

DIFERENÇAS E RELAÇÕES EM DADOS QUANTITATIVOS

Sally Barnes, Escola de Pós-graduação em Educação da Universidade de Bristol, Reino Unido.

Cathy Lewin, Instituto de Pesquisa Educacional e Social, Universidade Metropolitana de Manchester, Reino Unido.

Resumo

- Escolha de técnicas estatísticas adequadas.
- Avaliação de resultados estatísticos:
 - Testes de significância.
 - Teoria probabilística.
- Em busca de diferenças.
- Em busca de correlações.

Conceitos fundamentais

Introdução à estatística inferencial

Descrever conjuntos de dados é frequentemente apenas o primeiro passo na análise de dados. Em geral, o que mais nos interessa fazer é perguntar sobre os dados, estudar relações entre diferentes coisas que medimos. A *estatística inferencial* abrange todas as técnicas que nos permitem examinar em profundidade as relações entre variáveis. Elas são uma maneira muito eficaz de fazer perguntas sobre dados numéricos. Existem três enfoques principais: examinar as diferenças; exami-

nar a índole e a extensão das relações; classificar e fazer previsões (cf. JONES, cap. 27). Este capítulo tem por objetivo apresentar alguns dos conceitos fundamentais e alguns dos procedimentos mais comuns para examinar diferenças e relações. O uso de um pacote estatístico pode dar conta dos cálculos matemáticos necessários para cada técnica, mas nenhum desses pacotes pode decidir qual técnica usar numa determinada situação nem como interpretar o significado dos resultados.

Os procedimentos estatísticos inferenciais dividem-se em dois tipos principais: *paramétricos* e *não paramétricos*. As estatísticas paramétricas baseiam-se nos princípios da curva normal (cf. LEWIN, cap. 25). Portanto, os dados têm de estar normalmente distribuídos e ser de nível intervalar (algum tipo de contagem e não categorias nem uma resposta ordenada) para que se possam usar estatísticas paramétricas. Os tamanhos de amostra recomendados variam conforme o tipo de teste estatístico, mas a regra prática geral é ter uma amostra mínima de cerca de trinta. Quando as amostras têm menos de trinta elementos, é melhor usar estatísticas não paramétricas, isto é,

aquelas sem distribuição (não baseadas nos princípios da curva normal). A análise das variáveis de categoria – nominais ou ordinais (cf. LEWIN, cap. 25) – efetua-se com técnicas não paramétricas porque normalmente não se deveria calcular nem a média nem o desvio padrão.

Uma das tarefas mais difíceis ao analisar dados é a de escolher a técnica estatística mais adequada, que aborde a pergunta da pesquisa e se adéque aos dados que foram coletados. Howell (2002: 11) oferece um diagrama útil para escolher uma técnica apropriada em função de:

- o tipo de dados (categóricos, ordinais, intervalares);
- se o teste procura diferenças ou relações;
- o número de grupos de participantes (dois ou mais);
- se os grupos são *dependentes* ou *relacionados* (um único grupo exposto a diferentes condições ou testado em diferentes momentos no tempo) ou *independentes* (dois ou mais grupos de participantes não relacionados entre si);
- se o teste deveria ser paramétrico ou não paramétrico.

Diversos textos orientam também na escolha da técnica apropriada; por exemplo, Pallant (2007, cap. 10), Salkind (2007, cap. 9) e Field (2009).

Avaliação de resultados estatísticos

Realizar um procedimento estatístico é um processo de três etapas. Primeiro aplicamos diversas técnicas descritivas para certificar-nos de que os dados são confiáveis, válidos e adequados aos critérios estabelecidos para a análise estatística (cf. FIELD, 2009; SALKIND, 2007). Em seguida, escolhemos e computamos o teste estatístico apropriado e, finalmente, realizamos o correspondente teste de significância para ver qual a probabilidade de obtermos o resultado do teste supondo que não haja diferença nem relação alguma. Com o teste de significância estatística podemos dizer em que medida estamos confiantes de o resultado da análise dos dados extraídos da amostra ser um resul-

tado “real” ou se ele surgiu “por acaso” (em lugar de decorrer de erro de amostragem).

Com a teoria probabilística podemos perguntar quais as chances de obter um resultado similar a partir de outra amostra extraída da mesma população. Quanto melhores as chances, mas confiantes tendemos a estar na validade do nosso resultado. Contudo, é essencial lembrarmos que o teste de significância baseia-se em probabilidade e que estamos testando a hipótese nula (cf. FIELD, 2009, cap. 2; SALKIND, 2007, cap. 9). É possível supor que a hipótese nula é correta quando na verdade ela não é, no que se denomina erro Tipo 2. Inversamente, pode-se rejeitar a hipótese nula quando na verdade ela é correta, no que se denomina erro Tipo 1. O teste de significância (FREEDMAN et al., 2007, cap. 26) implica dois pressupostos importantes. O primeiro é que a seleção aleatória tem de ser usada para escolher a amostra original de uma população bem-definida. O segundo, que as medições têm de ser válidas e confiáveis, o que costuma ser difícil de conseguir nas ciências sociais (FREEDMAN et al., 2007).

O teste de significância utiliza “graus de liberdade” (para uma explicação matemática, cf. HOWELL, 2002: 56; FIELD, 2009: 37). Os graus de liberdade são o número de elementos num conjunto (valores de uma variável, p. ex.) que podem variar e seu cálculo difere segundo a técnica estatística escolhida. Em uma sala de aula com 30 carteiras, o primeiro aluno a chegar pode escolher onde se sentar, como pode também cada um dos seguintes 28 alunos, embora partindo de um número de possibilidade cada vez menor. Mas o último aluno a chegar não terá escolha porque então só restará uma carteira. Portanto, os graus de liberdade neste caso serão o número de carteiras menos uma: 29. Em termos simplistas, os graus de liberdade aproximam-se do tamanho da amostra e por isso dão informação útil para a leitura de relatórios de pesquisa quantitativa.

Os procedimentos estatísticos paramétricos têm uma fórmula para calcular os graus de liberdade. A fórmula é, em muitos casos

$n - 1$ (se há apenas um grupo de pessoas, n = número no grupo)

ou:

$n_1 + n_2 - 2$ (se há dois grupos, n_1 = número no grupo 1 e n_2 = número no grupo 2)

Os resultados de testes estatísticos são apresentados em notação convencional que geralmente inclui dados: o teste estatístico utilizado, o resultado real, os graus de liberdade e a probabilidade de se obter o resultado pressupondo que a hipótese nula é verdadeira (cf. COOLIDGE, 2006; FREEDMAN et al., 2007).

Para avaliarem a significância dos resultados os pesquisadores apresentam suas conclusões em função de diferentes níveis de probabilidade. Por exemplo, o nível de 0,05 ($p < 0,05$, onde “ p ” é o valor da probabilidade) significa que a probabilidade de se obter um resultado quando a hipótese nula é verdadeira é inferior a cinco em 100. Em ciências sociais costumamos usar os valores de probabilidade $p < 0,05$, $p < 0,01$ e $p < 0,001$, baseados na curva de distribuição normal.

Há várias precauções que é preciso tomar na interpretação de resultados estatísticos e suas probabilidades associadas (FREEDMAN et al., 2007).

- Os níveis de significância 0,05 e 0,01 são arbitrários. Eles são usados por simples convenção comum e só por esta razão. Com efeito, faz mais sentido dar o valor real da probabilidade do que dizer que $p < 0,05$ (os pacotes estatísticos informam o valor real da probabilidade).
- O valor de p tem relação com o tamanho da amostra. Quanto maior a amostra, menor a diferença ou mais fraca a relação necessária para alcançar significância estatística.
- Os pesquisadores sempre deveriam resumir seus dados por meio da estatística descritiva (cf. Lewin) para que os leitores possam tirar suas conclusões sobre a importância de quaisquer valores de p estatisticamente significativos.
- Os valores de p só têm importância para amostras baseadas em probabilidade, ou modelos ao acaso. Quando se dispõe de dados da

totalidade de uma população é inútil fazer um teste de significância porque seria só comparar essa população com ela mesma.

- Os testes de significância não verificam se há erros de projeto. Se o pesquisador escolheu a amostra incorretamente, usou medidas não válidas ou não dignas de confiança ou aplicou um teste inadequado, não há como detectar isto estatisticamente.

Um aspecto do teste de significância que muitas vezes traz dificuldade aos pesquisadores principiantes é a ideia de testes com uma ou duas caudas. Na curva normal há duas “caudas”, cada uma estendendo-se para um lado até o infinito. Quando se tem uma pergunta de pesquisa muito específica quanto à direção do resultado (i. é, espera-se um valor médio maior de uma variável num grupo em comparação com um segundo grupo), pode-se usar um teste de significância monocaudal. Por exemplo, a pergunta “Os garotos fazem mais abdominais superiores em cinco minutos do que as meninas?” é bem específica e direcional; podemos checar a estatística de teste gerada se os garotos fazem mesmo mais abdominais do que as meninas. Mas se perguntamos “Há diferença no número de abdominais superiores que garotos e meninas fazem em cinco minutos?”, isto sugere que qualquer dos grupos poderia fazer mais abdominais. Neste caso não podemos prever a direção de mudança e optamos por considerar ambas as possibilidades e aplicar um teste bicaudal (os garotos fazem mais abdominais do que as meninas ou as meninas fazem mais do que os garotos).

Fundamentação da análise em estatística inferencial: buscando diferenças

Quão diferentes têm de ser duas coisas antes de nos empolgarmos muito com os resultados? Ou seja, será que a diferença entre diversos grupos é real ou se apresentou por acaso? Para verificar isto nós fazemos testes estatísticos que estudam as diferenças e depois, com métodos de cálculo de significância estatística, avaliamos a probabilidade de se chegar a um determinado resultado supondo que a hipótese nula é verdadeira.

Técnicas paramétricas para detectar diferenças

Para números intervalares ou racionais, as principais técnicas paramétricas são o teste t (quando temos apenas dois grupos) e a Análise de Variância quando temos mais de dois grupos.

Aplicamos o teste t quando queremos ver se há diferença significativa entre as médias de duas amostras, isto é, para ver se duas amostras podem ser consideradas provenientes da mesma população ou de duas diferentes. Nossa hipótese nula é que não há diferença entre as médias de duas amostras. Nossa hipótese alternativa é que há uma diferença estatisticamente significativa entre as médias.

Há duas fórmulas de teste t a se considerar quando se trata de amostras independentes, isto é, dois grupos de pessoas entre os quais não existe relação alguma (p. ex., homens/mulheres, médicos/pacientes, crianças de 10/12 anos). Usa-se o teste *t de modelo separado* quando os grupos têm o mesmo número de participantes. A fórmula de teste t usada quando os tamanhos das amostras são diferentes é o teste *t de modelo de variância combinada*. A diferença nas duas fórmulas está na maneira de calcular a variância (distribuição dos dados ao redor da média).

O teste t pressupõe que as variâncias dos dois grupos são iguais. O teste de Levene examina a igualdade de variâncias de dois grupos. A hipótese nula é que as duas variâncias são iguais. Se o resultado do teste de Levene é significativo em $p < 0,05$, a hipótese nula é rejeitada e a hipótese alternativa é aceita; as variâncias não são iguais. Se os resultados do teste de Levene não são significativos, aceita-se a hipótese nula e se admite que as variâncias são iguais. Cabe então ao pesquisador decidir qual teste t é adequado, com base nos resultados do teste de Levene.

O resultado do teste t é denominado “valor de t”. Quanto maior o valor resultante de t, maior a diferença entre as duas médias. Para interpretar o valor de t nós verificamos o valor de probabilidade associado com esse t para nossa amostra, levando em conta o tamanho da amostra. Depois podemos interpretar a conclusão e determinar se é ou não provável que haja uma diferença real entre os dois grupos e, nesse caso, qual o nosso grau de confiança em que tal diferença exista.

Ao apresentar os resultados de testes t é importante especificar qual foi a fórmula de teste t utilizada e incluir os graus de liberdade (df) para essa técnica, por exemplo:

$$t = 4,52, df = 40, p < 0,01$$

Isto nos diz que havia 42 pessoas ao todo nos dois grupos, uma vez que para esta técnica estatística os graus de liberdade são calculados somando as quantidades de pessoas nos dois grupos e diminuindo dois. Também nos diz que o valor real de t é 4,52 e que a probabilidade de aceitar que há uma diferença, supondo que a hipótese nula é verdadeira, é inferior a um em cem. Podemos concluir, portanto, que as médias de nossos dois grupos são diferentes e que esta diferença é estatisticamente significativa.

Há uma terceira fórmula de teste t que se utiliza quando há apenas um grupo de pessoas e os dados são colhidos dos mesmos participantes em duas condições diferentes (esquema *dependente* ou *relacionado*). Este é o teste t chamado de “amostra pareada” e seu uso é mais comum em projetos pré-teste/pós-teste. Por exemplo, os professores aplicam um teste na turma ao começar o ano letivo para avaliar o conhecimento dos alunos sobre um tema. No fim do ano letivo, os alunos são submetidos a um novo teste. Obviamente, os professores esperam que os alunos se saiam melhor no teste do fim do ano do que no teste inicial. Um teste t de amostra pareada mostraria se houve ou não uma diferença estatisticamente significativa entre as médias de pontos, indicando uma melhora.

Para testes em busca de diferenças entre três ou mais grupos se aplica a análise de variância (Anova) em lugar do teste t (cf. FIELD, 2009; SALKIND, 2007). Uma Anova detectará se há ou não uma diferença significativa entre as médias de vários grupos diferentes. A hipótese nula para este teste é que não há diferença entre as médias para os distintos grupos. Por exemplo, seria preciso fazer uma Anova para procurar qualquer variação entre três corretores de provas, cada um dos quais corrige 100 provas. Aqui, a variável testada seria a nota dada a cada aluno e a variável de agrupamento seria o corretor de provas. Neste caso, seria desejável que não houvesse diferença (i. é, que o resultado do

teste não fosse significativo), o que demonstraria a coincidência entre os corretores de provas.

Com a Anova procuramos ver se a diferença entre os grupos é maior que a diferença dentro deles. O resultado do teste de Anova é chamado de razão F. Esta razão compara a variabilidade (variância ou soma de quadrados) entre grupos (as diferenças que são previsíveis devido ao fator de agrupamento) com a variabilidade dentro dos grupos (as diferenças causadas por fatores eventuais, independentemente do grupo em questão). Quanto maior a variabilidade entre grupos, menor a variabilidade dentro dos grupos, maior será a razão F e com mais confiança poderemos aceitar que existe uma diferença entre os grupos.

Tabela 26.1 Exemplo de resultados de Anova

	df	Soma de quadrados	Quadrado médio	F	p
Entre grupos	2	198,38	99,19	12,45	<0,01
Dentro dos grupos	72	95,62	7,97		
Total	74				

Para esta técnica é necessário calcular dois graus de liberdade (df) diferentes. Os graus de liberdade para a medição entre grupos são o número de grupos menos um. Os graus de liberdade para a medição dentro dos grupos são o número total na amostra menos os graus de liberdade na medição entre grupo. Costuma-se apresentar os resultados da Anova em uma tabela (cf. Tabela 26.1).

O resultado da Anova indica se há diferença significativa entre as médias dos grupos como um todo, mas não diz se essa significância se dá entre todos os grupos ou só entre alguns deles. Para descobrir quais são os pares de grupos significativamente diferentes nós efetuamos uma análise *post hoc* (posterior à análise inicial). Os testes *post hoc* de uso mais habitual são a correção de Bonferroni, o teste HSD de Tukey e o teste de Scheffé (cf. FIELD, 2009, cap. 10).

Técnicas não paramétricas para detectar diferenças

No caso de dados nominais ou ordinais não é adequado empregar testes t, Anova e outras técnicas paramétricas. Recorremos aqui a técnicas que são isentas de distribuição, isto é, não se baseiam nos princípios da curva normal. Há muitos testes de diferença não paramétricos (cf. FIELD, 2009; SALKIND, 2007; SIEGEL & CASTELLON, 1988). Cada técnica tem uma finalidade específica e requisitos concretos quanto ao nível de medição (i. é, dados nominais ou ordinais), número de categorias, número de grupos e tipo de diferença estudada. É fundamental saber que tipos de diferenças estão sendo estudados, levando em consideração os tipos de dados para escolher a técnica mais adequada. Também se devem usar estas técnicas com dados intervalares quando se conta com amostras muito pequenas ou não se cumprem os pressupostos subjacentes aos testes paramétricos (p. ex., quando os dados não têm distribuição normal, cf. LEWIN, cap. 25). De modo geral, os testes não paramétricos se valem da classificação de dados para comparar grupos e baseiam-se em menos pressupostos. O teste U de Mann-Whitney é o equivalente não paramétrico do teste t com amostra independente. O teste de ordem com signo de Wilcoxon é o equivalente não paramétrico do teste t de amostra pareada. O teste de Kruskal-Wallis compara diferenças entre diversos grupos e é similar à Anova.

Fundamentação da análise em estatística: buscando relações

Para estudar as relações entre variáveis nós aplicamos um processo chamado de correlação. Os procedimentos de correlação são muito comuns na pesquisa em ciências sociais. Podemos usá-los para comparar grupos de indivíduos quanto a diferentes atributos. Por exemplo, podemos correlacionar dois grupos diferentes de desempenho de crianças de 10 anos num exame de matemática, ou comparar os resultados de um grupo de crianças em leitura e escrita com seus resultados em aritmética, ou estudar a relação entre estatura e peso. Em todos os casos, estamos analisando como se relacionam as formas das distribuições de duas variáveis.

Na condição paramétrica, o coeficiente de correlação de Pearson (r) indica o grau de associação linear entre as duas variáveis, isto é, a força da relação. Por linear nos referimos a em que medida a representação gráfica da relação entre essas variáveis é reta. Uma das implicações do estudo de relações lineares é que as coisas que estão sendo medidas devem poder formar uma linha reta. Só se pode examinar a linearidade utilizando números intervalares e racionais onde é possível calcular médias e desvios padrão.

O valor do coeficiente de correlação (r) será entre -1 e +1. Quando $r = 1$ ou -1 , os dados formam uma linha perfeitamente reta. Quando r é positivo, a correlação também é positiva, o que significa que os valores de ambas as variáveis aumentam juntos. Quando r é negativo a correlação é negativa, o que significa que ao aumentar o valor de uma variável, o valor da outra variável diminui. Quando $r = 0$, sabemos que não há associação alguma entre as duas variáveis. Valores de r entre 0 e +/- 0,33 mostram as diferentes intensidades da relação entre as duas variáveis. Não existe uma regra definitiva quanto à interpretação da força de uma correlação. Em geral, contudo, uma relação entre 0 e +/- 0,33 é considerada fraca, um valor r entre 0,34 e 0,66 indica uma relação de força média e, se r está entre 0,67 e 0,99, a relação é forte.

Técnicas de gráfico de dispersão

Uma maneira de entender o que queremos dizer com relações lineares é desenhar um gráfico de dispersão das duas variáveis. Quanto mais de "linha" ou cigarro for a forma do gráfico de dispersão, maior a relação linear. Se o gráfico de dispersão é uma linha reta, a correlação é perfeita. Por exemplo, o gráfico de dispersão na Figura 26.1 mostra uma relação não significativa entre horas de sono e ano de nascimento num conjunto de 90 pessoas que frequentam um curso de estatística.

Quando o gráfico de dispersão resulta num monte de pontos espalhados pela folha de papel, é improvável que haja uma correlação e cabe supor que o coeficiente de correlação seja próximo de zero.

O seguinte exemplo, na Figura 26.2, mostra uma relação linear entre uma nota dada pelo tra-

balho do curso e a nota final de módulo para 193 alunos. Seria previsível este gráfico de dispersão ser quase uma linha reta, uma vez que é óbvio que uma nota no trabalho do curso seria incluída na nota final.

Requisitos de cálculo

Para calcular o coeficiente de correlação é preciso ter duas amostras de valores que, geralmente chamadas de variáveis X e Y.

Figura 26.1 Exemplo de gráfico de dispersão que indica não haver relação entre horas de sono e idade

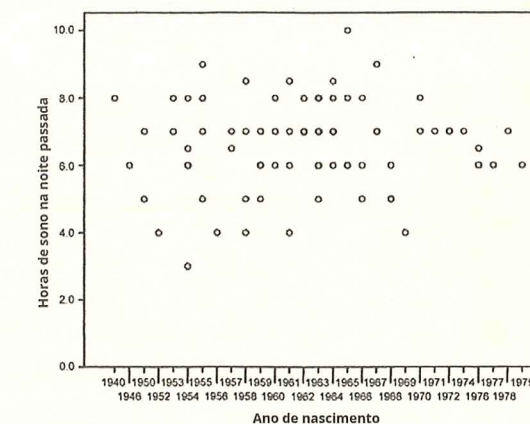
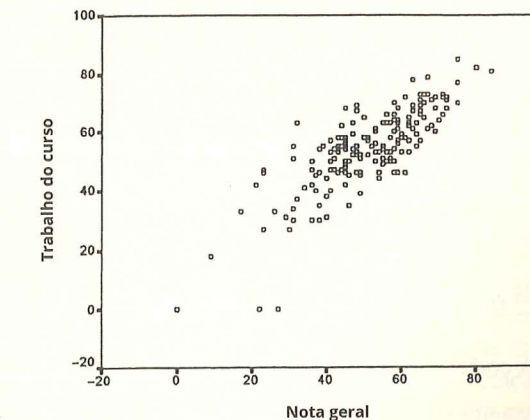


Figura 26.2 Exemplo de gráfico de dispersão que sugere uma relação linear entre a nota dada por trabalho do curso e a nota final do módulo



Os valores estão pareados de alguma maneira. Na maioria dos casos de pesquisa em ciências sociais este pareamento é por indivíduo. Comumente temos valores para um grupo de pessoas sobre duas medidas. Por exemplo, poderíamos correlacionar o desempenho de uma determinada turma de alunos no exame do primeiro ano e do segundo ano. Isto nos mostraria o grau de associação entre os resultados dos dois exames. Neste caso, cada indivíduo teria dois valores. Outra possibilidade é, se aplicarmos o mesmo exame dois anos consecutivos, calcular a correlação entre os alunos do ano passado e os deste ano.

Confirma-se que a relação entre a nota do trabalho do curso e a nota geral para 193 alunos conforme a Figura 26.2 é muito forte e o resultado de Pearson mostra que:

$$r = 0,908, n = 193, p < 0,01$$

Neste caso, normalmente se menciona o tamanho da amostra (n) em lugar dos graus de liberdade. (Às vezes se opta pelos graus de liberdade, calculados como o número de pares menos dois – neste caso, $df = 191$). É importante dar o tamanho da amostra por ser um fator que muito influi na significância estatística da correlação (cf. Figura 26.3).

Há diversos métodos correlacionais. A escolha do método depende da escala de medição das duas variáveis. O método mais comum é o coeficiente de correlação produto-momento de Pearson, usado quando ambas as variáveis são intervalares ou racionais. O método correlacional adequado

quando uma ou as duas variáveis são ordinais é o coeficiente de correlação de ordem de classificação de Spearman, chamado *r* de Spearman ou, se o conjunto de dados é pequeno e com muitas observações igualmente classificadas, tau de Kendall (cf. FIELD, 2009: 179-181). Também se podem aplicar estas técnicas quando não se cumprem os pressupostos para testes paramétricos (i. é, os dados não apresentam distribuição normal). Existem também métodos correlacionais para usar com dados nominais (qui-quadrado, a seguir).

Interpretação de correlações

Para interpretar uma correlação nós usamos três dados: r (o coeficiente de correlação que indica a intensidade da relação, descrito anteriormente), a significância estatística de r (a probabilidade de aceitar que há uma relação, supondo que a hipótese nula é verdadeira) e o tamanho da amostra. Todos estes fatores podem ser levados em consideração para interpretar correlações. A significância estatística é pertinente, por certo, mas também o é a força da relação. E o tamanho da amostra influi grandemente na correlação. Com amostras muito grandes, coeficientes de correlação fracos podem ser estatisticamente significativos; já com amostras pequenas, só correlações muito fortes chegarão a ter significância estatística.

Como ocorre com outros testes, nós começamos por decidir se estamos ou não supondo um resultado em determinada direção, para então usarmos

Correlações

		Trabalho do curso	Nota geral
Trabalho do curso	Correlação de Pearson	1	0,908**
	Sig. (bicaudal)		0,000
	N	193	193
Nota geral	Correlação de Pearson	0,908**	1
	Sig. (bicaudal)	0,000	
	N	193	193

** A correlação é significativa no nível 0,01 (bicaudal)

Figura 26.3 Exemplo de resultado de teste de correlação

um nível de significância monocaudal ou bicaudal. Em seguida decidimos qual é o nível mínimo de significância que estamos dispostos a aceitar – 0,05, 0,01, 0,001 ou qualquer um. A escolha do nível de significância a se adotar como padrão para um determinado estudo depende da confiabilidade e da validade dos dados analisados, bem como da pergunta original da pesquisa que está sendo tratada. Em todos os casos será preciso explicar as razões que levaram a estas decisões ao apresentar os resultados.

Qui-quadrado

A técnica de qui-quadrado é uma das mais habituais para examinar relações com dados nominais e/ou ordinais. Trata-se de uma estatística muito infrequente porque faz duas coisas num só teste. É um teste de independência e também de associação. No teste de qui-quadrado estamos testando formalmente a hipótese nula de duas variáveis serem independentes. Se rejeitamos a hipótese nula porque temos um resultado significativo (isto é, se confiamos em que há uma relação entre as duas variáveis) podemos fazer mais testes para estudar o tipo de relação que existe entre duas variáveis e a força dessa relação.

Embora seja um teste não paramétrico, o qui-quadrado depende de pressupostos que precisam ser cumpridos. Primeiro, cada caso ou pessoa deve contribuir para apenas uma célula na tabela de contingência (i. é, as características para cada variável devem ser mutuamente exclusivas). Isto pode ser testado estatisticamente em muitos casos. Se colhemos informação sobre gênero e cor de olhos de 100 alunos, é de se esperar que cada pergunta receba apenas uma resposta (cada aluno será classificado como pertencente a um gênero e dotado de olhos de determinada cor). Em segundo lugar, o teste de qui-quadrado funciona comparando a distribuição de observações em células da tabela de contingência com a distribuição que se poderia ter previsto se não houvesse associação, gerando uma *contagem prevista* em cada célula. Se alguma das contagens esperadas for inferior a 5, o teste de qui-quadrado será inválido, ainda que em grandes tabelas de contingência seja aceitável ter até 20% de células com contagens esperadas inferiores a 5.

Tabulação cruzada de Gênero* Uso de computador fora da escola

Contagem

		Usa um computador fora da escola		Total
		Não	Sim	
GÊNERO	Masculino	322	2.963	3.285
	Feminino	288	2.566	2.854
Total		610	5.529	6.139

Figura 26.4 Exemplo de uma tabulação cruzada simples

O primeiro passo ao interpretar um teste qui-quadrado significativo é observar cuidadosamente os valores e/ou as porcentagens numa tabela de frequência bidirecional para compreender a associação entre diferentes variáveis. A tabulação cruzada (também chamada de tabela de contingência) indica quantos casos (pessoas, p. ex.) apresentam cada uma das possíveis combinações das duas variáveis nominais que estão sendo testadas. O exemplo dado na Figura 26.4 mostra quantos garotos e meninas usam computadores fora da escola e quantos não o fazem.

No entanto, as porcentagens em cada coluna e fileira podem facilitar a interpretação da relação entre gênero e uso de computador fora da escola na amostra que vemos na Figura 26.5.

As porcentagens ajudam-nos a levar em conta que são muito mais os alunos que dizem usar computadores fora da escola do que os que não usam, e que o número de garotos é ligeiramente maior que o de meninas nesta amostra. A estatística qui-quadrado mais usada é a de Pearson (χ^2). No exemplo aqui apresentado (cf. Figura 26.5), ela é:

$$\chi^2 = 0,142, df = 1, p = 0,706$$

Os df para qui-quadrado são (número de fileiras – 1) multiplicado por (número de colunas – 1). Para uma tabela com duas fileiras e duas colunas (Tabela 2 x 2), $df = 1$. A estatística também nos diz que para esta distribuição a probabilidade de se aceitar que existe uma associação, supondo que a hipótese nula é verdadeira, é muito alta porque $p > 0,05$ ($p = 0,706$). Portanto, podemos estar certos de que as duas variáveis são independentes e não há associação alguma entre gênero e uso declarado de computadores fora da escola.

Tabulação cruzada de Gênero* Uso de computador fora da escola

			Usa um computador fora da escola		Total
			Não	Sim	
GÊNERO	Masculino	Contagem	322	2.963	3.285
		% no GÊNERO	9,8%	90,2%	100%
		% em Usa um computador fora da escola	52,8%	53,6%	53,5%
GÊNERO	Feminino	Contagem	288	2.566	2.854
		% no GÊNERO	10,1%	89,9%	100%
		% em Usa um computador fora da escola	47,2%	46,4%	46,5%
Total		Contagem	610	5.529	6.139
		% no GÊNERO	9,9%	90,1%	100%
		% em Usa um computador fora da escola	100%	100%	100%

Figura 26.5 Exemplo de tabulação cruzada com porcentagens de fileira e coluna

É da estatística de qui-quadrado que a pesquisa em ciências sociais faz maior uso e abuso. Abuso por parte de pesquisadores que elaboram interpretações sobre a associação entre variáveis para além do que o resultado indica. A qui-quadrado só nos diz se há associação ou independência entre duas coisas. Para examinar a intensidade da relação é preciso efetuar uma das análises *post hoc* em medições de associação como Φ ou V de Cramer, Λ de Goodman e Kruskal (cf. FIELD, 2009: 695). Similares aos coeficientes de correlação, estes testes geram resultados que dão uma medida da força entre 0 e 1. Quanto mais próxima de 1, mais forte a associação.

Implicações para o projeto de pesquisa

Uma das principais questões a considerar ao fazer testes em busca de diferenças ou relações é se convém optar por um teste paramétrico ou não paramétrico. Esta decisão será baseada no tamanho da amostra, nas técnicas aplicadas para a coleta de dados, no modo de medir os dados e no sistema de codificação eventualmente utilizado.

Por exemplo, é possível perguntar a idade dos inquiridos de diversas maneiras que resultarão em diferentes tipos de números. Se a pergunta for formulada de maneira aberta, “Quantas horas de

sono você teve na noite passada?” poderá ensejar respostas com um número específico de horas, isto é, um número intervalar no qual seria possível usar técnicas paramétricas. Mas também se poderia fazer a pergunta de modo a ensejar uma escala ordinal que precisará de estatística não paramétrica. Por exemplo:

Quantas horas de sono você teve na noite passada?

- 0-2 -----
- 3-4 -----
- 5-6 -----
- 7-8 -----
- 9-10 -----
- 11 + -----

Os testes paramétricos são mais complexos e, na opinião de muitos, mais sensíveis que os não paramétricos (e, portanto, com maior probabilidade de detectarem diferenças entre grupos, mesmo que estes sejam relativamente pequenos). Entretanto, só é possível usar testes paramétricos com dados de nível intervalar e apenas se certos pressupostos se cumprirem. Os dados devem estar normalmente distribuídos (cf. LEWIN, cap. 25), o que pode ser verificado mediante procedimentos estatísticos como o teste de Kolmogorov-Smirnov (cf. FIELD, 2009: 144-148). Além disso, a variabilidade de valores tem de ser aproximadamente igual em todos

os grupos; isto é denominado *homogeneidade de variância* e se verifica de diferentes maneiras conforme o procedimento.

Os testes não paramétricos são mais sólidos, mas menos sensíveis. Às vezes são chamados de testes sem pressupostos. As técnicas baseiam-se na ordem e não em diferenças exatas, isto é, se contagens ou valores de variáveis são maiores ou menores que outros. Os testes não paramétricos são adequados a dados ordinais ou nominais. Também se pode recorrer a eles quando não se apresentam os parâmetros necessários para testes paramétricos (p. ex., quando a distribuição da amostra não é normal). Salkind (2007, cap. 16) dá uma visão geral das técnicas não paramétricas.

Previsão e causação – uma precaução

Um dos erros mais comuns em que os pesquisadores incorrem ao usar correlações é falar em causa e efeito. Se há uma associação entre duas variáveis, isto não necessariamente significa que uma cause a outra. Três condições têm de ser preenchidas para provar relações entre causa e efeito:

- Tem de haver uma correlação significativa entre as variáveis “causa” e “efeito”.
- A correlação deve ser “real”, sem ter a ver com algum outro fator (nós chamamos esta de “espúria”).
- A variável causa deve preceder à variável efeito em tempo real.

A correlação significativa não é evidência de causalidade. Por exemplo, há uma associação positiva entre estatura e peso, pois pessoas mais altas tendem a pesar mais do que as mais baixas e pessoas de menor peso tendem a ser mais baixas do que aquelas que pesam mais. Não podemos dizer que as pessoas altas pesam mais porque comem mais, uma vez que é igualmente possível que elas pesem mais porque crescem mais e necessitam comer mais. Associação não implica causação.

Também é preciso verificar que a correlação significativa entre as variáveis seja “real”. Freedman e colegas (2007) dão o exemplo da busca da causa da pólio nos anos de 1950, quando houve uma epidemia mundial. Os pesquisadores reuniram enormes quantidades de informação fornecida por

vítimas da pólio e seus familiares e descobriram duas correlações muito significativas:

- um aumento da incidência da pólio e um aumento do consumo de refrigerantes;
- um aumento da incidência da pólio e um aumento da temperatura.

Apenas uma destas relações é “real”. A outra é “espúria”. O vírus da pólio dissemina-se quando a temperatura aumenta e esta é uma relação real. No entanto, quando a temperatura sobe a gente tende a beber mais líquido, daí que a ligação entre pólio e refrigerantes é espúria porque tem a ver com a temperatura e não existe por si mesma.

A condição final que se deve cumprir para examinar previsão e causalidade é que as variáveis de causa devem preceder às variáveis de efeito. Por exemplo, a intervenção para melhorar o ensino de leitura deve ocorrer antes que se efetue o teste para medir a habilidade de leitura. Nas ciências sociais, porém, às vezes é mais difícil deslindar a causa e o efeito. Pensemos na relação entre satisfação no emprego e produtividade. O que vem primeiro?

Referências anotadas

COOLIDGE, F.L. (2006). *Statistics: A gentle Introduction*. 2. ed. Londres: Sage. [O livro oferece uma boa introdução à estatística e seus princípios básicos. Os exemplos e o texto de Coolidge facilitam a abordagem de alguns destes conceitos complexos. Sua descrição das peculiaridades do teste t e da Anova é muito útil. Seus cálculos passo a passo podem ajudar a entender o que esses testes fazem realmente. Ele usa um estilo de notação que muitos leitores acham acessível.]

FIELD, A. (2009). *Discovering Statistics using SPSS*. 3. ed. Londres: Sage. [A terceira edição é igualmente amena e até mais acessível que as anteriores. Há capítulos destinados ao pesquisador iniciante que faz seu primeiro curso de estatística, bem como muitos outros para pesquisadores mais experientes que preparam projetos complexos. O livro dá muitas dicas úteis para usar o pacote SPSS/PASW. Sumamente recomendável.]

FREEDMAN, D.; PISANI, R. & PURVES, R. (2007). *Statistics*. 4. ed. Nova York: Norton. [Continua a ser um livro excelente focado em testes de probabilidade e sig-

nificância e no uso do valor quadrático médio (RMS) no desenvolvimento de avanços estatísticos.]

HOWELL, D.C. (2002). *Statistical Methods for Psychology*. 5. ed. Belmont, CA: Duxbury Press. [Um texto muito abrangente com cuidadosas explicações conceituais e matemáticas acompanhadas de exemplos úteis. Conduz o leitor pelos princípios subjacentes das técnicas estatísticas até métodos avançados como a regressão múltipla.]

MENDENHALL, W. (1987). *Introduction to Probability and Statistics*. Boston: Duxbury Press, cap. 1: "What is statistics?" [Mendenhall apresenta seus conceitos fundamentais valendo-se de um estudo de caso. Desenvolve a ideia e o uso de estatísticas inferenciais nas seções 1.2 e 1.4 e depois os aplica a seu estudo de caso em 1.5. Uma técnica interessante e útil.]

PALLANT, J. (2007). *SPSS Survival Manual*. 3. ed. Buckingham: Open University Press. [Escrito em estilo ameno e detalhado, com uma boa explicação de como interpretar e apresentar os resultados dos testes estatísticos.]

cos. Abrange quase tudo. Uma das melhores coisas deste livro é a tabela que resume os testes estatísticos, mostrando onde e como podem ser aplicados (116-117).]

SALKIND, N.J. (2007). *Statistics for People Who (Think They) Hate Statistics*. 3. ed. Londres: Sage. [Para os principiantes, este livro é o lugar para começarem. É ameno, fácil de entender e cheio de informação útil. Os capítulos estão em bom inglês claro, com procedimentos fáceis passo a passo. Começa supondo que o leitor não sabe e talvez não queira saber o que precisa saber! Os três capítulos sobre o teste T e Anova são um bom avanço. Sumamente recomendável.]

SIEGE, S. & CASTELLON, J. (1988). *Non-Parametric Statistics for the Social Sciences*. 2. ed. Nova York: McGraw Hill. [Siegel escreveu o guia definitivo para a estatística não paramétrica na década de 1950. Essa obra é um clássico. O texto atual é uma atualização do original e, embora escrito em estilo estatístico, ele descreve e explica a maioria dos testes que porventura se possam aplicar.]

UMA INTRODUÇÃO À MODELAGEM ESTATÍSTICA

Kelvyn Jones, Escola de Ciências Geográficas da Universidade de Bristol, Reino Unido.

Resumo

- Modelagem de regressão.
 - Pesquisa de relações de "causa e efeito" que não são necessárias nem suficientes.
- Modelagem em múltiplos níveis.
 - Pesquisa de um problema específico e como ele se relaciona com diferentes formas de estruturas de múltiplos níveis.
- Principais teóricos e autores:
 - Paul D. Allison
 - Harvey Goldstein
 - Stephen Raudenbush

Modelagem estatística é um tema muito vasto. No espaço de que disponho, eu me concentrarei na razão pela qual fazemos a modelagem e no que se pode conseguir. Examinarei que tipo de perguntas ela pode responder, que tipos de dados parecem ser um problema de "regressão" e que passos dar para garantir que os resultados obtidos sejam válidos. Escrevi esta introdução com a perspectiva avançada do modelo linear generalizado (McCULLAGH & NELDER, 1989) e incluí uma discussão substan-

cial sobre o enfoque da modelagem em múltiplos níveis, hoje em desenvolvimento, em razão de sua grande potencialidade para a análise de questões de pesquisa social.

Conceitos fundamentais: modelagem de regressão

Nas ciências sociais, nós pesquisamos relações de "causa e efeito" que não são necessárias (o resultado só ocorre se o fator causal tiver agido) nem suficientes (a ação de um fator sempre produz o resultado). Além do mais, a variação ou "ruído" inerente pode fazer o "sinal" sumir e nós precisamos de técnicas quantitativas para revelar os padrões subjacentes para obter provas críveis de uma relação (cf. JONES, cap. 23). Achamos um bom exemplo na epidemiologia. Há vítimas do câncer de pulmão que nunca fumaram, bem como pessoas que fumaram a vida inteira sem jamais adoecer. Já houve dúvidas quanto à ligação, mas hoje temos evidências inequívocas. Os homens que fumam têm risco 22 vezes maior de morrer de câncer de