



Estatística I

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Introdução

Intervalos de Confiança para a Média: σ conhecido

Intervalos de confiança para a média: σ desconhecido

Intervalo de Confiança para Proporção

Determinação do tamanho amostral (σ conhecido)

Determinação do tamanho amostral (σ desconhecido)

Introdução

Por que não confiar só na média?

Imagine que você está com muita fome e abre o aplicativo de delivery. A informação aparece assim:

"O tempo médio de entrega é de 30 minutos."

Você pensa:

"Perfeito! Em meia horinha eu tô comendo!"

Por que não confiar só na média?

Mas... será que é tão simples assim?

Pense nos dados que o aplicativo usa para calcular essa média:

- Algumas entregas foram muito rápidas (15, 20 minutos).
- Outras demoraram bastante (60, 70, até 90 minutos).

A **média** de 30 minutos parece bonita... mas esconde toda essa variabilidade!

A verdade por trás da média

Se o aplicativo dissesse:

"Com 95% de confiança, seu pedido chegará entre 20 minutos e 1 hora e 10 minutos."

Agora sim, você entende o jogo!

Isso significa que:

- É possível que chegue rápido (20 min).
- Mas também existe uma chance real de demorar mais de uma hora.

Percebe a diferença?

"A média é uma informação solitária. O intervalo de confiança é uma informação honesta."

"A média é uma informação solitária. O intervalo de confiança é uma informação honesta."

Confiar só na média é como dirigir olhando apenas pelo retrovisor... Parece informação, mas não te mostra o que vem pela frente.

Dois tipos de estimativas

Estimativa Pontual

É quando usamos um único número, calculado a partir da amostra, para estimar um parâmetro populacional.

Exemplos:

- Média amostral (\bar{x}) para estimar a média populacional (μ).
- Proporção amostral (\hat{p}) para estimar a proporção populacional (p).

Limitação: Fornece apenas um valor. Não diz nada sobre a incerteza ou confiabilidade desse valor.

Dois tipos de estimativas

Estimativa Intervalar (Intervalo de Confiança)

Em vez de fornecer um único número, fornece um **intervalo de valores plausíveis** para o parâmetro populacional.

Exemplo:

"Com 95% de confiança, a média populacional está entre 25 e 35."

Vantagem: Expressa não só a estimativa, mas também a incerteza associada a ela.

Por que precisamos de um intervalo de confiança?

Todo estimador (como a média amostral) é uma **variável aleatória**.

- Se coletarmos outra amostra, vamos obter outro valor.
- A cada amostra possível, temos uma média diferente.

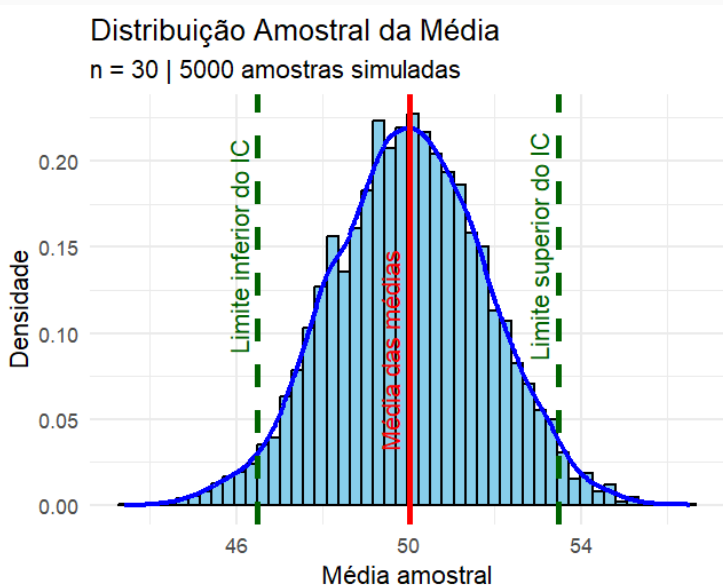
Por isso, o estimador possui uma **distribuição de probabilidade**, chamada de **distribuição amostral**.

E é exatamente a partir dessa distribuição que construímos o intervalo de confiança.

O intervalo de confiança nos permite afirmar algo do tipo:

"Se eu repetir esse processo muitas vezes, 95% dos intervalos conterão o verdadeiro valor do parâmetro."

Visualizando a incerteza



Intervalos de Confiança para a Média: σ conhecido

Suposições Necessárias

Para construirmos um intervalo de confiança para a média (com σ conhecido), precisamos garantir:

- A amostra é uma **amostra aleatória simples (AAS)**.
- O desvio padrão da população (σ) é conhecido.
- E uma das seguintes condições:
 - A população tem distribuição normal;
 - ou o tamanho da amostra é suficientemente grande ($n > 30$).

Erro Amostral: Sempre Existe!

Ao coletar uma amostra, a média amostral (\bar{X}) dificilmente será exatamente igual à média populacional (μ).

Essa diferença é chamada de erro amostral:

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

Sabemos que a média amostral segue uma distribuição:

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

Ou seja, as médias amostrais **variam** de amostra para amostra!

O que é a Margem de Erro?

Se padronizarmos a média amostral, obtemos:

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

Z responde: Quantos desvios padrão minha média amostral está distante da média populacional.

A **margem de erro** (e) representa o erro máximo aceitável, dentro de um grau de confiança (γ):

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

Onde $z_{\gamma/2}$ é o **valor crítico da normal padrão**.

Construindo o Intervalo de Confiança

Raciocínio

Queremos capturar o valor de μ dentro de um intervalo simétrico ao redor da média amostral \bar{x} .

$$P(-z_{\gamma/2} < Z < z_{\gamma/2}) = \gamma$$

Substituindo Z pela padronização da média:

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

Pronto! Este é o intervalo de confiança para μ com confiança γ .

Valor Crítico $z_{\gamma/2}$

O valor $z_{\gamma/2}$ é o ponto da distribuição normal padrão que deixa uma área de $\gamma/2$ em cada cauda.

Por exemplo, para $\gamma = 0,95$:

- A área central é 95%.
- Sobra 5% para as caudas \rightarrow 2,5% em cada lado.
- Buscamos na tabela da normal padrão a área acumulada até 0,975.
- Resultado: $z_{0,025} = 1,96$.

A área central corresponde ao nível de confiança γ .

Fórmula do Intervalo de Confiança

O intervalo de confiança para μ , com nível de confiança γ , é dado por:

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

Interpretação: Uma faixa de valores plausíveis para a média populacional, considerando a variabilidade natural das amostras.

Passos para construir o Intervalo de Confiança

1. Verificar as suposições:
 - AAS
 - σ conhecido
 - População normal ou $n > 30$
2. Escolher o nível de confiança γ e determinar $z_{\gamma/2}$.
3. Calcular a margem de erro:

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

4. Construir o intervalo:

$$\bar{X} \pm e$$

Interpretação do Intervalo de Confiança

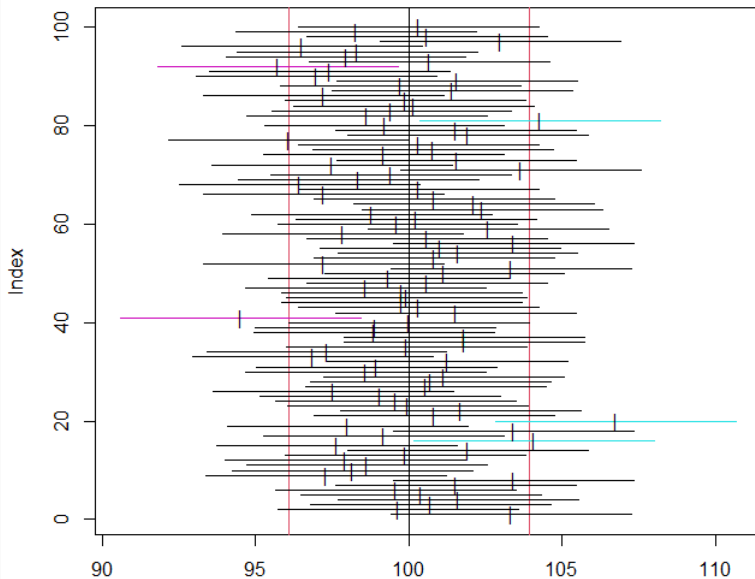
Atenção: O que significa γ ?

- O **parâmetro** (μ) é fixo. - O que varia é o **intervalo**, porque ele depende da amostra.

Se construirmos 100 intervalos de confiança de 95%, usando 100 amostras diferentes, **esperamos que 95 deles contenham** μ , e 5 não contenham.

Importante: *Não dizemos que "a chance de μ estar no intervalo é 95%". μ não é aleatório. O que é aleatório é o intervalo.*

Confidence intervals based on z distribution



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

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

$$(20.36164 \leq \mu \leq 24.43836)$$

Intervalos de confiança para a média: σ desconhecido

Intervalos de confiança para a média: σ desconhecido

Estimativa da variância amostral

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio padrão populacional σ . Se o desvio padrão é desconhecido, ele precisa ser estimado. Sendo (X_1, \dots, X_n) VAs onde $X \sim N(\mu, \sigma^2)$, vimos que o “melhor” estimador para σ^2 é a variância amostral

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

que é não viciada e consistente para σ^2 .

A distribuição t de Student

Definindo a variável padronizada

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

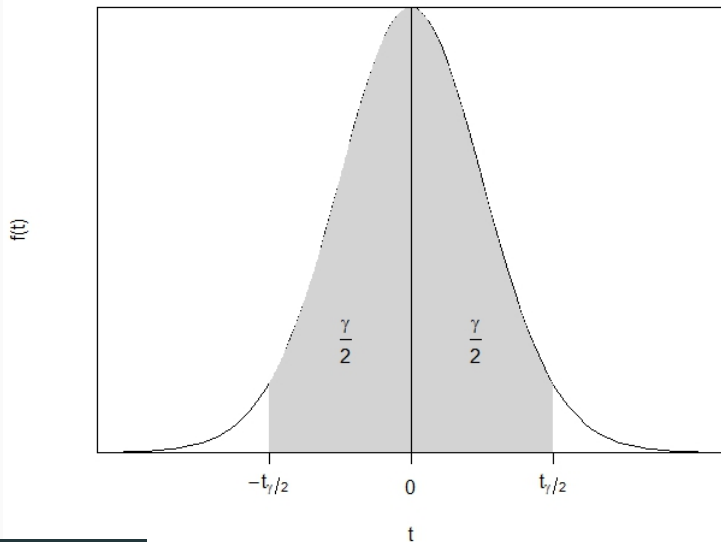
o denominador S^2 fará com que a função densidade de T seja diferente da Normal. Essa nova densidade é denominada **t de Student**, e seu parâmetro é denominado **graus de liberdade**, que nesse caso é $n - 1$. Assim:

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

Valores Críticos de t

Com a definição do **nível de confiança** e sabendo o tamanho da amostra n , sabemos então o valor de γ e dos gl, e devemos encontrar o **valor crítico** de $t_{\gamma/2}$. Usando como exemplo $\gamma = 0,95$ 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 coluna correspondente ao valor de $1 - \gamma = 1 - 0,95 = 0,05 = 5\%$
- O 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.



Intervalo de confiança

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

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

Procedimentos para a construção de intervalos de confiança

1. Verifique se as suposições necessárias estão satisfeitas
 - Temos uma AAS
 - Temos uma estimativa de s
 - A população tem distribuição normal ou $n > 30$
2. Determine o nível de confiança γ , e encontre o valor crítico $t_{\gamma/2}$
3. Calcule a margem de erro $e = t_{\gamma/2} \cdot (s/\sqrt{n})$
4. Calcule IC(μ, γ)

Exemplo

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.

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

Intervalo de Confiança para Proporção

Intervalo de Confiança para Proporção

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 . Através 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**

Distribuição amostral da proporção \hat{p}

Assim, sabemos que

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

É possível mostrar que a quantidade

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

Intervalo de Confiança

Logo, podemos construir um **intervalo de confiança** para p , com **coeficiente de confiança** γ

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

Procedimentos para a construção de intervalos de confiança

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 $np(1 - p) \geq 5$
2. Determine o nível de confiança γ , e encontre o valor crítico $z_{\gamma/2}$

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 responderam que sim. Com isso:

- a) 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.
- b) Com base nesses resultados, podemos concluir que a maioria dos adultos acredita no aquecimento global?

Determinação do tamanho amostral (σ conhecido)

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**

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

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

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

Note que, em

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

- O tamanho amostral n **não** depende do tamanho populacional N ;
- O tamanho amostral depende:
 - do nível de confiança desejado (expresso pelo valor crítico $z_{\gamma/2}$);
 - 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

Exemplo

Seja $X \sim N(\mu, 36)$

- 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%

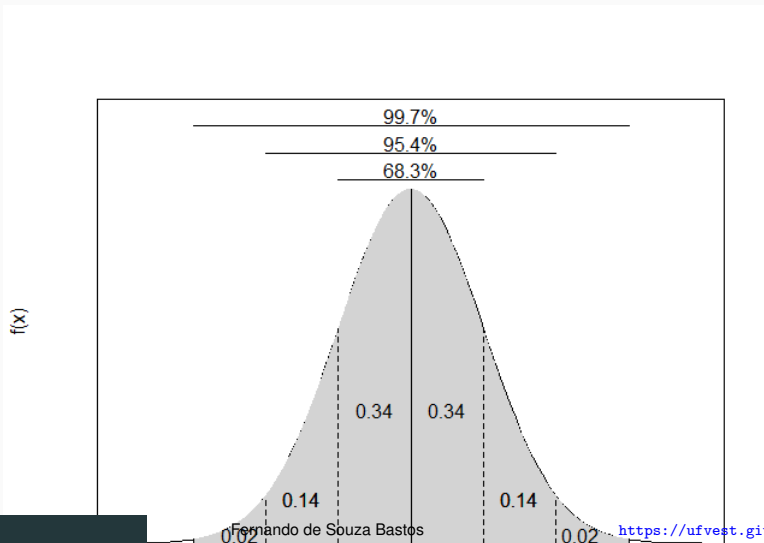
Determinação do tamanho amostral (σ desconhecido)

Determinação do tamanho amostral (σ desconhecido)

Se σ for desconhecido?

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

Regra empírica para uma distribuição normal



Regra empírica para uma distribuição normal

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

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




pode ser utilizado como uma estimativa para σ .

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

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

Referências

-  Bastos, Fernando de Souza (2025). ***Apostila Interativa***. Disponível online: <https://ufvest.shinyapps.io/ApostilaInterativa/>.
-  Ferreira, Eric Batista e Marcelo Silva de Oliveira (2020). ***Introdução à Estatística com R***. Editora Universidade Federal de Alfenas. URL: https://www.unifal-mg.edu.br/bibliotecas/wp-content/uploads/sites/125/2021/12/32-EBR_Unifal.pdf.
-  Meyer, Paul L (1982). ***Probabilidade: aplicações à estatística***. Livros Técnicos e Científicos.

Referências ii

-  Montgomery, D. C. e G. C Runger (2016). ***Estatística Aplicada E Probabilidade Para Engenheiros***. 6ª ed. São Paulo: Grupo Gen-LTC.
-  Morettin, P.A. e W.O Bussab (2023). ***Estatística básica***. 10ª ed. São Paulo: Editora Saraiva.
-  Peternelli, Luiz Alexandre (s.d.). ***Apostila (EST 106)***. Formato slide Disponível no PVANet - Moodle.