BIOESTATÍSTICA

M.I. Eng. Biomédica

2015-2016

Aula Teórica 10

Teste bionomial para 1 amostra

- Exemplo:
  - Numa amostra de 10000 mulheres americanas, com idades compreendidas entre os 50 e os 54 anos, 400 delas cujas mães tinham tido cancro da mama apresentavam elas próprias cancro da mama.
  - Sabendo que a prevalência de cancro da mama é de 2% qual a relevância da prevalência de cancro da mama quando existem casos familiares prévios?

Teste bionomial para 1 amostra

- Exemplo:
  - Outra forma de colocar a questão é sob a forma de hipóteses:

$$\begin{cases} H_0: & p = 0,02 = p_0 \\ H_1: & p \neq 0.02 \end{cases}$$

Teste bionomial para 1 amostra

- O teste é então efectuado à proporção.
  - Assumindo uma aproximação à distribuição normal (o que é razoável, dado que $np_o q_o \geq 5$)

$$\hat{p} \sim N\left(p, \frac{p_0 q_0}{n}\right)$$

  - Mas, sob H_o

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$$

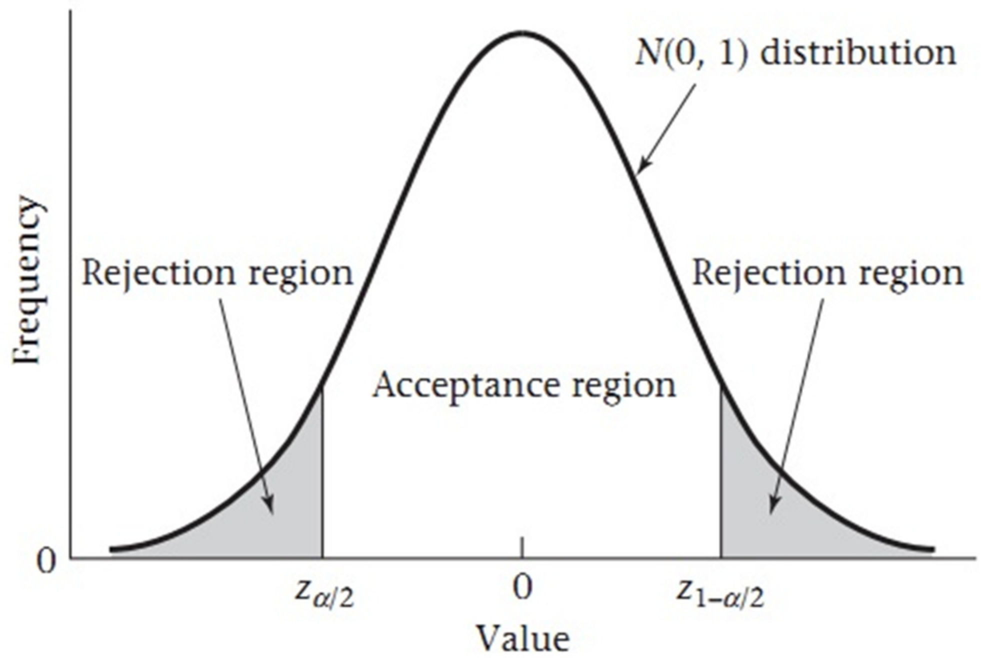
Teste bionomial para 1 amostra

- O teste é então:
  - Determinar z;
  - Se $z < z_{\alpha/2}$ ou $z > z_{1-\alpha/2}$ rejeita-se H_o;
  - Se $z_{\alpha/2} \leq z \leq z_{1-\alpha/2}$ não se rejeita H_o;

- Nota: o teste é válido se $np_oq_o \geq 5$

Teste bionomial para 1 amostra

- Graficamente:



in Fundamentals of Biostatistics, B. Rossner, pp 245

Teste bionomial para 1 amostra

- Método exacto:
  - Seja X uma variável aleatória que segue uma distribuição binomial de parâmetros n e p_o;
  - Seja $\hat{p} = x/n$, em que x é o número de eventos observados;

Teste bionomial para 1 amostra

- Método exacto
  - Se $\hat{p} \leq p_o$:

$$p/2 = P\left(\leq x \text{ sucessos em n tentativas} \mid H_0\right)$$

$$= \sum_{k=0}^{x} \binom{n}{k} p_0^k \left(1 - p_0\right)^{n-k}$$

Teste bionomial para 1 amostra

- Método exacto
  - Se $\hat{p} > p_o$:

$$p/2 = P\left(\geq x \text{ sucessos em n tentativas} | H_0\right)$$

$$= \sum_{k=x}^{n} \binom{n}{k} p_0^k \left(1 - p_0\right)^{n-k}$$

Teste bionomial para 1 amostra

- Exemplo de aplicação:

Suppose, for example, that 13 deaths have occurred among 55- to 64-year-old male workers in a nuclear-power plant and that in 5 of them the cause of death was cancer. Assume, based on vital-statistics reports, that approximately 20% of all deaths can be attributed to some form of cancer. Is this result significant?

in Fundamentals of Biostatistics, B. Rossner, pp 248

Teste bionomial para 1 amostra

- Exemplo de aplicação:

$$n\, p_0\, q_0 = 13 \times 0,2 \times 0,8 = 2,1 < 5$$

$$\hat{p} = \frac{5}{13} = 0,38 > 0,20$$

$$p = 2\sum_{k=5}^{13} \binom{13}{k} 0,2^k\, 0,8^{13-k} = 2\left[1 - \sum_{k=0}^{4} \binom{13}{k} 0,2^k 0,8^{13-k}\right] = 0,198$$

- Logo, a proporção de mortes por cancro não é significativamente diferente à obtida para outros trabalhadores da mesma idade;

Teste binomial para 2 amostras

- Exemplo:
  - Para testar a hipótese de que o risco de sofrer de cancro da mama pode estar relacionado com o intervalo de tempo entre a menarca e o 1º filho, foram constituídos dois grupos de mulheres com e sem cancro da mama. Em ambos os grupos foi determinada a proporção de mulheres que deu à luz o 1º filho depois dos 30 anos de idade.

Teste binomial para 2 amostras

- Exemplo:
  - Pretende-se saber se as proporções são diferentes nos dois grupos.

$$\begin{cases} H_0: & p_1 = p_2 = p \\ H_1: & p_1 \neq p_2 \end{cases} \quad \text{ou} \quad \begin{cases} H_0: & p_1 - p_2 = 0 \\ H_1: & p_1 - p_2 \neq 0 \end{cases}$$

Teste binomial para 2 amostras

- Sob H_o:

$$\hat{p}_1 \sim N\left(p, pq/n_1\right)$$

$$\hat{p}_2 \sim N\left(p, pq/n_2\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(0, pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

Teste binomial para 2 amostras

- Sob H_o:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p\,q\left(1/n_1 + 1/n_2\right)}} \sim N(0,1)$$

  - como p é desconhecido, o melhor estimador é

$$\hat{p} = \frac{n_1\,\hat{p}_1 + n_2\,\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

Teste binomial para 2 amostras

- Para melhor aproximação da distribuição normal à distribuição binomial usa-se uma correcção de continuidade

$$z = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{p\,q\left(1/n_1 + 1/n_2\right)}}$$

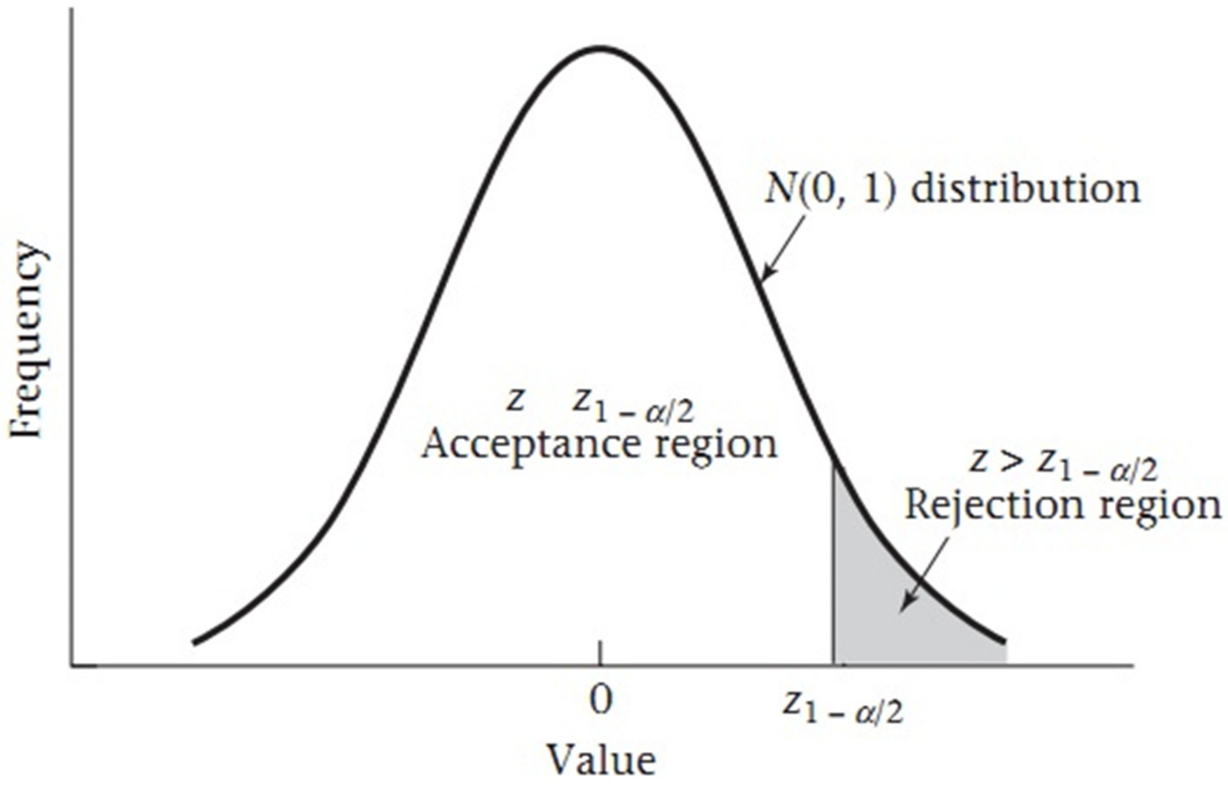
Teste binomial para 2 amostras

- O teste é então:
  - Determinar z;
  - Se $z > z_{1-\alpha/2}$ rejeita-se H_0;
  - Se $z \leq z_{1-\alpha/2}$ não se rejeita H_0;

• Nota: o teste é válido se $n_1\hat{p}\hat{q} \geq 5$ e $n_2\hat{p}\hat{q} \geq 5$

Teste binomial para 2 amostras

- Graficamente:



in Fundamentals of Biostatistics, B. Rossner, pp 355

Tabela de contigência

Data for the international study in Example 10.4 comparing age at first birth in breast-cancer cases with comparable controls

	Age at first birth		
Status	≥30	≤29	Total
Case	683	2537	3220
Control	1498	8747	10,245
Total	2181	11,284	13,465

Source: Reprinted with permission from *WHO Bulletin*, *43*, 209–221, 1970.

in Fundamentals of Biostatistics, B. Rossner, pp 357

Tabela de contigência

- A significância estatística resulta da comparação entre os valores observados e valores esperados;
- A tabela de valores esperados é calculada a partir dos valores marginais;
  - O valor esperado na célula, $E_{i,j}$, é obtido pelo produto da margem da linha i pela margem da coluna j, dividido pelo total;

Teste de qui-quadrado

- Resumindo:
  - Determinar

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(\left|O_{ij} - E_{ij}\right| - 0{,}5\right)^2}{E_{ij}} \sim \chi_1^2 \qquad E_{ij} = \frac{L_i \times C_j}{N}$$

  - Se $X^2 > \chi^2_{1,\,1-\alpha}$ rejeita-se H_o
    - Pode ser aplicado se nenhum $E_{ij} < 5$

Teste Exacto de Fisher

- Quando não é possível usar a aproximação à distribuição normal, é possível, para tabelas 2x2, determinar os níveis exactos de significância;
- Fixando os valores marginais pode determinar-se a probabilidade exacta de observar uma determinada tabela com determinadas margens;

Teste Exacto de Fisher

- Dada a tabela:

Variável	Amostras ou grupos		Linhas
	1	2	
1	a	b	a+b
2	c	d	c+d
Colunas	a+c	b+d	N

Teste Exacto de Fisher

- A probabilidade de observar a tabela é:

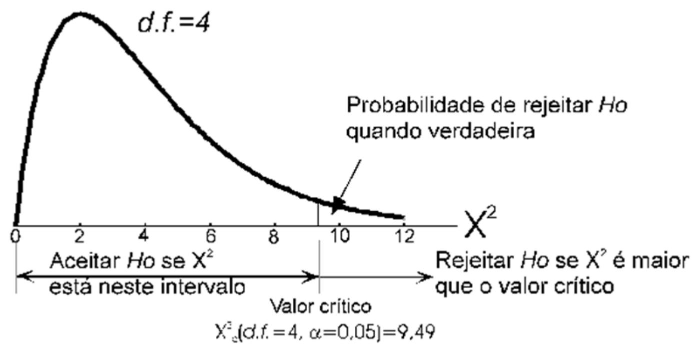
$$P(a,b,c,d)=\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Tabela de contigência LxC

- Teste Qui-quadrado (χ^2)

$$Z = \sum_{i=1}^{L}\sum_{j=1}^{C}\frac{\left(O_{i\,j} - E_{i\,j}\right)^2}{E_{i\,j}} \sim \chi^2_{(L-1)\times(C-1)} \qquad E_{i\,j} = \frac{L_i \times C_j}{N}$$

- Só pode ser aplicado se (regras Cochran):
  - $E_{ij} > 1$
  - Pelo menos 80% $E_{ij} \geq 5$



Métodos não paramétricos

- Testes paramétricos – dados cujas distribuições subjacentes são conhecidas (binomial, normal);
- Teste não-paramétricos – a distribuição da população não é conhecida (ou a família especificada é posta é dúvida) e as inferências processam-se em quadro menos restritivo;
  - Podem aplicar-se em variáveis ordinais
  - Comparam-se postos, ordenações e medianas

Teste do sinal

- O teste de sinal é um dos teste mais conhecidos e dá uma resposta independente da função de distribuição.
- Seja μ_e a mediana de uma população que como é usual:
  - E considere-se o teste

$$H_0 : \mu_e = \mu_{e0}$$
$$H_1 : \mu_e \neq \mu_{e0}$$

  - Realize-se ainda a transformação:

$$(X_1, X_2, \ldots, X_n) \to (Z_1, Z_2, \ldots, Z_n), Z_i = X_i - \mu_{e0}$$

Teste do sinal

- O teste pode então ser reformulado:

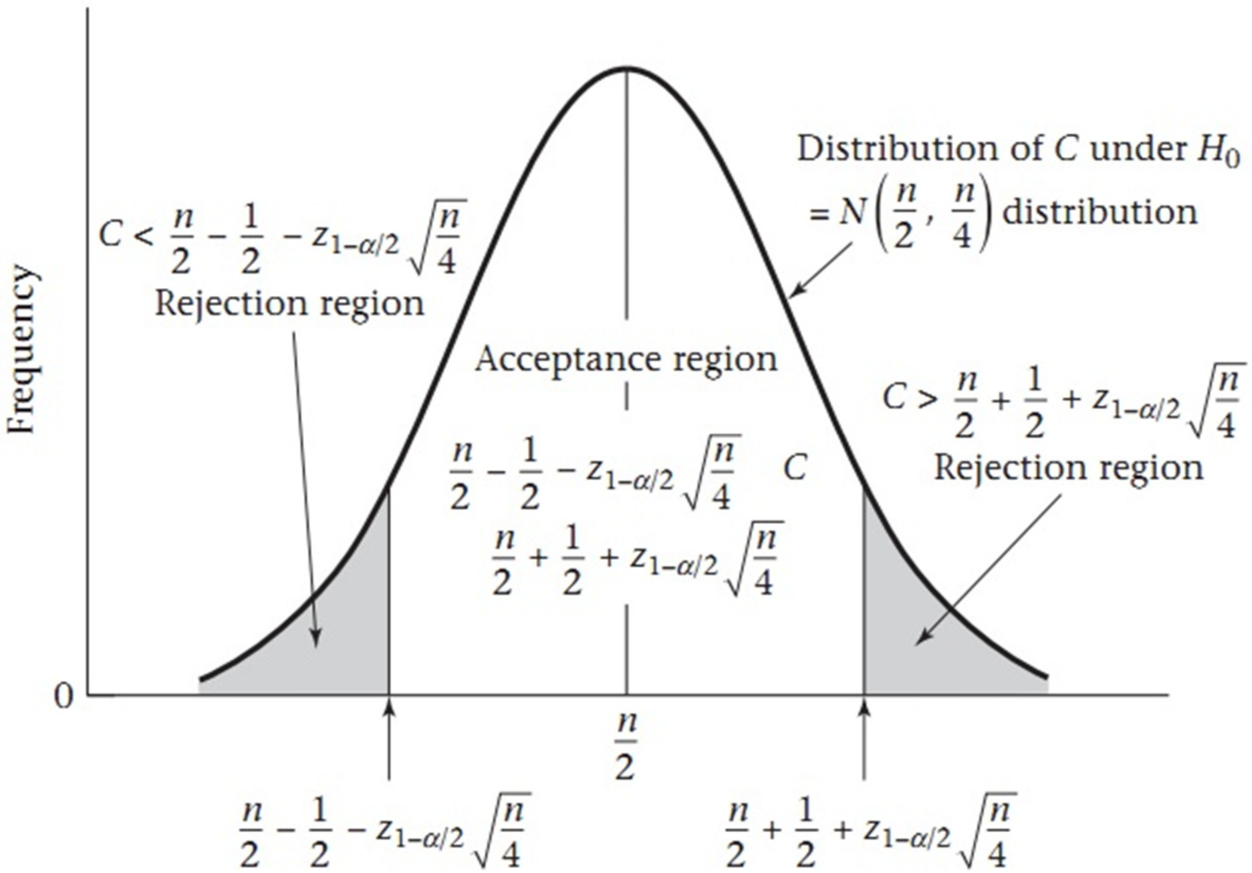
$$H_0 : \Delta = 0$$
$$H_1 : \Delta \neq 0$$

- Com Δ a mediana de Z.
- Para n ≥ 20 e C = número de $Z_i > 0$
- Se

$$C > c_2 = \frac{n}{2} + \frac{1}{2} + z_{1-\alpha/2}\sqrt{n/4} \qquad \text{ou}$$
$$C < c_1 = \frac{n}{2} - \frac{1}{2} - z_{1-\alpha/2}\sqrt{n/4}$$

- H_0 é rejeitada.

Teste do sinal



in Fundamentals of Biostatistics, B. Rossner, pp 330

Teste de Wilcoxon

- O teste de Wilcoxon ou de ordem-sinal utiliza-se para testar uma hipótese sobre a mediana quando se considera a distribuição simétrica.
- O teste de Wilcoxon representa uma melhoria relativamente ao teste de sinal pois não despreza a informação contida na ordem das diferenças.
- O teste de Wilcoxon baseia-se na estatística do mesmo nome e baseia-se na obtenção das diferenças e da sua ordem.

Teste de Wilcoxon

- Seja $Z_i = 1$ ($Z_i = 0$) se o i-ésimo valor da sucessão ordenada dos módulos está associado a uma diferença Z_i positiva (negativa) e

$$Z = \sum_{i=1}^{n} Z_i \qquad R = \sum_{i=1}^{n} i\, Z_i$$

- Quando existem valores iguais de $|Z_i|$ (empates), atribui-se a cada um deles um número de ordem igual à média das ordens que lhe caberia.

- $|Zi|$ = 1.1, 1.1, 2.2, 3.4,3.4,3.4
- As ordens neste caso seriam
- 1.5,1.5,3,5,5,5

Teste de Wilcoxon

- Se R ≠ n(n+1)/4 e não há empates então

$$T = \left[\left| R - \frac{n(n+1)}{4} \right| - \frac{1}{2} \right] \Big/ \sqrt{n(n+1)(2n+1)/24}$$

- Se R ≠ n(n+1)/4 e há empates então

$$T = \left[\left| R - \frac{n(n+1)}{4} \right| - \frac{1}{2} \right] \Big/ \sqrt{n(n+1)(2n+1) \Big/ 24 - \sum_{i=1}^{g} \left(t_i^3 - t_i \right) \Big/ 48}$$

- com t_i o número de diferenças com o mesmo valor no grupo empatado i e g é o número de empates.

- Se R = n(n+1)/4 então T = 0.

Teste de Wilcoxon

- Finalmente, se

$$T > z_{1-\alpha/2}$$

- Então rejeita-se H_o.

- O valor de p é dado por:

$$p = 2 \times \left[1 - \Phi(T)\right]$$

- Nota: este teste deve ser usado se o número de diferenças não nulas é superior a 16 e se a distribuição subjacente é simétrica