Teste de Hipótese:
Teste Qui-Quadrado

Prof. Fermín Alfredo Tang Montané

Teste Qui-Quadrado
Descrição

É um teste não paramétrico, pois independe dos parâmetros populacionais (média, variância, etc.);

É utilizado quando se deseja comparar frequências observadas com frequências esperadas, estas baseadas em distribuições de probabilidades conhecidas.

O método é aplicado a dois tipos de testes:

1º) Teste de aderência ou teste de bondade de ajustamento;
2º) Teste de independência de duas variáveis aleatórias.

Teste de Aderência
Descrição

É utilizado para testar a natureza de uma distribuição de probabilidade amostral, ou seja, se os dados da amostra aderem a uma determinada distribuição de probabilidade (Binomial, Poisson, Hipergeométrica, Normal, etc.).
Compara frequências observadas às frequências esperadas para uma distribuição de probabilidade proposta. Para isso utiliza a seguinte estatística:

$$\chi^2_{\text{calc}} = \sum_{i=1}^{k} \frac{(\text{fo}_{\text{i}} - \text{fe}_{\text{i}})^2}{\text{fe}_{\text{i}}}$$

Onde:
k: é o número de observações o classes;
fo_i : é a frequência observada da classe i;
fe_i : é a frequência esperada da classe i;

$\chi^2{}_{calc}$ tem aproximadamente distribuição qui-quadrado com (k-1-m) graus de liberdade (m é o número de parâmetros estimados se houver).
O valor do $\chi^2{}_{calc}$ será menor quanto mais próximos sejam os valores observados dos valores esperados.

Teste de Aderência
Procedimento

1°) Definir as hipóteses nula e alternativa:

H_o: os dados da amostra **aderem** à distribuição proposta

H_1: os dados da amostra **não aderem** à distribuição

2°) Fixar o nível de significância (α);

3°) Determinar a região de aceitação (RA) de H_o através do valor crítico (abscissa) da distribuição qui-quadrado, $\chi^2_{\text{crit}} = \chi^2_{\varphi}(\alpha)$, para φ=(k-1-m) graus de liberdade e nível de significância α; onde k é o número de classes ou observações e m é o número de parâmetros estimados a partir da amostra.

4°) Calcular as frequências esperadas fe_i utilizando a distribuição proposta (Probabilidade x Numero Total de Observações).

5°) Determinar o valor da estatística:

$$\chi^2_{\text{calc}} = \sum_{i=1}^{k} \frac{\left(\text{fo}_i - \text{fe}_i\right)^2}{\text{fe}_i}$$

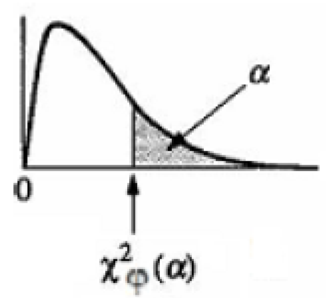
6°) Se $\chi^2_{\text{calc}} \leq \chi^2_{\text{crit}}$ aceita-se H_o, caso contrário rejeita-se H_o.

Teste de Aderência
Procedimento

A hipótese nula será aceita se: $\chi^2_{\text{calc}} \le \chi^2_{\text{crit}}$

Sendo que, $\chi^2_{\text{crit}} = \chi^2_{\varphi}(\alpha)$

corresponde ao valor crítico (abscissa) da distribuição qui-quadrado com φ graus de liberdade e um nível de significância α.



Teste de Aderência

Exemplo

Um livro foi impresso com 1000páginas. Acredita-se que o número de erros por página observado no livro, apresentado na tabela abaixo, estão distribuídos segundo a distribuição de Poisson. Utilize o nível de significância de 1% para avaliar esta hipótese.

Número de erros por página	0	1	2	3	4
Número de páginas observado (f_{oi})	500	340	120	30	10
Número de páginas esperado (f_{eij})	492	349	124	29	5

Teste de Aderência
Exemplo

Definem-se as hipóteses nula e alternativa:

H_o: as frequências observadas distribuem-se segundo Poisson
H_1: as frequências observadas não distribuem-se segundo Poisson

Calcula-se a estatística: $\chi^2_{\text{calc}} = \sum_{i=1}^{k} \frac{(\text{fo}_\text{i} - \text{fe}_\text{i})^2}{\text{fe}_\text{i}}$

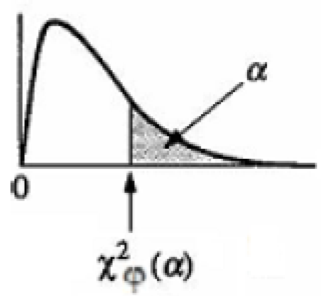
$$\chi^2_{\text{calc}} = \frac{(500-492)^2}{492} + \frac{(340-349)^2}{349} + \frac{(120-124)^2}{124} + \frac{(30-29)^2}{29} + \frac{(10-5)^2}{5} = 5{,}53$$

Calcula-se o valor crítico para a distribuição qui-quadrado: $\chi^2_{\text{crit}} = \chi^2_{\varphi}(\alpha)$

Onde:
O número de graus de liberdade φ corresponde ao número de observações ou classes menos um, (k-1)=4 (Não foram estimados parâmetros).
O nível de significância α corresponde a 1%, ou seja, 0,01.
O valor crítico é obtido a partir da tabela da distribuição qui-quadrado.

Teste de Aderência
Exemplo

Resposta: O valor críticos é obtido da tabela qui-quadrado:



$$\chi^2_{\text{crit}} = \chi^2_\varphi(\alpha) =$$

$$\chi^2_4(0,01) = 13,28$$

Como: $\chi^2_{\text{calc}} \leq \chi^2_{\text{crit}}$

$$5,53 \leq 13,28$$

φ	α = 0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.015	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.050	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	$\boxed{13.28}$	14.86
5	0.412	0.554	0.831	1.145	1.61	2,67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00

Logo, aceita-se a hipótese nula, ou seja, ao nível de significância de 1%, o número de erros por páginas do livro estão distribuídos segundo a distribuição de Poisson.

Teste de Aderência
Exemplo 2

- Utilize o teste χ^2 com nível de significância de 5% ou α=0,05 para testar se a distribuição abaixo, obtida em uma amostragem, é uma distribuição de Poisson, com média estimada de λ=2 chegadas/minuto.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

$$P(n=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Nº Chegadas /Min	Freq. Observada	P (Nº Chegadas = k)	Freq. Esperada
0	39	$2^0e^{-2}/0!$ = 0,135	300x 0,135 = 40,6
1	91	$2^1e^{-2}/1!$ = 0,271	300x 0,271= 81,2
2	67	$2^2e^{-2}/2!$ = 0,271	300x 0,271 = 81,2
3	59	$2^3e^{-2}/3!$ = 0,180	300x 0,180 = 54,1
4	28	$2^4e^{-2}/4!$ = 0,090	300x 0,090 = 27,1
5	10	$2^5e^{-2}/5!$ = 0,036	300x 0,036 = 10,8
6	4	$2^6e^{-2}/6!$ = 0,012	300x 0,012 = 3,6
7	2	$2^7e^{-2}/7!$ = 0,003	300x 0,003 = 1,0
Soma	**300**		

- Como a freqüência observada é menor do que 5, agrupamos a última e penúltima linhas, tanto na freqüência observada como na freqüência esperada.

Teste de Aderência
Exemplo 2

- Calculamos: $$\chi^2_{\text{calc}} = \sum_{i=1}^{k} \frac{(\text{fo}_\text{i} - \text{fe}_\text{i})^2}{\text{fe}_\text{i}}$$

Nº Chegadas /Min	Freq. Observada	P (Nº Chegadas = k)	Freq. Esperada
0	39	0,135	40,6
1	91	0,271	81,2
2	67	0,271	81,2
3	59	0,180	54,1
4	28	0,090	27,1
5	10	0,036	10,8
6 ou 7	6	0,015	4,6
Soma	**300**		

$$\chi^2_{calc} = \frac{(39-40,6)^2}{40,6} + \frac{(91-81,2)^2}{81,2} + \frac{(67-81,2)^2}{81,2} + \frac{(59-54,1)^2}{54,1}$$
$$+ \frac{(28-27,1)^2}{27,1} + \frac{(10-10,8)^2}{10,8} + \frac{(6-4,6)^2}{4,6} = 4,88$$

Teste de Aderência
Exemplo 2

- Procuramos na tabela: χ^2_{crit} com φ graus de liberdade e $\alpha = 0,05$
- Onde: $\varphi = (\text{N}^{\underline{o}} \text{ de Classes}) - (\text{N}^{\underline{o}} \text{ parâmetros estimados}) - 1$

  $\varphi = 7 - 1 - 1 = 5$
- A média λ é o único parâmetro, logo com φ=5 e α=0,05 temos:

$$\chi^2_{crit} = \chi^2_5(0,05) = 11,07.$$

Como: $\chi^2_{calc} = 4,88 \leq \chi^2_{crit} = 11,07.$

- Aceita-se a hipótese de que a distribuição das chegadas segue uma distribuição de Poisson.

Teste de Aderência
Exemplo 3

- Utilize o teste χ^2 com nível de significância de 5% ou $(1-\gamma)=0,05$ para testar se a distribuição abaixo, obtida em uma amostragem, segue uma distribuição exponencial, com média estimada de $\mu=3$ clientes/minuto.

$$\boxed{P(t_1 \leq t \leq t_2) = e^{-\mu t_1} - e^{-\mu t_2}}$$

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Duração do Serviço/Segs	Freq. Observada	P (t)	Freq. Esperada
0-30	480	$e^{-3x0}-e^{-3x0,5}$ = 0,7769	540 x 0,7769 = 419,53
30-60	33	$e^{-3x0,5}-e^{-3x1}$ = 0,1733	540 x 0,1733 = 93,58
60-90	19	$e^{-3x1}-e^{-3x1,5}$ = 0,0387	540x0,0387=20,90
90-120	7	$e^{-3x1,5}-e^{-3x2}$ = 0,0086	540x0,0086=4,64
120-150	1	$e^{-3x2}-e^{-3x2,5}$ = 0,0019	540x0,0019=1,03
Soma	**540**		

- Como a freqüência observada é menor do que 5, agrupamos a última e penúltima linhas, tanto na freqüência observada como na freqüência esperada.

Teste de Aderência
Exemplo 3

- Calculamos: $\chi^2_{\text{calc}} = \sum_{i=1}^{k} \frac{(\text{fo}_i - \text{fe}_i)^2}{\text{fe}_i}$

Duração do Serviço/Segs	Freq. Observada	P (t)	Freq. Esperada
0-30	480	0,7769	419,53
30-60	33	0,1733	93,58
60-90	19	0,0387	20,90
90-150	8	0,0105	5,67
Soma	**540**		

$$\chi^2_{calc} = \frac{(480-419,53)^2}{419,53} + \frac{(33-93,58)^2}{93,58} + \frac{(19-20,9)^2}{20,9} + \frac{(8-5,67)^2}{5,67}$$

$$= 49,066$$

Teste de Aderência
Exemplo 3

- Procuramos na tabela: χ^2_{crit} com φ graus de liberdade e $\alpha = 0,05$
- Onde: $\varphi = (\text{N}^{\underline{o}} \text{ de Classes}) - (\text{N}^{\underline{o}} \text{ parâmetros estimados}) - 1$

  $\varphi = 4 - 1 - 1 = 2$
- A média λ é o único parâmetro, logo com φ =2 e α=0,05 temos:

$$\chi^2_{crit} = \chi^2_2(0,05) = 5,99.$$

Como: $\chi^2_{calc} = 49,066 > \chi^2_{crit} = 5,99.$

- Rejeita-se a hipótese de que a distribuição da duração do atendimento segue uma distribuição exponencial.

Teste de Aderência

Exemplos

1) No lançamento de uma moeda 200 vezes, ocorreram 80 caras e 120 coroas. Testar se a moeda é honesta ao nível de significância de 1%.

2) Um dado foi lançado 300 vezes e o resultado observado está apresentado na tabela abaixo. Verifique se o dado é honesto ao nível de significância de 2,5%.

	1	2	3	4	5	6
f_{oi}	58	55	52	43	40	52
f_{ei}	50	50	50	50	50	50

Teste de Independência

Definição

Este teste é utilizado para verificar se há independência entre duas variáveis aleatórias X e Y, sendo que X tem "n" amostras e Y tem "m" amostras.

A tabela abaixo, chamada de tabela de contingência apresenta as frequências observadas.

	X_1	X_2	...	X_n
Y_1	f_{o11}	f_{o12}	...	f_{o1n}
Y_2	f_{o21}	f_{o22}	...	f_{o2n}
...	...	...	...	...
Y_m	f_{om1}	f_{om2}	...	f_{omn}

Essa tabela é comparada a outra tabela de contingência para frequências esperadas.

Teste de Independência
Procedimento

1°) Definir as hipóteses nula e alternativa:
H_o: as variáveis X e Y são independentes
H_1: as variáveis X e Y não são independentes

2°) Fixar o nível de significância (α);

3°) Determinar a região de aceitação (RA) de H_o através do valor crítico (abscissa) da distribuição qui-quadrado, $\chi^2_{\text{crit}} = \chi^2_{\varphi}(\alpha)$, para φ=(n-1)x(m-1) graus de liberdade e nível de significância α;

4°) Montar a tabela de contingência das freqüências esperadas, onde cada elemento fe_{ij} é calculado como a soma dos elementos da linha i multiplicado pela soma dos elementos da coluna j dividido pelo total de observações;

5°) Determinar o valor da estatística:

$$\chi^2_{\text{calc}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\left(\text{fo}_{\text{ij}} - \text{fe}_{\text{ij}}\right)^2}{\text{fe}_{\text{ij}}}$$

6°) Se $\chi^2_{\text{calc}} \leq \chi^2_{\text{crit}}$ aceita-se H_o, caso contrário rejeita-se H_o.

Teste de Independência
Exemplo

Pesquisou-se a preferência de 4 emissoras de rádio em 3 bairros diferentes com 500 pessoas, numa determinada cidade. O resultado está apresentado na tabela abaixo. Verifique se há independência entre o bairro onde foi realizada a pesquisa e a preferência pela emissora, ao nível de significância de 10%.

Tabela de contingência das freqüências observadas (f_{oij})

Emissora	Bairro			Σ
	B1	B2	B3	
E1	56	42	102	200
E2	39	30	56	125
E3	19	9	22	50
E4	36	19	70	125
Σ	150	100	250	500

Teste de Independência

Exemplo

Definem-se as hipóteses nula e alternativa:

H_o: as variáveis Bairro e Emissora **são** independentes
H_1: as variáveis Bairro e Emissora **não são** independentes

Calcula-se a tabela de contingência para as frequências esperadas. Cada elemento da tabela fe_{ij} é calculado como a soma dos elementos da linha i multiplicado pela soma dos elementos da coluna j dividido pelo total de observações.

Tabela de contingência das freqüências esperadas (f_{eij})

Emissora	Bairro			Σ
	B1	**B2**	**B3**	
E1	60	40	100	200
E2	37,5	25	62,5	125
E3	15	10	25	50
E4	37,5	25	62,5	125
Σ	150	100	250	500

Teste de Independência
Exemplo

Calcula-se a estatística: $\chi^2_{calc} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left(fo_{ij} - fe_{ij}\right)^2}{fe_{ij}}$

$$\chi^2_{calc} = \frac{(56-60)^2}{60} + \frac{(42-40)^2}{40} + \frac{(102-100)^2}{100} + \frac{(39-37,5)^2}{37,5} + \frac{(30-25)^2}{25} + \frac{(56-62,5)^2}{62,5}$$
$$+ \frac{(19-15)^2}{15} + \frac{(9-10)^2}{10} + \frac{(22-25)^2}{25} + \frac{(36-37,5)^2}{37,5} + \frac{(19-25)^2}{25} + \frac{(70-62,5)^2}{62,5} = 6,07$$

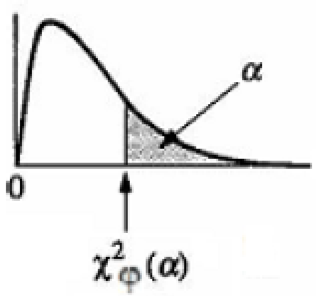
Calcula-se o valor crítico para a distribuição qui-quadrado: $\chi^2_{crit} = \chi^2_{\varphi}(\alpha)$

Onde:
O número de graus de liberdade φ corresponde a (n-1)x(m-1)=3x2=6 (Não foram estimados parâmetros).
O nível de significância α corresponde a 10%, ou seja, 0,1.
O valor crítico é obtido a partir da tabela da distribuição qui-quadrado.

Teste de Independência

Exemplo

Resposta: O valor críticos é obtido da tabela qui-quadrado:



φ	α 0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.015	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.050	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	2,67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	$\boxed{10.64}$	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00

$$\chi^2_{\text{crit}} = \chi^2_\varphi(\alpha) =$$

$$\chi^2_6(0,1) = 10,64$$

Como: $\chi^2_{\text{calc}} \leq \chi^2_{\text{crit}}$

$$6,07 \leq 10,64$$

Logo, aceita-se a hipótese nula, ou seja, ao nível de significância de 10%, existe independência entre o bairro onde foi realizado a pesquisa e a preferência pela emissora de rádio.