Intervalos de Confiança para a Média

ESTAT0078 – Inferência I

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

http://sadraquelucena.github.io/inferencia1

UNIVERSIDADE FEDERAL DE SERGIPE

Introdução

- Na aula anterior, definimos o conceito de Quantidade Pivotal: uma função $Q(\mathbf{X}, \theta)$ que combina dados e o parâmetro, mas cuja distribuição não depende de θ.
- Agora, utilizaremos essa ferramenta para construir Intervalos de Confiança (IC) para a média (μ).
- O objetivo é identificar qual quantidade pivotal utilizar em cada cenário, garantindo que nossa estimativa tenha um fundamento probabilístico sólido.

Cenário 1: População Original Normal

- Se a variável aleatória X na população segue uma distribuição Normal, $X \sim N(\mu, \sigma^2)$, então a média amostral $\overline{X}$ será exatamente Normal para qualquer tamanho de amostra n:

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

- **Vantagem:** Podemos trabalhar com amostras pequenas $(n < 30)$ sem perder a validade estatística.
- **Resultado:** O intervalo de confiança construído aqui é chamado de **exato**.

Cenário 2: Teorema Central do Limite (TCL)

- Se a distribuição da população for desconhecida ou não-normal, recorremos ao TCL.
- **Princípio:** À medida que o tamanho da amostra (n) cresce, a distribuição de $\overline{X}$ se aproxima de uma Normal, não importa o formato original dos dados.
- **Nota sobre Assimetria:** Se os dados originais forem muito assimétricos, precisaremos de um n maior que 30 para que a aproximação seja segura.
- **Resultado:** O intervalo de confiança construído aqui é chamado de assintótico (ou aproximado).

Formalização do TCL

- Se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de uma população com média μ e variância σ^2 finita, então:

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1) \quad \text{quando } n \to \infty$$

- **Implicação Prática:** O TCL é o que permite fazer estatística no "mundo real", onde raramente sabemos a distribuição exata dos dados. Ele garante que, com dados suficientes, a média sempre se proxima da Normal.

Como saber se os dados são Normais?

- **Métodos visuais:** histograma e qq-plot.
- **Testes Estatísticos:** Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, etc.
- **Medidas Descritivas:** Verifique se Média ≈ Mediana e se os valores de Assimetria e Curtose estão próximos de zero.

Cenário 3: Amostras Pequenas e Não-Normais

- E se os dados não forem normais e a amostra for pequena ($n < 30$)?
- Neste caso, os métodos que veremos hoje não são confiáveis.
- A média $\overline{X}$ não terá comportamento normal e os níveis de confiança (ex: 95%) serão falsos.
- **Solução:** Nestes casos, entra-se no campo da Estatística Não-Paramétrica.

Decidindo a Quantidade Pivotal

- Uma vez garantida a normalidade de $\overline{X}$ (via cenário 1 ou 2), a escolha da quantidade pivotal depende do nosso conhecimento sobre a dispersão da população (σ^2):
  - **Variância (σ^2) Conhecida:**
    - Situação rara (ex: processos industriais muito estáveis).
    - Usamos a distribuição Normal Padrão (Z).
  - **Variância (σ^2) Desconhecida:**
    - Situação padrão na prática.
    - Precisamos estimar a variância usando S^2.
    - Usamos a distribuição t de Student.

Caso 1: Variância (σ^2) Conhecida

- Quando a variância σ^2 é uma constante conhecida (ex: processos industriais padronizados), utilizamos a Distribuição Normal Padrão como base para a quantidade pivotal:

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

- **O Intervalo de Confiança (IC)** então é

$$IC(\mu; 1 - \alpha) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

UNIVERSIDADE FEDERAL DE SERGIPE

Caso 2: Variância (σ^2) Desconhecida

- Na prática, σ^2 é raramente conhecido, sendo necessário estimá-lo através da variância amostral S^2.
- Ao substituir σ por S, a quantidade pivotal assume uma nova distribuição:

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

- onde t_{n-1} representa a Distribuição t de Student com $n - 1$ graus de liberdade.

Caso 2: Variância (σ^2) Desconhecida

- **O Intervalo de Confiança (IC)** então é

$$IC(\mu; 1-\alpha) = \left[\bar{x} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\ ;\ \bar{x} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

A Distribuição t de Student

Desenvolvida por William Gosset (pseudônimo *Student*) em 1908.

- **Características:**
  - Simétrica em relação à média (zero) e em forma de sino.
  - Possui **caudas mais largas**, refletindo a incerteza de não conhecer σ.
  - O parâmetro é chamado de **graus de liberdade** ($\nu = n - 1$).
  - Conforme $n \to \infty$, a distribuição t converge para a $N(0, 1)$.
    - Na prática, para $n > 30$, as distribuições são muito parecidas, mas com o poder computacional atual, usamos t para qualquer n.

Decisão entre as Distribuições Z e t

- A escolha do modelo probabilístico depende do conhecimento dos parâmetros populacionais e do tamanho amostral:

Condição	Distribuição	Quantidade Pivotal
σ conhecido, Pop. Normal ou $n \geq 30$	Normal (Z)	$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$
σ desconhecido, Pop. Normal ou $n \geq 30$	t-Student (t)	$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$

Exemplo 17.1

Uma indústria de laticínios produz embalagens de manteiga cujo peso segue uma distribuição normal. O controle de qualidade afirma que o desvio padrão histórico é de $10g$. Uma amostra de 25 embalagens revelou um peso médio de $505g$. Construa um IC de 95% para o peso médio real (μ).

Exemplo 17.2

Um atuário analisa o custo de sinistros de um novo produto de seguro. Como o produto é novo, não se conhece a variância populacional. Uma amostra aleatória de 9 processos resultou em uma média de $R\$1.200,00$ e um desvio padrão de $R\$300,00$. Assumindo normalidade, determine o IC de 90% para a média dos sinistros.

Exemplo 17.3

Uma operadora de cartões de crédito deseja estimar o gasto médio mensal em lazer dos seus usuários. Sabe-se que a distribuição de gastos costuma ser muito assimétrica (muitas pessoas gastam pouco, poucas pessoas gastam muito). Uma amostra aleatória de 100 usuários apresentou uma média de $R\$450,00$ e um desvio padrão de $R\$120,00$. Construa um IC de 99% para a média populacional μ.

Interpretação Frequentista do IC

- O parâmetro (μ) é uma constante fixa, porém desconhecida.
- O intervalo $[L(\mathbf{X}), U(\mathbf{X})]$ é uma variável aleatória, pois seus limites dependem da amostra.
- A Confiança $(1 - \alpha)$ é a taxa de sucesso do método no longo prazo.

Checklist: Posso usar o IC para a Média?

- Para usar as fórmulas desta aula, você precisa atender a pelo menos um desses critérios:

  1. **A população original é Normal?** (Se sim, siga em frente independente do n).
  2. **A amostra é grande ($n \geq 30$)?** (Se sim, o TCL garante a normalidade da média).

- **E se não atender nenhum dos dois?** Se a amostra for pequena e a população não for normal, o método clássico falha. Entramos no campo da **Estatística Não-Paramétrica**.

Aplicação no R

No R (pacote `DescTools`), a função `MeanCI` automatiza esse processo:

```r
# Exemplo no R
library(DescTools)
dados <- c(102, 98, 105, 99, 101) # Pequena amostra (n=5)
                                  # vamos supor normalidade
MeanCI(dados, conf.level = 0.95)
```

```
     mean    lwr.ci    upr.ci
101.00000  97.59956 104.40044
```

Fim