Exemplo Teste Qui-Quadrado de Independência ex 5, Seção 8.5 Magalhães e Lima

Em um experimento para verificar se existe relação entre crise de asma e incidência de gripe, 150 crianças foram escolhidas ao acaso dentre aquelas acompanhadas pelo posto de saúde do bairro. Os dados referentes a uma particular semana encontram-se na tabela a seguir

Asma \ Gripe	Sim	Não	
Sim	27	34	61
Não	42	47	89
	69	81	150

Teste a hipótese de que a ocorrência de gripe e asma são independentes, ao nível $\alpha = 4\%$.

H_0: Presença de Asma e de Gripe são independentes

H_a: Presença de Asma e gripe não são independentes

$$e_{11} = \frac{69 \times 61}{150} = 28{,}06 \qquad \underbrace{e_{21} = 40{,}94 \quad e_{12} = 32{,}94 \quad e_{22} = 48{,}06}_{\text{poderiam ser obtidos por diferença}}$$

$69 - 28{,}06$

$$Q^2 = \frac{(27-28{,}06)^2}{28{,}06} + \frac{(34-32{,}94)^2}{32{,}94} +$$

$$+ \frac{(42-40{,}94)^2}{40{,}94} + \frac{(47-48{,}06)^2}{48{,}06} = 0{,}1112$$

$1\ gl = (2-1)(2-1)$ $\quad \alpha = 0{,}04$ $\quad a = 4{,}218$

RC : $Q^2 \geq 4{,}218$

$Q^2_{obs} < 4{,}218$

Não rejeitamos Ho. Os dados sugerem independência entre as variáveis.

Teste Qui-Quadrado de Homogeneidade

Dispomos de r populações e deseja-se verificar se uma particular variável categórica se comporta de forma homogênea nessas populações. Tomadas amostras de tamanhos n_1, n_2 ... n_r, os dados são dispostos em uma tabela de dupla entrada da forma a seguir.

Categorias da Variável

Populações	1	2	...	s	
1	o_{11}	o_{12}		o_{1s}	n_1
2	o_{21}	o_{22}		o_{2s}	n_2
⋮					
r	o_{r1}	o_{r2}		o_{rs}	n_r

H_0: A probabilidade de cada uma das s categorias é a mesma para as r populações

H_a: Pelo menos uma das populações apresenta uma distribuição de probabilidades diferente.

Estatística do teste:

$$Q^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$e_{ij} = \frac{\text{total da linha } i \times \text{total da coluna } j}{n}$$

$$n = \sum_{i=1}^{r} n_i$$

Justificativa:

Sob a hipótese de igualdade das probabilidades das categorias para as r populações, a probabilidade da j-ésima categoria seria estimada por

$$\frac{\text{total da coluna } j}{n}$$

$\therefore$ Sob H_0, espera-se na categoria ij a frequência

$$e_{ij} = n_{i} \cdot \frac{\text{total da coluna } j}{n} =$$

$$\frac{\text{total de linha } i \times \text{total da coluna } j}{n}$$

Rejeita-se H_0 para $Q^2 \geq a$ onde a é tal que $P(\chi^2_{(r-1)(s-1)} \geq a) = \alpha$

Distribuição Qui-Quadrado com $(r-1)(s-1)$ g.l.

Exemplo

Em uma faculdade, deseja-se comparar o desempenho desportivo dos cursos de Administração e Economia. Com esse objetivo foram colhidas amostras de 180 e 150 alunos de cada curso respectivamente, que forneceu os seguintes dados

Curso \ Desempenho	Bom	Regular	Ruim	
administração	65	70	45	180
Economia	27	103	20	150
	92	173	65	330

Teste as hipóteses de interesse ao nível $\alpha = 0,05$.

H_0: A probabilidade de cada uma das categorias de desempenho é a mesma para os dois cursos

H_a: Não há homogeneidade entre os cursos com relação ao desempenho esportivo (a probabilidade de cada uma das categorias não é a mesma para os dois cursos)

$$e_{11} = \frac{92 \cdot 180}{330} = 50,18 \qquad e_{12} = \frac{173 \cdot 180}{330} = 94,36$$

$$e_{13} = \frac{65 \cdot 180}{330} = 35,45$$ Poderia ser obtido por diferença: $180 - 94,36 - 50,18$ *

$$e_{21} = \frac{92 \cdot 150}{330} = 41,81 \; * \qquad e_{22} = \frac{173 \cdot 150}{330} = 78,63 \; *$$

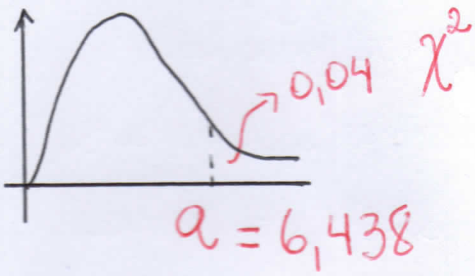
$$e_{23} = \frac{65 \cdot 150}{330}$$ Poderia ser obtido por diferença *

$$Q^2 = \frac{(65 - 50,18)^2}{50,18} + \frac{(70 - 94,36)^2}{94,36} + \frac{(45 - 35,45)^2}{35,45} +$$

$$+ \frac{(27 - 41,81)^2}{41,81} + \frac{(103 - 78,63)^2}{78,63} + \frac{(20 - 29,54)^2}{29,54}$$

$$= 29,12$$

Adotando $\alpha = 0,04$, $(r-1)(s-1)\,gl = (3-1)(2-1) = 2$



$RC: Q^2 \geq 6,438$

$Q^2_{obs} = 29,12 \in RC$

* Poderiam ser obtidos por diferença uma vez conhecidos os totais de linha e de coluna

B	Reg	R	
50,18 *	94,36*	*	180
*	*	*	150
92	173	65	

* 2 graus de liberdade

Conclusão: Rejeitamos H_0. Os dados sugerem que a probabilidade das categorias de desempenho não é a mesma para os dois cursos.

Continuação

Construir um intervalo com 96% de confiança para a diferença de proporções de bom desempenho entre alunos dos cursos de Administração e Economia

p_1 - proporção de bom desempenho no curso de Administração

p_2 - proporção de bom desempenho no curso de Economia

$\hat{p}_1 = \frac{65}{180} = 0{,}3611$ $\hat{p}_2 = \frac{27}{150} = 0{,}18$

$$IC: \hat{p}_1 - \hat{p}_2 \pm z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0{,}1811 \pm 1{,}96 \sqrt{\frac{0{,}3611(1-0{,}3611)}{180} + \frac{0{,}18(1-0{,}18)}{150}}$$

$$0{,}1811 \pm 0{,}0932$$

$$[0{,}0868 \, ; \, 0{,}2743]$$

4 - Tópicos Especiais

Teste de Comparação de duas médias

Figura pag 311

- Dados Pareados *
- Amostras Independentes
  - σ_x^2, σ_y^2 conhecidos
  - Desconhecidos Iguais *
  - Desiguais

a) Comparações de duas médias com base em amostras dependentes (dados pareados)