

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Porque os usamos?

Quando devemos usar?

O que é que mostram?

Como Calcular e interpretar?

Testes Ajustamento

Testes homogeneidade

Testes Independência

ENSAIOS NÃO PARAMÉTRICOS

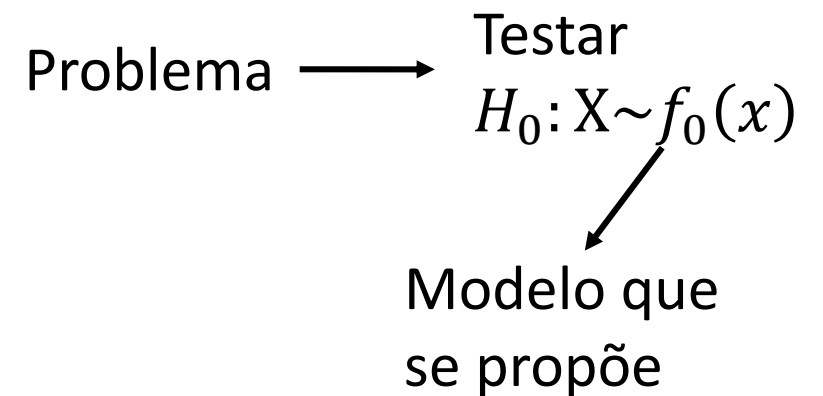


ENSAIOS DE AJUSTAMENTO

Ideia base:



Testar a aderência de um modelo ao comportamento de uma população



ENSAIOS NÃO PARAMÉTRICOS

Teste pode ser formulado com uma:

Hipótese simples - $f_0(x|\theta)$ é completamente especificada

Propõe-se um modelo

Propõe-se um valor para o parâmetro

Exemplos: $X \sim Po(10)$, $X \sim Bi(5, 0.3)$, $X \sim Ex(1/5)$, $X \sim N(2, 16)$

Hipótese composta - $f_0(x|\theta)$ não é completamente especificada

Propõe-se um modelo

desconhecido

Exemplos: $X \sim Po(\lambda)$, $X \sim Bi(n, \theta)$, $X \sim Ex(\lambda)$, $X \sim N(\mu, \sigma^2)$

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Ensaio de ajustamento

Porque os usamos?



Amostra aleatória de 20 alunos

$(X_1, X_2, \dots, X_{20})$

Para testar a afirmação de que a escola tem igual número de alunos do sexo feminino e masculino



Proporção de alunos sexo feminino = 50%



$H_0: X \sim B(1, 0.5)$

ENSAIOS NÃO PARAMÉTRICOS

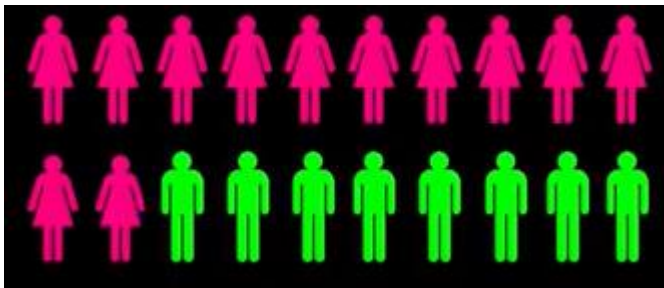
Testes do Qui-Quadrado



11: 9



Quais os valores para os quais rejeitaríamos H_0 ?



12: 8



Não podemos confiar na intuição\palpite!



14: 6



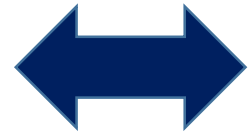
Precisamos de instrumentos analíticos mais robustos.

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

O que é que ele mostra?

Dados observados na amostra



Dados previstos pelo modelo

X - aluno sexo feminino

Hipótese nula - $H_0 =$ Modelo proposto \longrightarrow

Previsão

DADOS OBSERVADOS $\begin{matrix} \text{???} \\ \uparrow \\ \text{L-shaped arrow} \end{matrix}$

$$H_0: X \sim B(1, 0.5) \longrightarrow$$
$$\underbrace{10}_{\underbrace{20 * 0.5}_n} : \underbrace{10}_{\underbrace{20 * 0.5}_{1-p_0}}$$

Mostra a distância que separa valores observados e esperados

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como Calcular?

Estatística teste

Dados observados
na amostra

Dados previstos
pelo modelo

$$Q = \sum_{j=1}^m \frac{(\text{Valores Observ.} - \text{Valores Esperados})^2}{\text{Valores Esperados}}$$

Soma

$Q \sim \chi_{(?)^2}$ quando H_0 é verdadeira

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como calcular?

	Valores Observ. (1)	Valores previstos. (2)	Val.Obs. - Val. Prev. (1)-(2)	$[(1)-(2)]^2$	$\frac{[(1)-(2)]^2}{(2)}$
Alunas	13	10	3	9	0.9
Alunos	7	10	-3	9	0.9
				$q_{obs} =$	1.8

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como interpretar?

Para interpretar o teste do Qui-Quadrado precisamos da tabela da Distribuição do $\chi^2_{(?)}$

df	Probability level (alpha)					
	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

Qual a linha do quadro a fixar? A escolha depende dos graus de liberdade da $\chi^2_{(?)}$



ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como interpretar?

Atributo – género – 2 categorias $\Rightarrow m = 2 \Rightarrow m - 1 = 1 \Rightarrow \chi^2_{(1)}$

α

	Probability level (alpha)					
df	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

A região de rejeição de dimensão 0.05 é: $W = \{q: q > q_{0.05}\}$ com

$q_{0.05}: P(Q > q_{0.05}) = 0.05 \Rightarrow q_{0.05} = 3.841$. Como $q = 1.8 < 3.841$

Não se
rejeita H_0

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como calcular e interpretar?

	Valores Observ. (1)	Valores previstos. (2)	Val.Obs. - Val. Prev. (1)-(2)	$[(1)-(2)]^2$	$\frac{[(1)-(2)]^2}{(2)}$
Alunas	16	10	6	36	3.6
Alunos	4	10	-6	36	3.6
				$q_{obs} =$	7.2

A região de rejeição de dimensão 0.05 é: $W = \{q: q > 3.841\}$

Como $q = 7.2 > 3.841$ Rejeita-se $H_0: X \sim B(1, 0.5)$

ENSAIOS NÃO PARAMÉTRICOS

Como testar $H_0: X \sim f_0(x)$?

Uma possível solução: \longrightarrow Teste do Qui-Quadrado à Bondade do Ajustamento

Aplicação do teste em três circunstâncias distintas:

1ª situação: X corresponde a um **atributo qualitativo** com m categorias

Exemplo: um aspirador vendido em 5 cores A_1, A_2, A_3, A_4, A_5

• Notação:

Número de categorias

A_1, A_2, \dots, A_m \longrightarrow Categorias que o atributo pode assumir

$p_j = P(A_j)$ \longrightarrow Probabilidade (desconhecida) de um elemento da população, escolhido ao acaso, apresentar a modalidade A_j ($j = 1, 2, \dots, m$)

ENSAIOS NÃO PARAMÉTRICOS

- Hipótese nula $H_0: p_j = p_{0j}$ ($j = 1, 2, \dots, m$) contra $H_1: p_j \neq p_{0j}$, para algum j

$p_{0_1}, p_{0_2}, \dots, p_{0_m}$ conhecidos $p_{0j} > 0$ ($j = 1, 2, \dots, m$), $\sum_{j=1}^m p_{0j} = 1$

- O teste:

N_j - v.a. que representa o número de observações na amostra (de dimensão n) que assumem a modalidade A_j ($\sum_{j=1}^m N_j = n$)

Frequência observada da modalidade j Frequência esperada da modalidade j

Estatística teste: $Q = \sum_{j=1}^m \frac{(N_j - n * p_{0j})^2}{n * p_{0j}}$ mede o afastamento entre os dados observados e esperados

Quanto maior for o valor observado Q_{obs} menos plausível é a hipótese em teste.

Quando H_0 é verdadeira $Q = \sum_{j=1}^m \frac{(N_j - n * p_{0j})^2}{n * p_{0j}} \sim \chi^2_{(m-1)}$

ENSAIOS NÃO PARAMÉTRICOS

- O teste (continuação):

A região de rejeição de dimensão α é: $W = \{q: q > q_\alpha\}$ onde $q_\alpha: P(Q > q_\alpha) = \alpha$

Frequência da modalidade j
observada na amostra

$$\text{Rejeita-se } H_0 \text{ quando } q_{obs} = \sum_{j=1}^m \frac{(n_j - n * p_{0j})^2}{n * p_{0j}} > q_\alpha$$

Ou, utilizando o valor-p:

$$p_{obs} = P(Q > q_{obs} | H_0) \text{ e rejeita-se } H_0 \text{ quando } p_{obs} < \alpha$$

Observação importante

A distribuição de Q é válida quando $n \rightarrow +\infty$

Para que a aproximação no caso finito seja válida, deve-se garantir que:

$$n * p_{0j} \geq 5 \quad (\text{número esperado de elementos em cada classe/modalidade ser pelo menos 5})$$

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Exemplo: um aspirador vendido em 5 cores A_1, A_2, A_3, A_4, A_5

Um aspirador é vendido em cinco cores: verde (A_1), castanho (A_2), vermelho (A_3), azul (A_4) e branco (A_5). Num estudo de mercado para apreciar a popularidade das várias cores analisou-se uma amostra casual de 300 vendas recentes com o seguinte resultado

A_1	A_2	A_3	A_4	A_5	Total
88	65	52	40	55	300

Pretende testar-se a hipótese de que os consumidores não manifestam preferência por qualquer das cores ($\alpha = 0.05$)

$$H_0: p_{0_1} = p_{0_2} = \dots = p_{0_5} = \frac{1}{5}$$

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Solução:

1. Formalizar a hipótese nula $H_0: p_{0_1} = p_{0_2} = \dots = p_{0_5} = \frac{1}{5}$
2. Calcular as frequências esperadas de cada uma das modalidades

Modalidades	Freq. Observada n_j	Freq. Esperada $n * p_{0_j}$	$\frac{(n_j - n * p_{0_j})^2}{n * p_{0_j}}$
A_1	88	60	13.07
A_2	65	60	0.42
A_3	52	60	1.07
A_4	40	60	6.67
A_5	55	60	0.42
	300	300	21.65

3. Efectuar o teste

$$\alpha = 0.05; m - 1 = 4$$

$$q_{0.05}: P(\chi_{(5-1)}^2 > q_{0.05}) = 0.05$$

$$q_{0.05} = CHISQ.INV.RT(0.05,4) = 9.49$$

$$q_{obs} = 21.65 > 9.49$$

$$p_{obs} = P(Q > 21.65)$$

$$= CHISQ.DIST.RT(21.65,4) \\ = 0,000235 < 0.05$$

Rejeita-se H_0

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

- **2ª situação:** H_0 é uma hipótese simples $H_0: X \sim f_0(x)$ Não envolve qualquer parâmetro desconhecido

Ideia base \longrightarrow Adaptar esta situação para aplicar a metodologia anterior

- Construir uma partição do domínio de X em m classes A_1, A_2, \dots, A_m
- Calcular os valores $p_{0j} = P(A_j)$ $j = 1, 2, \dots, m$ recorrendo a $f_0(x)$
 - Quando a partição é dada parte-se dela;
 - Quando a partição fica ao nosso cuidado:
 - ▲ Variável contínua: constroem-se, tanto quanto possível, classes equiprováveis
 - ▲ Variável discreta: classes formadas pelos valores de D_X

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Exemplo (9. 2 do livro) – Um estudo sobre o tempo de vida, em dias, de uma amostra de 1000 tubos electrónicos deu os seguintes resultados:

Tempo de vida	$X \leq 150$ A_1	$150 < X < 300$ A_2	$300 < X < 450$ A_3	$450 < X < 600$ A_4	$600 < X < 750$ A_5	$X \geq 750$ A_6	Total
Freq.Observ.	543	258	120	48	20	11	1000

$$H_0: X \sim Ex(\lambda) \quad e \quad \mu_X = 200 = \frac{1}{\lambda} \Leftrightarrow \lambda = \frac{1}{200}$$

1. Formalizar a hipótese nula $H_0: X \sim Ex(1/200)$

2. Definir as classes: neste caso já estão definidas

3. Calcular $p_{0j} = P(X \in A_j) \quad j = 1, 2, \dots, 6$ $F_X(x) = 1 - e^{-\lambda x} \quad x > 0, \lambda > 0$

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

3. Calcular $p_{0_j} = P(X \in A_j)$ $j = 1, 2, \dots, 6$ $F_X(x) = 1 - e^{-\lambda x}$ $x > 0, \lambda > 0$

$$p_{0_1} = P(X \in A_1) = P(X \leq 150) = 1 - e^{-\frac{1}{200} * 150} = 0.52763$$

$$p_{0_2} = P(X \in A_2) = P(150 < X < 300) = F_X(300) - F_X(150) = 0.24924$$

$$p_{0_3} = P(X \in A_3) = P(300 < X < 450) = F_X(450) - F_X(300) = 0.11773$$

$$p_{0_4} = P(X \in A_4) = P(450 < X < 600) = F_X(600) - F_X(450) = 0.05561$$

$$p_{0_5} = P(X \in A_5) = P(600 < X < 750) = F_X(750) - F_X(600) = 0.02627$$

$$p_{0_6} = P(X \in A_6) = P(X \geq 750) = 1 - F_X(750) = 0.02352$$

$H_0: X \sim Ex(1/200) \longrightarrow H'_0: p_{0_1} = 0.52763, \dots, p_{0_6} = 0.02352$

↑
Hipótese aparentada

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

4. Obter as frequências esperadas das 6 classes Freq. Esperada da classe $j = n * p_{0j}$

	Frequência Observada	Frequência Esperada	$\frac{(n_j - n * p_{0j})^2}{n * p_{0j}}$
$X \leq 150$	543	527.63	0.4477
$150 < X < 300$	258	249.20	0.3108
$300 < X < 450$	120	117.73	0.0438
$450 < X < 600$	48	55.61	1.0414
$600 < X < 750$	20	26.27	1.4965
$X \geq 150$	11	23.52	6.6646
Total	1000	1000	10.0047

5. Efectuar o teste

$$\alpha = 0.05; m - 1 = 5$$

$$Q_{0.05} = 11.1$$

$$q_{obs} = 10.0047 < 11.1$$

ou

$$p_{obs} = P(Q > 10.005) \\ = 0,075 > 0.05$$

Não se rejeita $H'_0 \Rightarrow$ pode-se admitir que não será de pôr em causa H_0

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Observação importante

De facto, a hipótese testada não foi, $H_0: X \sim Ex(1/200)$

mas a hipótese “aparentada” $H'_0: p_{0_1} = 0.52763, \dots, p_{0_6} = 0.02352$

Se se rejeita H'_0 então não há dúvida de que também se deve rejeitar H_0 .

Quando não se rejeita H'_0 , sobretudo se se tiverem poucas classes, não se pode afirmar com tanta certeza que não se deve rejeitar H_0

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

- **3ª situação:** H_0 é uma hipótese composta $H_0: X \sim f_0 \left(x \mid \overbrace{\theta_1, \theta_2, \dots, \theta_k}^{\text{Desconhecidos}} \right)$

Exemplo (9.6 do livro) – Numa amostra de 100 peças de fazenda observou-se o número de defeitos por peça tendo-se obtido os resultados seguintes:

Defeitos por peça	0	1	2	3	4	5	Total
Freq. observada	20	30	25	10	10	5	100

Será de aceitar ($\alpha = 0.05$)
uma distribuição de Poisson?

1. Formalizar a hipótese nula $H_0: X \sim Po(\lambda)$ ↙ Desconhecido
2. Estimar o(s) parâmetro(s) desconhecido(s) $\tilde{\lambda} = \bar{X} \Rightarrow \tilde{\lambda}(x_1, x_2, \dots, x_{100}) = 1.75$

3. Calcular $\widehat{p}_{0j} = P(\widehat{X} = j)$
 $j = 0, 1, 2, \dots, 5$

Defeitos/peça	0	1	2	3	4	5+	
\widehat{p}_{0j}	0,17	0,30	0,27	0,16	0,07	0,03	1

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

4. Obter as frequências esperadas das 5 classes Freq. Esperada da classe $j = n * \widehat{p}_{0j}$

Defeitos peça por	0	1	2	3	4 e 5	Total
Freq. observada	20	30	25	10	10+5	100
Freq. esperada	17.38	30.41	26.61	15.52	10.08	
$\frac{(n_j - n * \widehat{p}_{0j})^2}{n * \widehat{p}_{0j}}$	0.40	0.01	0.10	1.96	2.4	4.87

5. Efectuar o teste

$$\alpha = 0.05; m - 1 - 1 = 3$$

$$Q_{0.05} = 7.815$$

$$q_{obs} = 4.87 < 7.815$$

ou

$$p_{obs} = P(Q > 7.815)$$

$$= 0,1816 > 0.05$$

Não se rejeita $H'_0 \Rightarrow$ pode-se admitir que não será de pôr em causa H_0

ENSAIOS NÃO PARAMÉTRICOS

Variáveis  2 atributos de uma população

Objectivo: Testar a independência entre 2 variáveis

Porque os usamos?

TESTE DE INDEPENDÊNCIA

Exemplos:

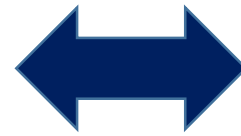
População	Atributo A	Atributo B
Alunos Ensino Superior	Sucesso escolar	Nº retenções Sec.
Atletas de salto em altura	Altura do salto	Género
Atletas que correm a maratona	Género	Idade

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

O que é que ele mostra?

Dados observados na amostra



Dados previstos

Observações das 2 variáveis

Independência entre as variáveis

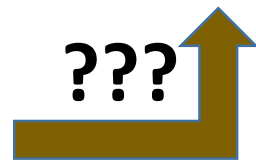
Tabela de contingência

Hipótese nula $H_0: X, Y$ independentes



Frequência esperada

DADOS OBSERVADOS



Frequência observada

Mostra a distância que separa valores observados e esperados

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

- **TABELA DE CONTINGÊNCIA**

Observa-se uma amostra à luz de 2 atributos:

Atributo A com r modalidades A_1, A_2, \dots, A_r

Atributo B com s modalidades B_1, B_2, \dots, B_s

Na célula (A_i, B_j) da tabela de contingência regista-se o número de elementos da amostra com a modalidade i do atributo A e a modalidade j do atributo B .

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Tabela de contingência ($r \times s$) Antes de observar a amostra

	B_1	B_2	...	B_s	Totais
A_1	N_{11}	N_{12}	...	N_{1s}	$N_{1\bullet}$
A_2	N_{21}	N_{22}	...	N_{2s}	$N_{2\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_r	N_{r1}	N_{r2}	...	N_{rs}	$N_{r\bullet}$
Totais	$N_{\bullet 1}$	$N_{\bullet 2}$...	$N_{\bullet s}$	N

Não é aleatório
porque dimensão
da amostra é fixa

N_{ij} ($i = 1, 2, \dots, r; j = 1, 2, \dots, s$) - frequência de elementos com modalidade i do atributo A e modalidade j do atributo B é uma **variável aleatória**.

$$N_{\bullet j} = \sum_{i=1}^r N_{ij} \quad (j = 1, 2, \dots, s), \quad N_{i\bullet} = \sum_{j=1}^s N_{ij} \quad (i = 1, 2, \dots, r)$$

São variáveis
aleatórias

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Como Calcular?

Tabela de contingência observada ($r * s$)

	B_1	B_2	...	B_s	Totais
A_1	n_{11}	n_{12}		n_{1s}	$n_{1\bullet}$
A_2	n_{21}	n_{22}		n_{2s}	$n_{2\bullet}$
\vdots					
A_r	n_{r1}	n_{r2}		n_{rs}	$n_{r\bullet}$
Totais	$n_{\bullet 1}$	$n_{\bullet 2}$		$n_{\bullet s}$	n

n_{ij} ($i = 1, 2, \dots, r; j = 1, 2, \dots, s$) - frequência observada de elementos com modalidade i do atributo A e modalidade j do atributo B .

$$n_{\bullet j} = \sum_{i=1}^r n_{ij} \quad (j = 1, 2, \dots, s)$$

$$n_{i\bullet} = \sum_{j=1}^s n_{ij} \quad (i = 1, 2, \dots, r)$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Como Calcular?

- Em termos do universo, as probabilidades (desconhecidas) das células (A_i, B_j) representam-se por,

$$p_{ij} = P(A_i, B_j) \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, s), \text{ verificando } \sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$$

- As respectivas probabilidades marginais são dadas por,

$$p_{i\cdot} = \sum_{j=1}^s p_{ij} \quad (i = 1, 2, \dots, r) \text{ verificando } \sum_{i=1}^r p_{i\cdot} = 1$$

$$p_{\cdot j} = \sum_{i=1}^r p_{ij} \quad (j = 1, 2, \dots, s) \text{ verificando } \sum_{j=1}^s p_{\cdot j} = 1$$

- Assumir a **independência entre os 2 atributos** equivale a assumir

$$P(A_i, B_j) = P(A_i) * P(B_j) = p_{i\cdot} * p_{\cdot j}$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

logo, a hipótese a testar vai ser:

Como Calcular?

$$H_0: p_{ij} = p_{i\cdot} p_{\cdot j} \quad \forall (i, j) \quad \text{contra} \quad H_0: p_{ij} \neq p_{i\cdot} p_{\cdot j} \quad \text{algum } (i, j)$$

- Assumindo H_0 , pode estimar-se p_{ij} a partir de $p_{i\cdot}$ e $p_{\cdot j}$
- As estimativas de máxima verosimilhança de $p_{i\cdot}$ e $p_{\cdot j}$ são dados por:

$$\widehat{p}_{i\cdot} = \frac{n_{i\cdot}}{n} \quad (i = 1, 2, \dots, r) \quad \text{e} \quad \widehat{p}_{\cdot j} = \frac{n_{\cdot j}}{n} \quad (j = 1, 2, \dots, s) \Rightarrow \widehat{p}_{ij} = \widehat{p}_{i\cdot} \widehat{p}_{\cdot j}$$

- A estatística teste vai avaliar a diferença entre a frequência observada e esperada

Frequência observada

Frequência esperada

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n \widehat{p}_{i\cdot} \widehat{p}_{\cdot j})^2}{n \widehat{p}_{i\cdot} \widehat{p}_{\cdot j}} \sim \chi^2_{[(r-1)(s-1)]}$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Como interpretar?

Notas:

- Os graus de liberdade obtêm-se verificando que existem rs células e se estimaram $(r - 1)$ parâmetros referentes ao atributo A (o último valor está pré-fixado) e $(s - 1)$ parâmetros referentes ao atributo B . Tem-se assim,

$$rs - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1)$$

- A região de rejeição vai situar-se, pelas mesmas razões que no teste do qui-quadrado à bondade do ajustamento na aba direita da distribuição
- Para que o teste seja válido mantem-se a restrição de um número mínimo esperado de elementos de cada célula (A_i, B_j) dado por $n \hat{p}_{\cdot i} \hat{p}_{\cdot j}$.

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Exemplo (9.10 do livro) – No quadro que se segue apresenta-se uma tabela 3 construída considerando os 86441 casamentos realizados em 1977 (que se podem considerar uma amostra dos casamentos realizados durante um período de alguns anos), em Portugal Continental (Anuário Estatístico, INE, 1980). Nela são apresentados, para cada sexo, o estado civil dos cônjuges anterior ao casamento.

Atributo A – estado civil da mulher

Modalidades do Atributo A :

1 – solteira

2 – viúva

3 - divorciada

Atributo B – estado civil do homem

Modalidades do Atributo B :

1 – solteiro

2 – viúvo

3 - divorciado

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

A hipótese a testar é a existência de independência entre o estado civil e o género de cada cônjuge no momento do casamento.

$$H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \quad \forall (i, j = 1, 2, 3) \quad \text{contra} \quad H_0: p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j} \quad \text{algum} (i, j = 1, 2, 3)$$

$$\widehat{p}_{\cdot 1} = \frac{79588}{86441} = 0,92$$

$$\widehat{p}_{\cdot 2} = \frac{2785}{86441} = 0,03$$

$$\widehat{p}_{\cdot 3} = \frac{4098}{86441} = 0,05$$

	Mulheres	Homens			Totais
		Solteiros	Viúvos	Divorciados	
Solteiras	77670	1573	3115	82358	
Viúvas	545	796	350	1691	
Divorciadas	1343	416	633	2392	
Totais	79558	2785	4098	86441	

$$\widehat{p}_{1\cdot} = \frac{82358}{86441} = 0,95$$

$$\widehat{p}_{2\cdot} = \frac{1691}{86441} = 0,02$$

$$\widehat{p}_{3\cdot} = \frac{2392}{86441} = 0,03$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Cálculo das estimativas para a probabilidade de $(A_i, B_j) \longrightarrow \hat{p}_{ij} = \hat{p}_{i\cdot} \hat{p}_{\cdot j}$

	B_1	B_2	B_3	$\hat{p}_{i\cdot}$
A_1	0,88	0,03	0,05	0,95
A_2	0,02	0,00	0,00	0,02
A_3	0,03	0,00	0,00	0,03
$\hat{p}_{\cdot j}$	0,92	0,03	0,05	

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

A azul, frequências esperadas na hipótese de os atributos serem independentes

Frequência esperada = $n \hat{p}_{i\cdot} \hat{p}_{\cdot j} = 86441 * \hat{p}_{ij} \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, s)$

$$Q \sim \chi^2_{\left[\frac{(3-1)(3-1)}{4} \right]}$$

valor - p
 $= P(Q > Q_{obs.})$
 $= P(\chi^2_{(4)} > 16509.74)$
 $\approx 0 \Rightarrow rej. H_0$

Mulheres	Homens			Totais
	Solteiros	Viúvos	Divorciados	
Solteiras	77670 <i>(75800.12)</i>	1573 <i>(2653.45)</i>	3115 <i>(3904.43)</i>	82358
Viúvas	545 <i>(1556.35)</i>	796 <i>(54.48)</i>	350 <i>(80.17)</i>	1691
Divorciadas	1343 <i>(2201.53)</i>	416 <i>(77.07)</i>	633 <i>(113.40)</i>	2392
Totais	79558	2785	4098	86441

Não existe independência entre género e estado civil no momento do casamento

$$Q_{obs} = \frac{(77670 - 75800.12)^2}{75800.12} + \frac{(1573 - 2653.45)^2}{2653.45} + \dots + \frac{(633 - 113.4)^2}{113.4} = 16509.74$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

A azul, frequências esperadas na hipótese de os atributos serem independentes

Frequência esperada = $n \hat{p}_{i\cdot} \hat{p}_{\cdot j} = 86441 * \hat{p}_{ij} \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, s)$

$$Q \sim \chi^2_{\left[\frac{(3-1)(3-1)}{4} \right]}$$

valor - p
 $= P(Q > Q_{obs.})$
 $= P(\chi^2_{(4)} > 16509.74)$
 $\approx 0 \Rightarrow rej. H_0$

Mulheres	Homens			Totais
	Solteiros	Viúvos	Divorciados	
Solteiras	77670 <i>(75800.12)</i>	1573 <i>(2653.45)</i>	3115 <i>(3904.43)</i>	82358
Viúvas	545 <i>(1556.35)</i>	796 <i>(54.48)</i>	350 <i>(80.17)</i>	1691
Divorciadas	1343 <i>(2201.53)</i>	416 <i>(77.07)</i>	633 <i>(113.40)</i>	2392
Totais	79558	2785	4098	86441

Não existe independência entre género e estado civil no momento do casamento

$$Q_{obs} = \frac{(77670 - 75800.12)^2}{75800.12} + \frac{(1573 - 2653.45)^2}{2653.45} + \dots + \frac{(633 - 113.4)^2}{113.4} = 16509.74$$