BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP USP

BIOESTATÍSTICA

II. Teste de hipóteses para a proporção populacional p

GLEICE M S CONCEIÇÃO

MARIA DO ROSÁRIO D D LATORRE

FSP USP

Inferência

$$X \sim Bernoulli(p), \text{ i.e., } \quad X = \begin{cases} 1 \ \cdots sucesso \quad \cdots \quad p \\ 0 \ \cdots fracasso \ \cdots 1-p \end{cases} \quad \Rightarrow \quad \begin{array}{l} E(X) = p \\ \text{VAR}(X) = p(1-p) \end{array}$$

BIOESTATÍSTICA

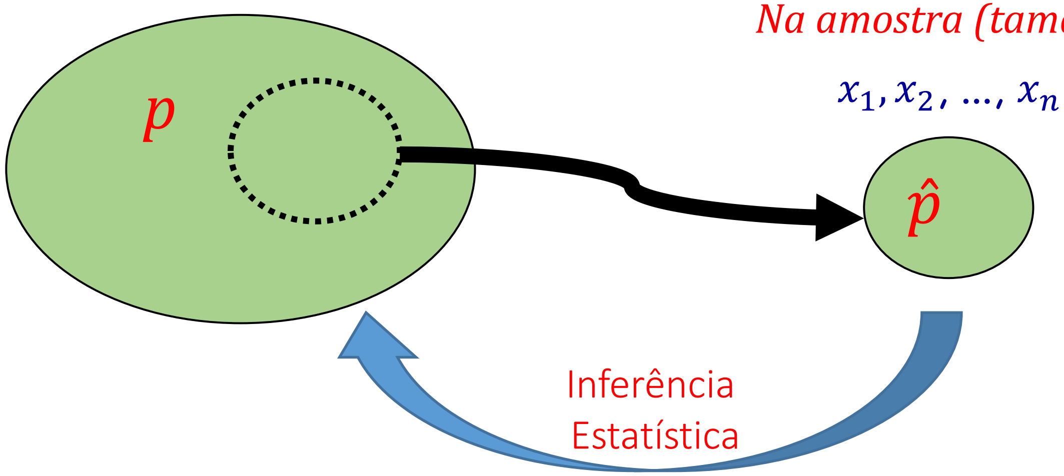
- ✓ *"Sucesso" significa o indivíduo ser portador de uma determinada característica de interesse na população (p. ex., tuberculose, asma, COVID-19, etc.).*
- ✓ *Ou seja, p é a probabilidade (constante) de um indivíduo da população ter a característica.*

Na população (tamanho N):

$X_1, X_2, \ldots, X_N$

Na amostra (tamanho n):

$x_1, x_2, \ldots, x_n$



- ✓ Tirar conclusões para a população com base na amostra
- ✓ Necessário associar probabilidades a essas conclusões

Além disso,

- ✓ *a proporção de indivíduos na população com a característica é*

$$p = \frac{\sum_{i=1}^{N} X_i}{N}$$

- ✓ *a proporção de indivíduos na amostra com a característica é*

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Inferência

$$X \sim Bernoulli(p), \text{ i.e., } \quad X = \begin{cases} 1 \cdots sucesso & \cdots \quad p \\ 0 \cdots fracasso & \cdots 1-p \end{cases} \quad \Rightarrow \quad \begin{array}{l} E(X) = p \\ \text{VAR}(X) = p(1-p) \end{array}$$

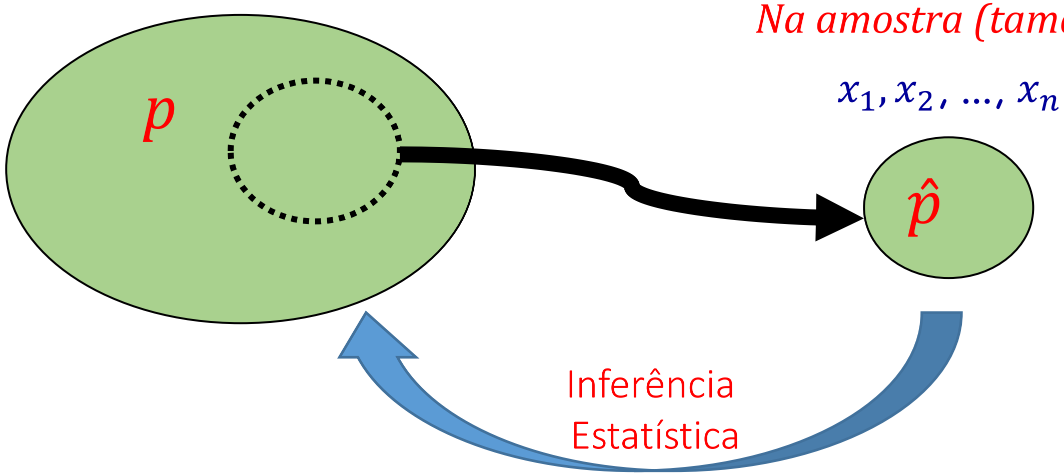
- ✓ *"Sucesso" significa o indivíduo ser portador de uma determinada característica de interesse na população (p. ex., tuberculose, asma, COVID-19, etc.).*
- ✓ *Ou seja, p é a probabilidade (constante) de um indivíduo da população ter a característica.*

Na população (tamanho N):

$X_1, X_2, \ldots, X_N$

Na amostra (tamanho n):

$x_1, x_2, \ldots, x_n$



- ✓ Tirar conclusões para a população com base na amostra
- ✓ Necessário associar probabilidades a essas conclusões

Inferência Estatística:

Tomar decisões sobre p, com base em $\hat{p}$:

- ✓ $IC(p, \gamma)$ *- intervalo de confiança para p*
- ✓ *Teste de hipóteses para p*

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Teste de hipóteses

BIOESTATÍSTICA

- ✓ *Já aprendemos a obter uma estimativa pontual e estimativas por intervalo para o parâmetro populacional p.*
- ✓ *Agora, vamos aprender como tomar decisões sobre o parâmetro populacional p quando conhecemos apenas seu estimador amostral $\hat{p}$.*
- ✓ *De novo, vamos tomar decisões sobre a população quando conhecemos apenas a amostra (tomar decisões para o todo, com base em apenas uma parte).*
- ✓ *Essas decisões estão sujeitas a erros - necessário conhecer a probabilidade de cometer tais erros e associar essas probabilidades às nossas decisões.*

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Teste de hipóteses

- ✓ *Só será possível se pudermos associar uma distribuição de probabilidades conhecida ao estimador. Qual?*
- ✓ *Sim , a Normal.*
- ✓ *Como sempre, vamos utilizar o TLC.*

Lembrando ...

Teorema Limite Central (TLC)

BIOESTATÍSTICA

- ✓ $X \sim Bernoulli(p)$, i.e., $X = \begin{cases} 1 \cdots sucesso & \cdots \quad p \\ 0 \cdots fracasso & \cdots 1 - p \end{cases}$
- ✓ $E(X) = p, \; VAR(X) = p(1-p)$
- ✓ *p é a probabilidade (constante) de um indivíduo da população ter a característica e é também a proporção de indivíduos na população com a característica:*

$$p = \frac{\sum_{i=1}^{N} X_i}{N}, \text{ ou seja, } p \text{ é a média populacional } (p = \mu)$$

- ✓ *$\hat{p}$ é a proporção de indivíduos na amostra com a característica*

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}, \text{ ou seja, } \hat{p} \text{ é a média amostral } (\hat{p} = \bar{X})$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Lembrando ...

Teorema Limite Central (TLC)

BIOESTATÍSTICA

- ✓ $X \sim Bernoulli(p)$, i.e., $X = \begin{cases} 1 & \cdots sucesso & \cdots p \\ 0 & \cdots fracasso & \cdots 1-p \end{cases}$
- ✓ $E(X) = p,\ VAR(X) = p(1-p)$
- ✓ Vimos, pelo TLC, que $\bar{X} \xrightarrow{n \to \infty} N\left(E(X), \frac{VAR(X)}{n}\right)$
- ✓ Consequentemente, $\hat{p} \xrightarrow{n \to \infty} N\left(p, \frac{p(1-p)}{n}\right)$
- ✓ Isto é, para n grande:
  - $\hat{p}$ tem distribuição Normal
  - Com esperança (média) igual à da variável original $\rightarrow E(X) = p$
  - Com variância igual à da variável original dividida por n $\rightarrow VAR(X) = \frac{p(1-p)}{n}$

Lembrando ...

Teorema Limite Central (TLC)

Então, para *n* grande:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$\Rightarrow Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Já temos uma estatística para testar hipóteses!

II. Teste de hipóteses para a proporção populacional *p*

Vamos utilizar os mesmos passos descritos anteriormente para construir um teste de hipóteses para a proporção:

1. Definir a(s) variável(eis) de interesse e o(s) parâmetro(s) que a(s) caracteriza(m).

$$X \sim Bernoulli(p), \text{ i.e.,} \quad X = \begin{cases} 1 & \cdots \, sucesso \quad \cdots \quad p \\ 0 & \cdots \, fracasso \quad \cdots 1-p \end{cases} \quad \Rightarrow \quad \begin{array}{l} E(X) = p \\ \mathrm{VAR}(X) = p(1-p) \end{array}$$

em que sucesso significa o indivíduo ser portador de uma determinada característica de interesse

2. Estabelecer as hipóteses nula e alternativa

$$H_o: p = p_0 \qquad \text{x} \qquad \begin{cases} H_a: p \neq p_0 \text{ ou} \\ H_a: p > p_0 \text{ ou} \\ H_a: p < p_0 \text{ ou} \end{cases}$$

3. Escolher o estimador que será usado e definir a forma da região crítica, com base na hipótese alternativa, p.ex., para Ha: $p > p_0$:

$$RC = \{\hat{p} \geq \hat{p}_c\}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

II. Teste de hipóteses para a proporção populacional *p*

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador; definir a estatística do teste (que deve conter parâmetro e estimador e ter uma distribuição conhecida) e sua distribuição. Especificar as suposições assumidas.

- ✓ Parâmetro: p
- ✓ Estimador e distribuição do estimador: $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$,
- ✓ Estatística do teste e sua distribuição: $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0,1)$
- ✓ Suposições assumidas: À princípio, nenhuma.
  Se n for suficientemente grande, $\hat{p} \sim N$

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

II. Teste de hipóteses para a proporção populacional *p*

5. Fixar α

Três soluções possíveis:

6. Obter a região crítica com base no valor de $\hat{p}_c$

$$\alpha = P(Erro\ I) = P(rejeitar\ H_0 \mid H_0\ \text{é}\ verdadeira)$$

$$= P(\hat{p} \geq \hat{p}_c \mid p = p_0) \quad \Rightarrow \hat{p}_c = \ ...$$

$$\Rightarrow RC = \{\hat{p} \geq \hat{p}_c\}$$

7. Tomar a decisão, comparando o valor de $\hat{p}_{obs}$ com a região crítica

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

II. Teste de hipóteses para a proporção populacional *p*

BIOESTATÍSTICA

Três soluções possíveis:

ou

6. Obter a região crítica com base no valor de z_c
7. Tomar a decisão, comparando o valor de z_{obs} com a região crítica

ou

6. Obter o nível descritivo (p-valor)
7. Tomar a decisão, comparando o p-valor com o valor de α

Exemplo

Um tratamento para diminuir a dor utilizando uma droga padrão é eficaz em 65% dos casos em que é utilizado. Um médico afirma que o tratamento com acupuntura pode produzir melhores resultados.

Para testar tal hipótese, 50 pacientes foram selecionados aleatoriamente e submetidos ao tratamento com acupuntura. Sabendo que, ao final do tratamento, 39 pacientes ficarem livres da dor:

a) Construa um intervalo de confiança otimista para p, com um coeficiente de confiança igual a 95%. Interprete-o e tire conclusões iniciais.

b) Conduza o teste de hipóteses.

Exemplo

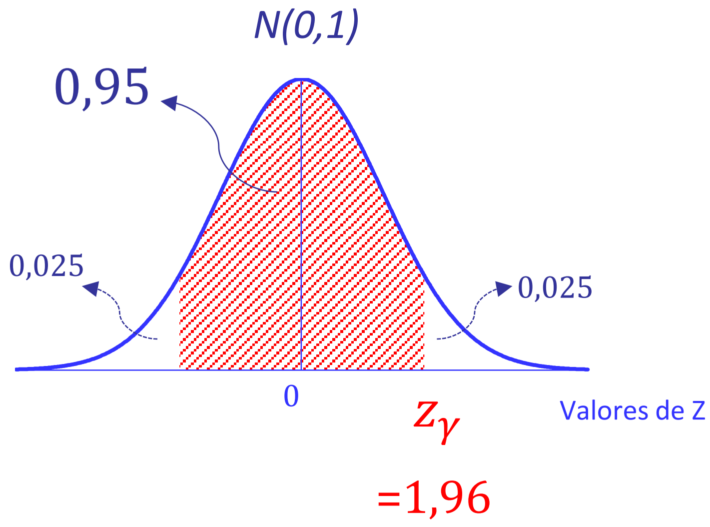
a) Intervalo de confiança otimista para p:

$$IC(p,\gamma) = \hat{p} \pm z_\gamma \sqrt{\frac{\hat{p}\,(1-\hat{p}\,)}{n}}$$

$$\hat{p} = \frac{39}{50} = 0{,}78$$

$$IC(p,\gamma) = 0{,}78 \pm 1{,}96 \sqrt{\frac{0{,}78 * 0{,}22}{50}}$$

$$= 0{,}78 \pm 0{,}115 \; = [0{,}665 : 0{,}895]$$



- ✓ Interpretação?
- ✓ Conclusões iniciais?

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exemplo

b) Teste de hipóteses

1. Definir a(s) variável(eis) de interesse e o(s) parâmetro(s) que a(s) caracteriza(m).

X – paciente tratado com acupuntura livre da dor (1=sim, 0=não),

$E(X) = p$ e $VAR(X) = p(1-p)$

2. Estabelecer as hipóteses nula e alternativa e interpretá-las.

$H_0: p = 0{,}65$ (o tratamento com acupuntura não é melhor do que o padrão)

$H_a: p > 0{,}65$ (o tratamento com acupuntura é melhor do que o padrão)

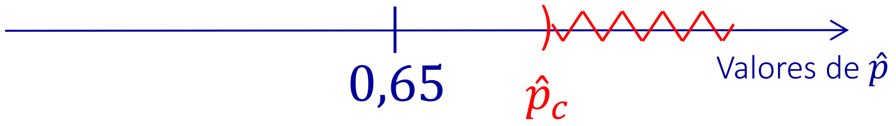
GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exemplo

b) Teste de hipóteses

3. Definir a forma da região crítica, com base na hipótese alternativa

- ✓ $H_a\!: p > 0{,}65$



$$RC = \{\bar{X} \in \mathcal{R} \mid \hat{p} \geq \hat{p}_c\}$$

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

II. Teste de hipóteses para a proporção populacional *p*

BIOESTATÍSTICA

b) Teste de hipóteses

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador; definir a estatística do teste (que deve conter parâmetro e estimador e ter uma distribuição conhecida) e sua distribuição. Especificar as suposições assumidas.

- ✓ Parâmetro: p
- ✓ Estimador e distribuição do estimador: $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$,
- ✓ Estatística do teste e sua distribuição: $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0,1)$
- ✓ Suposições assumidas: À princípio, nenhuma.
  Se n for suficientemente grande, $\hat{p} \sim N$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exemplo

b) Teste de hipóteses

5. Fixar α.

$\alpha = 0{,}05$

6. i) Obter a região crítica com base no valor de $\hat{p}_c$

$$\alpha = P(Erro\ I) = P(rejeitar\ H_0 \mid H_0\ é\ verdadeira)$$

$$\alpha = P(\hat{p} \geq \hat{p}_c \mid p = 0{,}65)$$

$$0{,}05 = P(\hat{p} \geq \hat{p}_c \mid p = 0{,}65)$$

$$0{,}05 = P\left(\frac{\hat{p} - 0{,}65}{\sqrt{\frac{0{,}65\ x\ 0{,}35}{50}}} \geq \frac{\hat{p}_c - 0{,}65}{\sqrt{\frac{0{,}65\ x\ 0{,}35}{50}}}\right) = P\left(Z \geq \frac{\hat{p}_c - 0{,}65}{\sqrt{\frac{0{,}65\ x\ 0{,}35}{50}}}\right)$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

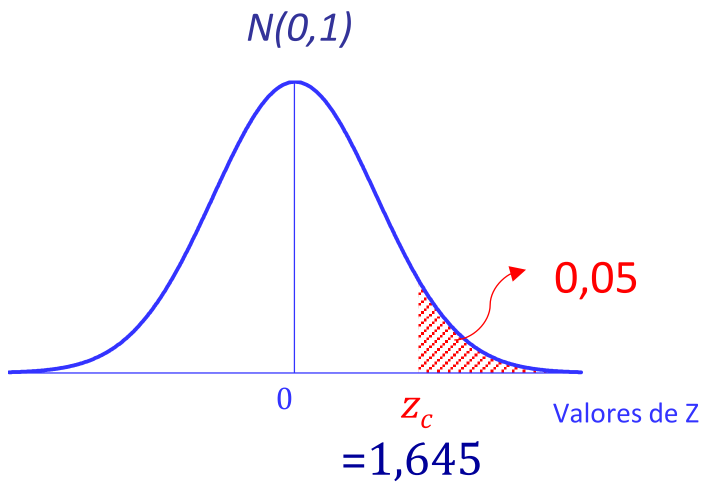
Exemplo

b) Teste de hipóteses

Da Tabela da N(0,1):

$$\frac{\hat{p}_c - 0,65}{\sqrt{\frac{0,65 \ x \ 0,35}{50}}} = 1,645 \Rightarrow \hat{p}_c = 0,7601$$

Logo, RC=$\{\hat{p} \in \mathcal{R} \mid \hat{p} \geq 0,7601\}$



7. i) Tomar a decisão, comparando o valor de $\hat{p}_{obs}$ com a região crítica

Como $\hat{p}_{obs}$ =39/50=0,78, $\hat{p}_{obs} \in RC$, então rejeito H_0 e decido que a proporção de cura populacional do tratamento com acupuntura é maior do que 0,65 (isto é, o tratamento com acupuntura é melhor do que o padrão).

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exemplo

b) Teste de hipóteses

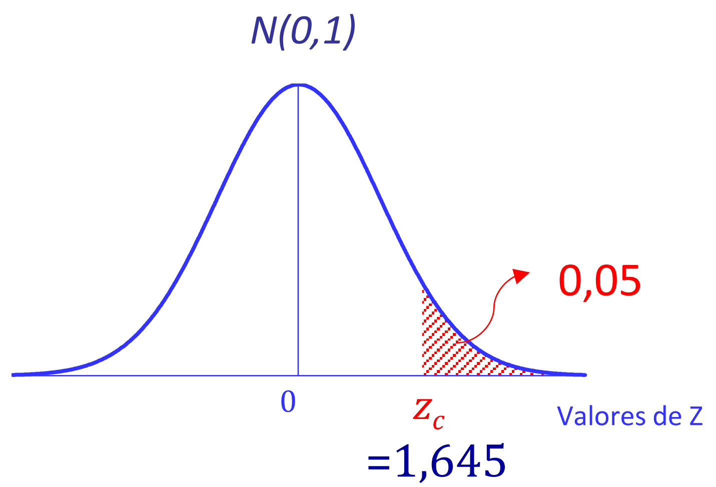
6. ii) Obter a região crítica com base no valor de z_c

Da Tabela da N(0,1):

RC={Z ∈ $\mathcal{R}$ | $Z \geq 1{,}645$}

Sob Ho, o valor de Z observado é:

$$Z_{obs} = \frac{\hat{p}_{obs} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0{,}78 - 0{,}65}{\sqrt{\frac{0{,}65 \, x \, 0{,}35}{50}}} = 1{,}927$$



7. ii) Tomar a decisão, comparando o valor de z_{obs} com a região crítica

Como $Z_{obs} = 1{,}927$, $Z_{obs} \in RC$, então rejeito H_0 e decido que a proporção de cura populacional do tratamento com acupuntura é maior do que 0,65 (isto é, o tratamento com acupuntura é melhor do que o padrão).

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exemplo

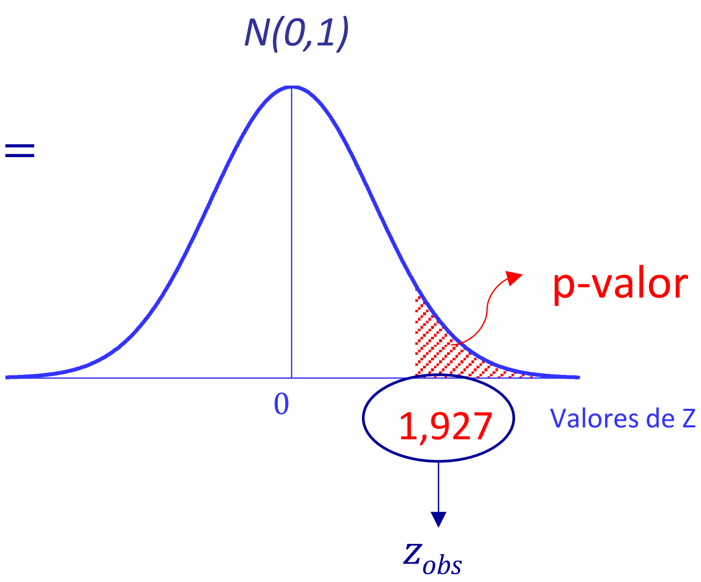
b) Teste de hipóteses

6. ii) Obter o p-valor

$$p - valor = P(\hat{p} \geq \hat{p}_{obs} | H_0 \text{ é } verdadeira) =$$

$$= P(\hat{p} \geq 0{,}78 | p = 0{,}65) = P\left(\frac{\hat{p} - 0{,}65}{\sqrt{\frac{0{,}65 \ x \ 0{,}35}{50}}} \geq \frac{0{,}78 - 0{,}65}{\sqrt{\frac{0{,}65 \ x \ 0{,}35}{50}}}\right) =$$

$$= P(Z \geq \underbrace{1{,}927}_{z_{obs}}) \quad = 1 - 0{,}9732 = 0{,}0268$$



7. ii) Tomar a decisão, comparando o p-valor com o valor de α

Como p-valor =0,0268, p-valor <α, então rejeito H_0 e decido que a proporção de cura populacional do tratamento com acupuntura é maior do que 0,65 (isto é, o tratamento com acupuntura é melhor do que o padrão).

Exemplo

b) Teste de hipóteses

Note que:

$$p - valor = 0{,}0268$$

Então:

- ✓ Para α=0,05, rejeito H_0 e decido que a proporção de cura populacional do tratamento com acupuntura é maior do que 0,65 (isto é, o tratamento com acupuntura é melhor do que o padrão).
- ✓ Mas, para α=0,01, não rejeito H_0 e decido que a proporção de cura populacional do tratamento com acupuntura é igual a 0,65 (isto é, o tratamento com acupuntura não é melhor do que o padrão).

Exercícios extras

5) Certa comunidade apresentou num período de vários anos incidência da doença X de 12 por 10.000 hab.. Em 1999, a incidência foi de 70 casos e a população estimada foi igual a 50.000 habitantes. Nestas condições, em um nível de significância de 1% (ou mais próximo) diga se concorda com as autoridades sanitárias que consideraram a situação dentro do esperado.

6) Desejando-se conhecer a prevalência de determinada doença na cidade A, selecionou-se uma amostra aleatória de 500 pessoas. Nesta amostra encontrou-se 20 doentes. Teste a hipótese de que a prevalência é semelhante ao que é descrito na literatura (p=10%).(α=5%)

7) Para se determinar a letalidade da doença B , acompanhou-se uma amostra de 30 doentes durante um ano. Após esse período, 5 deles haviam morrido. Testar a hipótese de que essa letalidade é igual a 20%.(α=10%)

8) Em uma amostra de 88 pacientes atendidos no ambulatório do Departamento de Oncologia Clínica, verificou-se que 38 eram fumantes. Teste a hipótese de que a porcentagem de fumantes atendidos neste ambulatório é igual ao referido na literatura (50%).(α=4%)

9) Estima-se que um medicamento A provoque efeitos colateral em 55% dos pacientes. Deseja-se testar se uma nova droga tem menos efeitos colaterais que A. Para tanto, tratou-se 50 pacientes com a nova droga e 30 deles apresentaram efeitos colaterais. Há diferença entre as proporções de pacientes com efeitos colaterais nos dois medicamentos? (α=1%)

10) Sabe-se que na cidade Y, 40% dos homens são obesos. Estudou-se uma amostra de 200 mulheres desta mesma cidade e verificou-se que havia 50 obesas. A prevalência de obesos entre os homens é igual à das mulheres? (α=2%)

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Gabarito

BIOESTATÍSTICA

5)

$$\begin{cases} H_0 : \pi_{85} = \pi_{anterior} \\ H_a : \pi_{85} \neq \pi_{anterior} \end{cases}$$

$\alpha \cong 1\%$; $z_{crítico} = \pm 2{,}58$; $z_{observado} = 1{,}29 \Rightarrow$ **Decisão** : Aceita-se H_0

Conclusão : Não há evidência estatística para discordar das autoridades sanitárias, com nível de significância próximo a 1%.

6) Rejeita Ho
7) Aceita Ho
8) Aceita Ho
9) Aceita Ho
10) Rejeita Ho

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP