ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

Ideia base:

↓

Testar a aderência de um modelo ao comportamento de uma população

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

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%

$\Updownarrow$

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

Amostra aleatória de 20 alunos

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

ENSAIOS NÃO PARAMÉTRICOS

Testes do Qui-Quadrado

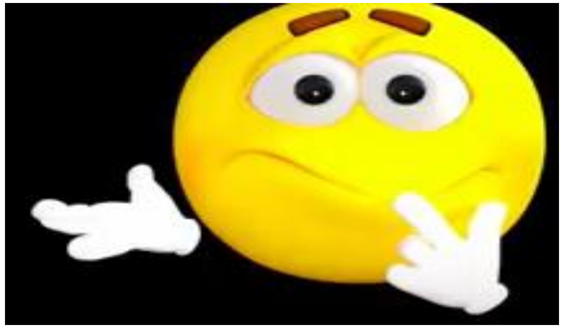
11: 9

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 ⟷ **Dados previstos pelo modelo**

X - aluno sexo feminino

Hipótese nula - H_0 $=$ **Modelo proposto** ⟶ **Previsão**

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

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

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

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
				$q_{Obs}=$	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
				$q_{Obs} =$	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: $\mathrm{X} \sim f_0(x)$?

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

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

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

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 assumem a modalidade A_j $(\sum_{j=1}^{m} N_J = n)$

Frequência observada da modalidade j

Frequência esperada da modalidade j

Estatística teste: $Q = \sum_{j=1}^{m} \frac{\left(N_j - n * p_{0_j}\right)^2}{n * p_{0_j}}$ 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 $Q = \sum_{j=1}^{m} \frac{\left(N_j - n * p_{0_j}\right)^2}{n * p_{0_j}} \overset{\cdot}{\sim} \chi^2_{(m-1)}$

ENSAIOS NÃO PARAMÉTRICOS

- O teste (continuação):

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

Frequência da modalidade j observada na amostra

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:

$n * p_{0_j} \geq 5$ (**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 ($\alpha = 0.05$)

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

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 as frequências esperadas de cada uma das modalidades

Modalidades	**Freq. Observada** n_j	**Freq. Esperada** $n * p_{0_j}$	$\frac{\left(n_j - n * p_{0_j}\right)^2}{n * p_{0_j}}$
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

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

$q_{0.05}: P\left(\chi^2_{(5-1)} > q_{0.05}\right) = 0.05$

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

$q_{Obs} = 21.65 > 9.49$

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

$= CHISQ.DIST.RT(21.65{,}4)$

=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 $\boldsymbol{m}$ classes $A_1, A_2, \cdots, A_{\boldsymbol{m}}$
- Calcular os valores $p_{0_j} = P(A_j) \quad j = 1, 2, \cdots, m$ 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

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

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

2. Definir as classes: neste caso já estão definidas

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

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

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

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

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

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

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

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

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

Hipótese aparentada

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

ENSAIOS NÃO PARAMÉTRICOS

ENSAIOS DE AJUSTAMENTO

4. Obter as frequências esperadas das 6 classes

Freq. Esperada da classe $j = n * p_{0_j}$

	Frequência Observada	Frequência Esperada	$\frac{\left(\boldsymbol{n_j} - \boldsymbol{n} * p_{0_j}\right)^2}{\boldsymbol{n} * p_{0_j}}$
$\boldsymbol{X \leq 150}$	543	527.63	0.4477
$\boldsymbol{150 < X < 300}$	258	249.20	0.3108
$\boldsymbol{300 < X < 450}$	120	117.73	0.0438
$\boldsymbol{450 < X < 600}$	48	55.61	1.0414
$\boldsymbol{600 < X < 750}$	20	26.27	1.4965
$\boldsymbol{X \geq 150}$	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' ⇒ 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, $H_0\!: \mathrm{X} \sim Ex(1/200)$

mas a hipótese "aparentada" $H_0'\!: p_{0_1} = 0.52763, \cdots, p_{0_6} = 0.02352$

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 $\quad H_0\colon X \sim f_0\left(x \mid \overbrace{\theta_1, \theta_2, \cdots, \theta_k}^{\textbf{Desconhecidos}}\right)$

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 $\quad H_0\colon X \sim Po(\boldsymbol{\lambda})$ **Desconhecido**

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

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 as frequências esperadas das 5 classes

Freq. Esperada da classe $j = n * \widehat{p_{0_j}}$

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

5. Efectuar o teste

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

$Q_{0.05} = 9.488$

$q_{Obs} = 8.09 < 11.1$

ou

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

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

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

Objectivo: Testar a independência entre 2 variáveis

Variáveis → 2 atributos de uma população

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

- **TABELA DE CONTINGÊNCIA**

Observa-se uma amostra à luz de 2 atributos:

Atributo A com $\boldsymbol{r}$ modalidades $A_1, A_2, \cdots, A_{\boldsymbol{r}}$

Atributo B com $\boldsymbol{s}$ modalidades $B_1, B_2, \cdots, B_{\boldsymbol{s}}$

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 (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 modalidade i do atributo A e modalidade 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

Tabela de contingência observada ($r * s$)

	B_1	B_2	$\cdots$	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} $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$ - **frequência observada** de elementos com modalidade i do atributo A e modalidade j do atributo B.

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

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

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

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

- As respectivas probabilidades marginais são dadas por,

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

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

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

$$P\left(A_i, B_j\right) = P(A_i) * P\left(B_j\right) = p_{i\bullet}\, p_{\bullet j}$$

ENSAIOS NÃO PARAMÉTRICOS

TESTE DE INDEPENDÊNCIA

logo, a hipótese a testar vai ser:

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

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

$$\widehat{p_{i\bullet}} = \frac{n_{i\bullet}}{n} \quad (\boldsymbol{i = 1, 2, \cdots, r}) \quad \text{e} \quad \widehat{p_{\bullet j}} = \frac{n_{\bullet j}}{n} \quad (\boldsymbol{j = 1, 2, \cdots}, s) \Rightarrow \widehat{p_{ij}} = \widehat{p_{i\cdot}}.\widehat{p_{\bullet j}}$$

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

Frequência observada

Frequência esperada

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(\boldsymbol{n_{ij}} - n.\widehat{p_{i\bullet}}\,\widehat{p_{\bullet j}}\right)^2}{n\,\widehat{p_{\bullet i}}\ \ \widehat{p_{\bullet j}}} \dot{\sim} \chi^2_{[(r-1)(s-1)]}$$

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 <u>número mínimo esperado de elementos de cada célula</u> $\left(A_i, B_j\right)$ dado por $n\,\widehat{p_{\bullet i}}\ \widehat{p_{\bullet j}}$.

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.

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

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

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

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

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, frequências esperadas na hipótese de os atributos serem independentes

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

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

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

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

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