Aula 03. Estimação Pontual e Intervalos de Confiança

Estatística Inferencial

MBA CDIA

ENAP - Escola Nacional de Administração Pública

2025

Agenda

1. **Estimação Pontual**
   - Conceitos básicos
   - Propriedades dos estimadores
   - Viés, variância e erro quadrático médio
2. **Função de Verossimilhança**
   - Método de máxima verossimilhança
   - Estimadores para média e variância
   - Outros métodos de estimação
3. **Intervalos de Confiança**
   - IC para média populacional
   - IC para proporção populacional
4. **Exercícios práticos em Python**

Estimação Pontual

Definição

Um **estimador pontual** é uma estatística (função da amostra) usada para estimar um parâmetro populacional desconhecido.

Notação:

- θ = parâmetro populacional (desconhecido)
- $\hat{\theta}$ = estimador (estatística calculada da amostra)
- $\hat{\theta}_{\text{obs}}$ = estimativa (valor numérico obtido)

Exemplos comuns:

- $\bar{X}$ estima μ (média populacional)
- s^2 estima σ^2 (variância populacional)
- $\hat{p}$ estima p (proporção populacional)

Um bom estimador deve ter as seguintes propriedades:

1. **Não-viesado (ou não-tendencioso):**

$$E[\hat{\theta}] = \theta$$

   O valor esperado do estimador é igual ao parâmetro.
2. **Eficiente:** Entre todos os estimadores não-viesados, possui a menor variância.
3. **Consistente:** À medida que $n \to \infty$, $\hat{\theta} \to \theta$ em probabilidade.
4. **Suficiente:** Utiliza toda a informação relevante contida na amostra.

Viés, Variância e Erro Quadrático Médio

- **Viés** de um estimador:

$$\text{Viés}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

- **Variância** de um estimador:

$$\text{Var}(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$$

Exemplos de Estimadores

1. **Média amostral** $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
   - $E[\bar{X}] = \mu$ (não-viesado)
   - $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ (diminui com n)
2. **Variância amostral** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
   - $E[s^2] = \sigma^2$ (não-viesado)
   - Divisor $(n-1)$ corrige o viés
3. **Proporção amostral** $\hat{p} = \frac{X}{n}$ (X = número de sucessos)
   - $E[\hat{p}] = p$ (não-viesado)
   - $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$

Definição

A **função de verossimilhança** $L(\theta|x_1, x_2, ..., x_n)$ é a probabilidade (ou densidade) dos dados observados, vista como função do parâmetro θ.

Importante:

- $L(\theta|dados) = P(dados|\theta)$ **NÃO** é $P(\theta|dados)$
- É a probabilidade dos **dados** dado o **parâmetro**
- Mas usamos como função de θ para encontrar o melhor valor

Para uma amostra aleatória: $L(\theta|x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$

Ideia central: Valores de θ que tornam os dados observados mais prováveis são candidatos melhores para estimar o parâmetro verdadeiro.

Método de Máxima Verossimilhança (MLE)

Estimador de Máxima Verossimilhança

O EMV de θ é o valor $\hat{\theta}_{MLE}$ que maximiza a função de verossimilhança:
$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta | x_1, x_2, ..., x_n)$

Observações importantes:

- O método se aplica a **qualquer distribuição** de probabilidade
- O EMV depende da distribuição assumida para os dados
- Para diferentes distribuições, obtemos diferentes estimadores

Procedimento geral:

1. Assumir uma distribuição para os dados
2. Escrever a função de verossimilhança $L(\theta)$
3. Maximizar $L(\theta)$ (geralmente via log-verossimilhança)
4. Resolver $\frac{d\ell}{d\theta} = 0$ onde $\ell(\theta) = \ln L(\theta)$

Exemplo: EMV para Distribuição Normal

Assumindo $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$:

1. EMV da média (σ^2 **conhecido):** $\hat{\mu}_{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

2. EMV da variância (μ **conhecido):** $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

3. EMV conjunto de μ **e** σ^2**:** $\hat{\mu}_{MLE} = \bar{X}, \quad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Observações:

- Para a normal, EMV da média = média amostral
- O EMV da variância é viesado: $E[\hat{\sigma}^2_{MLE}] = \frac{n-1}{n}\sigma^2$
- Para outras distribuições, os EMVs podem ser diferentes!

EMV para Outras Distribuições

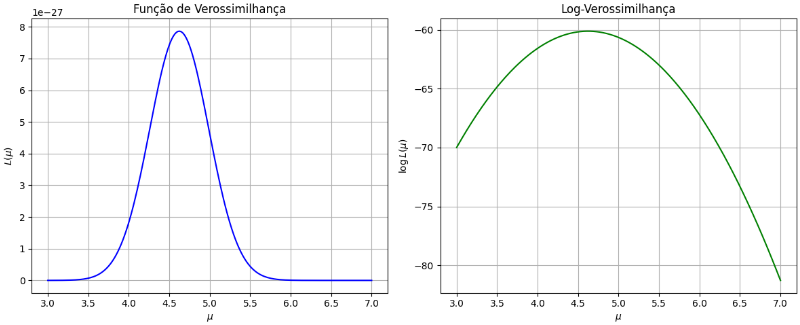
O EMV depende da distribuição assumida:

Distribuição	**Parâmetro**	**EMV**
Normal(μ, σ^2)	μ	$\bar{X}$
Exponencial(λ)	λ	$1/\bar{X}$
Poisson(λ)	λ	$\bar{X}$
Bernoulli(p)	p	$\bar{X}$
Uniforme($0, \theta$)	θ	$\max\{X_1, ..., X_n\}$
Gama(α, β)	α, β	Não tem forma fechada

Pontos importantes:

- O EMV da média nem sempre é $\bar{X}$ (ex: Uniforme)
- Alguns EMVs não têm solução analítica (ex: Gama)
- EMVs podem ser viesados (ex: variância da Normal)

Ilustração: Função de Verossimilhança



- Curva mostra $L(\mu)$ para dados de uma $N(\mu, 1)$
- Pico indica o valor de μ mais compatível com os dados
- EMV corresponde ao máximo da função

Outros Métodos de Estimação

1. **Método dos Momentos (MM):**
   - Iguala momentos amostrais aos momentos populacionais
   - Exemplo: $\bar{X} = \mu$ e $s^2 = \sigma^2$
   - Mais simples mas geralmente menos eficiente que MLE
2. **Estimadores Bayesianos:**
   - Incorporam informação prévia sobre θ
   - Posteriori $\propto$ Verossimilhança $\times$ Priori
3. **Mínimos Quadrados:**
   - Minimiza soma dos quadrados dos erros
   - Muito usado em regressão

Nota: Em muitos casos, diferentes métodos levam aos mesmos estimadores!

Exercício 1: Comparando Estimadores de Variância

Implemente uma simulação para comparar o viés dos estimadores:

- $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum(X_i - \bar{X})^2$
- $s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$

Passos:

1. Gere 1000 amostras de tamanho $n = 10$ de $N(0, 1)$
2. Calcule ambos estimadores para cada amostra
3. Compare as médias dos estimadores com $\sigma^2 = 1$
4. Visualize as distribuições dos estimadores

```
import numpy as np
import matplotlib.pyplot as plt
```

Gabarito - Exercício 1 - Parte 1

```
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(42)
n = 10
n_simulacoes = 1000
sigma2_verdadeiro = 1
# Armazenar resultados
mle_estimates = []
unbiased_estimates = []

for _ in range(n_simulacoes):
    # Gerar amostra
    amostra = np.random.normal(0, 1, n)
    # Calcular estimadores
    media = np.mean(amostra)
    mle = np.sum((amostra - media)**2) / n
    unbiased = np.sum((amostra - media)**2) / (n - 1)

    mle_estimates.append(mle)
    unbiased_estimates.append(unbiased)

# Calcular viés
vies_mle = np.mean(mle_estimates) - sigma2_verdadeiro
vies_unbiased = np.mean(unbiased_estimates) - sigma2_verdadeiro

print(f"Viés do EMV: {vies_mle:.4f}")
print(f"Viés do estimador não-viesado: {vies_unbiased:.4f}")
print(f"Média do EMV: {np.mean(mle_estimates):.4f}")
print(f"Média do não-viesado: {np.mean(unbiased_estimates):.4f}")
```

Gabarito - Exercício 1 $_parte2$

```
# Visualização
plt.figure(figsize=(10, 6))
plt.hist(mle_estimates, bins=50, alpha=0.5, label='EMV', density=True)
plt.hist(unbiased_estimates, bins=50, alpha=0.5, label='Não-viesado', density=True)
plt.axvline(sigma2_verdadeiro, color='red', linestyle='--', label='Valor verdadeiro')
plt.xlabel('Estimativa')
plt.ylabel('Densidade')
plt.title('Distribuição dos Estimadores de Variância')
plt.legend()
plt.show()
```

Exercício 2: Visualizando a Função de Verossimilhança

Enunciado: Implemente e visualize a função de verossimilhança para estimar a média de uma distribuição normal com variância conhecida.

Passos:

1. Gere 20 observações de $N(\mu = 5, \sigma^2 = 4)$
2. Calcule a função de verossimilhança para $\mu \in [2, 8]$
3. Encontre o EMV numericamente
4. Plote a função de verossimilhança e log-verossimilhança

Dica: Use `scipy.stats.norm` para a densidade normal.

Gabarito - Exercício 2

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Gerar dados
np.random.seed(42)
mu_verdadeiro = 5
sigma = 2
n = 20
dados = np.random.normal(mu_verdadeiro, sigma, n)

# Função de verossimilhança
def verossimilhanca(mu, dados, sigma):
    return np.prod(norm.pdf(dados, mu, sigma))

def log_verossimilhanca(mu, dados, sigma):
    return np.sum(norm.logpdf(dados, mu, sigma))

# Calcular para vários valores de mu
mu_valores = np.linspace(2, 8, 200)
L_valores = [verossimilhanca(mu, dados, sigma) for mu in mu_valores]
logL_valores = [log_verossimilhanca(mu, dados, sigma) for mu in mu_valores]

# EMV (média amostral)
mu_emv = np.mean(dados)

# Visualização
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 5))

ax1.plot(mu_valores, L_valores)
```

Intervalos de Confiança - Introdução

- Vimos que estimadores pontuais fornecem uma única estimativa para o parâmetro.
- É improvável que uma estimativa pontual $\hat{\theta}$ seja exatamente igual ao parâmetro populacional θ.
- Muitas vezes é mais informativo fornecer um **intervalo de valores** em vez de uma única estimativa pontual.

Exemplo: Uma pesquisa com 120 servidores públicos resultou em satisfação média de 7.5.

- Estimativa pontual: $\bar{x} = 7.5$
- Intervalo de confiança: $7.3 < \mu < 7.7$ (com 95% de confiança)

Intervalo de Confiança (IC)

Definição

Um **intervalo de confiança** fornece um intervalo de valores que contém o parâmetro populacional com um certo nível de confiança.

O intervalo de confiança tem a forma:

$$\text{Estimativa pontual} \pm \text{Margem de erro}$$

ou equivalentemente:

$$(\text{Estimativa} - \text{Margem de erro}, \text{Estimativa} + \text{Margem de erro})$$

Componentes:

- **Nível de confiança:** $100(1-\alpha)\%$ (ex: 90%, 95%, 99%)
- **Margem de erro:** Depende da variabilidade e do tamanho da amostra

IC para Média Populacional - σ conhecido

Quando a variância populacional σ^2 é conhecida:

Condições:

- Amostra grande ($n > 30$), ou
- População normalmente distribuída

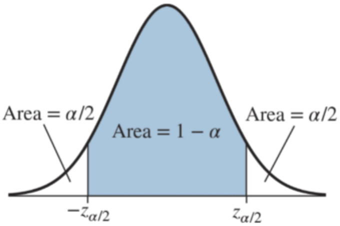
Intervalo de Confiança $100(1-\alpha)\%$**:**

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

onde:

- $\bar{x}$ = média amostral
- $z_{\alpha/2}$ = valor crítico da distribuição normal padrão
- $\frac{\sigma}{\sqrt{n}}$ = erro padrão da média

Valores Críticos da Distribuição Normal



Nível de Confiança	**Valor Crítico** $z_{\alpha/2}$
90%	1.645
95%	1.960
99%	2.576

IC para Média Populacional - σ desconhecido

Quando a variância populacional σ^2 é desconhecida (caso mais comum):

Condições:

- Amostra grande ($n > 30$), ou
- População normalmente distribuída

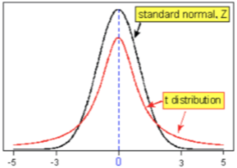
Intervalo de Confiança $100(1-\alpha)\%$**:** $\bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$

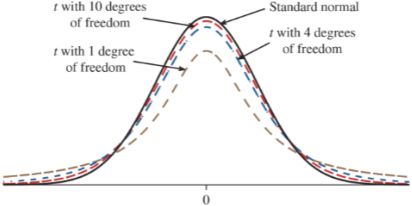
onde:

- s = desvio padrão amostral
- $t_{\alpha/2,n-1}$ = valor crítico da distribuição t de Student
- $n-1$ = graus de liberdade

Distribuição t de Student

- Forma similar à normal padrão (simétrica, centrada em 0)
- Caudas mais pesadas que a normal
- Depende dos graus de liberdade (g.l. $= n - 1$)
- Quando $n \to \infty$, converge para a normal padrão





Exemplo: IC para Média

Problema: Uma amostra de 37 beneficiários de uma política pública teve tempo médio de execução dos recursos financeiros de 45 semanas com desvio padrão de 12 semanas. Construa um IC de 95% para o tempo médio de execução orçamentária para todos os beneficiados pelo programa em questão.

Solução:

- $n = 37$, $\bar{x} = 45$, $s = 12$
- Como $n > 30$ e σ desconhecido, usar distribuição t
- g.l. $= n - 1 = 36$
- Para 95% de confiança: $t_{0.025,36} = 2.042$
- Margem de erro $= 2.042 \times \frac{12}{\sqrt{37}} = 4.03$
- IC: 45 ± 4.03 ou $(40.97, 49.03)$

Interpretação: Temos 95% de confiança que o tempo médio de execução de todos os beneficiados está entre 40.97 e 49.03 minutos.

Resolução em Python

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

# Dados fornecidos
n = 37
media_amostral = 45
desvio_padrao = 12
confianca = 0.95

# Calcula o valor crítico da distribuicao t
t_critico = stats.t.ppf(1 - (1 - confianca)/2, df=n-1)

# Calcula a margem de erro
margem_erro = t_critico * (desvio_padrao / np.sqrt(n))

# Calcula o intervalo de confianca
limite_inferior = media_amostral - margem_erro
limite_superior = media_amostral + margem_erro

print(f"IC de 95%: ({limite_inferior:.2f}, {limite_superior:.2f})")
```

IC para Proporção Populacional

Para estimar uma proporção populacional p:

Condições:

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

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

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

onde:

- $\hat{p} = \frac{x}{n}$ = proporção amostral
- x = número de sucessos na amostra
- n = tamanho da amostra

Exemplo: IC para Proporção

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

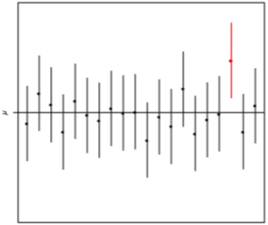
Solução:

- $n = 500$, $x = 320$, $\hat{p} = \frac{320}{500} = 0.64$
- Verificar condições: $500(0.64) = 320 \geq 10$
- $500(0.36) = 180 \geq 10$
- Para 99%: $z_{0.005} = 2.576$
- Erro padrão $= \sqrt{\frac{0.64 \times 0.36}{500}} = 0.0215$
- Margem de erro $= 2.576 \times 0.0215 = 0.055$
- 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.

Interpretação Correta do IC

Interpretação teórica: Se repetíssemos o processo de amostragem muitas vezes e construíssemos um IC de 95% em cada vez, aproximadamente 95% desses intervalos conteriam o verdadeiro parâmetro populacional.



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

Cuidado: Não diga que há 95% de probabilidade do parâmetro estar no intervalo!

Para estimar uma média: Para obter margem de erro E com confiança $100(1-\alpha)\%$:
$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$

Para estimar uma proporção: $n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot p(1-p)$

Se p é desconhecido, usar $p = 0.5$ (caso mais conservador).

Exemplo: Para estimar uma proporção com margem de erro de 3% e 95% de confiança:
$n = \left(\frac{1.96}{0.03}\right)^2 \cdot 0.5 \cdot 0.5 = 1067$

Resumo

Estimação Pontual:

- Estimadores devem ser não-viesados, eficientes e consistentes
- Trade-off entre viés e variância

Máxima Verossimilhança:

- Método geral para obter estimadores
- Propriedades assintóticas desejáveis

Intervalos de Confiança:

- Fornecem faixa de valores plausíveis para o parâmetro
- Dependem do nível de confiança e tamanho da amostra
- Usar distribuição t quando σ é desconhecido

Próxima aula: Testes de Hipóteses