ENSAIOS NÃO PARAMÉTRICOS

Ideia base:

↓

Testar a aderência de um modelo ao comportamento de uma população

Problema ⟶ Testar $H_0: X \sim f_0(x)$

Modelo que se propõe

ENSAIOS NÃO PARAMÉTRICOS

Teste pode ser formulado com uma:

Propõe-se um modelo

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

Propõe-se um valor para o parâmetro

Exemplos: $X \sim Po(10)$, $X \sim Bi(5, 0.3)$, $X \sim Ex(^1/_5)$, $X \sim N(2, 16)$

Propõe-se um modelo

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

desconhecido

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

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Porque os usamos?

O que é que ele mostra?

Como Calcular e interpretar?

Quando o devemos usar?

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Ensaios de ajustamento

Porque os usamos?

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

$\Updownarrow$

Proporção de alunos sexo feminino = 50%

$\Updownarrow$

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

Amostra aleatória de 20 alunos

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

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

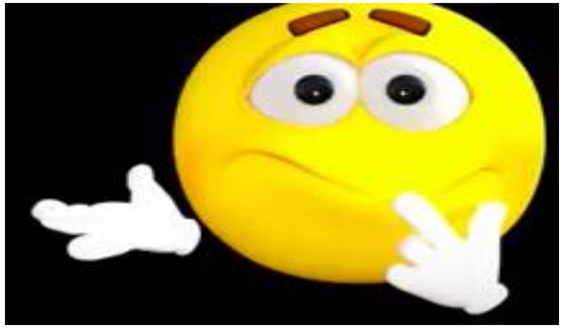
$11:9$

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 $\Leftrightarrow$ **Dados previstos pelo modelo**

X- aluno sexo feminino$\sim B(1,\theta)$

Hipótese nula - H_0 $=$ **Afirmação feita** $\longrightarrow$ **Previsão**

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

DADOS OBSERVADOS **???**

Mostra a distância que separa valores observados e esperados

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado **Como Calcular?**

Dados observados na amostra

Dados previstos pelo modelo

Estatística teste

$$Q = \sum \frac{(Observed - Expected)^2}{Expected}$$

Soma

$Q \dot{\sim} \chi^2_{(?)}$ 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
				$q = Q_{Obs}$	1.8

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como interpretar?

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

	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

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 $\Rightarrow \boldsymbol{m} = 2 \Rightarrow \boldsymbol{m} - \mathbf{1} = 1 \Rightarrow \chi^2_{(1)}$

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

$q_{0.05}: \mathrm{P}(Q > q_{0.05}) = 0.05 \Rightarrow q_{0.05}$=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
				$q = Q_{Obs}$	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

ENSAIOS DE AJUSTAMENTO

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

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

Aplicação do teste em duas 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

$A_1, A_2, \cdots, A_{\boldsymbol{m}}$ $\longrightarrow$ Categorias que o atributo pode assumir

$p_j = \mathrm{P}(A_j)$ $\longrightarrow$ Probabilidade (desconhecida) de um elemento da população, escolhido ao acaso, pertencer à categoria $A_j \quad (j = 1, 2, \cdots, m)$

ENSAIOS NÃO PARAMÉTRICOS

,

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

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

- O teste: N_j - v.a. que representa o número de observações na amostra (de dimensão $\boldsymbol{n}$) que pertencem à categoria A_j $(\sum_{j=1}^{m} N_j = n)$

Valores observados na amostra

Dados previstos pelo modelo

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

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

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

ENSAIOS NÃO PARAMÉTRICOS

- O teste (continuação):

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

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

Ou, utilizando o valor-p:

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

Observação importante

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

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

$n * p_{0_j} \geq 5$ (**valor previsto** para os elementos em cada classe\modalidade deve ser de 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 ($\alpha = 0.05$)

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

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Solução:

1. Formalizar a hipótese nula $H_0: p_{0_1} = p_{0_2} = \cdots = p_{0_5} = \frac{1}{5}$

2. Calcular os valores previstos para cada uma das modalidades

Categorias	**Valores Observados** n_j	**Valores previstos** $n * p_{0_j}$	$\frac{\left(n_j - n * p_{0_j}\right)^2}{n * p_{0_j}}$
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

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

$Q_{0.05} = 9.49$

$Q_{Obs} = 21.65 > 9.49$

$p_{obs} = P(Q > 21.65)$
$= 0{,}000235 < 0.05$

Rejeita-se H_0

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

- Construir uma partição do domínio de X em $\boldsymbol{m}$ classes $A_1, A_2, \cdots, A_{\boldsymbol{m}}$
  - Quando a partição fica ao nosso cuidado:
    - Variável contínua: constroem-se, tanto quanto possível, classes equiprováveis ou de igual amplitude.
    - Variável discreta: classes formadas pelos valores de D_X
- Calcular os valores $p_{0_j} = P(A_j) \quad j = 1, 2, \cdots, m$ recorrendo a $f_0(x)$
- Quando a v.a. é continua, para cada classe $A_j = [a, b)$ os valores

$P(A_j) = F_X(b) - F_X(a) \quad j = 1, 2, \cdots, m$ recorrendo a $F_0(x)$

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

- **2ª situação:** H_0 é uma hipótese composta $H_0: X \sim f_0\left(x \mid \overbrace{\theta_1, \theta_2, \cdots, \theta_k}^{\text{Desconhecidos}}\right)$

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**
Val. observados	20	30	25	10	10	5	100

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

Desconhecido

1. Formalizar a hipótese nula $H_0: X \sim Po(\boldsymbol{\lambda})$

2. Estimar o(s) parâmetro(s) desconhecido(s) $\tilde{\boldsymbol{\lambda}} = \bar{X} \Rightarrow \tilde{\boldsymbol{\lambda}}(x_1, x_2, \cdots, x_{100}) = 1.75$

3. Calcular $\widehat{p_{0_j}} = P(\widehat{X = j})$
$j = 0,1,2,\cdots,5$

Defeitos/peça	**0**	**1**	**2**	**3**	**4**	**5**
$\widehat{p_{0_j}}$	0,17	0,30	0,27	0,16	0,07	0,02

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

4. Obter os valores previstos para as 5 classes

Valor previsto da classe $j = n * \widehat{p_{0_j}}$

Defeitos por peça	**0**	**1**	**2**	**3**	**4 e 5**	**Total**
Val. observados	20	30	25	10	15	100
Val. previstos	17.38	30.41	26.61	15.52	9.17	
$\frac{\left(\boldsymbol{n_j - n * \widehat{p_{0_j}}}\right)^2}{\boldsymbol{n * \widehat{p_{0_j}}}}$	0,40	0,01	0,10	1,96	2,73	8,09

5. Efectuar o teste

$\alpha = 0.05;$

$m - k - 1 = 5 - 2 = 3$

$Q_{0.05} = 7.815$

$Q_{Obs} = 8.09 > 7.815$

ou

$p_{obs} = P(Q > 8.09)$
$= 0{,}088 > 0.05$

Não se rejeita H_0

Nota: agregaram-se as colunas 4 e 5 para satisfazer o critério de que valor previsto da classe ≥ 5

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

População	**Atributo A**	**Atributo B**
Alunos Superior	Sucesso escolar	Nota acesso Ens. Sup.
Atletas Olimpicos	Idade	Género

Amostra aleatória de dimensão $\boldsymbol{n}$

$$(X_1, X_2, \cdots, X_{\boldsymbol{n}})$$

Testes de Independência

Porque os usamos?

Para testar a independência entre duas características\atributos de uma população

$\Updownarrow$

Considerem-se $\boldsymbol{A}$ e B atributos

$\Updownarrow$

$$H_0: P\left(\boldsymbol{A}_i, \boldsymbol{B}_j\right) = P(\boldsymbol{A}_i) * P\left(\boldsymbol{B}_j\right) \quad \forall(\boldsymbol{i}, \boldsymbol{j})$$

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado | **O que é que ele mostra?**

Dados observados na amostra $\Longleftrightarrow$ **Dados previstos pelo modelo**

Hipótese nula → **Independência** → **Previsão**

H_0: A, B independentes → $P(\boldsymbol{A_i}, \boldsymbol{B_j}) = P(\boldsymbol{A_i}) * P(\boldsymbol{B_j})\ \forall(i,j)$

$\Updownarrow$

$\widehat{p_{ij}} = \widehat{p_{i\cdot}}\,\widehat{p_{\cdot j}}\ \forall(i,j)$

DADOS OBSERVADOS ???

Mostra a distância que separa valores observados e previstos

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

- **TABELA DE CONTINGÊNCIA**

Observa-se uma amostra à luz dos 2 atributos:

Atributo $\boldsymbol{A}$ com $\boldsymbol{r}$ categorias $\boldsymbol{A_1, A_2, \cdots, A_r}$

Atributo $\boldsymbol{B}$ com $\boldsymbol{s}$ categorias $\boldsymbol{B_1, B_2, \cdots, B_s}$

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

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Tabela de contingência (rxs) Antes de observar a amostra

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

Não é aleatório porque dimensão da amostra é fixa

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

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

São variáveis aleatórias

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Inquérito às preferências clubistas em certa região forneceu a ss informação:

Tabela de contingência observada $r * s \quad (2 * 3)$

	Clubes (categorias atributo A)			
Idade (categorias atributo B)	Porto	Benfica	Sporting	**Total**
35 ou menos	75	75	50	200
Mais de 35	75	125	100	300
Total	150	200	150	500

$\boldsymbol{n_{1.}}$ $\boldsymbol{n_{2.}}$ $\boldsymbol{n}$

$\boldsymbol{n_{.1}}$ $\boldsymbol{n_{.2}}$ $\boldsymbol{n_{.3}}$

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

$$\boldsymbol{n_{\bullet j} = \sum_{i=1}^{r} n_{ij} \; (j = 1, 2, \cdots, s),}$$

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

Como Calcular?

Estatística teste

Valores observados

Valores previstos pelo modelo

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(\boldsymbol{n_{ij}} - \overbrace{n * \widehat{p_i} * \widehat{p_j}}\right)^2}{n * \widehat{p_i} * \widehat{p_j}} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(\boldsymbol{n_{ij}} - \overbrace{n * \widehat{p_{ij}}}\right)^2}{n * \widehat{p_i} * \widehat{p_j}}$$

$Q \dot{\sim} \chi^2_{(\boldsymbol{r}-1)(\boldsymbol{s}-1)}$ quando H_0 é verdadeira

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

	1	2	3
$\widehat{p_{i\cdot}} = \frac{n_{i\cdot}}{n} \quad (i = 1, 2)$	$\frac{200}{500} = \frac{2}{5}$	$\frac{300}{500} = \frac{3}{5}$	
$\widehat{p_{\cdot j}} = \frac{n_{\cdot j}}{n} \quad (j = 1, 2, 3)$	$\frac{150}{500} = \frac{3}{10}$	$\frac{200}{500} = \frac{2}{5}$	$\frac{150}{500} = \frac{3}{10}$

$Se\ H_0: P\left(A_i, B_j\right) = P(A_i) * P\left(B_j\right)$ $\forall(i, j)$ verd. $\Rightarrow \widehat{p_{ij}} = \widehat{p_{i\cdot}}.\widehat{p_{\cdot j}}$

	A			
B	Porto	Benfica	Sporting	**Total**
35 ou menos	$\frac{2}{5} * \frac{3}{10} = \frac{3}{25}$	$\frac{2}{5} * \frac{2}{5} = \frac{4}{25}$	$\frac{2}{5} * \frac{3}{10} = \frac{3}{25}$	$10/25$
Mais de 35	$\frac{3}{5} * \frac{3}{10} = \frac{9}{50}$	$\frac{3}{5} * \frac{2}{5} = \frac{6}{25}$	$\frac{3}{5} * \frac{3}{10} = \frac{9}{50}$	$15/25$
Total	$3/10$	$2/5$	$3/10$	1

Tabela dos $\widehat{p_{ij}}$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

		Clubes (categorias atributo A)			
Idade (categorias atributo B)		Porto	Benfica	Sporting	**Total**
35 ou menos	Val.Observ	75	75	50	200
	Val. Previsto	60	80	60	
Mais de 35	Val.Observ	75	125	100	300
	Val. Previsto	90	120	90	
Total		150	200	150	500

Valores previstos=$\boldsymbol{n}^{*}\widehat{p_{ij}} = 500 * \widehat{p_{ij}}$

$$Q_{Obs.} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\boldsymbol{n_{ij}} - n.\widehat{p_i}\ \widehat{p_j}\right)^2}{n\,\widehat{p_i}\ \widehat{p_j}}$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Como interpretar?

$$Q_{Obs.} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\boldsymbol{n_{ij}} - n.\widehat{p_i}.\widehat{p_j}\right)^2}{n.\widehat{p_i}.\ \widehat{p_j}} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\boldsymbol{n_{ij}} - n.\widehat{p_{ij}}\right)^2}{n\,\widehat{p_{ij}}}$$

$$Q_{Obs.} = \frac{(75-60)^2}{60} + \frac{(75-80)^2}{80} + \frac{(75-60)^2}{60} + \frac{(75-90)^2}{90} + \frac{(125-120)^2}{120} + \frac{(100-90)^2}{90}$$

$$= 9.5486$$

$$Q \dot{\sim} \chi^2_{(2-1)(3-1)}$$ A região de rejeição de dimensão 0.05 é: $W = \{q: q > 5.991\}$

Como $Q_{Obs.} = q = 9.5486 < 5.991$ ou $p_{obs} = P(Q > 9.5486\,) = 0{,}008 < 0.05$

Rejeita-se $H_0: X \sim B(1, 0.5)$ **A Idade e Preferência clubista não são independentes**

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

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 $\left(A_i, B_j\right)$ dado por $n\,\widehat{p_{ij}} \geq 5$.

Quando esta condição não se verifica agregam-se colunas\linhas adjacentes até que seja verificada.

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.

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

$$\widehat{p_{\bullet 1}} = \frac{79588}{86441} = 0{,}92$$

$$\widehat{p_{\bullet 2}} = \frac{2785}{86441} = 0{,}03$$

$$\widehat{p_{\bullet 3}} = \frac{4098}{86441} = 0{,}05$$

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

$$\widehat{p_{1\bullet}} = \frac{82358}{86441} = 0{,}95$$

$$\widehat{p_{2\bullet}} = \frac{1691}{86441} = 0{,}02$$

$$\widehat{p_{3\bullet}} = \frac{2392}{86441} = 0{,}03$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Cálculo das estimativas para a probabilidade de $(A_i, B_j) \longrightarrow \widehat{p_{ij}} = \widehat{p_{i\bullet}}\, \widehat{p_{\bullet j}}$

	B_1	B_2	B_3	$\widehat{p_{i\bullet}}$
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
$\widehat{p_{\bullet j}}$	0,92	0,03	0,05	

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

A azul, valores previstos na hipótese de os atributos serem independentes

Valores previstos = $n\, \widehat{p_{i\bullet}}\ \widehat{p_{\bullet j}} = 86441 * \widehat{p_{ij}} \qquad (\boldsymbol{i = 1, 2, \cdots, r; j = 1, 2, \cdots, s})$

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

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

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

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

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