examinaram a proporção, nos conjuntos integrados e segregados, que mostrava atitudes amistosas com relação a negros no conjunto residencial. Verificaram que 31 por cento e 72 por cento dos politicamente liberais nos conjuntos residenciais integrados, e zero por cento e 17 por cento dos politicamente liberais nos dois conjuntos residenciais segregados, foram classificados como possuidores de sentimentos amistosos com relação a negros de seu conjunto residencial. O exame separado dos "intermediários", dos conservadores, dos com diferentes religiões e diferentes proporções de educação mostrou os mesmos resultados: entre os entrevistados emparelhados nessas outras características, os dos conjuntos residenciais integrados mostravam, consistentemente, sentimentos mais amistosos com relação aos negros do conjunto residencial.¹⁹

LIMITAÇÃO DA AMOSTRA. Em vez de considerar uma amostra heterogênea e comparar subgrupos emparelhados dentro dela, o pesquisador pode certificar-se de que seus sujeitos estão emparelhados sob certos aspectos, se incluir em sua amostra apenas pessoas com determinadas características. Esse processo também pode ser usado em estudos experimentais; neste último caso, é empregado, geralmente, por razões de facilidade administrativa ou para aumentar a sensibilidade do experimento, através da eliminação de fatores que poderiam influir tão intensamente na variável dependente que "apagariam" os efeitos da variável independente em que o pesquisador está interessado (ver págs. 117-119). Em estudos não-experimentais, a limitação da amostra tem outra função: separar a variável independente — em que o pesquisador está interessado — de outras variáveis com as quais está geralmente ligada, de forma que quaisquer efeitos encontrados possam ser atribuídos, justificadamente, a essa variável. Naturalmente, nos experimentos essa separação é obtida através da distribuição casual de pessoas pelo grupo experimental e pelo grupo de controle, e pelo controle de algumas das outras variáveis contemporâneas.

Por exemplo, suponha-se que estamos interessados em verificar se o fato de ter um lugar quieto em casa, para estu-

¹⁹ No Capítulo 11 são discutidos os processos usados na análise para comparar tais subgrupos.

dar, influi nas notas de alunos do ginásio. O fato de ter um lugar quieto para estudo tende a ser parte de um conjunto de fatores. O excesso de pessoas tende a ser maior no caso de famílias com menos recursos; por isso, em suas casas tende a haver menos possibilidade de espaço quieto para estudo. Ao mesmo tempo, as famílias de poucos recursos tendem a dar menos valor aos resultados na escola que os pais de crianças de classe média e classe alta. Além disso, nos lares de classe mais baixa tende a haver menos oportunidade para as crianças conseguirem o tipo de conhecimento que contribuiria para melhor aproveitamento escolar. Se o pesquisador se limitar a escolher um grupo de ginasianos, perguntar a cada um se tem ou não um lugar quieto para estudar em casa (ou conseguir essa informação de alguma outra forma), e depois comparar as notas dos que têm um lugar quieto para estudo com as dos que não o têm, pode chegar a uma conclusão muito enganadora. A diferença em notas pode não ser devida a diferenças no espaço para estudo, mas a diferenças na valorização de aproveitamento escolar, em oportunidades para obter conhecimento importante para a escola, ou a alguma combinação de todos esses fatores ligados à classe sócio-econômica.

Se o problema pudesse ser tratado em pesquisa experimental, o experimentador poderia selecionar uma amostra de estudantes e fazer com que alguns deles (casualmente selecionados) tivessem um lugar quieto para estudo, outros (também casualmente selecionados, mas talvez também emparelhados quanto a status sócio-econômico) não o tivessem. Sem essa possibilidade, o pesquisador pode eliminar os efeitos de diferenças de classes, ao limitar seu estudo a alunos de apenas um nível sócio-econômico. Poderia estudar apenas crianças de classe baixa, ou apenas crianças de classe média, e comparar as notas dos que têm com as dos que não têm um lugar quieto para estudo. Ao limitar dessa forma a sua amostra, teria separado sua variável independente de algumas variáveis ligadas a classe social, com as quais geralmente está associada. Evidentemente, não teria eliminado algumas dificuldades — por exemplo, a possibilidade de que na classe baixa, ou na classe média, os pais preocupados com o êxito escolar arranjam um lugar quieto para o estudo dos filhos. Em outras palavras, nesta amostra limitada, se verificasse uma relação entre lugar quieto para estudo e notas na escola, estaria certo de que as diferenças em notas não seriam devidas

a diferenças em status sócio-econômico, embora pudessem ser devidas a algum outro fator, ou fatores não controlados.

Um simples aspecto lógico é importante aqui. A frase "não devidas a diferenças em . . ." significa, literalmente, "não devidas a *diferenças* em . . ." Não significa que o status sócio-econômico não desempenhe um papel nas relações observadas. Assim, se seu estudo incluiu apenas crianças de classe baixa, a verificação de que as crianças com um lugar quieto para estudo obtinham notas mais altas poderia valer apenas para esse grupo sócio-econômico; entre crianças de classe média, é possível que alguma outra relação fosse encontrada, ou que não se encontrasse qualquer relação. Vale a pena lembrar esse aspecto porque freqüentemente se supõe, embora de forma incorreta, que a manutenção da constância de certas variáveis em um estudo *elimina* o seu efeito. Considere-se o seguinte exemplo: um dos gêmeos idênticos é criado num ambiente sócio-econômico muito pobre, enquanto o outro é criado em condições muito melhores. Quando chegam à idade de doze anos, verifica-se uma diferença de vinte pontos de QI entre eles. Afirma-se que, como sua hereditariedade era idêntica, a diferença no QI é *devida* à diferença de ambiente — e, por implicação, que a hereditariedade comum não exerceu qualquer influência. No entanto, pode-se pensar que, se essa hereditariedade comum fosse, sob algum aspecto importante, diversa da que é na realidade, as duas crianças poderiam ter reagido de maneira muito diferente da sua reação real a seus respectivos ambientes, e que se poderia observar uma outra diferença de QI, ou mesmo não encontrar qualquer diferença.

Embora esse processo de limitação da amostra tenha a vantagem de eliminar variações em outras características especificadas como condições contribuintes para as diferenças observadas na variável dependente, tem a desvantagem correspondente de limitar a população para a qual os resultados podem ser generalizados. Assim, em nosso exemplo da relação entre lugar quieto para estudo e notas na escola, se limitamos os sujeitos a crianças de classe baixa que freqüentam escolas públicas, quaisquer resultados sobre o efeito de lugar para estudo em notas não podem com qualquer confiança, ser supostos para crianças de classe média, etc. Para que se possa ter confiança na possibilidade de generalizar os resultados, o estudo precisa ser repetido com grupos de sujeitos que tenham outras características. Esse processo de

repetição com outros grupos tem, naturalmente, uma outra função: ajuda a especificar as condições *contingentes* sob as quais ocorre uma relação entre duas variáveis.

O EMPREGO DE UMA COMBINAÇÃO DE ABORDAGENS

Como nenhum desses substitutos dos processos experimentais apresenta tanta proteção quanto a distribuição casual de sujeitos por grupos de controle e grupos experimentais — na qual o pesquisador controla a aplicação da variável independente —, é sempre aconselhável planejar o maior número possível de abordagens diferentes, a fim de conseguir uma base para inferências de causalidade. Quase todos os estudos que mencionamos usaram várias verificações: não descrevemos todas. Muitas vezes, o assunto de um estudo determinará as verificações mais importantes e mais realizáveis.

Quase todos os recursos que possibilitam as inferências causais em estudos não-experimentais incluem exigências específicas na análise dos dados. Por isso, uma seção do Capítulo 11, que se refere a análise de dados, é dedicada à discussão de processos analíticos, úteis quando desejamos fazer inferências a respeito de causalidade em tais estudos.

Resumo

Nestes dois capítulos interessamo-nos pelas exigências de processos de pesquisa, propostos por diferentes tipos de problemas de pesquisa. Acentuamos que nem mesmo a pesquisa cuidadosamente controlada dará conhecimento com absoluta certeza. No entanto, apesar de suas limitações, o método científico, mais que qualquer outro processo conhecido pelo homem, pode reduzir a informação errada. O ceticismo, a autocritica vigilante, a verificação constante de hipóteses através de pesquisa empírica, a consciência das limitações que caracterizam o empreendimento científico, fazem com que os resultados de pesquisa, quando usados com inteligência, sejam a fonte mais segura de que podemos depender para informação.

No Capítulo 3 indicamos que a pesquisa não é apenas uma questão de verificação experimental de hipóteses bem formuladas. A criação de hipóteses produtivas não ocorre

num vazio, nem depende apenas de sorte ou capacidade criadora. Pode ser auxiliada por estudos descritivos e exploratórios cuidadosamente planejados, cujo objetivo é a reunião de informação básica, necessária para uma formulação adequada do problema. Indicamos que tais estudos, sobretudo numa ciência jovem, exigem os esforços de mais imaginação do cientista, e não devem ser desprezados porque não têm o elegante rigor do estudo experimental. Foi também examinada a função dos estudos descritivos na avaliação das características de determinada situação e na verificação de hipóteses que não envolvem relações causais. Neste capítulo discutimos pormenorizadamente os estudos destinados a verificar hipóteses causais. Nesse contexto, apresentamos as considerações lógicas existentes nas inferências sobre causalidade, e descrevemos vários processos para a obtenção de dados que podem dar uma base para tais inferências.

Agora, é correto passar da estrutura de um estudo de pesquisa para os métodos que dão substância e corpo ao planejamento de pesquisa — os processos de coleta de dados. No capítulo seguinte, discutiremos algumas considerações gerais de mensuração que influem em todas as coletas de dados. Os Capítulos 6, 7 e 8 discutirão métodos para a coleta de dados primários — métodos de observação, questionários e entrevistas, técnicas projetivas. No Capítulo 9, será discutido o emprego de fontes secundárias de dados — registros estatísticos, documentos pessoais, conteúdo da comunicação. O Capítulo 10 discutirá as técnicas para colocar os indivíduos em escalas, a partir de dados coligidos por qualquer um desses métodos.

5

ALGUNS PROBLEMAS GERAIS DE MENSURAÇÃO

Variações nos Resultados com Instrumentos de Medida

A Validade das Mensurações

A Precisão das Mensurações

Escalas de Mensuração

Resumo