Exemplo - Teste de Aderência

Acredita-se que um dado é viciado, de modo que a probabilidade de face par é o dobro da probabilidade de face ímpar. Em 450 lançamentos desse dado observou-se

Face	1	2	3	4	5	6
Frequência	50	90	50	100	80	80

Teste ao nível de significância de 0,01 a hipótese de que a probabilidade de face par é o dobro da de face ímpar.

H_0: $P(1) = P(3) = P(5) = 1/9 \quad P(2) = P(4) = P(6) = \frac{2}{9}$

H_a: Pelo menos uma ($\Rightarrow$ duas) das probabilidades acima não se verifica

Frequências esperadas sob H_0:

$e_1 = e_3 = e_5 = 450 \cdot \frac{1}{9}$

$e_2 = e_4 = e_6 = 450 \cdot \frac{2}{9} = 100$

2

$$Q^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \frac{(50-50)^2}{50} + \frac{(90-100)^2}{100} +$$

$$+ \frac{(50-50)^2}{50} + \frac{(100-100)^2}{100} + \frac{(80-50)^2}{50} + \frac{(80-100)^2}{100}$$

$$= 0 + \frac{100}{100} + 0 + 0 + \frac{900}{50} + \frac{400}{100} = 23$$

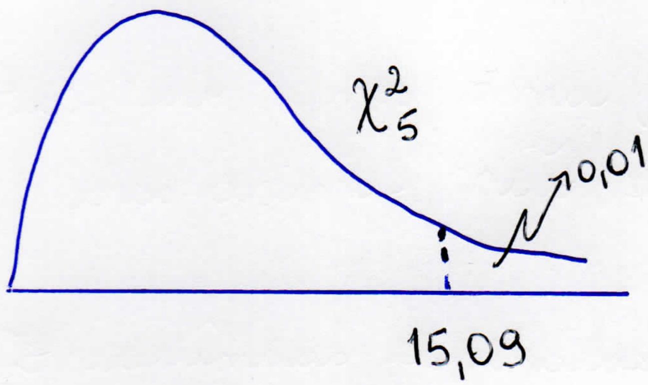
Fixando $\alpha = 0{,}01$ $\quad k = 6$

$k - 1 = 6 - 1 = 5$ gl

p	1,0%
5	15,09

Distribuição Qui-Quadrado

$P(\chi^2_5 \geq 15{,}09) = 0{,}01$



RC: $Q^2 \geq 15{,}09$ $\qquad Q^2_{obs} \in 23 \in RC$

Rejeitamos H_0. Os dados sugerem que a probabilidade de face par não é o dobro da prob. de face ímpar.

Exemplo 8.9 pág 285 Magalhães e Lima

X - número de impactos anteriores à falha em um equipamento eletrônico

Foi obtida uma amostra de 80 ensaios, cada ensaio correspondendo aos testes feitos até a interrupção por falha no equipamento, resultando em 80 observações da variável X.

Acredita-se que X tem distribuição geométrica com parâmetro $p = 0,4$.

Distribuição Geométrica

X - nº de falhas antes do primeiro sucesso

p - probabilidade de sucesso

$q = 1 - p$ - probabilidade de fracasso

$$P(X = x) = q^x p$$

No exemplo: impacto - fracasso

falha - sucesso

$H_0: P(X=x) = 0{,}6^x \cdot 0{,}4 \qquad x = 0, 1, 2, 3, \ldots$

H_a: X tem outra distribuição

Teste as hipóteses de interesse sabendo que as frequências observadas para essa amostra foram

nº de impactos X	0	1	2	3	4	mais de 4
freq. obs. o_i	30	26	10	5	5	4
freq. esp. e_i	32	19,2	11,5	6,9	4,1	6,3

juntar

Cálculo das frequências esperadas sob H_0

$P(X=0) = 0{,}6^0 \cdot 0{,}4 = 0{,}4 = p_1$

$P(X=1) = 0{,}6^1 \cdot 0{,}4 = 0{,}24 = p_2$

$P(X=2) = 0{,}6^2 \cdot 0{,}4 = 0{,}144 = p_3$

$P(X=3) = 0{,}6^3 \cdot 0{,}4 = 0{,}0864 = p_4$

$P(X=4) = 0{,}6^4 \cdot 0{,}4 = 0{,}05184 = p_5$

$$P(X>4) = 1 - \sum_{l=0}^{4} P(X=l) = 0{,}07776 = p_6$$

p_i - probabilidade da categoria i

e_1 - frequência esperada da categoria 1

$= n p_1 = 80 . 0{,}4 = 32$

$e_2 = 80 . 0{,}24 = 19{,}2$ $\quad$ $e_3 = 80 . 0{,}144 = 11{,}5$

$e_4 = 80 . 0{,}0864 = 6{,}9$ $\quad$ $e_5 = 80 . 0{,}05184 = 4{,}1$

$e_6 = 80 . 0{,}07776 = 6{,}3$

$$Q^2_{obs} = \frac{(30-32)^2}{32} + \frac{(26-19{,}2)^2}{19{,}2} + \frac{(10-11{,}5)^2}{11{,}5} +$$

$$+ \frac{(5-6{,}9)^2}{6{,}9} + \frac{(9-10{,}4)^2}{10{,}4} = 3{,}44$$

Adotando $\alpha = 0{,}05$ $\quad$ $k = 5$ $\quad$ $gl = 4$

5%

4 - - - - 9,49

RC: $Q^2 \geq 9{,}49$

$Q^2_{obs} = 3{,}44 < 9{,}49$ $\quad$ $Q^2_{obs} \notin RC$. Não rejeitamos H_0.

Concluimos pela aceitação do modelo proposto dist. Geométrica com $p = 0{,}4$.

Obs: Quando os parâmetros da distribuição de probabilidades especificada em H_0 são desconhecidos, devem ser estimados através da amostra. Nesse caso, o número de graus de liberdade da distribuição Qui-quadrado a ser utilizada é

$$K - 1 - p$$

p - número de parâmetros estimados.

Referência: Exemplo 8.14 pag 287
Magalhães e Lima

Teste Qui-Quadrado de Independência

Objetivo: Verificar se duas variáveis (qualitativas ou quantitativas categorizadas), medidas na mesma unidade amostral são independentes.

Essas variáveis possuem r e s categorias respectivamente.

As hipóteses para o problema são

H_0: As variáveis são independentes

H_a: Existe uma relação entre as variáveis
(as variáveis são dependentes)

Tomada uma amostra de tamanho n, classifica-se cada elemento segundo as duas variáveis.

Var.1 / Var.2	1	2	...	s	
1	o_{11}	o_{12}		o_{1s}	
2	o_{21}	o_{22}		o_{2s}	
$\vdots$					
r	o_{r1}	o_{r2}		o_{rs}	
					n

Estatística de teste:

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

e_{ij} - frequência esperada da casela (i, j) sob a hipótese de independência

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

$i = 1, 2 \ldots r$

$j = 1, 2 \ldots s$

Justificativa:

Se as variáveis 1 e 2 são independentes, então p_{ij} = Probabilidade da casela (i, j)

$p_{ij} = P(\text{categoria } i \text{ var } 2)\, P(\text{categoria } j \text{ var } 1)$

∴ Sob a hipótese de independência, espera-se na categoria ij a frequência

$$e_{ij} = n \underbrace{\frac{\text{total da linha } i}{n} \; \frac{\text{total da coluna } j}{n}}_{\hat{p}_{ij} \text{ sob a hipótese de independência}}$$

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 $(r-1) \times (s-1)$ graus de liberdade.

Exemplo 8.11 Magalhães e Lima

Uma amostra de 528 estudantes foi classificada segundo seu desempenho em provas de Física e Matemática. Deseja-se testar se existe dependência entre as notas nas duas disciplinas, que foram categorizadas em alta, média e baixa.

Física \ Matemática	Alta	Média	Baixa	Total
alta	56	71	12	139
média	47	163	38	248
Baixa	14	42	85	141
Total	117	276	135	528

H_0: As notas de Física e Matemática são independentes

H_a: Existe uma relação entre essas notas

Frequências esperadas

$$e_{11} = \frac{139 \times 117}{528} = 30,80 \qquad e_{12} = \frac{139 \times 276}{528} = 72,66$$

$$e_{13} = \frac{139 \times 135}{528} = 35,54 \quad \ldots \quad e_{33} = \frac{141 \times 135}{528} = 36,05$$

Tabela das Frequências Esperadas

F \ M	Alta	Média	Baixa
Alta	30,80	72,66	35,54
Média	54,95	129,64	63,41
Baixa	31,25	73,70	36,05

$\alpha = 0{,}01 \qquad r = 3 \qquad s = 3$

RC: $Q^2 \geqslant a \qquad P(\chi^2_4 \geqslant a) = 0{,}01$

$a = 13{,}28$

RC: $Q^2 \geqslant 13{,}28$

$$Q^2_{obs} = \frac{(56-30{,}80)^2}{30{,}80} + \frac{(71-72{,}66)^2}{72{,}66} + \frac{(12-35{,}54)^2}{35{,}54}$$

$$+ \dots + \frac{(85-36{,}05)^2}{36{,}05} = 145{,}78 > a$$

Rejeitamos H_0.

Os dados sugerem que existe dependência entre as notas de Física e Matemática