

TESTE ESTATÍSTICO PARA COMPARAR PROPORÇÕES EM PROBLEMAS DE CITOGENÉTICA

Carlos Alberto de Bragança Pereira

INTRODUÇÃO

As características especiais de alguns dados coletados por citogeneticistas tornam, na análise de tais dados, os métodos estatísticos aqui discutidos mais convenientes do que aqueles tradicionalmente divulgados na literatura.

Em uma amostra investigada, para cada uma de suas unidades (geralmente uma célula) registra-se a ocorrência ou não de determinada anomalia (uma aberração cromossômica, por exemplo). Em Citogenética, a anomalia é, normalmente, um evento raro, isto é, necessita-se de um grande número de unidades amostrais para que um número significativo de ocorrências do evento seja observado. Há casos em que necessitamos pesquisar uma grande amostra de células (10000 por exemplo) para detectarmos a ocorrência de determinada aberração cromossômica. Em resumo, o resultado experimental tratado aqui é a frequência de ocorrências de um evento raro, quando se investiga uma grande amostra.

Consideremos um exemplo de ilustração (para casos realmente estudados veja Rabello et alii, 1975 e Lessa et alii, 1976). Para investigar o possível efeito citogenético de certo inseticida, foram estudados 5 operários que lidam com a droga (expostos) e 5 outros, da mesma fábrica, que não têm contacto com a droga (controles). Prepararam-se lâminas de culturas de linfócitos de cada operário e analisaram-se 1000 células de cada indivíduo, registrando-se o número de células com anomalias cromossômicas. Nas 5000 células dos expostos foram encontradas 6 com anomalias e nas 5000 dos controles encontrou-se uma única célula anômala. A Tabela 1 apresenta esses resultados.

O método tradicional de análise de diferença de proporções é o teste χ^2 de homogeneidade. A teoria mostra, entretanto, que esse teste é adequado para casos em que as amostras são grandes, mas a probabilidade do evento não se afasta muito de 0,5. Ora, no nosso exemplo, a frequência observada de anomalias, 7 em 10.000, sugere uma

Pereira, CAB. Teste Estatístico Para Comparar Proporções em Problemas de Citogenética. In: Rabello-Gay MN, Rodrigues MAR, Monteleone-Neto R. Mutagênese, Teratogênese e Carcinogênese: Métodos e Critérios de Avaliação. Ribeirão Preto: SBG; 1991. p.113-121.

probabilidade baixa para a ocorrência desse evento. Assim, o uso do teste poderia, teoricamente, ser questionado.

Tabela 1

Operários	Frequência de células		Tamanho da amostra
	com quebra	normais	
Expostos	6	4994	5000
Controles	1	4999	5000
Total	7	9993	10000

É verdade que existem correções que são utilizadas em situações de frequências baixas. Contudo, mesmo essas correções são teoricamente delineadas para os casos de amostras não muito grandes (equivalente ao caso onde a probabilidade não se afasta de 0,5), o que certamente não é o caso aqui discutido. Para analisar experimentos com grandes amostras e eventos de frequência baixa, os argumentos apresentados nos levam a recomendar um teste de significância alternativo, na linha do teste exato de Fisher (Kalbfleisch, 1979).

Comparando Proporções de Eventos Raros

Consideremos os dados da Tabela 1.

A hipótese que se pretende comprovar, H_1 , é a de que a frequência de células com anomalias nos indivíduos expostos é significativamente maior que nos controles. Consideramos que o total de células anômalas, 7 em nosso exemplo, não é uma quantidade informativa sobre a relação entre as probabilidades. O que é informativo é a relação entre as frequências, isto é, 1 para 6. Conseqüentemente, podemos admitir que a função de probabilidade sob a hipótese nula (H_0 : as duas proporções são iguais) é uma binomial com parâmetros $n = 7$ e $p = 0,5$. Isto é, no caso de não haver diferença entre expostos e controle, esperaríamos que os resultados obtidos na investigação fossem equivalentes ao número observado de "caras" obtidas após 7 lançamentos de uma moeda equilibrada. Assim, sob a hipótese nula, dado o total 7 de células com aberrações, a probabilidade condicional de que 6 células pertençam aos expostos é igual à probabilidade de obtermos 6 caras no lançamento da moeda. Usando a distribuição binomial, obtemos $7(0,5)^7 = 7/128$ para esta probabilidade. Um resultado mais extremo que o obtido seria o caso de observarmos as 7 células com aberrações nos indivíduos

expostos e nenhuma nos controles. A probabilidade neste caso seria $1/128$. O nível de significância descritivo para o teste seria então, $1/128 + 7/128 = 8/128 = 1/16 = 0,0625$. (Nível de significância descritivo de um teste é a probabilidade, sob a hipótese nula, do evento formado pelo que se observou e por todos os casos mais extremos que fossem passíveis de serem observados. Por mais extremos entendemos aqueles resultados que contrariam a hipótese nula mais do que os dados efetivamente observados. Nos casos mais comuns apresentados em livros de estatística, nível de significância descritivo é a cauda da curva, iniciada a partir do valor efetivamente observado). Concluímos, então, que a hipótese de igualdade entre proporções seria rejeitada a um nível de pelo menos 7%.

Consideremos uma anomalia que pode ser observada em unidades populacionais (células, no exemplo acima). Suponhamos que, nas populações P_1 e P_2 , as proporções de unidades com tal anomalia são representadas por p_1 e p_2 , respectivamente. No nosso exemplo, P_1 é a população formada por células de indivíduos expostos e P_2 por células de indivíduos que possam ser considerados como controle. O objetivo aqui é estudar a validade da hipótese $H_0: p_1 = p_2 = p$. Para este estudo, seleciona-se da população P_1 uma amostra de tamanho n_1 e da população P_2 uma amostra de tamanho n_2 . Em cada uma dessas amostras, registra-se a frequência de unidades com a anomalia. Representemos essas frequências por x_1 e x_2 e a soma por $t = x_1 + x_2$ (no exemplo, $n_1 = n_2 = 5000$, $x_1 = 6$, $x_2 = 1$ e $t = 7$). Antes de observarmos os valores de x_1 e x_2 esperaríamos que seus valores estivessem em torno de $n_1 p_1$ e $n_2 p_2$, respectivamente. É importante ressaltar aqui que, embora neste primeiro exemplo os tamanhos das amostras sejam iguais, isto não é uma restrição para o método. No exemplo da tabela 2, consideramos um caso onde $n_1 = 1000$ e $n_2 = 5000$.

Com a descrição acima, não seria difícil concluirmos que as variáveis x_1 e x_2 são independentes e distribuídas segundo duas binomiais com parâmetros (n_1, p_1) e (n_2, p_2) , respectivamente. Isto é, a função de probabilidade conjunta para (x_1, x_2) tem a seguinte expressão:

$$\binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2}$$

Lembremos que o primeiro e o quarto fator representam o coeficiente combinatório. Isto é, o 1º fator é o fatorial de n_1 dividido pelo produto do fatorial de x_1 vezes o de (n_1-x_1) . Analogamente definiríamos o outro coeficiente. No caso de a hipótese $H_0: p_1 = p_2 = p$ ser verdadeira, o total $t = x_1 + x_2$ tem distribuição binomial com parâmetro (n, p) onde $n = n_1 + n_2$. Isto é, a função de probabilidades da variável t , sob a hipótese H_0 , tem a seguinte forma:

$$\binom{n}{t} p^t (1-p)^{n-t}$$

Para simplificar os cálculos estatísticos, o uso de aproximações é recomendável, visto que estamos trabalhando com grandes valores de n_1 e n_2 . Como o contexto é o de eventos raros, isto é, p_1 e p_2 são valores pequenos (próximos de zero), produzindo valores de x_1 e x_2 baixos em relação a n_1 e n_2 , a aproximação ideal é o das distribuições binomiais para as distribuições de poisson.

Com estas aproximações, x_1 e x_2 serão consideradas variáveis de poisson com médias $n_1 p_1$ e $n_2 p_2$, respectivamente. Por outro lado, sob a hipótese H , t é considerada como uma poisson de média np . As expressões acima serão substituídas, respectivamente, por

$$\frac{(n_1 p_1)^{x_1} (n_2 p_2)^{x_2}}{x_1! x_2!} \exp\{-(n_1 p_1 + n_2 p_2)\}$$

e

$$\frac{(np)^t}{t!} \exp\{-np\},$$

onde $\exp\{a\}$ indica o número e elevado à potência a e $a!$ é o fatorial de a .

Usando a suposição de que t é uma quantidade não informativa no estudo da relação entre p_1 e p_2 , recomenda-se usar a distribuição condicional de (x_1, x_2) dado t na construção de um teste para H . Assim, sob a hipótese H , a função de probabilidades de x_1 condicionada ao valor de t é uma função de probabilidades binomial com parâmetro $(t, n_1/n)$. No exemplo apresentado, temos que $t = 7$ e $n_1/n = 5000/10000 = 0,5$. Sob a hipótese H , a probabilidade condicional de x_1 , dado o valor de t , será então

$$\binom{t}{x_1} \left(\frac{n_1}{n}\right)^{x_1} \left(\frac{n_2}{n}\right)^{x_2}$$

Concluindo a discussão acima, temos o seguinte procedimento para testar a hipótese H :

- (i) Considera-se t como sendo um valor pré-fixado.
- (ii) Após conhecido o valor de x_1 , procuramos quais os pontos amostrais possíveis que consideramos mais extremos que x_1 ; isto é, que contrariam H mais do que x_1 contraria.

(iii) Para cada um desses pontos, calcula-se sua probabilidade condicional (dado t), assim como foi feito para x_1 .

(iv) Considera-se como nível de significância descritivo a soma dessas probabilidades. Caso este valor seja superior ao nível de significância pré-fixado, aceitaríamos H . Caso contrário, rejeitaríamos H .

Voltando ao exemplo, lembremos que $t = 7$ e que o único ponto mais extremo que $x_1 = 6$ seria o ponto 7 o que acarretaria em ter-se $x_2 = 0$, no lugar de $x_2 = 1$. Como $n_1/n = 5000/10000 = 0,5$, temos que o nível descritivo do exemplo é 0,0625, exatamente como calculado anteriormente.

Vamos supor que nos expostos apenas 1000 células foram investigadas e foram encontradas 2 células com anomalias. A Tabela 2 apresenta esses resultados.

Na análise dos dados da Tabela 2 teríamos $t = 3$, $x_1 = 2$ e $n_1/n = 1000/6000 = 1/6$. Notemos que um ponto mais extremo que $x_1 = 2$ seria $x_1 = 3$. As probabilidades condicionais desses pontos (dado $t = 3$) são, respectivamente,

$$3 \binom{1}{6}^2 \left(\frac{5}{6}\right) = 0,0694 \quad e \quad \binom{1}{6}^3 = 0,0046.$$

O nível descritivo, neste caso, seria $0,0694 + 0,0046 = 0,0740$. Com os dados da Tabela 2, rejeitaríamos a hipótese H em um nível de pelo menos 7,5%.

É interessante ressaltar que os dados da Tabela 2 produzem um nível de significância (7,5%) ligeiramente maior do que o produzido pelos dados da Tabela 1 (7%). A intuição poderia nos enganar pois a proporção 2/1000 (da tabela 2) é 10 vezes maior do que 1/5000 enquanto 6/5000 (da Tabela 1) é apenas 6 vezes maior. Isto poderia nos levar a pensar que a Tabela 2 apresenta uma maior significância do que a Tabela 1. A inversão dos papéis se deve ao menor tamanho de amostra na Tabela 2.

Tabela 2

Operários	Frequência de células		Tamanho da amostra
	com quebra	normais	
Expostos	2	998	1000
Controles	1	4999	5000
Total	3	5997	6000

Aproximações e Extensões

O exemplo agora discutido refere-se a um estudo sobre o efeito de um composto no índice mitótico de células de medula óssea. O índice mitótico - razão entre a frequência de metáfases em células tratadas sobre a frequência total de células - serve para avaliação da toxicidade de um composto, na medida em que qualquer alteração no valor desse índice pode indicar toxicidade para a medula.

O experimento consistiu na observação dos índices mitóticos de células da medula óssea de camundongos, divididos em dois grupos distintos, tratados (expostos ao composto) e controles (não expostos ao composto). Cada grupo foi dividido em 3 subgrupos. No primeiro, os camundongos foram sacrificados após 24 horas de exposição; no segundo, após 48 horas e no terceiro, após 72 horas. Os resultados do experimento são descritos na Tabela 3.

Tabela 3

Grupos	Exposição	Nº animais	Nº células	Nº de células em metáfase
Tratado	24 h	6	18000	358
Controle	24 h	5	15000	424
Tratado	48 h	6	18000	256
Controle	48 h	5	15000	378
Tratado	72 h	6	18000	342
Controle	72 h	4	12000	257

Consideremos inicialmente comparações (tratados/controles) dentro de cada exposição. Em seguida, comparemos as diferentes exposições (24 h, 48 h, 72 h) dentro dos tratados e dentro dos controles. Como estas tabelas parciais já evidenciaram diferenças significantes entre tratado e controle não há necessidade de uma comparação múltipla na tabela completa. A Tabela 3 apresenta todos os resultados, parciais e completos.

Our objective initial is the comparison between treated and controls in 3 subgroups, 24 h, 48 h and 72 h. If we concentrate on 24 hours and remember the analysis described in the previous section, we have $x_1 = 358$, $x_2 = 424$ and $t = x_1 + x_2 = 782$. Again, we are in the presence of rare events (compare the values of x_1 and x_2 with the sizes of sample n_1 and n_2 , respectively) and thus all that was considered in the previous section

se aplica neste caso. Isto é, se não existe efeito do tratamento, a distribuição condicional de x_1 dado t é binomial com parâmetros t e $p = n_1/(n_1 + n_2) = 18000/33000 = 6/11$. A Tabela 4 apresenta os valores observados e os valores esperados sob a hipótese nula de não existência de efeito de tratamento.

Tabela 4

	Observado	Esperado	$\frac{(o - e)^2}{e}$
Tratados	358	427	11.15
Controles	424	355	13.41
Total	782 (t)	782	24,56 (χ^2)

Neste caso, como o número total de metáfases é relativamente alto e as frequências esperadas são altas, podemos usar a estatística qui-quadrado no lugar de calcularmos o nível descritivo do teste, como fizemos na seção anterior. Temos assim o clássico teste χ^2 com um grau de liberdade. O valor dessa estatística neste caso é alto, 24,56, o que nos leva a concluir por uma diferença altamente significativa entre tratados e controles, na exposição de 24 h.

Um cálculo análogo para exposição de 48 h produziria a seguinte tabela:

Tabela 5

	Observado	Esperado	$\frac{(o - e)^2}{e}$
Tratados	256	346	23.41
Controles	378	288	28,12
Total	634 (t)	634	51,53 (χ^2)

Novamente, o valor da estatística qui-quadrado é grande, levando-nos a rejeitar a hipótese de igualdade entre as duas proporções.

O cálculo para o caso de exposição de 72 h é análogo, exceto quanto ao fato de que $p = n_1/(n_1 + n_2) = 18000/30000 = 3/5$. Os valores observados e esperados para esta exposição são apresentados na Tabela 6:

Tabela 6

	Observado	Esperado	$\frac{(o - e)^2}{e}$
Tratados	342	359	0,81
Controles	257	240	1,20
Total	599 (t)	599	2,01 (χ^2)

Neste caso o valor da estatística qui-quadrado é baixo, levando-nos a não rejeição da hipótese de igualdade entre as duas proporções.

Suponhamos agora que desejamos estudar o efeito do tempo de exposição sobre o índice mitótico. Devemos assim estudar as diferenças de exposição entre os tratados e também entre os controles. A necessidade de dois estudos deve-se ao fato de termos concluído que existem diferenças entre os tratados e controles, dentro de cada tipo de exposição. É intuitivo que se construirmos os mesmos argumentos anteriores, sob a hipótese de igualdade entre proporções, a distribuição conjunta das frequências de metáfases nas 3 diferentes exposições, condicionada em t (total de metáfases do grupo), é uma trinomial com parâmetros t e $(n_1/n, n_2/n, n_3/n)$. Assim, no grupo dos tratados temos $t = 956$ e $n_1/n = n_2/n = n_3/n = 18000/54000 = 1/3$. (*Consideremos n seleções, com reposição, de uma urna com bolas de três cores. A distribuição conjunta das frequências amostrais das 3 cores é denominada trinomial.*). No grupo controle temos $t = 1059$, $n_1/n = n_2/n = 15000/42000 = 5/14$ e $n_3/n = 12000/54000 = 4/14$. Os valores observados e esperados para os grupos dos tratados e dos controles são apresentados respectivamente nas Tabelas 7 e 8.

Tabela 7

Exposição	Observado	Esperado	$\frac{(o - e)^2}{e}$
Tratados			
24 h	358	318	5,03
48 h	256	319	12,44
72 h	342	319	1,66
Total	956 (t)	956	19,13 (χ^2)

Tabela 8

Exposição	Observado	Esperado	$\frac{(o - e)^2}{e}$
Controles			
24 h	424	378	5,60
48 h	378	378	0,00
72 h	257	303	6,98
Total	1059 (t)	1059	12,58 (χ^2)

Lembremos que a estatística qui-quadrado nestes casos tem dois graus de liberdade e, com os níveis padrões, rejeitaríamos a hipótese de igualdade em ambas as tabelas. A única tabela que não nos leva a concluir pela não significância é a Tabela 6. Intencionalmente não discutimos os aspectos básicos da estatística. Existem muitos textos elementares de estatística onde pode-se encontrar um farto material sobre testes de significância; recomendamos o excelente texto de Noether (1976).

REFERÊNCIAS BIBLIOGRÁFICAS

- KALBFLEISCH J.G. 1979. Probability and statistical inference. Vols. I & II. New York, Springer-Verlag.
- LESSA J.M.M., BEÇAK W., RABELLO M.N., PEREIRA C.A. DE B., UNGARO M.T. 1976. Cytogenetic study of DDT on human lymphocytes *in vitro*. *Mutation Res.*, 40: 131-138.
- NOETHER G.E. 1976. Introduction to statistics: A nonparametric approach. 2nd edition. (Tradução para o português por Rodrigues, F.W. & Cordani, L.K., s.d., Guanabara Dois, Introdução à estatística: uma abordagem não paramétrica). Boston, Houghton Mifflin.
- RABELLO M.N., BEÇAK W., DE ALMEIDA W.F., PIGATI P., UNGARO M.T., MURATA T., PEREIRA C.A. DE B. 1975. Cytogenetic study on individuals occupationally exposed to DDT. *Mutation Res.*, 28: 449-454.
- RIBEIRO L.R., RABELLO-GAY M.N., SALVADORI D.M.F., PEREIRA C.A. DE B., BEÇAK W. 1987. Cytogenetic effects of inhaled ethylene oxide in somatic and germ cells of mice. *Arch. Toxicol.* 59: 332-335.
- SALVADORI D.M.F., RIBEIRO L.R., PEREIRA C.A. DE B., BEÇAK W. 1988. Cytogenetic effects of malathion insecticide on somatic and germ cells of mice. *Mutation Res.*, 204: 283-287.