Estimação: (B) Estimação por intervalo

Wagner H. Bonat
Fernando P. Mayer
Elias T. Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

08/05/2018

Sumário

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.

Sumário

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 $n > 30$

Erro amostral

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,

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

Margem de erro

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

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

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

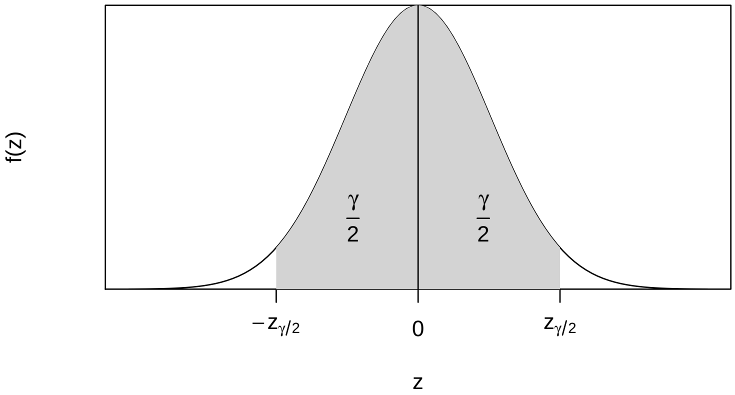
Intervalo de confiança

Fixando um valor γ tal que $0 < \gamma < 1$, podemos encontrar um valor $z_{\gamma/2}$ tal que:

$$
\begin{aligned}
P[|Z| < z_{\gamma/2}] &= \gamma \\
P[-z_{\gamma/2} < Z < z_{\gamma/2}] &= \gamma \\
P[-z_{\gamma/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\gamma/2}] &= \gamma \\
P[\bar{x} - z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)] &= \gamma \\
P[\bar{x} - e < \mu < \bar{x} + e] &= \gamma
\end{aligned}
$$

Intervalo de confiança

O valor crítico $z_{\gamma/2}$ é o valor de γ dividido por 2, uma vez que a "massa" γ deve ser distribuída igualmente em torno de 0.



Coeficiente de confiança γ

A área γ determina o **coeficiente de confiança** associado ao intervalo de confiança que estamos construindo.

O valor $z_{\gamma/2}$ pode ser obtido da tabela da Normal padrão, localizando o valor de $\gamma/2$ no corpo da tabela e obtendo o valor $z_{\gamma/2}$ nas margens correspondentes.

Exemplo: $\gamma = 0,95$:

- Temos que $\gamma/2 = 0,475$ é a área que devemos procurar no corpo da tabela
- O valor de $z_{\gamma/2}$ será determinado pelos valores correspondentes nas margens da tabela. Nesse caso, $z_{\gamma/2} = 1,96$ é o valor crítico procurado.

Intervalo de confiança

Com estas definições, podemos construir um **intervalo de confiança** para μ, com **coeficiente de confiança** γ:

$$\text{IC}(\mu, \gamma) = \left[\bar{X} - z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right); \bar{X} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

Outras notações:

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

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

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

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 encontre o valor crítico $z_{\gamma/2}$
3. Calcule a margem de erro $e = z_{\gamma/2} \cdot (\sigma/\sqrt{n})$
4. Calcule IC(μ, γ)

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança:
$\text{IC}(\mu, 95\%) = [52; 58]$

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:
$\text{IC}(\mu, 95\%) = [52; 58]$

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

Interpretação de um intervalo de confiança

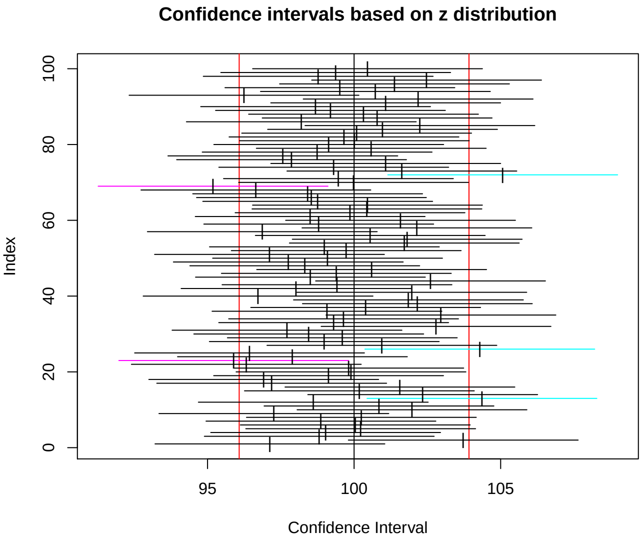
Como o intervalo de confiança é calculado a partir de uma **amostra aleatória**, este intervalo **também é aleatório**!

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.

Interpretação de um intervalo de confiança



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.

Construa um intervalo de confiança para a média μ com coeficiente de confiança de 95%.

Amplitude de um intervalo

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

$$
\begin{aligned}
\text{AMP}_{IC} &= \left[\bar{x} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)\right] - \left[\bar{x} - z_{\gamma/2} \cdot \left(\frac{\gamma}{\sqrt{n}}\right)\right] \\
&= 2 \times z_{\gamma/2} \cdot (\sigma/\sqrt{n})
\end{aligned}
$$

- Note que, claramente, um intervalo de confiança depende conjuntamente de três componentes:
  - Coeficiente de confiança γ, expresso pelo valor crítico $z_{\gamma/2}$
  - Desvio-padrão populacional σ
  - Tamanho da amostra n

Amplitude de um intervalo

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

$\sigma \rightarrow$ Um grande desvio padrão indica a possibilidade de um considerável distanciamento dos valores amostrais em relação à média populacional

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

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) 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.

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?

Determinação do tamanho amostral

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

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

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

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.