Testes χ^2 de Pearson

Testes χ^2 de Pearson

Aviso: Esta matéria não faz parte do Modelo Linear.

Ferramenta para o estudo de dados de contagem (valores em $\mathbb{N}_0$): frequências observadas de certo fenómeno.

Objectivo: testar se as contagens observadas são compatíveis com determinadas hipóteses.

Nestes Testes de Hipóteses, existe uma estatística de teste comum: a estatística de Pearson.

Os testes são também chamados testes χ^2, uma vez que a estatística de teste segue, assintoticamente, uma distribuição qui-quadrado.

Teste χ^2 com tabelas de contingência

Consideramos apenas o caso de contagens em tabelas de dupla entrada, conhecidas por tabelas de contingência: classificam-se observações em categorias que resultam de combinar os níveis de 2 factores.

No contexto dum exemplo de biodiversidade, admita-se que:

- há *a* locais geográficos, os níveis de um factor A;
- interessa observar *b* espécies, os níveis dum factor B.

Uma tabela de contingência (com *ab* células) pode dar o número de observações de cada espécie, em cada local (para um total fixo *N* de observações).

Assim, as *ab* contagens são classificadas de acordo com dois diferentes factores (o que obrigará a usar uma dupla indexação).

Tabelas de contingência

Chama-se tabela de contingência a uma tabela com o número O_{ij} de observações em cada célula (i,j) (nível i do factor A e j do factor B):

Níveis do Factor A	Níveis do Factor B					Marginal de A
	1	2	3	$\cdots$	b	
1	O_{11}	O_{12}	O_{13}	$\cdots$	$O_{1,b}$	$N_{1\cdot}$
2	O_{21}	O_{22}	O_{23}	$\cdots$	$O_{2,b}$	$N_{2\cdot}$
3	O_{31}	O_{32}	O_{33}	$\cdots$	$O_{3,b}$	$N_{3\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
a	O_{a1}	O_{a2}	O_{a3}	$\cdots$	$O_{a,b}$	$N_{a\cdot}$
Marginal de B	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$\cdots$	$N_{\cdot b}$	N

Serão úteis os totais marginais de linha e coluna e o total global:

$$N_{i.} = \sum_{j=1}^{b} O_{ij} \qquad N_{.j} = \sum_{i=1}^{a} O_{ij} \qquad N = \sum_{i=1}^{a}\sum_{j=1}^{b} O_{ij}$$

A estatística de Pearson

Nos testes χ^2 comparam-se:

- as contagens observadas (indicadas por O_{ij}); com
- as correspondentes contagens esperadas ao abrigo de alguma hipótese, indicadas por E_{ij}.

A maior ou menor proximidade global entre contagens observadas O_{ij} e esperadas E_{ij} contém informação sobre a plausibilidade da hipótese que gerou os valores esperados E_{ij}.

Estatística de Pearson para tabelas de contingência

No contexto agora descrito, a estatística de Pearson tem a forma

$$X^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Cálculo dos valores esperados

Frequentemente, as hipóteses associadas aos valores esperados são formuladas especificando probabilidades de pertença à célula (i,j), que designamos π_{ij}.

Nesse contexto, se N indica o número total de observações a distribuir pelas células da tabela de contingência, os valores esperados em cada célula (i,j) da tabela são calculados por uma regra simples:

$$E_{ij} \quad = \quad N \times \pi_{ij} \ .$$

Na definição das probabilidades π_{ij} surgem diferentes situações, que veremos em seguida.

Diferentes contextos para tabelas de contingência

- As probabilidades π_{ij} de recair em cada célula (i, j) podem ser totalmente especificadas. Vamos exemplificar esta situação com exemplos ligados à genética.

- As probabilidades π_{ij} de recair em cada célula (i, j) podem ser não inteiramente conhecidas, mas estimáveis. Vamos exemplificar esta situação com dois contextos frequentes:
  - testes de independência
  - testes de homogeneidade

Estatística do teste para tabelas bidimensionais

No contexto agora descrito, a estatística de Pearson tem a forma

$$X^2 \quad = \quad \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

e segue assintoticamente uma distribuição χ^2_{ab-r}.

Os graus de liberdade são dados pela diferença entre

- o número de células *ab* com contagens; e
- o número de restrições *r* associadas ao problema.

As restrições são em geral de dois tipos:

- quando se fixam dimensões de amostra (soma de contagens); e
- quando se estimam probabilidades.

Validade da distribuição assintótica

A distribuição da estatística de Pearson é apenas assintótica, ou seja, aproximada, para grandes amostras.

Há critérios diferentes para quando se considera a aproximação adequada.

Critério de Cochran

Um critério, proposto por Cochran, é:

- nenhum E_{ij} inferior a 1;
- não mais do que 20% dos E_{ij}s inferiores a 5.

Caso estas condições não se verifiquem, devem-se agrupar classes (linhas e/ou colunas da tabela) de forma a satisfazer o critério.

Região Crítica

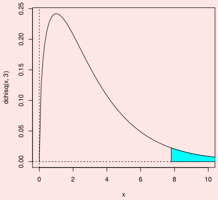
Discrepâncias grandes entre os valores esperados (E_{ij}) e observados (O_{ij}) dão origem a valores grandes da estatística.

Logo, tem-se uma Região Crítica unilateral direita.

Regra de rejeição

Sendo H_0 a hipótese que gerou os valores esperados E_{ij},

$$\text{Rejeita-se } H_0 \text{ se } \quad X^2_{calc} > \chi^2_{\alpha;(ab-r)},$$



Situação 1: probabilidades totalmente especificadas

Exemplo: Pêlo de coelhos

Supõe-se que, em coelhos, existe:

- um gene que controla a cor do pêlo, com:
  - um alelo determinante do cinzento (dominante);
  - um alelo determinante do branco (recessivo).
- outro gene que controla o tipo de pelagem, com:
  - um alelo determinante do pêlo normal (dominante);
  - um alelo determinante da pelagem tipo Rex (recessivo).

Para avaliar esta hipótese, realiza-se uma experiência cruzando coelhos duma população inicial que são heterozigóticos nos dois genes, i.e., têm um alelo de cada cor e um alelo de cada tipo de pelagem.

Pêlo de coelhos

Se a segregação dos genes for independente, isto é, se o alelo da cor for independente do alelo do tipo de pelagem, as probabilidades π_{ij} das quatro possíveis combinações de características nos descendentes serão:

- $\pi_{11} = 9/16$ coelhos cinzentos de pelagem normal;
- $\pi_{12} = 3/16$ coelhos cinzentos de pelagem tipo Rex;
- $\pi_{21} = 3/16$ coelhos brancos de pelagem normal;
- $\pi_{22} = 1/16$ coelhos brancos de pelagem tipo Rex;

O número esperado de observações em cada célula será dado por $E_{ij} = N \times \pi_{ij}$, sendo π_{ij} a probabilidade associada à cor i e pelagem j:

E_{ij}		Pêlo	
		Normal	Rex
Cor	Cinzento	$N \times \frac{9}{16}$	$N \times \frac{3}{16}$
	Branco	$N \times \frac{3}{16}$	$N \times \frac{1}{16}$

Exemplo: Pêlo de coelhos

Existem $a=2$ cores do pêlo e $b=2$ tipos de pelagem, num total de $ab=4$ células. Numa descendência de $N=232$ coelhos, esperar-se-iam:

E_{ij}		Pêlo	
		Normal	Rex
Cor	Cinzento	$232 \times \frac{9}{16} = 130.5$	$232 \times \frac{3}{16} = 43.5$
	Branco	$232 \times \frac{3}{16} = 43.5$	$232 \times \frac{1}{16} = 14.5$

É admissível a distribuição assintótica: o menor E_{ij} tem valor 14.5.
Observaram-se:

O_{ij}		Pêlo	
		Normal	Rex
Cor	Cinzento	134	44
	Branco	42	12

Os valores observados são compatíveis com os esperados ao abrigo da hipótese de segregação independente e dominância/recessividade dos alelos de cada gene?

Distribuição assintótica se π_{ij} especificados

No caso de probabilidades totalmente especificadas, há apenas uma restrição: o número total de observações N que foi utilizado na experiência.

Logo,

$$X^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_{ab-1}$$

No exemplo do pêlo dos coelhos tem-se $ab-1=3$, logo rejeita-se H_0 (as probabilidades resultantes do mecanismo genético proposto) se X^2_{calc} recair na cauda direita duma distribuição χ^2_3.

Exemplo: Pêlo de coelhos

O valor da estatística de Pearson para a amostra referida é

$$X^2_{calc} = \frac{(134-130.5)^2}{130.5} + \frac{(44-43.5)^2}{43.5} + \frac{(42-43.5)^2}{43.5} + \frac{(12-14.5)^2}{14.5} = 0.5823755 \, .$$

A fronteira da região crítica ao nível $\alpha = 0.05$ é

$$\chi^2_{0.05(3)} = 7.814728 \, .$$

Logo, não se rejeita H_0, isto é, não se rejeitam as hipóteses genéticas referidas (dominância/recessividade e segregação independente dos genes).

Segunda situação: Testes de homogeneidade

Consideremos agora situações onde as hipóteses nulas a testar não estão totalmente especificadas, e exigem a estimação de parâmetros. Veremos dois casos particulares frequentes, desta situação.

Inicialmente, admitimos que o número de observações em cada nível de um dos factores é previamente fixado pelo experimentador.

Para fixar ideias, admita-se que os a totais de linha, $N_{i.}$, foram previamente determinados pelo experimentador.

Neste caso, objectivo de interesse pode ser o de ver se as $N_{i.}$ observações de cada linha (nível do factor A) se distribuem de forma análoga (homogénea) pelas b colunas (níveis do factor B).

Um teste com este objectivo chama-se um teste de homogeneidade.

Exemplo - Teste de Homogeneidade

Exemplo: larvas de insectos

Nos solos duma dada região foi assinalada a presença de larvas de 4 espécies de insectos que afectam as principais culturas da região.

Pretende-se investigar se as frequências relativas das espécies são, ou não, iguais nos vários tipos de solos.

Classificaram-se os solos em três tipos: arenosos, limosos e argilosos (Factor A, com a=3 níveis).

Em cada tipo de solos foram recolhidas 100 larvas ($N_{i.} = 100$, $\forall\ i$), sendo classificadas de acordo com a respectiva espécie (Factor B, com b=4 níveis).

Exemplo - Teste de Homogeneidade (cont.)

Exemplo das larvas (teste de homogeneidade)

Feita a classificação das larvas, obtiveram-se os seguintes resultados:

		Espécie de larva				Total
		1	2	3	4	
Tipos	Arenosos	27	24	23	26	100
de	Limosos	20	32	18	30	100
solos	Argilosos	13	37	16	34	100
	Total	60	93	57	90	300

O objectivo é saber se se pode admitir que, nos três tipos de solos, a distribuição das quatro espécies de larvas é idêntica, ou seja, se há homogeneidade nas distribuições pelas espécies, em cada tipo de solo.

Hipóteses num teste de homogeneidade

A hipótese nula é que as probabilidades de cada espécie, condicionais ao solo i ($\pi_{1|i}$, $\pi_{2|i}$, $\pi_{3|i}$ e $\pi_{4|i}$) não dependem do tipo de solo i dado.

Designando por $\pi_{j|i}$ a probabilidade da espécie j, dado o solo i:

$$H_0 \ : \ \left\{ \begin{array}{ll} \pi_{1|1} = \pi_{1|2} = \pi_{1|3} & [\, = \pi_{.1}\,] \\ \pi_{2|1} = \pi_{2|2} = \pi_{2|3} & [\, = \pi_{.2}\,] \\ \pi_{3|1} = \pi_{3|2} = \pi_{3|3} & [\, = \pi_{.3}\,] \\ \pi_{4|1} = \pi_{4|2} = \pi_{4|3} & [\, = \pi_{.4}\,] \end{array} \right.$$

A hipótese alternativa H_1 é que pelo menos uma das igualdades acima referidas não é verdadeira.

Mas que valores usar para as probabilidades $\pi_{.j}$ $(j = 1, 2, 3, 4)$?

A estimação das probabilidades

As frequências relativas $\frac{N_{\cdot j}}{N}$ de cada espécie estimam as probabilidades das espécies, caso haja homogeneidade.

Exemplo de Teste de homogeneidade (larvas)

		Espécie de larva				
		1	2	3	4	Total
Tipos	Arenosos	27	24	23	26	100
de	Limosos	20	32	18	30	100
solos	Argilosos	13	37	16	34	100
	Total	60	93	57	90	300

A probabilidade estimada da espécie j será $\hat{\pi}_{\cdot j} = \frac{N_{\cdot j}}{N}$, ou seja:

$$\hat{\pi}_{\cdot 1} = \frac{60}{300} = 0.20 \qquad \hat{\pi}_{\cdot 2} = \frac{93}{300} = 0.31 \qquad \hat{\pi}_{\cdot 3} = \frac{57}{300} = 0.19 \qquad \hat{\pi}_{\cdot 4} = \frac{90}{300} = 0.30$$

Os valores esperados (estimados)

Uma vez que em cada tipo de solo há $N_{i.} = 100$ observações, o número esperado de observações na célula (i,j) é estimado por

$$\hat{E}_{ij} \quad = \quad N_{i.} \times \hat{\pi}_{.j} \quad = \quad N_{i.} \times \frac{N_{.j}}{N}$$

Exemplo das larvas (teste de homogeneidade)

A tabela com os valores esperados estimados entre parenteses:

		Espécie de larva				Total
		1	2	3	4	
Tipos de solos	Arenosos	27 (20)	24 (31)	23 (19)	26 (30)	100
	Limosos	20 (20)	32 (31)	18 (19)	30 (30)	100
	Argilosos	13 (20)	37 (31)	16 (19)	34 (30)	100
	Total	60	93	57	90	300

Entre as observações O_{ij} e os correspondentes valores esperados estimados ($\hat{E}_{ij}$), existe concordância suficiente para admitir que as espécies se distribuem de forma análoga nos três tipos de solos?

As restrições - Teste de Homogeneidade

Os graus de liberdade na distribuição assintótica da estatística de Pearson são o número de células (ab) menos o número de restrições (r).

Existem:

- a restrições resultantes de fixar o número de observações em cada linha (níveis do factor A, tipos de solo).
- $b-1$ restrições resultantes da necessidade de estimar as probabilidades de espécie (níveis do factor B).

NOTA: Não são b restrições pois a soma dos $\hat{\pi}_i$ tem de ser 1, logo estimar $b-1$ probabilidades determina a última estimativa.

Ao todo foram impostas $r = a+b-1$ restrições.

A estatística de Pearson em testes de homogeneidade

Nos testes de homogeneidade, os graus de liberdade da distribuição χ^2 são:

$$ab - r \;=\; ab - (a+b-1) \;=\; a(b-1) - (b-1) \;=\; (a-1)(b-1)$$

Nos Testes de Homogeneidade, a estatística de Pearson tem a distribuição assintótica $\chi^2_{(a-1)(b-1)}$:

$$X^2 \;=\; \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \;\sim\; \chi^2_{(a-1)(b-1)}$$

Rejeita-se a hipótese de homogeneidade H_0 se $X^2_{calc} > \chi^2_{\alpha;(a-1)(b-1)}$.

Exemplo de Teste de Homogeneidade

Exemplo: larvas de insectos

A distribuição assintótica é admissível: o menor E_{ij} é 19.

A estatística de Pearson calculada no exemplo das larvas tem valor

$$X^2_{calc} = 10.10928 \ .$$

Os g.l. da distribuição χ^2 são $(a-1)(b-1) = 2 \times 3 = 6$. A fronteira da região crítica, para $\alpha = 0.05$, é:

$$\chi^2_{0.05(6)} = 12.591 \ .$$

Como $X^2_{calc} < \chi^2_{0.05(6)}$ não se rejeita H_0, a hipótese de homogeneidade das distribuições de espécies de larva, nos três tipos de solos.

Situação 3: Testes de independência

Hipótese de independência

Numa tabela bidimensional, há independência quando as probabilidades conjuntas são o produto das probabilidades marginais:

$$H_0 \; : \quad \pi_{ij} \quad = \quad \pi_{i.} \times \pi_{.j} \; , \qquad \forall \, i,j$$

onde

- π_{ij} indica a probabilidade duma observação recair na célula (i,j);
- $\pi_{i.}$ indica a probabilidade marginal duma observação recair no nível i do factor A (seja qual for o nível do outro factor);
- $\pi_{.j}$ indica a probabilidade marginal duma observação recair no nível j do factor B (seja qual for o nível do outro factor);

Estimação das probabilidades

Pode haver situações onde as probabilidades marginais sejam conhecidas, mas em geral não o são.

É possível estimar as probabilidades marginais (das duas margens) a partir das frequências relativas marginais (como foi feito nos testes de homogeneidade, para o factor B):

$$\hat{\pi}_{i.} = \frac{N_{i.}}{N} \quad , \quad \forall\, i = 1,2,...,a$$

$$\hat{\pi}_{.j} = \frac{N_{.j}}{N} \quad , \quad \forall\, j = 1,2,...,b\,,$$

onde:

- N – número total de observações (fixo);
- $N_{i.}$ – número (livre) de observações no nível *i* do factor A;
- $N_{.j}$ – número (livre) de observações no nível *j* do factor B.

Valores esperados

Caso se verifique a independência, o número esperado de observações na célula (i,j) é dado por:

$$E_{ij} \quad = \quad N \times \pi_{ij} \quad = \quad N \times \pi_{i.} \times \pi_{.j} \qquad \forall \, i,j.$$

Estimando as probabilidades marginais, o número esperado estimado de observações na célula (i,j), ao abrigo da hipótese de independência, é:

$$\hat{E}_{ij} = N\,\hat{\pi}_{ij} = N\,\hat{\pi}_{i.}\,\hat{\pi}_{.j} = N \times \frac{N_{i.}}{N} \times \frac{N_{.j}}{N} = \frac{N_{i.} \times N_{.j}}{N} \quad , \quad \forall \, i,j \,.$$

A fórmula dos $\hat{E}_{ij}$ é igual à dos testes de homogeneidade.

As restrições

Foram estimadas:

- $a-1$ probabilidades marginais do factor A (a última tem de dar a soma 1); e
- $b-1$ probabilidades marginais do factor B (a última tem de dar a soma 1).

Juntamente com

- 1 restrição imposta pelo número total de observações (N),

tem-se um total de $r = (a-1)+(b-1)+1 = a+b-1$ restrições.

As restrições, logo os graus de liberdade da distribuição χ^2 são iguais aos dos testes de homogeneidade.

Testes χ^2 de independência (cont.)

Estes valores esperados estimados $\hat{E}_{ij}$ em cada uma das *ab* células serão comparados com os valores observados, O_{ij}, com base na estatística de Pearson.

NOTA: Repare-se que, embora com motivações diferentes,

- as expressões de cálculo dos $\hat{E}_{ij}$ são iguais, nos testes de homogeneidade e nos testes de independência;
- o número de restrições impostas é igual nos dois tipos de teste.

Logo, a estatística X^2 de Pearson terá uma expressão idêntica, e uma distribuição assintótica idêntica, quer nos testes de homogeneidade, quer nos testes de independência.

Mas importa não perder de vista que se trata de contextos diferentes, com hipóteses de referência diferentes e conclusões diferentes.

A estatística do teste

Quer no contexto de testes de homogeneidade, quer no contexto de testes de independência, a estatística de Pearson tem a forma

$$X^2 \quad = \quad \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(O_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}} \quad = \quad \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(O_{ij}-\frac{N_{i.}\,N_{.j}}{N}\right)^2}{\frac{N_{i.}\,N_{.j}}{N}}$$

e segue assintoticamente uma distribuição $\chi^2_{(a-1)(b-1)}$.

Em ambos os casos tem-se uma Região Crítica unilateral direita:

$$\text{Rejeita} - \text{se} \quad H_0 \text{ se} \quad X^2_{calc} > \chi^2_{\alpha;(a-1)(b-1)} \; .$$

Exemplo - Teste de independência

Exemplo: cores cabelo/olhos (teste de independencia)

Um estudo de $N = 6800$ alemães do sexo masculino analisou a cor do cabelo e a cor dos olhos de cada indivíduo. Os resultados foram:

	Cabelo				
Olhos	Louro	Castanho	Preto	Ruivo	Total
Azuis	1768	807	189	47	2811
Cinz./Verde	946	1387	746	53	3132
Castanhos	115	438	288	16	857
Total	2829	2632	1223	116	6800

Pretende-se testar se existe independência entre as características cor do cabelo e cor dos olhos (sendo natural que se rejeite esta hipótese).

Exemplo (cont.)

Exemplo: cores cabelo/olhos (teste de independencia)

As frequências marginais de linha dão estimativas das probabilidades marginais de cada cor de olhos ($\hat{\pi}_{i.} = \frac{N_{i.}}{N}$):

$$\hat{\pi}_{1.} = \frac{2811}{6800} = 0.4134 \qquad \hat{\pi}_{2.} = \frac{3132}{6800} = 0.4606 \qquad \hat{\pi}_{3.} = \frac{857}{6800} = 0.1260$$

De forma análoga se obtêm estimativas das probabilidades marginais de cores de cabelo ($\hat{\pi}_{.j} = \frac{N_{.j}}{N}$):

$$\hat{\pi}_{.1} = \frac{2829}{6800} = 0.416\,, \quad \hat{\pi}_{.2} = \frac{2632}{6800} = 0.387\,, \quad \hat{\pi}_{.3} = \frac{1223}{6800} = 0.180\,, \quad \hat{\pi}_{.4} = \frac{116}{6800} = 0.017$$

Os valores esperados estimados em cada célula, caso haja independência, são dados por:

$$\hat{E}_{ij} = N\,\hat{\pi}_{ij} = N\,\hat{\pi}_{i.}\,\hat{\pi}_{.j} = \frac{N_{i.} \times N_{.j}}{N}\,.$$

Por exemplo, $\hat{E}_{11} = \frac{2811 \times 2829}{6800} = 1169.4587$.

Exemplo (cont.)

Exemplo: cores cabelo/olhos (teste de independencia)

A tabela com os valores esperados (estimados) entre parenteses é:

Olhos	Cabelo				Total
	Louro	Castanho	Preto	Ruivo	
Azuis	1768 (1169.46)	807 (1088.02)	189 (505.57)	47 (47.95)	2811
Cin/Verde	946 (1303.00)	1387 (1212.27)	746 (563.30)	53 (53.43)	3132
Castanho	115 (356.54)	438 (331.71)	288 (154.13)	16 (14.62)	857
Total	2829	2632	1223	116	6800

A estatística de Pearson será então:

$$X^2_{calc} = \sum_{i=1}^{3}\sum_{j=1}^{4} \frac{(O_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}} = \frac{(1768-1169.46)^2}{1169.46} + \ldots + \frac{(16-14.62)^2}{14.62} = 1073.508\,.$$

A fronteira da região crítica (com $\alpha = 0.05$) é: $\chi^2_{0.05(6)} = 12.591$.
Como esperado, rejeita-se claramente a hipótese de independência.

Analisando as parcelas da estatística

Em qualquer dos contextos considerados, a região de rejeição é unilateral direita, isto é, são os valores grandes da estatística que rejeitam a hipótese nula, num teste baseado na estatística de Pearson.

No caso de rejeição de H_0, e como a estatística X^2 de Pearson é uma soma de parcelas não-negativas, é possível identificar a(s) célula(s) que contribuem com parcelas de maior valor e que são, por isso mesmo, maiormente responsáveis pela rejeição de H_0.

Ainda o exemplo de teste de independência

Exemplo: cores cabelo/olhos (teste de independencia)

As parcelas individuais da estatística de Pearson, $\frac{(O_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}$, no caso do teste de independência acima referido, são:

Olhos	Cabelo			
	Louro	Castanho	Preto	Ruivo
Azuis	306.340	72.585	198.222	0.019
Cin./Verde	97.814	25.185	59.257	0.003
Castanhos	163.630	34.059	116.263	0.130

Uma vez que $\chi^2_{0.05(6)} = 12.592$, quase todas as células (excepto as referentes aos ruivos) são, só por si, responsáveis pela rejeição de H_0, com destaque para as associações de olhos azuis com cabelo louro e de olhos azuis com cabelo preto.

Ainda o exemplo da independência (cont.)

No entanto, o sentido destas duas associações é diferente:

- para olhos azuis/cabelo louro, tem-se

$$1768 \;=\; O_{11} \;\gg\; \hat{E}_{11} \;=\; 1169.46\,.$$

  Trata-se duma associação positiva.

- para olhos azuis/cabelo preto, tem-se

$$189 \;=\; O_{13} \;\ll\; \hat{E}_{13} \;=\; 505.57\,.$$

  Trata-se duma associação negativa.

A identificação das parcelas que mais contribuem para uma rejeição de H_0 pode ajudar a identificar outras hipóteses, mais realistas, subjacentes às contagens observadas.