ADC: Aula 3 - Alguns tópicos especiais sobre tabelas bidimensionais e teste qui-quadrado:

1. Correção de continuidade de Yates para valores esperados pequenos

$$\chi^2_{Yates} = \sum_{i=1}^{l}\sum_{j=1}^{c}\frac{\left(|n_{ij}-e_{ij}|-0{,}5\right)^2}{e_{ij}}$$

2. Partição de tabelas *2 × c* em tabelas *2 × 2* não independentes

Exemplo: Comparação de diferentes tratamentos com placebo – H_0: $\Theta = \Theta_0$; H_1: $\Theta \neq \Theta_0$

TRATAMENTO DE DEPRESSÃO

	Placebo	Droga1	Droga2	Droga3	Droga4	Droga5	Total
Melhor	8	12	21	15	14	19	89
Igual/Pior	22	18	9	15	16	11	91
Total	30	30	30	30	30	30	180

Temos uma tabela bidimensional *2 × 6*, com 5 G.L., o que nos dá um valor para a estatística de teste $\chi^2_P = 14{,}78$, acima do valor crítico 11,071 da distribuição para α = 0,05.

Sub-tabelas obtidas:

8	12
22	18

$\chi^2 = 1{,}20$

8	21
22	9

$\chi^2 = 11{,}28$

8	15
22	15

$\chi^2 = 3{,}45$

8	14
22	16

$\chi^2 = 2{,}58$

8	19
22	11

$\chi^2 = 8{,}15$

Se realizarmos 5 testes repetidos, devemos atentar para o nível de significância real, uma vez que o valor inicial α = 0,05, que indica a $P(Rej.H_0|\Theta = \Theta_0)$, passa a ser α = 0,2262 para a sequência de 5 comparações não independentes[1].

Brunden[2] sugere uma correção para o nível de significância: $\alpha' = \frac{\alpha}{2\,(c-1)}$. No caso teremos o novo nível de significância $\alpha' = 0{,}005$ com o valor crítico 7,88 para 1 G.L. e podemos verificar onde existem diferenças estatisticamente significativas.

3. O caso de amostras pareadas (observações repetidas, observações correlacionadas): $\chi^2_{McNemar}$

Exemplo: Painel de 250 eleitores sem filiação partidária que indicam sua intenção de voto em plebiscito antes e após um debate público televisionado e também transmitido pela internet ao vivo.

Antes do debate	Após o debate: a favor	Após o debate: contra	Total
a favor	96	24	120
contra	45	85	130
Total	141	109	250

A estatística correta a ser utilizada, com 1 G.L. seria: $\chi^2_{McNemar} = \chi^2_P$, mas apenas para as células onde houve mudança de opinião (n*), sendo o valor esperado e_{ij} igual à média dos valores observados nestas células. No exemplo:

$$\chi^2_{McNemar} = \frac{(24-34{,}5)^2}{34{,}5} + \frac{(45-34{,}5)^2}{34{,}5} = 6{,}4 \; ou \; \frac{(|24-34{,}5|-0{,}5)^2}{34{,}5} + \frac{(|45-34{,}5|-0{,}5)^2}{34{,}5} = 5{,}8$$

com correção de continuidade[3], onde H_0 = as mudanças são iguais nas duas direções.

---

[1] Se H_0 é verdade, em 5 comparações seguidas temos $0{,}95^5 = 0{,}77378$ como probabilidade de não obter nenhum resultado significativo. Logo a probabilidade de pelo menos 1 resultado significativo será 1 – 0,77378 = 0,2262.

[2] R.D. Brunden 1972. The analysis of non-independent 2 X 2 tables using rank sums. *Biometrics,* **28**,603-607

[3] A estatística é calculada como o quadrado de $|Z| = n_{12} - \left(\frac{1}{2}\right) n^* / \sqrt{n^*\,(1/2)(1/2)} = n_{12} - n_{21} / \sqrt{n_{12} + n_{21}}$, para $n^* > 10$.

B.S. Everitt – The Analysis of Contingency Tables – Second Edition. Monographs on Statistics and Applied Probability 45. London: Chapman & Hall, 1992.

Tradução livre[4] do item 3.3, pp 39-40.

Frequências esperadas pequenas

A derivação da distribuição qui-quadrado como uma aproximação para a distribuição da estatística de teste, χ^2, quando a hipótese de independência é verdade, é feita sob o pressuposto de que os valores esperados não sejam muito pequenos. Esta expressão vaga tem sido utilizada tipicamente para interpretar uma aproximação satisfatória como sendo aquela onde as frequências esperadas são iguais ou superiores a cinco. Tal restrição parece, no entanto, ser arbitrária, baseada muito mais em tradição do que em evidência matemática ou empírica, e parece não haver uma justificativa maior para a regra de cinco-ou-mais do que para, por exemplo, uma regra de um-ou-mais.

Cochran (1954) foi o primeiro a chamar a atenção para o fato de que tal 'regra' usual é muito rigorosa, sugerindo que se relativamente poucos valores esperados são menores que cinco (digamos, uma célula em cada cinco), um valor esperado igual à unidade é permitido. Mesmo esta sugestão pode ser considerada muito restritiva, uma vez que os trabalhos de Lewontin e Felsenstein (1965), Slakter (1966), Roscoe e Byars (1971) e Larntz (1978) apontam que muitos dos valores esperados podem ser iguais à unidade sem que o teste seja muito afetado. Lewontin e Felsenstein apresentam a seguinte regra conservadora para tabelas com número de linhas igual a dois: 'A tabela 2 × *c* pode ser testada pelo critério qui-quadrado convencional se todos os valores esperados forem iguais ou maiores que 1.' Os autores argumentam que mesmo esta regra é extremamente conservadora e que na maioria dos casos o teste qui-quadrado pode ser utilizado para tabelas com valores esperados acima de 0,5 na menor célula.

No entanto, para dados assimétricos, esparsos ou pequenos a teoria assintótica pode não ser válida, embora seja frequentemente difícil prever *a priori* se um determinado conjunto de dados pode causar problemas. Um teste **exato** para a hipótese de que as classificações de linha e de coluna são independentes em uma tabela de contingência $\ell \times c$ pode ser realizado em princípio pela generalização do tratamento exato dado por Fisher a tabelas 2 × 2. O esforço computacional necessário, no entanto, limitou muito, até recentemente, tal enfoque. Na última década[5], o desenvolvimento de algoritmos rápidos e a disponibilidade de capacidade de processamento a um baixo custo permitiu um avanço considerável das fronteiras onde o teste exato é viável. Detalhes a repeito de tais algoritmos são, infelizmente, fora do escopo desse texto e os leitores interessados devem procurar as referências Mehta e Patel (1983, 1986a, 1986b) para uma descrição completa. No entanto, é possível ilustrar as diferenças que podem ocorrer entre o enfoque exato e o procedimento usual que envolve a distribuição qui-quadrado. Por exemplo, considere a seguinte tabela de contingência 3 × 9:

0	7	0	0	0	0	0	1	1
1	1	1	1	1	1	1	0	0
0	8	0	0	0	0	0	0	0

A estatística qui-quadrado para esta tabela apresenta o valor 22,9. O valor *p* correspondente é 0,1342. O valor verdadeiro de *p*, no entanto, é igual a 0,0013. A análise exata indica que as classificações de linha e de coluna **não** são independentes. A análise assintótica não consegue mostrar tal relação.

Um procedimento que vem sendo quase que rotineiramente utilizado por muitos anos para enfrentar o problema de frequências esperadas pequenas é a agregação de categorias. Tal procedimento pode ser criticado por diversas razões. Primeiramente, uma quantidade considerável de informação pode ser perdida pela combinação de categorias, e isto pode levar a um grande desvio do interesse e utilidade do estudo. Em segundo lugar, a aleatoriedade da amostra pode ser afetada. O princípio básico do teste qui-quadrado se apoia completamente na aleatoriedade da amostra e na escolha prévia das categorias onde as observações podem cair. Combinar categorias após os dados terem sido colhidos e vistos pode afetar a natureza aleatória da amostra com consequências

[4] Feita pelo Prof. Ronaldo R. Bastos apenas com fins didáticos, para utilização em sala de aula com a turma de Análise de Dados Categóricos (EST043)

[5] Lembrar que este livro é de 1992!

desconhecidas. Por último, a forma como as categorias são combinadas pode ter um efeito importante nas inferências realizadas. Como exemplo, considere os seguintes dados, apresentados por Baglivo *et al* (1988):

	Coluna				
	1	**2**	**3**	**4**	**5**
Linha 1	2	3	4	8	9
Linha 2	**0**	**0**	11	10	11

Ao se testar esta tabela para independência utilizando os métodos aproximados usuais a significância calculada é 0,086, que está de acordo com a probabilidade exata de Fisher até duas casas decimais, embora um pacote estatístico convencional geralmente apresente uma advertência nos moldes: 'Alguns dos valores esperados são menores que 2, o teste pode não ser apropriado.' No entanto, se as duas primeiras colunas forem ignoradas a significância muda para 0,48, e se as duas primeiras colunas forem combinadas com a terceira a significância é 1,00.

A prática de combinar categorias de classificação deve ser evitada e não é necessária atualmente, devido à existência dos testes exatos citados acima.

Referências Bibliográficas:

Baglivo, J., Oliver, D. e Pagano, M. 1988. Methods for the analysis of contingency tables with large and small cell counts. *J. Am. Statist. Assn*. **83**, 1006-1013.

Cochran, W.G. 1954. Some methods for strenghtening the common chi-square tests. *Biometrics* **10**, 417-451.

Larntz, K. 1978. Small-sample comparison of exact levels for chi-squared goodness-of-fit statistics. *J. Am. Statist. Assn*. **73**, 253-263.

Lewontin, R.C. e Felsenstein, J. 1965. The robustness of homogeneity tests in 2 × N tables. *Biometrics* **21**, 19-33.

Mehta, C.R. e Patel, N.R. 1983. A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *J. Am. Statist. Soc*. **78**, 427-434.

Mehta, C.R. e Patel, N.R. 1986a. A hybrid algorithm for Fisher's exact test on unordered $r \times c$ contingency tables. *Communications in Statistics* **15**, 387-403.

Mehta, C.R. e Patel, N.R. 1986b. A Fortran subroutine for Fisher's exact test on unordered $r \times c$ contingency tables. *ACM Trans. Math. Software* **12**, 154-161.

Roscoe, J.T. e Byars, J.A. 1971. An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square test. *J. Am. Statist. Assn*. **66**, 755-759.

Slakter, M.J. 1966. Comparative validity of the chi-square and two modified goodness- of-fit tests for small but equal expected frequencies. *Biometrika* **53**, 619-623.

Perguntas de interesse:

1. O que deve ser feito na presença de frequências esperadas muito pequenas em tabelas de contingência para se testar a independência das classificações de linha e de coluna?

2. O que devemos saber sobre a prática de combinar categorias de variáveis após a classificação em tabelas de contingência? Por quê?