Estimação

Introdução

- ***Estimação***:
  - ***Objectivo***: ***Inferir os valores dos parâmetros de uma população*** (μ, σ, p, etc.) a partir dos dados de uma amostra.

- ***Estimação Pontual***:
  - ***Def***: Determinação de um ***valor*** aproximado do parâmetro, usando uma amostra.
  - ***Ex***: $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$
  - ***Desvantagem***: ***Erro*** da estimativa ***não quantificado***.

- ***Estimação intervalar***:
  - ***Def***: Determinação de um ***intervalo*** com uma dada ***probabilidade de conter o valor do parâmetro***, usando uma amostra.

- ***Intervalo de confiança***:
  - ***Def***: Um ***intervalo de confiança*** (I.C.) de **100γ% *para o parâmetro* θ** é um intervalo **[*a*,*b*]** com ***probabilidade* γ** de conter o valor de θ. Portanto,

$$P(a \leq \theta \leq b) = \gamma$$

  - ***Notas***:
    - γ designa-se por ***nível de confiança***.
    - $\alpha = 1 - \gamma$ designa-se por ***nível de significância***.
    - θ é fixo mas os limites ***a*** e ***b variam*** de amostra para amostra.

I.C. para a média, μ, duma população

- ***I. C. para μ quando σ é conhecido***:
  - *Teor*: O ***I.C. de 100γ% para a média, μ, de uma população, X***, com ***desvio padrão, σ, conhecido*** é dado por

$$\bar{X} - \lambda \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \lambda \frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad \mu = \bar{X} \pm \Delta, \quad \Delta = \lambda \sigma_{\bar{X}} = \lambda \frac{\sigma}{\sqrt{n}}$$

Δ : ***Erro máximo absoluto*** da estimativa $\bar{x}$

$\bar{X}$: Média duma amostra aleatória i.i.d. dessa população.

n : Tamanho da amostra.

λ : Factor calculado a partir de γ da seguinte forma:

$$P(a \le \mu \le b) = \gamma \Leftrightarrow P(-\lambda \le Z \le \lambda) = \gamma \Leftrightarrow \Phi(\lambda) = \frac{1+\gamma}{2} \Leftrightarrow \Phi(-\lambda) = \frac{1-\gamma}{2}$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$$n \ge 30 \underset{TLC}{\Rightarrow} Z \sim N(0,1)$$

ou

$$X \sim \mathrm{N}(\mu, \sigma^2) \underset{TAN}{\Rightarrow} Z \sim N(0,1)$$

I.C. para a média, μ, duma população

- ***I. C. para μ quando σ é desconhecido***:
  - *Teor*: Sendo ***S*** o desvio padrão amostral, o ***I.C. de 100γ% para a média, μ, de uma população, X***, com ***desvio padrão, σ, desconhecido*** é dado por

$$\bar{X} - \lambda \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + \lambda \frac{S}{\sqrt{n}} \quad \Leftrightarrow \quad \mu = \bar{X} \pm \Delta\,, \quad \Delta = \lambda \frac{S}{\sqrt{n}}$$

sendo λ calculado a partir de γ da seguinte forma:

$$P(a \leq \mu \leq b) = \gamma \Leftrightarrow P(-\lambda \leq T \leq \lambda) = \gamma$$

$$\boxed{T = \frac{\bar{X} - \mu}{S/\sqrt{n}}}$$

i) <u>População normal</u>:

$$X \sim \mathrm{N}(\mu, \sigma^2) \;\Rightarrow\; T \sim \mathrm{T}(n-1) \;\Rightarrow\; P(T \leq \lambda) = F(\lambda) = \frac{1+\gamma}{2}$$

ii) <u>Grandes amostras e população qualquer</u>: R: $F^{-1}(p) = qt(p, df)$

$$n \geq 30 \;\Rightarrow\; T \sim \mathrm{N}(0,1) \;\Rightarrow\; \Phi(\lambda) = \frac{1+\gamma}{2}$$

Exercício

O tempo que um trabalhador especializado leva a realizar uma dada tarefa é normalmente distribuído, com desvio padrão de 12 minutos.

a) Determine um I.C. de 95% para o tempo médio gasto nessa tarefa, sabendo que numa amostra de 36 trabalhadores a média foi de 45 minutos.

b) Qual seria a melhoria na precisão se fossem seleccionados 64 trabalhadores?

c) Qual deve ser o tamanho mínimo da amostra para que o erro da estimativa da média não exceda 3 minutos?

a) X:"Tempo, em minutos, que um trabalhador leva a realizar a tarefa"

$$X \sim N(\mu, \sigma^2) \quad ; \quad \begin{cases} \mu = ? \\ \sigma = 12 \end{cases}$$

$\overline{X}$:"Tempo ***médio*** gasto por tarefa, em minutos, numa ***amostra de 36 tarefas***"

$$n = 36 \quad ; \quad \bar{x} = 45 \quad ; \quad \gamma = 0.95$$

Exercício

$$P(a \le \mu \le b) = \gamma \Leftrightarrow P(-\lambda \le Z \le \lambda) = \gamma \Leftrightarrow \Phi(\lambda) = \frac{1+\gamma}{2} \Leftrightarrow \Phi(\lambda) = \frac{1+0.95}{2} \Leftrightarrow$$

$$\boxed{Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}} \qquad \boxed{X \sim \mathrm{N}(\mu, \sigma^2) \underset{\text{TAN}}{\Rightarrow} Z \sim \mathrm{N}(0,1)}$$

$$\Leftrightarrow \quad \Phi(\lambda) = 0.975 \quad \Leftrightarrow \quad \lambda = \Phi^{-1}(0.975) \quad \Leftrightarrow \quad \lambda = 1.96$$

$qnorm(0.975)$

$$\Delta = \lambda \frac{\sigma}{\sqrt{n}} = 1.96 \frac{12}{\sqrt{36}} = 3.92 \quad \therefore \quad IC_{\mu}^{95\%} = [\bar{x} - \Delta\, , \bar{x} + \Delta] = [41.08\, , 48.92]$$

b) $\Delta_b = \lambda \dfrac{\sigma}{\sqrt{n_b}} = 1.96 \dfrac{12}{\sqrt{64}} = 2.94$

$\dfrac{\Delta_b}{\Delta} = \dfrac{2.94}{3.92} = 0.75 \quad \therefore$ O novo I.C. teria 75% do tamanho do anterior.

c) $\Delta \le 3 \quad \Leftrightarrow \quad \lambda \dfrac{\sigma}{\sqrt{n}} \le 3 \quad \Rightarrow \quad n \ge \left(\dfrac{\lambda\sigma}{3}\right)^2 \quad \Leftrightarrow \quad n \ge \left(\dfrac{1.96 \times 12}{3}\right)^2 \approx 61.4 \quad \therefore \quad n_{\min} = 62$

I.C. para a proporção de sucessos, *p*

- ***I. C. para p***:
  - *Teor*: O ***I.C. de 100γ % para a proporção de sucessos, p, numa população*** é dado por

$$p = \hat{P} \pm \Delta \qquad , \quad \Delta = \lambda \sigma_{\hat{P}} = \lambda \sqrt{\frac{p(1-p)}{n}} \approx \lambda \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

Δ : ***Erro máximo absoluto*** da estimativa $\hat{p}$

$\hat{P}$: Proporção de sucessos numa amostra aleatória i.i.d. dessa população.

n : Tamanho da amostra.

λ : Factor calculado a partir de γ da seguinte forma:

$$P(a \le p \le b) = \gamma \Leftrightarrow P(-\lambda \le Z \le \lambda) = \gamma \Leftrightarrow \Phi(\lambda) = \frac{1+\gamma}{2}$$

$$\boxed{Z = \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}} \qquad \boxed{n \ge 30 \underset{TLC}{\Rightarrow} Z \sim N(0,1)}$$

Numa amostra de 400 peças produzidas por certa máquina, 20 tinham defeito.

a) Determine um I.C. de 99% para a percentagem de peças defeituosas.

b) Que dimensão deve ter a amostra para que o I.C. de 99% tenha uma amplitude máxima de 0.04?

a) $\widehat{P}$:"Proporção de peças defeituosas, em 400."

$$\widehat{p} = \frac{20}{400} = 0.05$$

$$P(a \leq p \leq b) = \gamma \Leftrightarrow P(-\lambda \leq Z \leq \lambda) = \gamma \Leftrightarrow \Phi(\lambda) = \frac{1+\gamma}{2} = \frac{1+0.99}{2} = 0.995 \Leftrightarrow$$

$$\boxed{Z = \frac{\widehat{P} - p}{\sqrt{\frac{\widehat{P}\left(1-\widehat{P}\right)}{n}}}}$$

$$\boxed{n \geq 30 \underset{\text{TLC}}{\Rightarrow} Z \sim \mathrm{N}(0,1)}$$

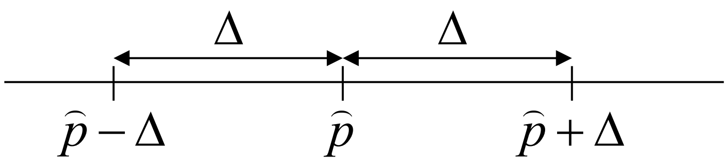
$$\Leftrightarrow \quad \lambda = \Phi^{-1}(0.995) \quad \Leftrightarrow \quad \lambda = 2.58$$

Exercício

$$\Delta = \lambda\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2.58\sqrt{\frac{0.05(1-0.05)}{400}} = 0.028$$

$$\therefore \quad IC_p^{99\%} = [\hat{p}-\Delta\,, \hat{p}+\Delta] = [0.022\,, 0.078]$$

b) $n = ? \quad : \quad 2\Delta \leq 0.04$



$$2\Delta \leq 0.04 \quad \Leftrightarrow \quad \Delta \leq 0.02 \quad \Leftrightarrow \quad \lambda\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.02 \Leftrightarrow$$

$$\Leftrightarrow \quad 2.58\sqrt{\frac{0.05(1-0.05)}{n}} \leq 0.02 \quad \Rightarrow \quad \frac{0.0475}{n} \leq \left(\frac{0.02}{2.58}\right)^2 \Leftrightarrow$$

$$\Leftrightarrow \quad n \geq 0.0475\left(\frac{2.58}{0.02}\right)^2 \approx 790.4 \quad \therefore \quad n_{\min} = 791$$