Estimação intervalar

Estimação intervalar

Fernando de Pol Mayer

Laboratório de Estatística e Geoinformação (LEG)
Departamento de Estatística (DEST)
Universidade Federal do Paraná (UFPR)

Sumário

Plano de aula

Estimação

Existem dois tipos de estimativas que podemos obter a partir de uma **amostra aleatória**:

Estimativa pontual

Fornecem como estimativa um único valor numérico para o parâmetro de interesse

Estimativa intervalar

Fornece um intervalo de valores "plausíveis" para o parâmetro de interesse

Estimação

Por serem **variáveis aleatórias**, os estimadores pontuais possuem uma distribuição de probabilidade (distribuições amostrais)

Com isso, podemos apresentar uma estimativa mais informativa para o parâmetro de interesse, que inclua uma medida de **precisão** do valor obtido $\rightarrow$ **estimativa intervalar** ou **intervalo de confiança**

Os **intervalos de confiança** são obtidos a partir da **distribuição amostral** de seus estimadores

Plano de aula

Intervalos de confiança para a média: σ conhecido

Suposições necessárias

- A amostra é uma **amostra aleatória simples**. (Todas as amostras de mesmo tamanho tem a mesma probabilidade de serem selecionadas)
- O valor do desvio-padrão populacionalm σ, é conhecido
- Uma ou ambas das seguintes condições são satisfeitas:
  - A população é normalmente distribuída
  - A amostra possui $n > 30$

Intervalos de confiança para a média: σ conhecido

Quando coletamos uma **amostra aleatória** e calculamos uma média, sabemos que o valor da média possui um desvio natural, em relação ao verdadeiro valor da média populacional (**erro amostral**), ou seja

$$e = \bar{X} - \mu \quad \Rightarrow \quad \bar{X} = \mu + e$$

Sabemos que a **distribuição amostral da média** é uma distribuição normal, com média μ e variância σ^2/n,

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

Intervalos de confiança para a média: σ conhecido

Usando a transformação

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

podemos determinar o **erro máximo provável** que assumimos para a média amostral que estamos calculando.

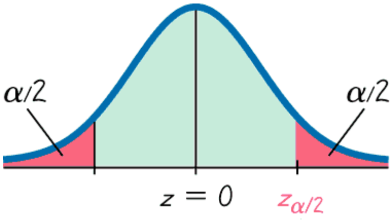
O **erro máximo provável** ou **margem de erro** da média é definido por

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

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

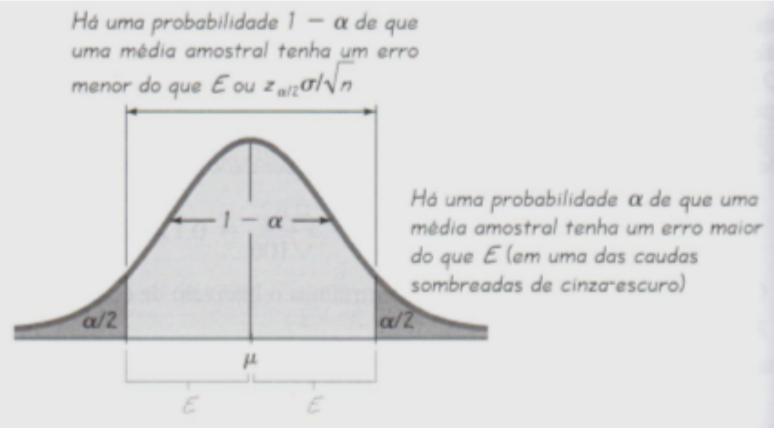
Valores críticos

O valor crítico $z_{\alpha/2}$ é o valor de Z que separa uma área de $\alpha/2$ da cauda da distribuição normal padrão



Como estamos interessados nos valores **mais prováveis** da média, então nosso interesse está no centro da distribuição Z, que concentra uma área $\gamma = 1 - \alpha$, que determina o **nível de confiança** do intervalo

Valores críticos



Nível de confiança

A área $\gamma = 1 - \alpha$ determina o **nível de confiança** associado ao intervalo de confiança que estamos construindo

O valor de α é o complemento do nível de confiança

Exemplo:

- Para um nível de confiança de 0,95 (ou 95%), $\alpha = 0,05$
- Para um nível de confiança de 0,99 (ou 99%), $\alpha = 0,01$

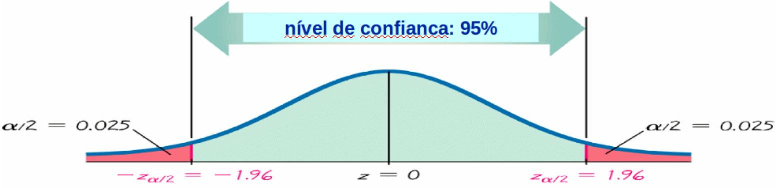
Importante!

O **nível de confiança** é a probabilidade $1 - \alpha$, que é a proporção de vezes que o intervalo de confiança realmente contém o parâmetro populacional, supondo que a amostragem pudesse ser repetida um grande número de vezes

Encontrando valores críticos

Com a definição do **nível de confiança**, sabemos então o valor de α, e devemos encontrar o valor de $z_{\alpha/2}$. Usando como exemplo $\gamma = 0,95 = 1 - 0,05 \Rightarrow \alpha = 0,05$:

- Temos que $\alpha/2 = 0,025$ é a área em cada cauda
- Na tabela da distribuição normal padrão, procure a **área**, no corpo da tabela, que corresponde a $0,5 - 0,025 = 0,475$
- O valor de $z_{\alpha/2}$ será determinado pelos valores correspondentes nas margens da tabela. Nesse caso, $z_{\alpha/2} = 1,96$ é o valor crítico procurado.



Encontrando valores críticos

Encontre os valores críticos para os níveis de confiança

- $\gamma = 0,90 \Rightarrow \alpha = 0,10$
- $\gamma = 0,99 \Rightarrow \alpha = 0,01$

Tabela: Níveis de confiança e valores críticos mais comuns

Nível de confiança γ	α	Valor crítico $z_{\alpha/2}$
0,90	0,10	1,645
0,95	0,05	1,96
0,99	0,01	2,575

Intervalos de confiança para a média: σ conhecido

Com estas definições, podemos construir um **intervalo de confiança** para uma **estimativa amostral da média com σ conhecido** através de

$$\bar{x} - e < \mu < \bar{x} + e$$

com

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

Outras notações

$$\bar{x} \pm e$$

$$[\bar{x} - e; \bar{x} + e]$$

Porque podemos fazer isso?

$$\Pr[-z_{\alpha/2} < Z < z_{\alpha/2}] = \gamma$$
$$\Pr[-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}] = \gamma$$

Isolando μ nessa inequação,

$$\Pr[\bar{x} - z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)] = \gamma$$
$$\Pr[\bar{x} - e < \mu < \bar{x} + e] = \gamma$$

Procedimentos gerais para a construção de intervalos de confiança

1. Verifique se as suposições necessárias estão satisfeitas
   - Temos uma AAS
   - σ é conhecido
   - A população tem distribuição normal ou $n > 30$
2. Determine o nível de confiança γ, e identifique α
3. Com o valor de α definido, encontre o valor crítico de $z_{\alpha/2}$
4. Calcule a margem de erro $e = z_{\alpha/2} \cdot (\sigma/\sqrt{n})$
5. Coloque em um dos formatos gerais para intervalo de confiança

$$\bar{x} - e < \mu < \bar{x} + e$$
$$\bar{x} \pm e$$
$$[\bar{x} - e; \bar{x} + e]$$

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de $52 < \mu < 58$

Interpretação 1

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 52 e 58

Interpretação 2

Temos 95% de confiança de que o intervalo entre 52 e 58 realmente contém a verdadeira média populacional μ

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de $52 < \mu < 58$

Interpretação 1 — ERRADA

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 52 e 58

Interpretação 2 — CERTA

Temos 95% de confiança de que o intervalo entre 52 e 58 realmente contém a verdadeira média populacional μ

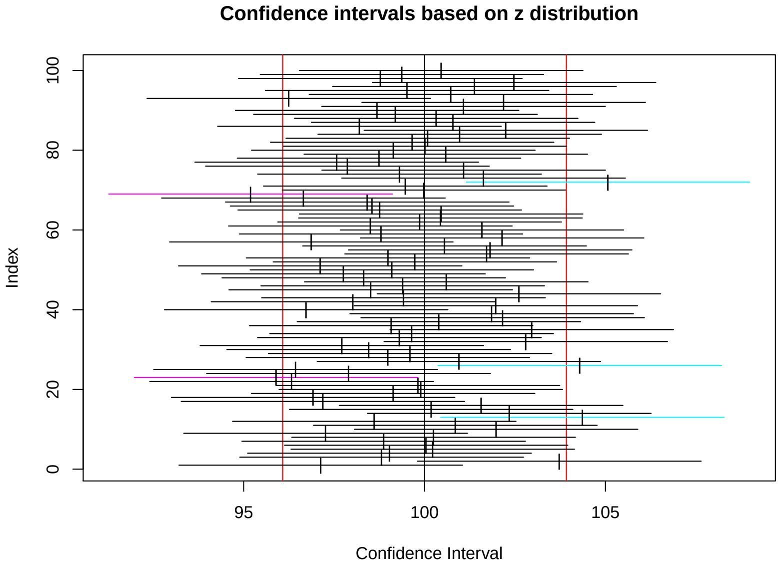
Intervalos de confiança para a média: σ conhecido

Como o intervalo de confiança é calculado a partir de uma **amostra aleatória**, este intervalo **também é aleatório**!

Isso significa que para cada amostra aleatória que tivermos, um intervalo **diferente** será calculado.

Como o valor de μ é fixo, é o intervalo que deve conter o valor de μ, e não o contrário.

Isso significa que se pudessemos obter 100 amostras diferentes, e calcularmos um intervalo de confiança de 95% para cada uma das 100 amostras, esperariamos que 5 destes intervalos **não** contenham o verdadeiro valor da média populacional μ.


Confidence intervals based on z distribution
Index
0
20
40
60
80
100
95
100
105
Confidence Interval

Exemplo: Uma empresa de computadores deseja estimar o tempo médio de horas semanais que as pessoas utilizam o computador. Uma amostra aleatória de 25 pessoas apresentou um tempo médio de uso de 22,4 horas. Com base em estudos anteriores, a empresa assume que $\sigma = 5,2$ horas, e que os tempos são normalmente distribuídos.

a) Verifique as suposições necessárias para o cálculo de um intervalo de confiança
b) Para um nível de confiança de 95%, encontre o valor crítico $z_{\alpha/2}$
c) Calcule o erro máximo provável
d) Construa o intervalo de confiança
e) Escreva a interpretação do resultado

```
Resp.: [20.362; 24.438]
```

Intervalos de confiança para a média: σ conhecido

A **amplitude** de um intervalo de confiança é dada pela diferença entre o limite superior e inferior, ou seja,

$$\text{AMP}_{IC} = \left[\bar{x} + z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)\right] - \left[\bar{x} - z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

Note que, claramente, um intervalo de confiança depende conjuntamente de três componentes:

- nível de confiança γ, expresso pelo valor crítico $z_{\alpha/2}$
- desvio-padrão populacional σ
- tamanho da amostra n

Intervalos de confiança para a média: σ conhecido

$z_{\alpha/2}$ Cada vez que aumentamos a confiança γ, o valor de $z_{\alpha/2}$ fica maior, e consequentemente a amplitude do intervalo aumenta.

> Intervalos maiores tem maior possibilidade de "captura" do verdadeiro valor de μ

σ Um grande desvio-padrão indica a possibilidade de um considerável distanciamento dos valores amostrais em relação à média populacional

> Ainda deve-se considerar que tanto $\bar{x}$ quanto σ podem ser influenciados pela presença de valores extremos

n Quanto maior for o tamanho da amostra, maior será a quantidade de informação disponível. Com isso, valores maiores de n produzem intervalos mais informativos

> Para valores fixos de γ e σ, valores maiores de n produzem intervalos menores

Exemplo: Seja $X \sim \mathrm{N}(\mu, 36)$

a) Para uma amostra de tamanho 50, obtivemos média amostral 18,5. Construa intervalos de confiança de

   (i) 90% (ii) 95% (iii) e 99%

b) Calcule as amplitudes dos intervalos acima e explique a diferença.

c) Para um nível de confiança de 95%, construa intervalos de confiança (admita a mesma média amostral 18,5) supondo tamanhos de amostra

   (i) $n = 15$ (ii) $n = 100$

d) Calcule as amplitudes dos intervalos acima e explique a diferença.

Respostas do exercício anterior:

```
## a)

       (i)   (ii)  (iii)
LI 17.104 16.837 16.314
LS 19.896 20.163 20.686

## b)

  (i)  (ii) (iii)
2.792 3.326 4.372

## c)

       (i)   (ii)
LI 15.464 17.324
LS 21.536 19.676

## d)

  (i)  (ii)
6.072 2.352
```

Plano de aula

Determinação do tamanho amostral

Nosso objetivo é coletar dados para estimar a **média populacional** μ

A questão é:

Quantos elementos (itens, objetos, pessoas, ...) devemos amostrar?

Já vimos que, de maneira (bem) geral, $n > 30$ é um tamanho de amostra mínimo para a maioria dos casos.

Será que podemos ter uma estimativa melhor de quantos elementos devem ser amostrados para estimarmos a média populacional com uma precisão conhecida?

A partir da equação do **erro máximo provável**

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

podemos isolar n e chegar na seguinte equação para a determinação do tamanho amostral

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

Determinação do tamanho amostral

Note que, em

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

- O tamanho amostral n **não** depende do tamanho populacional N
- O tamanho amostral depende
  - do nível de confiança desejado (expresso pelo valor crítico $z_{\alpha/2}$)
  - do erro máximo *desejado*
  - do desvio-padrão σ (embora veremos que não é estritamente necessário)
- Como o tamanho amostral precisa ser um número inteiro, arredondamos sempre o valor para o **maior** número inteiro mais próximo

Determinação do tamanho amostral

Exemplo: Seja $X \sim \mathrm{N}(\mu, 36)$

a) Calcule o tamanho da amostra, para que com 95% de probabilidade, a média amostral não difira da média populacional por mais de

   (i) 0,5 unidades   (ii) 2 unidades

b) Qual o impacto do erro máximo assumido para o tamanho da amostra?

c) Calcule o tamanho da amostra, para que a diferença da média amostral para a média populacional (em valor absoluto) seja menor ou igual a 2 unidades, com níveis de confiança de

   (i) 90%   (ii) 95%

d) Compare as estimativas do item anterior e analise o impacto do nível de confiança para a determinação do tamanho amostral.

Plano de aula

Estimação intervalar

Introdução

Intervalos de confiança para a média: σ conhecido

Determinação do tamanho amostral

Referências

Referências

- Bussab, WO; Morettin, PA. **Estatística básica**. São Paulo: Saraiva, 2006. [Cap. 11]
- Magalhães, MN; Lima, ACP. **Noções de Probabilidade e Estatística**. São Paulo: EDUSP, 2008. [Cap. 7]
- Montgomery, DC; Runger, GC. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC Editora, 2012. [Cap. 8]