

# **UNIVERSIDADE FEDERAL DA PARAÍBA**

## **Teste de Independência**

Luiz Medeiros de Araujo Lima Filho  
Departamento de Estatística

# Introdução

Um dos principais objetivos de se construir uma tabela de contingência, com o objetivo de se analisar a distribuição conjunta de duas variáveis qualitativas, é descrever a associação entre elas.

Ou seja, de certo modo esperamos que haja uma certa dependência entre as variáveis, por exemplo, sexo e ramo de atividade. Desta forma, nosso foco será buscar evidência estatística de que duas variáveis possuem certo grau de associação.

Ao fazer esse tipo de investigação em busca de evidência estatística, estamos realizando um Teste de Hipóteses. Estamos buscando o confronto entre duas hipóteses: Existência de associação contra Inexistência de associação.

**Exemplo:** Suponha que desejamos verificar se existe associação entre as variáveis tipo de cooperativa e estado, como dado na tabela a seguir.

Estado	Tipo de cooperativa				Total
	Consumidor	Produtor	Escola	Outras	
São Paulo	214(33%)	237(37%)	78(12%)	119(18%)	648(100%)
Paraná	51(17%)	102(34%)	126(42%)	22(7%)	301(100%)
Rio Grande do Sul	111(18%)	304(51%)	139(23%)	48(8%)	602(100%)
<b>Total</b>	376(24%)	643(42%)	343(22%)	189(12%)	1551(100%)

- Neste caso, um teste Qui-Quadrado pode ser usado para determinar se as duas variáveis (gênero e desempenho profissional, por exemplo) são independentes. Duas variáveis são independentes se a ocorrência de uma não afeta a ocorrência da outra.

# Teste Qui-Quadrado de independência

- O teste de independência Qui-Quadrado é usado para descobrir se existe uma associação entre a variável da linha e a variável da coluna em uma tabela de contingência construído à partir de dados da amostra.
- Para realização do teste, se faz necessário calcular o valor esperado de cada célula. Supondo-se que as variáveis sejam independentes, o valor esperado de cada célula será:

$$E_{r,c} = \frac{(\text{total na linha})(\text{total na coluna})}{\text{total na amostra}}$$

- $E_{1,1} = (648)(376)/1551 = 157,09$
- $E_{1,2} = (648)(643)/1551 = 268,64$

Podemos calcular todos os outros valores de forma similar.

# Teste Qui-quadrado de independência

- Utilizaremos uma medida global para verificar se existe associação entre as variáveis. Esta medida será dada através do afastamento global entre valores observados e valores esperados.
- Esta medida é chamada de  $\chi^2$  de Pearson (Qui-quadrado de Pearson) e sua estatística de teste é dada pela expressão:

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

em que  $O_{ij}$  e  $E_{ij}$  são, respectivamente, as frequências observadas e esperadas da  $i$ -ésima linha e  $j$ -ésima coluna. Se a hipótese de independência (não-associação) for verdadeira, o valor da estatística de teste será próximo de zero.

# Importante

Para validação do teste, se faz necessário que sejam respeitados alguns critérios:

- Os dados serem selecionados aleatoriamente.
- Todas as frequências esperadas sejam maiores ou igual a 1.
- Não mais de 20% das frequências esperadas sejam inferiores a 5.

Obs: O teste está baseado na comparação entre duas hipóteses, denominadas, respectivamente de, hipótese nula e hipótese alternativa. A hipótese nula é de que as variáveis não estão associadas, em outras palavras, eles são independentes. A hipótese alternativa é de que as variáveis estão associadas, ou dependentes.

# Etapas do Teste

Etapa 1: Definição das hipóteses

$H_0$ : As variáveis são independentes.

$H_1$ : As variáveis não são independentes.

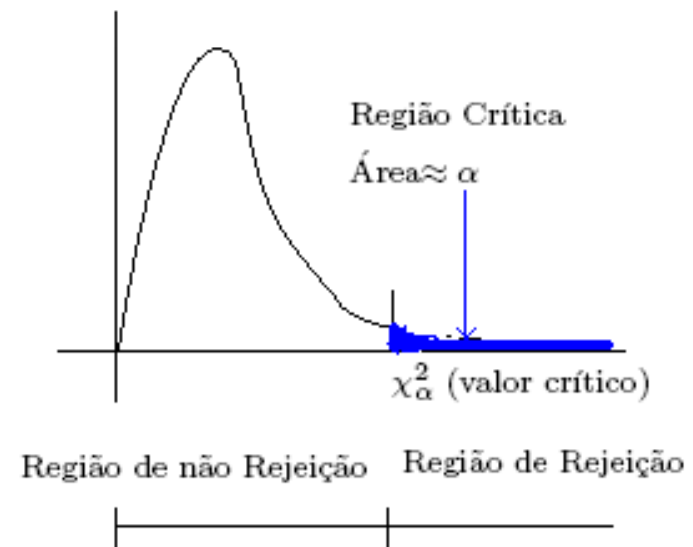
Etapa 2: Estabelecer o nível de significância ( $\alpha$ )  
(Definida pelo pesquisador)

Etapa 3: Determinar a distribuição amostral

$$\chi^2_{[\alpha; (r-1)(c-1)]}$$

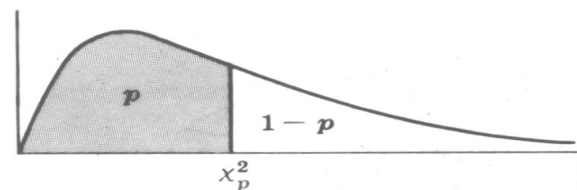
Etapa 4: Determinar o valor crítico  
(Tabela qui-quadrado)

Etapa 5: Determinar a região de rejeição (Ver gráfico)



## APÊNDICE E

Valores dos Percentis ( $\chi_p^2$ )  
para a Distribuição Qui-Quadrado  
com  $\nu$  graus de liberdade



$\nu$	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.25}$	$\chi^2_{.50}$	$\chi^2_{.75}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$	$\chi^2_{.999}$
1	,0000	,0002	,0010	,0039	,0158	,102	,455	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	,0100	,0201	,0506	,103	,211	,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	,0717	,115	,216	,352	,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	,207	,297	,484	,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	,412	,554	,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	,676	,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	29,6
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3	32,9
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8	34,5
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8	37,7
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3	39,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2	42,3
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0	45,3
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8	48,3
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2	49,7
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6	51,2
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6	55,5
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0	56,9
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7	59,7
40	20,7	22,2	24,4	26,5	29,1	33,7	39,3	45,6	51,8	55,8	59,3	63,7	66,8	73,4
50	28,0	29,7	32,4	34,8	37,7	42,9	49,3	56,3	63,2	67,5	71,4	76,2	79,5	86,7
60	35,5	37,5	40,5	43,2	46,5	52,3	59,3	67,0	74,4	79,1	83,3	88,4	92,0	99,6
70	43,3	45,4	48,8	51,7	55,3	61,7	69,3	77,6	85,5	90,5	95,0	100	104	112
80	51,2	53,5	57,2	60,4	64,3	71,1	79,3	88,1	96,6	102	107	112	116	125
90	59,2	61,8	65,6	69,1	73,3	80,6	89,3	98,6	108	113	118	124	128	137
100	67,3	70,1	74,2	77,9	82,4	90,1	99,3	109	118	124	130	136	140	149

FONTE: PEARSON E.S. e HARTLEY H.O. *Biometrika Tables for Statisticians*,  
Vol. 1 (1966), Tábua 8, págs. 137 e 138. Com permissão.



# Etapas do Teste

Etapa 6: Calcular a estatística do teste (Valor )

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

Etapa 7: Tomada de decisão. Verificar se a estatística do teste cai na região de rejeição ou não.

Etapa 8: Interpretação do teste

**Exemplo 1:** Verificar se existe associação entre as variáveis tipo de cooperativa e estado. Considere  $\alpha=0,1$ .

Estado	Tipo de cooperativa				Total
	Consumidor	Produtor	Escola	Outras	
São Paulo	214(33%)	237(37%)	78(12%)	119(18%)	648(100%)
Paraná	51(17%)	102(34%)	126(42%)	22(7%)	301(100%)
Rio Grande do Sul	111(18%)	304(51%)	139(23%)	48(8%)	602(100%)
<b>Total</b>	376(24%)	643(42%)	343(22%)	189(12%)	1551(100%)

**Exemplo 2:** Na tabela a seguir, observa-se o gênero e a avaliação do desempenho profissional de 220 estatísticos. Teste a alegação de que o gênero e o desempenho profissional são independentes. Use  $\alpha = 0,05$ .

	<b>Baixo</b>	<b>Médio</b>	<b>Superior</b>	<b>Total</b>
<b>Homem</b>	22	81	9	112
<b>Mulher</b>	14	75	19	108
<b>Total</b>	36	156	28	220

**Exemplo 3:** Verificar se existe associação entre as variáveis gênero e fumante. Considere  $\alpha=1\%$ .

empregados	gênero		total
	masculino	feminino	
fumantes	40	10	50
não fumantes	80	70	150
total	120	80	200

**Exemplo 4:** Verificar se existe associação entre as variáveis sexo e tipo de consumo de álcool. Considere  $\alpha = 0,1$ .

Tabela 1 - Tipo de consumo de álcool atual entre idosos por sexo e de acordo com o número de copos

Sexo	Consumo de Álcool							
	Sensato		Arriscado		Abusivo		Total	
	f	%	f	%	f	%	f	%
Feminino	19	90.4	1	4.8	1	4.8	21	100
Masculino	8	28.6	12	42.8	8	28.6	28	100
Total	27	55.1	13	26.5	9	18.4	49	100

Fonte: CDD e MAST-G  $\chi^2=18,61$ ;  $p < ,001$