

Tabela de frequências

Sejam X, Y variáveis aleatórias cujos valores observados são B_1, B_2, \dots, B_l e A_1, A_2, A_k , respectivamente.

Observam-se os seguintes dados

ind.	X	Y
1	B_2	A_1
2	B_7	A_3
\vdots	\vdots	\vdots
n	B_1	A_5

Colocamos nossos dados numa tabela de frequências absolutas observadas

$X \setminus Y$	A_1	A_2	\dots	A_k	Total X
B_1	O_{11}	O_{12}	\dots	O_{1k}	$O_{1\cdot}$
B_2	O_{21}	O_{22}	\dots	O_{2k}	$O_{2\cdot}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
B_l	O_{l1}	O_{l2}	\dots	O_{lk}	$O_{l\cdot}$
Total Y	$O_{\cdot 1}$	$O_{\cdot 2}$	\dots	$O_{\cdot k}$	n

Temos nossa tabela de frequências esperadas **sob** H_0 (Independência)

$X \setminus Y$	A_1	A_2	\dots	A_k	Total X
B_1	E_{11}	E_{12}	\dots	E_{1k}	$O_{1\cdot}$
B_2	E_{21}	E_{22}	\dots	E_{2k}	$O_{2\cdot}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
B_l	E_{l1}	E_{l2}	\dots	E_{lk}	$O_{l\cdot}$
Total Y	$O_{\cdot 1}$	$O_{\cdot 2}$	\dots	$O_{\cdot k}$	n

Em que

$$E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}$$

Note que, sob independência

$$P(B_i \cap A_j) = P(B_i) \cdot P(A_j)$$

$$E_{ij} = n \cdot P(B_i \cap A_j) \stackrel{\text{ind.}}{=} nP(B_i) \cdot P_{A_j}$$

Estimando $P(B_i), P(A_j)$ temos

$$\widehat{P(B_i)} = \frac{O_{i.}}{n}, \widehat{P(A_j)} = \frac{O_{.j}}{n}$$

Logo, o valor esperado estimado é

$$\widehat{E_{ij}} = n \cdot \widehat{P(B_i)} \cdot \widehat{P(A_j)} = \frac{O_{i.} \cdot O_{.j}}{n}$$

Em ambos testes, usaremos a seguinte estatística para testar suas hipóteses (independência e homogeneidade)

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sob H_0 , ou seja,

$$\chi_{obs}^2 \sim \chi_{(k-1)(l-1)}^2$$

Dessa forma, rejeitamos a hipótese H_0 a α graus de liberdade se

$$\chi_{obs}^2 > c_p$$

em que c_p satisfaz $P(\chi_{(k-1)(l-1)}^2 > c_p) = \alpha$.

Observação

Essa aproximação com a χ^2 só funciona de modo razoável quando cada $E_{ij} > 5$

Teste de Homogeneidade

Usamos esse teste para verificar se as medidas de probabilidade de vários grupos diferentes são iguais (seguem uma mesma distribuição).

Os totais marginais para cada grupo devem ser fixados antes de executarmos o experimento.

$$\begin{cases} H_0 : \text{Os grupos são independentes} \\ H_1 : \text{Pelo menos um dos grupos não é independente} \end{cases}$$

Teste de independência

Usamos esse teste para verificar se os eventos são independentes.

Aqui, apenas o tamanho amostral (total dos totais) é fixado.

$$\begin{cases} H_0 : \text{Os grupos se distribuem de forma equivalente} \\ H_1 : \text{Pelo menos um dos grupos não se distribui de forma equivalente} \end{cases}$$

Exemplos

Primeiro exemplo (homogeniedade)

510 segurados foram amostrados, sendo 200 de São Paulo, 100 do Ceará e 210 de Pernambuco. O objetivo é verificar se o número de acidentes se distribui igualmente entre os estados.

Indivíduos	Estado	Sinistralidade
1	SP	1
\vdots	\vdots	\vdots
200	SP	0
1	CE	1
\vdots	\vdots	\vdots
100	CE	0
1	PE	1
\vdots	\vdots	\vdots
210	PE	0

Tabela Observada

Estado	Sinistralidade		Total
	1	0	
SP	60	140	200
CE	10	90	100
PE	50	160	210
Total	120	390	510

Tabela esperada

Estado	Sinistralidade		Total
	1	0	
SP	47	153	200
CE	24	76	100
PE	49	161	210
Total	120	390	510

Temos nossa estatística qui-quadrado

$$\chi_{obs}^2 = \frac{(60 - 47)^2}{47} + \frac{(140 - 153)^2}{153} + \frac{(10 - 24)^2}{24} + \frac{(90 - 76)^2}{76} + \frac{(50 - 49)^2}{49} + \frac{(160 - 161)^2}{161} = 15.47$$

Concluiremos o teste tomando $\alpha = 1\%$ de significância estatística. Sabemos que, sob H_0 ,

$$\chi_{obs}^2 \sim \chi_{(3-1)(2-1)}^2$$

Logo, devemos encontrar c_p tal que

$$P(\chi_2^2 > c_p) = 1\%$$

Pela tabela, $c_p = 9.21$

Como $15.47 > 9.21$, concluímos que, a sinistralidade não se distribui de forma homogênea entre os estados de SP, CE e PE a 1% de significância.

Outro Exemplo (independência)

Temos nossa tabela de valores observados:

Opinião	1ª Tent	2ª Tent	3ª Tent	Total
Excelente	62	36	12	110
Satisfatório	84	42	14	140
Insatisfatório	24	22	24	70
Total	170	100	50	320

Nossa tabela de valores esperados (arredondados):

Opinião	1ª Tent	2ª Tent	3ª Tent	Total
Excelente	58	34	17	110
Satisfatório	74	44	22	140
Insatisfatório	37	22	11	70
Total	170	100	50	320

$$\chi_{obs}^2 = \frac{(62 - 58)^2}{58} + \frac{(36 - 34)^2}{34} + \frac{(12 - 17)^2}{17} + \frac{(84 - 74)^2}{74} + \frac{(42 - 44)^2}{44} + \frac{(24 - 22)^2}{22} + \frac{(24 - 37)^2}{37} + \frac{(22 - 22)^2}{22} + \frac{(24 - 11)^2}{11} = 26.14$$

Considerando $\alpha = 5\%$, precisamos encontrar c_p tal que

$$P(\chi_4^2 > c_p) = 5\% \Rightarrow c_p = 9.49$$

Como $26.14 > 9.49$, podemos concluir que, a 5% de significância estatística, existem evidências que o número da tentativa tem

influência sobre a opinião do cliente.