DGI
2019

8.1. Estimadores

No âmbito da inferência estatística, ao calcular uma estatística (por exemplo uma média amostral) o objectivo passa a ser o de ***caracterizar*** um conjunto mais vasto de dados: a ***população a partir da qual a amostra foi retirada***.

- A ***forma mais directa*** de se proceder a uma inferência estatística corresponde à ***estimação pontual de parâmetros***.
- ***Designa-se por estimador (pontual) de um parâmetro uma estatística cujo valor constitui uma estimativa do parâmetro em causa.***

Exemplos

A média amostral $\left(\bar{X}\right)$ é um estimador do valor esperado (μ_X) da população

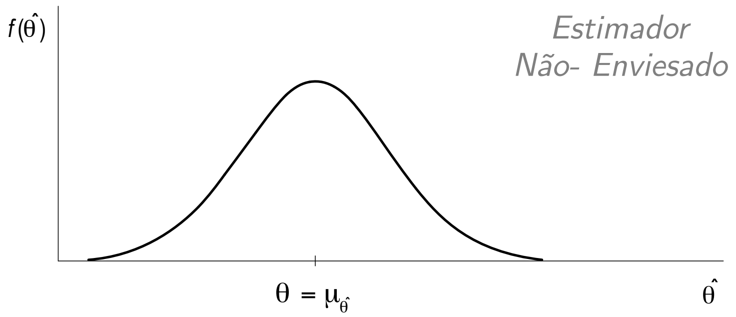
O erro quadrático médio (EQM_X) é um estimador da variância $\left(\sigma_X^2\right)$ da população

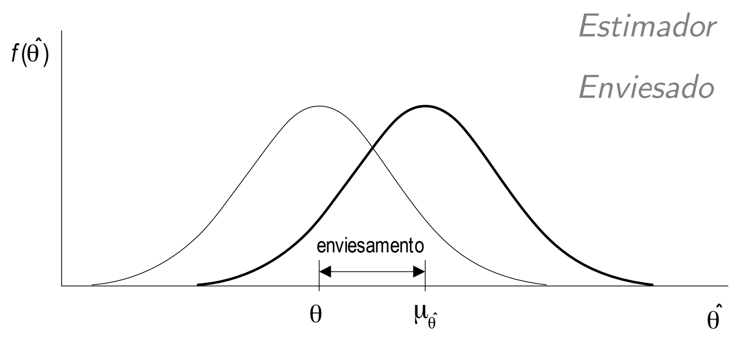
8.2. Propriedades Desejáveis dos Estimadores

8.2.1. Enviesamento

Seja θ um parâmetro qualquer de uma população e $\hat{\theta}$ um estimador desse parâmetro.

- *O estimador diz-se não* ***enviesado quando o seu valor esperado coincidir com o valor do parâmetro****.*





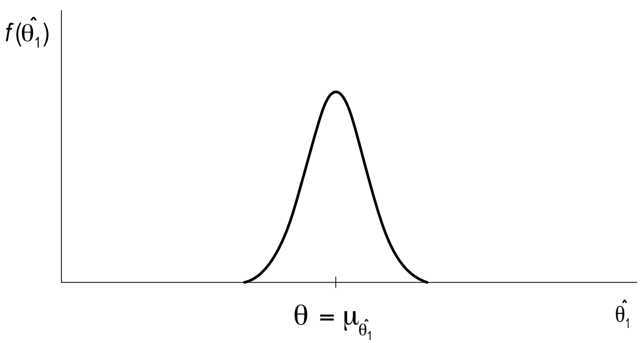
$$\mathrm{E}\left[\hat{\theta}\right] = \theta \qquad\qquad \text{Enviesamento} = \mathrm{E}\left[\hat{\theta}\right] - \theta$$

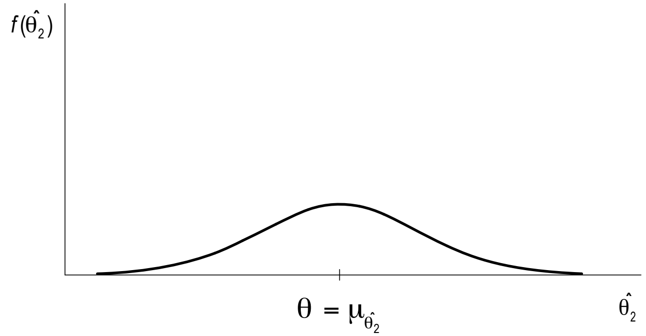
- *O estimador será enviesado se a diferença $\mathrm{E}\left(\hat{\theta}\right) - \theta$ for maior ou menor do que zero.* Tal diferença designa-se por enviesamento.

DGI
2019

8.2.2. ***Eficiência***

Considere-se o caso de dois estimadores não enviesados $(\hat{\theta}_1,\ \hat{\theta}_2)$ de um mesmo parâmetro θ. Admita-se que, para amostras de dimensão n, as variâncias daqueles estimadores são $\sigma^2{}_1$ e $\sigma^2{}_2$, com $\sigma^2{}_1 < \sigma^2{}_2$.





- Neste caso, o estimador $\hat{\theta}_1$ é melhor, a sua distribuição é mais concentrada em torno de θ ou seja tem uma menor variância – $\hat{\theta}_1$ é pois um estimador mais eficiente.

$$\text{Eficiência relativa entre } \hat{\theta}_1 \text{ e } \hat{\theta}_1 = \frac{VAR(\hat{\theta}_1)}{VAR(\hat{\theta}_2)}$$

- *Quando* ***dois estimadores são não enviesados, o melhor estimador é aquele que tem menor variância.***

Exemplo

Com base numa amostra de duas observações, considere os dois estimadores seguintes para a média:

$$U = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

$$V = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

(a) U e V são estimadores enviesados?

(b) Qual o estimador mais eficiente?

Generalize-se agora a comparação entre estimadores, admitindo que estes podem ser enviesados.

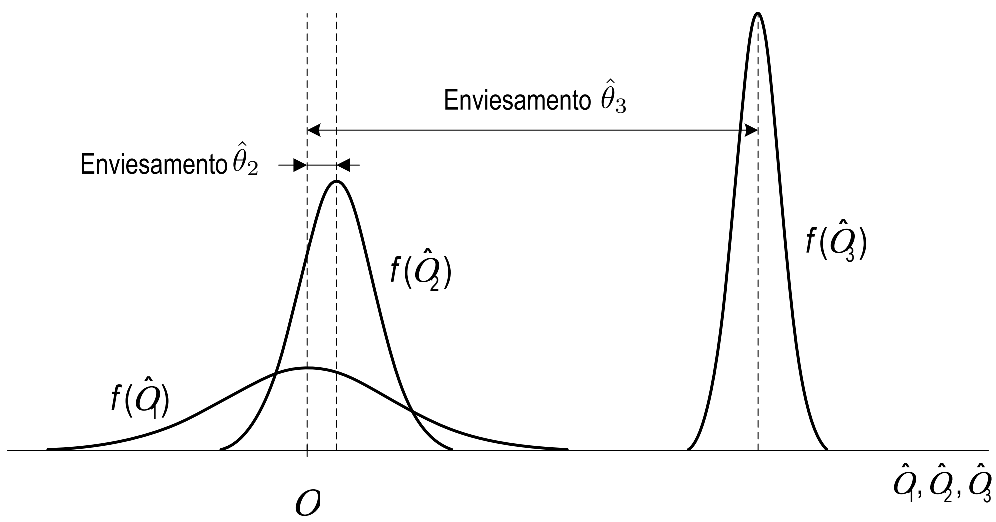
A medida da eficiência do estimador $\hat{\theta}$ será o desvio quadrático médio (DQM), dado por:

$$DQM = E\left[(\hat{\theta} - \theta)^2\right] = \sigma^2_{\hat{\theta}} + (\text{enviesamento})^2$$

- Para quaisquer dois estimadores, enviesados ou não enviesados:

$$\text{Eficiencia relativa de } \theta_1 \text{e } \theta_2 = \frac{DQM(\theta_1)}{DQM(\theta_2)}$$

Considere-se o exemplo da figura. Dos três estimadores $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$, o de menor variância $\hat{\theta}_3$ está longe de ser o melhor, dado que é muito mais enviesado do que os outros dois. O melhor também não será o que apresenta enviesamento nulo $(\hat{\theta}_1)$, dado que tem uma variância muito grande. Restará possivelmente $\hat{\theta}_2$ como o melhor compromisso.



- Se existir um estimador que seja mais eficiente do que qualquer outro então diz-se que o estimador é eficiente (ou absolutamente eficiente).

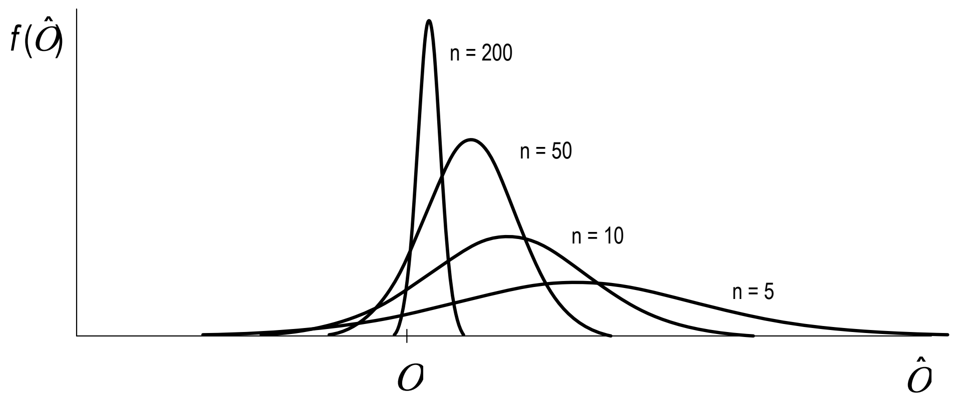
8.2.3. Consistência

Seja $\hat{\theta}$ um estimador do parâmetro θ e n a dimensão da amostra com base na qual $\hat{\theta}$ é calculado.

- O estimador $\hat{\theta}$ diz-se consistente quando, para qualquer valor positivo ∂, se verifica que:

$$\lim_{n\to\infty} PROB\left(\left|\hat{\theta}-\theta\right| < \partial\right) = 1$$

Isto é, quando $n \to \infty$ *o estimador consistente concentra-se perfeitamente sobre o seu alvo* (o valor do parâmetro estimado).



8.3. Métodos de Estimação

8.3.1. Método da Máxima Verosimilhança

Considere-se uma população com função densidade que contenha um parâmetro populacional, θ, a ser estimado por meio de determinada estatística. A função densidade pode neste caso ser denotada por $f(x,\ \theta)$. Admitindo que haja *n observações independentes*, X_1, X_2, ..., X_n, a função densidade conjunta para essas observações é:

$$L = f(x_1,\ \theta)\ f(x_2,\ \theta)\ \ldots\ f(x_n,\ \theta) \quad ; \quad (L \equiv verosimilhanca)$$

A máxima verosimilhança é obtida, derivando L em ordem a θ e igualando a zero:

$$\frac{dL}{d\theta} = 0$$

- Para o caso de vários parâmetros (k parâmetros), deverão ser tomadas as várias derivadas parciais em relação a cada um deles:

$$L = f(x_1,\ \theta_1)\ f(x_2,\ \theta_2)\ \dots\ f(x_n,\ \theta_k) \qquad\qquad \frac{\partial L}{\partial \theta_i} = 0 \quad ; \quad (i = 1,\ 2,\ \dots,\ k)$$

Na generalidade dos casos, os estimadores de máxima verosimilhança são dos melhores que se podem obter, apresentando as seguintes características:

(1) Em geral são consistentes;

(2) Embora nem sempre sejam não enviesados e eficientes para amostras pequenas, tendem a possuir aquelas propriedades à medida que a dimensão da amostra cresce;

(3) As suas distribuições aproximam-se de distribuições normais quando a dimensão da amostra tende para infinito.

8.3.2. *Método da Estimação Linear com Variância mínima*

Para que o ***método da máxima verosimilhança possa ser aplicado, há que especificar previamente o tipo de distribuição da população. No método da estimação linear com variância mínima tal especificação não é necessária***.

- O método consiste em ***combinar linearmente um conjunto de estimadores não-enviesados com variância mínima***.

No caso geral de uma *combinação linear de k estimadores*, obtem-se:

$$\hat{\theta} = \sum_{i=1}^{k} a_i \times \hat{\theta}_i \qquad \text{com} \qquad a_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{k} \left(1/\sigma_j^2\right)}$$