Aula 12 - Estimação de Parâmetros

Estimação

Avaliar características da população com base em informações da amostra

$$\Downarrow$$

Estimar os parâmetros

Mais utilizadas:

- média (μ)
- proporção (π)
- variância (σ^2)

Exemplos:

- produção média de determinada cultura;
- proporção média de área foliar atacada por uma praga;
- parâmetros estatísticos genéticos (variância genética, ambiental e fenotípica)...

- proporção de crescimento de um microorganismo;
- proporção de alimento estragado/vencido na prateleira;

Propriedades dos estimadores

Estimadores

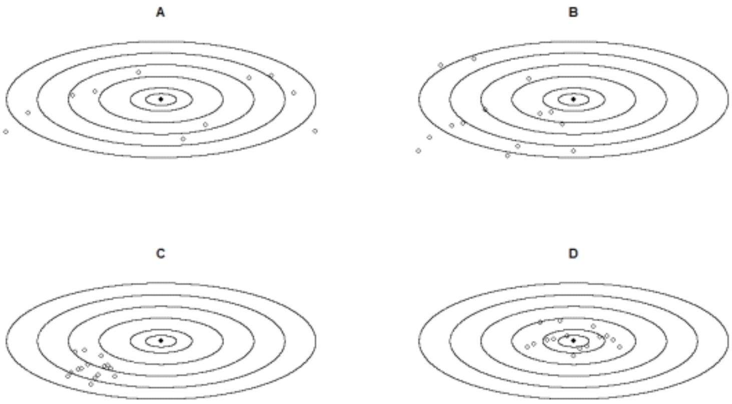
$$\text{Média: } \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\text{Proporção: } P = \frac{\text{número de sucessos}}{n}$$

$$\text{Variância: } S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Propriedades

- **não viesado**: média da distribuição amostral igual ao parâmetro
- **preciso**: variância amostral pequena
- **acurado**: erro amostral pequeno



A: não viesado, pouca precisão e pouca acurácia
B: viesado, pouca precisão e pouca acurácia
C: viesado, boa precisão e baixa acurácia
D: não viesado, boa precisão e boa acurácia

Estimativas pontuais e intervalares

Modelo probabilístico

$\Downarrow$

Estimar os parâmetros da distribuição

$\Downarrow$

Amostra

Estimadores $\Rightarrow$ **Estatísticas**

Estimativas pontuais

$$\text{Média: } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{Proporção: } p = \frac{\text{número de sucessos}}{n}$$

$$\text{Variância: } s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Estimativas intervalares

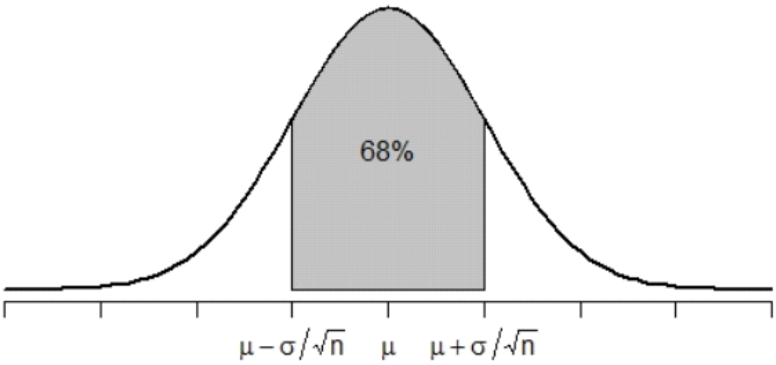
Intervalo de confiança

Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n de uma população e θ o parâmetro de interesse. Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ estatísticas tais que:

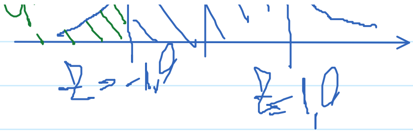
$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha.$$

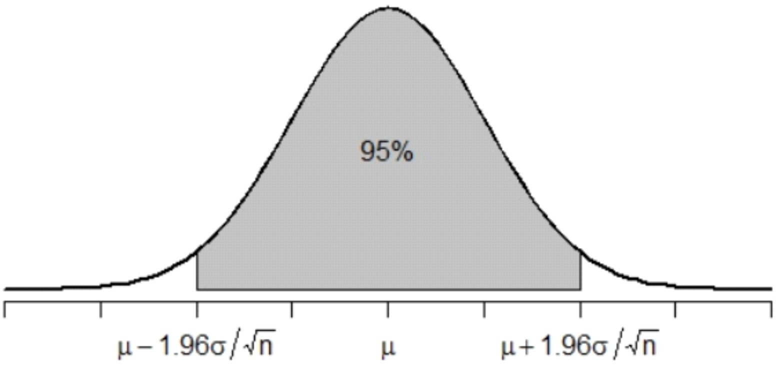
Então o intervalo $(\hat{\theta}_1; \hat{\theta}_2)$ é chamado intervalo de **100(1-α)% de confiança** para o parâmetro θ. Usualmente toma-se $1 - \alpha$=0,95 ou 0,99.

Distribuição normal

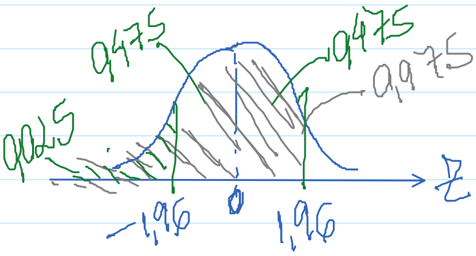


Podemos dizer que 68% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $\sigma/\sqrt{n}$.





Podemos dizer que 95% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $1,96\sigma/\sqrt{n}$.



99%
$Z_T = 2,58$

Intervalos de confiança para média populacional

Casos

- População Normal e Variância da população conhecida;
- População Normal e Variância da população desconhecida;
- População não Normal, grandes amostras (n>30).

Intervalo de confiança para média

- **População normal e variância populacional conhecida**

$$X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

$$P\left(-z_T < \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < z_T\right) = 1-\alpha$$

$$\cdots$$

$$P\left(\bar{X} - z_T\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + z_T\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$$

$$\boxed{IC(\mu)_{1-\alpha} = \left(\bar{X} - z_T\sqrt{\frac{\sigma^2}{n}} ; \bar{X} + z_T\sqrt{\frac{\sigma^2}{n}}\right)}$$

Exemplo: A distribuição dos pesos de pacotes de determinadas sementes, enchidos automaticamente por uma certa máquina, é normal, com desvio padrão (σ) conhecido e igual a 0,20 kg. Uma amostra de 15 pacotes retirada ao acaso apresentou os seguintes pesos, em kg:

20,05	20,10	20,25	19,78	19,69	19,90	20,20	19,89
19,70	20,30	19,93	20,25	20,18	20,01	20,09	

Construir os intervalos de confiança de 95% e 99% para o peso médio dos pacotes de sementes.

$\bar{x} = 20,02$ $n = 15$ $\sigma = 0,20$ kg

$$IC(\mu)_{95\%}: 20,02 \pm 1,96 \cdot \frac{0,20}{\sqrt{15}}$$

$$\boxed{IC(\mu)_{95\%}: 19,92 < \mu < 20,12}$$

$$IC(\mu)_{99\%}: 20,02 \pm 2,58 \cdot \frac{0,20}{\sqrt{15}}$$

$IC(\mu)_{99\%}: 19,89 < \mu < 20,15$

Exercício:

Para uma amostra de 36 observações de uma população normal com variância populacional conhecida $\sigma^2 = 0{,}36kg^2$ *e média μ desconhecida, seja* $\bar{x} = 59{,}6\ kg$ *a média amostral. Construir os intervalos de 95 e 99% de confiança para μ.*

$n = 36 \quad \sigma^2 = 0{,}36\,kg^2 \quad \bar{x} = 59{,}6\,kg$

95%

$Z_T = 1{,}96 \quad IC(\mu)_{95\%} = \bar{x} \pm Z_T\sqrt{\frac{\sigma^2}{n}}$

$= 59{,}6 \pm 1{,}96 \cdot \sqrt{\frac{0{,}36}{36}}$

$IC(\mu)_{95\%}: 59{,}404 < \mu < 59{,}796$

99%

$Z_T = 2{,}58 \quad IC(\mu)_{99\%} = 59{,}6 \pm 2{,}58\sqrt{\frac{0{,}36}{36}}$

$IC(\mu)_{99\%}: 59{,}342 < \mu < 59{,}858$

Exercício:

De um povoamento de eucaliptos, sortearam-se 10 árvores e determinaram-se os diâmetros, em cm, com a finalidade de estimar o diâmetro médio do povoamento, que possui desvio padrão conhecido de 7,8cm. Estes diâmetros foram:

10,1	15,8	18,5	22,3	23,5	17,2	17,8	18,7	16,7	28,0

$\bar{x} = 18{,}86$ cm

$\sigma = 7{,}8$ cm

Com base nessa amostra, calcule:

a) intervalo com grau de confiança de 95%;

$Z_T = 1{,}96$ $\quad IC: \bar{x} \pm Z_T \sqrt{\frac{\sigma^2}{n}}$

$: 18{,}86 \pm 1{,}96 \cdot \frac{7{,}8}{\sqrt{10}}$

$\boxed{IC(\mu)_{95\%}: 14{,}03 < \mu < 23{,}69}$

b) intervalo com grau de confiança de 99%.

$Z_T = 2{,}58$ $\quad IC: \bar{x} \pm Z_T \cdot \sqrt{\frac{\sigma^2}{n}}$

$: 18{,}86 \pm 2{,}58 \cdot \frac{7{,}8}{\sqrt{10}}$

$\boxed{IC(\mu)_{99\%}: 12{,}50 < \mu < 25{,}22}$

Exercícios:

erro máximo

$$IC(\mu)_{1-\alpha} = \left(\bar{X} - z_T\sqrt{\frac{\sigma^2}{n}}; \bar{X} + z_T\sqrt{\frac{\sigma^2}{n}}\right)$$

1) *Um antropólogo mediu as alturas de uma amostra aleatória de 100 homens de determinada população, encontrando uma média amostral de 173 cm. Se o desvio padrão da população for de 9 cm, calcular:*

$n = 100 \quad \bar{x} = 173\,cm \quad \sigma = 9\,cm$

a) um intervalo de 95% de confiança para a altura média de toda a população; interpretar o intervalo de confiança;

$95\% \Rightarrow Z_T = 1{,}96 \quad 173 \pm 1{,}96 \cdot \frac{9}{\sqrt{100}}$

$\boxed{IC(\mu)_{95\%}: 171{,}24 < \mu < 174{,}76}$

b) um intervalo de 99% de confiança para a altura média de toda a população; interpretar o

$IC(\mu)_{95\%}$

b) um intervalo de 99% de confiança para a altura média de toda a população; interpretar o intervalo de confiança.

$99\% \Rightarrow Z_t = 2,58$ $\quad 173 \pm 2,58 \cdot \frac{9}{\sqrt{100}}$

$$\boxed{IC(\mu)_{99\%}: 170,68 < \mu < 175,32}$$

2) Uma máquina enche pacotes de café com um desvio padrão igual a 10g. Ela estava regulada para enchê-los com 500g, em média. Agora ela está desregulada e queremos saber qual a nova média verdadeira (populacional). Uma amostra de 25 pacotes apresentou média igual a 485g. $\quad \sigma = 10g \quad n = 25 \quad \bar{x} = 485g$

a) Construir intervalos de confiança de 95% e de 99% para a média verdadeira; interpretar o intervalo de confiança;

$95\% \Rightarrow Z_t = 1,96$ $\quad 485 \pm 1,96 \cdot \frac{10}{\sqrt{25}}$

$$\boxed{IC(\mu)_{95\%}: 481,08 < \mu < 488,92}$$

$99\% \Rightarrow Z_t = 2,58$ $\quad 485 \pm 2,58 \cdot \frac{10}{\sqrt{25}}$

$$\boxed{IC(\mu)_{99\%}: 479,84 < \mu < 490,16}$$

Pode-se dizer que a máquina realmente está desregulada e não está atingindo o peso necessário de 500g por pacote.

b) Qual o erro máximo associado aos intervalos encontrados em "a"?

$95\% \Rightarrow e_{max} = 1,96 \cdot \frac{10}{\sqrt{25}} = 3,92g$

$99\% \Rightarrow e_{max} = 2,58 \cdot \frac{10}{\sqrt{25}} = 5,16g$

c) Qual o tamanho da amostra será necessário para produzir um intervalo de confiança

c) Qual o tamanho da amostra será necessário para produzir um intervalo de confiança para a verdadeira média populacional, com precisão de 3,5g de café para mais e para menos?

95%

$$e_{max} = Z_T \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow 3{,}5 = 1{,}96 \cdot \frac{10}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1{,}96 \cdot 10}{3{,}5} \Rightarrow n = \left(\frac{1{,}96 \cdot 10}{3{,}5}\right)^2 = 31{,}36$$

$$\boxed{n \cong 32 \text{ unidades}}$$

99%

$$n = \left(\frac{Z_T \cdot \sigma}{e_{max}}\right)^2 = \left(\frac{2{,}58 \cdot 10}{3{,}5}\right)^2 = 54{,}34$$

$$\boxed{n \cong 55 \text{ unidades}}$$

3) De 1000 lavouras de arroz, foi levantada uma amostra de 25 lavouras e a informação a respeito da produtividade permitiu o cálculo do rendimento médio, por hectare, que foi de 3400kg com desvio padrão populacional conhecido de 150kg.

$n = 25$ lavouras $\quad \bar{x} = 3400$ kg/ha $\quad \sigma = 150$kg

a) Determine intervalos com grau de confiança de 95% e de 99% para o verdadeiro rendimento médio;

95% $\quad Z_T = 1{,}96 \qquad 3400 \pm 1{,}96 \cdot 150/\sqrt{25} \Rightarrow e_{max} = 59$

$$\boxed{IC(\mu)_{95\%}: 3341{,}2 < \mu < 3458{,}8}$$

99% $\quad Z_T = 2{,}58 \qquad 3400 \pm 2{,}58 \cdot 150/\sqrt{25}$

$$\boxed{IC(\mu)_{99\%}: 3322{,}6 < \mu < 3477{,}4}$$

b) Qual tamanho deve ter a amostra para que seja de 95% o grau de confiança na estimativa intervalar 3400 ± 100?

b) Qual tamanho deve ter a amostra para que seja de 95% o grau de confiança na estimativa intervalar 3400 ± 100?

$e_{max} = 100g$ $\qquad$ $95\% \Rightarrow Z_T = 1{,}96$

$$e_{max} = Z_T \cdot \frac{\sigma}{\sqrt{n}}$$

$$100 = 1{,}96 \cdot \frac{150}{\sqrt{n}}$$

$$n = \left(\frac{1{,}96 \cdot 150}{100}\right)^2 = 8{,}64$$

$$\boxed{n \cong 9 \text{ unidades}}$$