

# Estimação Intervalar

Luan D. Fiorentin

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

08/11/2019



# Sumário

## 1 Introdução

## 2 Intervalos de Confiança para a Média: $\sigma$ Conhecido

- IC para Média:  $\sigma$  Conhecido
- Dimensionamento de amostra

## 3 Intervalos de Confiança para a Média: $\sigma$ Desconhecido

- IC para Média:  $\sigma$  Desconhecido
- Dimensionamento de amostra

## 4 Proporção amostral

## 5 Exercícios recomendados

# Introdução

- Existem dois tipos de **estimativas** que se pode 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.
- Por serem **variáveis aleatórias**, os estimadores pontuais possuem uma **distribuição de probabilidade** (distribuições amostrais).

# Introdução

- É possível 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.
  - Por exemplo: é possível obter um intervalo de confiança para a média amostral.

# Sumário

## 1 Introdução

## 2 Intervalos de Confiança para a Média: $\sigma$ Conhecido

- IC para Média:  $\sigma$  Conhecido
- Dimensionamento de amostra

## 3 Intervalos de Confiança para a Média: $\sigma$ Desconhecido

- IC para Média:  $\sigma$  Desconhecido
- Dimensionamento de amostra

## 4 Proporção amostral

## 5 Exercícios recomendados

## Pressuposição inicial

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

# Erro amostral

- Quando coleta-se uma amostra aleatória e calcula-se uma média, o valor da média possui um **desvio natural** em relação ao verdadeiro valor da média populacional (**erro amostral**), ou seja:

$$e = \bar{X} - \mu$$

$$\bar{X} = \mu + e.$$

- Sabe-se que a distribuição amostral da média é uma **distribuição normal**, com média  $\mu$  e variância  $\sigma^2/n$ :

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

# Margem de erro

- Usando a transformação

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

pode-se 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  $\gamma$  tal que  $0 < \gamma < 1$ , é possível encontrar um valor  $z_{\gamma/2}$  tal que:

$$P[|Z| < z_{\gamma/2}] = \gamma$$

$$P[-z_{\gamma/2} < Z < z_{\gamma/2}] = \gamma$$

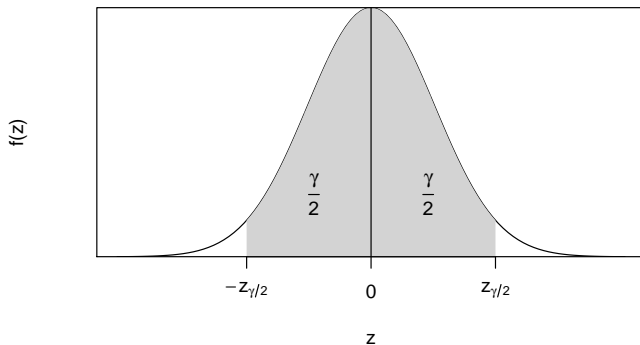
$$P\left[-z_{\gamma/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\gamma/2}\right] = \gamma$$

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

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

## Intervalo de confiança

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



# Intervalo de confiança

- A área  $\gamma$  representa o **coeficiente de confiança** associado ao **intervalo de confiança** que está sendo construído.
- O valor  $z_{\gamma/2}$  pode ser obtido da tabela da **Normal Padrão**, localizando o valor de  $\gamma/2$  obtido na tabela e determinando o valor  $z_{\gamma/2}$  nas margens.
- Exemplo: considere o coeficiente de confiança  $\gamma = 0,95$ :
  - Tem-se que  $\gamma/2 = 0,475$  é a área a ser procurada na tabela.
  - $\gamma/2 = 0,475 = 1,96$  é o valor crítico.

## Construção do intervalo de confiança

- Podemos construir um intervalo de confiança com um coeficiente de confiança para  $\mu$ , com coeficiente de confiança  $\gamma$ :

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

Com outras notações:

$$IC[\mu, \gamma] = [\bar{X} - e; \bar{X} + e]$$

$$IC[\mu, \gamma] = \bar{x} \pm e$$

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

# Como construir um intervalo de confiança???

- Verifique se as suposições são satisfeitas:
  - É uma amostra aleatória;
  - Desvio padrão é conhecido;
  - População tem distribuição normal ou  $n > 30$ .
- Determinar o nível de confiança  $\gamma$  e obter o valor crítico  $z_{\gamma/2}$ .
- Determinar a margem de erro  $e$ .
- Calcular o  $IC[\mu, \gamma]$ .

## Interpretação do intervalo de confiança

- Suponha que foi obtido um intervalo de 95% de confiança:  $IC[\mu; 0, 95] = [95; 105]$

### Interpretação 1

Temos 95% de confiança de que a verdadeira média populacional  $\mu$  se encontra entre 95 e 105.

### Interpretação 2

Temos 95% de confiança de que o intervalo entre 95 e 105 realmente contém a verdadeira média populacional  $\mu$ .

## Interpretação do intervalo de confiança

- Suponha que foi obtido um intervalo de 95% de confiança:  $IC[\mu; 0, 95] = [95; 105]$

### Interpretação 1 — ERRADA

Temos 95% de confiança de que a verdadeira média populacional  $\mu$  se encontra entre 95 e 105.

### Interpretação 2 — CERTA

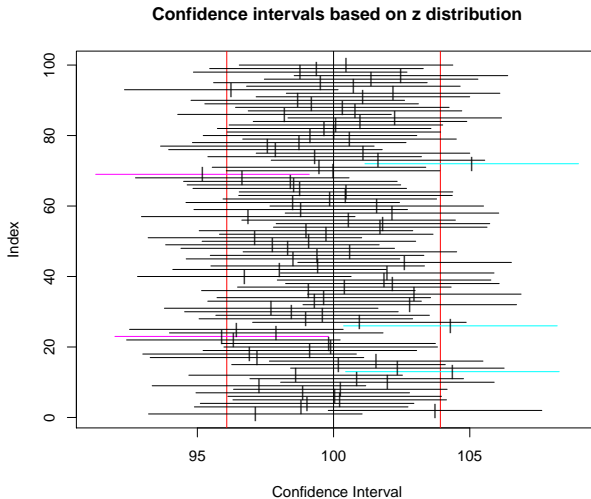
Temos 95% de confiança de que o intervalo entre 95 e 105 realmente contém a verdadeira média populacional  $\mu$ .

# Interpretação do 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  $\mu$  é fixo, é o intervalo que deve conter o valor de  $\mu$ , e não o contrário.
- Isso significa que se pudermos obter 100 amostras diferentes, e calcularmos um intervalo de confiança de 95% para cada uma das 100 amostras, esperamos que 5 destes intervalos **não** contenham o verdadeiro valor da média populacional  $\mu$ .



# Interpretação do intervalo de confiança



## Exemplo 1

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  $\mu$  com coeficiente de confiança de 95%.

## Exemplo 1

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

$$IC[\mu; 0,95] = \left[ 22,4 - 1,96 \cdot \left( \frac{5,2}{\sqrt{25}} \right); 22,4 + 1,96 \cdot \left( \frac{5,2}{\sqrt{25}} \right) \right]$$

$$IC[\mu; 0,95] = [22,4 - 2,04; 22,4 + 2,04]$$

$$IC[\mu; 0,95] = [20,36 < \mu < 24,44]$$

$$IC[\mu; 0,95] = [20,36; 24,44].$$

## Amplitude do intervalo

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

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

$$AMP_{IC} = 2 \cdot z_{\gamma/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

## Amplitude do intervalo

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

- $z_{\gamma/2}$ : cada vez que aumenta-se a confiança  $\gamma$ , o valor de  $z_{\gamma/2}$  fica maior e, conseqüentemente, a amplitude do intervalo aumenta.
- $\sigma$ : um grande desvio padrão indica a possibilidade de um considerável distanciamento dos valores amostrais em relação à média populacional.
- $n$ : quanto maior for o tamanho da amostra, maior será a quantidade de informação disponível. Com isso, valores maiores de  $n$  produzem intervalos mais informativos.

## Exemplo 2

- Seja  $X \sim N(\mu, 36)$ .
- ❶ Para uma amostra de tamanho 50, obtivemos média amostral de 18,5. Construa intervalos de confiança de: i) 90%; ii) 95%; e iii) 99%.
- ❷ Calcule as amplitudes dos intervalos acima e explique a diferença.
- ❸ Para um nível de confiança de 95%, construa intervalos de confiança (admita a mesma média amostral de 18,5) supondo tamanhos de amostra de: i) 15; e ii) 100.
- ❹ Calcule as amplitudes dos intervalos acima e explique a diferença.

## Exemplo 2

❶  $IC[\mu; 90\%] = [17, 11; 19, 89]$

$$IC[\mu; 95\%] = [16, 84; 20, 16]$$

$$IC[\mu; 99\%] = [16, 33; 20, 67].$$

❷  $AMP_{IC}[\mu; 90\%] = 2, 78$

$$AMP_{IC}[\mu; 95\%] = 3, 32$$

$$AMP_{IC}[\mu; 99\%] = 4, 34.$$

❸  $AMP_{IC}[\mu; 95\%] = [15, 46; 21, 54]$

$$AMP_{IC}[\mu; 95\%] = [17, 32; 19, 68].$$

❹  $AMP_{IC}[\mu; 95\%] = 6, 08$

$$AMP_{IC}[\mu; 95\%] = 2, 36.$$

# Dimensionamento de amostra

- O objetivo do estudo é **estimar a média populacional  $\mu$  desconhecida**. Logo, a pergunta é:

Quantos elementos (itens, objetos, pessoas, ...) eu devo amostrar?

- **Em geral, recomenda-se inicialmente  $n > 30$  como um tamanho de amostra mínimo para grande parte dos problemas.**
- No entanto, pode-se obter uma **estimativa mais adequada** de quantos elementos amostrar.



# Dimensionamento de amostra

- A partir do **erro máximo provável**

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

pode-se isolar  $n$  e determinar uma **expressão** para estimar o tamanho amostral

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

# Dimensionamento de amostra

- Tamanho da amostra  $n$  não depende do tamanho da **população**  $N$ .
- No entanto, o **tamanho amostral depende** do:
  - Nível de confiança desejado ( $\gamma$ );
  - Erro máximo desejado ( $e$ );
  - Desvio padrão ( $\sigma$ ).

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

- O tamanho amostral precisa ser um **número inteiro**, arredonda-se sempre o valor para o maior número inteiro mais próximo.

## Exemplo 3

Desejamos coletar uma amostra de uma variável aleatória  $X$ , com distribuição Normal, de média desconhecida e variância 30. Qual deve ser o tamanho da amostra para que, com 95% de probabilidade:

- 1 A média amostral não difira da média populacional por mais de 1 unidade? E por mais de 4 unidades?
- 2 Qual o impacto do erro máximo assumido para o tamanho da amostra?
- 3 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 90% e 95%.
- 4 Analise o impacto do nível de confiança para a determinação do tamanho amostral.

## Exemplo 3

1

$$n = \left[ \frac{1,96 \cdot \sqrt{30}}{1} \right]^2 = 116.$$

$$n = \left[ \frac{1,96 \cdot \sqrt{30}}{4} \right]^2 = 8.$$

2 Quanto menor o erro, maior o tamanho amostral.

3

$$n = \left[ \frac{1,64 \cdot \sqrt{30}}{2} \right]^2 = 21.$$

$$n = \left[ \frac{1,96 \cdot \sqrt{30}}{2} \right]^2 = 29.$$

4 Quanto menor o nível de confiança, menor o tamanho amostral.

# Sumário

- 1 Introdução
- 2 Intervalos de Confiança para a Média:  $\sigma$  Conhecido
  - IC para Média:  $\sigma$  Conhecido
  - Dimensionamento de amostra
- 3 Intervalos de Confiança para a Média:  $\sigma$  Desconhecido
  - IC para Média:  $\sigma$  Desconhecido
  - Dimensionamento de amostra
- 4 Proporção amostral
- 5 Exercícios recomendados

## Estimativa da variância amostral

- Na maioria das situações práticas, não sabemos o **verdadeiro** valor do desvio padrão populacional  $\sigma$ .
- Se o desvio padrão é desconhecido, ele precisa ser estimado.
- Considere variáveis aleatórias  $(X_1, X_2, \dots, X_n)$ , onde  $X \sim N(\mu, \sigma^2)$ , sabe-se que o “melhor” estimador para  $\sigma^2$  é a variância amostral

$$S^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right),$$

a qual não é viciada e é consistente para  $\sigma^2$ .

# A distribuição $t$ de Student

- A **variável padronizada** é

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

- Logo, o denominador  $S^2$  fará com que  $T$  seja diferente da Normal.
- Essa densidade é denominada **t de Student**, e seu parâmetro é denominado **graus de liberdade**, que nesse caso é  $n - 1$ . Portanto,

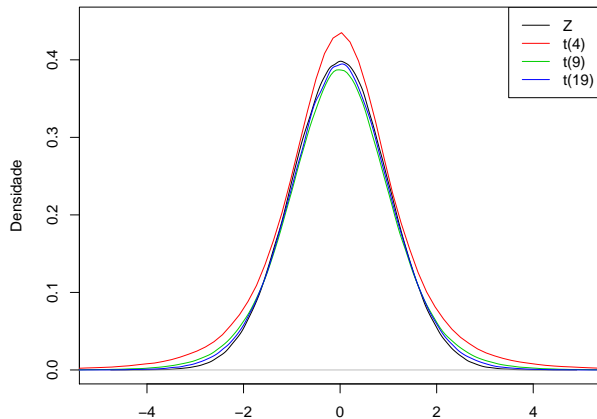
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

## Características da distribuição $t$

- É **simétrica** com média  $t = 0$  (assim como  $z = 0$ ).
- É **diferente** para tamanhos de amostra diferentes.
- Maior área nas caudas e menor área no centro, quando comparada com a distribuição normal, serve para incorporar as **incertezas**.
- **Desvio padrão** da distribuição  $t$  varia com o tamanho da amostra (ao contrário da distribuição  $z$  onde  $\sigma = 1$ ):
  - Se  $n$  diminui,  $\sigma$  aumenta.
  - Se  $n$  aumenta,  $\sigma$  diminui.
- A medida que o  $n$  amostral aumenta, a distribuição  $t$  se aproxima cada vez mais de uma **distribuição normal padrão  $Z$** .



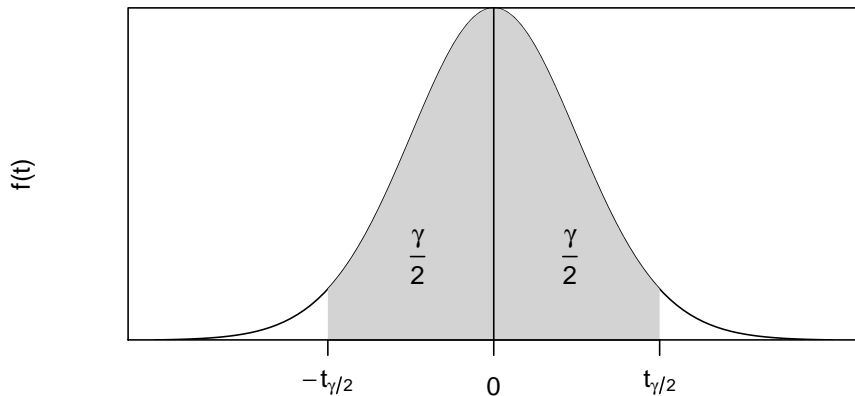
# Características da distribuição $t$



## Como encontrar os valores críticos de $t$ ?

- Com a definição do **nível de confiança** e sabendo o **tamanho da amostra**  $n$ , sabe-se então o valor de  $\gamma$  e dos G.L., podendo encontrar o **valor crítico** de  $t_{\gamma/2}$ .
- Exemplo:  $\gamma = 0,95$  e uma amostra de tamanho  $n = 7$ :
  - Se  $n = 7$ , o G.L. associado é  $n - 1 = 6$ .
  - Na tabela da distribuição  $t$  de Student procura-se a linha correspondente aos G.L. e a coluna correspondente ao valor de  $1 - \gamma = 1 - 0,95 = 0,05 = 5\%$ .
  - Valor de  $t_{\gamma/2}$  será determinado pelos valores correspondentes no corpo da tabela. Nesse caso,  $t_{\gamma/2} = 2,447$  é o valor crítico procurado.

## Como encontrar os valores críticos de $t$ ?



## Construção do intervalo de confiança

- Podemos construir um intervalo de confiança com um coeficiente de confiança:

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

$$IC[\mu, \gamma] = [\bar{x} - e; \bar{x} + e] = \bar{x} \pm e$$

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

## Como construir um intervalo de confiança?

- Verifique se as suposições são satisfeitas:
  - É uma amostra aleatória.
  - Desvio padrão é **estimado**.
  - População tem distribuição normal ou  $n > 30$ .
- Determinar o nível de confiança  $\gamma$  e o valor crítico  $t_{\gamma/2}$ .
- Determinar a margem de erro  $e = t_{\gamma/2} \cdot (S/\sqrt{n})$ .
- Calcular o  $IC[\mu, \gamma]$ .

## Exemplo 4

Em um teste da eficácia do alho na dieta para a redução do colesterol, 51 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.

- 1 Para um nível de confiança de 95%, calcule o intervalo para a verdadeira média das mudanças no nível de colesterol.
- 2 O que o intervalo de confiança sugere sobre a eficácia do uso do alho na dieta para a redução do colesterol?
- 3 Resolva o mesmo exemplo supondo que o  $\sigma = s$  é conhecido (usando a distribuição  $Z$ ). Compare os resultados.

## Exemplo 4

1

$$IC[\mu; 0,95] = \left[ 0,4 - 2,01 \cdot \left( \frac{21}{\sqrt{51}} \right); 0,4 + 2,01 \cdot \left( \frac{21}{\sqrt{51}} \right) \right]$$
$$IC[\mu; 0,95] = [-5,51; 6,31].$$

2

O intervalo de confiança sugere que o alho não é eficiente na dieta.

3

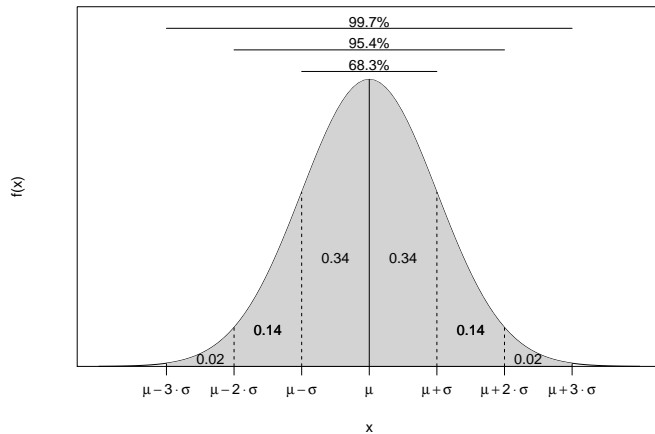
$$IC[\mu; 0,95] = \left[ 0,4 - 1,96 \cdot \left( \frac{21}{\sqrt{51}} \right); 0,4 + 1,96 \cdot \left( \frac{21}{\sqrt{51}} \right) \right]$$
$$IC[\mu; 0,95] = [-5,36; 6,16].$$

# Dimensionamento de amostra

- Se  $\sigma$  é **desconhecido**:
- Estimar o valor de  $\sigma$  com base em estudo preliminar.
- Faça uma amostra piloto e estime o desvio padrão amostral  $s$  e use-o como uma aproximação para o desvio-padrão populacional  $\sigma$ .
- Use a **regra empírica** da amplitude para dados com distribuição (aproximadamente) normal.



# Regra empírica para uma distribuição normal



## Regra empírica para uma distribuição normal

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

$$4\sigma = (\max - \min)$$

$$\hat{\sigma} = s = \frac{(\max - \min)}{4}.$$

## Exemplo 5

Um professor deseja estimar o salário médio de professores do ensino médio de uma cidade.

Quantos professores devem ser selecionados para termos 95% 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.

## Exemplo 5

Um professor deseja estimar o salário médio de professores do ensino médio de uma cidade.

Quantos professores devem ser selecionados para termos 95% 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.

$$n = \left[ \frac{z_{\gamma/2} \cdot \sigma}{e} \right]^2 = \left[ \frac{1,96 \cdot 100}{30} \right]^2 = 43.$$

# Sumário

- 1 Introdução
- 2 Intervalos de Confiança para a Média:  $\sigma$  Conhecido
  - IC para Média:  $\sigma$  Conhecido
  - Dimensionamento de amostra
- 3 Intervalos de Confiança para a Média:  $\sigma$  Desconhecido
  - IC para Média:  $\sigma$  Desconhecido
  - Dimensionamento de amostra
- 4 **Proporção amostral**
- 5 Exercícios recomendados

# Proporção amostral

- A **proporção amostral** é dada como:

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

onde  $\hat{p}$  é o melhor estimador para a proporção populacional  $p$ .

- Por meio do estudo da distribuição amostral da proporção chegamos aos seguintes resultados:
  - A proporção amostral  $\hat{p}$  tende para o valor da proporção populacional  $p$ ;
  - A distribuição das proporções amostrais tende a ser uma **distribuição normal**;
  - $E[\hat{p}] = \mu_{\hat{p}} = p$ ;
  - $V[\hat{p}] = \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$ .

# Proporção amostral

- Assim, sabe-se que

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

- Ainda, pode-se mostrar que a quantidade

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

# Erro amostral da proporção

- O **erro amostral da proporção** pode ser definido por

$$e = p - \hat{p} \quad \rightarrow \quad p = \hat{p} + e,$$

então,

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

- Portanto, o **erro máximo provável ou margem de erro da proporção** é dado como

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



# Intervalo de Confiança para Proporção

- Com estas definições, podemos construir um **intervalo de confiança para  $p$** , com coeficiente de confiança  $\gamma$ .

$$IC(p, \gamma) = \left[ \hat{p} - z_{\gamma/2} \cdot \sqrt{\frac{p(1-p)}{n}}; \hat{p} + z_{\gamma/2} \cdot \sqrt{\frac{p(1-p)}{n}} \right].$$

Outras notações:

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

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

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

# Como construir um intervalo de confiança?

- Verifique se as suposições são satisfeitas:
  - É uma **amostra aleatória simples** (AAS).
  - As condições para uma **distribuição binomial** são satisfeitas:
    - As tentativas são independentes;
    - Há duas categorias de resultado (“sucesso”, “fracasso”);
    - Probabilidade de sucesso **p** permanece **constante**;
    - A distribuição normal pode ser usada como **aproximação** para a distribuição binomial, ou seja,  $np \geq 5$  e  $np(1 - p) \geq 5$ .
- Determinar o nível de confiança  $\gamma$  e o valor crítico  $z_{\gamma/2}$ .
- Determinar a margem de erro  $e$ .
- Calcular o  $IC[p, \gamma]$ .

# Estimação do $p$

- Uma possível dificuldade nessa abordagem é que em geral não conhecemos o verdadeiro valor de  $p$  para calcular o IC.
- Quando não conhecemos a proporção populacional  $p$ , temos duas alternativas:
  - Usar  $\hat{p}$  no lugar de  $p$  (**estimativa otimista**).
  - Usar  $p = 0,5$  (**estimativa conservadora**). Quando  $p = 0,5$ , o termo  $p(1 - p)$  terá valor máximo:

$p$	$(1 - p)$	$p(1 - p)$
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

## Exemplo 6

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 responderam que sim. Com isso:

- 1 Para um nível de confiança de 95%, calcule o intervalo de confiança para a verdadeira proporção de pessoas que acreditam no aquecimento global, utilizando i)  $p = \hat{p}$  e ii)  $p = 0,5$  e compare os resultados.
- 2 Com base nesses resultados, podemos concluir que a maioria dos adultos acredita no aquecimento global?

## Exemplo 6

1

$$IC(p; 95\%) = \left[ 0,7 \pm 1,96 \cdot \sqrt{\frac{0,7(1 - 0,7)}{1500}} \right]$$

$$IC(p; 95\%) = [0,68; 0,72].$$

$$IC(p; 95\%) = \left[ 0,5 \pm 1,96 \cdot \sqrt{\frac{0,5(1 - 0,5)}{1500}} \right]$$

$$IC(p; 95\%) = [0,48; 0,53].$$

- 2 Pode-se afirmar que a maioria dos adultos acreditam no aquecimento global quando usamos uma proporção estimada com base na amostra.

# Erro máximo

- A partir da equação do erro máximo provável

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

se pode isolar  $n$  e chegar na seguinte equação para a determinação do tamanho amostral para uma proporção populacional:

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

- Quando não conhecemos  $p$ , usamos  $\hat{p}$  (estimativa otimista) ou  $p = 0,5$  (estimativa conservadora) como estimativas de  $p$ .

## Exemplo 7

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:

- 1 O fabricante tem uma estimativa de que em uma amostra anterior, aproximadamente 10% das peças eram defeituosas.
- 2 O fabricante não tem nenhuma informação prévia sobre a proporção de peças defeituosas.

## Exemplo 7

1

$$n = \left( \frac{2,57}{0,03} \right)^2 \cdot 0,1(1 - 0,1) = 661.$$

2

$$n = \left( \frac{2,57}{0,03} \right)^2 \cdot 0,5(1 - 0,5) = 1835.$$



# Sumário

## 1 Introdução

## 2 Intervalos de Confiança para a Média: $\sigma$ Conhecido

- IC para Média:  $\sigma$  Conhecido
- Dimensionamento de amostra

## 3 Intervalos de Confiança para a Média: $\sigma$ Desconhecido

- IC para Média:  $\sigma$  Desconhecido
- Dimensionamento de amostra

## 4 Proporção amostral

## 5 Exercícios recomendados

# Exercícios recomendados

- Seção 7.4: Ex. 1, 2, 3, 4 e 5.
- Seção 7.5: Ex. 17, 18, 19, 21, 23 e 25 e 29.