Capítulo 7 — Estimação Pontual e Intervalos de Confiança

Estimação Pontual

No âmbito da inferência estatística, as estatísticas permitem estimar os parâmetros da população a que pertence a amostra. À estatística, $\widehat{\Theta}$, cujos valores particulares $\hat{\theta}$ constituem estimativas de um parâmetro θ, designa-se estimador pontual do parâmetro em causa.

Propriedades desejáveis dos estimadores pontuais

Não-Enviesamento

$$\text{Enviesamento}_{\widehat{\Theta}} = E(\Theta) - \theta = \mu_{\widehat{\Theta}} - \theta.$$

Eficiência

$$\text{Eficiência}_{\widehat{\Theta}} = E\left[\left(\widehat{\Theta} - \theta\right)^2\right] = \sigma^2_{\widehat{\Theta}} + (\text{Enviesamento}_{\widehat{\Theta}})^2.$$

$$\text{Eficiência relativa}_{\widehat{\Theta}1/\widehat{\Theta}2} = \frac{E[\left(\widehat{\Theta}_1 - \theta\right)^2]}{E[\left(\widehat{\Theta}_2 - \theta\right)^2]}$$

Consistência

Condições suficientes (não necessárias):

$$\lim_{N\to\infty} (\mu_{\widehat{\Theta}} - \theta) = 0 \quad e \quad \lim_{N\to\infty} \sigma^2_{\widehat{\Theta}} = 0 \quad ou \quad \lim_{N\to\infty} E[\left(\widehat{\Theta}_1 - \theta\right)^2] = 0.$$

Suficiência

Capacidade de condensar toda a informação que a amostra possui relativamente ao parâmetro.

Robustez

Capacidade de um estimador se manter aproximadamente não-enviesado e eficiente para uma vasta gama potencial de distribuições populacionais.

Métodos de estimação

Método dos momentos

Propriedades dos estimadores obtidos: (i) são consistentes; (ii) seguem uma distribuição Normal quando a amostra tende para infinito; (iii) são, para amostras de grande dimensão, menos eficientes que os obtidos pelo método de máxima verosimilhança.
Seja uma população representada pela variável Z, cuja distribuição é conhecida a menos de R parâmetros $\theta_1, \theta_2, \ldots, \theta_R$. Geralmente, os momentos populacionais ordinários são funções dos parâmetros a estimar $\mu' = \mu'(\theta_1, \theta_2, \ldots, \theta_R)$.
Seja ainda $\{Z_1, Z_2, \ldots, Z_N\}$ uma amostra aleatória simples. Os momentos amostrais ordinários são dados por:

$$M'_i = \frac{1}{N} \cdot \sum_{n=1}^{N} (Z_n)^i$$

Os <u>estimadores **Mom**</u> (designação abreviada para os estimadores calculados através deste método) obtêm-se resolvendo o seguinte sistema de equações em ordem a $\theta_1, \theta_2, \ldots, \theta_R$:

$$M'_i = \mu'_i(\theta_1, \theta_2, \ldots, \theta_R) \quad para \;\; i = 1,2, \ldots, R$$

Método da máxima verosimilhança

Propriedades dos estimadores obtidos: (i) são, em geral, consistentes; (ii) para dimensões de amostras crescentes, tendem a ser não-enviesados e eficientes; (iii) seguem, frequentemente, uma distribuição assimptoticamente Normal.
Seja Y uma variável discreta, cuja distribuição é conhecida a menos de R parâmetros $\theta_1, \theta_2, \ldots, \theta_R$. A *função de verosimilhança* para uma amostra aleatória $\{y_1, y_2, \ldots, y_N\}$ é dada por:

$$L = Probabilidade(Y_1 = y_1, \ldots, Y_N = y_N | \theta_1, \theta_2, \ldots, \theta_R) = \prod_{n=1}^{N} Probabilidade(Y_N = y_n | \theta_1, \theta_2, \ldots, \theta_R)$$

As estimativas de máxima verosimilhança correspondem aos valores dos parâmetros que maximizam *L*.

$$\max_{\theta_1,\theta_2,\ldots,\theta_R} : L(\theta_1, \theta_2, \ldots, \theta_R) \Rightarrow \frac{\partial L}{\partial \theta_i} = 0 \quad (i = 1, 2, \ldots, R)$$

Analogamente, sendo *X* uma variável contínua, cuja distribuição é conhecida a menos de *R* parâmetros $\theta_1, \theta_2, \ldots, \theta_R$. A *função de verosimilhança* para uma amostra aleatória $\{x_1, x_2, \ldots, x_N\}$ é dada por:

$$L = f_{X_1,X_2,\ldots,X_N,\theta_1,\theta_2,\ldots,\theta_R}(x_1, x_2, \ldots, x_N) = \prod_{n=1}^{N} f_{X_n,\theta_1,\theta_2,\ldots,\theta_R}(x_n)$$

As estimativas de máxima verosimilhança (que se designam abreviadamente por estimativas **MV**) são aos valores dos parâmetros que maximizam *L*.

$$\max_{\theta_1,\theta_2,\ldots,\theta_R} : L(\theta_1, \theta_2, \ldots, \theta_R) \Rightarrow \frac{\partial L}{\partial \theta_i} = 0 \quad (i = 1, 2, \ldots, R)$$

Método de estimação linear com variância mínima

Propriedades dos estimadores obtidos: (i) são não-enviesado; (ii) têm a variância mínima possível

A partir deste método obtém-se um estimador que resulta da combinação linear de um conjunto de estimadores independentes, não-enviesados e com variância conhecida. O estimador **LVM** (designação abreviada para o estimador assim obtido) calcula-se a partir da expressão:

$$\hat{\Theta} = \sum_{n=1}^{N} k_n \cdot \hat{\Theta}_n \quad com \quad k_n = \frac{1/\sigma_n^2}{\sum_{n=1}^{N} 1/\sigma_n^2}$$

Método dos mínimos quadrados

Propriedades dos estimadores obtidos: (i) são, em geral, não-enviesado; (ii) têm, em geral, variância mínima.

Considerem-se diferentes réplicas de uma variável *X* expressas sob a forma:

$$X_n = f_n(\theta_1, \theta_2, \ldots, \theta_R) + E_n$$

onde $f_n(\theta_1, \theta_2, \ldots, \theta_R)$ $(n = 1, \ldots, N)$ representam funções conhecidas de *R* parâmetros $\theta_1, \theta_2, \ldots, \theta_R$ e E_n $(n = 1, \ldots, N)$ representam réplicas independentes de uma variável aleatória *E*, habitualmente designada por erro, com valor esperado nulo e variância σ_E^2.

Dispondo de um conjunto de *N* (*N*>*R*) observações de *X*, $\{X_1, X_2, \ldots, X_N\}$ os estimadores obtidos segundo o método dos mínimos quadrados (os estimadores **MQ**) dos parâmetros $\theta_1, \theta_2, \ldots, \theta_R$ são calculados minimizando o somatório do quadrado dos erros:

$$\min_{\theta_1,\theta_2,\ldots,\theta_R} : SEQ(\theta_1, \theta_2, \ldots, \theta_R) = \sum_{n=1}^{N} E_n^2 = \sum_{n=1}^{N} [X_n - f_n(\theta_1, \theta_2, \ldots, \theta_R)]^2$$

Resolvendo o seguinte sistema de equações em ordem a $\theta_1, \theta_2, \ldots, \theta_R$ obtêm-se os estimadores pretendidos.

$$\frac{\partial SEQ}{\partial \theta_i} = 0 \quad (i = 1, 2, \ldots, R$$

Estimação por Intervalo

Todos os intervalos que seguidamente se definem possuem um nível de confiança de (100-α)%.

Intervalo de confiança para o Valor Esperado (μ)

Amostra de grande dimensão ($s \approx \sigma$)

$$\left[\bar{X} - Z_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + Z_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{N}}\right]$$

Amostra de pequena dimensão (e população Normal)

$$\left[\bar{X} - t_{N-1\,(\alpha/2)} \cdot \frac{s}{\sqrt{N}}, \bar{X} + t_{N-1\,(\alpha/2)} \cdot \frac{s}{\sqrt{N}}\right]$$

Intervalo de confiança para a Proporção Binomial (p)

Amostra de grande dimensão (população infinita ou amostragem com reposição)

$$\left[\hat{p} - Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N}}, \hat{p} + Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N}}\right]$$

Amostra de grande dimensão (população finita de dimensão *M* e amostragem sem reposição)

$$\left[\hat{p} - Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{M-N}{M-1}}, \hat{p} + Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{M-N}{M-1}}\right]$$

Intervalo de confiança para a Variância de uma população Normal (σ^2)

$$\left[\frac{(N-1) \cdot s^2}{\chi^2_{N-1\,(\alpha/2)}}, \frac{(N-1) \cdot s^2}{\chi^2_{N-1\,(1-\alpha/2)}}\right]$$

Intervalo de confiança para a diferença de valores esperados de duas populações ($\mu_A - \mu_B$)

Amostras independentes de grande dimensão ($s_A^2 \approx \sigma_A^2$ e $s_B^2 \approx \sigma_B^2$)

Admitindo que as populações têm variâncias diferentes ($\sigma_A^2 \not\approx \sigma_B^2$):

$$\left[(\bar{X}_A - \bar{X}_B) - Z_{(\alpha/2)} \cdot \sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}}, (\bar{X}_A - \bar{X}_B) + Z_{(\alpha/2)} \cdot \sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}}\right]$$

Admitindo que as populações têm variâncias iguais ($\sigma_A^2 \approx \sigma_B^2$):

$$\left[(\bar{X}_A - \bar{X}_B) - Z_{(\alpha/2)} \cdot \sigma \cdot \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}, (\bar{X}_A - \bar{X}_B) + Z_{(\alpha/2)} \cdot \sigma \cdot \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}\right]$$

em que: $\sigma = \sqrt{\frac{(N_A-1) \cdot s_A^2 + (N_B-1) \cdot s_B^2}{N_A + N_B - 2}}$

Amostras independentes de pequena dimensão (e populações Normais)

Admitindo que as populações têm variâncias diferentes ($\sigma_A^2 \not\approx \sigma_B^2$):

$$\left[(\bar{X}_A - \bar{X}_B) - t_{GL\,(\alpha/2)} \cdot \sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}, (\bar{X}_A - \bar{X}_B) + t_{GL\,(\alpha/2)} \cdot \sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}\right]$$

em que: $GL = \dfrac{\left(s_A^2/N_A + s_B^2/N_B\right)^2}{\frac{\left(s_A^2/N_A\right)^2}{N_A - 1} + \frac{\left(s_B^2/N_B\right)^2}{N_B - 1}}$

Admitindo que as populações têm variâncias iguais ($\sigma_A^2 \approx \sigma_B^2$):

$$\left[(\bar{X}_A - \bar{X}_B) - t_{GL\,(\alpha/2)} \cdot s \cdot \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}, (\bar{X}_A - \bar{X}_B) + t_{GL\,(\alpha/2)} \cdot s \cdot \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}\right]$$

em que: $GL = N_A + N_B - 2$ e $S = \sqrt{\dfrac{(N_A - 1) \cdot s_A^2 + (N_B - 1) \cdot s_B^2}{N_A + N_B - 2}}$

Intervalo de confiança para a diferença de Proporções Binomiais ($\boldsymbol{p_A - p_B}$)

Amostras independentes de grande dimensão (populações infinitas ou amostragem com reposição)

$$\left[(\hat{p}_A - \hat{p}_B) - Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p}_A \cdot (1 - \hat{p}_A)}{N_A} + \frac{\hat{p}_B \cdot (1 - \hat{p}_B)}{N_B}}, (\hat{p}_A - \hat{p}_B) + Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p}_A \cdot (1 - \hat{p}_A)}{N_A} + \frac{\hat{p}_B \cdot (1 - \hat{p}_B)}{N_B}}\right]$$

Amostras independentes de grande dimensão (populações de dimensão finita (M_A e M_B) e amostragem sem reposição).

$$\left[(\hat{p}_A - \hat{p}_B) - Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p}_A \cdot (1 - \hat{p}_A)}{N_A} \cdot \frac{M_A - N_A}{M_A - 1} + \frac{\hat{p}_B \cdot (1 - \hat{p}_B)}{N_B} \cdot \frac{M_B - N_B}{M_B - 1}}, (\hat{p}_A - \hat{p}_B) + Z_{(\alpha/2)} \cdot \sqrt{\frac{\hat{p}_A \cdot (1 - \hat{p}_A)}{N_A} \cdot \frac{M_A - N_A}{M_A - 1} + \frac{\hat{p}_B \cdot (1 - \hat{p}_B)}{N_B} \cdot \frac{M_B - N_B}{M_B - 1}}\right]$$

Intervalo de confiança para a razão entre Variâncias de duas populações Normais ($\boldsymbol{\sigma_A^2/\sigma_B^2}$)

$$\left[\frac{1}{F_{N_A - 1, N_B - 1\,(\alpha/2)}} \cdot \frac{s_A^2}{s_B^2}, \frac{1}{F_{N_A - 1, N_B - 1\,(1 - \alpha/2)}} \cdot \frac{s_A^2}{s_B^2}\right]$$

Formulário adaptado de:
Estatística
Rui Campos Guimarães, José A. Sarsfield Cabral
Verlag Dashöfer