PÓS-GRADUAÇÃO EM 2025
PUC MINAS

Disciplina:

TÉCNICAS DE AMOSTRAGEM E REGRESSÃO LINEAR

Professora: Anaíle Mendes Rabelo

Inferência Estatística

COMPONENTES DE UM MODELO PROBABILÍSTICO

Variáveis Aleatórias

Resultado numérico da observação de um fenômeno aleatórios

Discreta

Contínuas

MODELOS PROBABILÍSTICOS

Parâmentro

Variável que é parte da distribuição de probabilidade

Discreta

Contínuas

INTRODUÇÃO A INFERÊNCIA ESTATÍSTICA

População

Característica

Amostra

Característica

População representada por meio de uma distribuição de probabilidade.

A partir da amostra queremos estimar os parametros populacionais

PROBLEMA:

Conhecer o parametro do modelo probab. -> Conseguimos caracterizar a V.A.

Ou seja, dizer a prob. de ocorrência de cada um dos seus potenciais valores

INFERÊNCIA ESTATÍSTICA

Contextualização:

- Qual a proporção da população que desenvolveu anticorpos contra a Dengue?

O que temos que identificar

- **Variável aleatória**
- **Valores que a v.a pode assumir**

- Qual a distribuição de probabilidade do espaço amostral?

$$P(Y = y) = p^y(1 - p)^{1-y}$$

PENSAMENTO ESTATÍSTICO

Determinar p

Censo (p – populacional)

Amostra (p – amostral)

A proporção obtida na amostra é diferente da população.

- **Incerteza** associada ao valor da proporção devido a termos apenas uma amostra.

- Como tomar uma decisão baseada apenas na amostra?

Descrição probabilística da estatística de interesse → Distribuição amostral

ESPECIFICANDO O PROBLEMA

- Y: desenvolveu anticorpos (v.a.).
- Modelo: Y ~ Ber (p) .
- Informação sobre p por meio de uma amostra da população.
- Objetivos da inferência estatística:
  - Estimar p baseado apenas na amostra **-> Estimativa pontual**
  - Medir a precisão do valor estimado **-> Estimativa intervalar**

ESPECIFICANDO O PROBLEMA

- Dado: uma amostra (n=10) , temos que 6 pessoas apresentaram anticorpos para dengue.

- Qual valor o parâmetro p?

- Assumindo observações independentes: Bernoulli -> n = 10 -> Binomial

- Qual a probabilidade de observar y = 6 para um valor de *p = 0,7*

$$P(Y = 6 \mid n = 10; p = 0{,}7) = \binom{10}{6} 0{,}7^{6} \, (1 - 0{,}7)^{10-6} = 0{,}14$$

PUC Minas

ESPECIFICANDO O PROBLEMA

Pode-se obter a probabilidade para qualquer outro valor de *p* :

$$P(Y = 6 \mid n = 10; p) = \binom{10}{6} p^6 (1-p)^{10-6}$$

A função de verossimilhança pode ser obtida ao variarmos *p* :

$$L(p) \equiv P(Y = 6 \mid n = 10; p) = \binom{10}{6} p^6 (1-p)^{10-6}$$

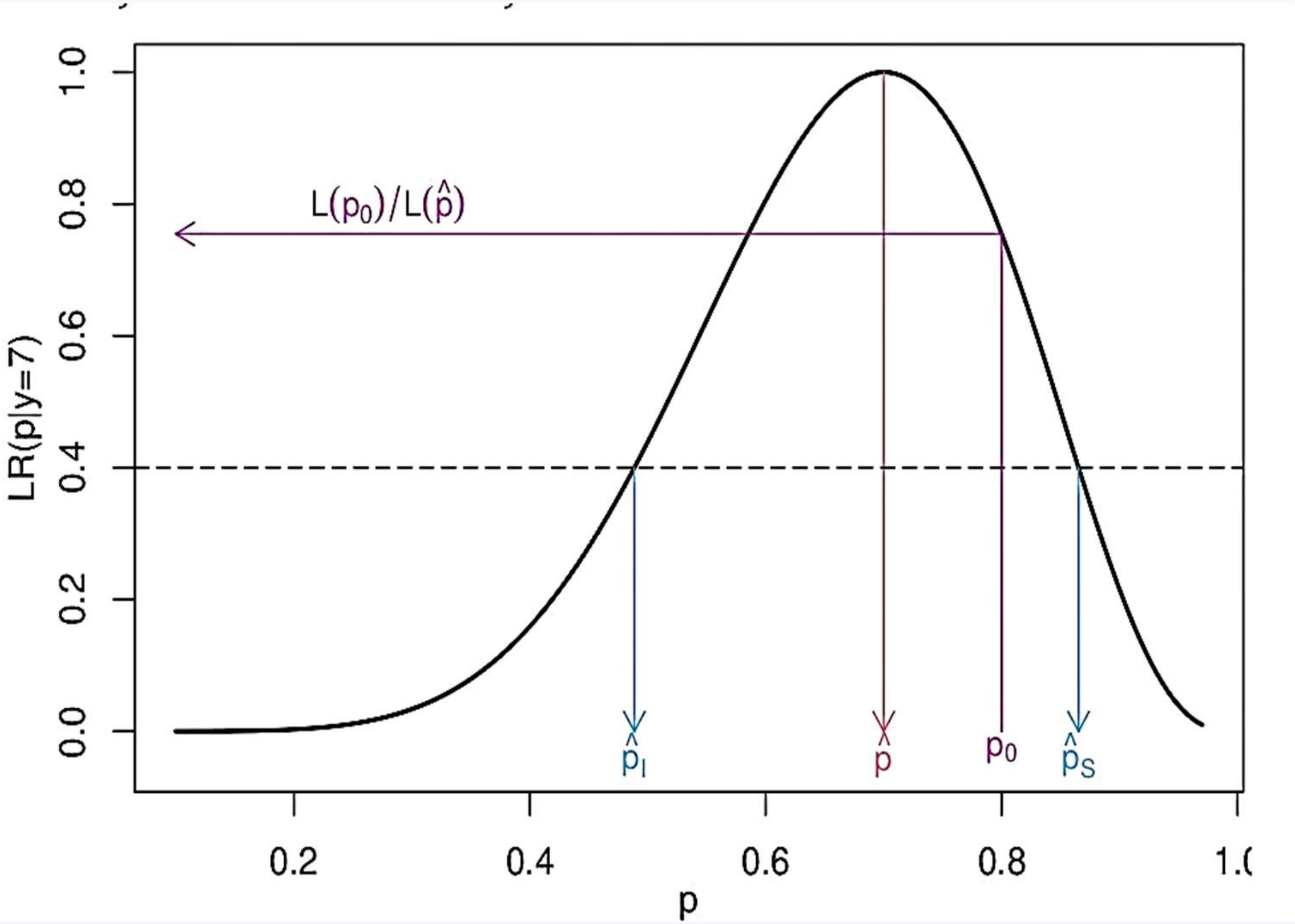
Queremos saber qual a probabilidade de obter os resultados da amostra dado um determinado valor de p.

VEROSSIMILHANÇA

Função de verossimilhança: probabilidade da amostra ser obtida dado diferentes valores de p:

- **Melhor estimador** e **estimativa**;
- **Incerteza associada** à estimativa obtida;
- Conjunto de **valores razoavelmente compatíveis com a amostra**;
- Decidir entre dois valores qual é o mais compatível com a amostra;
- Decidir se a amostra é compatível com certo valor de p de interesse.

VEROSSIMILHANÇA



- Quanto um valor particular de um parâmetro é compatível com a minha amostra.
- Estimativa pontual: Qual o valor mais compatível (máximo)
- Estimativa intervalar: Valores que são razoavelmente compatíveis com minha amostra

ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA

- Função de verossimilhança:

$$L(p) \equiv P_p[Y = 6] = \binom{10}{6} p^6 (1-p)^{10-6}$$

- Função de log-verossimilhança:

$$l(\theta) = \log(L(p)) = \log\binom{10}{6} + 6\log p + (10-6)\log(1-p)$$

- Derivando em relação a *p* (função escore):

$$U(p) = \frac{6}{p} - \frac{10-6}{1-p}$$

PUC Minas

ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA

- Igualando a zero, temos:

$$\hat{p} = \frac{6}{10} = 0,6$$

- Estimativa de máxima verossimilhança:

$$\hat{p} = \frac{y}{n}$$

- Estimador de máxima verossimilhança:

$$\hat{p} = \frac{Y}{n}$$

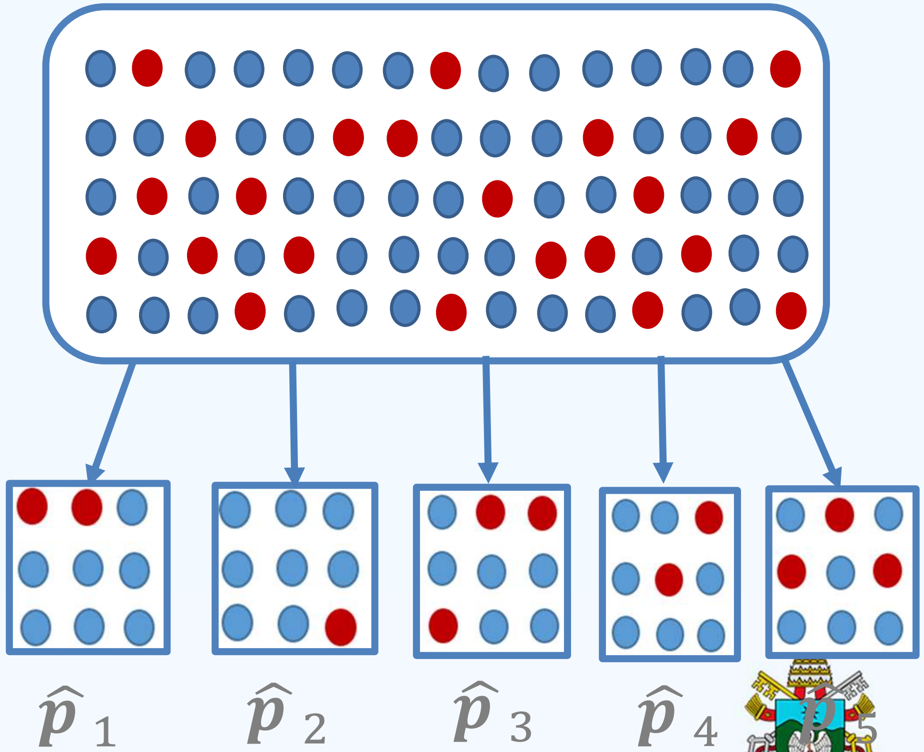
Estatística Frequentista

PENSAMENTO FREQUENTISTA

- Quando o experimento é **repetido** um **grande número de vezes** . Termos um p para cada realização. Logo, p estimado é uma variável aleatória.

- Em uma Variável aleatória temos que seu comportamento é dado pela **distribuição de probabilidade.**

- Definição da distribuição, e seus parâmetros



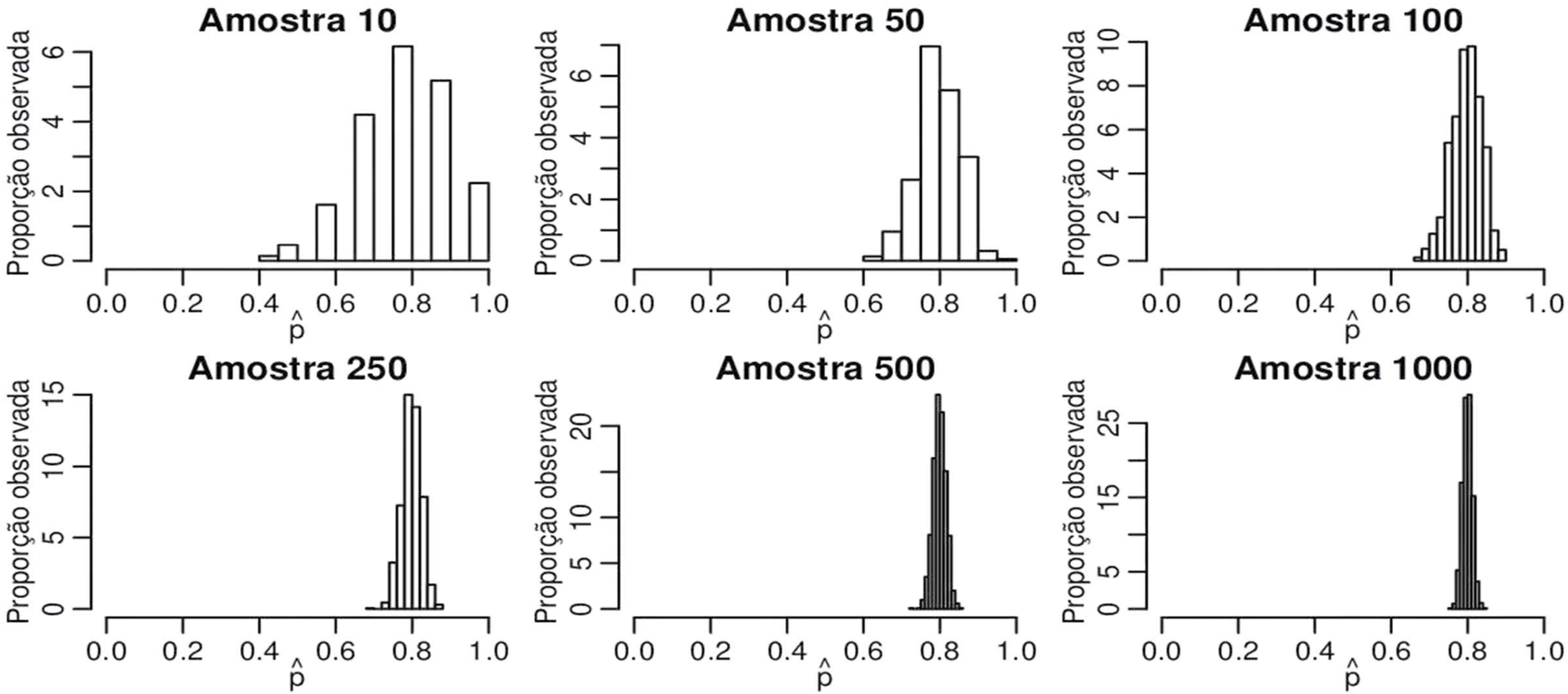
EXEMPLIFICANDO COMPUTACIONALMENTE

- Considerando p = 0,7 existe uma probabilidade considerável de observarmos “apenas” 6 pessoas imunizadas entre 10 pessoas.

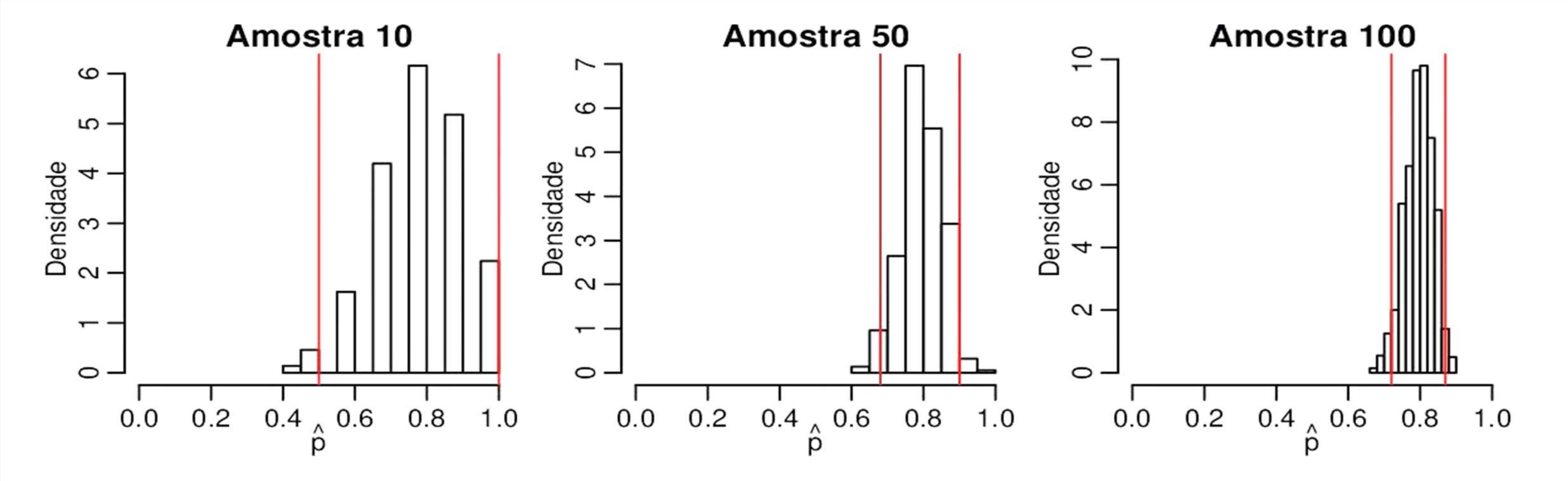
- A incerteza associada ao valor de p no caso de apenas 10 observações é grande.

DIMINUIR A INCERTEZA → AUMENTAR A AMOSTRA

EXEMPLIFICANDO COMPUTACIONALMENTE



INTERVALO DE CONFIANÇA



Qual o intervalo em que p estimado tem uma probabilidade, digamos (1-α) de pertencer. Valores comuns de α: 5% e 1%.

ESTATÍSTICA FREQUENTISTA

Inviabilidade de realizar replicações em termos práticos → Estimador é função da variável aleatória → Distribuição de Probabilidade

A distribuição amostral do estimador pode ser utilizada para estudar o resultado de múltiplas replicações do estudo.

Dificuldade em obter a distribuição exata de um estimador

O Teorema central do limite oferece uma aproximação para amostras grandes (assintótica).

PUC Minas

TEOREMA CENTRAL DO LIMITE(TCL)

Teorema de Lindeberg-Levy:

Seja $Y_1, Y_2, \ldots, Y_n$ uma amostra i.i.d com $E(Y_i) = \mu$ e $Var(Y_i) = \sigma^2 < \infty$:

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) \rightarrow Z \sim N(0,1), \text{ para } \mathbf{n} \rightarrow \infty$$

Ou seja, para todo $y \in \mathbb{R}$

$$\mathbf{P}(Y_n \leq \mathbf{y}) \rightarrow \Phi(y) \text{ quando } \mathbf{n} \rightarrow \infty$$

onde:

$$\left.\begin{aligned} &\Phi(y) = \int_{-\infty}^{y} \phi(z)dz \\ &\text{e } \phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}z^2\right) \end{aligned}\right\} \quad \bar{Y} \sim N(\mu, \sigma^2/n)$$

PUC Minas

APLICANDO O T.C.L. PARA A BINOMIAL

Estimador de máxima verossimilhança:

$$\hat{p} = \frac{Y}{n}$$

Aplicando as propriedade da distribuição binomial temos:

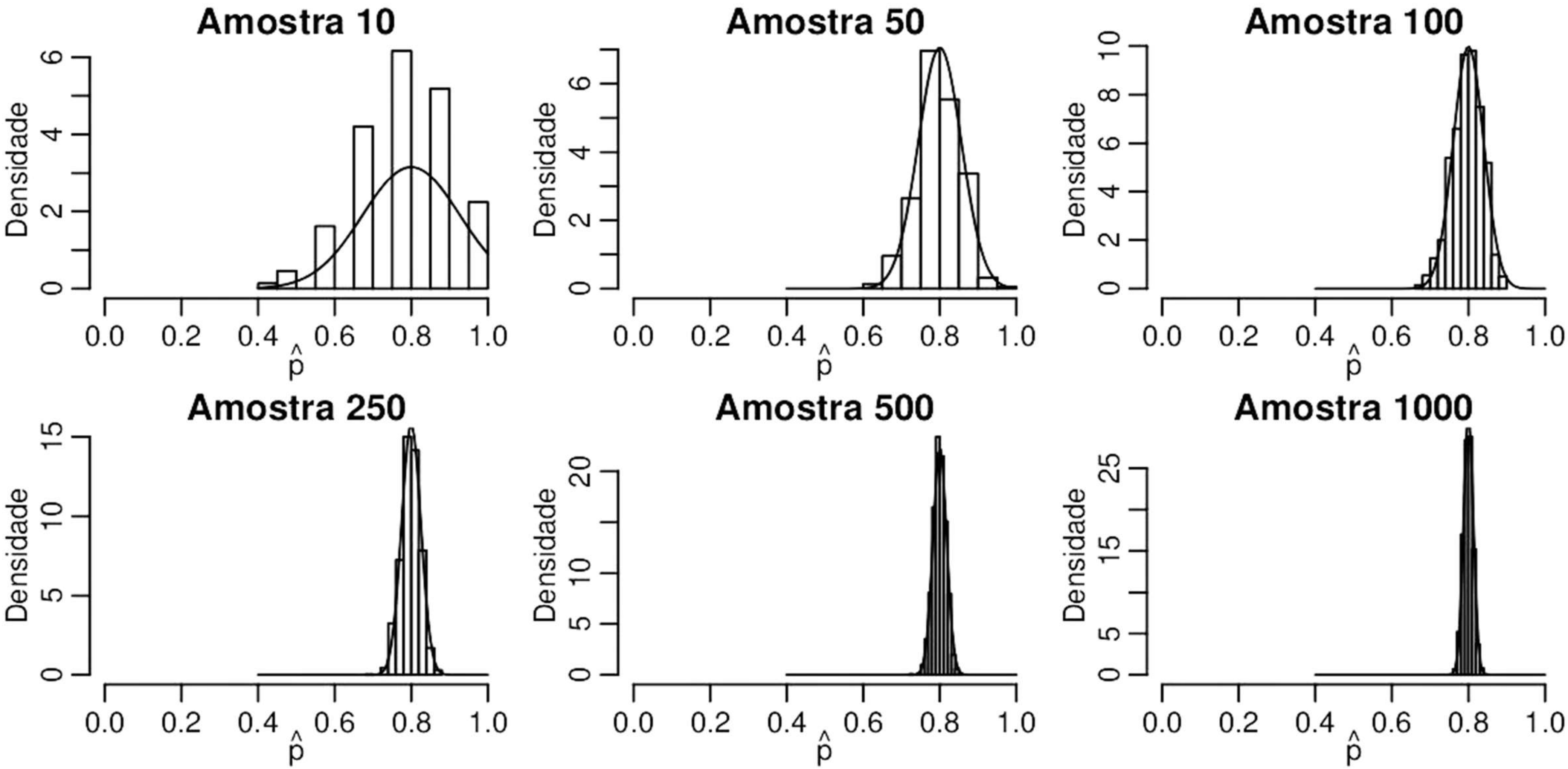
$$E(\hat{p}) = \frac{Y}{n} = \frac{1}{n} E(Y) = \frac{np}{n} = p$$

$$\text{Var}(\hat{p}) = \text{Var}\left(\frac{Y}{n}\right) = \frac{1}{n^2} Var(Y) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Utilizando o TCL, temos:

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

APLICAÇÃO COMPUTACIONAL



INTERVALO DE CONFIANÇA PARA A MÉDIA COM VARIÂNCIA CONHECIDA

I.C PARA A MÉDIA QUANDO σ^2 É CONHECIDO

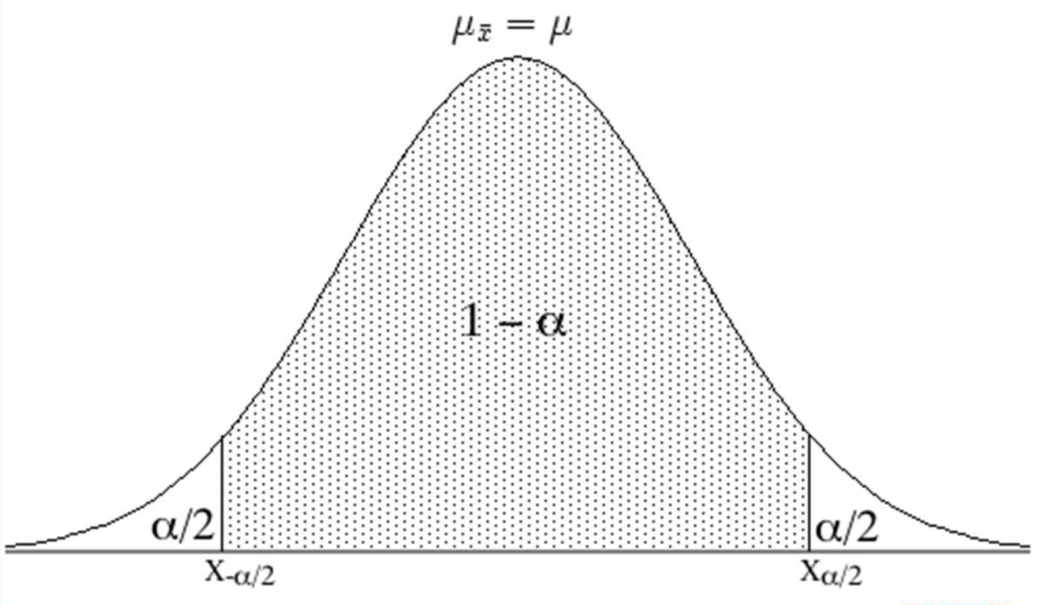
Seja $Y_i \sim N(\mu, \sigma^2)$ e suponha que σ^2 é conhecido.

Logo, temos que:

$$\bar{Y} \sim N(\mu, \sigma^2/n) \text{ ou } \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Fixando uma probabilidade 1 – α, podemos encontrar $\bar{y}_{LI}$ e $\bar{y}_{Ls}$, tal que :

$$P(\bar{y}_{LI} < \mu < \bar{y}_{Ls}) = 1 - \alpha$$



Fonte: https://pt.wikipedia.org/wiki/Intervalo_de_confian%C3%A7a

OBTENDO UM INTERVALO PARA μ

Definimos limites Z na distribuição amostral padronizada

$$P(z_{LI} < \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} < z_{Ls}) = 1 - \alpha$$

Isolamos μ,

$$P(\bar{y} - z_{LI}\frac{\sigma}{\sqrt{n}} < \mu < \bar{y} + z_{LS}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Como deseja-se intervalos simétricos, então abs(z_{LI}) = abs(z_{LS}) = $z_{\alpha/2}$.Assim,

$$P(\bar{y} - z_{LI}\frac{\sigma}{\sqrt{n}} < \mu < \bar{y} + z_{LS}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$z_{\alpha/2}$ é o quantil da distribuição normal padrão para o valor de 1 – α

PUC Minas

MARGEM DE ERRO E NÍVEL DE CONFIANÇA

A margem de erro pode ser definida por : $\mathbf{e} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$\mathbf{e} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad \mathbf{e} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{y} - z_{LI}\,\frac{\sigma}{\sqrt{n}} \qquad \bar{y} \qquad \bar{y} + z_{LS}\,\frac{\sigma}{\sqrt{n}}$$

Onde:

$z_{\alpha/2}$ é chamado de valor crítico.

$1 - \alpha$ é o nível de confiança do intervalo.

PUC Minas

EXEMPLO

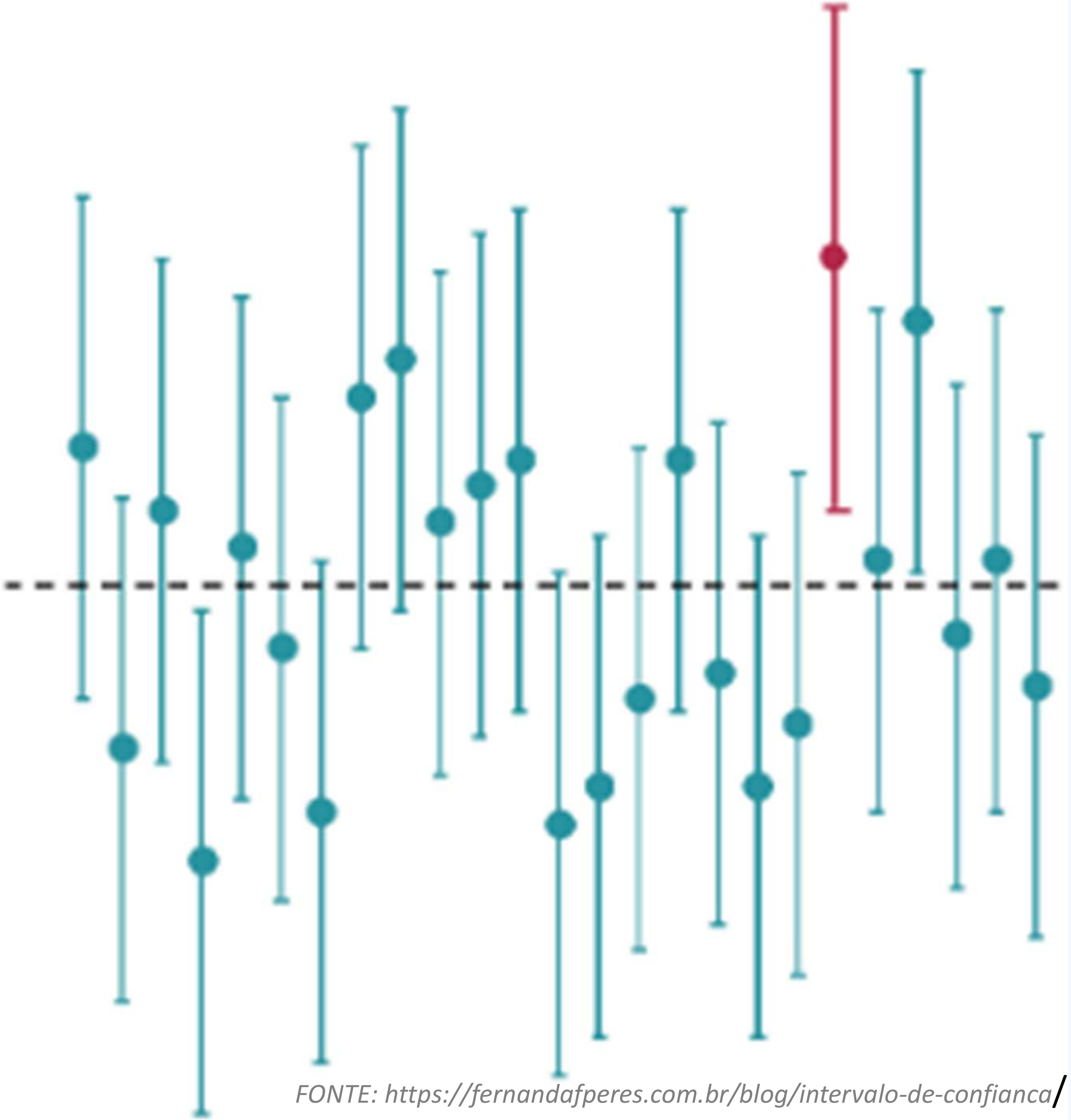
Y: Idade dos alunos de pós graduação // Sendo: y = (40,35,30,28,27)

$Y \sim N(\mu, \sigma^2)$ -> $\sigma^2 = 4^2$

- $\bar{y}$ = 32
- Ao nível de 95% de confiança : 1 – α = 0,95 ; α/2 = 0,025
- Valor – z : $z_{\alpha/2}$ = 1.96
- Margem de erro: e = $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ = 1.96 . $\frac{4}{\sqrt{5}}$ = 3.51

I.C. (95%): $\bar{y} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ = 32 $\pm$3.51 -> $\boldsymbol{IC}_{95\%}(\boldsymbol{\mu})$: $(\mathbf{28.5, 35.5})$

PUC Minas



FONTE: https://fernandafperes.com.br/blog/intervalo-de-confianca/

ENTENDIMENTO DO INTERVALO DE CONFIANÇA

Se repetirmos o experimento N vezes, e calculássemos o IC , **95% desses intervalos irão conter o verdadeiro parâmetro populacional**

PUC Minas

INTERPRETAÇÃO DO INTERVALO DE CONFIANÇA

Considerando um intervalo de confiança: $\boldsymbol{IC}_{95\%}(\boldsymbol{\mu})$: $[\bar{y}_{LI}; \bar{y}_{LS}]$

Interpretação Errada	Interpretação Certa
• Temos 95% de confiança de que a média populacional(μ) se encontra entre $\bar{y}_{LI}$ e $\bar{y}_{LS}$.	• Temos 95% de confiança que o intervalo entre $\bar{y}_{LI}$ e $\bar{y}_{LS}$ contém a média populacional(μ).

Aparentemente podem parecer iguais, porém é crucial observar que:

o intervalo é aleatório e o parâmetro é fixo.

PUC Minas

INTERVALO DE CONFIANÇA PARA A MÉDIA COM VARIÂNCIA DESCONHECIDA

IC PARA A MÉDIA QUANDO σ^2 É DESCONHECIDO

Seja $Y_i \sim N(\mu, \sigma^2)$ e suponha que σ^2 é desconhecido.
Logo, temos que:

$$\mathbf{t} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim t_{n-1}$$

Em que t_{n-1} é uma distribuição t-Student com *n-1* graus de liberdade.
Fixando uma probabilidade 1 – α, podemos encontrar $\bar{y}_{LI}$ e $\bar{y}_{LS}$, tal que :

$$P(\bar{y} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{y} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}) = 1 - \alpha$$

Onde s é o desvio padrão amostral

EXEMPLIFICANDO

Dado uma amostra de 15 empregados cujo salário médio desta amostra é de R$5.900,00 e o desvio padrão é de R$ 3.058,00. Calcule o IC com 95% de confiança.

Temos que $\boldsymbol{t_{\alpha/2} = t_{0,025} = 2,145}$ e com 15-1 = 14, graus de Liberdade.

Logo o interval de confiança é dado por ≈ (R$ 4.206,4; R$ 7.593,6)

$$\boldsymbol{IC_{1-0.95}(\mu) = \left(5.900 - 2.145 \,.\, \frac{3.058}{\sqrt{15}} < \mu < 5.900 + 2.145 \,.\, \frac{3.058}{\sqrt{15}}\right)}$$

PUC Minas

RESUMO DO MÉTODO

Verificação das suposições

- Amostra aleatória simples.
- Estimativa de s.
- A população tem distribuição Normal ou n>30

Determine o nível de confiança 1-α , e encontre o valor crítico t(α/2) .

Calcular a margem de erro.

Calcular o IC

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Seja $Y_i \sim Ber(p)$. Pelo teorema Central do limite temos que:
Logo, temos que:

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

Para o parâmetro p, temos que:

$$\mathrm{P}\left(\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} < p < \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right) = 1 - \alpha$$

PUC Minas

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Um fato que deve ser observada é que não conhecemos o verdadeiro valor de p(populacional) para calcular o IC.

Para resolver esse problema podemos:

- ❑ Utilizar o $\hat{p}$ -> estimativa Otimista
- ❑ Considerar $p = 0.5$ -> estimativa conservadora.

Quando $p = 0.5$ o termo p(1-p) terá o valor máximo

p	(1-p)	p(1-p)
0.1	0.9	0.01
0.3	0.7	0.21
0.5	0.5	0.25
0.6	0.4	0.24
0.8	0.2	0.16

PUC Minas

EXEMPLO

Uma amostra com 1500 brasileiros foi selecionada para entender se eles acreditavam ou não na cura do câncer. 1.050 responderam que sim.

Calcule o IC com 95% de confiança para :

- Estimativa Otimista -> p = p estimado
- Estimativa conservadora. -> p = 0.5

EXEMPLO

- Estimativa pontual: $\hat{p} = \frac{1.050}{1.500}$
- Intervalo 1: Otimista $\approx (0{,}677, 0{,}723)$

$$IC_{95\%}(p) = \left(0,7 \; - 1,96 \; . \sqrt{\frac{0,7\,(1-0,7)}{1.500}} < p < 0,7 + \; 1,96 \; . \sqrt{\frac{0,7\,(1-0,7)}{1.500}}\right)$$

- Intervalo 2: Conservador $\approx (0{,}675, 0{,}725)$

$$IC_{95\%}(p) = \left(0,7 \; - 1,96 \; . \sqrt{\frac{0,5\,(1-0,5)}{1.500}} < p < 0,7 + \; 1,96 \; . \sqrt{\frac{0,5\,(1-0,5)}{1.500}}\right)$$

RESUMO DO MÉTODO

Verificação das suposições

- Amostra aleatória simples.
- Dois resultados possíveis ("sucesso", "fracasso")
- Premissas da dist. binomial:
  - Tentativas independentes
  - *p* constante

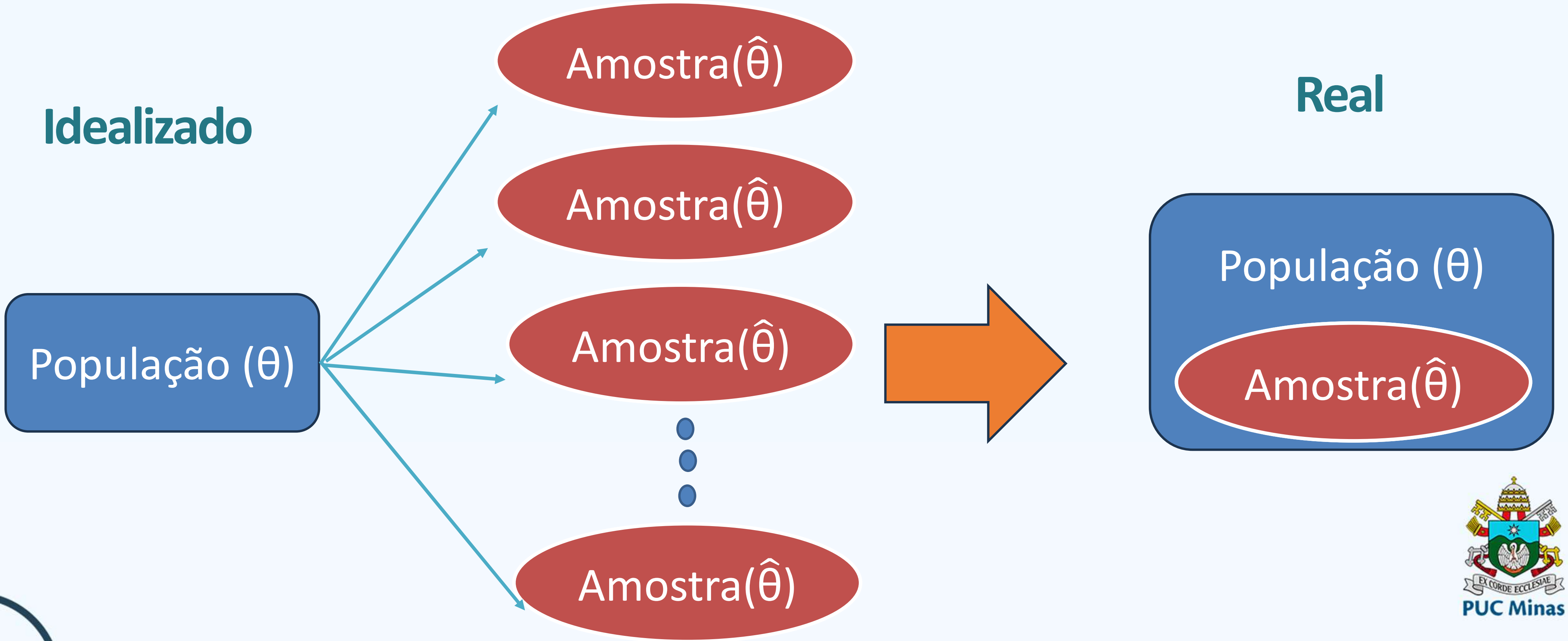
Determine o nível de confiança 1-α , e encontre o valor crítico z($\alpha/2$) .

Calcular a margem de erro com p=p estimado ou p=0.5

Calcular o IC

BOOTSTRAP

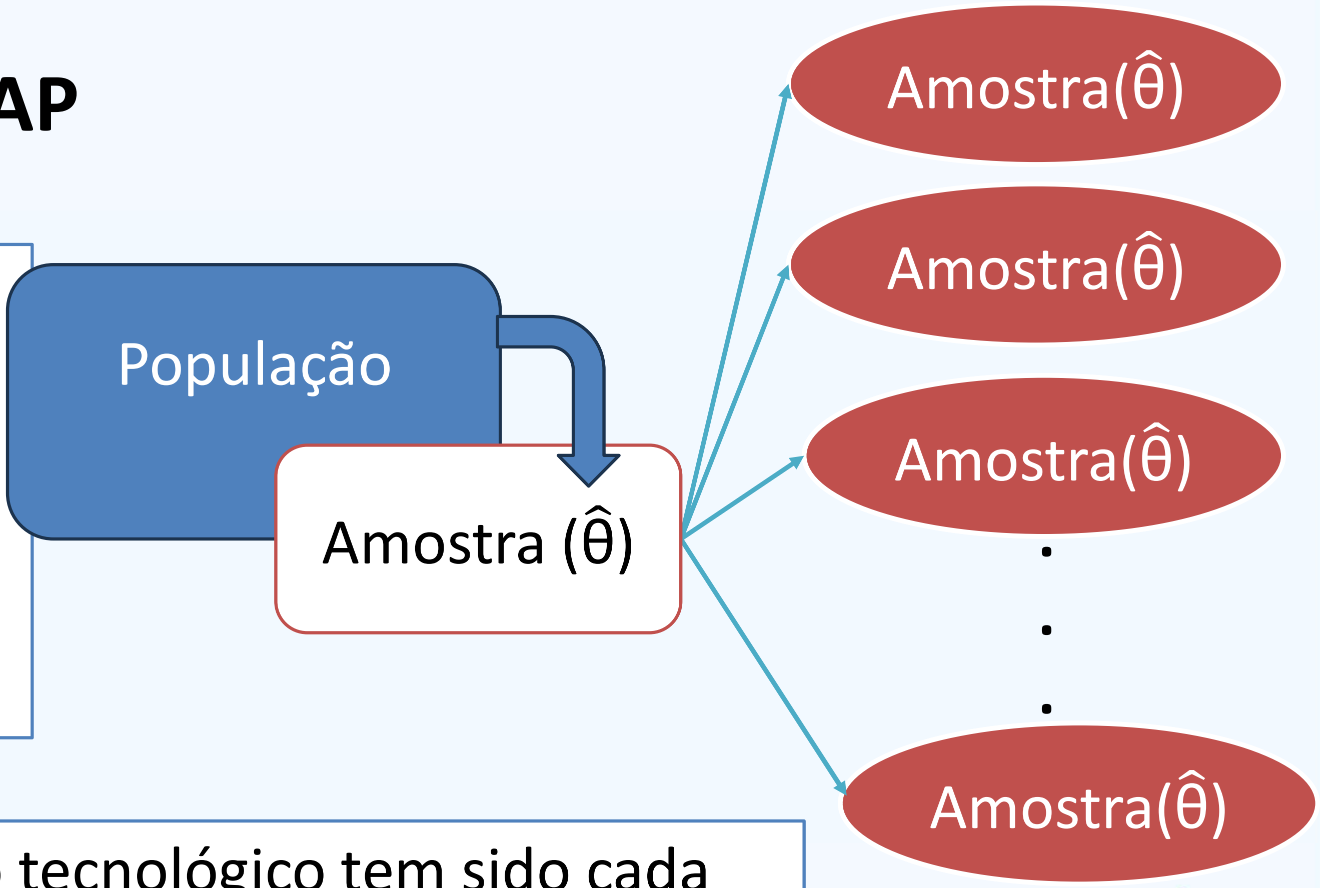
DEFINIÇÃO DE BOOTSTRAP



PUC Minas

DEFINIÇÃO DE BOOTSTRAP

- Técnica de **reamostragem**
- Sorteamos dados de uma amostra e **formamos novas amostras**.
- Reamostragem com reposição



Por conta do desenvolvimento tecnológico tem sido cada vez mais utilizada. Uma vez **que demanda uma grande quantidade de reamostragens.**

PUC Minas

REAMOSTRAGEM ALEATÓRIA COM REPETIÇÃO

Sabendo que temos 5 fichas na sacola “A”, “B”, “C”, “D”, “E”
Podemos extrair n amostras com reposição de tamanho 5:

- A,B,B,C,D
- A,B,C,A,E
- E,D,D,E,C

Espera-se um comportamento paramétrico próximo da amostra mestre.

Ilustração de como funciona o bootstrap:
Criamos um conjunto de mesmo tamanho da minha amostra a partir de seleção de amostra com reposição.

TIPOS DE BOOTSTRAP

Bootstrap

Paramétrico

Conhece-se a distribuição geradora das amostras. Parâmetros são estimados a partir da amostra original (AO) usada para estimar o chamado vício ou viés e corrigí-lo.

Não paramétrico

Dispensa o analista de premissas paramétricas para realizar inferências e fornece respostas a problemas para os quais não há soluções analíticas.

ETAPAS

Amostra Mestre

Amostra Original

Reamostragem

Seleção de várias reamostras com reposição

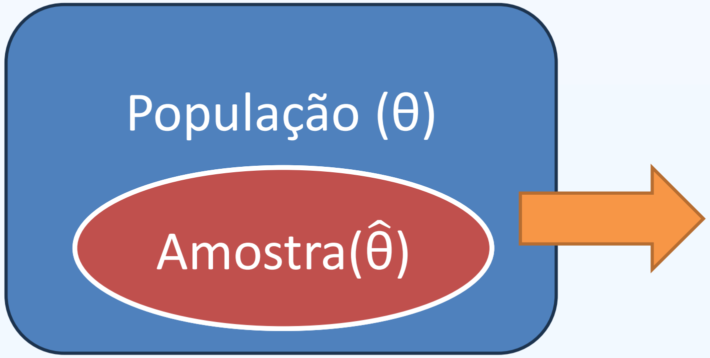
Estudos nas amostras

Cálculo das estatísticas de interesse em cada reamostra

Estimativa

Estimativas pontuais , intervalos de confiança

ESTATÍSTICAS



Amostra Original (AO)

$x_1, x_2, x_3, \dots, x_n \rightarrow \hat{\theta}$

Reamostragem:

$x_1^1, x_2^1, x_3^1, \dots, x_n^1 \rightarrow \hat{\theta}^1$

$x_1^2, x_2^2, x_3^2, \dots, x_n^2 \rightarrow \hat{\theta}^2$

.

.

$x_1^k, x_2^k, x_3^k, \dots, x_n^k \rightarrow \hat{\theta}^k$

$\bar{\hat{\theta}}$

$s_{\hat{\theta}}$

VIÉS = $\bar{\hat{\theta}} - \hat{\theta}$

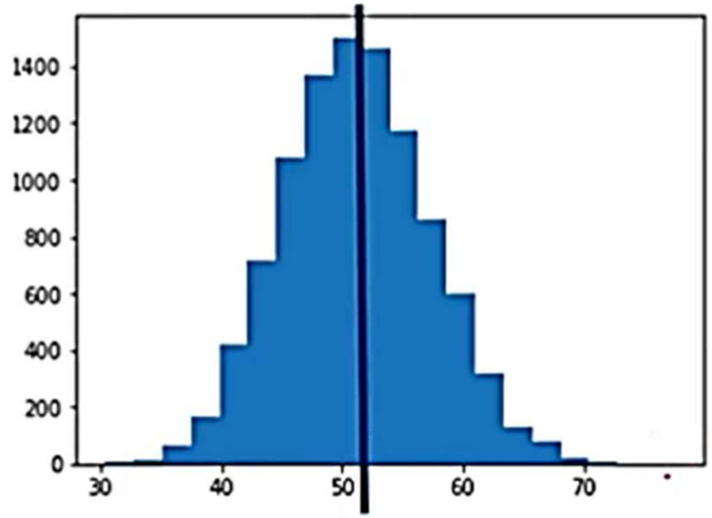
Requisito fundamental é que cada reamostra deve ser uma amostra independente e identicamente distribuída da distribuição empírica da amostra original

VIÉS

$$\text{viés} = \bar{\hat{\theta}} - \hat{\theta}$$

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

$$\hat{\theta}$$



$$\bar{\hat{\theta}} \quad \mu_{\bar{X}}$$

$$\text{viés} = \mu_{\bar{X}} - \bar{X}$$

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

$$\bar{X}$$

O vicio permite verificar se a distribuição bootstrap está centrada na estatística AO.

INTERVALO DE CONFIANÇA COM O BOOTSTRAP

Técnicas para estimar o intervalo de confiança via bootstrap:

- Bootstrap t
- Bootstrap percentil
- Bootstrap BCPB – com correção de vies
- Bootstrap BCa – com correção de viés acelerado

Se o viés e assimetria são muito fortes , recomendam-se métodos de correção

INTERVALOS DE CONFIANÇA COM BOOTSTRAP

PUC Minas

BOOTSTRAP T

$$IC\ \theta = \bar{\hat{\theta}} \pm t_c \,.\, s_{\hat{\theta}}$$

Aplicável para estimativa de locação:
- Média
- Mediana
- Quartis

Funciona bem quando a distribuição da estatística Bootstrap é aproximadamente normal e a estatística tem viés baixo

BOOTSTRAP PERCENTIL 1ª FORMA

$IC\ \theta = 95\%\ de\ conf.\ percentil\ 2{,}5\ e\ 97{,}5$

Funciona bem quando o viés baixo

$$
\left.
\begin{array}{l}
x_1^1, x_2^1, x_3^1, \ldots, x_n^1 \rightarrow \hat{\theta}^1 \\
x_1^2, x_2^2, x_3^2, \ldots, x_n^2 \rightarrow \hat{\theta}^2 \\
. \\
. \\
. \\
x_1^k, x_2^k, x_3^k, \ldots, x_n^k \rightarrow \hat{\theta}^k
\end{array}
\right\}
\begin{array}{l}
p_{2,5} \\
\\
p_{97,5}
\end{array}
$$

BOOTSTRAP PERCENTIL 2ª FORMA

$$IC\ \theta = \hat{\theta} - pd_{97,5}\ ;\ \hat{\theta} - pd_{2,5}$$

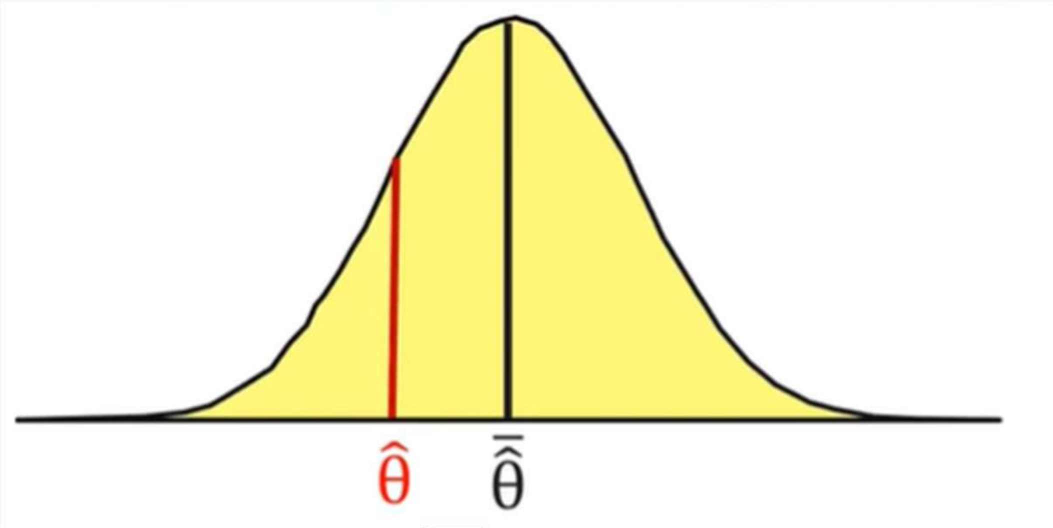
Funciona bem quando o viés baixo

$$\left.\begin{array}{l} x_1^1, x_2^1, x_3^1, \ldots, x_n^1 \rightarrow \hat{\theta}^1 - \hat{\theta} = d_1 \\ x_1^2, x_2^2, x_3^2, \ldots, x_n^2 \rightarrow \hat{\theta}^2 - \hat{\theta} = d_2 \\ . \\ . \\ . \\ x_1^k, x_2^k, x_3^k, \ldots, x_n^k \rightarrow \hat{\theta}^k - \hat{\theta} = d_k \end{array}\right\} \begin{array}{l} pd_{2,5} \\ \\ pd_{97,5} \end{array}$$

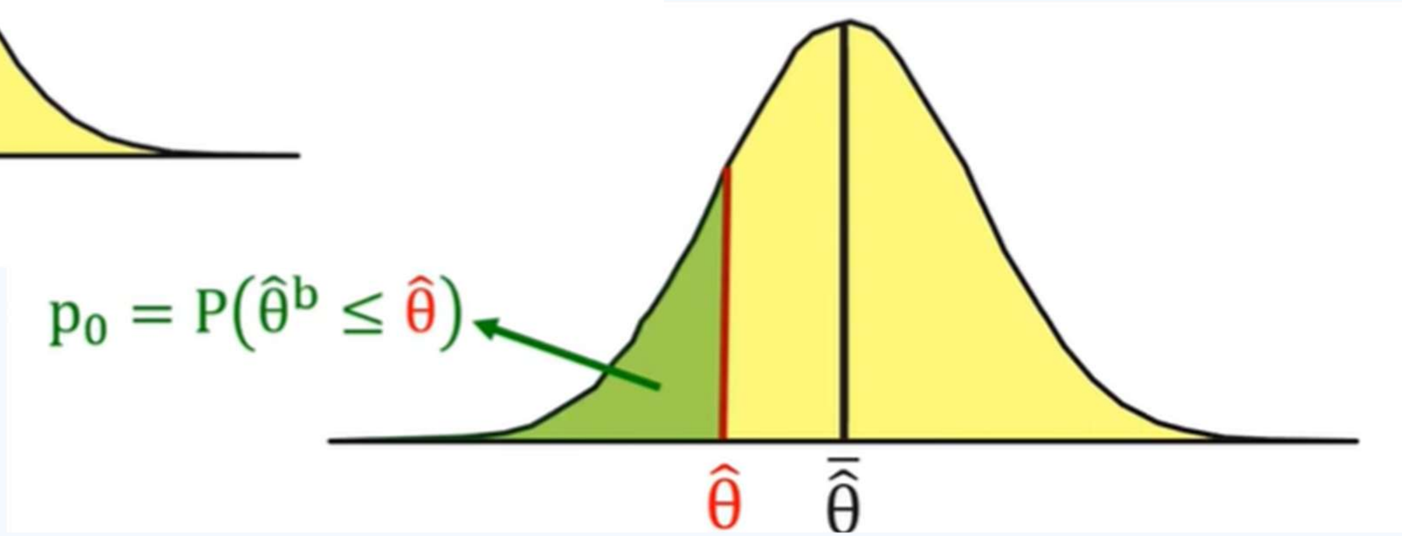
PUC Minas

BOOTSTRAP BCPB - COM CORREÇÃO DE VIÉS

Distribuição simulada a partir da reamostragem

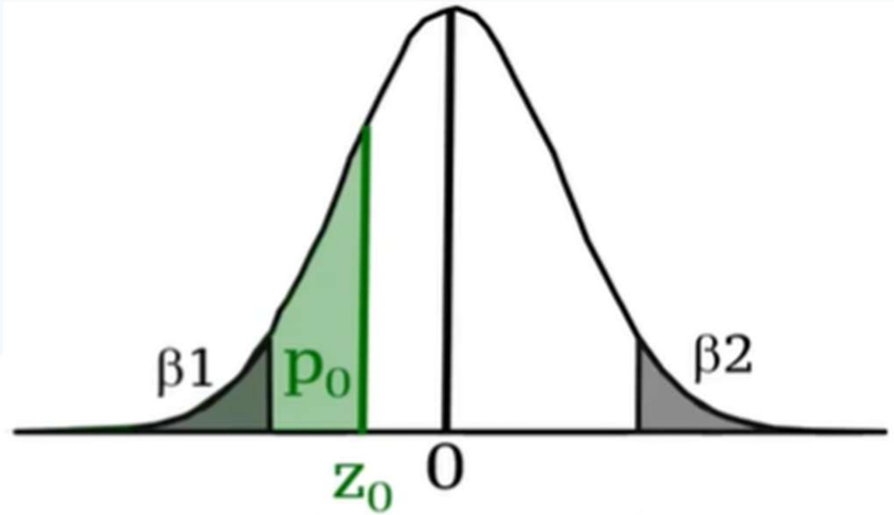


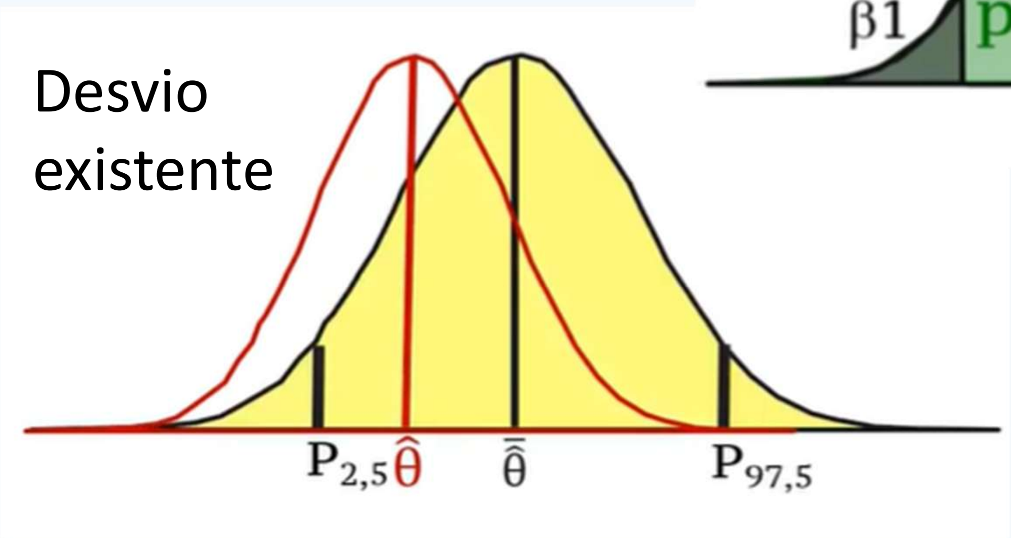
- z_0 será o corretor de viés ,
- Qual o P_0 corresponde a ele na curva normal padrão?

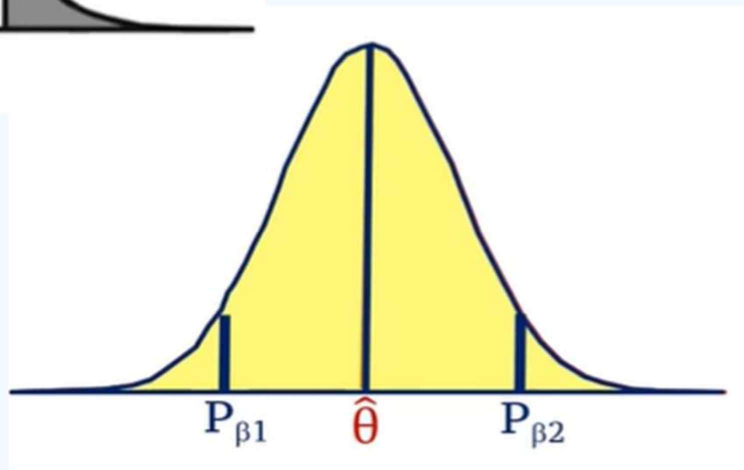


BOOTSTRAP BCPB - COM CORREÇÃO DE VIÉS

Com z_0, conseguiremos calcular os percentis ajustados (β_1 e β_2)







Sem desvio

BOOTSTRAP BCPB - COM CORREÇÃO DE VIÉS

Para nível de confiança de 95% : ajustar os percentis 2,5% e 97,5%, a fim de corrigir o viés e assimetria

1) Ordenar de forma crescente $\hat{\theta}^1$, $\hat{\theta}^2$, $\hat{\theta}^3$....
2) Calcular a probabilidade $p_0 = \ P(\hat{\theta}^{\mathrm{n}} \leq \ \hat{\theta})$ para n = 1,2, ...,n
3) Calcular o vício pela inversa da normal no ponto p_0

$$z_0 = \ \emptyset^{-1}(p_0)$$

4) Calcular os percentis β_1 e β_2

$$\beta_1 = \emptyset(2.\, z_0 - 1{,}96) \text{ e } \beta_2 = \emptyset(2.\, z_0 - 1{,}96)$$

Calcular o intervalo de confiança com PI e OS dos valores $\hat{\theta}^1$, $\hat{\theta}^2$, $\hat{\theta}^3$....

$$P_{\beta_1} \longleftrightarrow P_{\beta_2}$$

PUC Minas

BOOTSTRAP BCa - CORREÇÃO DE VIÉS ACELERADO

Obtido da mesma forma que o BCPB com os limites de Pβ_1 e Pβ_2 , adotando-se uma constante de aceleração

$$\beta_1 = \emptyset(z_0 - \frac{z_0 + 1{,}96}{1 - a\,(z_0 + 1{,}96)})\ ;\ \beta_2 = \emptyset(z_0 + \frac{z_0 + 1{,}96}{1 - a\,(z_0 + 1{,}96)})$$

Se a = 0 -> volta-se ao método BCPB

Se z_0 = 0 e a = 0-> volta-se para o método de percentil

PUC Minas

BOOTSTRAP BCa - CORREÇÃO DE VIÉS ACELERADO

Pode-se estimar a quando as v.a observadas são I.I.D

$$a = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\right)^3}{6 \cdot \left[\sum_{i=1}^{n}\left(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\right)^2\right]^{\frac{3}{2}}}$$

$\hat{\theta}_i$: O valor das estimativas do parâmetro estudado para cada amostra i

$\hat{\theta}_.$: Media dos valores de $\hat{\theta}_i$

PUC Minas

BOOTSTRAP BCa - CORREÇÃO DE VIÉS ACELERADO

Esse método faz três correções

1) Para não normalidade : através da função inversa de β_1 e β_2
2) Para o viés : através de z_0
3) Para o erro padrão não constante : através de a

TAMANHO DA AMOSTRA

DIMENSIONAMENTO DE AMOSTRAS

Dimensionar esforço, economizar recursos.

- Planejar pesquisas de opinião pública.
- Controle de qualidade.
- Estudos demográficos.
- Inspeções de qualidade da água.ultivadas.
- Teste de eficácia de vacinas

DEFINIÇÃO DA QUANTIDADE DE ELEMENTOS A SEREM SELECIONADOS

A determinação do tamanho de uma amostra é um problema de grande importância, porque **amostras desnecessariamente grandes acarretam desperdício de tempo e de dinheiro**; e **amostras demasiadamente pequenas podem levar a resultados não confiáveis**.

CÁLCULOS AMOSTRAIS

Tamanho da amostra para estimar a média.

- $Y \sim N(\mu, \sigma^2)$

Tamanho da amostra para estimar a proporção.

- $Y \sim Ber(p)$

TAMANHO DA AMOSTRA PARA ESTIMAR A MÉDIA

TAMANHO DA AMOSTRA PARA ESTIMAR A MÉDIA

Objetivo: estimar a média populacional μ.

Quantos elementos devemos amostrar?

Quantidade necessária que permita obter estimativas com uma certeza aceitável.

AMPLITUDE DO INTERVALO DE CONFIANÇA PARA A MÉDIA

O intervalo de confiança para média é:

$$\textbf{I.C. }(\boldsymbol{\mu})\textbf{:}\ \left(\bar{y} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ;\ \bar{y} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Amplitude do intervalo dada pela diferença entre o limite superior e inferior é:

$$\boldsymbol{A_{IC(\mu)} = 2 \,.\, z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

$$\bar{y} - z_{LI}\,\frac{\sigma}{\sqrt{n}}$$

$$\boldsymbol{z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

$$\boldsymbol{z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

$$\bar{y} + z_{LS}\,\frac{\sigma}{\sqrt{n}}$$

$$\bar{y}$$

COMPONENTES DO I.C. PARA A MÉDIA

Amplitude do intervalo tem 3 componentes:

$$A_{IC(\mu)} = 2 \, . \, z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

1 – Coeficiente de confiança 1 –α
2 – Desvio padrão populacional
3 – Tamanho da amostra n

PUC Minas

EFEITOS NA AMPLITUDE DO I.C. PARA A MÉDIA

$\uparrow 1 - \alpha \Rightarrow \uparrow z_{\frac{\alpha}{2}} \Rightarrow \uparrow A_{IC(\mu)}$

$\uparrow \sigma \Rightarrow \uparrow A_{IC(\mu)}$

$$A_{IC(\mu)} = 2 \,.\, z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\uparrow n \Rightarrow \downarrow A_{IC(\mu)}$

PUC Minas

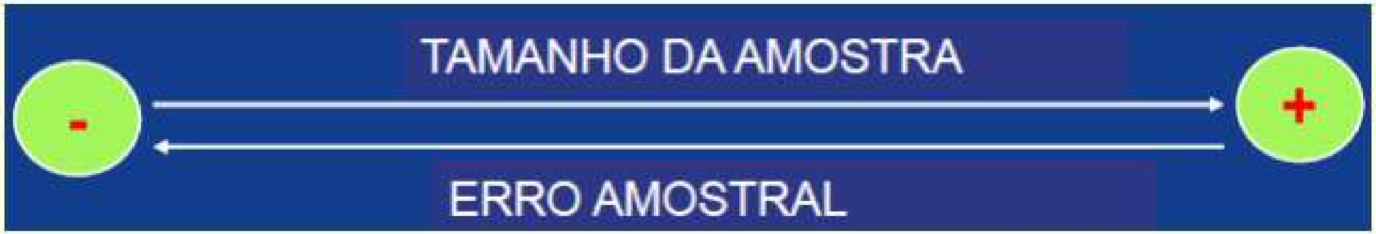
ERRO AMOSTRAL

Amostra não representa perfeitamente uma população -> aceitação de uma margem de erro que se denomina ERRO AMOSTRAL.

Erro Amostral é a diferença entre um resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais aleatórias.

EFEITOS NA AMPLITUDE DO I.C. PARA A MÉDIA

Não podemos evitar a ocorrência do ERRO AMOSTRAL, porém podemos limitar seu valor **através da escolha de uma amostra de tamanho adequado**.



PUC Minas

TRABALHANDO A MARGEM DE ERRO

Considerando a equação de erro máximo provável:

$$\boxed{e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

E fixando *e*, podemos obter *n* :

$$\boxed{n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e}\right)^2}$$

PUC Minas

TAMANHO DA AMOSTRA

$\uparrow 1 - \alpha \Rightarrow \uparrow z_{\frac{\alpha}{2}} \Rightarrow \uparrow n$

$\uparrow \sigma \Rightarrow \uparrow n$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e}\right)^2$$

$\uparrow e \Rightarrow \downarrow n$

O tamanho da amostra (n) depende do:

- **Nível de confiança** ($1-\alpha$) desejado (expresso pelo valor crítico $z_{\alpha/2}$).
- **Desvio-padrão** (σ) (embora veremos que não é estritamente necessário).
- **Erro máximo admitido** (e).

PUC Minas

EXEMPLO – TAMANHO DA AMOSTRA PARA MÉDIA

Considere : $Y \sim N(\mu\ , \sigma^2 = 25)$.

Calcule o tamanho da amostra, para que com 95% de probabilidade, a média amostral não distancie da média populacional por mais de 1 unidade e 2 unidades.

Logo temos $\sigma = 5$, $z_{0.025}$ = *1,96.* Dessa forma:

Para *e = 1 :*

$$n = \left(\frac{z_{\alpha/2} \cdot \ \sigma}{e}\right)^2 = \left(\frac{1.96 \ . \ \ 5}{1}\right)^2 \approx 96$$

Para e = 2 :

$$n = \left(\frac{z_{\alpha/2} \cdot \ \sigma}{e}\right)^2 = \left(\frac{1.96 \ . \ \ 5}{2}\right)^2 \approx 24$$

PUC Minas

TAMANHO DA AMOSTRA PARA ESTIMAR A PROPORÇÃO

TAMANHO DA AMOSTRA PARA ESTIMAR A PROPORÇÃO

Considerando a equação de erro máximo provável:

$$\boxed{e = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}}$$

E fixando *e*, podemos obter *n* :

$$\boxed{n = \left(\frac{z_{\alpha/2}}{e}\right)^2 . p(1-p)}$$

Quando não conhecemos p, usamos $\hat{p}$ (estimativa otimista) ou p =0.5 (estimativa conservadora).

EXEMPLO – TAMANHO DA AMOSTRA PARA PROPORÇÃO

Para $\mathrm{p} = 0.1$, $e = 0.03$, $z_{0.005} = 2.576$. Temos:

$$n = \left(\frac{2.576}{0.03}\right)^2 .0.1(1 - 0.1) \approx 664$$

Para $\mathrm{p} = 0.5$, $e = 0.03$, $z_{0.005} = 2.576$. Temos:

$$n = \left(\frac{2.576}{0.03}\right)^2 .0.5(1 - 0.5) \approx 1844$$

EX CORDE ECCLESIAE
PUC Minas