Sobre o uso do teste χ^2 em amostras pequenas. Teste exacto de Fisher

FERNANDO GALVÃO DE MELO

Uma tabela de contingência 2 × 2 pode ser obtida de diferentes maneiras. Os casos mais usuais são a dupla dicotomia [*a*)], a comparação de populações [*b*)] e a comparação de tratamentos [*c*)]. Na prática, nem sempre é óbvio como classificar um conjunto de dados num destes três casos. No entanto, este facto é irrelevante, na medida em que o método de análise é o mesmo em qualquer das situações. No caso *a*) é habitual falar-se num *teste de independência;* nos casos *b*) e *c*), num *teste de homogeneidade* (ou de equivalência de tratamentos). No entanto, o problema pode sempre pôr-se como um teste de independência dos atributos de linha e coluna. Nas situações *a*) e *b*) parece mais indicado o uso de um teste bilateral. Em *c*) o efeito de tratamento, se existir, é unidireccional, apontando para um teste unilateral.

□

Fernando Galvão de Melo é professor catedrático de Bioestatística e presidente do conselho científico da Escola Nacional de Saúde Pública.

a) *Dupla dicotomia.* — Os indivíduos de uma amostra aleatória da população em estudo são classificados de acordo com dois critérios dicotómicos, *A* e *B*. A dimensão *n* da amostra é fixada *a priori*, sendo os quatro totais marginais conhecidos após a classificação dos indivíduos.

b) *Comparação de populações (comparação de atributos).* — Observam-se duas amostras independentes, uma de cada população, calculando-se as proporções de um atributo *A*. Neste caso, utiliza-se previamente um dos critérios de classificação, o que permite separar as populações a comparar (homens e mulheres, indivíduos expostos e não expostos, vacinados e não vacinados, etc.). As dimensões n_1 e n_2 das duas amostras são fixadas *a priori*.

Este método de amostragem constitui a base dos estudos prospectivos e dos estudos caso-controlo.

c) Comparação de tratamentos. — Um grupo de n indivíduos é subdividido por aleatorização em dois subgrupos, de dimensões n_1 e n_2 $(n_1 + n_2 = n)$, aplicando-se o tratamento B_i ao grupo i, $i = 1, 2$, observando-se uma variável resposta A dicotómica ou dicotomizada.

1. O teste χ^2

Do ponto de vista formal, uma tabela 2×2 pode escrever-se com o aspecto da tabela que se segue (*Tabela 1*):

	B	$\bar{B}$	
A	a	b	m_1
$\bar{A}$	c	d	m_2
	n_1	n_2	n

em que $m_1 = a + b$, $m_2 = c + d$, $n_1 = a + c$, $n_2 = b + d$ e $n = a + b + c + d$.

Por conveniência de notação, representa-se a configuração

$$\begin{array}{cc|} a & b \\ c & d \\ \hline \end{array}$$

por (a, b, c, d).

A *Tabela 1* representa a repartição de n indivíduos pelas quatro células, determinadas por dois critérios de classificação dicotómicos, ou, em alternativa, representa as distribuições do atributo A em duas amostras independentes, de dimensões n_1 e n_2.

Neste caso, o critério B foi previamente usado para definir as subpopulações correspondentes às categorias B e $\bar{B}$.

O problema em estudo sugere, em geral, qual o método mais conveniente para a obtenção dos dados. Independentemente da técnica de amostragem utilizada, pretende-se, em última análise, verificar a existência de uma eventual relação entre as variáveis qualitativas A e B.

A estatística habitual para o teste da hipótese nula, H_0, de independência (ou de homogeneidade) é

$$X^2 = \frac{n\,(ad - bc)^2}{(a + c)\,(b + d)\,(a + b)\,(c + d)} = \frac{n\,(ad - bc)^2}{n_1\, n_2\, m_1\, m_2}$$

sendo a distribuição teórica de referência a distribuição χ^2 com 1 grau de liberdade (gl).

Este teste usa uma distribuição contínua de probabilidade (distribuição χ^2) como aproximação à distribuição de probabilidade discreta (distribuição multinomial) das frequências observadas.

É, portanto, um teste aproximado, cuja validade assintótica aumenta com n.

Para melhorar a aproximação Yates (1934) sugeriu a introdução na expressão de X^2 de uma *correcção de continuidade.*

A estatística X^2 corrigida é

$$X_c^2 = \frac{n\left(|ad - bc| - \frac{n}{2}\right)^2}{n_1\, n_2\, m_1\, m_2}$$

igualmente referida à distribuição χ^2 com 1 gl.

Note-se que $X_c^2 < X^2$ e para valores grandes de n X^2 aproxima-se de X_c^2.

O uso de X_c^2 tem gerado alguma controvérsia. Grizzle (1967) e Conover (1968, 1974) argumentam contra o uso da correcção de Yates, salientando que o teste baseado na estatística X_c^2 é demasiado conservativo, tornando o actual nível de significância inferior ao valor nominal fixado. Como uma redução do nível de significância arrasta uma redução da potência do teste, este torna-se menos sensível. Por outras palavras, há um enfraquecimento da capacidade do teste para detectar uma efectiva existência de associação (ou de diferença de proporções).

Por outro lado, a variável aleatória induzida por X^2 parece estar mais de acordo com uma variável χ^2, na medida em que a sua distribuição se aproxima mais de uma distribuição χ^2 do que a distribuição de amostragem de X_c^2.

Mantel e Greenhouse (1968) e Fleiss (1981) contestam esta posição, fazendo notar que, quer no caso de independência, quer no caso de homogeneidade, as distribuições teóricas (populacionais) contêm parâmetros perturbadores, que devem ser estimados a partir das frequências marginais observadas.

Este facto implica uma análise dos dados condicionada pela hipótese de trabalho de que os totais marginais da tabela 2×2 observada são fixos. Nestas condições, as probabilidades exactas das configurações (a, b, c, d) são dadas pela distribuição hipergeométrica (*Apêndice*). Como as probabilidades associadas com X_c^2 são mais concordantes com as probabilidades exactas do que as probabilidades associadas com X^2, concluem pela conveniência do uso da correcção de Yates. Mais precisamente, verifica-se que a estatística X_c^2 permite, em geral, obter valores p (a partir da distribuição χ^2) que aproximam melhor os valores p obtidos com a distribuição hipergeométrica.

Indicações sobre uma discussão mais aprofundada do uso da correcção de Yates podem ser encontradas em Conover (1980) e Fleiss (1981).

O teste assintótico χ^2 é válido desde que se verifiquem determinadas condições empíricas de aplicabilidade, referindo-se desde já que o teste apresenta uma certa robustez. De facto, mesmo com uma deficiente verificação destas condições, obtêm-se conclusões concordantes com as que decorrem do uso da distribuição nula exacta.

Historicamente, a condição de aplicabilidade mais frequentemente proposta requer que o valor esperado (sob H_0) em cada célula seja, pelo menos, igual a 5. Investigações recentes tendem a considerar esta condição demasiado exigente.

Larntz (1978) e Koehler e Larntz (1980) mostraram que a aproximação χ^2 é legítima mesmo para frequências esperadas muito pequenas.

A adopção do critério de Cochran (1952, 1954) tem sido recomendada pela generalidade dos autores:

1) Com $n > 40$, use-se a estatística X_c^2 ou a estatística X^2 se $n >> 40$;
2) Com $20 \leq n \leq 40$, use-se X_c^2 desde que nenhuma das frequências esperadas seja inferior a 5. Caso contrário, use-se o teste exacto de Fisher;
3) Com $n < 20$, use-se o teste de Fisher.

No entanto, mesmo para $n < 20$ há, em geral, convergência nas decisões tomadas usando o teste χ^2 e o teste de Fisher (*Quadro II*).

Contemplando o facto de o teste χ^2, baseado na estatística X_c^2, ser conservativo, Schwartz (1969) sugere, para valores pequenos de n, o seguinte procedimento:

- Fixado o nível de significância α, se $X^2 < \chi^2_{1-\alpha}$ (quantil de ordem $1-\alpha$ da distribuição χ^2 com 1 gl), aceite-se H_0 (hipótese nula de independência ou de homogeneidade);
- Com $X^2 \geq \chi^2_{1-\alpha}$, calcule-se o valor corrigido X_c^2;
- Se $|X_c^2 - \chi^2_{1-\alpha}| >> 0$, aceite-se ou rejeite-se H_0, conforme for $X_c^2 < \chi^2_{1-\alpha}$ ou $X_c^2 > \chi^2_{1-\alpha}$;
- Com $X_c^2 \approx \chi^2_{1-\alpha}$, use-se o método exacto de Fisher.

2. O teste exacto de Fisher

O teste exacto de Fisher é baseado na avaliação das probabilidades condicionais das diferentes configurações de uma tabela 2×2, admitindo fixos os totais marginais. Retome-se a *Tabela 1*, representando a repartição de n indivíduos de acordo com dois critérios de classificação (ou as distribuições de duas amostras independentes relativamente a uma variável qualitativa).

Sob a hipótese nula, H_0, de independência (homogeneidade), a probabilidade da configuração (a, b, c, d) é dada pela distribuição hipergeométrica:

$$P(a, b, c, d) = \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{n}{a+b}} = \frac{\binom{n_1}{a}\binom{n_2}{b}}{\binom{n}{m_1}} = \frac{n_1!n_2!m_1!m_2!}{n!a!b!c!d!}$$

Como, fixados os totais marginais, a configuração (a, b, c, d) é completamente determinada por a, escreve-se, mais simplesmente, $P(a, b, c, d) = P(a)$.

Observada uma tabela, há que identificar as configurações tão ou mais extremas. Dada a configuração (a, b, c, d), o valor esperado para a, sob H_0, é

$$\bar{a} = \frac{(a+c)\ (a+b)}{n} = \frac{n_1 m_1}{n}$$

Configurações que suportem com maior evidência a hipótese alternativa são as configurações tão ou mais extremas que a observada. Para um teste unilateral estas configurações obtêm-se substituindo a, sucessivamente, pelos valores (inteiros) x tais que $x \leq a$, se $a < \bar{a}$, ou $x \geq a$, se $a > \bar{a}$, isto é, $x = a, a-1, \ldots,$ máx. $(0, n_1 + m_1 - n)$, no primeiro caso, e $x = a, a+1, \ldots,$ mín. (n_1, m_1), no segundo.

[Escrevendo a configuração (a, b, c, d) na forma $(a, m_1 - a, n_1 - a, n - n_1 - m_1 + a)$, reconhece-se imediatamente que máx. $(0, n_1 + m_1 - n) \leq a \leq$ mín. (n_1, m_1).]

Note-se ainda que, com $x = a - j$ $(j = \pm 1, \pm 2, \ldots)$ a configuração correspondente é $(a-j, b+j, c+j, d-j)$.

Para um teste bilateral há que considerar as configurações mais extremas nos dois sentidos. Estas configurações são as que correspondem aos inteiros x, satisfazendo a

$$|x - \bar{a}| \geq \begin{cases} \bar{a} - a, & \text{se } a < \bar{a} \\ a - \bar{a}, & \text{se } a > \bar{a} \end{cases}$$

isto é, $x \leq a$ ou $x \geq 2\bar{a} - a$, se $a < \bar{a}$, e $x \geq a$ ou $x \leq 2\bar{a} - a$, se $a > \bar{a}$, correspondendo $x = a$ à configuração observada.

Então para um teste unilateral a probabilidade de significância p^* da configuração observada (a, b, c, d) ou do valor a obtém-se somando as probabilidades de todas as configurações com $x \leq a$ $(x \geq a)$. Para um teste bilateral a probabilidade de significância p^{**} obtém-se somando as probabilidades das configurações com $x \leq a$ ou $x \geq 2\bar{a} - a$ $(x \geq a$ ou $x \leq 2\bar{a} - a)$. Note-se que, no caso de tabelas com totais de linha ou coluna iguais ($n_1 = n_2$ ou $m_1 = m_2$ — distribuição hipergeométrica simétrica), a probabilidade de significância para um teste bilateral é o dobro da probabilidade de significância do teste unilateral: $p^{**} = 2p^{*}$. Na prática, e por economia de cálculo, mesmo no caso não simétrico, um valor aproximado para p^{**} ainda é dado por $2p^{*}$.

O cálculo das probabilidades das configurações mais extremas pode ser simplificado usando fórmulas de recorrência (Feldman e Klinger, 1963). Considere-se a sucessão de configurações com $x = a$, $a - 1 = a_1$, ..., $a - j = a_j$, ... e sejam $P(a)$, $P(a_1)$, ..., $P(a_j)$, ... as respectivas probabilidades.

Como se verifica facilmente, calculada a probabilidade $P(a)$, obtém-se, sucessivamente

$$P(a_1) = \frac{ad}{b_1 c_1} P(a), \; P(a_2) = \frac{a_1 d_1}{b_2 c_2} P(a_1), \ldots,$$

$$P(a_j) = \frac{a_{j-1} d_{j-1}}{b_j c_j} P(a_{j-1})$$

[Note-se que $a_{j-1} = a - (j-1)$, $d_{j-1} = d - (j-1)$, $b_j = b + j$, $c_j = c + j$.]

De um modo geral, as probabilidades das configurações (a, b, c, d) e (a_j, b_j, c_j, d_j) estão ligadas pela relação

$$P(a_j) = \frac{a \; d \; a_1 \; d_1 \ldots a_{j-1} \; d_{j-1}}{b_1 c_1 b_2 c_2 \ldots b_j c_j} P(a)$$

com $j = 1, 2, \ldots$

Exemplo 1. — Considere-se a configuração (1, 7, 5, 2) de probabilidade $P(1) = \frac{6!9!8!7!}{15!7!5!2!} = 0{,}336$, com $a = 1$ e $\bar{a} = 3$, 2. Configurações mais extremas, em ambos os sentidos, são (0, 8, 6, 1) e (6, 2, 0, 7) de probabilidades $P(0) = \frac{1 \times 2}{8 \times 6} P(1) = 0{,}0014$ e $P(6) = \frac{6!9!8!7!}{15!6!2!7!} = 0{,}0056$.

A probabilidade de significância (valor p) da configuração (1, 7, 5, 2) é $p^{*} = P(1) + P(0) = 0{,}035$ para um teste unilateral e $p^{**} = p^{*} + P(6) = 0{,}041$ para um teste bilateral.

A fórmula de recorrência de Feldman e Klinger permite obter facilmente a distribuição nula exacta (distribuição sob H_0) da população de configurações. As sete configurações possíveis, ordenadas pelos valores decrescentes de a, são (6, 2, 0, 7), (5, 3, 1, 6), (4, 4, 2, 5), (3, 5, 3, 4), (2, 6, 4, 3), (1, 7, 5, 2), (0, 8, 6, 1), de probabilidades $P(6) = 0{,}0056$, $P(5) = \frac{6 \times 7}{3 \times 1} P(6) = 0{,}0784$, $P(4) = \frac{5 \times 6}{4 \times 2} \times P(5) = 0{,}2940$, $P(3) = \frac{4 \times 5}{5 \times 3} P(4) = 0{,}3920$, $P(2) = \frac{3 \times 4}{6 \times 4} P(3) = 0{,}1960$, $P(1) = \frac{2 \times 3}{7 \times 5} P(2) = 0{,}0336$, $P(0) = \frac{1 \times 2}{8 \times 6} P(1) = 0{,}0014$ ($\sum_{j=0}^{6} P(j) = 1$).

Exemplo 2. — Para comparar a eficácia de dois tratamentos, T_1 e T_2, na cura de uma doença D, um grupo de 16 indivíduos com o mesmo estádio de evolução da doença foi subdividido, por aleatorização, em dois grupos de igual dimensão, aos quais se aplicou T_1 e T_2, respectivamente. Os resultados obtidos dispõem-se na tabela

		Tratamentos		
		T_1	T_2	
Resposta	C	5	1	6
	$\bar{C}$	3	7	10
		8	8	16

sendo C = curado e $\bar{C}$ = não curado.

H_0: T_1 e T_2 são equivalentes *vs.* H_1 : T_1 é superior a T_2.

Uma configuração mais extrema e no mesmo sentido que a observada (traduzindo uma maior evidência na rejeição de H_0 em favor da alternativa H_1) é (6, 0, 2, 8). Sob H_0 é $P(6, 0, 2, 8) = P(6) = \frac{8!8!6!10!}{16!6!2!8!} = 0{,}0035$ e, por recorrência, $P(5, 1, 3, 7) = P(5) = \frac{6 \times 8}{1 \times 3} P(6) = 0{,}056$.

Então, sob H_0, a probabilidade de obter uma configuração tão ou mais extrema que a observada é $p^{*} = 0{,}0035 + 0{,}056 = 0{,}0595$, probabilidade de significância (valor p) da tabela observada.

Para um teste bilateral (H_0 *vs.* H_1 : T_1 diferente de T_2), como a distribuição é simétrica, a probabilidade de significância é $p^{**} = 2p^{*} = 0{,}119$, probabilidade de uma configuração tão ou mais extrema que a observada, nos dois sentidos. Fixado o nível de significância $\alpha = 0{,}05$, a hipótese nula não é rejeitada em nenhum dos casos, não se podendo concluir que T_1 seja superior a T_2 (ou T_1 diferente de T_2). Note-se que a estatística X_c^2 conduz à mesma conclusão. De facto, $X_c^2 = 2{,}4$, valor não significativo ao nível $\alpha = 0{,}05$.

Quando $n > 20$ e o menor dos valores esperados não é inferior a 5, os testes de Fisher e χ^2 conduzem, em geral, às mesmas conclusões. Considere-se a família de configurações (a, b, c, d) com $a + b = 10$, $a + c = 12$ e $n = 24$. Fazendo a variar de 0 a 10, obtém-se a seguinte distribuição exacta:

$P(0) = P(10) = 0{,}000034$ $P(3) = P(7) = 0{,}08976$

$P(1) = P(9) = 0{,}00136$ $P(4) = P(6) = 0{,}23562$

$P(2) = P(8) = 0{,}01683$ $P(5) = 0{,}323136$

$$\sum_{j=0}^{10} P(j) = 1$$

Neste caso, a configuração esperada sob H_0 é $(\bar{a}, \bar{b}, \bar{c}, \bar{d}) = (5, 5, 7, 7)$.

O *Quadro I* permite comparar as decisões associadas ao teste de Fisher e ao teste χ^2.

Quando $p^{**} < 0{,}05$ é $X^2 > 3{,}84$ e quando $p^{**} > 0{,}05$ é $X^2 < 3{,}84$, implicando a igualdade de decisões. Este resultado mantém-se substituindo X^2 por X_c^2.

Quando $n < 20$, os dois testes (Fisher e χ^2) podem conduzir, nalguns casos, a decisões diferentes. Retome-se o exemplo 2 e construa-se o *Quadro II*, semelhante ao *Quadro I*, com $a = 6, 5, 4, 3$.

Note-se que, neste caso, $n < 20$ e dois dos valores esperados são inferiores a 5. Como se verifica, à configuração (5, 1, 3, 7) correspondem decisões distintas. No entanto, usando a estatística X_c^2, já há coincidência nas decisões.

De facto, os valores de X_c^2 para as quatro configurações do *Quadro II* são, respectivamente, 6,67, 2,40, 0,27, 0,27, e apenas o valor 6,67 leva à rejeição da hipótese nula.

Como foi referido, a estatística X_c^2 permite obter valores p próximos dos obtidos com as probabilidades hipergeométricas.

Por exemplo, para a configuração observada (5, 1, 3, 7) tem-se $p^{**} = 0{,}119 \approx P(X^2 \geq 2{,}40) = 0{,}121$.

Registe-se, no entanto, que há situações em que a estatística concordante com a decisão baseada no valor da probabilidade p^{**} é a estatística X^2. Por exemplo, para a configuração (1, 7, 5, 2), com $n = 15$ e todos os valores esperados entre 1 e 5 ($\bar{a} = 3{,}2$, $\bar{b} = 4{,}8$, $\bar{c} = 2{,}8$, $\bar{d} = 4{,}2$), tem-se $p^{**} = 0{,}041$, $X^2 = 5{,}40$, $X_c^2 = 3{,}23$ e, fixado o nível de significância $\alpha = 0{,}05$, p^{**} e X^2 levam à rejeição de H_0 e X_c^2 à aceitação de H_0. Para as restantes seis configurações há coincidência nas decisões.

Estes resultados permitem concluir que mesmo com $n < 20$ e $1 < \bar{a}, \bar{b}, \bar{c}, \bar{d} < 5$ há em geral, concordância nas decisões associadas ao teste de Fisher e ao teste χ^2, não parecendo haver uma vantagem nítida no uso da estatística X_c^2.

Em resumo, sugere-se a seguinte regra prática:

a) Com $n < 20$, use-se o teste exacto de Fisher;

Quadro I

Configuração	p^{**}	Decisão ($\alpha = 0{,}05$)	X^2	Decisão ($\alpha = 0{,}05$)
(10, 0, 2, 12)	0,000 07	Rej. H_0	17,14	Rej. H_0
(9, 1, 3, 11)	0,002 79	Rej. H_0	10,97	Rej. H_0
(8, 2, 4, 10)	0,036 45	Rej. H_0	6,17	Rej. H_0
(7, 3, 5, 9)	0,215 97	Ac. H_0	2,74	Ac. H_0
(6, 4, 6, 8)	0,687 21	Ac. H_0	0,69	Ac. H_0
(5, 5, 7, 7)	1	Ac. H_0	0	Ac. H_0

Quadro II

Configuração	p^{**}	Decisão ($\alpha = 0{,}05$)	X^2	Decisão ($\alpha = 0{,}05$)
(6, 0, 2, 8)	0,0070	Rej. H_0	9,60	Rej. H_0
(5, 1, 3, 7)	0,1188	Ac. H_0	4,27	Rej. H_0
(4, 2, 4, 6)	0,6084	Ac. H_0	1,07	Ac. H_0
(3, 3, 5, 5)	1	Ac. H_0	0	Ac. H_0

b) Com $20 \leq n < 40$ e nenhuma frequência esperada inferior a 1, use-se a estatística X^2. Se algum dos valores esperados for inferior à unidade, use-se a estatística X_c^2. Por exemplo, para as configurações (a, b, c, d), com $n = 20$, $a + b = 3$ e $a + c = 4$ (um valor esperado inferior à unidade e dois inferiores a 5), há concordância entre o teste de Fisher e o teste χ^2 baseado na estatística X_c^2;

c) Com $n \geq 40$, use-se a estatística X^2.

Em particular, nos casos *b*) e *c*), se, face à situação concreta em estudo, o investigador concluir pela necessidade de um teste mais conservativo, deverá optar pela estatística X_c^2.

Apêndice

Hipóteses compostas e testes condicionais

Uma hipótese que não permite especificar completamente a distribuição nula diz-se uma hipótese composta (de contrário, a hipótese diz-se simples). Por exemplo, na comparação de duas distribuições binomiais, a hipótese $H_0 : p_1 = p_2$ é composta.

O método clássico de rodear as dificuldades resultantes de uma hipótese nula composta consiste em trabalhar com uma versão condicionada conveniente da distribuição nula.

O teste exacto de Fisher é um exemplo de teste condicional cuja região crítica é definida dado o valor de uma determinada estatística T. Não contendo a estatística T informação sobre a veracidade ou falsidade da hipótese a testar, o teste de Fisher tem propriedades *óptimas*.

Numa tabela 2×2 há apenas uma pequena quantidade de informação nos totais marginais, justificando o uso de distribuições condicionais, que tornam relativamente fácil o cálculo da distribuição nula exacta.

Mais precisamente, Tocher (1950) mostrou que o teste exacto, baseado em totais marginais fixos, dá igualmente testes *óptimos* (no sentido de testes centrados uniformemente mais potentes) para os casos em que nenhuma ou apenas uma margem é fixada *a priori*. Em resumo, a situação em que ambas as margens são fixadas *a priori* proporciona testes *óptimos*, quer para o caso da comparação de proporções (homogeneidade), quer para o caso da dupla dicotomia (independência) (Kendall e Stuart, 1973).

Seja (A, B) um par categorial, definido sobre um universo U. Obtida uma amostra aleatória de dimensão n, designe-se por f_{ij} o número de ocorrências do par (A_i, B_j). Com A e B dicotómicas ($i, j = 1, 2$), escreve-se, em geral, $A_1 = A$, $A_2 = \tilde{A}$, $B_1 = B$, $B_2 = \tilde{B}$, $f_{11} = a$, $f_{12} = b$, $f_{21} = c$, $f_{22} = d$. No caso 2×2, quando a amostragem é feita a partir de uma população bivariada, a hipótese de independência é $H_0 :\ P(AB) = P(A)\ P(B)$.

Quando amostras independentes são obtidas das populações univariadas (condicionais) B e $\tilde{B}$, a hipótese de homogeneidade traduz-se por $H_0 : P(A/B = P(A/\tilde{B})$.

Estes, aparentemente, distintos modelos pretendem ambos descrever uma situação de *não existência de relação*, sendo, essencialmente, equivalentes. Esta equivalência é justificada pelo resultado, de demonstração imediata.

Teorema 1. — As variáveis A e B são independentes se e só se as variáveis A/B e $A/\tilde{B}$ são semelhantes.

Teorema 2. — A distribuição nula condicionada pela fixação das margens de uma tabela 2×2 é a distribuição hipergeométrica.

a) *Teste de homogeneidade* (teste de igualdade de proporções; uma margem fixada). — A tabela

	B	$\tilde{B}$
A	a	b
$\tilde{A}$	c	d
	n_1	n_2

representa a distribuição de duas amostras independentes de dimensões n_1 e n_2, respectivamente (valores fixados *a priori*).

Seja, na população B, $P(A) = p_1$ e, na população $\tilde{B}$, $P(A) = p_2$. A tabela das probabilidades é

	B	$\tilde{B}$
A	p_1	p_2
$\tilde{A}$	q_1	q_2
	1	1

com $q_1 = 1 - p_1$ e $q_2 = 1 - p_2$.

Admitindo que as dimensões populacionais são suficientemente grandes, de modo que a distribuição binomial seja um bom ajustamento, tem-se

$$P(a;\ n_1,\ p_1) = \binom{n_1}{a} p_1^a q_1^{n_1 - a}$$

$$P(b;\ n_2,\ p_2) = \binom{n_2}{b} p_2^b q_2^{n_2 - b}$$

Sob $H_0 : p_1 = p_2 = p$ (p desconhecido), a probabilidade da configuração (a, b, c, d) é dada por

$$P_0(a, b) = \binom{n_1}{a} \binom{n_2}{b} p^{a+b} (1 - p)^{n - (a+b)}$$

expressão contendo o parâmetro perturbador p. Sob H_0 $a + b = m_1$ é uma estatística suficiente para p (m_1/n é uma estimativa de p) e, portanto, a distribuição condicional de (a, b), dado $a + b = m_1$, não depende de p. De facto:

$$P(m_1) = \binom{n}{m_1} p^{m_1} (1-p)^{n-m_1}$$

$$g(a, b/m_1) = \frac{P_0(a, b)}{P(m_1)} \frac{\binom{n_1}{a}\binom{n_2}{b} p^{a+b} (1-p)^{n-(a+b)}}{\binom{n}{m_1} p^{m_1}(1-p)^{n-m_1}} =$$

$$= \frac{\binom{n_1}{a}\binom{n_2}{b}}{\binom{n}{m_1}}$$

com $a + b = m_1$.

Como $b = m_1 - a$, obtém-se a distribuição univariada

$$h(a/m_1) = \frac{\binom{n_1}{a}\binom{n_2}{m_1 - a}}{\binom{n}{m_1}}$$

com $n = n_1 + n_2$, máx. $(0,\ n_1 + m_1 - n) \leq a \leq$ mín. (n_1, m_1), distribuição livre do parâmetro perturbador p. A expressão de $h\ (a/m_1)$ mostra que a variável aleatória induzida por a tem uma distribuição hipergeométrica de parâmetros $(n, m_1, \frac{n_1}{n})$ *c. q. d.*

b) Teste de independência (dupla dicotomia; nenhuma margem fixada). — A tabela

	B	$\bar{B}$
A	a	b
$\bar{A}$	c	d

com $a + b + c + d = n$, representa um conjunto de n observações do par (A, B) de componentes dicotómicas (resultado da classificação de n indivíduos de acordo com uma dupla dicotomia). Neste caso, apenas n foi fixado *a priori*.

Seja $\alpha = P(A)$ e $\beta = P(B)$. A hipótese nula de independência, H_0, é traduzida formalmente por $P(AB) = P(A)\,P(B) = \alpha\beta$, sendo a distribuição bivariada, sob H_0, dada pela tabela

	B	$\bar{B}$	
A	$\alpha\beta$	$\alpha(1-\beta)$	α
$\bar{A}$	$(1-\alpha)\beta$	$(1-\alpha)(1-\beta)$	$1-\alpha$
	β	$1-\beta$	1

O modelo teórico que permite o cálculo da probabilidade da configuração (a, b, c, d) é a distribuição multinomial:

$$P_0(a, b, c, d) = \frac{n!}{a!b!c!d!} (\alpha\beta)^a [\alpha(1-\beta)]^b [(1-\alpha)\beta]^c$$

$$[(1-\alpha)(1-\beta)]^d = \frac{n!}{a!b!c!d!} \alpha^{a+b} (1-\alpha)^{c+d} \beta^{a+c}$$

$$(1-\beta)^{b+d}$$

com $a + b + c + d = n$, expressão que depende dos parâmetros perturbadores α e β.

Sob H_0, $a + b = m_1$ e $a + c = n_1$ são estatísticas conjuntamente suficientes para (α, β). Então a distribuição condicional de (a, b, c, d), dado $(a + b = m_1,\ a + c = n_1)$, é independente de α e β. De facto, com A e B independentes é

$$P(m_1, n_1) = \binom{n}{m_1}\alpha^{m_1}(1-\alpha)^{n-m_1} \binom{n}{n_1}\beta^{n_1}(1-\beta)^{n-n_1}$$

$$g(a, b, c, d/m_1, n_1) = \frac{P_0(a, b, c, d)}{P(m_1, n_1)} = \frac{\frac{n!}{a!b!c!d!}}{\binom{n}{m_1}\binom{n}{n_1}} = \frac{\binom{n_1}{a}\binom{n_2}{b}}{\binom{n}{m_1}}$$

com $a + b = m_1$, $a + c = n_1$.

Como $b = m_1 - a$, obtém-se novamente a distribuição (univariada) hipergeométrica

$$h(a/n_1, m_1) = \frac{\binom{n_1}{a}\binom{m_1}{m_1 - a}}{\binom{n}{m_1}}$$

com máx. $(0,\ n_1 + m_1 - n) \leq a \leq$ mín. $(n_1,\ m_1)$.

Então, em qualquer dos casos, fixados os totais marginais, a distribuição de probabilidade das configurações (a, b, c, d) é a distribuição hipergeométrica. Esta distribuição, livre de parâmetros desconhecidos, é a requerida versão condicionada da distribuição nula.

□ Referências bibliográficas

COCHRAN, W. G.
The χ^2 test of goodness of fit.
«Annals of Maths. Statistics», 23, 1952, pp. 31-345.

COCHRAN, W. G.
Some methods for strengthening the common χ^2 tests.
«Biometrics», 10, 1954, pp. 417-451.

CONOVER, W. J.
Uses and abuses of the continuity correction.
«Biometrics», 24, 1968, p. 1028.

CONOVER, W. J.
Some reasons for not using the Yates continuity correction on 2 × 2 contingency tables.
«J. Am. Stat. Assoc.», 69, 1974, pp. 374-376.

CONOVER, W. J.
Practical nonparametric statistics, New York: J. Wiley & Sons, 1980.

FELDMAN, S. E.; KLINGER, E.
Short-cut calculation of the Fisher-Yates exact test.
«Psychometrika», 28, 1963, pp. 289-291.

FLEISS, J. L.
Statistical methods for rates and proportions, New York: J. Wiley & Sons, 1981.

GRIZZLE, J. E.
Continuity correction in the χ^2 test for 2 × 2 tables.
«Am. Stat.», 21, 1967, pp. 28-32.

KENDALL, M. G.; STUART, A.
The advanced theory of statistics, 2, London: Charles Griffin, 1973.

KOEHLER, K.; LARNTZ, K.
An empirical investigation of goodness-of-fit statistics for sparse multinomials.
«J. Am. Stat. Assoc.», 75, 1980, pp. 336-344.

LARNTZ, K.
Small-sample comparison of exact levels for χ^2 goodness-fit statistics.
«J. Am. Stat. Assoc.», 73, 1978, pp. 253-263.

MANTEL, N.; GREENHOUSE, S. W.
What is the continuity correction?
«Am. Stat.», 22, 1968, pp. 27-30.

SCHWARTZ, D.
Méthodes statistiques à l'usage des médicins et des biologistes, Paris: Flammarion, 1969.

TOCHER, K. D.
Extension of the Neyman-Pearson theory of tests to discontinuous variates.
«Biometrika», 37, 1950, pp. 130-144.

YATES, F.
Contingency tables involving small numbers and the χ^2 test.
«J. Roy. Stat. Soc.», 1, 1934, pp. 217-235.

□ Résumé

UTILISATION DU TEST χ^2 POUR DE PETITS ÉCHANTILLONS. LE TEST EXACT DE FISHER.

Cet article met en évidence la robustesse du test χ^2 et démontre que, même dans les cas de petits échantillons et de transgression des conditions classiques d'applicabilité, il y a un accord raisonnable entre le test asymptotique χ^2 et le test exact de Fisher. Ce dernier test est décrit en détail. L'utilisation de la correction de continuité d'Yates est mise en discussion et quelques règles sont suggérées en vue du choix de la statistique la plus convenable à chaque cas concret étudié.

□ Summary

ON THE USE OF THE χ^2 TEST WITH SMALL SAMPLES. FISHER'S EXACT TEST

This article emphasizes the robustness of χ^2 test and shows that, even with small samples and some transgression of classic applicability conditions, there is reasonable agreement between the asymptotic χ^2 test and Fisher's exact test. The latter is described in detail. The use of Yates' continuity correction is discussed and some practical rules are suggested for the choice of the appropriate statistic to be used in each specific case under consideration.