ACH2053 – Introdução à Estatística

Aula 13: Intervalo de Confiança

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2025

Intervalo de Confiança

Seja $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição que depende do parâmetro θ e $g(\theta)$ uma função real de θ. Se $A \leq B$ são duas estatísticas que apresentam a seguinte propriedade para todos valores de θ:

$$\Pr(A < g(\theta) < B) \geq \gamma,$$

então o intervalo aleatório (A, B) é chamado de intervalo de confiança para $g(\theta)$ com coeficiente γ. Comumente especifica-se $\gamma = 1 - \alpha$, onde α é chamado de nível de significância.

Intervalo de Confiança - Resumo

Se $X_1, \ldots, X_n$ seguem distribuição normal, temos:

- σ conhecido, mas μ desconhecido

$$\mu = \bar{x}_n \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- σ e μ desconhecido

$$\mu = \bar{x}_n \pm t_{n-1,\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

- σ e μ desconhecido

$$\sigma^2 \in \left[\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} \right]$$

Intervalo de Confiança

Depois que os valores das variáveis aleatórias $a \sim A$ e $b \sim B$ são calculados, o intervalo (a, b) é o intervalo observado do intervalo de confiança.

Note que: $\Pr(A < g(\theta) < B) \in [0, 1]$, enquanto $\Pr(a < g(\theta) < b) \in \{0, 1\}$.

Exemplo

Considere n amostras de uma distribuição de Bernoulli com chance p. Considere também as estatísticas $A = \bar{X} - \frac{1}{\sqrt{n}}$ e $B = \bar{X} + \frac{1}{\sqrt{n}}$. Calcule:

1. Calcule $\Pr(A < p < B)$, quando $p = 1.0$ e $n = 25$.
2. Calcule $\Pr(A < p < B)$, quando $p = 0.1$ e $n = 25$.
3. Calcule $\Pr(A < p < B)$, quando $p = 0.5$ e $n = 25$.

Exemplo

Note que $\bar{X} = \frac{X}{n}$, onde X é uma variável aleatória binomial com parâmetros n e p, então:

$$\begin{aligned}\Pr(A < p < B) &= \Pr\left(\frac{X}{n} - \frac{1}{\sqrt{n}} < p < \frac{X}{n} + \frac{1}{\sqrt{n}}\right) \\ &= \Pr(np - \sqrt{n} < X < np + \sqrt{n}) \\ &= \sum_{np - \sqrt{n} < x < np + \sqrt{n}} \binom{n}{x} p^x (1-p)^{n-x}\end{aligned}$$

Para $p = 1.0$ temos que $\Pr(A < p < B) = 1.0$.
Para $p = 0.1$ temos que $\Pr(A < p < B) = 0.9977$.
Para $p = 0.5$ temos que $\Pr(A < p < B) = 0.9567$.

Pivotal

Seja $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição que depende do parâmetro θ.

Seja $V(\mathbf{X}, \theta)$ uma variável aleatória cuja distribuição é a mesma para todo θ. Então V é chamado de pivotal.

Exemplo: Seja $X_1, \ldots, X_n$ amostras i.i.d. de uma distribuição normal com esperança μ e desvio padrão σ. Então:

$$V(\mathbf{x}, \mu, \sigma) = \frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\sigma}$$

tem distribuição normal padrão.

Pivotal

Seja $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição que depende do parâmetro θ e $g(\theta)$ uma função real de θ. Suponha que exista um pivotal V e uma função $r(v, \mathbf{x})$ estritamente crescente em v tal que:

$$r(V(\mathbf{x}, \theta), \mathbf{x}) = g(\theta)$$

Seja G a c.d.f. de V (assuma que G é contínua) e $0 < \gamma < 1$. Escolha $\gamma_2 > \gamma_1$ tal que $\gamma_2 - \gamma_1 = \gamma$, então as seguintes estatísticas formam um intervalo de confiança para $g(\theta)$ com coeficiente γ:

$$A = r(G^{-1}(\gamma_1), \mathbf{X}),$$
$$B = r(G^{-1}(\gamma_2), \mathbf{X}).$$

Intervalo de Confiança - Resumo

Se $X_1, \ldots, X_n$ seguem distribuição normal, temos:

- σ conhecido, mas μ desconhecido

$$\mu = \bar{x}_n \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- σ e μ desconhecido

$$\mu = \bar{x}_n \pm t_{n-1,\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

- σ e μ desconhecido

$$\sigma^2 \in \left[\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} \right]$$

Exemplo

Se $\sigma = 1.5$ é conhecido e X segue a distribuição normal, crie um intervalo de confiança para a média μ com coeficiente $\gamma = 0.95$ para a amostra: $\mathbf{x} = (2.4, 1.6, 3.0, 1.8, 3.2)$.

1. Defina quem é $g(\theta)$.

$$g(\mu, \sigma) = \mu$$

2. Defina um pivotal V.

$$V(\mathbf{X}, \mu, \sigma) = \frac{\mu - \bar{X}_n}{\sigma/\sqrt{n}}$$

Nesse caso, temos que $V(\cdot)$ é uma distribuição normal padrão.

3. Encontre a função inversa $r(v, \mathbf{X})$.

$$r(v, \mathbf{x}) = \bar{x}_n + \frac{\sigma}{\sqrt{n}} v$$

Exemplo

4. Defina γ_1 e γ_2.
   Vamos escolher γ_1 e γ_2 tal que $1 - \gamma_2 = \gamma_1$, então temos:

   $$\gamma_2 - \gamma_1 = \gamma \Rightarrow 1 + \gamma = 2\gamma_2 \Rightarrow \gamma_2 = \frac{1+\gamma}{2} = 1 - \frac{\alpha}{2}$$

   e $\gamma_1 = \frac{1-\gamma}{2} = \frac{\alpha}{2}$. Logo, $\gamma_1 = 0.025$ e $\gamma_2 = 0.975$
5. Encontre $G^{-1}(\gamma_1)$ e $G^{-1}(\gamma_2)$.
   Buscando na tabela da distribuição normal padrão, temos:

   $$G^{-1}(\gamma_1) = -1.96 \text{ e } G^{-1}(\gamma_2) = 1.96$$

Exemplo

6. Se a função $r(v, \mathbf{x})$ é crescente em v, calcule $a = r(G^{-1}(\gamma_1), \mathbf{x})$ e $b = r(G^{-1}(\gamma_2), \mathbf{x})$.
   Então:

$$a = r(-1.96, \mathbf{x}) = \bar{x}_n + \frac{\sigma}{\sqrt{n}}(-1.96) = 2.4 - 1.3148 = 1.0852$$

$$b = r(1.96, \mathbf{x}) = \bar{x}_n + \frac{\sigma}{\sqrt{n}}(+1.96) = 2.4 + 1.3148 = 3.7148$$

Intervalo de Confiança - Resumo

Se $X_1, \ldots, X_n$ seguem distribuição normal, temos:

- σ conhecido, mas μ desconhecido

$$\mu = \bar{x}_n \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- σ e μ desconhecido

$$\mu = \bar{x}_n \pm t_{n-1,\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

- σ e μ desconhecido

$$\sigma^2 \in \left[\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} \right]$$

Função Gamma

Para cada número positivo α, a função gamma é definida pela integral:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Algumas propriedades da função gamma:

- Se $\alpha > 1$, então $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- Se n é inteiro, então $\Gamma(n) = (n - 1)!$.
- Fórmula de Stirling:

$$\lim_{x\to\infty} \frac{\sqrt{(2\pi)} x^{x-0,5} e^{-x}}{\Gamma(x)} = 1.$$

Distribuição Gamma

A função de probabilidade

$$f(x;\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{se } x > 0 \\ 0 & \text{caso contrário} \end{cases}$$

é chamada de distribuição Gamma com parâmetros $\alpha > 0$ e $\beta > 0$.

$$\mathsf{E}(X) = \frac{\alpha}{\beta}$$

$$\mathsf{Var}(X) = \frac{\alpha}{\beta^2}$$

Distribuição Gamma

A distribuição exponencial com parâmetro λ é igual a distribuição gamma com parâmetros $\alpha = 1$ e $\beta = \lambda$.

Suponha que chegadas ocorram de acordo com um processo de Poisson com taxa λ. Seja Z_k o tempo até a k-ésima chegada para $k = 1, 2, \ldots$. A distribuição de Z_k é a distribuição gamma com parâmetros $\alpha = k$ e $\beta = \lambda$.

Distribuição χ^2

Para cada número positivo m, a distribuição gamma com parâmetros $\alpha = m/2$ e $\beta = 1/2$ é chamada de distribuição χ^2 (chi quadrado) com m graus de liberdade e possui p.d.f.:

$$f(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-x/2}$$

$$\mathsf{E}(X) = m$$

$$\mathsf{Var}(X) = 2m$$

Seja X uma variável aleatória com distribuição normal padrão. Então a variável aleatória $Y = X^2$ tem distribuição χ^2 com um $(m = 1)$ grau de liberdade.

Distribuição χ^2

Se as variáveis aleatórias $X_1, \ldots, X_k$ são independentes e se X_i tem distribuição χ^2 com m_i graus de liberdade, então a soma $X_1 + \cdots + X_k$ tem distribuição χ^2 com $m_1 + \cdots + m_k$ graus de liberdade.

Suponha que $X_1, \ldots, X_n$ formam uma amostra aleatória da distribuição normal com média μ e variância σ^2. Então a média da amostra $\overline{X}_n$ e a variância da amostra $(1/n)\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ são variáveis aleatórias independentes tal que:

1. $\overline{X}_n$ tem distribuição normal com média μ e variância σ^2/n, e
2. $\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2}$ tem distribuição χ^2 com $n-1$ graus de

Exemplo

Se X segue a distribuição normal, crie um intervalo de confiança para a variância σ^2 com coeficiente $\gamma = 0.95$ para a amostra: $\mathbf{x} = (2.4, 1.6, 3.0, 1.8, 3.2)$.

1. Defina quem é $g(\theta)$.

$$g(\mu, \sigma) = \sigma^2$$

2. Defina um pivotal V.

$$V(\mathbf{X}, \mu, \sigma) = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2}$$

Nesse caso, temos que $V(\cdot)$ é uma distribuição χ^2 com $n-1$ graus de liberdade.

Exemplo

3. Encontre a função inversa $r(v, \mathbf{X})$.

$$r(v, \mathbf{x}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}{v}$$

4. Defina γ_1 e γ_2.
   Vamos escolher $\gamma_1 = 0.025$ e $\gamma_2 = 0.975$.
5. Encontre $G^{-1}(\gamma_1)$ e $G^{-1}(\gamma_2)$.
   Buscando na tabela da distribuição χ^2, temos:

$$G^{-1}(\gamma_1) = 0.484 \text{ e } G^{-1}(\gamma_2) = 11.143$$

6. Se a função $r(v, \mathbf{x})$ é crescente em v, calcule $a = r(G^{-1}(\gamma_1), \mathbf{x})$ e $b = r(G^{-1}(\gamma_2), \mathbf{x})$.
   Como $r(\cdot)$ não é crescente em v, temos que:

$$a = r(11.143, \mathbf{x}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}{11.143} = \frac{2}{11.143} = 0.179$$

$$b = r(0.484, \mathbf{x}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}{0.484} = \frac{2}{0.484} = 4.132$$

Intervalo de Confiança - Resumo

Se $X_1, \ldots, X_n$ seguem distribuição normal, temos:

- σ conhecido, mas μ desconhecido

$$\mu = \bar{x}_n \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- σ e μ desconhecido

$$\mu = \bar{x}_n \pm t_{n-1,\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

- σ e μ desconhecido

$$\sigma^2 \in \left[\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} \right]$$

Distribuição t de Student

Considere duas variáveis independentes Y e Z, tal que Y tem distribuição χ^2 com m graus de liberdade e Z tem a distribuição normal padrão. Suponha que a variável aleatória X é definida pela equação:

$$X = \frac{Z}{\left(\frac{Y}{m}\right)^{1/2}},$$

então a distribuição de X é chamada de distribuição t de Student com m graus de liberdade.

Distribuição t de Student

Considere que X segue a distribuição t de Student com m graus de liberdade, então X possui p.d.f.:

$$f(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}.$$

$$\mathsf{E}(X) = 0$$

$$\mathsf{Var}(X) = \frac{m}{m-2}, \text{ para } m > 2$$

Distribuição t de Student

Suponha que $X_1, \ldots, X_n$ são amostras aleatórias da distribuição normal com média μ e variância σ^2. Seja $\overline{X}_n$ a média da amostra e defina:

$$\sigma' = \left[\frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{n-1} \right]^{\frac{1}{2}}.$$

Então $\sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma'}$ tem a distribuição t de Student com $n-1$ graus de liberdade.

Exemplo

Se X segue a distribuição normal, crie um intervalo de confiança para a média μ com coeficiente $\gamma = 0.95$ para a amostra: $\mathbf{x} = (2.4, 1.6, 3.0, 1.8, 3.2)$.

1. Defina quem é $g(\theta)$.

$$g(\mu, \sigma) = \mu$$

2. Defina um pivotal V.

$$V(\mathbf{X}, \mu, \sigma) = \sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma'}$$

Nesse caso, temos que $V(\cdot)$ é uma distribuição t de Student com $n-1$ graus de liberdade.

Exemplo

3. Encontre a função inversa $r(v, \mathbf{X})$.

$$r(v, \mathbf{x}) = \bar{x}_n - \frac{\sigma'}{\sqrt{n}} v$$

4. Defina γ_1 e γ_2.
   Vamos escolher $\gamma_1 = 0.025$ e $\gamma_2 = 0.975$.
5. Encontre $G^{-1}(\gamma_1)$ e $G^{-1}(\gamma_2)$.
   Buscando na tabela da distribuição t de Student, temos:

$$G^{-1}(\gamma_1) = -2.776 \text{ e } G^{-1}(\gamma_2) = 2.776$$

Exemplo

6. Se a função $r(v, \mathbf{x})$ é crescente em v, calcule $a = r(G^{-1}(\gamma_1), \mathbf{x})$ e $b = r(G^{-1}(\gamma_2), \mathbf{x})$.
Como $r(\cdot)$ não é crescente em v, temos que:

$$a = r(2.776, \mathbf{x}) = \bar{x}_n - \frac{\sigma'}{\sqrt{n}}(2.776) = 2.4 - 2.776\frac{\sqrt{0.5}}{\sqrt{5}} = 1.522$$

$$b = r(-2.776, \mathbf{x}) = \bar{x}_n - \frac{\sigma'}{\sqrt{n}}(-2.776) = 2.4 + 2.776\frac{\sqrt{0.5}}{\sqrt{5}} = 3.278$$

Intervalos em Distribuições Exponencial
Suponha que $X_1, \ldots, X_n$ são amostras aleatórias da distribuição exponencial com taxa λ. Defina a estatística $T = \sum_{i=1}^{n} X_i$. Então λT tem a distribuição gamma com $\alpha = n$ e $\beta = 1$.

Pivotais para Outras Distribuições

Pivotais Assintóticos

Lembre-se do Teorema Central do Limite:

Teorema (Teorema do Limite Central)

Se as variáveis aleatórias $X_1, \ldots, X_n$ formam uma amostra aleatória de tamanho n de uma distribuição com média μ e desvio padrão σ $(0 < \sigma < \infty)$, então para cada x

$$\lim_{n\to\infty} \Pr\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x),$$

onde $\Phi(x)$ denota a c.d.f. da distribuição normal padrão.

No limite, pode-se sempre obter um pivotal aproximado para a média μ.

Exemplo

Se X segue a distribuição de Bernoulli, crie um intervalo de confiança para a taxa de sucesso p com coeficiente $\gamma = 0.95$ para uma amostra com $n = 100$ tentativas e 40 sucessos.

Temos que:

$$\bar{x} = \frac{40}{100} = 0.4 \text{ e } \sigma' = \sqrt{\frac{40 \times 0.36 + 60 \times 0.16}{100 - 1}} = 0.492$$

Então:

$$a = \bar{x}_n - \frac{\sigma'}{\sqrt{n}}(2.0) = 0.4 - 2.0\frac{0.492}{10} = 0.3016$$

$$b = \bar{x}_n - \frac{\sigma'}{\sqrt{n}}(-2.0) = 0.4 + 2.0\frac{0.492}{10} = 0.4984$$

Intervalos de Confiança a partir de Testes com Nível de significância
Seja $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição que depende do parâmetro θ e $g(\theta)$ uma função real de θ. Suponha que para cada valor possível g_0 de $g(\theta)$, existe um teste δ_{g_0} com nível de significância α_0 para as hipóteses:

$$H_{0,g_0} : g(\theta) = g_0 \text{ e } H_{1,g_0} : g(\theta) \neq g_0.$$

Para cada valor $\mathbf{x}$ de $\mathbf{X}$, defina:

$$w(\mathbf{x}) = \{g_0 : \delta_{g_0} \text{ não rejeita } H_{0,g_0} \text{ se } \mathbf{X} = \mathbf{x} \text{ é observado}\}.$$

Seja $\gamma = 1 - \alpha_0$. Então, o conjunto aleatório $w(\mathbf{X})$ satisfaz

$$\Pr[g(\theta_0) \in w(\mathbf{X}) | \theta = \theta_0] \geq \gamma,$$

para todo $\theta_0 \in \Omega$.

Testes com Nível de significância a partir de Intervalos de Confiança
Seja $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição que depende do parâmetro θ e $g(\theta)$ uma função real de θ. Suponha que para cada valor possível $\mathbf{x}$ de $\mathbf{X}$, exista um conjunto de confiança (intervalo de confiança) $w(\mathbf{X})$ com confiança γ. Para cada possível valor g_0 de $g(\theta)$ construa o teste δ_{g_0} para as hipóteses:

$$H_{0,g_0} : g(\theta) = g_0 \text{ e } H_{1,g_0} : g(\theta) \neq g_0,$$

no qual H_{0,g_0} não é rejeitada se e somente se $g_0 \in w(\mathbf{X})$. Então δ_{g_0} é um teste com nível de confiança $\alpha_0 = 1 - \gamma$.

Resumo do Curso

1. Experimento, resultado e espaço amostral
2. Probabilidade
3. Probabilidade Condicional
4. Variáveis Aleatórias
5. Esperança e Variância
6. Distribuições Típicas
7. Distribuições Multivariadas
8. Estimadores: viés, variância e erro quadrático médio
9. Estimadores Pontuais
10. Teorema do Limite Central
11. Informação de Fisher
12. Teste de Hipótese
13. Intervalo de Confiança