Intervalos de Confiança para a Variância e para a Proporção

ESTAT0078 – Inferência I

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

http://sadraquelucena.github.io/inferencia1

UNIVERSIDADE FEDERAL DE SERGIPE

Introdução

- Nas aulas passadas, focamos em estimar **onde** os dados estão centrados (μ).
- **Hoje:** Vamos estimar **quanto** os dados variam (σ^2) e a **frequência** de um evento (p).
- **Atenção Atuários e Estatísticos:**
  - Saber a média de um sinistro (μ) define o preço puro do seguro.
  - Saber a variância (σ^2) define o **Risco de Ruína**. Uma variância alta exige reservas muito maiores.
  - *Não basta saber a média; precisamos medir a imprevisibilidade de um fenômeno.*

Parte 1: Intervalo de Confiança para a Variância (σ^2)

Para construir um IC para a variância populacional, assumimos que a população original é **Normal**.

- **A Quantidade Pivotal:** Utilizamos a relação entre a variância amostral (S^2) e a populacional (σ^2):

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

- **Essa quantidade segue uma distribuição Qui-quadrado com $n-1$ graus de liberdade.**

A Distribuição Qui-quadrado (χ^2)

Diferente da Normal e da t-Student, a χ^2: 1. **Não é simétrica:** Ela é “torta” para a direita. 2. **Só assume valores positivos:** Variância nunca é negativa. 3. **Consequência para o IC:** Os valores críticos da tabela não são iguais ($\pm z$). Precisamos olhar dois valores distintos: χ^2_{inf} e χ^2_{sup}.

Construção do IC para σ^2

Partimos da probabilidade:

$$P(\chi^2_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}) = 1 - \alpha$$

Ao isolar σ^2 (invertendo as frações), obtemos:

$$IC(\sigma^2; 1-\alpha) = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \; ; \; \frac{(n-1)S^2}{\chi^2_{\alpha/2}} \right]$$

- **Detalhe Crítico:** O valor *maior* da tabela fica no denominador do limite *inferior*, e vice-versa.

Atenção: A Fragilidade do IC para σ^2

O método da Qui-quadrado **não é robusto**. Isso significa que ele é muito sensível à suposição de normalidade.

- Diferente da média (onde o TCL ajuda em grandes amostras), para a variância, a **curtose** (peso das caudas) distorce gravemente a cobertura do intervalo.
- **Se os dados não forem normais:**
  1. O nível de confiança real pode ser muito menor que 95% (o intervalo fica "mentiroso").
  2. **Solução:** Utilizamos métodos computacionais de reamostragem, como o **Bootstrap**.

> **Regra de Ouro:** Só use a fórmula da χ^2 se o teste de normalidade (ex.: Shapiro-Wilk) e o QQ-Plot confirmarem que

os dados são normais.

Exemplo 18.1

Um analista de risco monitorou o retorno diário de uma ação por 20 dias e encontrou uma variância amostral igual a 0,04. Construa um IC de 95% para a variância real σ^2.

Exemplo 18.2

Uma farmacêutica precisa garantir que a variabilidade na dosagem de um comprimido seja mínima. Uma amostra de 15 comprimidos revelou uma variância amostral de 4 mg^2. Supondo que a dosagem segue uma distribuição Normal, construa um IC de 95% para a variância populacional (σ^2).

Parte 2: Intervalo para a Proporção (p)

- Agora mudamos o foco de variáveis contínuas (dinheiro, tempo) para variáveis binárias (Sucesso/Fracasso).
- **Estimador Pontual:** $\widehat{p} = \frac{X}{n}$ (número de sucessos / total).
- A distribuição exata é a Binomial, que é complexa para calcular intervalos.
- Usamos como solução para grandes amostras o Teorema Central do Limite (TCL).

Justificativa Assintótica

- Para n suficientemente grande, a distribuição de $\hat{p}$ se aproxima de uma Normal:

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

- A quantidade pivotal aproximada é: $Z = \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0, 1)$
- Isolando p, chegamos ao **Intervalo de Wald**:

$$IC(p) = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

UNIVERSIDADE
FEDERAL DE
SERGIPE

Validade do IC para p

- O Método de Wald é uma **aproximação**. Para usá-lo, devemos garantir que a normalidade "pegou".
- **Regra Prática:** o intervalo é confiável se o número esperado de sucessos e fracassos for pelo menos 5 (alguns autores usam 10):
  - $n \cdot \hat{p} \geq 5$
  - $n \cdot (1 - \hat{p}) \geq 5$

> **Caso contrário:** Devemos usar métodos "Exatos" (como Clopper-Pearson), que o R calcula automaticamente se pedirmos.

Exemplo 18.3

Em uma pesquisa com 400 consumidores, 160 afirmaram preferir a marca A. Construa um IC de 95% para a proporção real de preferência.

Exemplo 18.4

Uma seguradora deseja estimar a taxa de renovação de sua carteira de automóveis para projetar o fluxo de caixa do próximo ano. Em uma amostra aleatória de 500 apólices que venceram no mês passado, 425 foram renovadas. Construa um IC de 90% para a verdadeira proporção de renovação (p).

Aplicação no R: `VarCI`

- Para a variância, utilizamos a função `VarCI` do pacote `DescTools`.

```
library(DescTools)

# Dados do Exemplo 18.2
# Supondo que temos os dados brutos ou simulando com s^2=4
set.seed(123)
dosagem <- rnorm(n = 15, mean = 50, sd = 2) # sd=2 -> var=4

# Cálculo do IC para a Variância
VarCI(dosagem, conf.level = 0.95, method = "classic")
```

```
     var   lwr.ci   upr.ci
2.858395 1.532126 7.109517
```

```
# Se você quiser o IC para o Desvio Padrão, basta usar
# o argumento sd = TRUE dentro da mesma função (ou tirar
# a raiz quadrada dos limites manualmente).
# Se os dado não forem normais, use method = "boot"
```

Aplicação no R: `BinomCI`

- Para proporções, o pacote `DescTools` oferece a função `BinomCI`.

```
library(DescTools)

# x = sucessos, n = total
# method = "wald" (para grandes amostras)
BinomCI(x = 160, n = 400, conf.level = 0.95, method = "wald")
```

```
      est    lwr.ci    upr.ci
[1,] 0.4 0.3519909 0.4480091
```

```
# Se a amostra for pequena, use na função
# method = "clopper-pearson"
```

Resumo

Parâmetro	Cenário	O que usar	Ação / Fórmula
σ^2	População **Normal**	**Qui-quadrado** (χ^2)	$\left[\frac{(n-1)S^2}{\chi^2_{\text{sup}}}; \frac{(n-1)S^2}{\chi^2_{\text{inf}}}\right]$
σ^2	População **Não-Normal**	**Bootstrap** (Reamostragem)	Usar métodos computacionais.
p	Grandes Amostras ($n\hat{p} \geq 5$)	**Normal** (Z) [Wald]	$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
p	Pequenas Amostras	**Exato** [Clopper-Pearson]	Usar função do R.

Fim