



5. Estimação Intervalar

USP-ICMC-SME
2013

Estimação por intervalos

X_1, \dots, X_n é uma amostra aleatória de uma variável cuja distribuição depende do parâmetro θ .

Se $L(X_1, \dots, X_n)$ e $U(X_1, \dots, X_n)$ são duas funções tais que $L < U$ e $P(L \leq \theta \leq U) = 1 - \alpha$,

o intervalo $[L, U]$ é chamado de **intervalo de confiança (IC)** de $100(1-\alpha)\%$ para θ .

$100(1-\alpha)\%$ é o **coeficiente de confiança** do intervalo. Deve ser “alto”.

O coeficiente de confiança é escolhido (**90%**, **95%** e **99%** são comuns). Em seguida **calculamos** L e U .

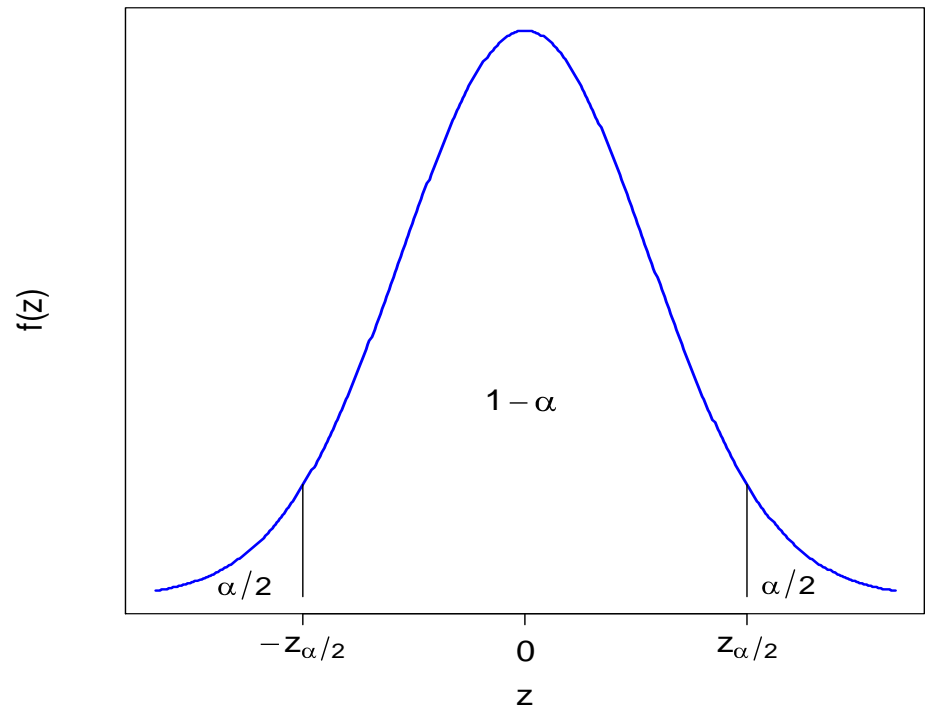
IC para uma média populacional

X_1, \dots, X_n é uma amostra aleatória de tamanho n de uma **população normal** com média μ (desconhecida) e variância σ^2 (**conhecida**). Vimos que a média amostral \bar{X} , tem distribuição **normal** com **média μ** e **variância σ^2/n** . Isto é,

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

Se a distribuição de X **não é** normal, o resultado acima é válido **aproximadamente**.

Logo, **fixando** um coeficiente de confiança **$(1-\alpha)$** , pode-se determinar $z_{\alpha/2}$ (consultando a tabela normal):



IC para uma média populacional

Sendo assim, $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$,

que equivale a $P(-z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z_{\alpha/2}) = 1 - \alpha$

$$\Leftrightarrow P\left(-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Logo, um **IC de 100 (1- α)%** para a média μ é dado por

$$[L; U] = \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = [\bar{X} - E; \bar{X} + E],$$

sendo que $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ é o erro máximo

e a **amplitude** do IC é $U - L = 2E$.

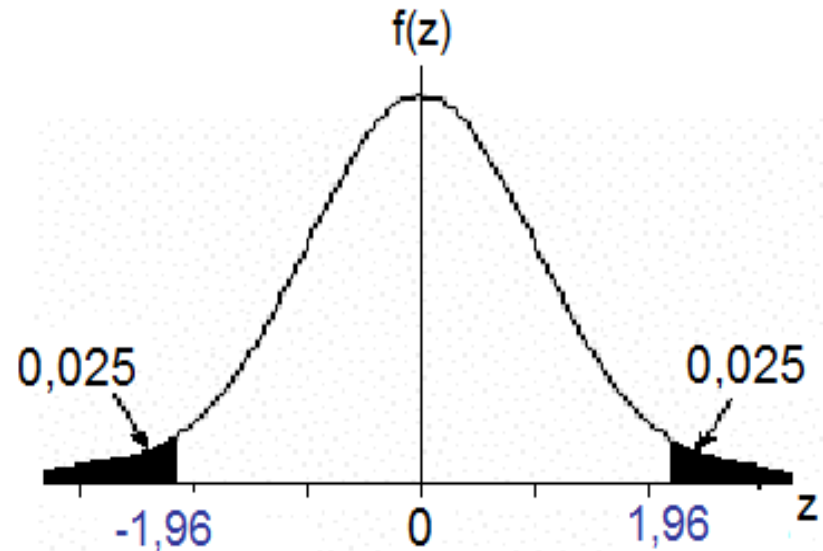
Exemplo

Em uma fábrica de cerveja a quantidade de cerveja em latas seguia uma distribuição normal com média 350 ml e desvio padrão 3 ml. Após alguns problemas na linha de produção, suspeita-se que houve alteração na média. Uma **amostra de 20** latas forneceu uma média de 346 ml. Obtenha um intervalo de 95% para a quantidade média de cerveja envasada supondo que não tenha ocorrido alteração na variabilidade.

Como $1-\alpha = 0,95$, temos da tabela normal padrão $z_{0,025} = 1,96$.

Obtemos IC = [L; U]

$$\begin{aligned} &= \left[\bar{X} - 1,96 \times \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \times \frac{\sigma}{\sqrt{n}} \right] \\ &= \left[346 - 1,96 \times \frac{3}{\sqrt{20}}; 346 + 1,96 \times \frac{3}{\sqrt{20}} \right] = [346 - 1,31; 346 + 1,31] = [344,69; 347,31], \text{ em ml.} \end{aligned}$$



Determinação do tamanho da amostra para estimação de μ

Erro máximo (E) na estimação de μ :
$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

$z_{\alpha/2}$ é obtido da tabela normal após a escolha do coeficiente de confiança ($1 - \alpha$).

(a) Especificamos o erro máximo. Se o desvio padrão (σ) for conhecido, podemos calcular n:

$$n = \frac{z_{\alpha/2}^2 \times \sigma^2}{E^2}.$$

(b) Especificamos o erro máximo. Se o desvio padrão (σ) não for conhecido, podemos utilizar o desvio padrão obtido de uma amostra piloto com n_0 observações:

$$n \cong \frac{z_{\alpha/2}^2 \times s_0^2}{E^2}, \text{ sendo que } s_0^2 \text{ é a variância amostral da amostra piloto.}$$

(c) Especificamos o erro máximo em função do desvio padrão como $E = k \sigma$:

$$n = \frac{z_{\alpha/2}^2}{k^2}.$$

Exemplo

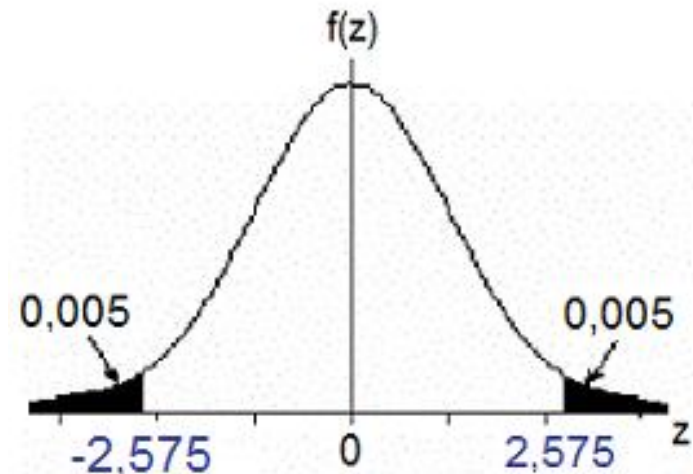
Em uma siderúrgica estuda-se a resistência média de barras de aço utilizadas na construção civil. Qual o tamanho amostral necessário para garantir que um erro máximo de 8 kg seja superado com probabilidade igual a 0,01? O desvio padrão da resistência para este tipo de barra é de 25 kg.

Solução. Do enunciado tem-se $\sigma = 25$ kg, $E = 8$ kg e

$1 - P(\bar{X} - E \leq \mu \leq \bar{X} + E) = 0,01 \Rightarrow P(\bar{X} - E \leq \mu \leq \bar{X} + E) = 1 - 0,01$,
ou seja, $\alpha = 0,01$ (o coeficiente de confiança do IC é $1 - \alpha = 99\%$).

Consultando a tabela normal encontramos $z_{\alpha/2} = 2,575$.

$$\begin{aligned} \text{Portanto, } n &= \frac{z_{\alpha/2}^2 \times \sigma^2}{E^2} \\ &= \frac{2,575^2 \times 25^2}{8^2} = 65. \end{aligned}$$



IC para uma média populacional (σ desconhecido)

Se a variável de interesse (X) tem **distribuição normal**, então

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{s} \sim t_{n-1}, \quad : \text{distribuição t de Student com } n - 1 \text{ g.l.,}$$

sendo que s é o desvio padrão amostral.

Se a distribuição de X **não é** normal, o resultado acima é válido **aproximadamente**.

Um IC de $100(1-\alpha)\%$ para μ é dado por

$$[L;U] = [\bar{X} - E; \bar{X} + E], \quad \text{em que } E = t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}.$$

IC para a média e testes de hipóteses

O teste da hipótese $H_0: \mu = \mu_0$ contra $H_1: \mu \neq \mu_0$ a um nível de significância α pode ser efetuado utilizando um IC com coeficiente de confiança igual a $1 - \alpha$.

Construímos o IC de $100(1-\alpha)\%$ para μ , dado por

$$[L;U] = [\bar{X} - E; \bar{X} + E], \text{ em que } E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} : \sigma \text{ conhecido.}$$

$$\text{ou } E = t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} : \sigma \text{ desconhecido.}$$

Se $\mu_0 \notin \text{IC}$, rejeitamos H_0 ; **caso contrário**, não rejeitamos H_0 .

Exemplo (p. 28 em Montgomery *et al.*, 2004)

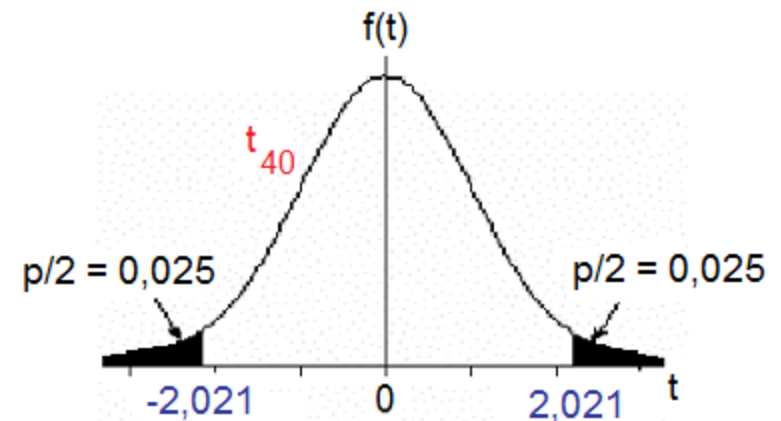
Foram coletados dados de viscosidade (em u.v.) de um líquido produzido em batelada. Resultados de 40 amostras encontram-se abaixo. Apresente um IC de 95% para a viscosidade média. Podemos concluir que a viscosidade média da população é igual a 15,5?

Dados: 13,3 14,5 15,3 15,3 14,3 14,8 15,2 14,5 14,6 14,1
14,3 16,1 13,1 15,5 12,6 14,6 14,3 15,4 15,2 16,8 14,9 13,7
15,2 14,5 15,3 15,6 15,8 13,3 14,1 15,4 15,2 15,2 15,9 16,5
14,8 15,1 17,0 14,9 14,8 14,0.

Solução. Inicialmente calculamos

$$\bar{X} = \frac{1}{40} \sum_{i=1}^{40} X_i = 14,875 \text{ e } s = \sqrt{\frac{1}{40-1} \sum_{i=1}^{40} (X_i - 14,875)^2} = 0,948.$$

Pelo enunciado, $n = 40$ e $1 - \alpha = 0,95$, de modo que $\alpha = 0,05$. Da tabela (Tábua III) com 40 g.l. (39 g.l. não estão na Tábua III) e $p = 5\%$, obtemos $t_{\alpha/2} = 2,021$.



Exemplo (p. 28 em Montgomery *et al.*, 2004)

Logo, o erro máximo (E) é igual

$$E = t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} = 2,021 \frac{0,948}{\sqrt{40}} = 0,303$$

e o IC de 95% para a média da população é dado por

$$[L; U] = [\bar{X} - E; \bar{X} + E] = [14,875 - 0,303; 14,875 + 0,303] = [14,57; 15,18].$$

Como $\mu_0 = 15,5 \notin \text{IC}$, rejeitamos $H_0: \mu = 15,5$ em favor de $H_1: \mu \neq 15,5$.

Conclusão. Com base nos dados coletados e adotando um nível de significância de 5%, não há evidência de que a viscosidade média da população seja igual a 15,5.

Exemplo. Solução em Excel

	A	B	C	D	E
1	13,3		Média =	14,9	
2	14,5		Desvio padrão =	0,9483	
3	15,3				
4	15,3		Confiança =	0,95	
5	14,3		n =	40	
6	14,8		$t_{\alpha/2}$ =	2,023	
7	15,2		Erro máximo =	0,303	
	14,5				
	14,6		Limites do	Inferior	Superior
	14,1		IC	14,57	15,18
	14,3				
	16,1				
	13,1				
	15,5				
	12,6				
	14,6				
	14,3				
	15,4				
	15,2				
	16,8				
	14,9				
	13,7				
	15,2				
	14,5				
	15,3				
	15,6				
	15,8				
	13,3				
	14,1				
	15,4				
	15,2				
	15,2				
	15,9				
	16,5				
	14,8				
	15,1				
	17,0				
	14,9				
39	14,8				
40	14,0				

Média: =MÉDIA(A:A)

Desvio padrão: =DESVPADA(A:A)

n: =CONT.SE(A:A; "> 0")

$t_{\alpha/2}$: =INVT(1 - \$D\$4; \$D\$5 - 1)

Erro máximo: =\$D\$6 * \$D\$2 / RAIZ(\$D\$5)

Limite inferior: =\$D\$1 - \$D\$7

Limite superior: =\$D\$1 + \$D\$7

IC para uma proporção populacional

Cada observação pode ser classificada como **sucesso** ($X = 1$) ou **insucesso** ($X = 0$) e a **probabilidade de sucesso** é p . Dispomos de uma amostra aleatória X_1, \dots, X_n . Vimos que

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

sendo que $\bar{p} = \frac{1}{n} \sum_{i=1}^n X_i$: proporção **amostral** de sucessos.

Para um nível confiança fixado em $100(1-\alpha)\%$, obtemos (veja lâmina 4)

$$P\left(\bar{p} - z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{p} + z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}\right) \cong 1 - \alpha.$$

IC para uma proporção populacional

(a) Abordagem otimista

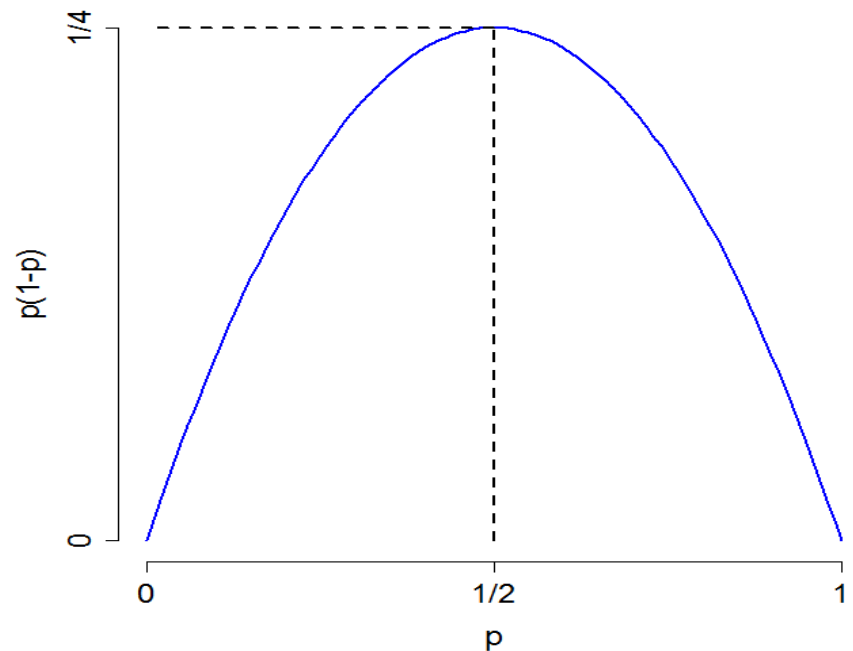
Substituir $p(1-p)$ por $\bar{p}(1-\bar{p})$:

$$\text{IC} \cong \left[\bar{p} - z_{\alpha/2} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{\alpha/2} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right].$$

(b) Abordagem conservativa

Substituir $p(1-p)$ por $\frac{1}{4}$, que corresponde ao valor máximo de $p(1-p)$.

$$\text{IC} \cong \left[\bar{p} - z_{\alpha/2} \times \frac{1}{\sqrt{4n}}; \bar{p} + z_{\alpha/2} \times \frac{1}{\sqrt{4n}} \right].$$



Exemplo

Um estudo foi realizado para determinar a **proporção** de componentes de um certo tipo que resistem durante um certo período a condições de uso mais rigorosas do que as especificadas. Em uma **amostra** de **200** componentes selecionados ao acaso, **160** resistiram. Apresente um intervalo de **95%** de confiança para a proporção de componentes que resistem.

Solução. Estimativa pontual de p : $\bar{p} = \frac{160}{200} = 0,8$ (80%).

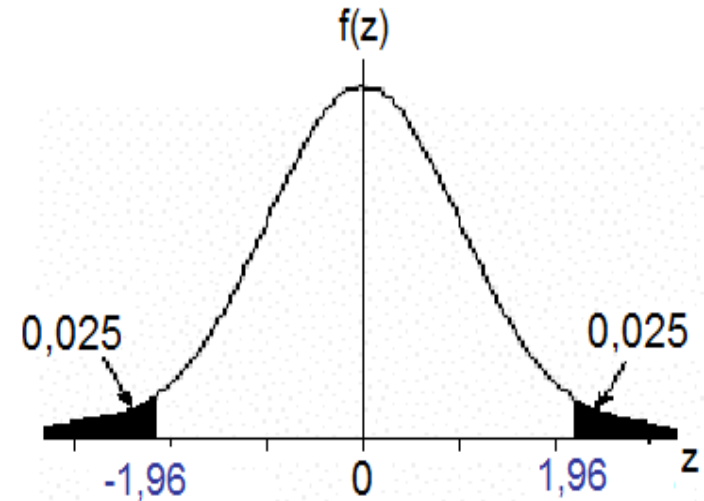
Como $1 - \alpha = 0,95$, obtemos da tabela normal padrão $z_{0,025} = 1,96$.

Abordagem **otimista**:

$$\text{IC} \cong \left[0,8 - 1,96 \times \sqrt{\frac{0,8(1-0,8)}{200}}; 0,8 + 1,96 \times \sqrt{\frac{0,8(1-0,8)}{200}} \right]$$
$$= [0,745; 0,855].$$

Abordagem **conservativa**:

$$\text{IC} \cong \left[0,8 - 1,96 \times \frac{1}{\sqrt{4 \times 200}}; 0,8 + 1,96 \times \frac{1}{\sqrt{4 \times 200}} \right] = [0,731; 0,869].$$



Determinação do tamanho da amostra para estimação de p

Erro máximo de estimação de p é fixado:

$$E = z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} \quad \Rightarrow \quad n = \frac{z_{\alpha/2}^2 \times p(1-p)}{E^2}.$$

(a) Há informação sobre p: p^* (estudos anteriores, especialistas, amostra piloto, etc):

$$n = \frac{z_{\alpha/2}^2 \times p^* (1-p^*)}{E^2}.$$

(b) Não há informação sobre p:

$p(1-p)$ é substituído pelo valor máximo, igual a $\frac{1}{4}$ (veja lâmina 15):

$$n = \frac{z_{\alpha/2}^2}{4E^2} = \left(\frac{z_{\alpha/2}}{2E} \right)^2.$$

Coeficiente de confiança de 95%: $\alpha = 5\%$, $z_{\alpha/2} = 1,96 \cong 2$ e

$$n \cong 1 / E^2.$$

Exemplo

Uma equipe pretende estimar a **proporção** de avarias ocorridas no transporte de um produto. Estudos **anteriores** indicam que esta proporção **não ultrapassa 20%**. Que tamanho de amostra é necessário para assegurar com uma **confiança** de **99%** que o **erro** de estimação desta proporção seja no **máximo** igual a **0,05**?

Solução. Do enunciado obtemos $p \leq 0,2$, $1 - \alpha = 0,99$ e $E = 0,05$. Da tabela normal padrão, $z_{0,005} = 2,575$.

Proteção em relação à situação mais desfavorável: $p^* = 0,20$.

Finalmente,

$$\begin{aligned} n &= \frac{z_{\alpha/2}^2 \times p^* (1 - p^*)}{E^2} \\ &= \frac{2,575^2 \times 0,2 \times (1 - 0,2)}{0,05^2} \\ &= 424,4 \Rightarrow n = 425. \end{aligned}$$

