Aula 03. Intervalos de Confiança

Estatística Inferencial

MBA CDIA

ENAP - Escola Nacional de Administração Pública

2025

pop.

sample

$\mu\,(?)$

$\bar{x}$

$$EMV(\mu) = \frac{\sum x_i}{n} = \bar{x}$$

$\hat{p}$

μ

Agenda

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma}}$$

1. **Revisão: Estimação Pontual e EMV**
2. **Intervalos de Confiança**
   - Derivação do IC para média
   - IC para média com σ conhecido
   - IC para média com σ desconhecido
   - IC para proporção populacional
3. **Interpretação e Propriedades dos ICs**
4. **Determinação do Tamanho Amostral**
5. **Exercícios práticos em Python**

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\hat{\sigma}^2_{EMV} = \cancel{(n-1)}\,\sigma^2 \times \frac{1}{\cancel{n-1}}$$

Estimação Pontual:

- θ = parâmetro populacional
- $\hat{\theta}$ = estimador
- Propriedades desejáveis:
  - Não-viesado: $E[\hat{\theta}] = \theta$
  - Eficiente: menor variância
  - Consistente: $\hat{\theta} \to \theta$ quando $n \to \infty$

Máxima Verossimilhança (EMV):

- $L(\theta|dados) = \prod_{i=1}^{n} f(x_i|\theta)$
- $\hat{\theta}_{EMV} = \arg\max_{\theta} L(\theta)$
- Exemplos:
  - Normal: $\hat{\mu} = \bar{X}$
  - Exponencial: $\hat{\lambda} = 1/\bar{X}$
  - Bernoulli: $\hat{p} = \bar{X}$

Limitação: Uma estimativa pontual não fornece informação sobre a incerteza!

Intervalos de Confiança - Motivação

- Uma estimativa pontual $\hat{\theta}$ raramente será igual ao parâmetro populacional θ
- É mais informativo fornecer um **intervalo de valores plausíveis**
- O intervalo quantifica a incerteza da estimação

7.3 0.2 0.2 7.7

7.5

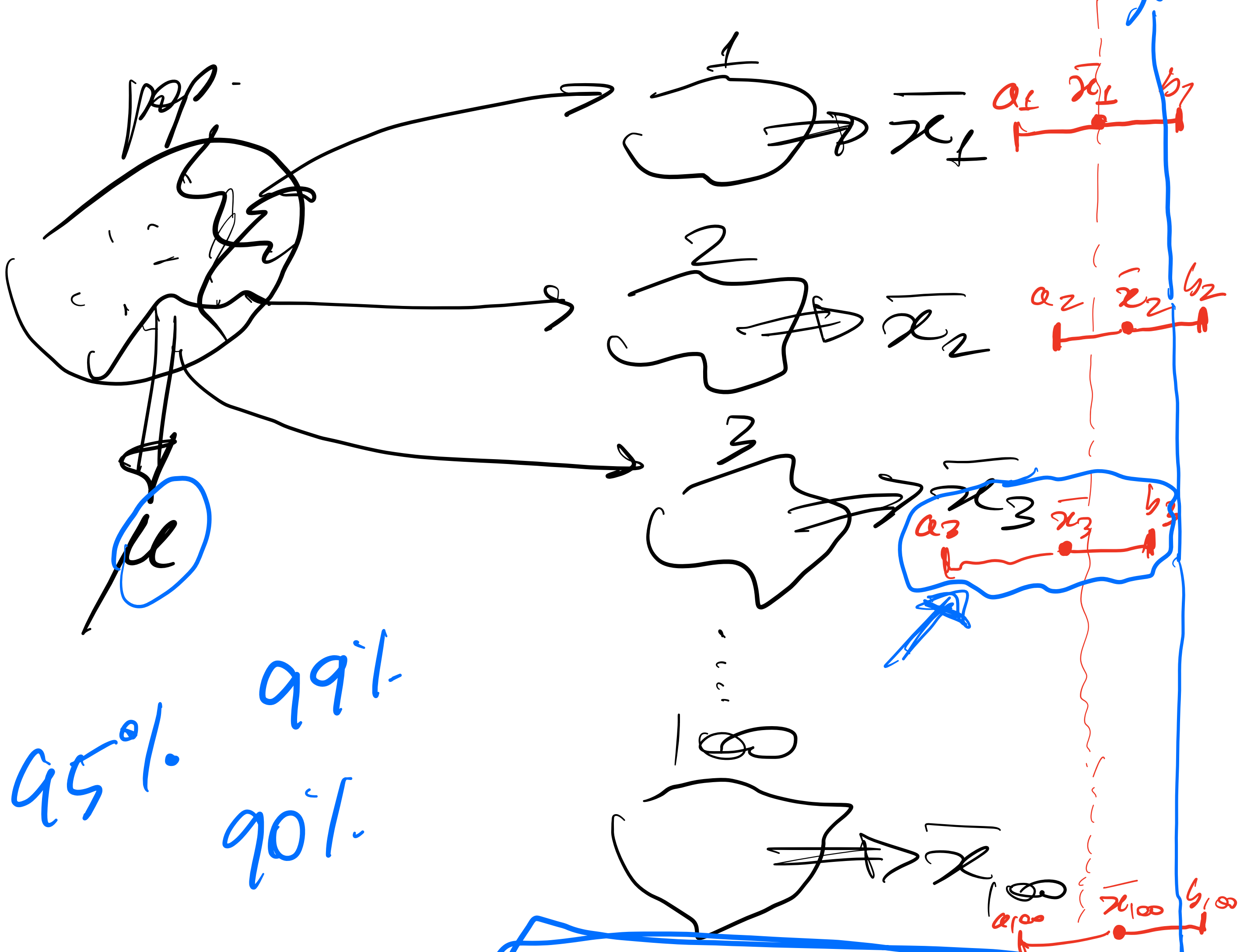
Exemplo prático:

- Estimativa pontual: "A satisfação média é 7.5"
- Intervalo de confiança: "Com 95% de confiança, a satisfação média está entre 7.3 e 7.7"

taxa de erro

Definição

Um intervalo de confiança de nível $(1 - \boxed{\alpha})$ para um parâmetro populacional é uma regra de construção de intervalos baseada em amostras aleatórias, tal que, se repetirmos o processo de amostragem muitas vezes, a proporção de intervalos que conterá o verdadeiro valor do parâmetro será $(1 - \alpha)$.



$(1-\alpha)\% = 95\%$

5%

BR adulto homem

μ

100%

$$\boxed{IC = \text{est pontual} \pm M.E}$$

Objetivo: Construir um IC para μ quando σ^2 é conhecido.

$$IC(\mu)_{0.95} = \bar{x} \underbrace{\pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{M.E.}$$

Sabemos que: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

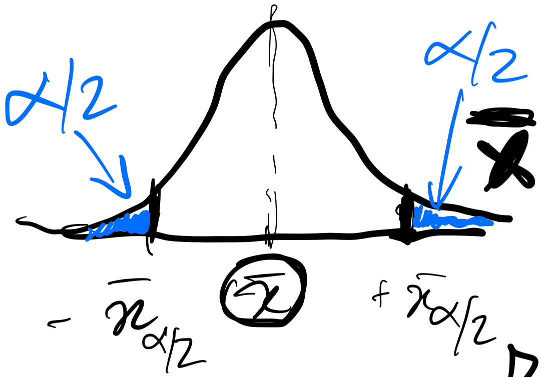
Padronizando: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Para 95% de confiança:

$$P(-1.96 < Z < 1.96) = 0.95$$

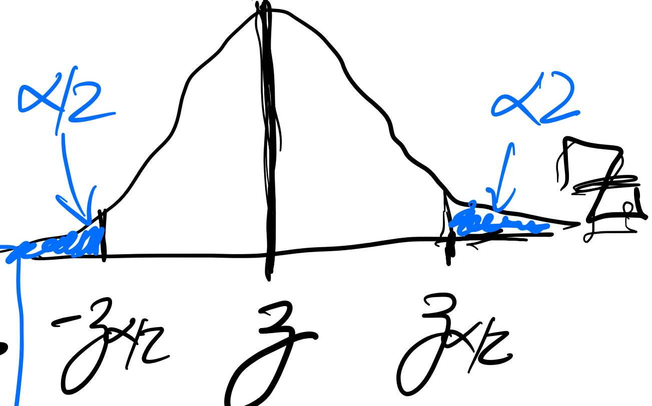
Substituindo Z**:**

$$\boxed{P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95}$$



$\Longleftrightarrow$

$$Z = \boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}$$



$$\Rightarrow P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

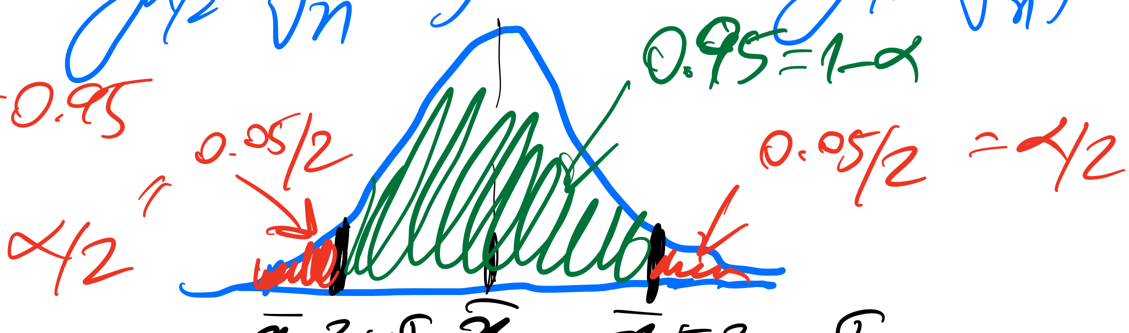
$$\times(-1) \qquad \times(-1) \qquad \times(-1)$$

$$\Rightarrow P\left(-\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} > \mu > \bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

ex: $1-\alpha = 0.95$



$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ $\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Multiplicando por $\sigma/\sqrt{n}$:

$$P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Subtraindo $\bar{X}$ e multiplicando por -1:

$$P\left(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

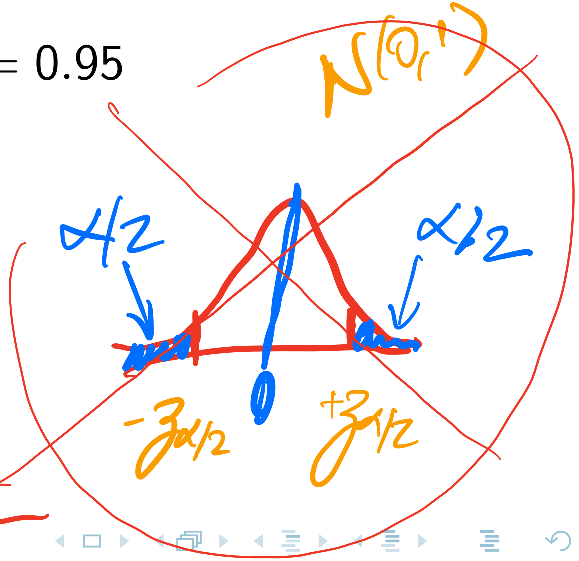
Forma geral do IC de $(1-\alpha)100\%$:

est. pontual

M.E.

$$\boxed{IC(\mu) = \bar{X} \pm z_{\alpha/2} \cdot \frac{\cancel{\sigma}}{\sqrt{n}}}$$

$\bar{x}$



onde $z_{\alpha/2}$ é o quantil da $N(0,1)$ que deixa $\alpha/2$ na cauda superior.

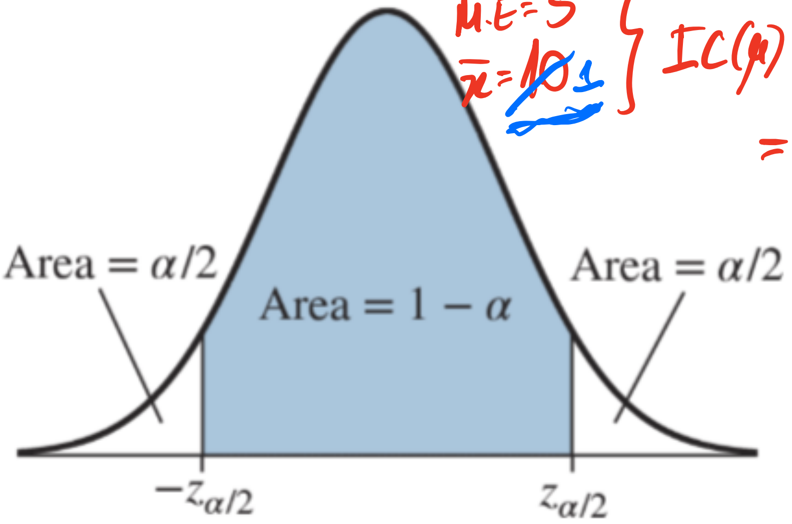
Valores Críticos da Distribuição Normal

$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = [LI, LS]$

M.E = 5

$\bar{x} = \cancel{10}\ 1$

$IC(\mu) = \cancel{10}^{\,1} \pm 5 = [\cancel{5}^{\,-4}, \cancel{15}^{\,6}]$



Confiança	α	$z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.960
99%	0.01	$\boxed{2.576}$

5

Interpretação:

- 95% dos valores de uma $N(0,1)$ estão entre -1.96 e 1.96
- Áreas nas caudas: 2.5% cada

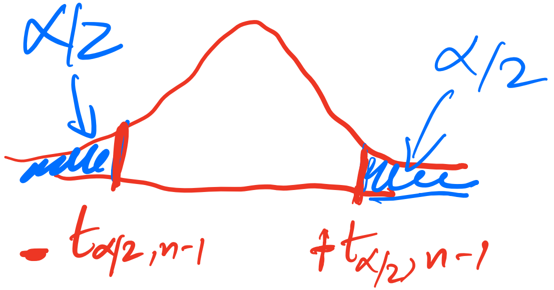
Problema: Na prática, raramente conhecemos σ

Solução: Usar o desvio padrão amostral s e a distribuição t de Student

Estatística: $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

IC de $(1-\alpha)100\%$**:**

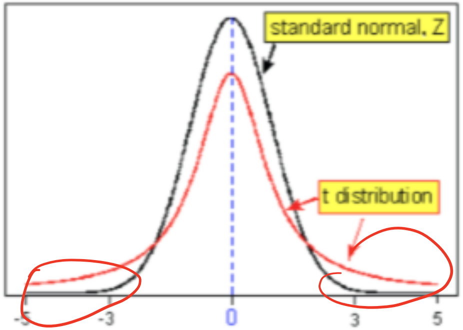
$$\boxed{IC(\mu) = \bar{X} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}}$$

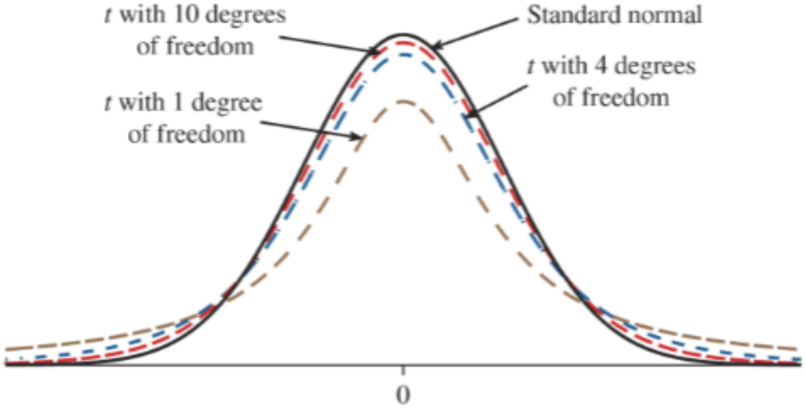


Propriedades da distribuição t**:**

- Simétrica, centrada em 0
- Caudas mais pesadas que a normal
- Converge para $N(0,1)$ quando $n \to \infty$
- Graus de liberdade: $gl = n - 1$

Distribuição Normal vs. t de Student





Quando usar cada distribuição:

- **Distribuição Normal:** quando σ é conhecido
- **Distribuição t:** quando σ é desconhecido (caso mais comum)
- Para $n > 30$, as distribuições são praticamente idênticas $\boxed{n=100}$

t, z

Interpretação:

- 95% dos valores de uma $N(0, 1)$ estão entre -1.96 e 1.96
- Áreas nas caudas: 2.5% cada

$n = 25$

Problema: Uma amostra de 25 tempos de atendimento em um órgão público teve média de 18.5 minutos e desvio padrão de 4.2 minutos. Construa um IC de 95% para o tempo médio de atendimento.

$IC(\mu)_{0.95} = \bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{\cancel{\sigma}\, s}{\sqrt{n}}$

Solução:

- Dados: $n = 25$, $\bar{x} = 18.5$, $s = 4.2$
- Como σ é desconhecido, usar distribuição t
- Graus de liberdade: $gl = 25 - 1 = 24$
- Para 95% de confiança: $t_{0.025,24} = 2.064$
- Erro padrão: $\frac{s}{\sqrt{n}} = \frac{4.2}{\sqrt{25}} = 0.84$
- Margem de erro: $2.064 \times 0.84 = 1.73$
- IC: 18.5 ± 1.73 ou $(16.77, 20.23)$

① $X \sim N$

$\bar{X} \sim N$

TCL

② $X \sim ?$ ou $X \nsim N$

→ n gde

amostra

$\bar{X} \sim N$

$\bar{X} \sim t_{n-1}$

na população

$1 - \alpha = 0.95$

$\alpha = 0.050$

$\alpha/2 = 0.025$

μ

Interpretação: Com 95% de confiança, o tempo médio de atendimento está entre 16.77 e 20.23 minutos.

16.77 20.23

Implementação em Python - IC para Média

```
import scipy.stats as stats
import numpy as np

# Dados do problema
n = 25
media_amostral = 18.5
desvio_padrao = 4.2
confianca = 0.95

# Método 1: Usando stats.t.interval() - mais simples!
erro_padrao = desvio_padrao / np.sqrt(n)
ic = stats.t.interval(confidence=confianca,
                      df=n-1,
                      loc=media_amostral,
                      scale=erro_padrao)

print(f"IC de 95%: ({ic[0]:.2f}, {ic[1]:.2f})")
# Saída: IC de 95%: (16.77, 20.23)

# Método 2: Usando stats.sem() para erro padrão
# erro_padrao = stats.sem(dados, ddof = 1)  # se tivéssemos os dados brutos
```

$$\bar{x} \pm t_{\alpha/2,\, n-1} \times \frac{s}{\sqrt{n}}$$

$$\frac{\text{np.sd(dados, df=1)}}{\text{np.sqrt(n)}} = \text{erro\_padrão}$$

IC para Proporção Populacional

Objetivo: Estimar uma proporção populacional p

Estimador: $\hat{p} = \frac{X}{n}$ onde X = número de sucessos

Distribuição aproximada: Para n grande, $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$

IC de $(1-\alpha)100\%$:

$$IC(p) = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

pontual ± M.E.

Condições de validade:

- $n\hat{p} \geq 10$ e $n(1-\hat{p}) \geq 10$
- Garante boa aproximação normal

Exemplo: IC para Proporção

Problema: Em uma pesquisa com 500 cidadãos, 320 aprovaram um serviço público. Construa um IC de 99% para a proporção de aprovação na população.

Solução:

$\alpha = 0.01$

$\alpha/2 = 0.005$

- Dados: $n = 500$, $X = 320$
- Proporção amostral: $\hat{p} = \boxed{\frac{320}{500}} = 0.64$
- Verificar condições:
  - $500 \times 0.64 = 320 \geq 10$
  - $500 \times 0.36 = 180 \geq 10$

$n\hat{p} \geq 10$

$n(1-\hat{p}) \geq 10$

- Para 99% de confiança: $z_{0.005} = 2.576$

norm.ppf(0.005...)

- Erro padrão: $\sqrt{\frac{0.64 \times 0.36}{500}} = 0.0215$
- Margem de erro: $2.576 \times 0.0215 = \boxed{0.055}$

M.E. - $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

- IC: 0.64 ± 0.055 ou $(0.585, 0.695)$

Interpretação: Com 99% de confiança, entre 58.5% e 69.5% da população aprova o serviço.

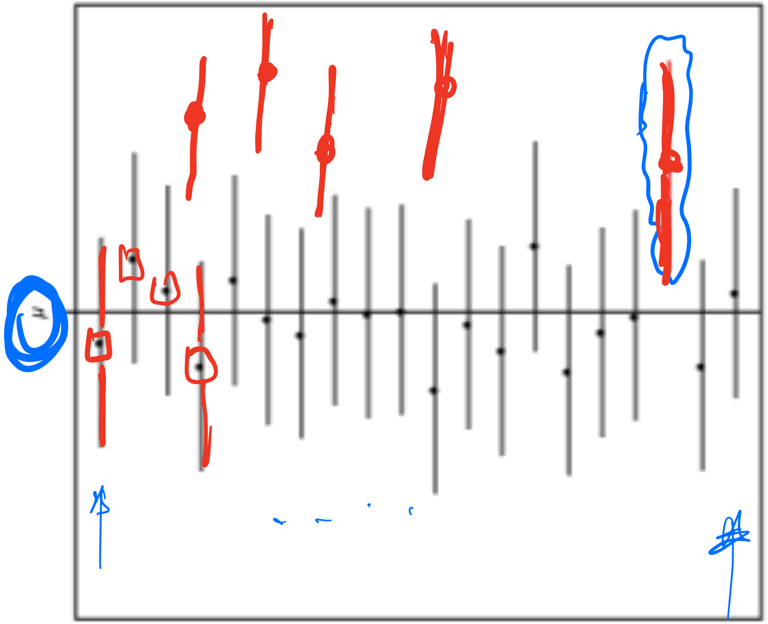
Dica Python: Use `statsmodels.stats.proportion.proportion_confint()` para cálculo direto!

Interpretação Correta do IC

0.99 0.95

O que o IC significa:

- Se repetíssemos o processo de amostragem muitas vezes
- E construíssemos um IC de 95% em cada vez
- Aproximadamente 95% desses intervalos conteriam θ



Interpretação prática: "Temos 95% de confiança que o verdadeiro valor do parâmetro está no intervalo"

Cuidado:

- NÃO diga: "Há 95% de probabilidade de θ estar no intervalo"
- O parâmetro é fixo, o intervalo é aleatório

Fatores que Afetam a Largura do IC

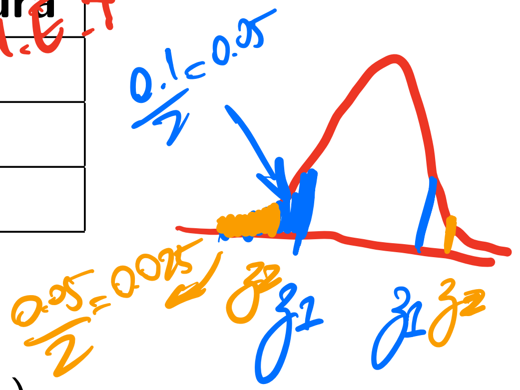
$1-\alpha = 0.9 \to \alpha = 0.1 \to \alpha = 0.05 \to 1-\alpha = 0.95$

A largura do IC é determinada pela margem de erro: $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$

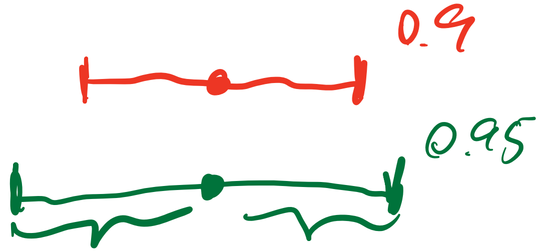
Fator	Mudança	Efeito na Largura
Tamanho da amostra (n)	Aumenta	Diminui
Nível de confiança ($1-\alpha$)	Aumenta	Aumenta
Variabilidade (σ)	Aumenta	Aumenta

M.E A



Trade-off fundamental:

- Maior confiança ⇒ Intervalo mais largo (menos preciso)
- Maior precisão ⇒ Menor confiança ou maior amostra necessária



Determinação do Tamanho da Amostra

Pergunta: Qual tamanho de amostra necessário para atingir uma margem de erro desejada?

Para estimar uma média: Dado margem de erro E e confiança $(1-\alpha)$:

$0.95 \to 1-\alpha$

$E = 0.1\,m$

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

$\bar{x} \pm \underbrace{\boxed{z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}}}_{E}$

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

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

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

Para estimar uma proporção:

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

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

Se p desconhecido, usar $\boxed{p = 0.5}$ (caso mais conservador)

Exemplo: Para estimar uma proporção com erro de 3% e 95% de confiança:

margem de erro

$0.37 \quad 0.03 \quad 0.03 \quad 0.43$

$0.4 = \hat{p}$

$[0.37, 0.43]$

$$n = \left(\frac{1.96}{0.03}\right)^2 \cdot 0.5 \cdot 0.5 = \boxed{1067}$$

Exercício 1: IC para Média com Dados Reais

Enunciado: Você tem os tempos de atendimento (em minutos) de 20 cidadãos em um órgão público. Calcule o IC de 95% para o tempo médio de atendimento.

leitura de csv.

```
import numpy as np
from scipy import stats

# Dados
tempos = [15.2, 18.5, 22.1, 16.8, 19.3,
          21.5, 17.2, 20.8, 18.9, 16.5,
          19.7, 23.2, 18.1, 17.9, 20.3,
          19.5, 18.8, 21.1, 17.6, 19.2]

# 1. Estatísticas descritivas
media = np.mean(tempos)
desvio = np.std(tempos, ddof=1)  # ddof=1 para amostra

# 2. IC de 95% usando stats.t.interval()
ic = stats.t.interval(confidence=0.95,
                      df=len(tempos)-1,
                      loc=media,
                      scale=stats.sem(tempos, ddof=1))

print(f"Média: {media:.2f} minutos")
print(f"Desvio padrão: {desvio:.2f} minutos")
print(f"IC de 95%: ({ic[0]:.2f}, {ic[1]:.2f})")
# Resultado: IC de 95%: (18.36, 20.22)
# Interpretação: Com 95% de confiança, o tempo médio
# de atendimento está entre 18.36 e 20.22 minutos.
```

σ ?

intervalo t

n-1

desvio / np.sqrt(len(tempos))

$\bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$

(18.17, 20.05)

18.17 20.05

Exercício 2: IC para Proporção

Enunciado: Em uma pesquisa de satisfação com 200 servidores públicos, 156 se declararam satisfeitos com o ambiente de trabalho. Calcule o IC de 95% para a proporção de servidores satisfeitos.

```
from statsmodels.stats.proportion import proportion_confint
from scipy import stats
import numpy as np

# Dados
n = 200   # tamanho da amostra
x = 156   # número de satisfeitos

# 1. Proporção amostral
p_hat = x / n
print(f"Proporção amostral: {p_hat:.3f}")  # 0.780

# 2. Erro padrão
erro_padrao = np.sqrt(p_hat * (1 - p_hat) / n)
print(f"Erro padrão: {erro_padrao:.4f}")  # 0.0293

# 3. IC de 95% usando aproximação normal
z_critico = stats.norm.ppf(0.975)  # 1.96 para 95%
margem_erro = z_critico * erro_padrao

ic_inferior = p_hat - margem_erro
ic_superior = p_hat + margem_erro

print(f"IC de 95%: ({ic_inferior:.3f}, {ic_superior:.3f})")
# Resultado: IC de 95%: (0.723, 0.837)
```

$1-\alpha=0.95$

$\alpha=0.05$

$\frac{\alpha}{2}=0.025$

na

mulher

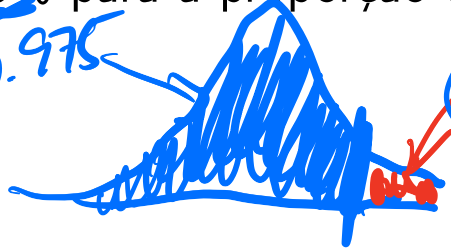
$z_{\alpha/2}=$

$z_{\alpha/2}$

z crítico

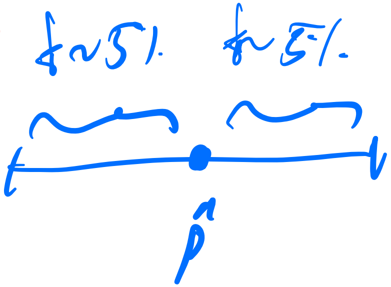
0.975

0.025



z

$\hat{p}(1-\hat{p})$



$1-\alpha = 0.95$

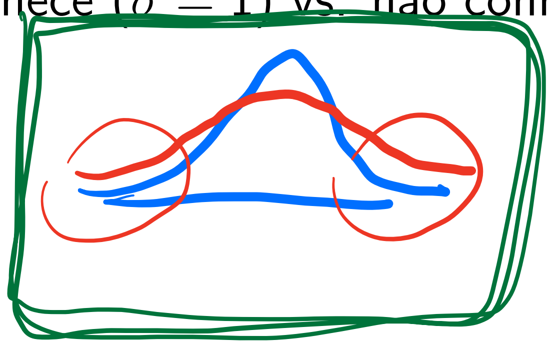
$\alpha = 0.05$

Usando statsmodels (mais fácil e robusto):

```
from statsmodels.stats.proportion import proportion_confint

# Mesmos dados do exercício anterior
n = 200   # tamanho da amostra
x = 156   # número de satisfeitos

# IC de 95% com apenas uma linha!
ic = proportion_confint(x, n, alpha=0.05, method='normal')

print(f"IC de 95%: ({ic[0]:.3f}, {ic[1]:.3f})")
# Resultado: IC de 95%: (0.723, 0.837)

# Outros métodos disponíveis:
# method='wilson' - melhor para amostras pequenas
# method='beta' - método exato
# method='agresti_coull' - ajuste para amostras pequenas
```

Enunciado: Compare os ICs quando você conhece ($\sigma = 1$) vs. não conhece o desvio padrão populacional.

$s < \sigma$

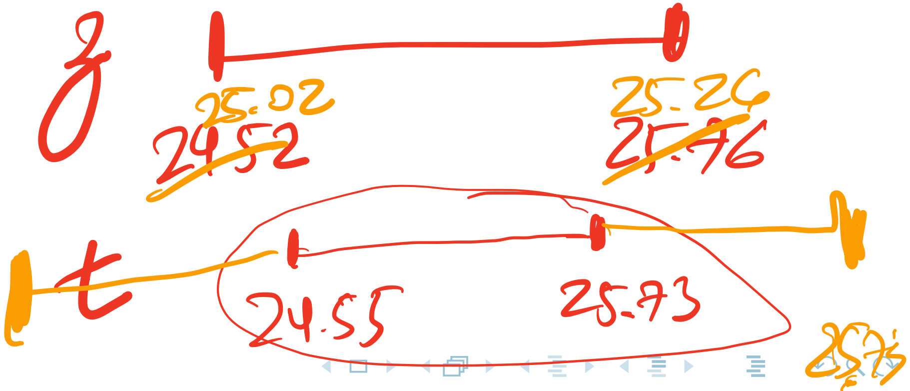


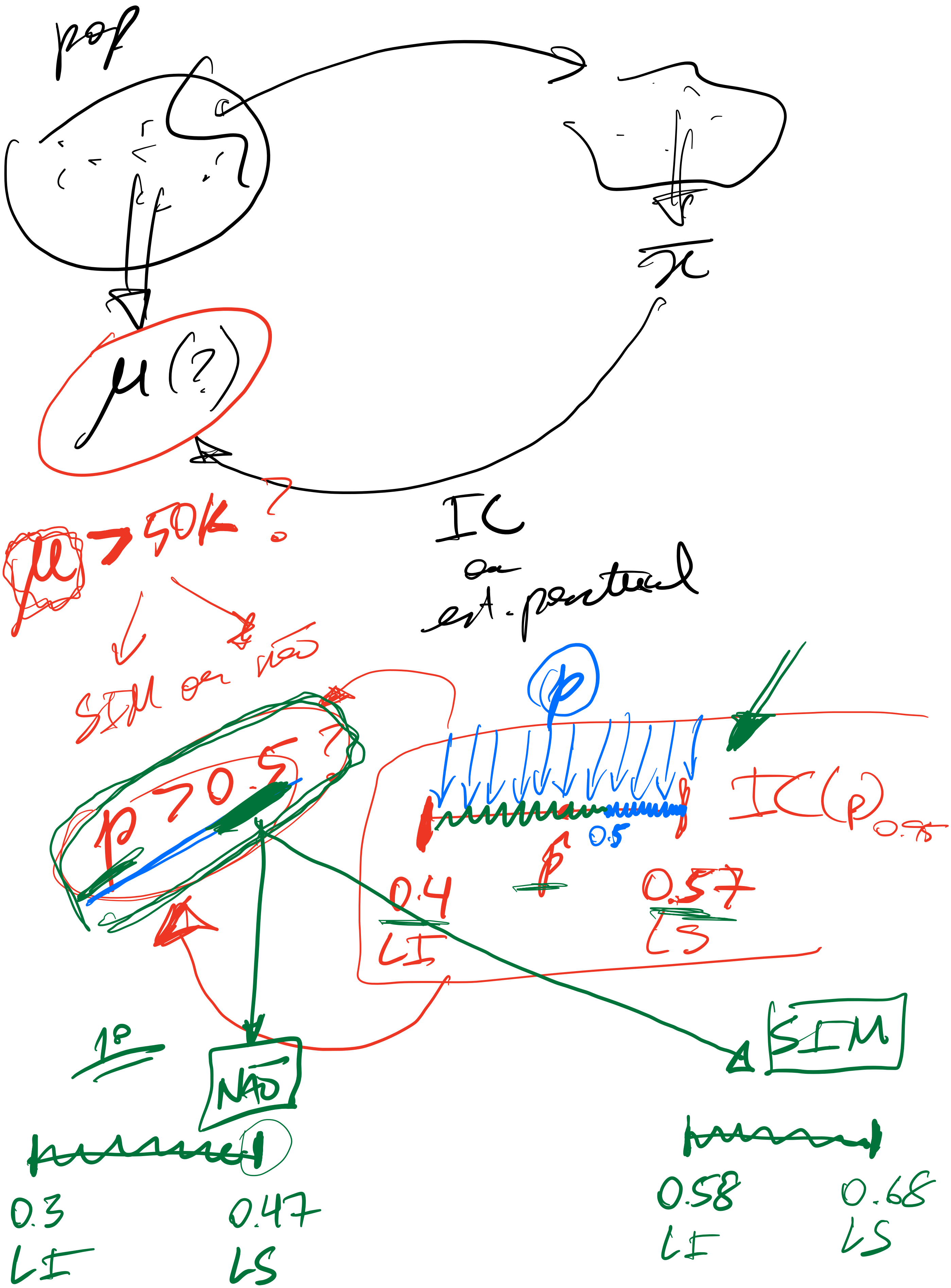
```
from scipy import stats
import numpy as np

# Dados de uma amostra
dados = [23.5, 25.1, 24.8, 26.2, 25.5,
         24.3, 25.8, 24.9, 25.3, 26.0]
media = np.mean(dados)
n = len(dados)

# 1. IC com sigma conhecido (usar distribuição normal)
sigma_conhecido = 1.0
ic_normal = stats.norm.interval(0.95, loc=media, scale=sigma_conhecido/np.sqrt(n))

# 2. IC com  desconhecido (usar distribuição t)
ic_t = stats.t.interval(0.95, df=n-1, loc=media, scale=stats.sem(dados))

print(f"Média amostral: {media:.2f}")
print(f"IC ( conhecido):   ({ic_normal[0]:.2f}, {ic_normal[1]:.2f})")
print(f"IC ( desconhecido): ({ic_t[0]:.2f}, {ic_t[1]:.2f})")

print(f"\nLargura IC normal: {ic_normal[1] - ic_normal[0]:.3f}")
print(f"Largura IC t: {ic_t[1] - ic_t[0]:.3f}")
# O IC usando t é mais largo (mais conservador)
```

0.2

$\sigma/\sqrt{n}$

$s/\sqrt{n}$




pop
μ (?)
x̄
IC
ou
est. pontual
μ > 50K ?
SIM ou NÃO
p > 0.5 ?
p
IC(p)0.95
0.5
p̂
0.4
LI
0.57
LS
1º
NÃO
SIM
0.3
LI
0.47
LS
0.58
LI
0.68
LS

Resumo da Aula

24.55

Conceitos principais:

- **Limitação da estimação pontual:** não quantifica incerteza
- **Intervalo de Confiança:** fornece faixa de valores plausíveis
- **Componentes do IC:** estimativa $\pm$ margem de erro

Casos importantes:

- **IC para média (σ conhecido):** usar distribuição normal
- **IC para média (σ desconhecido):** usar distribuição t
- **IC para proporção:** usar aproximação normal quando n grande

Interpretação:

- IC quantifica a incerteza amostral
- Trade-off entre confiança e precisão
- Tamanho da amostra afeta diretamente a precisão

Próxima aula: Testes de Hipóteses