

Teste de hipóteses de associação pelo Qui-quadrado de Pearson (χ^2)

O qui-quadrado é obtido somando-se razões dadas pelos quadrados das diferenças entre frequências observadas e as esperadas, divididos pelas frequências esperadas.

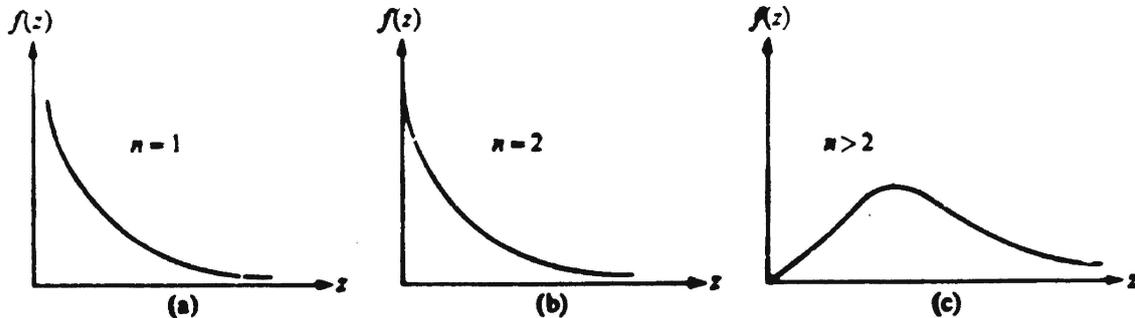
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Quando as variáveis são independentes, é equivalente a dizer que não existe associação, e neste caso, o valor do qui-quadrado será zero. O qui-quadrado não mede força de associação e não é suficiente para estabelecer relação de causa e efeito.

Distribuição qui-quadrado ($\chi^2_{(n-1)}$) com (n-1) graus de liberdade

Seja uma população com distribuição normal $N(\mu, \sigma)$. Se desta população se obter um número infinito de amostras de tamanho n , calculando-se as quantidades \bar{X} e S^2 em cada amostra, a variável aleatória $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$, onde $\chi^2_{(n-1)}$ se lê "qui-quadrado com n-1 graus de liberdade" Berquó (1981).

A distribuição qui-quadrado é assimétrica e se torna menos assimétrica à medida que os graus de liberdade aumentam. Os valores da distribuição são sempre positivos (maior ou igual a zero). Existe uma família de distribuições qui-quadrado, dependendo do número de graus de liberdade. Para grandes amostras, a distribuição qui-quadrado tende para uma distribuição normal.



Abordagem de Neyman e Pearson

Estabelecimento das hipóteses:

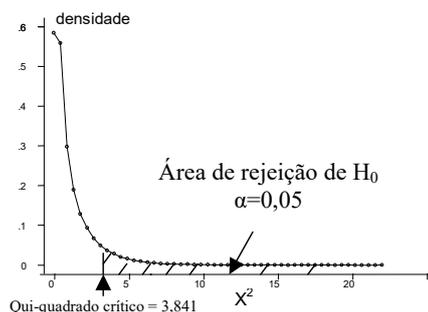
$$\begin{cases} H_0: \text{Não existe associação} \\ H_a: \text{Existe associação} \end{cases}$$

Fixando-se a probabilidade de erro tipo I:

Nível de significância (α) = 0,05

Para a tomada de decisão, utiliza-se a regra: rejeita-se H_0 se o valor calculado do qui-quadrado for maior do que o valor crítico para um nível de significância pré-definido.

Área de rejeição do teste:



Estatística do teste:

$$\text{Qui-quadrado} = \sum \frac{(O - E)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

onde r e c representam o número de linhas e de colunas, respectivamente.

Correção de continuidade:

$$\text{Qui-quadrado}_{\text{correcao de Yates}} = \sum \frac{(|O - E| - 0,5)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

Limitações:

Para $n < 20$, utilizar o teste exato de Fisher

Para $20 \leq n \leq 40$, utilizar o qui-quadrado somente se os valores esperados forem maiores ou iguais a 5

Exemplo

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo condição de sobrevivência e peso ao nascer (g). Local? Ano?

Peso ao nascer (g)	Óbito	Sobrevida	Total
Baixo peso (<2500g)	24	13	37
Não baixo peso (2500g e mais)	3	10	13
Total	27	23	50

Fonte: Hand DJ et al. A handbook of small data sets. Chapman & Hall, 1994.

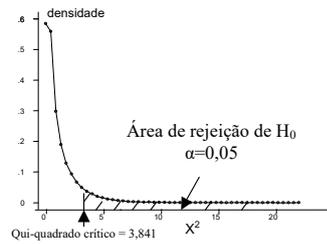
Hipóteses:

$$\begin{cases} H_0: \text{Não existe associação} \\ H_a: \text{Existe associação} \end{cases}$$

Fixando-se a probabilidade de erro tipo I:

Nível de significância (α) = 0,05

Área de rejeição do teste:



Cálculo do qui-quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$	$\frac{(O-E -0,5)^2}{E}$
24	19,98	4,02	16,16	0,809	0,62
3	7,02	-4,02	16,16	2,302	1,77
13	17,02	-4,02	16,16	0,949	0,73
10	5,98	4,02	16,16	2,702	2,07
				$\chi^2 = 6,762$	$\chi^2_{\text{corrigido}} = 5,19$

Decisão: rejeita-se H₀

Conclusão: As variáveis estão associadas para nível de significância de 5%

Coeficiente de associação de Yule

Coeficiente de associação de Yule – permite investigar a força (magnitude) da associação

$$Y = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}, \text{ onde: } -1 \leq Y \leq +1$$

$$Y = \frac{24 \times 10 - 3 \times 13}{24 \times 10 + 3 \times 13} = +0,72$$

Calculando-se os percentuais:

Peso ao nascer (g)	Óbito		Sobrevida		Total	
	n	%	n	%	n	%
Baixo peso (<2500g)	24	64,9	13	35,1	37	100
Não baixo peso (2500g e mais)	3	23,1	10	76,9	13	100
Total	27	54,0	23	46,0	50	100

A associação entre peso ao nascer e condição de sobrevivência é forte. A proporção de óbitos é maior entre recém-nascidos de baixo peso se comparados aos de não baixo peso.

Teste de associação pelo qui quadrado de Pearson - Abordagem do valor de p

Pela tabela da distribuição qui-quadrado, com 1 gl, $p = 0,010$

Calculando-se o valor de p pelo Excel, para 1 gl, o valor de p não corrigido = 0,009

No Excel utilizar a função DIST.QUI tendo como argumentos o valor calculado do qui-quadrado e o número de graus de liberdade: = DIST.QUI(6,762;1))

Conclusão: Existe forte evidência contrária à independência. Portanto a associação observada ocorre não devido ao acaso. Pode-se dizer que os dados são compatíveis com existência de associação entre peso ao nascer e condição de sobrevivência de recém-nascidos com SFDIG ($p = 0,009$).

Exemplo utilizando a estatística qui quadrado corrigido

Com o objetivo de investigar a associação entre história de bronquite na infância e presença de tosse diurna ou noturna em idades mais velhas, foram estudados 1319 adolescentes com 14 anos. Destes, 273 apresentaram história de bronquite até os 5 anos de idade sendo que 26 apresentaram tosse diurna ou noturna aos 14 anos.

Número de adolescentes segundo história de bronquite aos 5 anos e tosse diurna ou noturna aos 14 anos de idade. Local X, ano Y.

Tosse	Bronquite		Total
	Sim	Não	
Sim	26	44	70
Não	247	1002	1249
Total	273	1046	1319

Fonte: Holland, WW et al.. Long-term consequences of respiratory disease in infancy. *Journal of Epidemiology and Community Health* 1978; 32: 256-9.

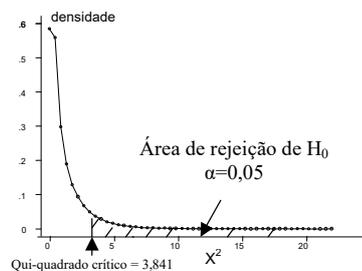
Hipóteses:

$$\left\{ \begin{array}{l} H_0: \text{Não existe associação (variáveis são independentes)} \\ H_a: \text{Existe associação} \end{array} \right.$$

Fixando-se a probabilidade de erro tipo I:

Nível de significância (α) = 0,05

Área de rejeição do teste:



Cálculo do qui quadrado

Valores observados (O)	Valores esperados (E)	(O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$	$\frac{(O-E -0,5)^2}{E}$
26	14,488	11,512	132,526	9,147	8,37
247	258,512	-11,512	132,526	0,513	0,469
44	55,512	-11,512	132,526	2,387	2,184
1002	990,488	11,512	132,526	0,134	0,122
				$\chi^2 = 12,181$	$\chi^2_{\text{corrigido}} = 11,145$

Decisão:

O valor do qui-quadrado calculado é maior do que o valor do qui-quadrado crítico para 1 grau de liberdade e nível de significância de 5%, portanto, rejeita-se H₀.

Conclusão: Pode-se dizer que na população existe associação entre bronquite na infância e tosse na adolescência.

Coefficiente de associação de Yule

Coefficiente de associação de Yule – permite investigar a força (magnitude) da associação

$$Y = \frac{a.d - b.c}{a.d + b.c}, \text{ onde: } -1 \leq Y \leq +1$$

$$Y = \frac{26 \times 1002 - 44 \times 247}{26 \times 1002 + 44 \times 247} = +0,41$$

Calculando-se os percentuais:

Tosse	Bronquite (Sim)		Bronquite (Não)		Total	
	n	%	n	%	n	%
Sim	26	9,5	44	4,2	70	5,3
Não	247	90,5	1002	95,8	1249	94,7
Total	273	100	1046	100	1319	100

A associação entre tosse na adolescência e bronquite, mas não é forte; entretanto, não pode ser desprezada. Adolescentes apresentam mais tosse na adolescência se tiveram bronquite na infância, comparados aos que não tiveram bronquite na infância.

Teste de associação pelo qui quadrado de Pearson – utilizando o valor de p

Pela tabela da distribuição qui-quadrado, com 1 gl, p<0,001 (na tabela, menor que 0,1%)

Calculando-se o valor de p pelo Excel, para 1 gl, o valor de p não corrigido = 0,0004829
 No Excel utilizar a função DIST.QUI tendo como argumentos o valor calculado do qui-quadrado e o número de graus de liberdade: = DIST.QUI(12,181;1))

Conclusão: Existe forte evidência contrária à independência. Portanto a associação observada ocorre não devido ao acaso. Pode-se dizer que os dados são compatíveis com existência de associação entre bronquite na infância e tosse na adolescência, na população.

Contingency Tables

A		B		Total
		1	2	
1	Observed	26	44	70
	Expected	14.5	55.5	70.0
	% within column	9.5 %	4.2 %	5.3 %
2	Observed	247	1002	1249
	Expected	258.5	990.5	1249.0
	% within column	90.5 %	95.8 %	94.7 %
Total	Observed	273	1046	1319
	Expected	273	1046	1319
	% within column	100.0 %	100.0 %	100.0 %

 χ^2 Tests

	Value	df	p
χ^2	12.2	1	< .001
χ^2 continuity correction	11.1	1	< .001
N	1319		

Exercício

Investigação de toxiinfecção alimentar

Tomou sorvete de baunilha	Toxiinfecção		Total
	Sim	Não	
Sim	43	11	54
Não	3	18	21
Total	46	29	75

Fonte: Epi Info, 2000.

Calcule

- O Qui quadrado de Pearson
- O coeficiente de associação de Yule
- Você diria que as variáveis estão associadas?