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 → **estimativa intervalar** ou **intervalo de confiança**

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

Plano de aula

1. Introdução
2. Intervalos de confiança para a média: σ conhecido
    - Determinação do tamanho amostral
3. Intervalos de confiança para a média: σ desconhecido
    - Determinação do tamanho amostral
4. Intervalo de confiança para a proporção p
    - Determinação do tamanho amostral
5. Referências

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

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

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

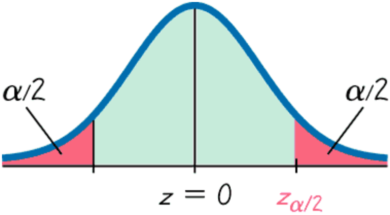
O **erro máximo provável** ou **margem de erro** da média é definido por

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

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

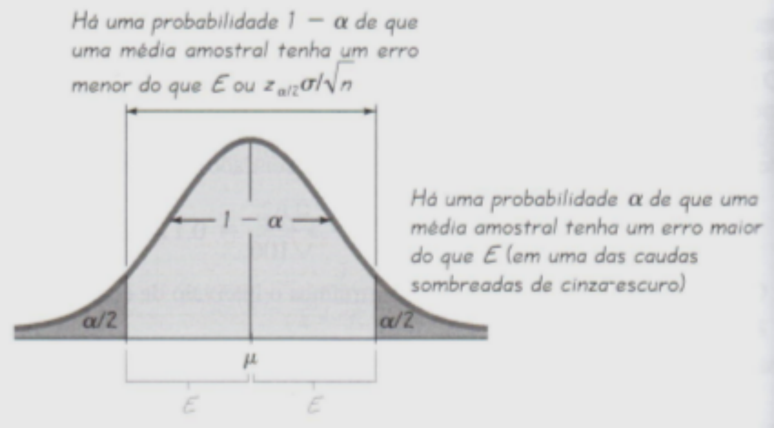
Valores críticos

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



Há uma probabilidade $1 - \alpha$ de que uma média amostral tenha um erro menor do que E ou $z_{\alpha/2}\sigma/\sqrt{n}$

Há uma probabilidade α de que uma média amostral tenha um erro maior do que E (em uma das caudas sombreadas de cinza-escuro)

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%), $\alpha = 0,01$

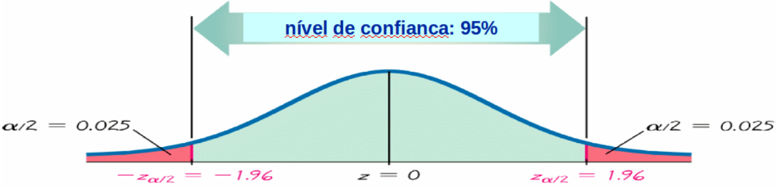
Importante!

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

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

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

Intervalos de confiança para a média: σ conhecido

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,

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

$$\Pr[\bar{x} - e < \mu < \bar{x} + e] = \gamma$$

Intervalos de confiança para a média: σ conhecido

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

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

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

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

Intervalos de confiança para a média: σ conhecido

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

Intervalos de confiança para a média: σ conhecido

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

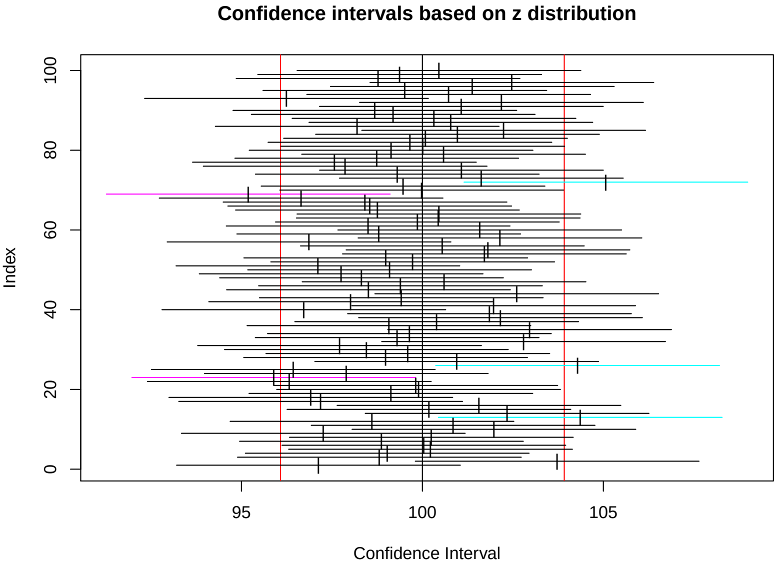
Como o intervalo de confiança é calculado a partir de uma **amostra aleatória**, este intervalo **também é aleatório**!

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.

Intervalos de confiança para a média: σ conhecido



Intervalos de confiança para a média: σ conhecido

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

Intervalos de confiança para a média: σ conhecido

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.

Intervalos de confiança para a média: σ conhecido

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?

Determinação do tamanho amostral

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.

Determinação do tamanho amostral

Respostas do exercício anterior

```
## a)
# (i)

[1] 554

# (ii)

[1] 35

## c)
# (i)

[1] 25

# (ii)

[1] 35
```

Plano de aula

Intervalos de confiança para a média: σ desconhecido

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio-padrão populacional σ

Se não conhecemos σ, então não podemos usar a distribuição normal padrão (Z) para estimarmos a verdadeira média populacional μ

Nesse caso, usaremos a distribuição t **de Student** que permite o uso da estimativa do **desvio-padrão amostral** s no lugar do valor desconhecido de σ.

A distribuição t de Student

Se uma população tem distribuição normal, então a distribuição da estatística

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

é uma **distribuição t de Student** (ou simplesmente distribuição t) com $n-1$ **graus de liberdade**.

A distribuição t de Student

Como não conhecemos σ, usamos o desvio-padrão amostral (não viesado):

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

mas isso introduz uma fonte de incerteza.

Para manter o nível de confiança desejado γ, compensamos essa incerteza calculando um intervalo de confiança um pouco maior: usando os valores críticos da distribuição $t_{\alpha/2;n-1}$ (ou simplesmente $t_{\alpha/2}$)

A distribuição t de Student

Para acharmos os valores críticos de $t_{\alpha/2;n-1}$, precisamos determinar o nível de confiança $\gamma = 1 - \alpha$ desejado e o valor dos **graus de liberdade** que indexa a distribuição t

Graus de liberdade (gl)

É o número de valores amostrais que podem variar depois que certas restrições tiverem sido impostas.

$$gl = n - 1$$

Dada uma média, apenas $n - 1$ valores podem ser associados **livremente**, antes que o último valor seja determinado.

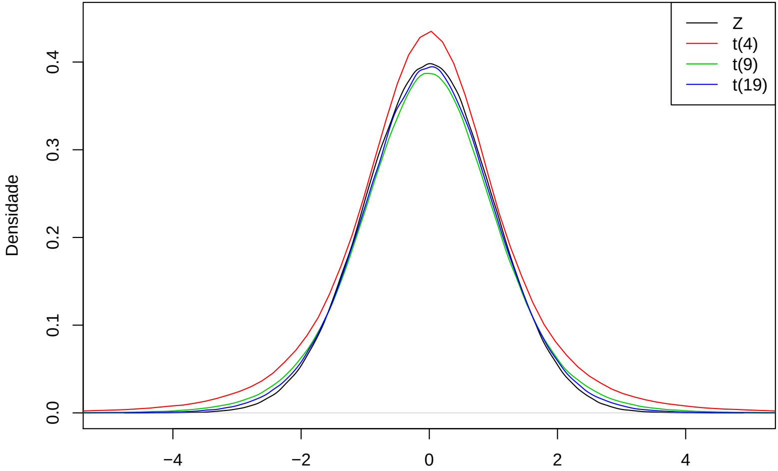
Exemplo: consideremos que 10 estudantes obtiveram média 8,0 em um teste. Assim, a soma das 10 notas deve ser 80 (restrição). Portanto, neste caso, temos $10 - 1 = 9$ graus de liberdade, pois as nove primeiras notas podem ser escolhidas aleatoriamente, contudo a 10^a nota deve ser igual a [80 - (soma das 9 primeiras)].

A distribuição t de Student

Características da distribuição t

- É simétrica com média $t = 0$ (assim como $z = 0$)
- É diferente para tamanhos de amostra diferentes
- Possui maior área nas caudas e menor área no centro (quando comparada com a distribuição normal) → para incorporar a incerteza
- O desvio-padrão da distribuição t varia com o tamanho da amostra (ao contrário da distribuição z onde $\sigma = 1$)
  - $n \downarrow \quad \sigma \uparrow$
  - $n \uparrow \quad \sigma \downarrow$
- Á medida que o n amostral aumenta, a distribuição t se aproxima cada vez mais de uma distribuição normal padrão Z
  - Por isso, para amostras grandes ($n > 30$) o resultado das duas é similar

A distribuição t de Student



Encontrando valores críticos de t

Com a definição do **nível de confiança** e sabendo o tamanho da amostra n, sabemos então o valor de α e dos gl, e devemos encontrar o valor de $t_{\alpha/2;n-1}$. Usando como exemplo $\gamma = 0,95 = 1 - 0,05 \Rightarrow \alpha = 0,05$ ou 5% e uma amostra de $n = 7$

- Temos que $n = 7 \Rightarrow gl = n - 1 = 6$
- Na tabela da distribuição t de Student procure a linha correspondente aos gl, e à coluna correspondente ao valor de α
- O valor de $t_{\alpha/2;n-1}$ será determinado pelos valores correspondentes **no corpo da tabela**. Nesse caso, $t_{\alpha/2;n-1} = 2,447$ é o valor crítico procurado.

Encontrando valores críticos de t

Sabemos que o valor crítico de 95% de confiança para a distribuição normal padrão é $z_{\alpha/2} = 1,96$. Quais são os valores críticos correspondentes para a distribuição t, se o n amostral for

- $n = 10$
- $n = 21$
- $n = 30$
- $n = 61$
- $n = 100$
- $n = 1000$

Verifique também os 4 primeiros valores para $\gamma = 0,9$ e $\gamma = 0,99$

Intervalos de confiança para a média: σ desconhecido

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

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

com

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

Outras notações

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

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

Intervalos de confiança para a média: σ desconhecido

Procedimentos gerais para a construção de intervalos de confiança com σ desconhecido

1. Verifique se as suposições necessárias estão satisfeitas
   - Temos uma AAS
   - Temos uma estimativa de s
   - *A população tem distribuição normal ou* $n > 30 \leftarrow$ este requisito não é mandatório, pois a distribuição t irá se "ajustar" para acomodar a incerteza de amostras pequenas
2. Determine o nível de confiança γ, e identifique α
3. Com o valor de α definido, e o valor dos **graus de liberdade** $gl = n - 1$, encontre o valor crítico de $t_{\alpha/2;n-1}$
4. Calcule a margem de erro $e = t_{\alpha/2} \cdot (s/\sqrt{n})$
5. Coloque em um dos formatos gerais para intervalo de confiança

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

   $\bar{x} \pm e$

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

Intervalos de confiança para a média: σ desconhecido

Exemplo: Em um teste da eficácia do alho na dieta para a redução do colesterol, 49 pessoas foram avaliadas e seus níveis de colesterol foram medidos antes e depois do tratamento. As **mudanças** nos níveis de colesterol apresentaram média de 0,4 e desvio-padrão de 21.

a) Verifique as suposições necessárias para o cálculo do intervalo de confiança
b) Para um nível de confiança de 95%, encontre o valor crítico $t_{\alpha/2;n-1}$
c) Calcule o erro máximo provável
d) Construa o intervalo de confiança
e) Escreva a interpretação do resultado
f) O que o intervalo de confiança sugere sobre a eficácia do uso do alho na dieta para a redução do colesterol?

Resolva o mesmo exemplo supondo que o $\sigma = s$ é conhecido (ou seja, usando a distribuição Z). Compare os dois métodos.

Intervalos de confiança para a média: σ desconhecido

Respostas do exercício anterior:

- Usando a distribuição t

```
$`valor critico`
[1] 2.0086

$`margem de erro`
[1] 5.9063

$`intervalo de confiança`
[1] -5.5063  6.3063
```

- Usando a distribuição z

```
$`valor critico`
[1] 1.96

$`margem de erro`
[1] 5.8799

$`intervalo de confiança`
[1] -5.4799  6.2799
```

Diferenças entre z e t

Para o mesmo exemplo anterior, assuma que foram avaliadas apenas 8 pessoas. Calcule intervalos de confiança para a média, $\bar{x} = 0,4$ usando

- a distribuição t (ou seja, assumindo σ desconhecido, portanto $s = 21$)
- a distribuição Z (ou seja, assumindo σ conhecido, portanto $\sigma = 21$)

Qual a sua conclusão referente à diferença observada entre os dois intervalos de confiança? Por que ocorre essa diferença?

Intervalos de confiança para a média: σ desconhecido

Respostas do exercício anterior:

- Usando a distribuição t

```
$`valor critico`
[1] 2.3646

$`margem de erro`
[1] 17.556

$`intervalo de confiança`
[1] -17.156  17.956
```

- Usando a distribuição z

```
$`valor critico`
[1] 1.96

$`margem de erro`
[1] 14.552

$`intervalo de confiança`
[1] -14.152  14.952
```

Escolha entre z e t

Método	**Condições**
Distribuição normal (Z)	σ conhecido e população normalmente distribuída ou σ conhecido e $n > 30$
Distribuição t	σ desconhecido e população normalmente distribuída ou σ desconhecido e $n > 30$

Escolha entre z e t

Nos exemplos abaixo, indique se é mais apropriado usar o valor crítico de $z_{\alpha/2}$ (da distribuição normal), de $t_{\alpha/2}$ (da distribuição t), ou nenhum deles

a) $n = 9$, $\bar{x} = 75$, $s = 15$, população com distribuição normal

b) $n = 5$, $\bar{x} = 20$, $s = 2$, população com distribuição muito assimétrica

c) $n = 12$, $\bar{x} = 98,6$, $\sigma = 0,6$, população com distribuição normal

d) $n = 75$, $\bar{x} = 98,6$, $\sigma = 0,6$, população com distribuição assimétrica

e) $n = 75$, $\bar{x} = 98,6$, $s = 0,6$, população com distribuição assimétrica

Plano de aula

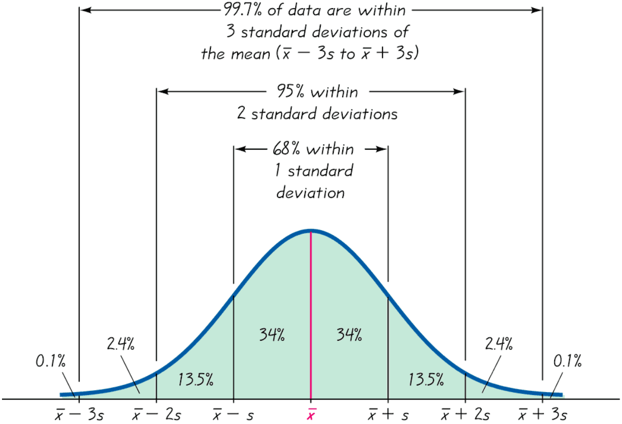
Determinação do tamanho amostral

Se σ for desconhecido?

1. Estime o valor de σ com base em algum estudo feito anteriormente
2. Faça uma amostra piloto e estime o desvio-padrão amostra s, e use-o como uma aproximação para o desvio-padrão populacional σ
3. Use a **regra empírica** da amplitude para dados com distribuição (aproximadamente) normal

Determinação do tamanho amostral

Regra empírica para uma distribuição (aproximadamente) normal



Determinação do tamanho amostral

Regra empírica para uma distribuição (aproximadamente) normal

Define-se **valores usuais** aqueles que são típicos e não muito extremos.

Como sabemos que em uma distribuição (aproximadamente) normal, 95% dos dados encontram-se a 2 desvios-padrões acima e abaixo da média, temos que

$$\begin{aligned} 4\sigma &= (\max - \min) \\ \sigma &= \frac{(\max - \min)}{4} \end{aligned}$$

Determinação do tamanho amostral

Exemplo: Um professor deseja estimar o salário médio de professores do ensino médio de uma cidade. Quantos professores devem ser selecionados para termos 90% de confiança que a média amostral esteja a menos de R$30,00 da média populacional? Sabe-se apenas que os salários variam entre R$800,00 e R$1.200,00.

Use

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

```
## Estimativa de sigma pela regra empírica

[1] 100

## Tamanho da amostra

[1] 31
```

Plano de aula

Intervalo de confiança para a proporção p

Em muitas situações, estamos interessados em estudar uma **proporção**, ao invés da média

Por exemplo:

- a proporção de eleitores de um determinado candidato
- a proporção de peças defeituosas em um processo industrial
- a proporção de alunos reprovados na disciplina de estatística

Nestes casos, a **distribuição binomial** deve ser utilizada no processo de inferência.

Intervalo de confiança para a proporção p

A proporção amostral

$$\hat{p} = \frac{x}{n} = \frac{\text{número de sucessos}}{\text{total de tentativas}}$$

é a "melhor estimativa" para a proporção populacional p

Exemplo: em 5 lançamentos de uma moeda considere que o evento "cara" (Ca) seja o sucesso ("sucesso" $=1$; "fracasso" $=0$). Um possível resultado seria o conjunto $\{Ca, Ca, Co, Co, Ca\}$. A proporção amostral seria

$$\hat{p} = \frac{\text{número de sucessos}}{\text{total de tentativas}} = \frac{3}{5} = 0,6$$

Distribuição amostral de uma proporção

Através do estudo da distribuição amostral da proporção, chegamos aos seguintes resultados

- $\text{E}(\hat{p}) = \mu_{\hat{p}} = p$
- $\text{Var}(\hat{p}) = \sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$

Ou seja,

$$p \sim \text{N}\left(p, \frac{p(1-p)}{n}\right)$$

Ainda podemos mostrar que a quantidade

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \text{N}(0, 1)$$

Quando não conhecemos p, usamos $\hat{p} = x/n$ como estimativa

Intervalo de confiança para a proporção p

Com estas definições, podemos construir um intervalo de confiança para uma estimativa da proporção amostral p através de

$$\hat{p} - e < p < \hat{p} + e$$

com

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

Outras notações

$$\hat{p} \pm e$$

$$[\hat{p} - e; \hat{p} + e]$$

Intervalo de confiança para a proporção p

Procedimentos gerais para a construção de intervalos de confiança para a proporção p

1. Verifique se as suposições necessárias estão satisfeitas
    - Temos uma AAS
    - As condições para a distribuição binomial são satisfeitas
        - as tentativas são independentes
        - há duas categorias de resultado ("sucesso", "fracasso")
        - a probabilidade de sucesso p permanece constante
    - A distribuição normal pode ser usada como aproximação para a distribuição binomial, ou seja, $np \geq 5$ e $n(1-p) \geq 5$
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 $e = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$
5. Coloque em um dos formatos gerais para intervalo de confiança

$$\hat{p} - e < \pi < \hat{p} + e$$

$$\hat{p} \pm e$$

$$[\hat{p} - e; \hat{p} + e]$$

Intervalo de confiança para a proporção p

Exemplo: em uma pesquisa realizada por um instituto de pesquisa Norte-Americano, 1500 adultos foram selecionados aleatoriamente para responder à pergunta se acreditam ou não no aquecimento global. 1050 entrevistados respoderam que sim. Com isso:

a) Verifique as suposições para o cálculo do intervalo de confiança
b) Para um nível de confiança de 95%, encontre o valor crítico de $z_{\alpha/2}$
c) Calcule o erro máximo provável
d) Construa o intervalo de confiança para p
e) Escreva a interpretação do resultado
f) Com base nesse resultado, podemos concluir que a maioria dos adultos acredita no aquecimento global?

Intervalo de confiança para a proporção p

Resposta do exercício anterior:

```
$`valor critico`
[1] 1.96

$`margem de erro`
[1] 0.023191

$`intervalo de confiança`
[1] 0.67681 0.72319
```

Plano de aula

Determinação do tamanho amostral

A partir da equação do erro máximo provável

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

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

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

Determinação do tamanho amostral

Quando conhecemos a verdadeira proporção populacional p, podemos usá-la para calcular o n a partir da expressão acima.

Quando **não conhecemos** a proporção poppulacional p, usamos como estimativa $p = 0,5$ porque

p	$(1-p)$	$p(1-p)$
0,1	0,9	0,09
0,3	0,7	0,21
0,5	0,5	**0,25**
0,6	0,4	0,24
0,8	0,2	0,16

Quando $p = 0,5$ teremos o maior tamanho de amostra possível.

Determinação do tamanho amostral

Exemplo: um fabricante de peças deseja estimar a verdadeira proporção de peças defeituosas no processo de fabricação, com um erro máximo de 3% e nível de confiança de 99%. Calcule o tamanho da amostra necessário para se estimar esta proporção se:

a) O fabricante acredita que aproximadamente 10% de seus produtos são defeituosos.

b) O fabricante não tem nenhuma informação prévia sobre a proporção de peças defeituosas.

```
## a)

[1] 664

## b)

[1] 1844
```

Plano de aula

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]