Intervalos de Confiança (IC)

5.4

Introdução

Ao fazer a estimação pontual, ou seja, quando estimamos um parâmetro através de um único valor numérico, a principal restrição é que toda a informação presente nos dados é resumida através deste número. Assim, é importante também encontrar um **intervalo de valores plausíveis para o parâmetro** e, assim, termos ideia da precisão da estimativa encontrada.

Queremos construir um intervalo em torno da estimativa pontual de modo que ele tenha uma **probabilidade conhecida de conter o verdadeiro valor do parâmetro**.

Utilizaremos as **distribuições amostrais** de estimadores dos parâmetros desconhecidos.

Estimação por intervalos

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável cuja distribuição depende do parâmetro θ.

Se $L(X_1, X_2, \ldots, X_n)$ e $U(X_1, X_2, \ldots, X_n)$ 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.

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

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

Procedimento Geral

1. Obter uma estatística que depende de θ, V = G(X,θ), mas cuja distribuição não depende de θ.

2. Usando a distribuição de V, encontrar as constantes a e b tais que

$$P(a \leq V \leq b) \geq 1- \alpha.$$

3. Definir $\{\theta : a \leq G(x,\theta) \leq b\}$ como o intervalo (ou região) de confiança 100(1- α)% para θ.

Observações:

A desigualdade em **2.** é útil principalmente no caso de distribuições discretas onde nem sempre é possível satisfazer a igualdade (queremos que o intervalo seja o menor possível).

A variável aleatória V é denominada **quantidade pivotal** ou **pivot**. Idealmente ela deve ter distribuição conhecida.

Este intervalo não pode ser interpretado como um intervalo de probabilidade para o parâmetro θ já que a aleatoriedade presente é devida a amostra. Ou seja, o procedimento leva a construção de um intervalo probabilístico para V e não para θ.

Na prática, dizemos que 100(1- α)% de todos os intervalos de confiança que construímos utilizando essa metodologia conterão o verdadeiro valor do parâmetro. Por exemplo se 1- α = 0,99 então somente 1 a cada 100 intervalos não conterão θ, em média.

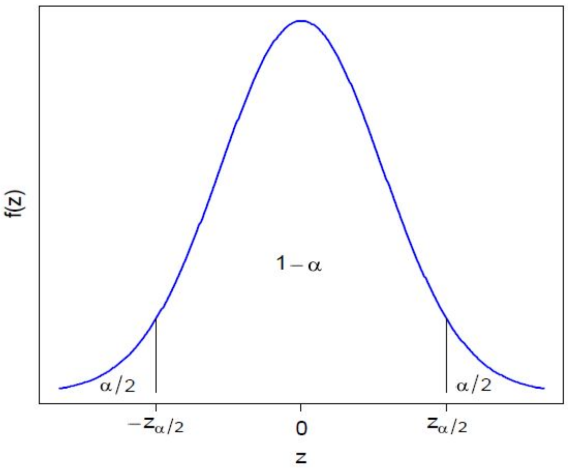
IC para uma média populacional (σ conhecido)

X_1, X_2, ..., 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 $\overline{X}$, tem distribuição **normal** com média μ e variância σ^2/n. Isto é,

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

Se a distribuição de X **não é** normal, o resultado acima é válido **aproximadamente** (para um tamanho de amostra grande).

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



Sendo assim,

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

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

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

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

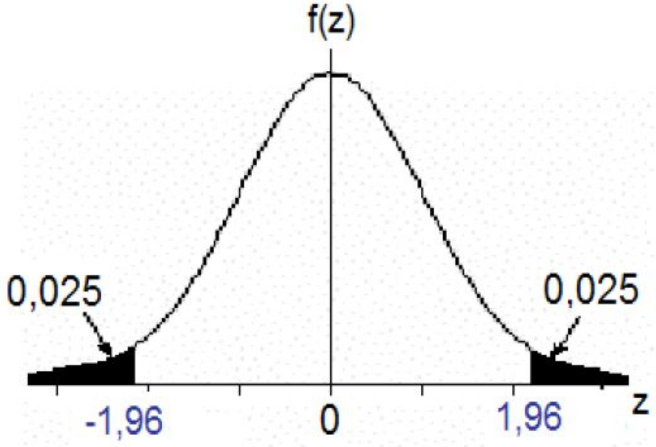
Logo, um **IC de 100(1- α)% para a média** μ é dado por

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

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.



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

Obtemos IC = [L; U] =

$$= \left[\bar{X} - 1{,}96 \times \frac{\sigma}{\sqrt{n}} ; \bar{X} + 1{,}96 \times \frac{\sigma}{\sqrt{n}} \right]$$

$$= \left[346 - 1{,}96 \times \frac{3}{\sqrt{20}} ; 346 + 1{,}96 \times \frac{3}{\sqrt{20}} \right] = [346 - 1{,}31 ; 346 + 1{,}31] = [344{,}69 ; 347{,}31], \text{ em ml.}$$

Determinação do tamanho da amostra para estimação de μ

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

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

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

$$\boxed{n = \frac{z^2_{\alpha/2} \times \sigma^2}{E^2}.}$$

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:

$\boxed{n \cong \dfrac{z^2_{\alpha/2} \times s_0^2}{E^2}}$, sendo que s_0^2 é a variância amostral da amostra piloto.

c. Especificamos o erro máximo em função do desvio padrão como **E = kσ**:

$$\boxed{n = \frac{z^2_{\alpha/2}}{k^2}.}$$

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 σ = 25 kg, E = 8 kg e

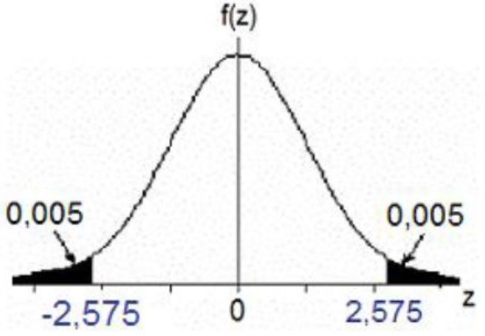
$$1 - \mathrm{P}(\overline{X} - E \leq \mu \leq \overline{X} + E) = 0{,}01 \Rightarrow \mathrm{P}(\overline{X} - E \leq \mu \leq \overline{X} + E) = 1 - 0{,}01,$$

ou seja, α = 0,01 (o coeficiente de confiança do IC é **1 – α = 99%**).

Consultando a tabela normal encontramos
$z_{0,005}$ = 2,575.

Portanto,

$$n = \frac{z^2{}_{\alpha/2} \times \sigma^2}{E^2} = \frac{2{,}575^2 \times 25^2}{8^2} = 65.$$



IC para uma média populacional (σ desconhecido)

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

$$T = \frac{\sqrt{n}(\overline{X} - \mu)}{s} \sim t_{n-1}$$; distribuição t de Student com n – 1 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 – α)% para** μ é dado por

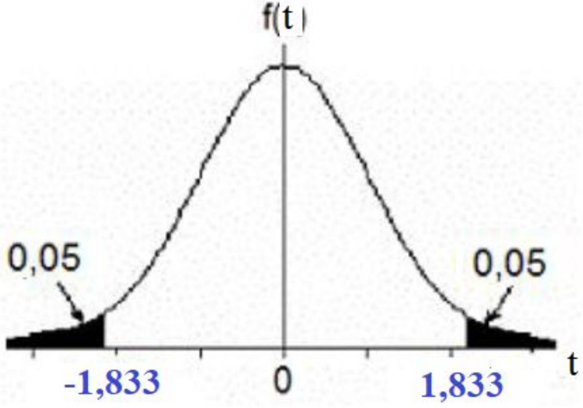
$$[L;U] = [\overline{X} - E; \overline{X} + E], \quad \text{em que } E = t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}.$$

Exemplo

A seguinte amostra foi extraída de uma população normal:

6, 6, 7, 8, 9, 9, 9, 10, 11, 12.

Obtenha um intervalo de 90% de confiança para μ.



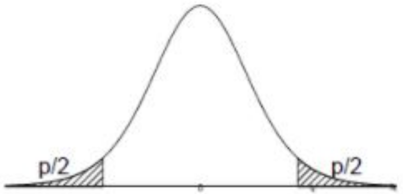
Solução. Como 1- α = 0,90 e g.l. = n-1= 9, temos da tabela da t: $t_{0,05;9}$ = 1,833. Obtemos da amostra uma média de 8,7 e desvio-padrão de 2,003

Assim, IC = [L; U] =

$$= \left[\overline{X} - 1{,}833 \times \frac{s}{\sqrt{n}} ; \overline{X} + 1{,}833 \times \frac{s}{\sqrt{n}} \right]$$

$$= \left[8{,}7 - 1{,}833 \times \frac{2{,}003}{\sqrt{10}} ;\ 8{,}7 + 1{,}833 \times \frac{2{,}003}{\sqrt{10}} \right] = \left[8{,}7 - 1{,}161 ;\ 8{,}7 - 1{,}161 \right] = \left[7{,}539 ; 9{,}861 \right]$$

Distribuição t de Student



	90%	80%	70%	60%	50%	40%	30%	20%	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%	0.5%	0.2%	0.1%
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	3.104	3.320	3.578	3.896	4.303	4.849	5.643	6.965	9.925	14.089	22.327	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	2.471	2.605	2.763	2.951	3.182	3.482	3.896	4.541	5.841	7.453	10.215	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.226	2.333	2.456	2.601	2.776	2.999	3.298	3.747	4.604	5.598	7.173	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.098	2.191	2.297	2.422	2.571	2.757	3.003	3.365	4.032	4.773	5.893	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.019	2.104	2.201	2.313	2.447	2.612	2.829	3.143	3.707	4.317	5.208	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	1.966	2.046	2.136	2.241	2.365	2.517	2.715	2.998	3.499	4.029	4.785	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	1.928	2.004	2.090	2.189	2.306	2.449	2.634	2.896	3.355	3.833	4.501	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	1.899	1.973	2.055	2.150	2.262	2.398	2.574	2.821	3.250	3.690	4.297	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	[illegible]	1.877	1.948	2.028	2.120	2.228	2.359	2.527	2.764	3.169	3.581	4.144	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	1.859	1.928	2.007	2.096	2.201	2.328	2.491	2.718	3.106	3.497	4.025	4.437