

1. (Q49) Questão do Mood-página 479.

Um soro que supostamente tem algum efeito na prevenção de resfriados foi testado em 500 indivíduos e seus registros de um ano foram comparados com os registros de 500 indivíduos não tratados, como segue:

	Nenhum	Um	Mais de Um
Tratados	252	145	103
Não Tratados	224	136	140

Teste, ao nível de significância de 5%, se as duas populações trinomiais podem ser consideradas iguais.

Solução: Vamos completar nossa tabela:

	Nenhum	Um	Mais de Um	Total
Tratados	252	145	103	500
Não Tratados	224	136	140	500
Total	476	281	243	1000

Vamos entender nosso processo de amostragem. De 500 pessoas não tratadas contamos:

N_{11} = número de pessoas não tratadas com nenhum resfriado.

N_{12} = número de pessoas não tratadas com um resfriado.

N_{13} = número de pessoas não tratadas com mais de um resfriado.

Assim

$$(N_{11}, N_{12}, N_{13}) \sim \text{Trinomial}(500, p_{11}, p_{12}, p_{13}).$$

vspace0.5cm

N_{21} = número de pessoas tratadas com nenhum resfriado.

N_{22} = número de pessoas tratadas com um resfriado.

N_{23} = número de pessoas tratadas com mais de um resfriado.

Assim

$$(N_{21}, N_{22}, N_{23}) \sim \text{Trinomial}(500, p_{21}, p_{22}, p_{23}).$$

Queremos testar a hipótese nula:

$$H_0 : (p_{11}, p_{12}, p_{13}) = (p_{21}, p_{22}, p_{23}).$$

contra a alternativa que há pelo menos uma diferença, isto é,

$$H_0 : (p_{11}, p_{12}, p_{13}) \neq (p_{21}, p_{22}, p_{23}).$$

Se H_0 é verdade temos:

$$p_{11} = p_{21} = p_1.$$

$$p_{21} = p_{22} = p_2.$$

$$p_{31} = p_{32} = p_3.$$

Vamos estimar

$$p_i, i = 1, 2, 3.$$

Sejam

$$N_{.i} = N_{1i} + N_{2i}, i = 1, 2, 3.$$

Assim

$$\hat{p}_i = \frac{N_{.i}}{n}.$$

Vamos pensar na tabela:

	Nenhum	Um	Mais de Um	Total
Tratados	N_{11}	N_{12}	N_{13}	n_1
Não Tratados	N_{21}	N_{22}	N_{23}	n_2
Total	$N_{.1}$	$N_{.2}$	$N_{.3}$	$n = n_1 + n_2$

com n_1, n_2 fixos a priori.

Assim

$$n = n_1 + n_2 = 500 + 500 = 1000.$$

$$\hat{p}_1 = \frac{N_{.1}}{n} = \frac{476}{1000} = 0,476.$$

$$\hat{p}_2 = \frac{N_{.1}}{n} = \frac{281}{1000} = 0,281.$$

$$\hat{p}_3 = \frac{N_{.1}}{n} = \frac{243}{1000} = 0,243.$$

Sempre é bom verificar:

$$\hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 0,476 + 0,281 + 0,243 = 1.$$

Vamos calcular as frequências esperadas:

Note que

$$N_{1j} \sim \text{Bin}(n_1, p_{1j}), j = 1, 2, 3.$$

$$E(N_{1j}) = n_1 \times p_{1j}, j = 1, 2, 3$$

Note que

$$N_{2j} \sim \text{Bin}(n_2, p_{2j}), j = 1, 2, 3.$$

$$E(N_{2j}) = n_2 \times p_{2j}, j = 1, 2, 3$$

Se H_0 é verdade temos

$$N_{1j} \sim \text{Bin}(n_1, p_j), j = 1, 2, 3$$

$$N_{2j} \sim \text{Bin}(n_1, p_j), j = 1, 2, 3.$$

As frequências esperadas estimadas são dadas por:

$$e_{1j} = n_1 \times p_{est} = 500 \times (0,476, 0,281, 0,2453) = c(238, 0,140, 5121, 55).$$

$$e_{2j} = n_2 \times p_{est} = 500 \times (0,476, 0,281, 0,2453) = c(238, 0,140, 5121, 55).$$

O número de graus de liberdade para comprar r populações multinomiais com c categorias é dado por:

$$gl(r-1) \times (c-1).$$

temos $r = 2$ e $c = 3$ logo

$$gl = 1 \times 2 = 2.$$

A estatística do teste, se H_0 é verdade, é dada por:

$$Q = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1)(c-1))$$

Agora a compahe a solução detalhada da questão:

```
> #####CC0291-Estatística Não Paramétrica.
>
> ##Exercício 49-pag do Mood
>
>
>
> dados = matrix(c(252,224,145,136,103,140), nrow = 2)
> dados
      [,1] [,2] [,3]
```

```
[1,] 252 145 103
[2,] 224 136 140
>
>
> #####Nomear linhas e colunas.
>
>
> rownames(dados)=c("Trat","NTrat")
>
>
>
> colnames(dados) = c('0','1','>=2');dados
      0    1 >=2
Trat  252 145 103
NTrat 224 136 140
>
> apply(dados,1,sum) #Calculando os totais por linha.
Trat NTrat
500   500
> #####
> apply(dados,2,sum) #Calculando os totais por coluna
0    1 >=2
476 281 243
> #####
>
> dados = rbind(dados,apply(dados,2,sum));dados

#Adicionando os totais à tabela

      0    1 >=2
Trat  252 145 103
NTrat 224 136 140
      476 281 243

> dados = cbind(dados,apply(dados,1,sum)); dados
      0    1 >=2
Trat  252 145 103  500
NTrat 224 136 140  500
      476 281 243 1000
>
>
> #####Completando a tabela
>
>
> colnames(dados) = c('0','1','>=2','Total')
> rownames(dados) = c('Trat','NTrat','Total'); dados
      0    1 >=2 Total
Trat  252 145 103  500
NTrat 224 136 140  500
Total 476 281 243 1000
>
```

```
>
>
> #####Fazer o teste de homogeneidade no R diretamente:
>
>
> dad = matrix(c(252,224,145,136,103,140), nrow = 2)
> dad
      [,1] [,2] [,3]
[1,]  252  145  103
[2,]  224  136  140
>
>
> mod1 = chisq.test(dad); mod1

Pearson's Chi-squared test

data:  dad
X-squared = 7.5691, df = 2, p-value = 0.02272

>
>
> alfa = 0.05
>
> qcal = mod1$statistic;qcal
X-squared
7.56906
>
> ###Retirar a parte literal!!!!
>
> qcal = qcal[[1]]; qcal
[1] 7.56906
>
> nd = 1-pchisq(qcal,2); nd
[1] 0.02271954
>
> mod1$p.value
[1] 0.02271954
>
> nd < alfa #Rejeitar H0
[1] TRUE
>
> qtab = qchisq(1-alfa,2); qtab
[1] 5.991465
>
> names(mod1)
[1] "statistic" "parameter" "p.value"    "method"    "data.name" "observed"
[7] "expected"  "residuals" "stdres"
> e = mod1$expected; e

      [,1] [,2] [,3]
[1,]  238 140.5 121.5
```

```
[2,] 238 140.5 121.5
> o = dad; o

      [,1] [,2] [,3]
[1,] 252 145 103
[2,] 224 136 140
>
>
> #####chisq.test #####Olhar a programa
> ###Veja que cada frequ
>
> #####Vamos calcular na mão:
>
> n11=252;n21=224;n12=145;n22=136;n13=103;n32=140
>
> N.1=n11+n21;N.1
[1] 476
>
> N.2=n12+n22;N.2
[1] 281
>
> N.3=n13+n23;N.3
[1] 243
>
> n1=n11+n12+n13;n1
[1] 500
>
> n2=n21+n22+n23;n2
[1] 500
> n=n1+n2;n
[1] 1000
>
>
> ###Se H_0 é verdade temos:
>
> p1_est=N.1/n; p1_est
[1] 0.476
>
> p2_est=N.2/n; p2_est
[1] 0.281
>
> p3_est=N.3/n; p3_est
[1] 0.243
>
> p_est=c(p1_est,p2_est,p3_est);p_est
[1] 0.476 0.281 0.243
>
>
> ###Cálculo das freq. esperadas da primeira trinomial:
>
> e1=n1*p_est;e1
[1] 238.0 140.5 121.5
```

```
>
> sum(e1);n1
[1] 500
[1] 500
>
> ###Cálculo das freq. esperadas da segunda trinomial:
>
> e2=n2*p_est;e2
[1] 238.0 140.5 121.5
>
> sum(e2);n2
[1] 500
[1] 500
>
>
>
> #####Vamos formar o vetor das frq. observada e o das esperadas.
>
> e=c(e1,e2);e
[1] 238.0 140.5 121.5 238.0 140.5 121.5
> o=c(n11,n12,n13,n21,n22,n23);o
[1] 252 145 103 224 136 140
>
>
> Qcal=sum((o-e)^2/e);Qcal
[1] 7.56906
>
>
>
> #####Graus de liberdade r=2;c=3
>
> r=2;c=3
>
> gl=(r-1)*(c-1);gl
[1] 2
>
>
> ###Nível descritivo ou nível mínimo de significância
>
>
> nms=1- pchisq(Qcal,gl);nms
[1] 0.02271954
>
>
> #####Comparando com a saída do R:
>
> mod1
```

Pearson's Chi-squared test

data: dad

X-squared = 7.5691, df = 2, p-value = 0.02272

>
>
>

2. Refaça a questão usando o Teste da Razão de Verossimilhança Generalizada.

Um dos métodos mais poderosos para a construção dos testes de hipóteses é o método da razão de verossimilhança, pois ele fornece uma definição específica para a estatística teste

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}$$

e uma forma específica para a região, $\{x; \lambda(\mathbf{x}) \leq c\}$. Depois dos $X = \mathbf{x}$ serem observados $L(\theta|\mathbf{x})$ é uma função definida na variável θ .

Teorema

Seja X_1, \dots, X_n uma amostra aleatória de uma v.a. com $f(x_i; \theta)$. Sob certas condições de regularidade sobre o modelo $f(x_i; \theta)$ e se $\theta \in \Theta_0$ então a distribuição $-2 \ln \lambda(\mathbf{x})$ converge para uma distribuição de qui-quadrado quando $n \rightarrow \infty$, o grau de liberdade da distribuição é a diferença entre o número de parâmetros livres especificada por $\theta \in \Theta$ e o número de parâmetros livres especificados por $\theta \in \Theta_0$.

As condições de regularidade e existência são a compatibilidade das derivadas da função de verossimilhança com relação ao parâmetro e o suporte da distribuição que não pode depender do parâmetro.

A rejeição de $H_0: \theta \in \Theta_0$ para pequenos valores de $\lambda(\mathbf{x})$ é equivalente a rejeitar para grandes valores $-2 \ln \lambda(\mathbf{x})$. Então H_0 é rejeitada se e somente se

$$-2 \ln \lambda(\mathbf{x}) > \chi^2_{(v, \alpha)}$$

com $v = d - m$, em que d é o número de parâmetros livres em Θ enquanto m é o número de parâmetros livres em Θ_0 .

No teste de homogeneidade temos a comparação de r multinomiais com c categorias.

Para cada multinomial temos $(c - 1)$ parâmetros livres. No total temos

$$d = r \times (c - 1).$$

Se H_0 é verdade temos só uma multinomial com $m = c - 1$ parâmetros livres

Assim

$$v = d - m = r \times (c - 1) - (c - 1) = (r - 1) \times (c - 1).$$

A estatística do teste é dada por:

$$Q_2 = 2 \times \sum_{i=1}^r \sum_{j=1}^c O_{ij} \log(O_{ij}/E_{ij})$$

Veja a solução pelo software R


```
> #####Comparando com a saída do R:  
>  
> mod1
```

Pearson's Chi-squared test

```
data: dad  
X-squared = 7.5691, df = 2, p-value = 0.02272
```

```
>  
>  
> D=2*sum(o*log(o/e));D  
[1] 7.592034  
> Dcal=2*sum(o*log(o/e));Dcal;Qcal  
[1] 7.592034  
[1] 7.56906  
>  
nd=1-pchisq(Q2cal,2);nd  
[1] 0.02246006
```