BIOESTATÍSTICA

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

BIOESTATÍSTICA

ESTIMAÇÃO

GLEICE M S CONCEIÇÃO

MARIA DO ROSÁRIO D O LATORRE

FSP USP

O que vamos aprender

BIOESTATÍSTICA

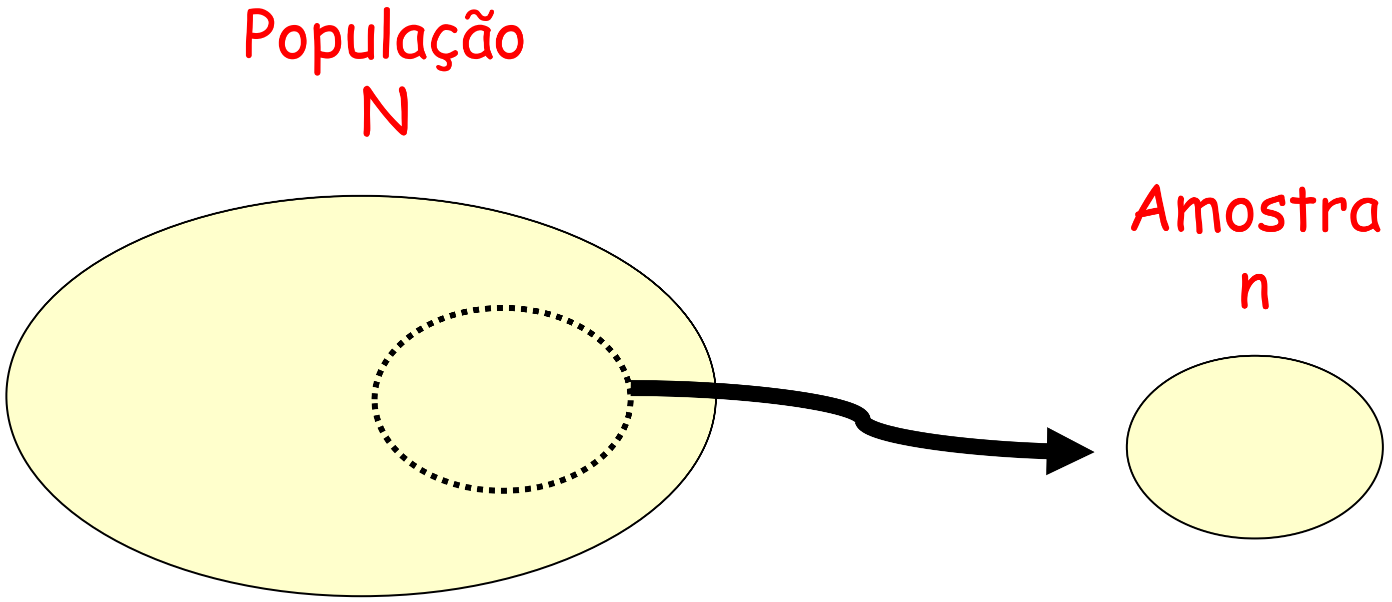
- ✓ Algumas considerações sobre população e amostra
- ✓ Diferença entre parâmetros e estimadores
- ✓ Fundamental para entender técnicas de inferência estatística:
    - ✓ Estimativa por ponto e por intervalo (Intervalos de Confiança)
    - ✓ Teste de Hipóteses

População e amostra

BIOESTATÍSTICA

Medidas populacionais são quase sempre desconhecidas.

Na maioria das vezes não é possível estudar a população toda.



A Estatística permite tirar conclusões sobre a população a partir do estudo de alguns de seus elementos (amostra).

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

População e amostra

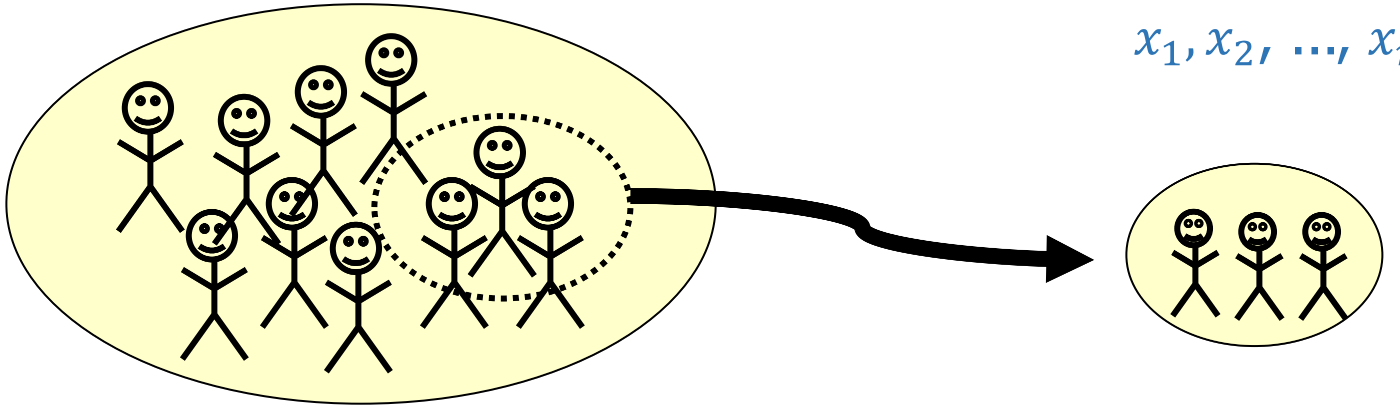
Seja X uma variável aleatória (quantitativa) de interesse

Na população (tamanho N):

$X_1, X_2, \ldots, X_N$

Na amostra (tamanho n):

$x_1, x_2, \ldots, x_n$



μ é a média de X na população $= E(X)$

σ^2 é a variância de X na população $= VAR(X)$

$\bar{X}$ é a média de X na amostra

S^2 é a variância de X na amostra

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

População e amostra

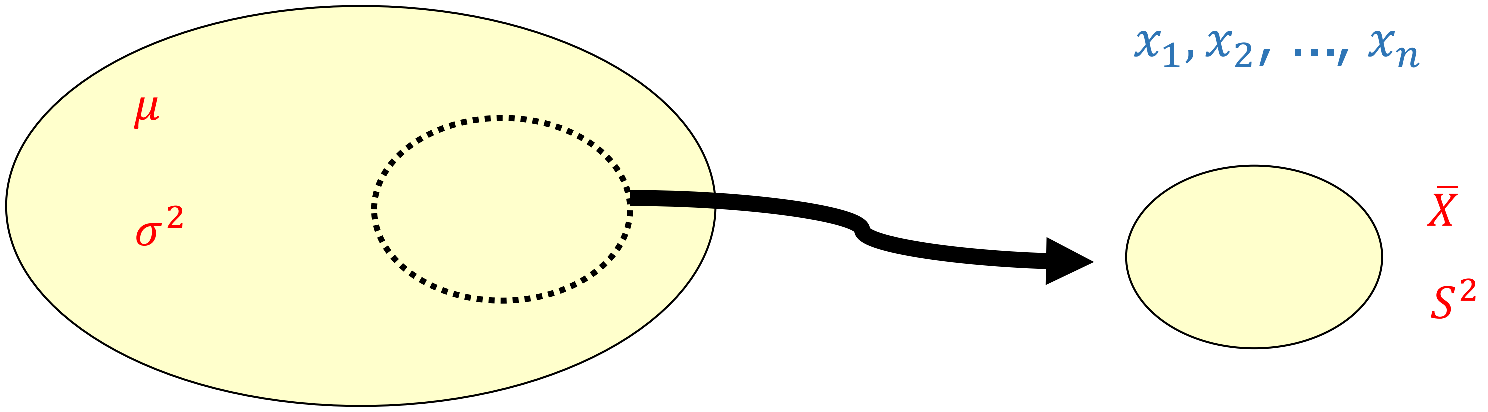
Seja X uma variável aleatória (quantitativa) de interesse

Na população (tamanho N):

$$X_1, X_2, \ldots, X_N$$

Na amostra (tamanho n):

$$x_1, x_2, \ldots, x_n$$



$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}$$

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

Parâmetros x Estimadores

BIOESTATÍSTICA

μ, σ^2 são **parâmetros** populacionais, constantes.

$\bar{X},\ S^2$ são estatísticas da amostra,

estimadores dos parâmetros populacionais,

são variáveis aleatórias.

→ $\bar{X}$ é um **estimador** para μ

→ S^2 é um **estimador** para σ^2

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

Estimação

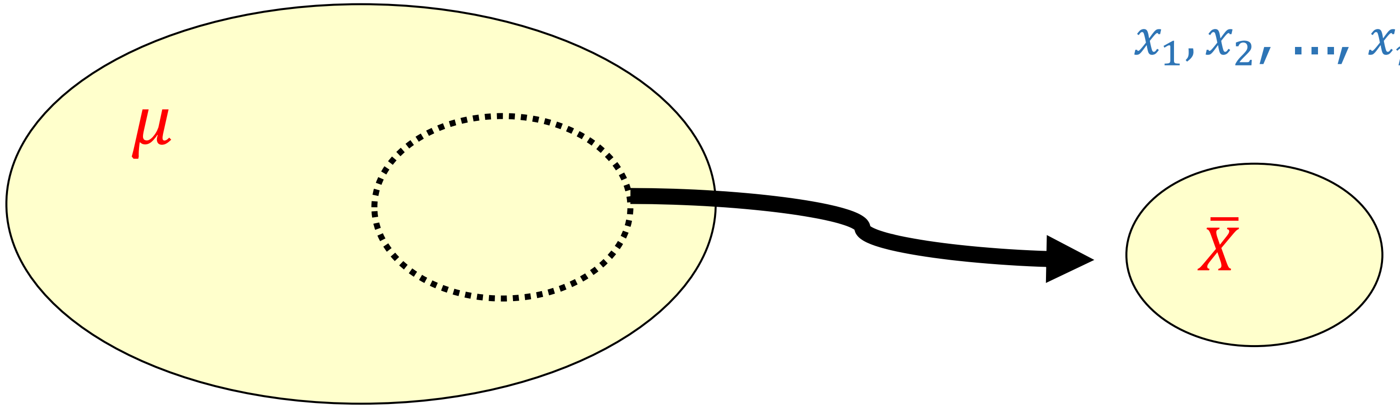
Seja X uma variável aleatória (quantitativa) de interesse

Na população (tamanho N):

$X_1, X_2, \ldots, X_N$

Na amostra (tamanho n):

$x_1, x_2, \ldots, x_n$



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

Inferência

BIOESTATÍSTICA

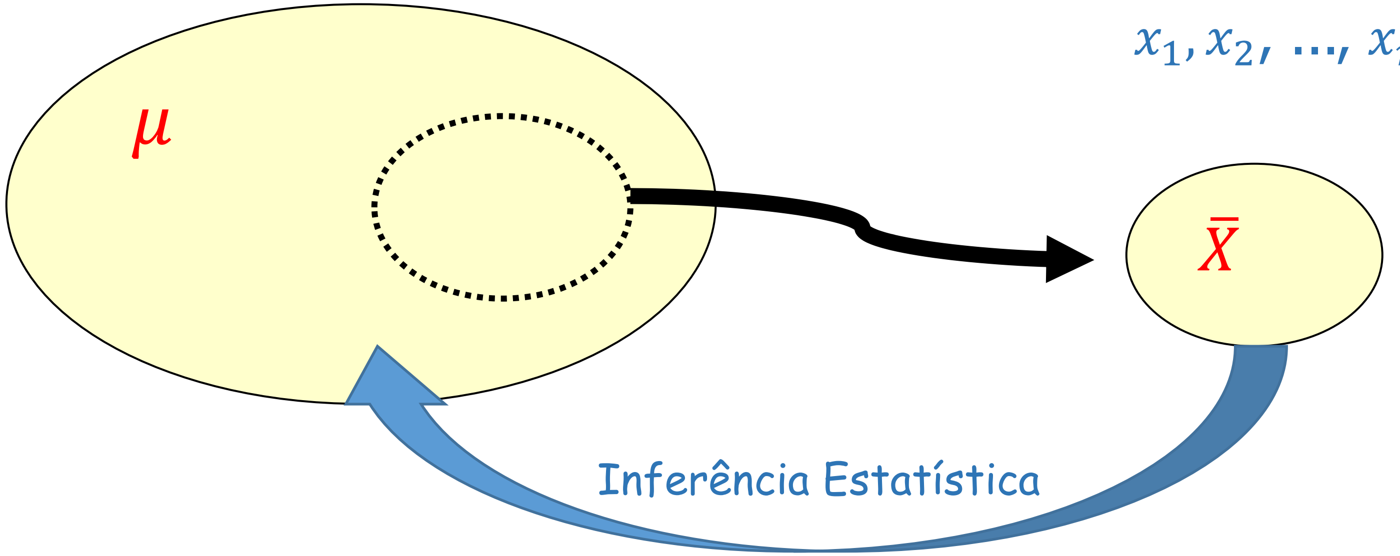
Seja X uma variável aleatória (quantitativa) de interesse

Na população (tamanho N):

$X_1, X_2, \ldots, X_N$

Na amostra (tamanho n):

$x_1, x_2, \ldots, x_n$



Inferência Estatística:

Tomar decisões sobre μ, com base em $\bar{X}$:

- ✓ $IC(\mu, \gamma)$ - intervalo de confiança para μ
- ✓ Teste de hipóteses para μ

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

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

Inferência

Tirar conclusões para a população com base na amostra

- ✓ *Se conhecêssemos o "todo" (a população) , poderíamos tirar conclusões sobre qualquer uma de suas "partes" (amostras) sem cometer nenhum erro.*
- ✓ *Mas estamos indo na direção contrária - queremos tirar conclusões sobre o "todo" (a população), quando observamos apenas uma "parte" (a amostra).*
- ✓ *Conclusões estão sujeitas a erros - necessário conhecer a probabilidade de cometer tais erros.*

Inferência

Tirar conclusões para a população com base na amostra

- ✓ *Tomar decisões sobre μ (parâmetro populacional)*

  com base em $\bar{X}$ (estimador, estatística da amostra).

- ✓ *Necessário associar probabilidades de erros (ou margens de erros) a essas decisões*
- ✓ *Mas, para conhecer essas probabilidades, seria necessário conhecer a população – PROBLEMA!*
- ✓ *A menos que ....*

  ... pudéssemos associar distribuições de probabilidades conhecidas aos dados

  ... ou ao estimador ($\bar{X}$)

- ✓ *Para saber se isto é possível, precisamos estudar as propriedades de $\bar{X}$*

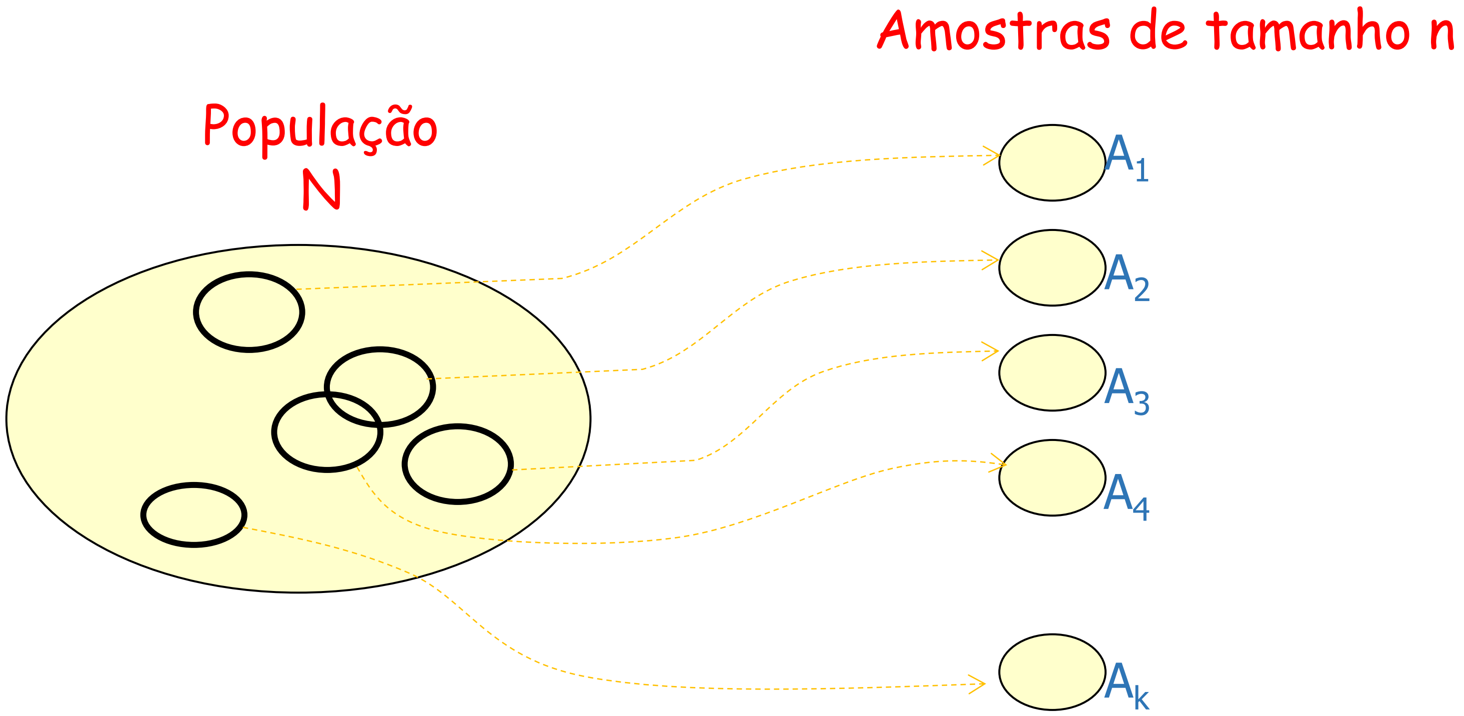
GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Propriedades de $\bar{X}$

X - variável aleatória (quantitativa) de interesse

$E(X) = \mu$

$VAR(X) = \sigma^2$



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

Propriedades de $\bar{X}$

BIOESTATÍSTICA

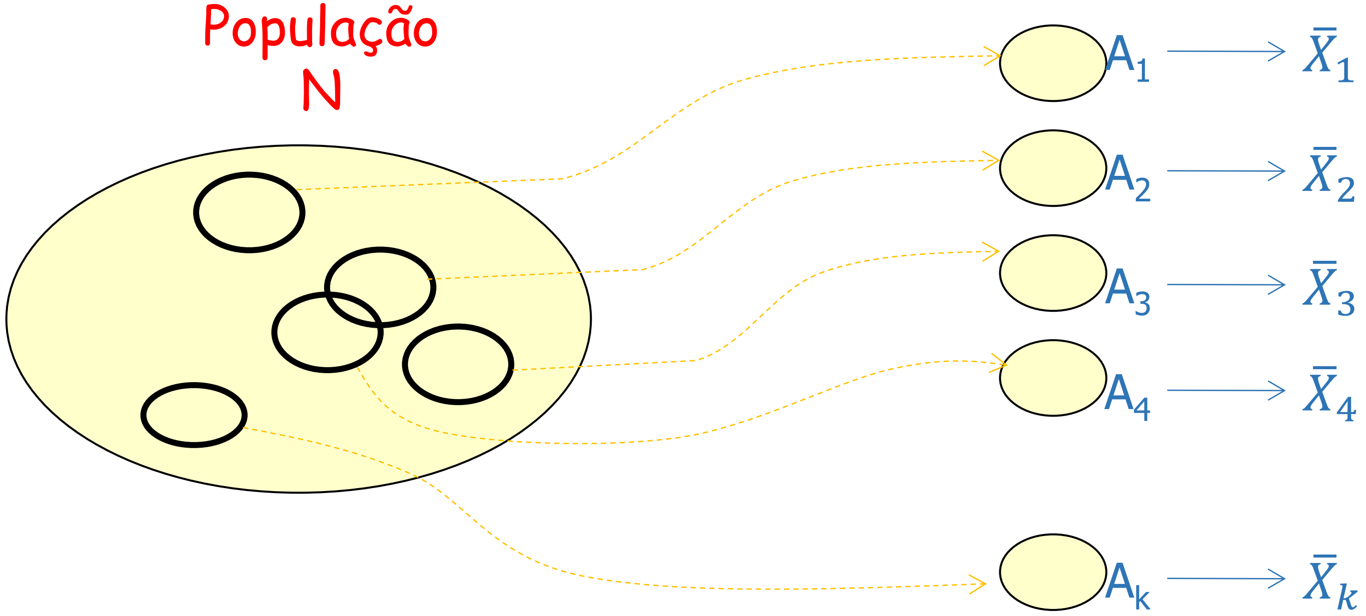
X - variável aleatória (quantitativa) de interesse

$E(X) = \mu$

$VAR(X) = \sigma^2$

Amostras de tamanho n

População
N



- ✓ $\bar{X}$ é constante ou variável aleatória?
- ✓ $\bar{X}$ é uma variável aleatória, pois é uma função da amostra
- ✓ Variáveis aleatórias tem E, VAR e distribuição
- ✓ $E(\bar{X}) =?$
- ✓ $VAR(\bar{X}) =?$
- ✓ Como seria a distribuição de $\bar{X}$?

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

Propriedades de $\bar{X}$

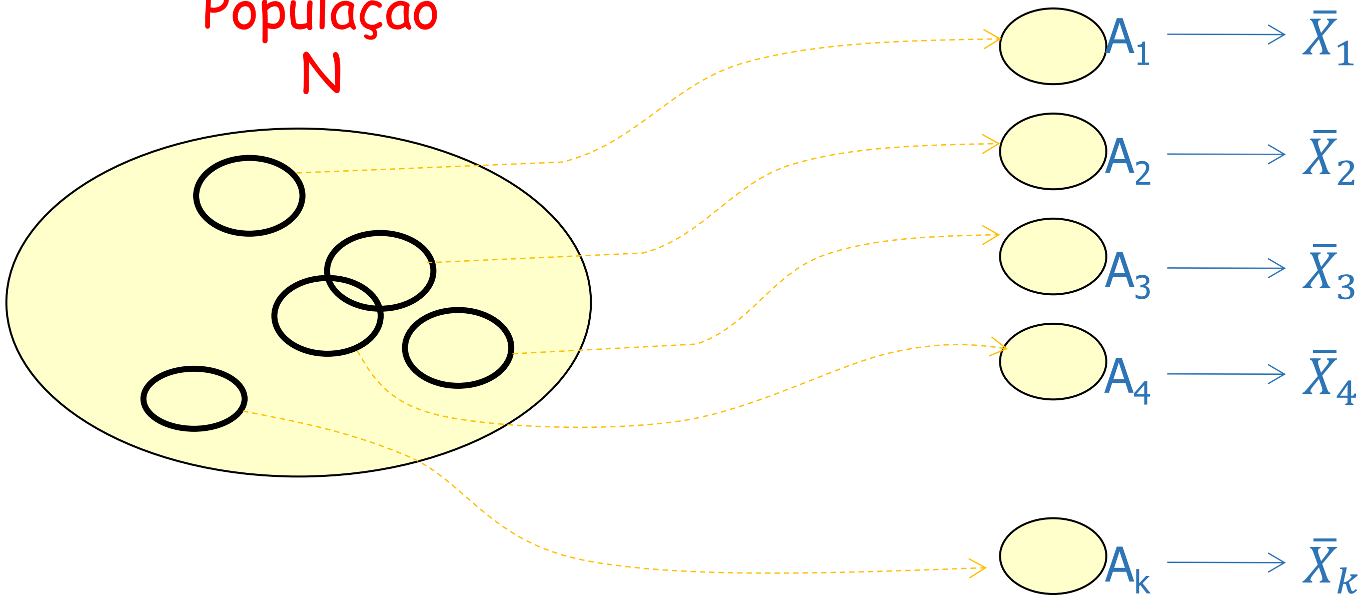
X - variável aleatória (quantitativa) de interesse

$E(X) = \mu$

$VAR(X) = \sigma^2$

Amostras de tamanho n

População
N



Resultados importantes:

- ✓ $E(\bar{X}) = \mu$
- ✓ $VAR(\bar{X}) = \frac{\sigma^2}{n}$
- ✓ **TLC:** Para n "grande", $\bar{X} \sim Normal$

Em resumo:

$$\bar{X} \xrightarrow{n \to \infty} N\left(\mu, \frac{\sigma^2}{n}\right)$$

BIOESTATÍSTICA

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

Distribuição amostral de $\bar{X}$

Vamos entender melhor isto por meio de um exemplo:

Considere uma população de tamanho N=5.

A variável de interesse X assume os seguintes valores na população:

1 3 5 5 7

- Como a população é conhecida, é possível obter E(X) e VAR(X)

$$E(X) = \mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{1+3+5+5+7}{5} = 4{,}2$$

$$VAR(X) = \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N} = 4{,}16$$

- Intuitivamente, amostras diferentes devem apresentar médias amostrais diferentes.
- Vamos retirar todas as amostras possíveis de tamanho n=2 desta população, com reposição, e construir a distribuição da média amostral $\bar{X}$.

Distribuição amostral de $\bar{X}$

População: (1,3,5,5,7)

A	(X_1, X_2)	$\bar{x}$
1	(1, 1)	1

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

Distribuição amostral de $\bar{X}$

BIOESTATÍSTICA

População: (1,3,5,5,7)

A	(X_1, X_2)	$\bar{x}$
1	(1, 1)	1
2	(1, 3)	2

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

Distribuição amostral de $\bar{X}$

População: (1,3,5,5,7)

A	(X_1, X_2)	$\bar{x}$
1	(1, 1)	1
2	(1, 3)	2
3	(1, 5)	3
4	(1, 5)	3
5	(1, 7)	4

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

Distribuição amostral de $\bar{X}$

BIOESTATÍSTICA

População: (1,3,5,5,7)

A	(X_1, X_2)	$\bar{x}$	A	(X_1, X_2)	$\bar{x}$	A	(X_1, X_2)	$\bar{x}$	A	(X_1, X_2)	$\bar{x}$	A	(X_1, X_2)	$\bar{x}$
1	(1, 1)	1	6	(3, 1)	2	11	(5, 1)	3	16	(5, 1)	3	21	(7, 1)	4
2	(1, 3)	2	7	(3, 3)	3	12	(5, 3)	4	17	(5, 3)	4	22	(7, 3)	5
3	(1, 5)	3	8	(3, 5)	4	13	(5, 5)	5	18	(5, 5)	5	23	(7, 5)	6
4	(1, 5)	3	9	(3, 5)	4	14	(5, 5)	5	19	(5, 5)	5	24	(7, 5)	6
5	(1, 7)	4	10	(3, 7)	5	15	(5, 7)	6	20	(5, 7)	6	25	(7, 7)	7

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

Distribuição amostral de $\bar{X}$

Distribuição de $\bar{X}$

$\bar{x}$	$P(\bar{X} = \bar{x})$
1	1/25
2	2/25
3	5/25
4	6/25
5	6/25
6	4/25
7	1/25
Total	1

A partir da **distribuição de $\bar{X}$**, vamos obter

- ✓ $E(\bar{X})$
- ✓ $VAR(\bar{X})$

$$E(\bar{X}) = \sum \bar{x} P(\bar{X} = \bar{x})$$

$$VAR(\bar{X}) = \sum \left(\bar{x} - E(\bar{X})\right)^2 P(\bar{X} = \bar{x})$$

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

Distribuição amostral de $\bar{X}$

Distribuição de $\bar{X}$

$\bar{x}$	$P(\bar{X} = \bar{x})$
1	1/25
2	2/25
3	5/25
4	6/25
5	6/25
6	4/25
7	1/25
Total	1

A partir da **distribuição de $\bar{X}$**, vamos obter

$$E(\bar{X}) = \sum \bar{x} P(\bar{X} = \bar{x})$$

$$E(\bar{X}) = 1\frac{1}{25} + 2\frac{2}{25} + 3\frac{5}{25} + 4\frac{6}{25} + 5\frac{6}{25} + 6\frac{4}{25} + 7\frac{1}{25}$$

$= 4{,}2$

$$VAR(\bar{X}) = \sum \left(\bar{x} - E(\bar{X})\right)^2 P(\bar{X} = \bar{x})$$

$$VAR(\bar{X}) = (1 - 4{,}2)^2 \frac{1}{25} + (2 - 4{,}2)^2 \frac{2}{25} + \cdots + (7 - 4{,}2)^2 \frac{1}{25}$$

$= 2{,}08$

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

Distribuição amostral de $\bar{X}$

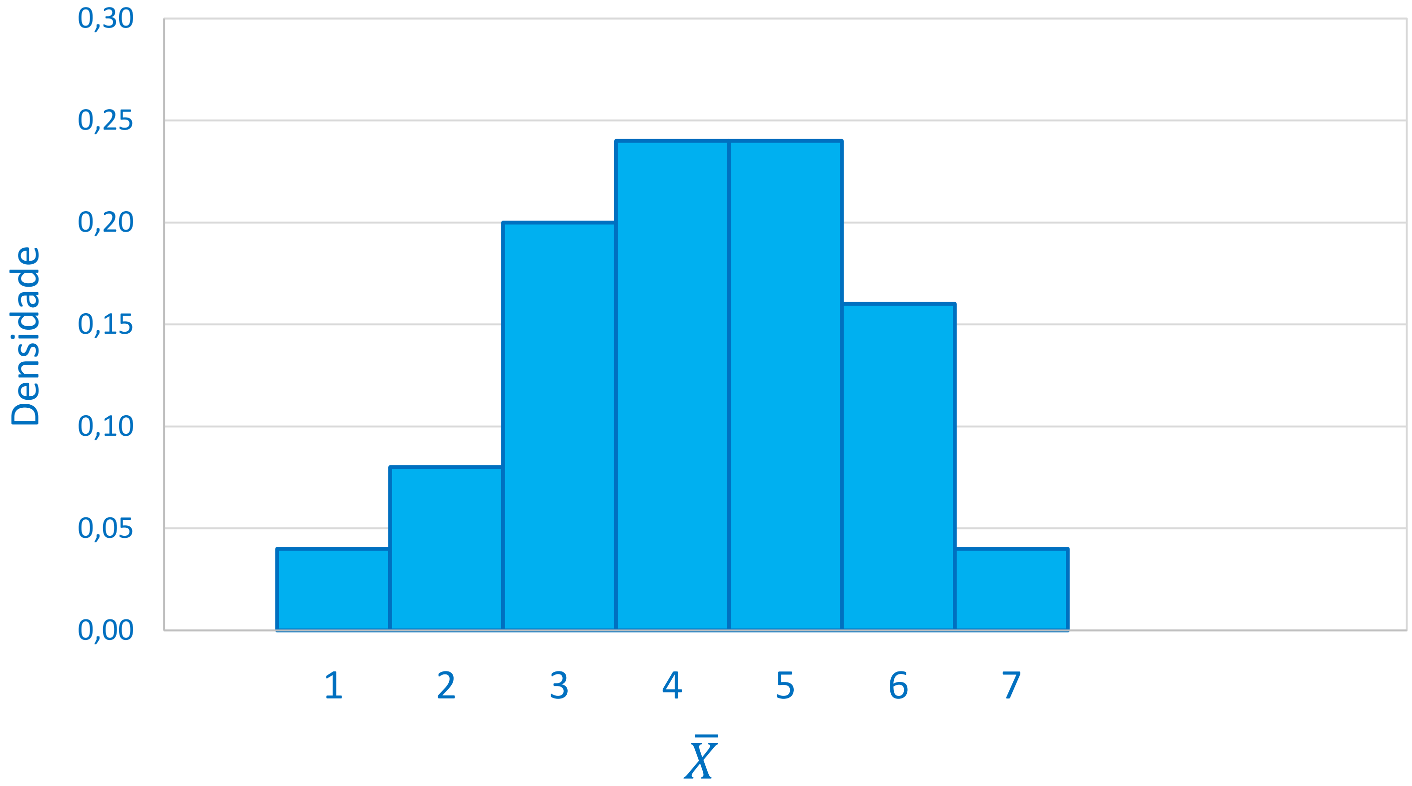
População	Amostras
(1,3,5,5,7)	
$E(X) = \mu = 4{,}2$	$E(\bar{X}) = 4{,}2$
$VAR(X) = \sigma^2 = 4{,}16$	$VAR(\bar{X}) = 2{,}08 = \dfrac{4{,}2}{2}$

- ✓ $E(\bar{X}) = \mu \quad \Rightarrow \bar{X}$ *é um estimador não viesado para* μ
- ✓ $VAR(\bar{X}) = \dfrac{\sigma^2}{n}$

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

Distribuição amostral de $\bar{X}$

Histograma para a distribuição de $\bar{X}$



$$E(\bar{X}) = \mu = 4{,}12$$

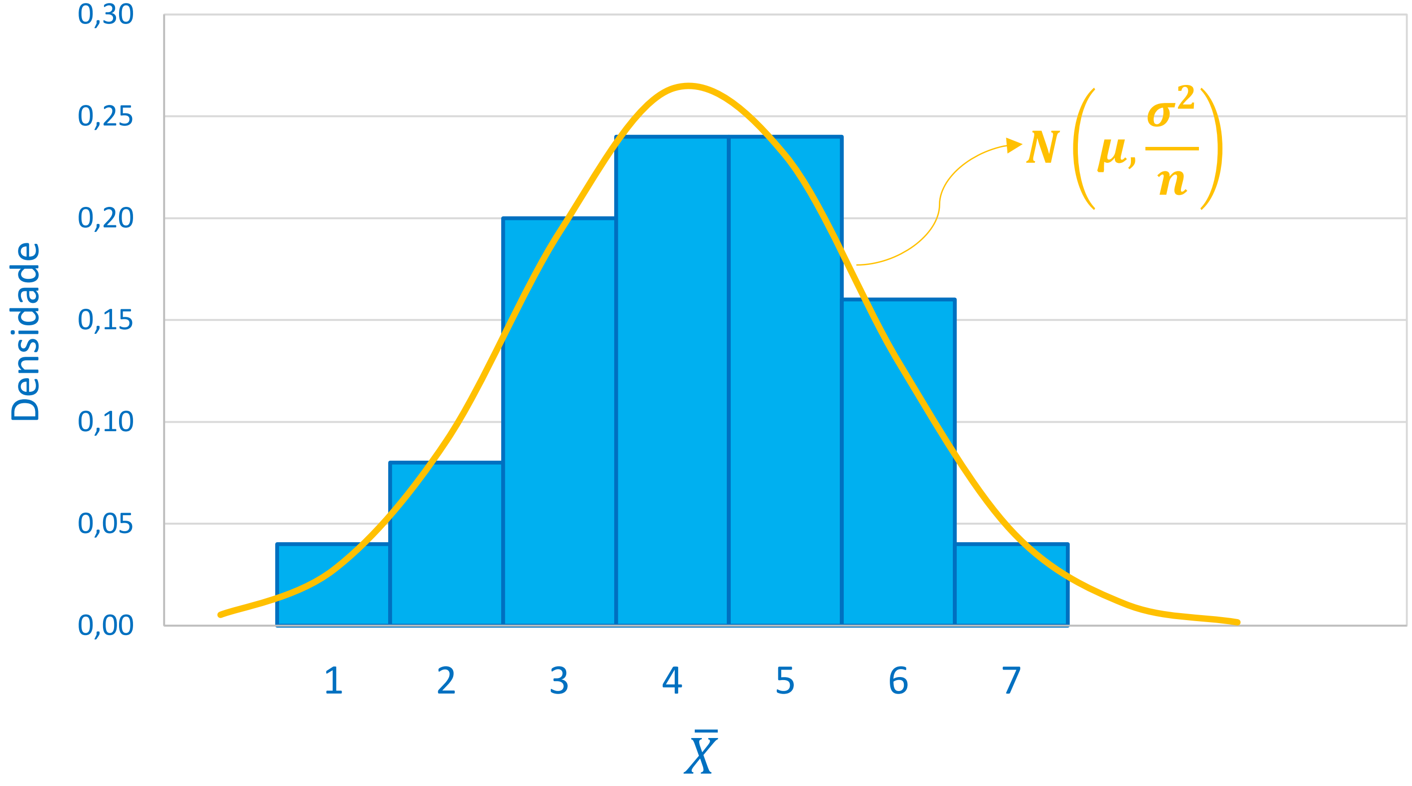
$$VAR(\bar{X}) = \frac{\sigma^2}{n} = 2{,}08$$

A distribuição de $\bar{X}$ assemelha-se à alguma distribuição conhecida?

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

Distribuição amostral de $\bar{X}$

Histograma para a distribuição de $\bar{X}$



$E(\bar{X}) = \mu = 4{,}12$

$VAR(\bar{X}) = \dfrac{\sigma^2}{n} = 2{,}08$

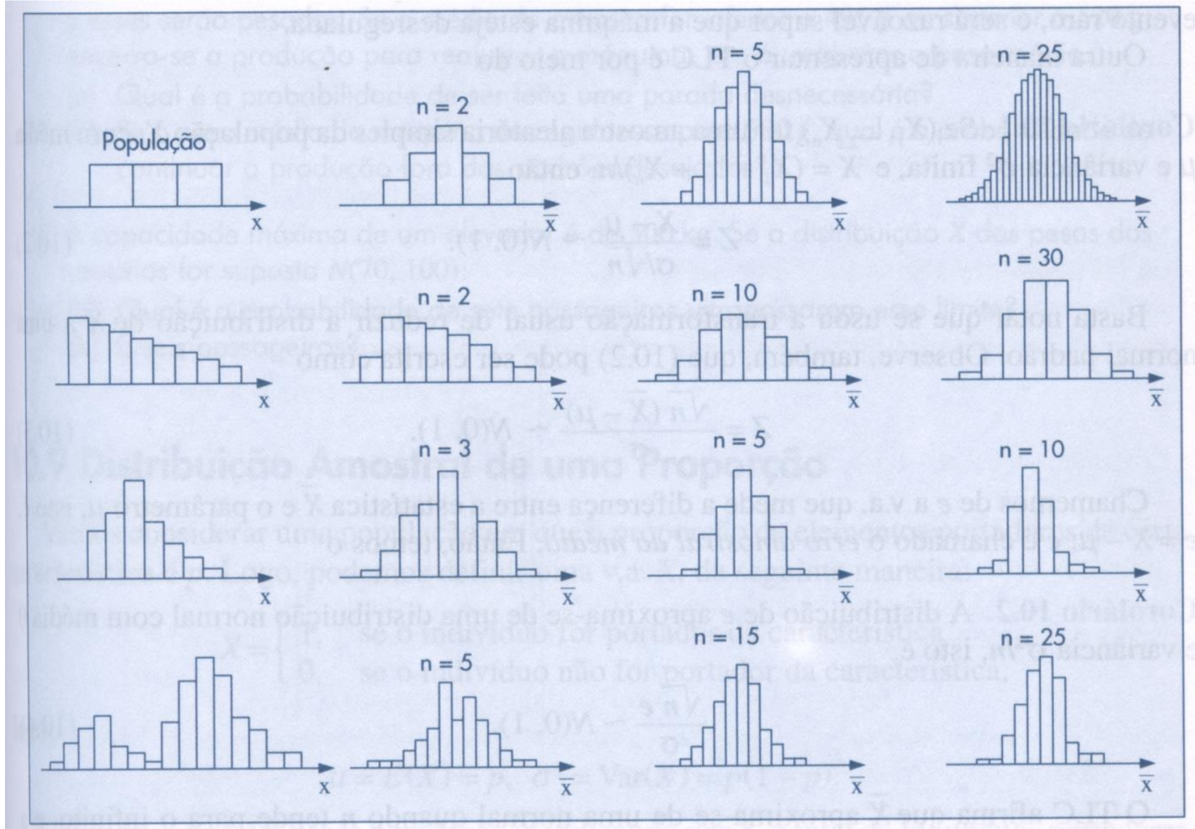
A distribuição de $\bar{X}$ assemelha-se à alguma distribuição conhecida?

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

Teorema Limite Central (TLC)

Histogramas correspondentes às distribuições amostrais de $\bar{X}$ para amostras extraídas de algumas populações.

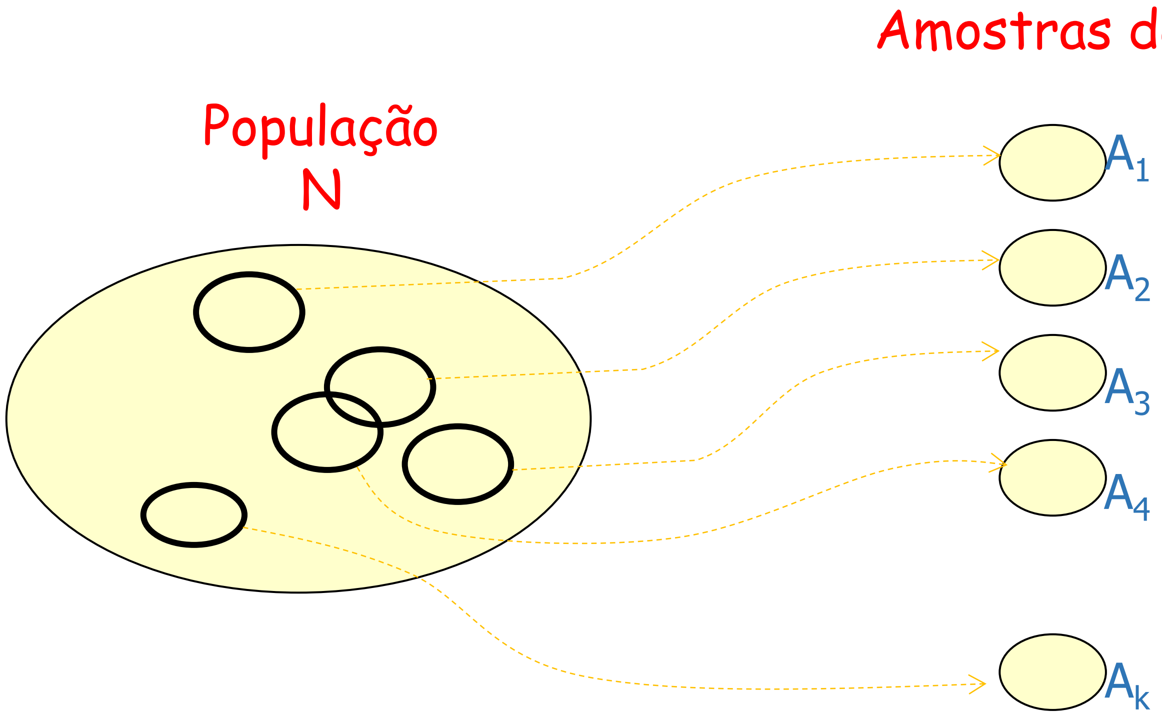
Distribuição de X | Distribuição de $\bar{X}$



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

Distribuição amostral de $\hat{p}$ (proporção)

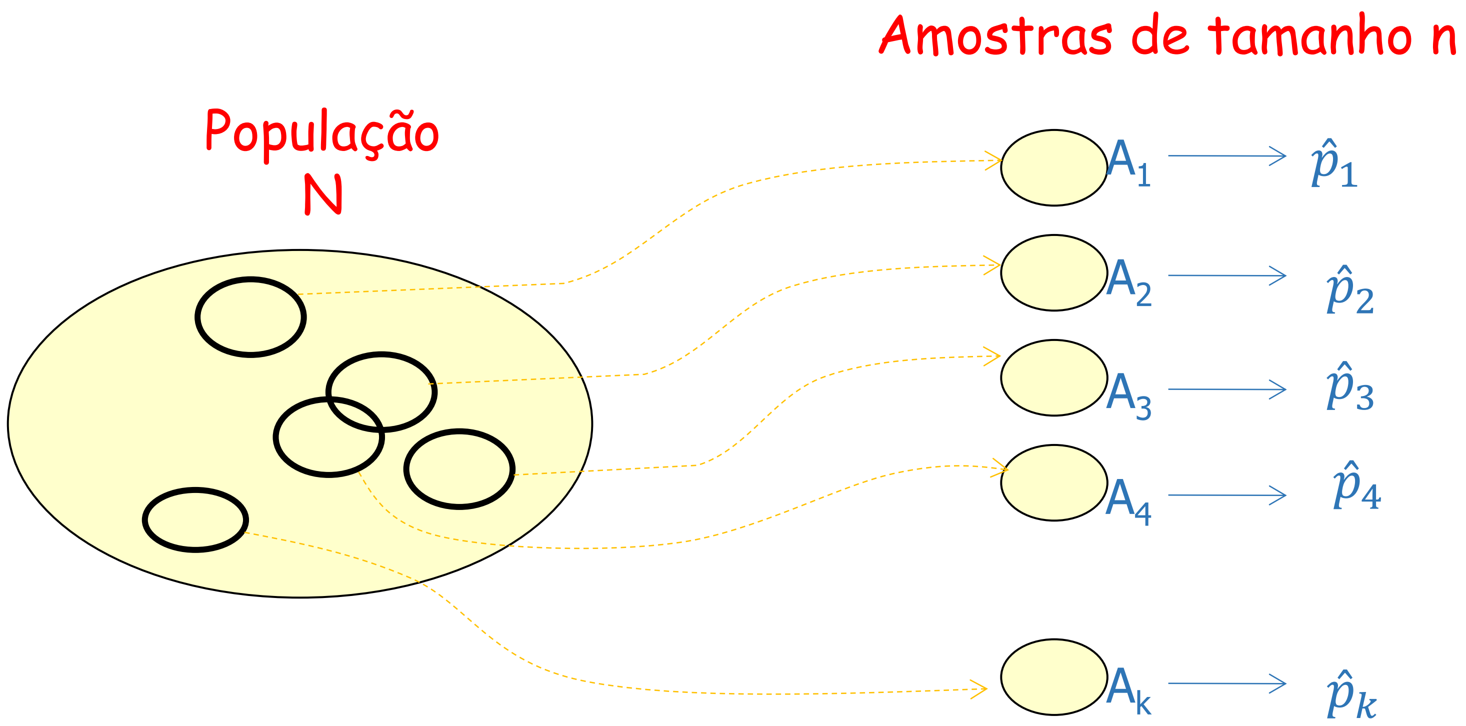
P (proporção) - variável aleatória (qualitativa) de interesse



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

Distribuição amostral de $\hat{p}$ (proporção)

P (proporção) - variável aleatória (qualitativa) de interesse



- ✓ $\hat{p}$ é constante ou variável aleatória?
- ✓ $\bar{p}$ é uma variável aleatória, pois é uma função da amostra
- ✓ Variáveis aleatórias tem E, VAR e distribuição
- ✓ $E(\bar{p}) = ?$
- ✓ $VAR(\bar{p}) = ?$
- ✓ Como seria a distribuição de $\bar{p}$?

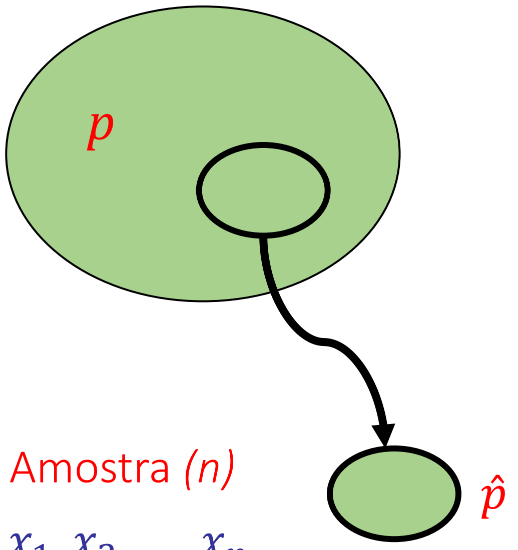
GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Distribuição amostral da proporção ($\hat{p}$)

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

População *(N)*

$X_1, X_2, \dots, X_N$



Amostra *(n)*

$x_1, x_2, \dots, x_n$

Em que sucesso significa o indivíduo ser portador de uma determinada característica de interesse na população (p. ex., tuberculose, asma, etc.).

Ou seja, *p* é a probabilidade (constante) de um indivíduo da população ter a característica.

Além disso,

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

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

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

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

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

Distribuição amostral da proporção ($\hat{p}$)

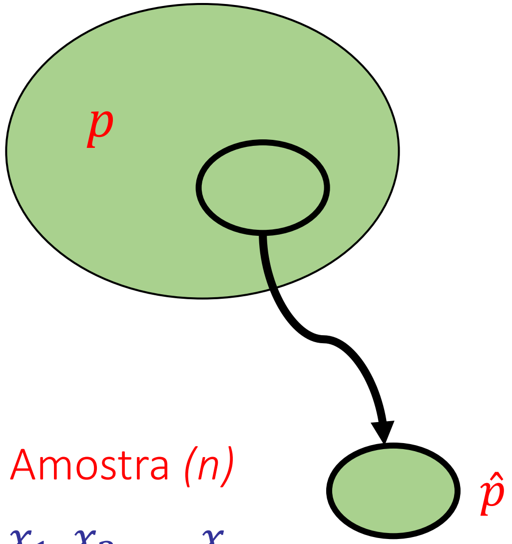
$X \sim Bernoulli(p)$, i.e., $X = \begin{cases} 1 \cdots sucesso \cdots p \\ 0 \cdots fracasso \cdots 1-p \end{cases}$

$E(X) = p$
$\mathrm{VAR}(X) = p(1-p)$

BIOESTATÍSTICA

População *(N)*

$X_1, X_2, \ldots, X_N$



Amostra *(n)*

$x_1, x_2, \ldots, x_n$

- ✓ na população: $p = \mu$
- ✓ na amostra: $\hat{p} = \bar{X}$

Vimos, pelo TLC, que $\bar{X} \xrightarrow{n\to\infty} N\left(E(X), \frac{VAR(X)}{n}\right)$

Consequentemente, $\hat{p} = \bar{X} \xrightarrow{n\to\infty} N\left(p, \frac{p(1-p)}{n}\right)$

Então, para n grande

- ✓ a distribuição de $\hat{p}$ pode ser aproximada pela Normal!
- ✓ podemos utilizar a Normal para construir intervalos de confiança e testar hipóteses para p.

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

Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

BIOESTATÍSTICA

Pelo TLC, para *n* grande:

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

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

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

Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

BIOESTATÍSTICA

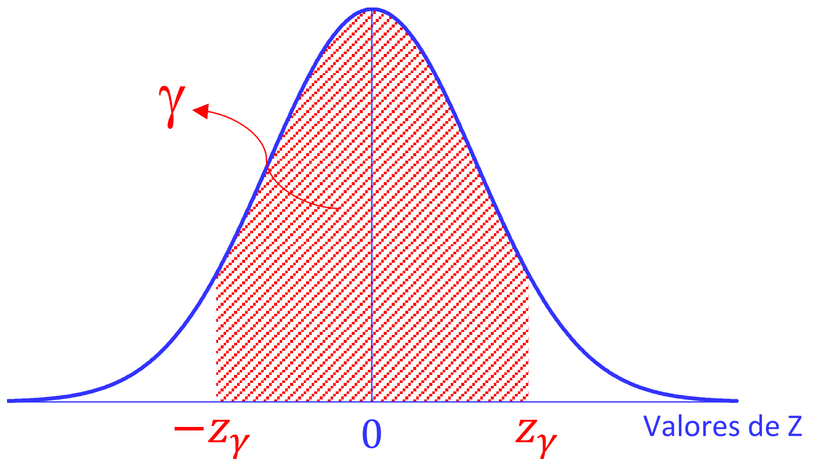
Na N(0,1), fixando γ, podemos encontrar um valor z_γ tal que

$$P(-z_\gamma \leq Z \leq z_\gamma) = \gamma$$

$$\Longrightarrow P\left(-z_\gamma \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_\gamma\right) = \gamma$$

$$\Longrightarrow P\left(-z_\gamma\sqrt{\frac{p(1-p)}{n}} \leq \hat{p}-p \leq z_\gamma\sqrt{\frac{p(1-p)}{n}}\right) = \gamma$$

$Z \sim N(0,1)$



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

Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

BIOESTATÍSTICA

$$\Rightarrow \mathrm{P}\left(-\hat{p} - z_{\gamma}\sqrt{\frac{p(1-p)}{n}} \leq -p \leq -\hat{p} + z_{\gamma}\sqrt{\frac{p(1-p)}{n}}\right) = \gamma$$

Multiplicando tudo por -1:

$$\Rightarrow \mathrm{P}\left(\hat{p} - z_{\gamma}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\gamma}\sqrt{\frac{p(1-p)}{n}}\right) = \gamma$$

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

Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

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

Problema: não conhecemos p.

Podemos proceder de duas maneiras:

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

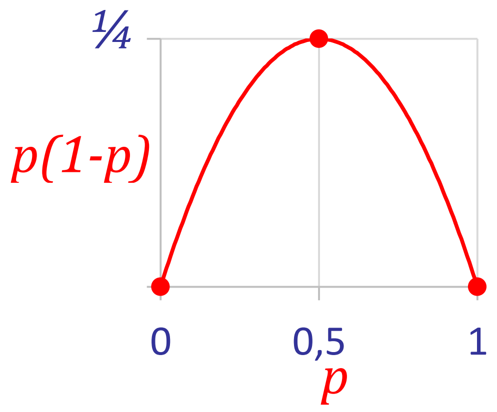
Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

i) Note que *p(1-p)* assume o valor máximo *1/4* (quando *p=1/2*).

Então, substituímos p(1-p) por ¼ :

$$\mathrm{IC}(\mathrm{p},\gamma) = \hat{p} \mp z_{\gamma}\sqrt{\frac{1}{4n}}$$



produzindo um intervalo de confiança **conservador** ou **conservativo** para *p.*

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

Intervalo de confiança para a proporção (p)

Utilizando a aproximação pela Normal

ii) Substituímos p por $\hat{p}$

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

produzindo um intervalo de confiança **otimista** para p.

Exercício 3

BIOESTATÍSTICA

Pretende-se estimar a proporção de cura, através do uso de um certo medicamento em doentes contaminados com cercária.

Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados.

a) Obtenha uma estimativa pontual para a proporção de cura.

b) Obtenha estimativas por intervalo para a proporção de cura (conservativo e otimista). Utilize um coeficiente de confiança de 95%.

c) Interprete os intervalos de confiança que você construiu.

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

Exercício 3

a) Estimativa pontual para o parâmetro p: $\hat{p} = \dfrac{160}{200} = 0{,}8$

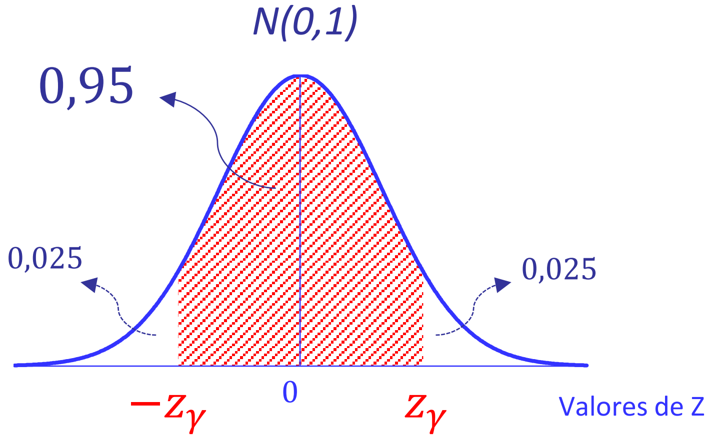
b) Estimativa por intervalo para o parâmetro p:

$\gamma = 0{,}95 \Rightarrow z_{\gamma} = 1{,}96$

✓ Conservativo

$$\text{IC}(\text{p}, 0{,}95) = \hat{p} \mp z_{\gamma}\sqrt{\frac{1}{4n}} = 0{,}8 \mp 1{,}96\sqrt{\frac{1}{4 * 200}}$$

$$= 0{,}8 \mp 0{,}069296 = [0{,}7307 : 0{,}8693]$$



$z_{\gamma} = 1{,}96$

✓ Otimista

$$\text{IC}(\text{p}, 0{,}95) = \hat{p} \mp z_{\gamma}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}8 \mp 1{,}96\sqrt{\frac{0{,}8 * 0{,}2}{200}}$$

$$= 0{,}8 \mp 0{,}055437 = [0{,}7446 : 0{,}8554]$$

BIOESTATÍSTICA

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

Intervalo de confiança para a proporção (p)

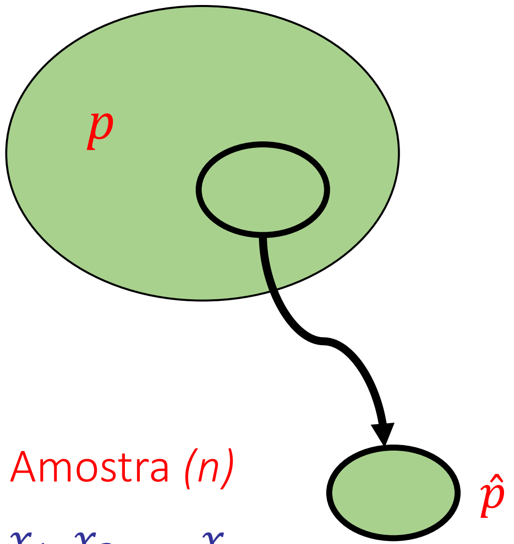
Utilizando a distribuição Binomial (Exata!)

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

$$E(X) = p$$
$$\mathrm{VAR}(X) = p(1-p)$$

População *(N)*

$X_1, X_2, \ldots, X_N$



Amostra *(n)*

$x_1, x_2, \ldots, x_n$

Na amostra, o estimador de *p* será:

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

onde $y \sim Binomial\ (n, p)$

Ou seja, a distribuição de $\hat{p}$ pode ser obtida

da distribuição de *y* (Binomial):

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

Intervalo de confiança para a proporção (p)

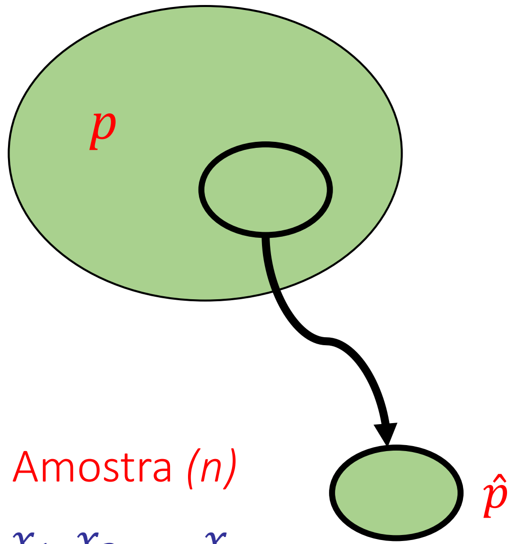
Utilizando a distribuição Binomial (Exata!)

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

$$E(X) = p$$
$$\text{VAR}(X) = p(1-p)$$

População *(N)*

$X_1, X_2, \ldots, X_N$



Amostra *(n)*

$x_1, x_2, \ldots, x_n$

Como $y \sim$ ***Binomial (n, p),*** consigo obter probabilidades:

$$P(y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Mas

$$P(y = k) = P\left(\frac{y}{n} = \frac{k}{n}\right) = P\left(\hat{p} = \frac{k}{n}\right)$$

- ✓ Difícil de fazer os IC à mão!
- ✓ Mas o computador faz!

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

Teorema Limite Central (TLC)

BIOESTATÍSTICA

- X – variável aleatória com uma distribuição **qualquer**
- $E(X) = \mu\ ,\ \mathrm{VAR}(X) = \sigma^2$
- Amostra aleatória simples, tamanho *n*: $x_1, x_2, \ldots, x_n$
- $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

Então: $$\bar{X} \xrightarrow{n \to \infty} N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Isto é, para n grande:
  - $\bar{X}$ tem distribuição Normal (independentemente da distribuição de X)
  - Com esperança (média) igual à da variável original: $E(X) = \mu$
  - Com variância igual à da variável original dividida pelo tamanho da amostra (n): $\mathrm{VAR}(X) = \frac{\sigma^2}{n}$
- Entretanto, se X já tiver distribuição Normal, $\bar{X}$ será Normal, não importa o valor de *n*

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

Teorema Limite Central (TLC)

BIOESTATÍSTICA

$$\bar{X} \xrightarrow{n\to\infty} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Longrightarrow \quad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- ✓ Frequentemente, precisamos trabalhar com esta quantidade, mas não conhecemos σ^2, apenas S^2.
- ✓ Então, utilizamos um resultado que é uma consequência do TLC:

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

- ✓ A distribuição t-Student não depende de μ e σ^2, apenas do número de graus de liberdade

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

Intervalo de confiança para a média (μ)

BIOESTATÍSTICA

- ✓ Já aprendemos a obter uma estimativa pontual para a média μ, que é $\bar{X}$
- ✓ Utilizando o TLC, vamos obter uma estimativa por intervalo para μ
- ✓ Queremos obter um intervalo que contenha μ , com uma probabilidade alta.
- ✓ Algo do tipo: dois valores v_1 e v_2 tais que $\mathrm{P}(v_1 \leq \mu \leq v_2) = \gamma$,

  onde γ é uma probabilidade alta!
- ✓ Vamos utilizar o TLC:

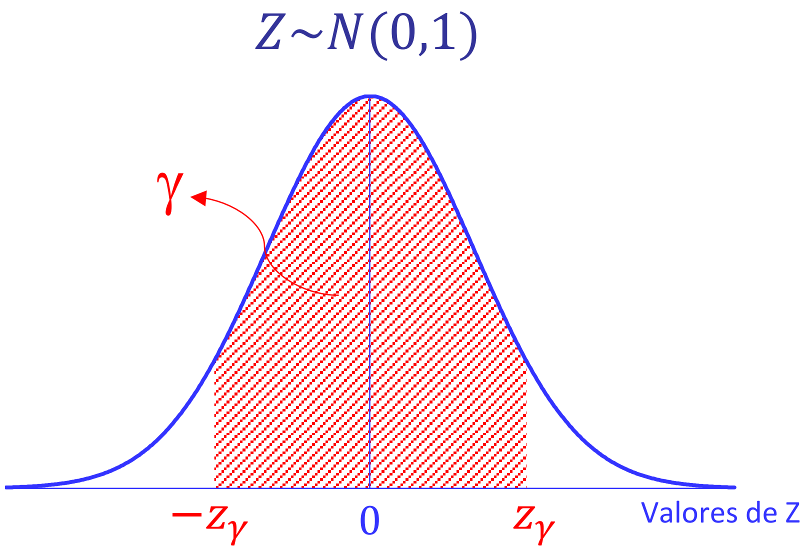
  Segundo o TLC, para n grande, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Intervalo de confiança para a média (μ)

Lembrando ...

Se $Z \sim N(0,1)$

- ✓ P(-z < Z < z) = 0,90 ⇒ z=1,645
- ✓ P(-z < Z < z) = 0,95 ⇒ z=1,96
- ✓ P(-z < Z < z) = 0,96 ⇒ z=2,055
- ✓ P(-z < Z < z) = 0,99 ⇒ z=2,575



Ou seja, fixando γ, podemos encontrar um valor z_γ tal que

$$P(-z_\gamma \leq Z \leq z_\gamma) = \gamma$$

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

Intervalo de confiança para a média (μ)

Fixando γ, podemos encontrar um valor z_γ tal que

$$P(-z_\gamma \leq Z \leq z_\gamma) = \gamma$$

Substituindo Z por $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$:

$$\Rightarrow P\left(-z_\gamma \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_\gamma\right) = \gamma$$

Agora basta isolar μ na expressão acima, de modo a obter

v_1 e v_2 tais que $P(v_1 \leq \mu \leq v_2) = \gamma$

e teremos o intervalo!

BIOESTATÍSTICA

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

Intervalo de confiança para a média (μ)

$$P\left(-z_\gamma \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_\gamma\right) = \gamma \quad \Rightarrow P\left(-z_\gamma\,\sigma/\sqrt{n} \le \bar{X}-\mu \le +z_\gamma\,\sigma/\sqrt{n}\right) = \gamma$$

$$\Rightarrow P\left(-\bar{X} - z_\gamma\,\sigma/\sqrt{n} \le -\mu \le -\bar{X} + z_\gamma\,\sigma/\sqrt{n}\right) = \gamma$$

Multiplicando tudo por -1:

$$\Rightarrow P\left(\bar{X} - z_\gamma\,\sigma/\sqrt{n} \le \mu \le \bar{X} + z_\gamma\,\sigma/\sqrt{n}\right) = \gamma$$

Então, $v_1 = \bar{X} - z_\gamma\,\sigma/\sqrt{n}$ e $v_2 = \bar{X} + z_\gamma\,\sigma/\sqrt{n}$

E representamos o intervalo assim:

$$IC(\mu,\gamma) = \bar{X} \mp z_\gamma\,\sigma/\sqrt{n}$$

Note que, para construir o IC, a variância populacional σ^2 deve ser conhecida.

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

Intervalo de confiança para a média μ, quando a variância σ^2 é conhecida

$$\mathrm{IC}(\mu,\gamma) = \bar{X} \mp z_\gamma\, \sigma/\sqrt{n}$$

Interpretação

- ✓ A probabilidade de que o IC contenha o verdadeiro valor da média populacional é dada por γ.
- ✓ Se obtivermos várias amostras de mesmo tamanho e para cada uma calcularmos os correspondentes intervalos de confiança com coeficiente de confiança γ, esperamos que a proporção de intervalos que contenham o valor μ seja igual a γ.

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

Exemplo

BIOESTATÍSTICA

Por analogia a produtos similares, o tempo de reação a um determinado medicamento pode ser considerado como tendo distribuição Normal com desvio padrão igual a 2 minutos (a média é desconhecida). Vinte pacientes foram sorteados, receberam o medicamento e tiveram seu tempo de reação anotado. Os dados foram os seguintes:

2,9	3,4	3,5	4,1	4,6	4,7	4,5	3,8	5,3	4,9
4,8	5,7	5,8	5,0	3,4	5,9	6,3	4,6	5,5	6,2

Obtenha uma estimativa pontual e uma estima por intervalo (intervalo de confiança) para o tempo médio de reação. Utilize um coeficiente de confiança igual a 95%.

Exemplo

- ✓ Estimativa pontual para μ:

  $\bar{X} = 4{,}668$ minutos

- ✓ Estimativa para μ por intervalo de confiança:

  - ✓ $\mathrm{IC}(\mu,\gamma) = \bar{X} \mp z_\gamma\, \sigma/\sqrt{n}$
  - ✓ $\sigma = 2$
  - ✓ Para $\gamma = 0{,}95 \Rightarrow z_\gamma = 1{,}96$
  - ✓ n=20

$$\mathrm{IC}(\mu,\gamma) = \bar{X} \mp z_\gamma\, \sigma/\sqrt{n}$$

$$\mathrm{IC}(\mu,\gamma) = 4{,}668 \mp 1{,}96 * 2/\sqrt{20}$$

$$\mathrm{IC}(\mu,\gamma) = 4{,}668 \mp 0{,}8763$$

$$\mathrm{IC}(\mu,\gamma) = [3{,}7915 : 5{,}5445]\ \mathrm{min}$$

- ✓ Interpretação: A probabilidade de que o intervalo $[3{,}8 : 5{,}5]$ minutos contenha o tempo **médio** de reação ao medicamento na **população** $(\boldsymbol{\mu})$ é 0,95 ou 95%.

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

Intervalo de confiança para a média μ, quando a variância σ^2 é conhecida

BIOESTATÍSTICA

$$\mathrm{IC}(\mu, \gamma) = \bar{X} \mp z_{\gamma}\, \sigma/\sqrt{n}$$

Problema!!!

- ✓ Na prática, é difícil obter este tipo de intervalo.
- ✓ Porque, em geral, não conhecemos σ^2, apenas S^2.

Teorema Limite Central (TLC)

BIOESTATÍSTICA

$$\bar{X} \xrightarrow{n\to\infty} N\left(\mu, \frac{\sigma^2}{n}\right) \implies Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- ✓ Frequentemente, precisamos trabalhar com esta quantidade, mas não conhecemos σ^2, apenas S^2.
- ✓ Então, utilizamos um resultado que é uma consequência do TLC:

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

- ✓ A distribuição t-Student não depende de μ e σ^2, apenas do número de graus de liberdade

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

Distribuição t de Student

Seja Z uma v.a. com distribuição N(0,1) e

Y uma v.a. com distribuição $\chi^2(\nu)$, com Z e Y independentes.

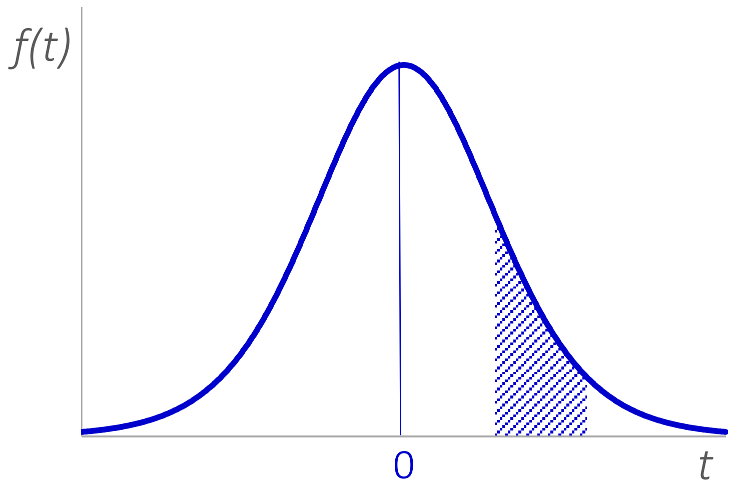
Então a v.a.

$$t = \frac{Z}{\sqrt{Y/\nu}}$$

tem função densidade dada por

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1+t^2/\nu)^{-(\nu+1)/2}$$

$$-\infty < t < \infty$$



$E(t) = 0$

$Var(t) = \nu/(\nu-2)$

A área sob a curva fornece probabilidades

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

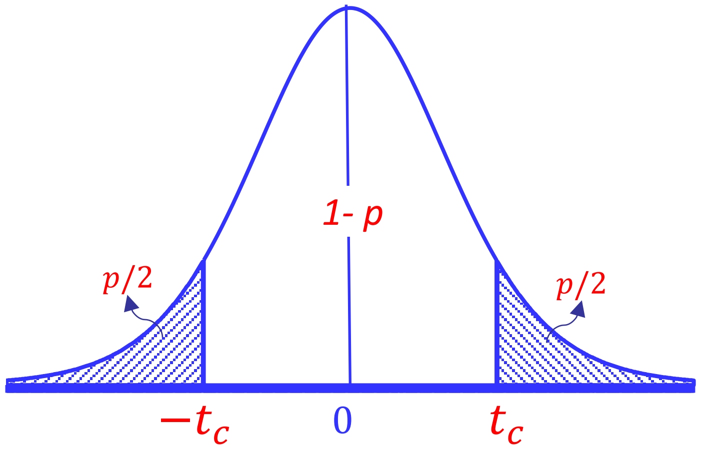
Tabela da Distribuição t de Student

Corpo dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$

g,l.	p=0,90	...	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
...	...	...	...	...	...	...	...	...	...	...
7	0,130	...	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	...	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	...	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	...	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	...	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	...	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	...	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	...	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	...	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	...	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	...	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	...	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	...	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	...	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	...	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
...	...	...	...	...	...	...	...	...	...	...
26	0,127	...	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	...	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	...	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	...	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	...	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	...	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	...	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	...	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∝	0,126	...	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

BIOESTATÍSTICA

- ✓ Para um determinado valor de p (soma das áreas em azul), o corpo da tabela fornece o valor de t_c.
- ✓ Para um determinado valor de t_c, a primeira linha da tabela fornece o valor de p (soma das áreas em azul).



$$P(-t_c < t < t_c) = 1 - p$$

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

Tabela da Distribuição t de Student

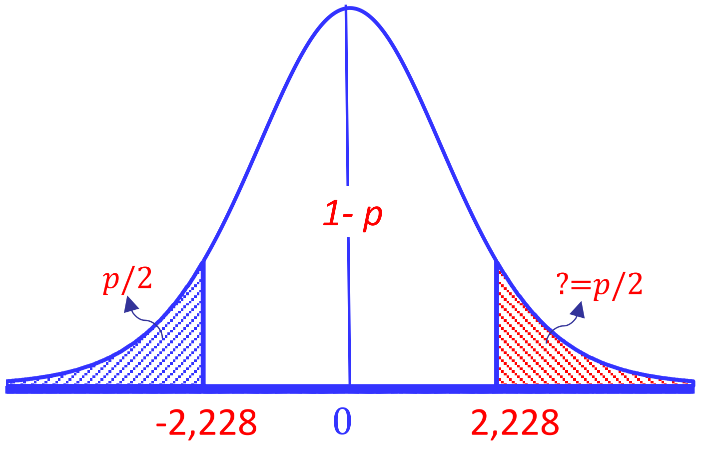
Corpo dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$

g,l.	p=0,90	...	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
...	...	...	...	...	...	...	...	...	...	...
7	0,130	...	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	...	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	...	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	...	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	...	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	...	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	...	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	...	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	...	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	...	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	...	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	...	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	...	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	...	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	...	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
...	...	...	...	...	...	...	...	...	...	...
26	0,127	...	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	...	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	...	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	...	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	...	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	...	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	...	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	...	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∝	0,126	...	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

BIOESTATÍSTICA

$$P(t_{10} > 2{,}228) = ?$$

$$P(t_{10} > 2{,}228) = \frac{0{,}05}{2} = 0{,}025$$



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

Tabela da Distribuição t de Student

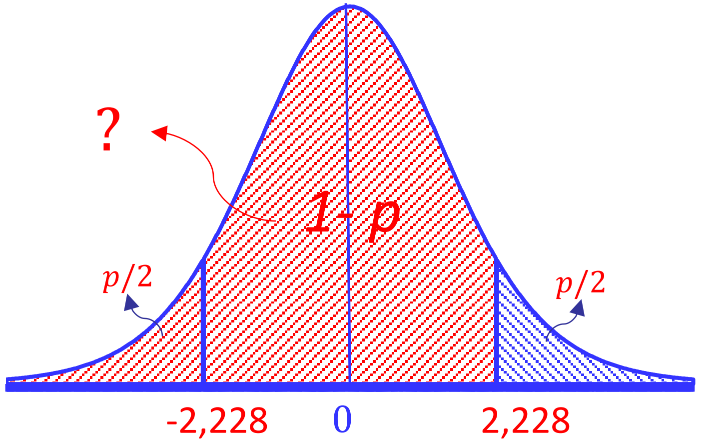
Corpo dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$

g,l.	p=0,90	...	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
...	...	...	...	...	...	...	...	...	...	...
7	0,130	...	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	...	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	...	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	...	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	...	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	...	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	...	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	...	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	...	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	...	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	...	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	...	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	...	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	...	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	...	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
...	...	...	...	...	...	...	...	...	...	...
26	0,127	...	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	...	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	...	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	...	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	...	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	...	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	...	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	...	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∝	0,126	...	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

BIOESTATÍSTICA

$$P(t_{10} < 2{,}228) = ?$$

$$P(t_{10} < 2{,}228) = 1 - \frac{0{,}05}{2} = 0{,}975$$



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

Tabela da Distribuição t de Student

Corpo dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$

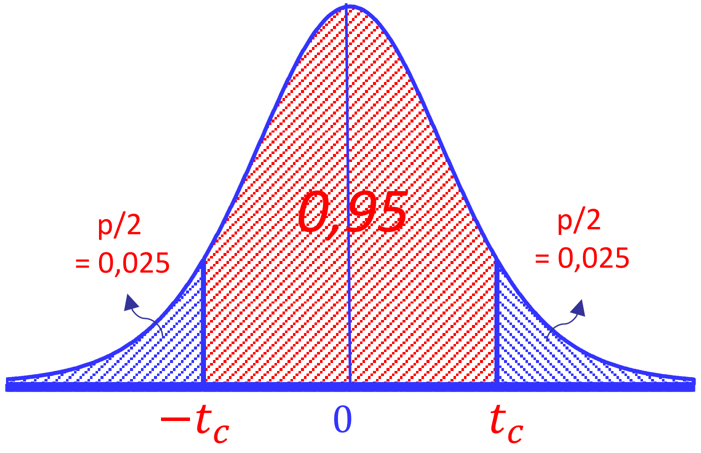
g,l.	p=0,90	...	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
...	...	...	...	...	...	...	...	...	...	...
7	0,130	...	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	...	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	...	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	...	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	...	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	...	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	...	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	...	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	...	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	...	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	...	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	...	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	...	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	...	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	...	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
...	...	...	...	...	...	...	...	...	...	...
26	0,127	...	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	...	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	...	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	...	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	...	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	...	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	...	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	...	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∝	0,126	...	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

BIOESTATÍSTICA

t_c *, tal que*

$P(-t_c < t < t_c) = 0{,}95$

$t_c = 2{,}228$



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

Exercício 1

Para uma t_{60}

✓ Obtenha as probabilidades:

a) $P(t < 2)$

b) $P(t > 2)$

c) $P(-2 < t < 2)$

d) $P(t > -2)$

✓ Obtenha o valor t_c tal que:

a) $P(-t_c < t < t_c) = 0{,}80$

b) $P(-t_c < t < t_c) = 0{,}85$

c) $P(-t_c < t < t_c) = 0{,}90$

d) $P(-t_c < t < t_c) = 0{,}95$

e) $P(-t_c < t < t_c) = 0{,}99$

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

Exercício 1

BIOESTATÍSTICA

Para uma t_{60}

✓ Obtenha as probabilidades:

a) $P(t < 2) = 1 - \frac{0,05}{2} = 0,975$

b) $P(t > 2) = \frac{0,05}{2} = 0,025$

c) $P(-2 < t < 2) = 1 - 0,05 = 0,95$

d) $P(t > -2) = 1 - \frac{0,05}{2} = 0,975$

✓ Obtenha o valor t_c tal que:

a) $P(-t_c < t < t_c) = 0,80 \Rightarrow t_c = 1,296$

b) $P(-t_c < t < t_c) = 0,85 \Rightarrow t_c = 1,4835$

c) $P(-t_c < t < t_c) = 0,90 \Rightarrow t_c = 1,671$

d) $P(-t_c < t < t_c) = 0,95 \Rightarrow t_c = 2,000$

e) $P(-t_c < t < t_c) = 0,99 \Rightarrow t_c = 2,660$

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

Lembrando ...

Teorema Limite Central (TLC)

$$\bar{X} \xrightarrow{n\to\infty} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Longrightarrow \quad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- ✓ Frequentemente, precisamos trabalhar com esta quantidade, mas não conhecemos σ^2, apenas S^2.
- ✓ Então, utilizamos um resultado que é uma consequência do TLC:

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

- ✓ A distribuição t não depende de μ e σ^2, apenas do número de graus de liberdade

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

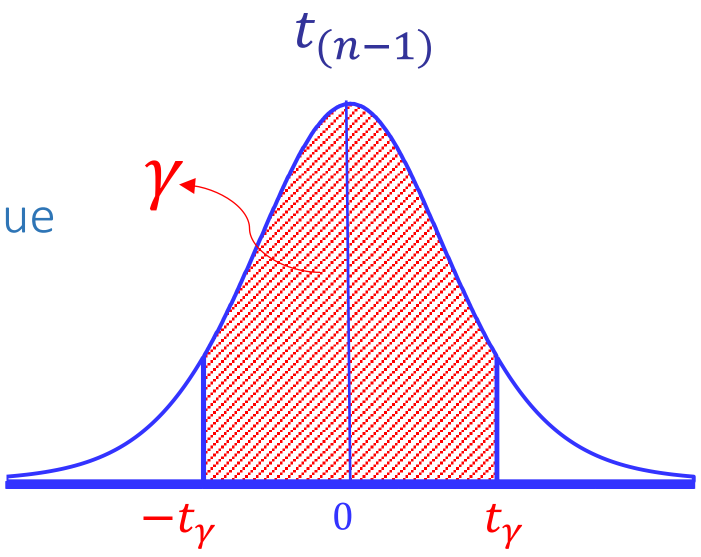
Intervalo de confiança para a média μ, quando a variância σ^2 é desconhecida

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Da mesma forma ...

Fixando γ, podemos encontrar um valor t_γ, da distribuição t_{n-1}, tal que

$$\mathrm{P}(-t_\gamma \leq t \leq t_\gamma) = \gamma \Rightarrow \mathrm{P}\left(-t_\gamma \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_\gamma\right) = \gamma$$



E o IC será dado por:

$$\mathrm{IC}(\mu,\gamma) = \bar{X} \mp t_\gamma\, S/\sqrt{n}$$

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

Exercício 2

BIOESTATÍSTICA

Um pesquisador avaliou os níveis plasmáticos de vitamina A em um grupo de 61 crianças diabéticas com idade até 12 anos, obtendo, para esse grupo, $\bar{X} = 25{,}5\ \mu g/dL$ e $S = 8{,}5\ \mu g/dL$.

a) Construa intervalos de confiança para o nível plasmático médio de vitamina A nessa população, com coeficientes de confiança 80%, 85%, 90%, 95% e 99%.

b) Interprete cada intervalo que você construiu.

c) O que acontece com o intervalo à medida que aumenta o coeficiente de confiança? Por quê?

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

Exercício 2

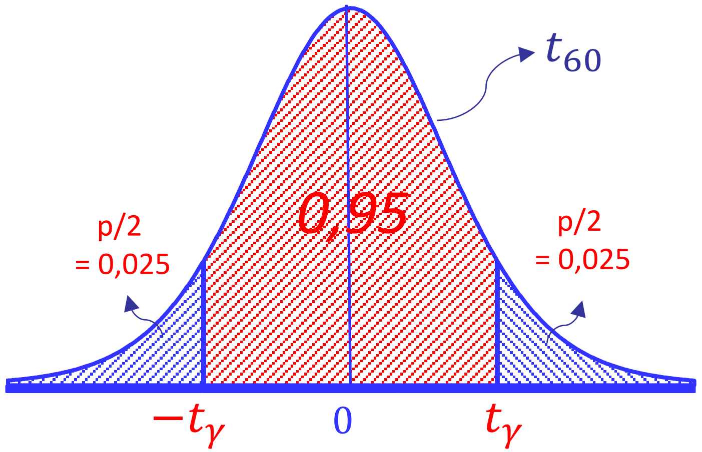
$$IC(\mu,\gamma) = \bar{X} \mp t_\gamma\, S/\sqrt{n}$$

$\bar{X} = 25{,}5$

$S = 8{,}5$

$n = 61$

$\gamma = 0{,}95 \Rightarrow t_\gamma = ?$



$\Rightarrow t_\gamma = 2{,}000$

$IC(\mu; 0{,}95) = 25{,}5 \mp 2 * 8{,}5/\sqrt{61}$

$IC(\mu; 0{,}95) = 25{,}5 \mp 2{,}177$

$IC(\mu; 0{,}95) = [23{,}323 : 27{,}677]\ \mu g/dL$

Interpretação: A probabilidade de que o intervalo $[23{,}3 : 27{,}7]\ \mu g/dL$ contenha o nível plasmático **médio** de vitamina A na **população** $(\boldsymbol{\mu})$ de crianças diabéticas até 12 anos é 0,95 ou 95%.

Exercício 2

BIOESTATÍSTICA

γ	t_γ	$t_\gamma \frac{S}{\sqrt{n}}$	$IC(\mu,\gamma) = \bar{X} \pm t_\gamma \frac{S}{\sqrt{n}}$
0,80	1,296	1,410	24,090 : 26,910
0,85	1,4835	1,615	23,885 : 27,115
0,90	1,671	1,819	23,681 : 27,319
0,95	2,000	2,177	23,323 : 27,677
0,99	2,660	2,895	22,605 : 28,395

- ✓ Nesse caso, à medida que o coeficiente de confiança aumenta, aumenta a amplitude do intervalo.
- ✓ Como poderíamos resolver este problema?

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

Intervalo de confiança para a média μ

Note que a amplitude do intervalo de confiança

- ✓ aumenta à medida que o coeficiente de confiança (γ) aumenta
- ✓ aumenta à medida que o desvio padrão (S) aumenta
- ✓ diminui à medida que o tamanho da amostra (n) aumenta

Intervalo de confiança para a média μ, quando a variância σ^2 é desconhecida

BIOESTATÍSTICA

Então, para construir um $\mathrm{IC}(\mu, \gamma)$, precisamos:

- ✓ definir γ
- ✓ procurar na tabela da t_{n-1} o valor t_γ tal que $\mathrm{P}(-t_\gamma \leq t \leq t_\gamma) = \gamma$
- ✓ coletar uma amostra de tamanho n: $x_1, x_2, \ldots, x_n$
- ✓ calcular $\bar{X}$ e S
- ✓ calcular o $\mathrm{IC}(\mu, \gamma) = \bar{X} \mp t_\gamma \, S/\sqrt{n}$

estimador para o parâmetro de interesse

percentil da $t_{(n-1)}$

desvio padrão do estimador

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

Em resumo...

O QUE VOCÊ PRECISA SABER...

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

Em resumo...

BIOESTATÍSTICA

Teorema Limite Central (TLC)

- ✓ X – variável aleatória com uma distribuição qualquer
- ✓ $E(X) = \mu \ , \ \mathrm{VAR}(X) = \sigma^2$
- ✓ Amostra aleatória simples, tamanho *n:* $x_1, x_2, \ldots, x_n$
- ✓ $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

Então: $\bar{X} \xrightarrow{n \to \infty} N\left(\mu, \frac{\sigma^2}{n}\right)$

- ✓ Isto é, para n grande:
  - ✓ $\bar{X}$ tem distribuição Normal
  - ✓ Com esperança (média) igual à da variável original $\rightarrow E(X) = \mu$
  - ✓ Com variância igual à da variável original dividida por n $\rightarrow \ \mathrm{VAR}(X) = \frac{\sigma^2}{n}$
- ✓ Entretanto, se X já tiver distribuição Normal, $\bar{X}$ será Normal, não importa o valor de n

Em resumo...

Teorema Limite Central (TLC)

$$\bar{X} \xrightarrow{n\to\infty} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Longrightarrow \quad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- ✓ Frequentemente, precisamos trabalhar com esta quantidade, mas não conhecemos σ^2, apenas S^2.
- ✓ Então, utilizamos um resultado que é uma consequência do TLC:

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

- ✓ A distribuição t não depende de μ e σ^2, apenas do número de graus de liberdade

BIOESTATÍSTICA

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

Em resumo...

Parâmetro	*Intervalo de confiança* *com coeficiente de confiança γ*
μ $\left(\begin{array}{l}\text{variância conhecida} \\ \quad - \text{pouco provável!}\end{array}\right)$	$\bar{X} \pm z_{\gamma} \dfrac{\sigma}{\sqrt{n}}$
μ (variância desconhecida)	$\bar{X} \pm t_{\gamma} \dfrac{S}{\sqrt{n}}$
p (otimista)	$\hat{p} \pm z_{\gamma} \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
p (conservativo)	$\hat{p} \pm z_{\gamma} \sqrt{\dfrac{1}{4n}}$

Interpretação de um IC

A probabilidade de que o IC contenha o verdadeiro valor do parâmetro é γ.

BIOESTATÍSTICA

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