



Instituto Tecnológico de Aeronáutica

MOQ-13: Probabilidade e Estatística

Prof. Mauri Aparecido de Oliveira

mauri@ita.br

2019

Desafio – Semana 14

Teste Qui-quadrado para Independência

Teste Qui-quadrado

O teste Qui-quadrado é um **teste de independência** para uma tabela de contingência $r \times c$ (r – linhas, por c - colunas), onde as hipóteses são:

H_0 : A variável A é independente da variável B

H_1 : A variável A não é independente da variável B

Sendo que duas variáveis são independentes se a ocorrência de uma não afeta a ocorrência da outra.

A estatística do teste qui-quadrado mede a diferença entre as frequências esperada e a observada

$$\chi^2_{calc} = \sum_{j=1}^r \sum_{k=1}^c \frac{(o_{jk} - e_{jk})^2}{e_{jk}}$$

Se as duas variáveis forem independentes, então o_{jk} deveria estar próxima de e_{jk} , levando a uma estatística do teste qui-quadrado próxima de zero. Reciprocamente, grandes diferenças entre o_{jk} e e_{jk} levarão a uma estatística grande do teste qui-quadrado. A estatística do teste qui-quadrado não pode ser negativa (devido ao quadrado), de modo que esse será sempre um teste unilateral à direita. Se a distância do teste estiver suficientemente longe na cauda direita, rejeita-se a hipótese de independência.

Distribuição Qui-quadrado

A estatística do teste é comparada com um valor crítico da distribuição de probabilidade qui-quadrado. Ela tem um parâmetro denominado graus de liberdade. Para a tabela de contingência $r \times c$, os graus de liberdade são:

$$g.l. = \text{graus de liberdade} = (r - 1)(c - 1)$$

em que

r = o número de linhas na tabela de contingência

c = o número de colunas na tabela de contingência

Exemplo

Depois do acidente no qual o senador norte-americano John F. Kennedy Jr. morreu enquanto pilotava seu avião, à noite, de Nova York até Cape Cod, foi feita uma pesquisa por telefone aleatoriamente, na qual se perguntou a 409 nova-iorquinos: “Deve ser permitido que os pilotos particulares voem à noite sem uma habilitação para navegação por instrumentos?” A mesma pergunta foi feita a 70 especialistas em aviação. Os resultados são mostrados na Tabela a seguir. Os totais excluem aqueles que responderam “Sem opinião” (1 especialista e 25 entre o público geral).

Pilotos sem habilitação devem voar á noite por instrumentos?

Opinião	Pilotos com Experiência	Público Geral	Total da Linha
Sim	40	61	101
Não	29	323	352
Total da Linha	69	384	453

Fonte: Siena College Research Institute

As hipóteses são:

H_0 : A opinião é independente da experiência em aviação

H_1 : A opinião não é independente da experiência em aviação

Independência

		Coluna 1	Coluna 2	Total
Linha 1	Observado	40	61	101
	Esperado			
	$O - E$			
	$(O - E)^2/E$			
Linha 2	Observado	29	323	352
	Esperado			
	$O - E$			
	$(O - E)^2/E$			
Total	Observado	69	384	453
	Esperado			
	$O - E$			
	$(O - E)^2/E$			

Independência

		Coluna 1	Coluna 2	Total
Linha 1	Observado	40	61	101
	Esperado	15,38411	85,61589	101,00000
	O – E			
	$(O - E)^2/E$			
Linha 2	Observado	29	323	352
	Esperado	53,61589	298,38411	352,00000
	O – E			
	$(O - E)^2/E$			
Total	Observado	69	384	453
	Esperado	69,00000	384,00000	453,00000
	O – E			
	$(O - E)^2/E$			

Independência

		Coluna 1	Coluna 2	Total
Linha 1	Observado	40	61	101
	Esperado	15,38411	85,61589	101,00000
	O – E	24,61589	-24,61589	0,00000
	$(O - E)^2/E$	39,38753	7,07745	46,46497
Linha 2	Observado	29	323	352
	Esperado	53,61589	298,38411	352,00000
	O – E	-24,61589	24,61589	0,00000
	$(O - E)^2/E$	11,30154	2,03075	13,33228
Total	Observado	69	384	453
	Esperado	69,00000	384,00000	453,00000
	O – E	0,00000	0,00000	0,00000
	$(O - E)^2/E$	50,68906	9,10819	59,79726

Qui-quadrado = 59,797 df = 1 p-value = 1,051e-14

```
# Construir a estrutura dos dados de entrada
dados <- data.frame(especialista = c(40,29),
                    geral = c(61,323),
                    row.names = c("Sim", "Não"))

dados

qui.test <- chisq.test(dados, correct=FALSE)
qui.test

qui.test$expected
qui.test$observed

figura <- par(mfrow=c(1,2), "mar"=c(1,1,3,1))
mosaicplot(qui.test$observed, cex.axis = 1, main = "Valores Observados")
mosaicplot(qui.test$expected, cex.axis = 1,
           main = "Valores Esperados\n(Se a opinião não teve influência)")
par(figura)
```

```
> # Construir a estrutura dos dados de entrada
> dados <- data.frame(especialista = c(40,29),
+                     geral = c(61,323),
+                     row.names = c("Sim", "Não"))
>
```

```
> dados
      especialista geral
Sim             40    61
Não             29   323
>
```

```
> qui.test <- chisq.test(dados, correct=FALSE)
> qui.test
```

Pearson's Chi-squared test

```
data: dados
X-squared = 59.797, df = 1, p-value = 1.051e-14
```

```
>
> qui.test$expected
      especialista      geral
Sim      15.38411  85.61589
Não      53.61589 298.38411
> qui.test$observed
      especialista geral
Sim             40    61
Não             29   323
>
> figura <- par(mfrow=c(1,2), "mar"=c(1,1,3,1))
> mosaicplot(qui.test$observed, cex.axis =1 , main = "valores observados")
> mosaicplot(qui.test$expected, cex.axis =1 ,
+           main = "Valores Esperados\n(Se a opinião não teve influência)")
> par(figura)
```

Independência

Valores Observados	
	Sim Não
especialista	
geral	

Valores Esperados (Se a opinião não teve influência)	
	Sim Não
especialista	
geral	

As hipóteses são:

H_0 : Opinião é independente da perícia em aviação

H_1 : Opinião não é independente da expertise em aviação

Nos resultados do teste, realizado no R, mostrados na Figura anterior temos:

Graus de liberdade são $df = (r - 1) (c - 1) = (2 - 1) (2 - 1) = 1$.

Uma Tabela de Distribuição mostra que o valor crítico do Qui-quadrado para $\alpha = 0,005$ é 7,879.

Como a estatística de teste $\chi^2 = 59,797$ excede muito a 7,879, rejeitamos firmemente a hipótese.

O p-valor ($\sim 0,0000$) gera forte evidência de que a opinião não é independente da experiência na aviação.