

Laboratório de Psicologia Experimental
Departamento de Psicologia - UFSJ
Disciplina: Método de Pesquisa Quantitativa
Professora: Marina Bandeira

TEXTO 6: DELINEAMENTOS QUASE-EXPERIMENTAIS

Autores: Selltitz – Wrightsman – Cook.. (1976). Métodos de pesquisa nas relações sociais.
Volume 1. Delineamentos de pesquisa. São Paulo. E.P.U.

A ciência não começa nem termina com os experimentos propriamente ditos, mas é um processo de descoberta no qual usamos os melhores instrumentos e ferramentas para responder as nossas questões. Quando a distribuição aleatória e o controle de laboratório são inviáveis, podemos escolher dentre ampla variedade de outras técnicas. O termo quase-experimentos abarca grande variedade de outros delineamentos de pesquisa. O termo se torna conhecido pela primeira vez com a publicação de um livro pequeno mas influente, escrito por Campbell e Stanley (1963), recentemente revisto e ampliado (Cook e Campbell, também Judd e Kenny, no prelo).

Quase-experimentos são delineamentos de pesquisa que não têm distribuição aleatória dos sujeitos pelos tratamentos, nem grupos-controle. Ao invés disso, a comparação entre as condições de tratamento e não-tratamento deve sempre ser feita com grupos não equivalentes ou com os mesmos sujeitos antes do tratamento. Se você não puder distribuir aleatoriamente pessoas ou grupos pelas condições experimentais, perderá a capacidade de controlar o que ocorre a quem. Entretanto, você ainda poderá *observar* o que ocorre, quando ocorre e a quem ocorre; e, ao decidir o que e quando medir, poderá planejar um dentre vários quase-experimentos. Você terá perdido o controle por não usar a distribuição aleatória, mas ainda assim poderá realizar uma pesquisa e analisar relações de causa-efeito sem um experimento. Ao coletar dados criteriosamente em situações e momentos adicionais você poderá criar um quase-experimento. Por exemplo, Lawler e Hackman (1969) estudaram os efeitos da participação de zeladores nas tomadas de decisões sobre sua assiduidade no trabalho. Eles não podiam distribuir os zeladores aleatoriamente por grupos que participariam na tomada de decisões ou grupos-controle e, portanto, não tinham grupos-controle equivalentes com os quais fazer comparações. No entanto, os pesquisadores tinham uma longa série de registros do absenteísmo dos homens, anteriores a sua nova condição de participantes nas tomadas de decisão, e poderiam comparar os índices de absenteísmo antes e depois de serem introduzidos os processos de participação. Este não foi um experimento propriamente dito, porque não houve distribuição aleatória. Foi um quase-experimento, denominado delineamento de série temporal.

Neste capítulo mostramos como você pode fazer uma pesquisa na qual avalie causas e efeitos mesmo que não possa distribuir as pessoas aleatoriamente pelas condições. Quase-experimentos bem concebidos permitem-lhe descartar muitas das ameaças à validade interna que discutimos no capítulo 2.

Neste capítulo discutiremos três tipos de quase-experimentos, sendo cada um dos quais uma extensão de um pré-experimento. Cada um deles alcança sua maior possibilidade de interpretação através do acréscimo de mais pontos de dados ou observações à base pré-experimental.

1. Delineamento de série temporal descontínua

<i>O1</i>	<i>O2</i>	<i>O3</i>	<i>O4</i>	<i>O5</i>	<i>O6</i>	<i>O7</i>	<i>O8</i>
<i>XO9</i>	<i>O10</i>	<i>O11</i>	<i>O12</i>	<i>O13</i>	<i>O14</i>	<i>O15</i>	<i>O16</i>

Delineamentos de série temporal são extensões do delineamento pré-experimental de pré e pós-teste com um grupo (*O1* *X* *O2*). Embora o antecessor pré-experimental seja sujeito a várias ameaças à validade interna, a série temporal, com sua longa sequência de *O*s, fornece informações que permitem descartar várias ameaças, principalmente maturação e testagem. Se você encontrou uma diferença marcante entre *O8* e *O9* e se perguntou se a diferença é realmente resultado do tratamento (*X*) ou da maturação, poderá inspecionar todos os intervalos anteriores e posteriores àquele ponto para procurar indícios de maturação. Presumivelmente, se estiver ocorrendo maturação, esta aparecerá como uma tendência a longo prazo, produzindo efeitos similares entre *O1* e *O2*, *O2* e *O3*, e assim sucessivamente, ao longo de toda a série. Se nenhum dos outros intervalos mostrou tal tendência e a única diferença encontrada foi entre *O8* e *O9*, maturação não será uma hipótese muito plausível a menos, é claro, que esteja estudando algum fenômeno que venha a coincidir com uma mudança maturacional particular, tal como puberdade, e que pudesse também, plausivelmente, ser afetado por ela. Apenas sob tal conjunto especial de circunstâncias e coincidências a maturação colocaria ameaças à validade interna de um estudo de série temporal.

O mesmo raciocínio se aplica à testagem como uma hipótese rival. Se você suspeitasse que a diferença entre *O8* e *O9* resultou não do tratamento mas de efeitos sensibilizadores do pré-teste (*O8*), poderia examinar todos os intervalos precedentes e subsequentes para ver se a testagem repetida produziu diferenças similares ao longo da série inteira. Se não houvesse diferenças em qualquer outros pontos, seria altamente implausível que a testagem apenas em *O8* tivesse produzido um efeito em *O9*.

Não raro o *X* ocorre apenas uma vez, e presume-se que seu efeito persista para sempre ou por tempo determinado. Uma inoculação contra sarampo deveria durar para sempre; uma injeção antigripal pode ter um período limitado de eficácia – algumas vezes o *X* sinaliza uma mudança permanente na situação – como quando um Estado muda suas leis sobre o divórcio para autorizar divórcios sem “justa” causa ou quando o governo introduz novos padrões de poluição atmosférica. Em casos como estes, quando o tratamento ocorre não apenas em um único ponto, mas continua em vigor, a série temporal é mais corretamente representada como se segue:

<i>O1</i>	<i>O2</i>	<i>O3</i>	<i>O4</i>	<i>O5</i>	<i>O6</i>	<i>O7</i>	<i>O8</i>
<i>XO9</i>	<i>XO10</i>	<i>XO11</i>	<i>XO12</i>	<i>XO13</i>	<i>XO14</i>	<i>XO15</i>	<i>XO16</i>

Em cada caso, com um tratamento de caso único ou contínuo, a virtude dos delineamentos de série temporal é que você pode examinar a *tendência* dos dados antes do tratamento, no momento da intervenção e após o tratamento. Isso lhe permite avaliar a plausibilidade de maturação como uma hipótese rival. Se maturação foi uma causa, ela deverá aparecer como uma tendência antes do tratamento, bem como após.

Quão facilmente você pode interpretar uma série temporal e descartar hipóteses rivais não depende das características formais do delineamento isoladamente, mas também do padrão dos resultados. Alguns resultados são relativamente fáceis de interpretar – você pode descartar a maioria das ameaças à validade e concluir que o tratamento causou o efeito. Outros padrões são mais vulneráveis a interpretações rivais.

Os dois fatores que os pesquisadores examinam nos resultados de um delineamento de série temporal descontínua são as inclinações das retas para as observações de pré e pós-teste e a intercepção no eixo dos Y – o ponto no qual cada uma interceptará o eixo vertical. Por exemplo, as figuras 3.1 e 3.3 mostram vários resultados esquematizados de delineamentos de série temporal descontínua – a primeira (3.1) é prontamente interpretável e mostra um efeito de tratamento; a segunda (3.2) é ambígua; a terceira (3.3) é um caso claro em que não houve efeitos do tratamento.

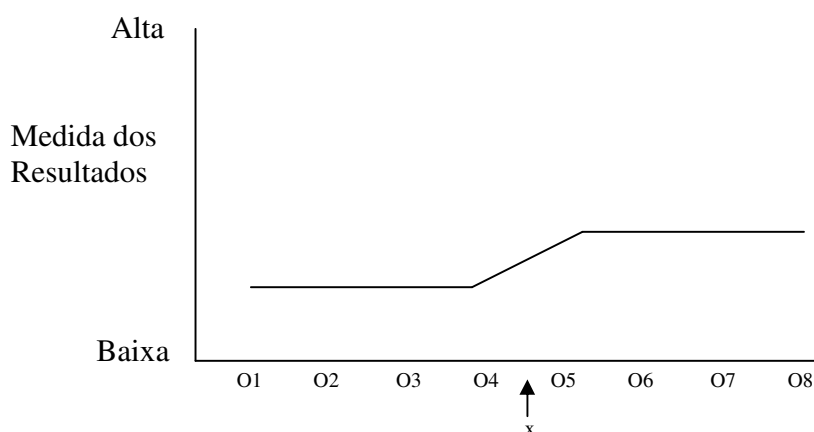


Figura 3.1. Intervalos temporais em que as medidas foram tomadas, sendo o tratamento introduzido entre O4 e O5.

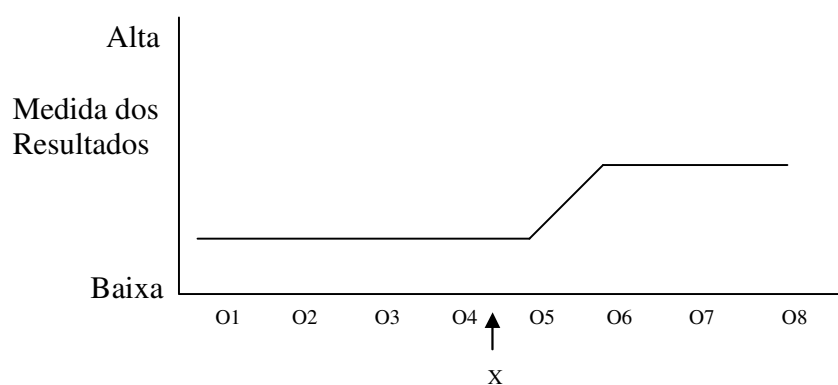


Figura 3.2. Intervalos temporais em que as medidas foram tomadas, sendo o tratamento introduzido entre O4 e O5.

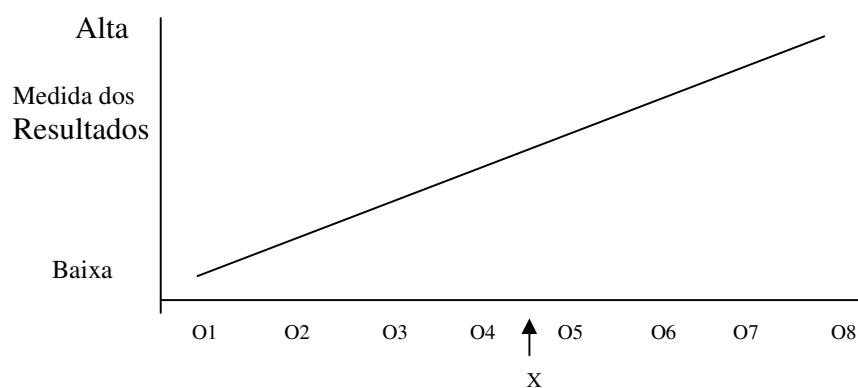


Figura 3.3. Intervalos temporais em que as medidas foram tomadas, sendo o tratamento introduzido entre O4 e O5.

Figuras 3.1, 3.2 e 3.3. Alguns resultados possíveis derivados da introdução de um tratamento (x) numa série temporal de mensurações.

Com resultados como os da figura 3.1 não há muitas hipóteses rivais. É bastante óbvio que o tratamento causou a mudança nos escores de pré para pós-teste. Não há tendência de maturação nas observações de pré nem de pós-teste; assim, maturação isoladamente ou a interação de maturação com seleção não são explicações alternativas convincentes. A ameaça mais problemática e plausível é história – algum evento que tenha coincidido com o tratamento. Quão plausível ela é depende totalmente do problema em estudo. Se estivermos estudando os efeitos de um filme sobre relações exteriores sobre as atitudes de estudantes norte-americanos em relação a pessoas de outros países, e se a apresentação do filme coincidir com acordos internacionais acerca da distribuição das fontes de energia mundiais ou com um incidente internacional no qual algum outro país fez reféns norte-americanos, estes eventos históricos seriam hipóteses rivais plausíveis. Se não houver eventos externos óbvios que coincidem com o tratamento e que poderiam ter produzido os mesmos efeitos, a história não se apresenta como um problema.

Com resultados como os da figura 3.2, não fica claro se o tratamento causou a mudança ascendente porque a mudança não coincide com o tratamento; ela aparece com atraso de um intervalo. Em alguns casos pode haver razões para esperarmos tal atraso e, portanto, atribuir o efeito ao tratamento. Por exemplo, uma alta no preço mundial de exportação de petróleo não seria sentida imediatamente no preço da gasolina nas bombas, por haver uma reserva de óleo para seis meses que seria vendida a preços mais baixos. Portanto, se você estivesse estudando os efeitos do aumento mundial no preço de exportação do petróleo sobre o consumo de gasolina pelos norte-americanos, o efeito poderia demorar seis meses para aparecer após a causa. Se as observações de série temporal forem feitas trimestralmente, com intervalos de três meses, o efeito apareceria não na primeira observação pós-tratamento, mas na segunda, e o padrão apresentado na figura 3.2 seria uma demonstração convincente do efeito do aumento de preço no consumo. Se não houver tal mecanismo de armazenamento e, portanto, nenhum atraso plausível para que a causa tenha este efeito, a figura 3.2 será mais difícil de interpretar, e outras hipóteses rivais poderão ser causas. Estas explicações alternativas provavelmente ficariam sob o rótulo de *história* – outros eventos que acompanharam o tratamento e produziram os efeitos.

A figura 3.3 é o caso mais claro em que não ocorre nenhum efeito. Os níveis mais altos que *O* atingiu após o tratamento refletem apenas a tendência predominante que também produziu níveis de *O* sucessivamente mais altos antes do tratamento. Esta figura também mostra mais claramente por que o delineamento de série temporal é muito superior ao delineamento pré-experimental de pré e pós-teste com um grupo. Se o padrão observado na figura 3.3 fosse estudado com apenas pré-teste e um pós-teste (pontos *O4* e *O5*), o experimentador ver-se-ia tentado a concluir que o tratamento teve um efeito – fez com que *O5* fosse mais alto que *O4*. Sem outros pontos de dados, seria impossível distinguir entre um efeito real e a tendência predominante que vemos na figura 3.3.

Com o acréscimo de uma série de observações de pré e pós-teste, o delineamento de série temporal é muito mais interpretável que o delineamento de pré e pós-teste com um grupo. Mesmo que não inclua distribuição aleatória nem grupo-controle, ele será um delineamento experimental útil, pois as observações adicionais permitem ao investigador testar a plausibilidade de várias hipóteses rivais: maturação, testagem, instrumentação, mortalidade e regressão. A principal ameaça à validade com este delineamento é *história*.

O delineamento de série temporal descontínua apresenta diversas variações. Pode incluir mais de um grupo ou sujeito, formando-se um delineamento de série temporal múltipla.

2. O delineamento de amostras temporais equivalentes

Autores: Campbell e Stanley (1979), Delineamentos Experimentais e Quase-experimentais de pesquisa. São Paulo: EDUSP. pag. 75-78.

A forma mais usual de delineamento experimental emprega uma amostra equivalente de pessoas para fornecer uma linha de base em relação à qual se possam comparar os efeitos da variável experimental. Em contraste, a forma recorrente de experimentação de um grupo emprega duas equivalentes amostras de ocasiões, numa das quais a variável experimental está presente sem que o esteja na outra. Esse esquema pode ter o seguinte diagrama (embora se queira uma alteração mais casual do que regular):

X0 O X1O X0 O X1O X0 O (ou delineamento A-B-A-B-A)

Esse delineamento pode ser visto como uma forma do experimento de série temporal com a introdução repetida da variável experimental. O experimento é obviamente mais útil quando o efeito da variável experimental é antecipado como de caráter transitório ou reversível. Enquanto a lógica do experimento pode ser vista como uma extensão do experimento de série temporal, o modo de análise estatística é mais tipicamente semelhante ao do experimento de dois grupos em que é empregada a significância da diferença entre as médias de dois conjuntos de medidas. Usualmente, as medidas são bem especificamente emparelhadas com as apresentações da variável experimental, sendo frequentemente concomitantes, como no caso de estudos de aprendizagem, produção de trabalho, condicionamento, reação fisiológica, etc. Talvez o primeiro dos usos mais típicos desse esquema experimental, como nos estudos de eficiência do trabalho e de estudantes sob várias condições de autoria de Allport (1920) e Sorokin (1930), envolveram a comparação de duas variáveis experimentais uma com a outra, isto é, X1 *versus* X2, e não a de uma com um controle. Por muitos motivos, a simples alternância de condições e o emprego de um espaçamento temporal consistente são indesejáveis, particularmente quando possam introduzir confusão com um ciclo diário, semanal ou mensal, ou quando, através da periodicidade previsível de um condicionamento não querido ao intervalo temporal, possam acentuar a diferença entre uma apresentação e outra. Assim, Sorokin tomou precauções para que cada tratamento experimental ocorresse com igual frequência à tarde e pela manhã.

A maioria dos experimentos que empregaram esse delineamento têm usado relativamente poucas repetições de cada condição experimental, mas o tipo de extensão da teoria de amostragem apresentado por Brunswik (1956) chama atenção para a necessidade de amostragens casuais de períodos de tempo amplas, representativas e equivalentes. Kerr (1945) talvez tenha sido quem mais se aproximou desse ideal em seus experimentos sobre os efeitos da música na produção industrial. Cada um de seus vários experimentos envolveu um único grupo experimental com uma amostra casual e equivalente de dias ao longo de períodos mensais. Assim, num experimento ele pode comparar 56 dias de música com 51

dias sem música e, em outro, pode comparar três diferentes tipos de música, cada um dos quais representados por amostras equivalentes de 14 dias.

Da forma pela qual foi empregado por Kerr, por exemplo, o delineamento de amostras temporais equivalentes parece também internamente válido. *História*, a maior fraqueza da série temporal, é controlada pela apresentação de X em numerosas ocasiões distintas, tornando extremamente improvável qualquer explicação rival baseada na coincidência de eventos estranhos. As outras fontes de invalidade são controladas pela mesma lógica invocada em relação ao delineamento de série temporal descontínua. Com relação à validade externa, generalização é obviamente possível apenas para populações freqüentemente testadas. O efeito reativo de montagem (*arrangements*), a consciência de experimentação representa uma vulnerabilidade típica desse experimento. Quando grupos distintos são submetidos a XX distintos, é possível mantê-los no total desconhecimento da presença de um experimento ou dos tratamentos em cotejo. Isso não ocorre quando se acha envolvido apenas um grupo e quando esse grupo é repetidamente exposto a uma condição ou outra, por exemplo, a uma base para computar pagamento e outra, como ocorreu no experimento de Sorokin; a uma espécie de ventilação e a outra, como se deu nos estudos de Wyatt, Fraser e Stock (1926); e a uma espécie de música ou outra, como se verificou na pesquisa de Kerr (embora Kerr tomasse precauções elaboradas para que a programação variada fosse vista como uma parte natural do ambiente de trabalho). No que tange à interação de *seleção* e X: como de hábito, existe a limitação da generalização dos efeitos demonstrados de X para o tipo particular de população envolvida.

Esse delineamento quase-experimental traz consigo uma aleatoriedade quanto à validade externa, que será encontrada em todos os experimentos descritos neste ensaio em que múltiplos níveis de X são apresentados ao *mesmo* conjunto de pessoas. Esse efeito têm sido rotulado como “interferência de X múltiplo”. O efeito de X1, na situação mais simples em que está sendo comparado com X0, pode ser generalizado somente para condições de apresentações repetitivas ou espaçadas de X1. Nenhuma base aceitável existe para generalização a possíveis situações em que X1 se acha continuamente presente ou à condição em que é introduzido uma vez e apenas uma vez. Além disso, a condição X0 ou a ausência de X não é típica de períodos sem X em geral, mas é apenas representativa de ausências de X intercaladas entre presenças. Se X1 tiver algum efeito prolongado capaz de repercutir nos períodos de não-X, como pareceria usualmente provável, o plano experimental poderá subestimar o efeito de X1 se comparado com um estudo do delineamento experimental com grupo de controle e só pós-teste, por exemplo. Do outro lado, o simples fato de haver freqüentes mudanças pode aumentar o valor estimulante de X em relação ao que este valor seria sob uma apresentação contínua e homogênea. A música hawaiana no estudo de Kerr pode afetar o trabalho de modo completamente diferente se intercalada por um dia entre dias de outra música, do que o faria se constituísse uma programação contínua. Os delineamentos experimentais de Ebbinghaus (1885) podem ser encarados como essencialmente desse tipo e, como observou Underwood (1957 a), as leis por ele encontradas são limitadas em sua generalizabilidade a uma população de pessoas que aprenderam dúzias de outras listas altamente semelhantes. Muitas de suas conclusões, com efeito, não prevalecem no caso de pessoas que tenham aprendido uma única lista de sílabas sem sentido. Assim, embora o esquema seja válido internamente, sua validade externa pode ser seriamente limitada a alguns tipos de conteúdos. (Ver também Kempthorne, 1952, Cap.29)

Note-se, contudo, que muitos aspectos do ensino sobre os quais alguém gostaria de conduzir um experimento podem muito bem ter efeitos limitados, do ponto de vista prático, ao período de efetiva presença de X. Para tais fins, esse delineamento pode ser perfeitamente valioso. Suponhamos que um professor levante a questão do valor de recitação oral *versus* estudo individual silencioso. Ao variar esses dois processos ao longo de uma série de unidades de ensino, pode-se montar um experimento interpretável. O efeito da presença de um pai-observador na sala de aula sobre o debate voluntário de estudantes pode ser estudado dessa forma. Consciência de tais esquemas pode situar uma testagem experimental de alternativas dentro do alcance de um único professor. Isso poderia servir de teste-piloto para avaliar procedimentos que, se promissores, poderiam ser examinados por experimentos mais amplos e mais coordenados.

Essa abordagem pode ser aplicada a uma amostragem de ocasiões em relação a um único indivíduo. Embora não dê ensejo tipicamente à aplicação de testes de significância, é esse um delineamento recorrente em pesquisa fisiológica, em que um estímulo é aplicado repetidamente a um animal, com o cuidado de evitar qualquer periodicidade na estimulação. Esta última característica corresponde à exigência de aleatoriedade de ocasiões exigidas pela lógica do plano.

3. Delineamento de série temporal múltipla

O1 O2 O3 O4 O5 O6 X O7 O8 O9 O10 O11 O12

O1' O2' O3' O4' O5' O6' O7' O8' O9' O10' O11' O12'

Este delineamento fornece informações idênticas sobre um segundo grupo ou sujeito e lhe permite testar a ameaça mais problemática à validade interna que um delineamento de série temporal simples apresenta – *história*. Se tiver coincidido com algum evento histórico e se os dois grupos foram presumivelmente expostos às mesmas condições históricas, o efeito deste incidente histórico deveria aparecer em ambas as séries como uma diferença entre O6 e O7. Por exemplo, suponha que tenhamos tentado avaliar o efeito de um filme denominado *A morte de uma princesa*, sobre a atitude de norte-americanos e europeus em relação a árabes sauditas. O filme era uma narrativa romanceada de uma história verdadeira – uma princesa da Arábia Saudita fora executada por um pelotão de fuzilamento após ter sido acusada de cometer adultério. Oficiais do governo saudita protestaram contra a exibição deste filme na Europa e América do Norte, alegando que ele deturpava a sociedade árabe e provocaria hostilidade nos espectadores. Apesar disso, o filme foi exibido na televisão norte-americana e européia. Se tivéssemos estudado os efeitos disto – medindo os sentimentos anti-sauditas expressos em editoriais de jornais ou contando o número de adesivos anti-sauditas nos pára-choques dos carros, poderíamos ter encontrado um efeito como o da figura 3.1. Suponha, entretanto, que a exibição do filme tenha coincidido com um aumento no preço do petróleo ou com um ato de terrorismo. Qualquer evento deste tipo que ocorresse aproximadamente ao mesmo tempo poderia ser responsável pelo aumento de hostilidade observado na imprensa ou em nossa contagem de

adesivos de pára-choques. Como poderíamos determinar se foi o filme ou os outros eventos históricos que produziram este efeito? Se pudéssemos encontrar alguma localidade que tivesse sido exposta ao mesmo aumento no preço do óleo ou tivesse ouvido falar sobre o mesmo ato de terrorismo, mas não tivesse sido exposta ao filme, poderíamos avaliar as hipóteses rivais. Duas cidades que tivessem experienciado o aumento no preço do óleo, uma das quais tivesse assistido ao filme e outra não, forneceriam um teste para a hipótese rival. Se o sentimento anti-saudita expresso nos jornais da primeira cidade ou nos pára-choques dos carros aumentasse após a exibição do filme enquanto se mantivesse constante na outra cidade, descartaríamos as hipóteses rivais de *história* e concluiríamos que o filme produziu efeito.

Este exemplo apresenta não só um quase-experimento hipotético, mas um dilema real. Algumas vezes as questões para as quais são planejados delineamentos quase-experimentais são importantes questões políticas com implicações sérias. Mas as medidas disponíveis para respondê-las são extremamente inadequadas. Um delineamento de série temporal, em particular, freqüentemente se baseia em dados de arquivos – tais como registros das vendas do número de adesivos para pára-choques – que nunca foram planejados para serem indicadores sensíveis de uma variável social, tal como atitudes em relação a outra nação. Apesar disto, tais dados podem ser as únicas medidas disponíveis numa série suficientemente longa num delineamento quase-experimental, como o de série temporal. O capítulo 12 sobre dados observacionais e de arquivos discute o poder e a fraqueza de tais dados de arquivos disponíveis.

4. Delineamento de pré e pós-teste com grupo controle não equivalente

Autores: Selltitz – Wrightsman – Cook.. (1976). Métodos de pesquisa nas relações sociais. Volume 1. Delineamentos de pesquisa. São Paulo. E.P.U.

O delineamento pré e pós-teste com grupo controle não equivalente não inclui uma longa série de observações nem ao longo do tempo nem entre grupos.

Grupo 1	O1	X	O3

Grupo 2	O2		O4

Este não é uma extensão de qualquer dos delineamentos pré-experimentais mas sim uma combinação da comparação com grupo estático com o pré-experimento de pré e pós-teste com um grupo. Combinando as características de ambos, possui mais possibilidade de interpretação do que qualquer um deles. Inclui informações de pré-teste sobre os níveis de linha de base existentes nos grupos e fornece um grupo-controle. Entretanto, o grupo controle não é formado através da distribuição aleatória. É um grupo preexistente ou pré-selecionado, que escolhemos por ser similar, mas não equivalente. Todavia, a simples inclusão de um grupo-controle e de um pré-teste possivelmente permite-nos descartar várias das ameaças à validade interna.

Uma grande vantagem deste delineamento sobre a comparação com grupo estático é que ele pode medir diferenças preexistentes entre grupos. A figura 3.6 ilustra os resultados que são facilmente interpretáveis graças à inclusão do pré-teste. É mais provável, como se

vê na figura 3.6, que pessoas que instalam aquecedor solar já gastassem de início menos energia que outras pessoas.

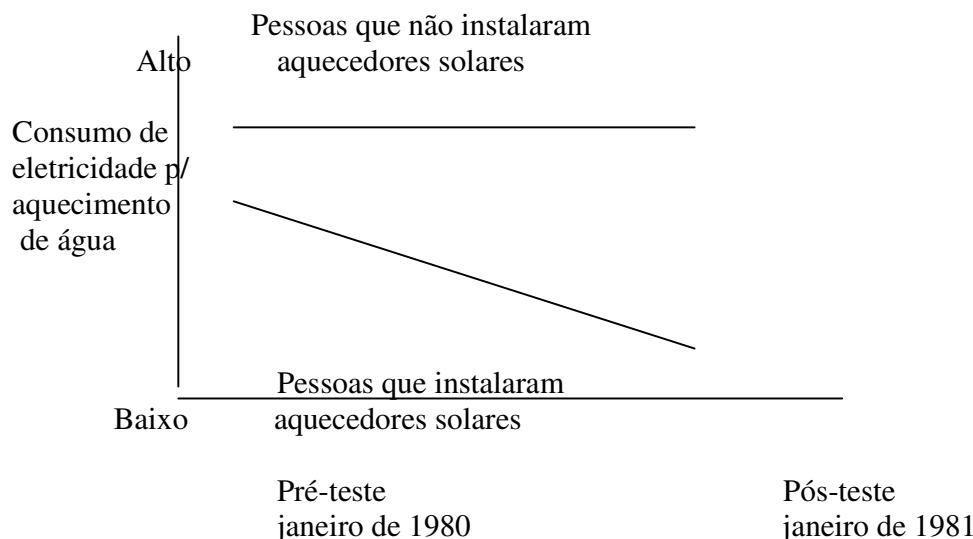


Figura 3.6. Efeitos dos aquecedores solares no consumo elétrico.

O seu interesse em energia solar é uma expressão decorrente de suas inclinações preexistentes. Assim, se tivéssemos apenas a informação do pós-teste, não saberíamos se a diferença refletia uma tendência natural à economia das pessoas que instalam aquecedor solar, ou se refletia as economias produzidas pela energia solar.

Quando olhamos para as diferenças preexistentes, vemos que aqueles que instalaram os aquecedores eram econômicos no início. Eles consumiam menos eletricidade mesmo antes de instalarem os aquecedores. Contudo, a diferença acirrou-se depois do tratamento, e isto sugere que as instalações solares tiveram um efeito.

Há alguma hipótese rival? Você poderia suspeitar que as pessoas que se tornaram interessadas em energia solar iriam naturalmente decrescer o consumo elétrico, mesmo que não houvessem instalado os aquecedores – porque eles estavam conscientes da necessidade de economizar energia e podem ter usado menos água quente em janeiro de 1981. Este delineamento não fornece a informação necessária para que descartemos esta possibilidade – uma interação entre seleção e maturação. Somente se tivéssemos uma série de observações longitudinais poderíamos ver se o grupo de tratamento estava com uma tendência decrescente tanto antes como depois da instalação solar.

Este exemplo ilustra o quanto precisamos considerar não apenas o delineamento, mas também o conteúdo do estudo e o padrão dos resultados para descartar hipóteses rivais. Para alguns conteúdos você pode não considerar plausível que o grupo de tratamento tenha mudado sem o tratamento, de modo a se tornar cada vez mais diferente do grupo-controle.

Para alguns padrões de resultados a hipótese mais plausível é a de que o tratamento tenha produzido os efeitos. Imagine um programa de economia de energia aplicado em pessoas residentes em Estados que usam ar condicionado a maior parte do tempo. Suponha

que o programa de economia aplicado contenha *as mais altas taxas de consumo elétrico* nos Estados de maior consumo. Para estudar os efeitos deste programa, compararíamos o consumo de energia elétrica de pessoas daqueles Estados com o consumo de pessoas de Estados vizinhos que não consumissem tanta eletricidade com ar condicionado. Sabemos que o grupo de tratamento consumia inicialmente mais eletricidade que aqueles residentes nos Estados de comparação, de modo que seus níveis de pré-teste seriam semelhantes a O1 e O2 na figura 3.7. Se o pós-teste mostrar que o grupo de tratamento reduziu seu consumo abaixo do nível do grupo-controle, a hipótese mais plausível é a de que o programa funcionou. Uma intercepção no eixo dos Y como a da figura 3.7 é mais difícil de explicar por qualquer uma das hipóteses rivais, como a interação entre maturação e seleção. Neste exemplo, a tendência constante do grupo controle pode ser considerada a tendência normal; a tendência decrescente do grupo de tratamento pode ser atribuída ao efeito do tratamento. Para descartar o efeito aparente do tratamento, teríamos que considerar as pessoas dos Estados de alto consumo elétrico como um extraordinário grupo de eminentes economizadores que não apenas alcançaram, mas ultrapassaram os níveis de economia das pessoas nos Estados vizinhos. Isto é tão pouco provável, neste caso e em quase todos, que o padrão de intercepção mostrado na figura 3.7 é usualmente interpretado como um efeito do tratamento.

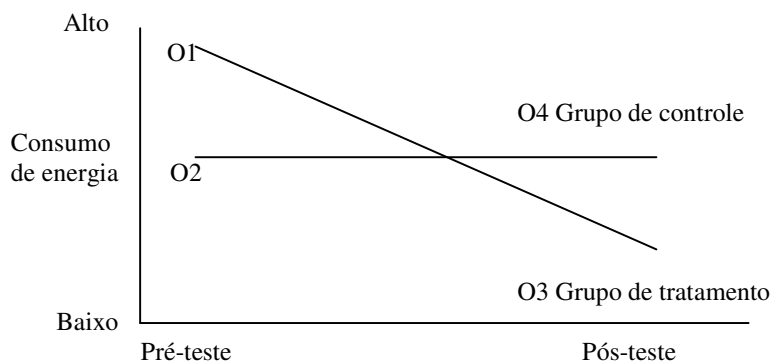


Figura 3.7. Níveis de consumo de energia elétrica por pessoas que vivem nos Estados de tratamento e de controle.

Quando realizar quase-experimentos

Dissemos neste capítulo e no anterior que se você quiser fazer afirmações causais, deverá conduzir um experimento e evitar pré-experimentos. Se você não puder usar distribuição aleatória para projetar um experimento, um quase-experimento é a segunda

melhor escolha. Um quase-experimento está numa posição inferior, não tão boa quanto um experimento, mas muito superior aos pré-experimentos.

Há muitas intervenções sociais – programas de moradia, escola, drogas, impostos – que são importantes de estudar mas que não são realizadas com distribuição aleatória. Para estudar os efeitos de tais inovações, precisamos regredir para alguns delineamentos quase-experimentais. Qual delineamento quase-experimental escolher depende do tipo de dados que se tem disponível. Os delineamentos de série temporal e de análise de descontinuidade da regressão requerem um grande número de pontos de dados e são mais facilmente utilizados no estudo de dados de arquivo – dados que são regularmente coletados, como os dados do censo, índice anual de preço, relatórios mensais, meteorológicos, etc... Embora estes dados houvessem sido originalmente coletados para outros fins, você poderá usá-los para testar hipóteses acerca dos efeitos de novos programas sociais. Às vezes, contudo, dados de arquivo coletados para um propósito não são adequados para responder a outras questões. Por exemplo, se desejássemos estudar os efeitos de um programa de prevenção de estupro sobre a incidência de estupro, deveríamos tentar o uso de um experimento quase-experimental de série-temporal para investigar os efeitos de um novo programa preventivo sobre o número de estupros registrados pela polícia em um período de 15 anos. Esperaríamos encontrar um decréscimo no número de estupros depois de iniciado o programa. Teoricamente, este dado parece ser derivado de um bom delineamento de série temporal, com úteis dados de arquivos. Na prática, os dados seriam difíceis de interpretar porque a incidência real de estupros é muito maior que o número de estupros registrados pela polícia. Os dados de arquivo, portanto, poderiam subestimar o número total de estupros. É concebível, também, que o número de registros policiais de estupros aumentasse, ao invés de decrescer com resultados do programa, se parte do programa consistisse em encorajar as vítimas de estupros a denunciar o crime.

Quando os dados de arquivo existentes não estão disponíveis ou não são adequados, um pesquisador poderia ainda planejar uma série temporal e coletar dados em várias semanas, meses ou anos. Quanto mais tempo levar para coletar tais dados, menor a probabilidade de que você ou alguém mais queira conduzir um quase-experimento de série-temporal porque seria custoso em tempo e esforços. A vantagem do delineamento de série-temporal é que ele inclui um grande número de observações para testar hipóteses rivais; a desvantagem é que é custoso coletar tantos dados se ainda não existem em arquivos.

Validade interna, validade externa e quase-experimentos

Quase-experimentos são intermediários entre um experimento com distribuição aleatória e um pré-experimento, o qual é freqüentemente ininterpretável. Também se constituem num intermediário entre a maximização da validade interna e da validade externa.

O que experimentos fazem para extrair inferência causal e maximizar a validade interna algumas vezes sacrifica a validade externa. Experimentos de laboratório algumas vezes não possuem a mais leve semelhança com os fenômenos do mundo real que eles pretendem reproduzir, e mesmo experimentos de campo às vezes perdem validade externa simplesmente porque os sujeitos sabem que são cobaias experimentais submetidas à distribuição aleatória. Quase-experimentos freqüentemente evitam estes prejuízos. Uma análise de série temporal dos efeitos de programas de prevenção de estupro não cria “efeitos de cobaia” porque as vítimas nem mesmo sabem que estão sendo estudadas. Quase-experimentos podem ser menos intrusivos que os experimentos, porque permitem a ocorrência dos processos de seleção natural. As pessoas estão acostumadas a escolher seus próprios tratamentos ou a serem selecionadas para tratamentos segundo algum critério, como uma expectativa de grande sucesso acadêmico. Não estão acostumadas a serem distribuídas aleatoriamente pelos tratamentos, exceto quando se trata, reconhecidamente, de algum tipo de loteria.

Os processos de distribuição do sujeito que os quase-experimentos incluem, são processos de seleção que naturalmente ocorrem no mundo. Estes processos de classificação não-aleatórios tornam difícil desvencilhar os efeitos do tratamento de outros efeitos rivais, particularmente a seleção, mas eles também fornecem aos estudos quase-experimentais um componente de validade externa que os experimentos carecem.

Eles nos possibilitam estudar os efeitos que tratamentos possuem sobre aquele segmento da população que também é o mais provável de escolher ou ser escolhido para o tratamento. Nem todos escolheriam ir a uma faculdade experimental, fazer psicoterapia ou trabalhar quatro horas por semana. Para tratamentos como estes, que as pessoas são capazes de escolher por elas próprias, não faz sentido descartar as tendências de seleção pessoal que existiriam naturalmente.

Poderíamos não querer distribuir aleatoriamente todas as pessoas para algum nível destes tratamentos. Poderíamos elaborar um experimento entre aquelas pessoas que normalmente escolheriam tais tratamentos, caso recrutássemos mais voluntários do que cada programa poderia comportar. Se selecionássemos aleatoriamente os principiantes entre os candidatos voluntários, poderíamos elaborar um experimento. Contudo, a maioria das pessoas não estão acostumadas a serem sorteadas para tais tipos de programas, o que requereria iludir o participante ou educá-lo para tornar o sorteio aceitável. Não defendemos o engodo. Gostaríamos de ver alguma forma de educação do público para que se tornasse possível o uso de sorteio na avaliação de certos programas – ainda que isto aumentasse a consciência das pessoas de que elas são cobaias e pudesse por isto, diminuir a validade externa de um estudo.

Resumo

Delineamentos quase-experimentais fornecem um meio de estudar alguns tratamentos sociais que naturalmente ocorrem. Eles são intermediários entre um experimento que possui alta validade interna e os fracos pré-experimentos que quase não possuem, no geral, validade interna. Quase-experimentos nos possibilitam descartar algumas ameaças à validade porque eles incluem mais pontos de dados que os pré-experimentos. O número de delineamentos quase-experimentais que um pesquisador criativo pode construir é ilimitado. Apresentamos três tipos que são extensões de pré-experimentos, os quais mostram como pontos de dados adicionais tornam um delineamento, antes de difícil interpretação, interpretável. Um determinado pesquisador pode delinear um quase-experimento, ainda não pensado, coletando dados de um número suficiente de sujeitos, num tempo suficiente, para descartar muitas ameaças à validade interna, de tal modo que mesmo sem a distribuição aleatória será possível inferir causas e efeitos.

Delineamento da análise de descontinuidade da regressão

Grupo 1 *O1*

 Grupo 2 *O2*

 Grupo 3 *O3*

 Grupo 4 *XO4*

 Grupo 5 *XO5*

 Grupo 6 *XO6*

O delineamento da análise de descontinuidade da regressão é uma extensão do delineamento pré-experimental de comparação com grupo estático *X O1* .

não X O2

O delineamento de grupo controle estático é inadequado porque *seleção* é uma hipótese rival plausível sempre presente para uma diferença entre os dois *O*s.

O delineamento da análise de descontinuidade da regressão, com sua longa série de grupo-controle, fornece informações sobre a plausibilidade de tais diferenças preexistentes entre os grupos e permite ao experimentador descartar seleção como uma hipótese rival. O pesquisador pode examinar as diferenças entre os vários grupos de não-tratamento para saber quais são as diferenças que ocorrem naturalmente entre os grupos. Os grupos num delineamento de análise de descontinuidade da regressão são todos classificados ou ordenados segundo um critério, tal como necessidade financeira, e aquelas acima ou abaixo de um ponto limite recebem o tratamento, por exemplo, uma bolsa de estudos.

O delineamento de análise de descontinuidade da regressão é um delineamento de corte transversal; ele examina os efeitos do tratamento olhando através de muitos grupos de pessoas e compara aqueles acima do ponto limite com os abaixo deste. Os delineamentos de série temporal são delineamentos longitudinais; eles examinam os efeitos do tratamento olhando *ao longo* da linha de tempo de uma única pessoa ou grupo e comparam as observações realizadas antes do tratamento com as realizadas após. Conforme esquematizado anteriormente, os dois tipos de delineamentos parecem muito diferentes. Conforme a representação gráfica das figuras 3.4 e 3.5, os dois parecem similares.

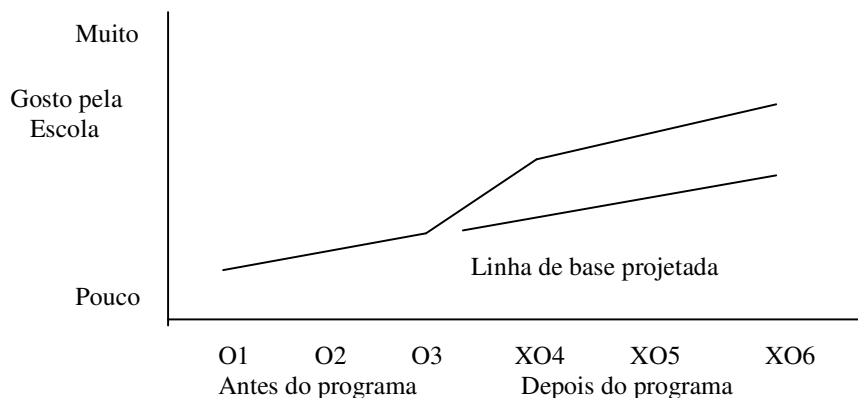


Figura 3.4. Série temporal para uma criança. Efeitos do programa de aprimoramento acadêmico no gosto das crianças pela escola.

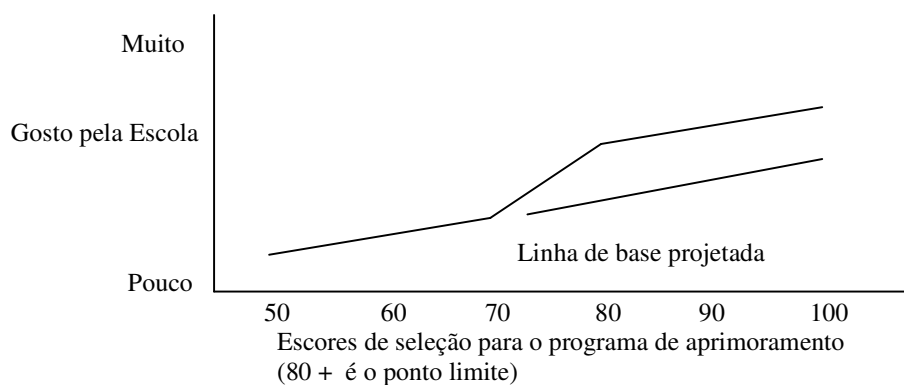


Figura 3.5. Análise de descontinuidade da regressão para crianças acima e abaixo do ponto limite. Efeitos do programa de aprimoramento acadêmico no gosto das crianças pela escola.

Embora a análise estatística seja diferente (o leitor interessado pode consultar Cook e Campbell, 1979 e Judd e Kenny, no prelo, para maiores detalhes), a lógica subjacente à interpretação desses dois delineamentos é semelhante. Em ambos os casos podemos projetar o ponto longo antes do tratamento e comparar a estimativa projetada com os resultados observados. Se estes últimos diferirem dos projetados, isto será uma evidência do efeito do tratamento. Em ambos os casos assumimos que, se houvesse uma tendência preexistente, esta tendência continuaria normalmente e não deveria ser confundida com um efeito de tratamento. Por exemplo, se a criança cuja série temporal que mostramos na figura 3.4 gostou cada vez mais da escola a cada ano que passa, esperamos que esta tendência continue. A mudança na intercessão mostra que

o gostar subiu para um novo nível após o programa de aprimoramento acadêmico; o programa deu um impulso na criança. O delineamento de análise de descontinuidade da regressão mostra o mesmo efeito (fig. 3.5). Se as crianças que recebem pontuações mais altas nos testes gostam mais de escola naturalmente e se as crianças são selecionadas com base nas pontuações que obtêm nos testes, esperaríamos que a tendência continuasse ao longo da linha programada mesmo que não houvesse efeito real do programa. O deslocamento ascendente da curva, a mudança na intercepção no eixo dos Y, demonstra o efeito do programa além da tendência natural.

Os delineamentos de série temporal e da análise da descontinuidade da regressão atingem seu alto nível de interpretabilidade graças ao acréscimo de observações. A série temporal acrescenta observações ao longo do tempo – estendendo a observações para diante e para trás. O delineamento de análise de descontinuidade da regressão acrescenta observações projetadas através de diferentes grupos, todos no mesmo momento. As observações adicionais permitem-nos, em ambos os casos, comparar a diferença entre observações imediatamente adjacentes ao tratamento com as diferenças entre pares de observações anteriores ao tratamento e pares posteriores a ele. As séries também nos permitem examinar tendências que ocorrem naturalmente e compara as tendências projetadas com as reais.