

M109

Estatística

Aula 7 – Testes Categorizados

Prof^a. Rosimara Beatriz Arci Salgado

Créditos:

Prof^a. Karina Perez Mokarzel Carneiro

Prof. Rausley Adriano Amaral de Souza

Inatel

CAMINHOS
QUE CONECTAM
COM O FUTURO



Bibliografia básica:

Arango HG. Bioestatística: teórica e computacional. 3ªed. Rio de Janeiro: Guanabara Koogan; 2011.

SPIEGEL, Murray Ralph; FARIA, Alfredo Alves De Probabilidade e estatística. São Paulo, SP: McGraw-Hill, 1978.

TESTES PARA DADOS CATEGORIZADOS



→ Amostras independentes

- $n > 40$ → Teste χ^2 Clássico.
- $20 < n < 40$ → Teste χ^2 com Correção de Yates.
- $n < 20$ → Teste Exato de Fisher

→ Amostras pareadas

- Teste McNemar

TESTES PARA DADOS AMOSTRAIS



Testes de Normalidade:

TESTES PARAMÉTRICOS → Teste de Student (t)

→ Amostras Independentes

Populações Homocedásticas e Populações Heterocedásticas

→ Amostras Pareadas

Teste de Fisher

TESTES NÃO PARAMÉTRICOS

→ Amostras pareadas → Testes dos sinais

OBS: Também utilizado na comparação dos resultados de uma amostra com a mediana de uma população e na comparação de dados qualitativos

→ Amostras independentes → Teste de wilcoxon-mann-whitney

TESTES PARA DADOS CATEGORIZADOS



→ Amostras independentes

- $n > 40$ → Teste χ^2 Clássico.
- $20 < n < 40$ → Teste χ^2 com Correção de Yates.
- $n < 20$ → Teste Exato de Fisher

→ Amostras pareadas

- Teste McNemar

TESTES CATEGORIZADOS

Conceitos Básicos

Dados categorizados (variáveis **categóricas** ou variáveis de **atributo**) referem-se à contagem de frequência de uma variável classificada ou subdividida em categorias ou atributos.

Embora esse procedimento seja típico de dados referentes a variáveis qualitativas, é possível também criar categorias para dados de variáveis quantitativas.



TABELA DE CONTINGÊNCIA



1º Passo: Tabela de contingência

		Fator discriminado (B)				TOTAIS
		B_1	B_2	\dots	B_s	
Fator discriminante (A)	A_1	o_{11}	o_{12}	\dots	o_{1s}	A_1
	A_2	o_{21}	o_{22}	\dots	o_{2s}	A_2
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	A_r	o_{r1}	o_{r2}	\dots	o_{rs}	A_r
	TOTAIS	B_1	B_2	\dots	B_s	T

Tabela 9.1 Tabela genérica de contingência com r linhas e s colunas

TESTES CATEGORIZADOS



O objetivo dos testes para dados categorizados é determinar, segundo algum critério válido de decisão, se o fator discriminante exerce alguma influência sobre o fator discriminado.

2º Passo: Hipóteses

H₀ → as categorias de A exercem igual influência sobre B

H₁ → pelo menos uma categoria de A exerce influência diferente sobre B

Tabelas 2 × 2 → identificação do fator responsável.

TESTE DE QUI-QUADRADO CLÁSSICO



- Amostras independentes
 - $n > 40$
- } Teste χ^2 clássico

3º Passo: Tabela de valores esperados, E:

$$E_{ij} = \frac{\sum_{j=1}^s O_{ij} \sum_{i=1}^r O_{ij}}{\sum_{i=1}^r \sum_{j=1}^s O_{ij}} = \frac{A_i \cdot B_j}{T}$$

4º Passo: Valor de qui-quadrado calculado:

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

TESTE DE QUI-QUADRADO CLÁSSICO



5º Passo: Valor de qui-quadrado tabelado → Decisão estatística:

$$\chi_t^2 = f(\alpha, gl) \quad \text{com} \quad gl = (r-1).(s-1)$$

Valores clássicos para tabelas 2 x 2:

OBS: Os passos 3º, 4º e 5º podem ser calculados por um software estatístico

α	χ_t^2
10%	2,7056
5%	3,8416
1%	6,6354

6º Passo: Regra de decisão:

$$\text{Se} \quad \begin{cases} \chi_c^2 \leq \chi_t^2 \rightarrow H_0 & \text{Aceita (não pode ser rejeitada)} \\ \chi_c^2 > \chi_t^2 \rightarrow H_0 & \text{Rejeitada} \end{cases}$$

7º Passo: Conclusão:

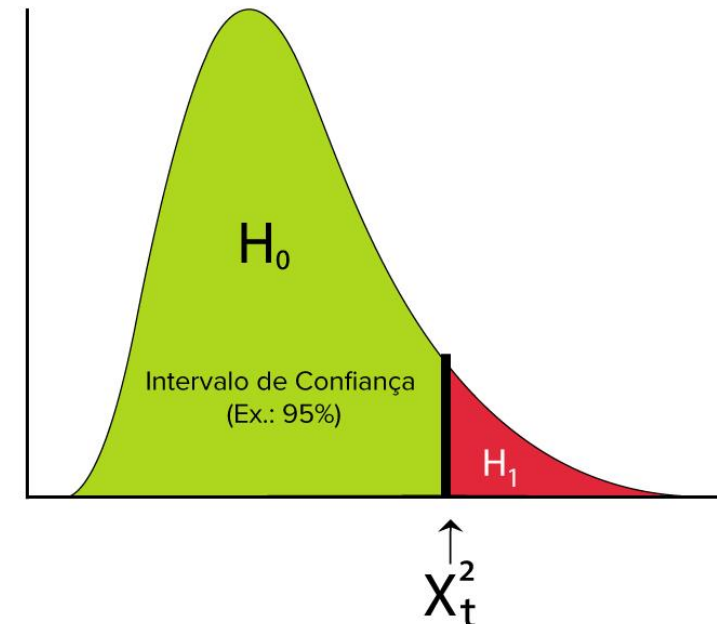
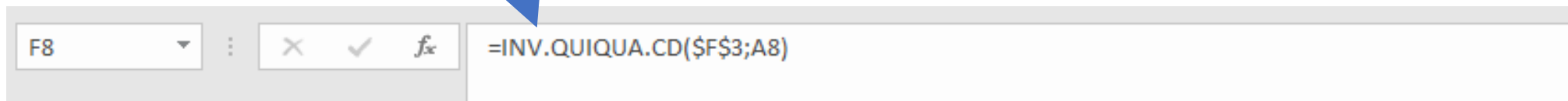


TABELA DA DISTRIBUIÇÃO QUI-QUADRADO NO EXCEL



	A	B	C	D	E	F	G	H	I	J	K
1	Valores da distribuição Qui-quadrado										
2		Nível de Significância									
3		0,10%	0,50%	1,00%	2,00%	4,00%	5,00%	6,00%	8,00%	10,00%	20,00%
4	Graus de Liberdade										
5	1	10,827566	7,879439	6,634897	5,411894	4,217885	3,841459	3,537385	3,064902	2,705543	1,642374
6	2	13,815511	10,596635	9,210340	7,824046	6,437752	5,991465	5,626821	5,051457	4,605170	3,218876
7	3	16,266236	12,838156	11,344867	9,837409	8,311171	7,814728	7,406880	6,758693	6,251389	4,641628
8	4	18,466827	14,860259	13,276704	11,667843	10,025519	9,487729	9,044368	8,336532	7,779440	5,988617
9	5	20,515006	16,749602	15,086272	13,388223	11,644332	11,070498	10,596232	9,836591	9,236357	7,289276
10	6	22,457744	18,547584	16,811894	15,033208	13,197815	12,591587	12,089578	11,283496	10,644641	8,558060
11	7	24,321886	20,277740	18,475307	16,622422	14,703047	14,067140	13,539734	12,691176	12,017037	9,803250
12	8	26,124482	21,954955	20,090235	18,168231	16,170776	15,507313	14,956339	14,068397	13,361566	11,030091
13	9	27,877165	23,589351	21,665994	19,679016	17,608277	16,918978	16,345918	15,421088	14,683657	12,242145
14	10	29,588298	25,188180	23,209251	21,160768	19,020743	18,307038	17,713124	16,753478	15,987179	13,441958
15	11	31,264134	26,756849	24,724970	22,617941	20,412034	19,675138	19,061413	18,068707	17,275009	14,631421
16	12	32,909490	28,299519	26,216967	24,053957	21,785109	21,026070	20,393435	19,369183	18,549348	15,811986
17	13	34,578170	29,819174	27,689250	25,474150	23,142007	22,362032	21,714126	20,655700	19,814020	16,994707



The chi-square test

7.34. In 200 tosses of a coin, 115 heads and 85 tails were observed. Test the hypothesis that the coin is fair using a level of significance of (a) 0.05, (b) 0.01. (c) Find the P value of the test.

Observed frequencies of heads and tails are, respectively, $x_1 = 115$, $x_2 = 85$.

Expected frequencies of heads and tails if the coin is fair are $np_1 = 100$, $np_2 = 100$, respectively. Then

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} = \frac{(115 - 100)^2}{100} + \frac{(85 - 100)^2}{100} = 4.50$$

Since the number of categories or classes (heads, tails) is $k = 2$, $\nu = k - 1 = 2 - 1 = 1$.

- (a) The critical value $\chi_{0.95}^2$ for 1 degree of freedom is 3.84. Then since $4.50 > 3.84$, we reject the hypothesis that the coin is fair at a 0.05 level of significance.
- (b) The critical value $\chi_{0.99}^2$ for 1 degree of freedom is 6.63. Then since $4.50 < 6.63$, we cannot reject the hypothesis that the coin is fair at a 0.01 level of significance.

We conclude that the observed results are *probably significant* and the coin is *probably not fair*. For a comparison of this method with previous methods used, see Method 1 of Problem 7.36.

- (c) The P value is $P(\chi^2 \geq 4.50)$. The table in Appendix E shows $0.025 < P < 0.05$. By computer software, $P = 0.039$.



Exemplo 1

A Tabela mostra os resultados de um ensaio com 154 pacientes que apresentavam dor abdominal. Ao grupo Tratamento (T) foi administrado Brometo de Pinavério (dois comprimidos/dia), enquanto ao grupo Controle (C) foi dado placebo.

Tabela de valores observados		Permanência da dor abdominal		
		Sim	Não	Total
Grupo	Tratamento	6	57	63
	Controle	30	61	91
	Total	36	118	154

(n > 40)

Para testar a eficiência do sal no tratamento da dor abdominal, deve ser efetuado o teste de qui-quadrado clássico.



Exemplo 1

1º Passo: Hipóteses

$$H_0 \rightarrow p_C = p_T$$

$$H_1 \rightarrow p_C \neq p_T$$

		Permanência da dor abdominal		
		Sim	Não	Total
Grupo	Tratamento	6	57	63
	Controle	30	61	91
	Total	36	118	154

A hipótese de nulidade se refere aos resultados dos grupos T e C serem iguais.

Dessa forma, do ponto de vista clínico, interessa a rejeição de H_0 (aceitação de H_1), que indicaria a eficiência terapêutica da droga.



Exemplo 1 - Usando Excel

2º Passo: Cálculos usando o Excel

4 Teste Qui-Quadrado.xls [Modo de Compatibilidade] - Excel

ARQUIVO PÁGINA INICIAL INSERIR LAYOUT DA PÁGINA FÓRMULAS DADOS REVISÃO EXIBIÇÃO

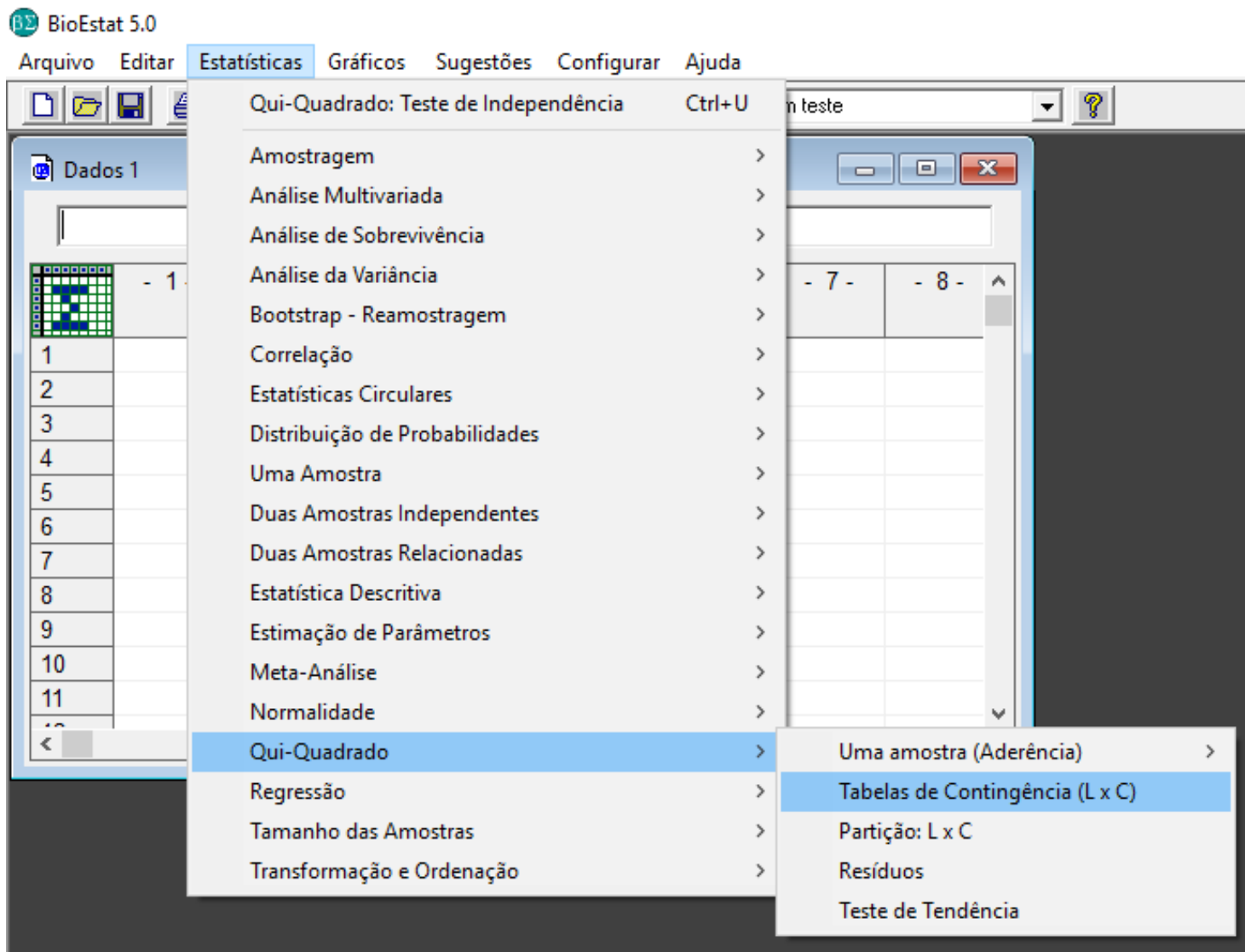
Área de Transferência Fonte Alinhamento Número Estilo Células

G5 : $=D5*B7/D7$

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Observado							Esperado				
3	Permanencia da dor											
4	Grupo	SIM	NÃO	Total			SIM	NÃO				
5	Tratamento	6	57	63	Trat.		14.72727	48.27273	63		Qui-Quadrado Calculado	p
6	Controle	30	61	91	Cont.		21.27273	69.72727	91		11.42228017	0.0726%
7	Total	36	118	154			36	118	154			
8												
9												
10											α	Qui-Quadrado Tabelado
11											10%	2.705543454
12											5%	3.841458821
13											1%	6.634896601
14												



Exemplo 1 usando BioEstat 5.0



Exemplo 1 usando BioEstat 5.0

BioEstat 5.0

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Escolha um teste

Dados 1

	- 1 -	- 2 -	- 3 -
1	6	57	
2	30	61	
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
<			

Qui-Quadrado: Teste de I...

Arquivo Editar Gráfico

	Resultados
Tabela de Contingência =	2 x 2
Qui-Quadrado =	11.422
Graus de liberdade =	1
(p) =	0.0007
Correção de Yates =	10.151
(p) =	0.0014

Exemplo 1

3º Passo: Regra de decisão:

Como o valor calculado de $p = 0,0007$ (0,07%) é menor do que 0,01 (1%), rejeita-se H_0 . Existe uma diferença altamente significativa entre o grupo tratamento e o grupo controle.

4º Passo: Conclusão:

Assim, fica comprovado o efeito terapêutico do Brometo de Pinavério no alívio das dores abdominais nos níveis tradicionais de significância. A afirmativa de que o Brometo de Pinavério é eficaz no combate das dores abdominais com base no ensaio efetuado envolve um erro de decisão de 0,0726%, isto é, uma segurança altamente significativa de que a decisão é correta.



Teste de Qui-Quadrado com Correção de Yates

- Amostras independentes
- $20 < n < 40$
- Valor esperado da célula maior que 5

1º Passo: Hipóteses

2º Passo: Tabela de valores esperados

3º Passo: Valor de qui-quadrado calculado:

$$\chi_c^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(|O_{ij} - E_{ij}| - 0,5)^2}{E_{ij}}$$

Correção
de Yates

4º Passo: Decisão estatística

5º Passo: Regra de decisão

Teste Exato de Fisher

→ Amostras independentes

→ $n < 20$

→ Tabelas 2×2

		Fator discriminado (B)		
		B ₁	B ₂	TOTAIS
Fator discriminante (A)	A ₁	o ₁₁	o ₁₂	A ₁
	A ₂	o ₂₁	o ₂₂	A ₂
TOTAIS		B ₁	B ₂	T < 20



Exemplo 2

Suponha um grupo de 16 ratos, divididos em dois grupos, experimental e normal. O grupo experimental é formado por 9 animais geneticamente modificados, por apresentarem uma disfunção pancreática com diminuição da capacidade de produção de insulina. Imagine que, após 1,5 ano em ambiente controlado, o número de ratos vivos do grupo experimental e do normal seja o seguinte:

		Sobrevida +1,5 ano		
		Vivos	Mortos	Total
Grupo	Normal	5	2	7
	Experimental	1	8	9
	Total	6	10	16



Exemplo 2

1º Passo: Hipóteses

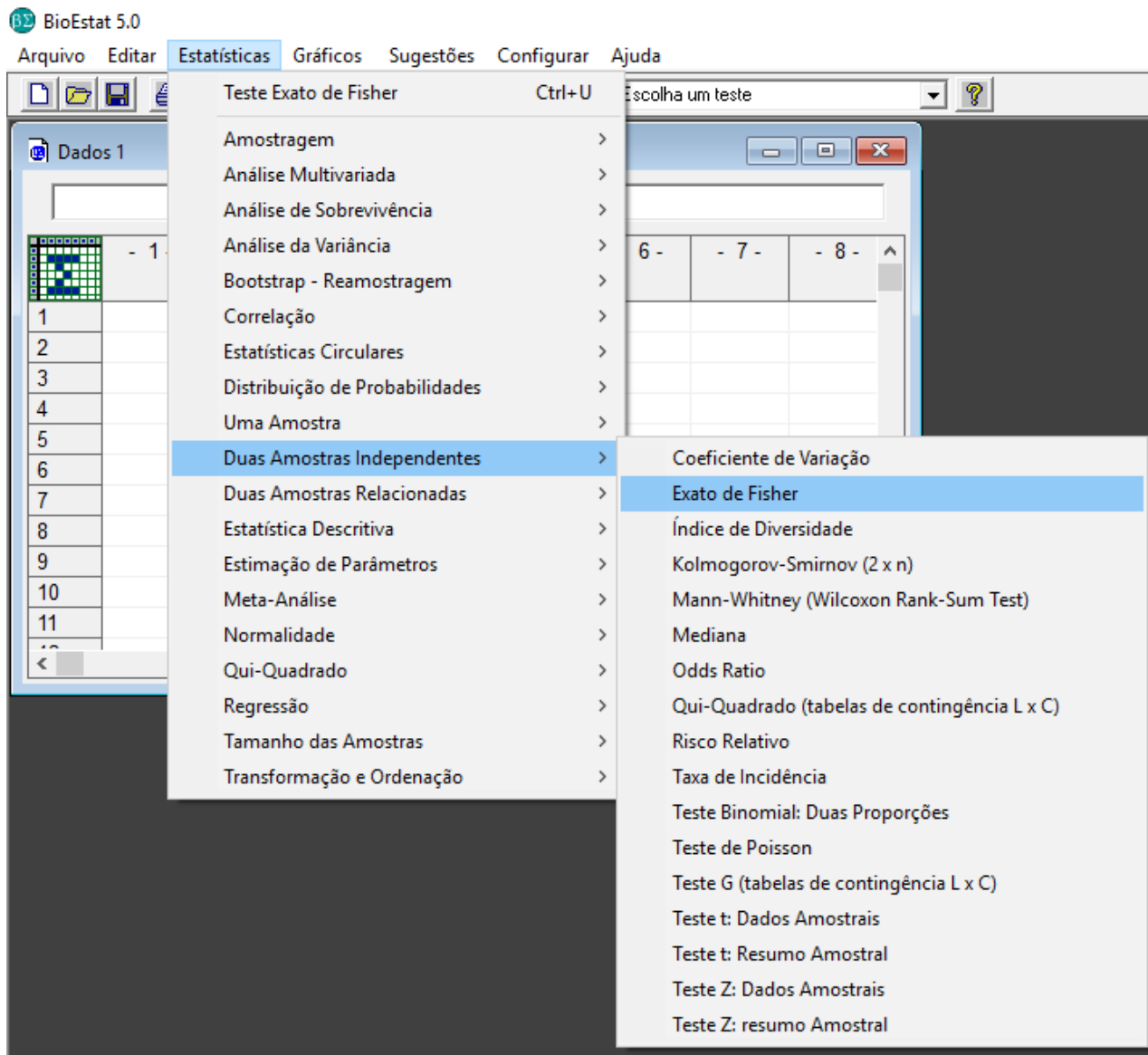
$$H_0 \rightarrow p_N = p_E$$

$$H_1 \rightarrow p_N \neq p_E$$

2º Passo: Cálculos



Exemplo 2 usando BioEstat 5.0



Exemplo 2 usando BioEstat 5.0

BioEstat 5.0

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Escolha um teste

Dados 1

	- 1 -	- 2 -	- 3 -
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			

Teste Exato de Fisher

Imprimir

Dados

	Colunas		Total / Linhas
	1	2	
1	5	2	7
2	1	8	9
Total / Colunas	6	10	16

Executar Sair

Resultados

Unilateral	Bilateral
$p = 0.0245$	$p = 0.0350$

Inatel

CAMINHOS
QUE CONECTAM
COM O FUTURO

Exemplo 2

3º Passo: Regra de decisão:

Como o valor calculado de $p = 0,035$ (3,5%) é menor do que 0,05 (5%), rejeita-se H_0 . Existe uma diferença significativa entre o grupo normal e o grupo experimental.

4º Passo: Conclusão:

Ou seja, a afirmação de que a sobrevivência dos ratos geneticamente alterados é diferente que a dos ratos normais envolve uma probabilidade de erro de 3,5%. Portanto, ao nível de 3,5% de significância, rejeita-se a hipótese de nulidade (ou seja, a hipótese de que as vidas dos ratos normais e dos transgênicos são iguais).



Teste de McNemar

→ Amostras pareadas

Emprega-se o teste de McNemar quando o fator discriminante é categorizado segundo duas situações que se referem ao mesmo grupo (grupos dependentes).



Exemplo 3

Numa experiência, 40 pacientes foram submetidos a um teste de memória que consistia em responder a uma lista de 15 palavras na sequência correta.

O teste foi aplicado antes e depois da ingestão de uma dose de 6 mg de Mesilato de Diidroergocristina.

Foram considerados corretos os testes em que o paciente conseguia responder toda a sequência corretamente (15 palavras).



CAMINHOS
QUE CONECTAM
COM O FUTURO

Exemplo 3

Para explicitar as informações de desempenho de cada par, a tabela deve ser da seguinte forma:

Tabela - Resultados do teste de memória individualizados

		Depois		
		Correto	Incorreto	Total
Antes	Correto	8	1	9
	Incorreto	14	17	31
	Total	22	18	40



Exemplo 3

1º Passo: Hipóteses:

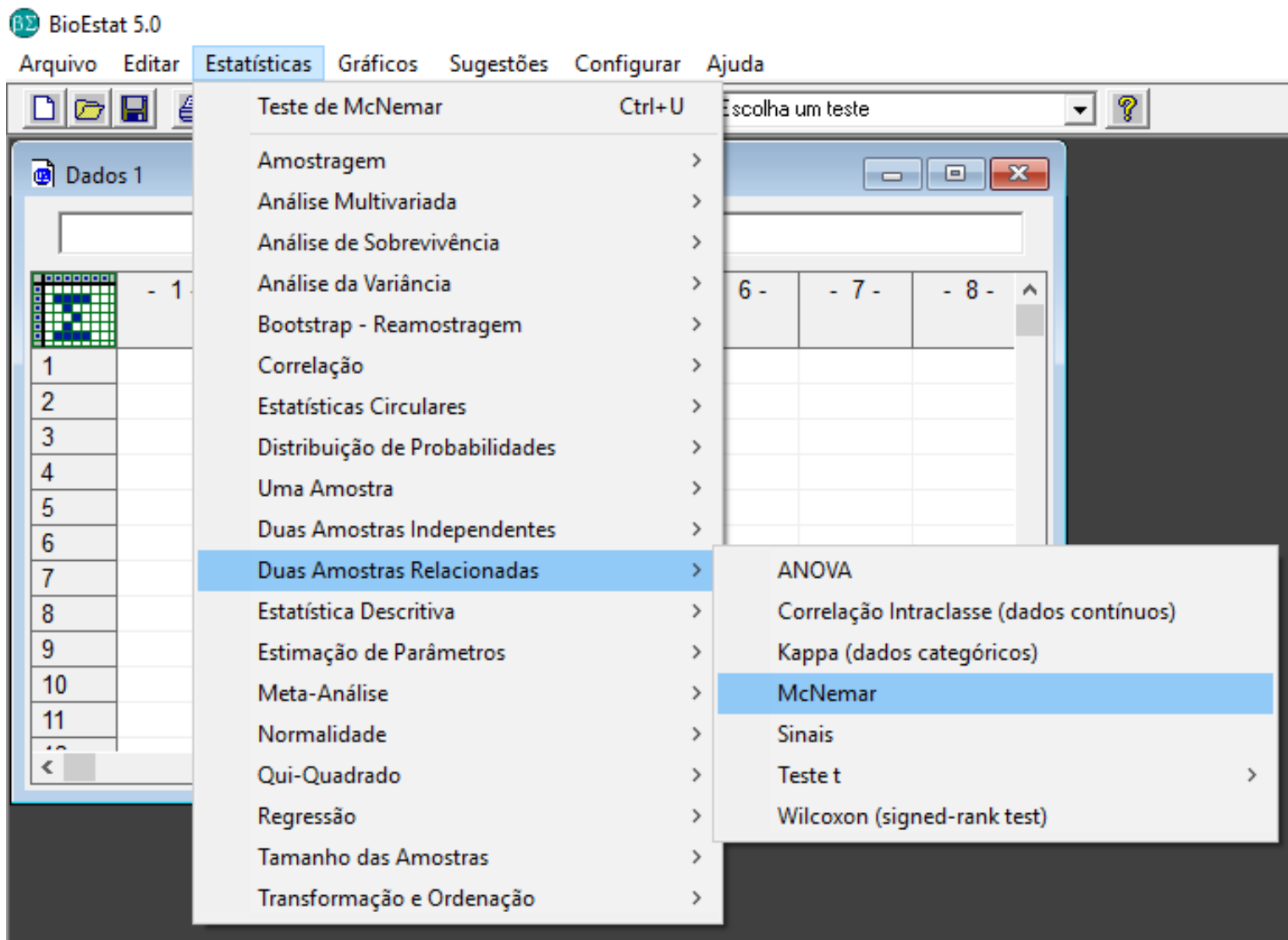
$$H_o \rightarrow p_a = p_d$$

$$H_1 \rightarrow p_a \neq p_d$$

2º Passo: Cálculos:



Exemplo 3 usando BioEstat 5.0



Exemplo 3 usando BioEstat 5.0

BioEstat 5.0

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Escolha um teste

Dados 1

	- 1 -	- 2 -	- 3 -
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			

Teste de McNemar

Imprimir

Dados

	A	B	Total / Linha
1	8	1	9
2	14	17	31
Total Coluna	22	18	40

Executar Sair

Resultados

Qui-Quadrado, se (A/D) são pares discordantes	Qui-Quadrado, se (B/C) são pares discordantes
—	—
p (A/D) =	p (B/C) =
—	0.0010

Calculado pelo Método Exato.



Exemplo 3

3º Passo: Decisão:

Como o valor calculado de $p = 0,001$ (0,1%) é menor do que 0,01 (1%), rejeita-se H_0 . Existe uma diferença altamente significativa entre o desempenho antes e depois da droga.

4º Passo: Conclusão:

Portanto, o emprego de Mesilato de Diidroergocristina provocou uma mudança no desempenho da memória dos pacientes, a um nível de significância de menos de 1%.

Rosimara Salgado

Professora
Coordenadora do NEaD

rosimara@inatel.br



Inatel

CAMINHOS
QUE CONECTAM
COM O FUTURO