Università degli Studi di Cassino

Esercitazione di Statistica 2 del 15.02.2007

Simona Balzano

Esercizio 1
Per stimare il valore atteso di una variabile casuale X con media μ e varianza σ^2, si estrae un campione (X_1, X_2, X_3, X_4) di ampiezza n = 4.
Si considerino i due stimatori T_1 e T_2 per μ definiti come segue:

$$T_1 = 0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4$$

e

$$T_2 = \frac{1}{2}(0,2 \cdot X_1 + 0,4 \cdot X_2 + 0,6 \cdot X_3 + 0,8 \cdot X_4)$$

a) si verifichi se i due stimatori sono corretti e, se non lo sono, si calcolino le rispettive distorsioni;
b) quale dei due è più efficiente?

Soluzione

a)
Uno stimatore T per un parametro incognito θ è corretto se il valore atteso di T è uguale a θ:

$$E(T) = \theta$$

Dobbiamo, dunque, verificare se sussiste le seguenti uguaglianze:

$$E(T_1) = \mu$$
$$E(T_2) = \mu$$

$$\begin{aligned} E(T_1) &= E(0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4) = \\ &= 0,1 \cdot E(X_1) + 0,2 \cdot E(X_2) + 0,3 \cdot E(X_3) + 0,4 \cdot E(X_4) = \\ &= 0,1 \cdot \mu + 0,2 \cdot \mu + 0,3 \cdot \mu + 0,4 \cdot \mu = \mu(0,1 + 0,2 + 0,3 + 0,4) = \mu \end{aligned}$$

$$\begin{aligned} E(T_1) &= E\left[\frac{1}{2}(0,2 \cdot X_1 + 0,4 \cdot X_2 + 0,6 \cdot X_3 + 0,8 \cdot X_4)\right] = \\ &= \frac{1}{2}\left[0,2 \cdot E(X_1) + 0,4 \cdot E(X_2) + 0,6 \cdot E(X_3) + 0,8 \cdot E(X_4)\right] = \\ &= \frac{1}{2}(0,2 \cdot \mu + 0,4 \cdot \mu + 0,6 \cdot \mu + 0,8 \cdot \mu) = \frac{1}{2}(0,2 + 0,4 + 0,6 + 0,8)\mu = \frac{1}{2} \cdot 2 \cdot \mu = \mu \end{aligned}$$

T_1 e T_2 sono entrambi corretti.

b)
Valutare quale dei due stimatori è più efficiente significa valutare l'efficienza relativa dei due, ossia:

$$\mathrm{Eff}\left(T_1 \mid T_2\right) = \frac{\mathrm{EQM}\left(T_2\right)}{\mathrm{EQM}\left(T_1\right)}$$

dove:

$$\mathrm{EQM}\ (T) = \mathrm{Var}(T) + d^2,$$

che se T è corretto si riduce a :

$$\mathrm{EQM}\ (T) = \mathrm{Var}(T)$$

Quindi, essendo T_1 e T_2 corretti:

$$\mathrm{Eff}\left(T_1 \mid T_2\right) = \frac{\mathrm{Var}\left(T_2\right)}{\mathrm{Var}\left(T_1\right)}$$

Occorre, dunque, calcolare le varianze dei due stimatori, ossia esprimere le due varianze campionarie in funzione della varianza della popolazione σ^2.

$$\begin{aligned}\mathrm{Var}\left(T_1\right) &= \mathrm{Var}\left(0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4\right) = \\ &= 0,1^2 \cdot \mathrm{Var}\left(X_1\right) + 0,2^2 \cdot \mathrm{Var}\left(X_2\right) + 0,3^2 \cdot \mathrm{Var}\left(X_3\right) + 0,4^2 \cdot \mathrm{Var}\left(X_4\right) = \\ &= 0,01 \cdot \sigma^2 + 0,04 \cdot \sigma^2 + 0,09 \cdot \sigma^2 + 0,16 \cdot \sigma^2 = \\ &= \sigma^2 \cdot \left(0,01 + 0,04 + 0,09 + 0,16\right) = 0,3\sigma^2\end{aligned}$$

$$\begin{aligned}\mathrm{Var}\left(T_2\right) &= \mathrm{Var}\left(0,2 \cdot X_1 + 0,4 \cdot X_2 + 0,6 \cdot X_3 + 0,8 \cdot X_4\right) = \\ &= 0,2^2 \cdot \mathrm{Var}\left(X_1\right) + 0,4^2 \cdot \mathrm{Var}\left(X_2\right) + 0,6^2 \cdot \mathrm{Var}\left(X_3\right) + 0,8^2 \cdot \mathrm{Var}\left(X_4\right) = \\ &= 0,04 \cdot \sigma^2 + 0,16 \cdot \sigma^2 + 0,36 \cdot \sigma^2 + 0,64 \cdot \sigma^2 = \\ &= \sigma^2 \cdot \left(0,04 + 0,16 + 0,36 + 0,64\right) = 0,3\sigma^2\end{aligned}$$

Ora è possibile calcolare l'efficienza relativa tra i due stimatori:

$$\mathrm{Eff}\left(T_1 \mid T_2\right) = \frac{\mathrm{Var}\left(T_2\right)}{\mathrm{Var}\left(T_1\right)} = \frac{0,3\sigma^2}{0,3\sigma^2} = 1$$

È del tutto indifferente utilizzare T_1 o T_2.

Nota
Se si desidera approfondire ulteriormente il confronto, bisogna passare ad esaminare le proprietà asintotiche.
Ad esempio, si può considerare la velocità con cui raggiungono il limite inferiore della varianza (consistenza), oppure la velocità con cui convergono alla distribuzione normale (normalità asintotica), che sicuramente è raggiunta da entrambi al crescere di n dal momento che, essendo costruiti come somme di v.c. indipendenti e identicamente distribuite, vale il teorema del limite centrale.

Esercizio 2

In riferimento alla stessa variabile casuale dell'esercizio 1 e allo stesso campione, considerare i due stimatori T_1 e T_2 per μ definiti come segue:

$$T_1 = \frac{1}{2}(0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4)$$

e

$$T_2 = \frac{1}{6}(X_1 + 2X_2 + 2X_3 + X_4)$$

a) si verifichi se i due stimatori sono corretti e, se non lo sono, calcolare le rispettive distorsioni;
b) quale dei due stimatori è più efficiente?

Soluzione

a)

$$\begin{aligned} E(T_1) &= E\left[\frac{1}{2}(0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4)\right] = \\ &= \frac{1}{2}\left[0,1 \cdot E(X_1) + 0,2 \cdot E(X_2) + 0,3 \cdot E(X_3) + 0,4 \cdot E(X_4)\right] = \\ &= \frac{1}{2}(0,1 \cdot \mu + 0,2 \cdot \mu + 0,3 \cdot \mu + 0,4 \cdot \mu) = \frac{1}{2}\mu(0,1 + 0,2 + 0,3 + 0,4) = \frac{1}{2}\mu \end{aligned}$$

Lo stimatore T_1 non è corretto. La sua distorsione è pari a:

$$d_1 = E(T_1) - \mu = \frac{1}{2}\mu - \mu = -\frac{1}{2}\mu$$

La distorsione di T_1 dipende dalla media incognita μ: per $\mu = 0$ T_1 è corretto e la distorsione aumenta al crescere di μ.

$$\begin{aligned} E(T_1) &= E\left[\frac{1}{6}(X_1 + 2X_2 + 2X_3 + X_4)\right] = \\ &= \frac{1}{6}\left[E(X_1) + 2 \cdot E(X_2) + 2 \cdot E(X_3) + E(X_4)\right] = \\ &= \frac{1}{6}(\mu + 2\mu + 2\mu + \mu) = \frac{1}{6}(1 + 2 + 2 + 1)\mu = \frac{1}{6} \cdot 6 \cdot \mu = \mu \end{aligned}$$

Lo stimatore T_2 è corretto.

b)
Nel calcolo dell'efficienza relativa dei due stimatori errori quadratici medi entra in gioco anche la distorsione (che per T_2 è nulla).
In generale, dunque:

$$\text{Eff}(T_1 \mid T_2) = \frac{\text{EQM}(T_2)}{\text{EQM}(T_1)} = \frac{\text{Var}(T_2) + d_2^2}{\text{Var}(T_1) + d_1^2}$$

$$\text{Var}(T_1) = \text{Var}\left[\frac{1}{2}(0,1 \cdot X_1 + 0,2 \cdot X_2 + 0,3 \cdot X_3 + 0,4 \cdot X_4)\right] =$$

$$= \left(\frac{1}{2}\right)^2 \cdot \left[0,1^2 \cdot \text{Var}(X_1) + 0,2^2 \cdot \text{Var}(X_2) + 0,3^2 \cdot \text{Var}(X_3) + 0,4^2 \cdot \text{Var}(X_4)\right] =$$

$$= \frac{1}{4} \cdot \left(0,01 \cdot \sigma^2 + 0,04 \cdot \sigma^2 + 0,09 \cdot \sigma^2 + 0,16 \cdot \sigma^2\right) =$$

$$= \frac{1}{4} \cdot \sigma^2 \cdot (0,01 + 0,04 + 0,09 + 0,16) = \frac{0,3}{4}\sigma^2 = 0,075\sigma^2$$

$$\text{Var}(T_2) = \text{Var}\left[\frac{1}{6}(X_1 + 2X_2 + 2X_3 + X_4)\right] =$$

$$= \left(\frac{1}{6}\right)^2 \cdot \left[\text{Var}(X_1) + 2^2 \cdot \text{Var}(X_2) + 2^2 \cdot \text{Var}(X_3) + \text{Var}(X_4)\right] =$$

$$= \frac{1}{36} \cdot \left(\sigma^2 + 4\sigma^2 + 4\sigma^2 + \sigma^2\right) =$$

$$= \frac{1}{36} \cdot \sigma^2 \cdot (1 + 4 + 4 + 1) = \frac{10}{16} \cdot \sigma^2 = 0,278\sigma^2$$

È ora possibile calcolare l'efficienza relativa:

$$\text{Eff}(T_1 \mid T_2) = \frac{\text{EQM}(T_2)}{\text{EQM}(T_1)} = \frac{\text{Var}(T_2)}{\text{Var}(T_1) + d_1^2} = \frac{0,278\sigma^2}{0,075\sigma^2 + (-0,5\mu)^2}$$

L'efficienza relativa non è dunque determinabile univocamente ma dipende dalla media e dalla varianza della popolazione.
Si può però considerare che:

- per $\mu = 0$ (valore in cui il fattore di distorsione si annulla) l'efficienza relativa vale 3,7 che, essendo maggiore di 1, indica che T_1 è migliore di T_2;

- al crescere di μ (in valore assoluto), l'errore quadratico medio di T_1 aumenta (aumentando la sua distorsione), quindi via via diminuisce la preferenza per T_1 rispetto a T_2.

È possibile determinare il valore di μ che definisce una soglia di scelta tra T_1 e T_2 come quel valore che rende l'efficienza relativa pari ad 1, ossia quel valore che soddisfa la seguente uguaglianza:

$$\frac{0,278\sigma^2}{0,075\sigma^2 + (-0,5\mu)^2} = 1$$

che può essere ottenuto come segue:

$$\begin{aligned} 0,075\sigma^2 + 0,25\mu^2 &= 0,278\sigma^2 \\ 0,25\mu^2 &= 0,278\sigma^2 - 0,075\sigma^2 \\ \mu &= \sigma\sqrt{\frac{0,278 - 0,075}{0,25}} = 0,9\sigma \end{aligned}$$

Ma neanche in questo caso il valore è determinato, in quanto varia al variare dello scarto quadratico medio incognito.

Esercizio 3

Si vuole stimare la lunghezza media delle foglie di una data specie di albero attraverso un campione di ampiezza n = 100 da cui risulta un valore medio pari a 98 mm.. Sapendo che la varianza è pari a 169, costruire un intervallo di confidenza per la media al livello di significatività del:

a) 95 %;
b) 99 %.

Soluzione

Si tratta in entrambi i casi di costruire un intervallo di confidenza per μ con varianza nota nel modo seguente:

$$IC_{1-\alpha} = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

ossia quell'intervallo che con probabilità pari a 0,95 contiene la media:

$$P\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

a) Popolazione non nota, varianza nota, n > 30

$\bar{x}$ = 98
σ^2 = 169
n = 100
α = 0,05
$\alpha/2$ = 0,025

$z_{\alpha/2} = z_{0,025} = \pm\ 1,96$

$$IC_{0,95} = 98 \pm 1,96 \frac{\sqrt{169}}{\sqrt{100}} = [98 - 1,96 \times 1,3; \quad 98 + 1,96 \times 1,3] =$$
$$= [95,5; \quad 100,5]$$

quindi:

$$P(95,5 \le \mu \le 100,5) = 0,95$$

b) Popolazione non nota, varianza nota, n > 30

$\bar{x} = 98$
$\sigma^2 = 169$
$n = 100$
$\alpha = 0,01$
$\alpha/2 = 0,005$

$z_{\alpha/2} = z_{0,005} = \pm 2,58$

$$IC_{0,95} = 98 \pm 2,58 \frac{\sqrt{169}}{\sqrt{100}} = [98 - 2,58 \times 1,3; \quad 98 + 2,58 \times 1,3] =$$
$$= [94,6; \quad 101,4]$$

quindi:

$$P(94,6 \leq \mu \leq 101,4) = 0,99$$

Esercizio 4

Un lotto di 25 bulloni presenta un diametro medio di 200 mm e uno scarto quadratico medio corretto pari a 0,5 mm. Sapendo che la lunghezza del diametro si distribuisce normalmente:

a) Determinare un intervallo di confidenza per la media dei bulloni prodotti, ad un livello di confidenza del 95%
b) Determinare l'intervallo per un lotto di 100 pezzi

Soluzione

a) Popolazione normale, varianza non nota, n < 30

$\bar{x} = 200$
$s = 0{,}5$
$n = 25$
$\alpha = 0{,}05$
$\alpha/2 = 0{,}025$

L'intervallo va determinato come:

$$IC = \bar{x} \pm t_{\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

$t_{\alpha/2;n-1} = t_{0,025;24} = 2{,}064$

$$IC_{0,95} = \left[\bar{x} \pm t_{0,025;24} \frac{s}{\sqrt{n}}\right] = \left[200 - 2{,}064 \frac{0{,}5}{\sqrt{25}};\quad 200 + 2{,}064 \frac{0{,}5}{\sqrt{25}}\right] =$$
$$= \left[199{,}79;\quad 200{,}21\right]$$

quindi:

$$P\left(199{,}79 \le \mu \le 200{,}21\right) = 0{,}95$$

b) Popolazione normale, varianza non nota, n > 30

$\bar{x} = 200$
$s = 0{,}5$
$n = 100$
$\alpha = 0{,}05$
$\alpha/2 = 0{,}025$

L'intervallo va determinato come:

$$IC = \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$z_{\alpha/2} = z_{0,025} = \pm\ 1{,}96$

$$IC_{0,95} = \left[\bar{x} \pm z_{0,025}\frac{s}{\sqrt{n}}\right] = \left[200 - 1,96\frac{0,5}{\sqrt{100}};\quad 200 + 1,96\frac{0,5}{\sqrt{100}}\right] =$$
$$= \left[199,9;\quad 200,1\right]$$

quindi:

$$P\left(199,9 \le \mu \le 200,1\right) = 0,95$$

Esercizio 5

Un campione di 20 automobilisti che percorrono un certo tratto di strada fa registrare una velocità media di 101 Km/h e uno scarto quadratico medio di 6,5 km/h. Ipotizzando che la velocità si distribuisce normalmente:

a) costruire un intervallo di confidenza al 95% per la velocità media delle auto che percorrono quel tratto;
b) costruire un intervallo di confidenza al 99%;
c) costruire gli stessi intervalli supponendo che quei valori di media e s.q.m. siano stati registrati su un campione di 200 automobilisti.

Soluzione

a) Popolazione normale, varianza non nota, n < 30

$\bar{x} = 101$ km/h
$\hat{\sigma} = 6{,}5$ km/h
$n = 20$
$\alpha = 0{,}05$
$\alpha/2 = 0{,}025$

$$P\left(\bar{x} - t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n-1}}\right) = 1-\alpha$$

quindi:

$$IC_{1-\alpha} = \bar{x} \pm t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n-1}}$$

$t_{\alpha/2;n-1} = t_{0,025;19} = 2{,}09$

$$IC_{0,95} = 101 \pm 2{,}09\frac{6{,}5}{\sqrt{19}} = \left[101 - 2{,}09\frac{6{,}5}{\sqrt{19}};101 + 2{,}09\frac{6{,}5}{\sqrt{19}}\right] =$$
$$= [97{,}88;104{,}12]$$

quindi:

$$P(97{,}88 \leq \mu \leq 104{,}12) = 0{,}95$$

b) Popolazione normale, varianza non nota, n < 30

$\bar{x} = 101$ km/h
$\hat{\sigma} = 6{,}5$ km/h
$n = 20$
$\alpha = 0{,}01$
$\alpha/2 = 0{,}005$

$$P\left(\bar{x} - t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n-1}}\right) = 1-\alpha$$

quindi:

$$IC_{1-\alpha} = \bar{x} \pm t_{\alpha/2;n-1} \frac{\hat{\sigma}}{\sqrt{n-1}}$$

$t_{\alpha/2;n-1} = t_{0,005;19} = 2,86$

$$IC_{0,99} = 101 \pm 2,86 \frac{6,5}{\sqrt{19}} = \left[101 - 2,86 \frac{6,5}{\sqrt{19}} ; 101 + 2,86 \frac{6,5}{\sqrt{19}}\right] =$$
$$= [96,73;105,26]$$

quindi:

$$P(96,73 \leq \mu \leq 105,26) = 0,99$$

c) Popolazione normale, varianza non nota, n > 30

$\bar{x} = 101$ km/h
$\hat{\sigma} = 6,5$ km/h
$n = 200$

$$IC = \bar{x} \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

1.
$\alpha = 0,05; \quad \alpha/2 = 0,025$

$Z_{0,025} = \pm 1,96$

$$IC_{0,95} = 101 \pm 1,96 \frac{6,5}{\sqrt{200}} = \left[101 - 1,96 \frac{6,5}{\sqrt{200}} ; 101 + 1,96 \frac{6,5}{\sqrt{200}}\right] =$$
$$= [100,1;101,9]$$

quindi:

$$P(100,1 \leq \mu \leq 101,9) = 0,95$$

2.
$\alpha = 0,01; \quad \alpha/2 = 0,005$

$Z_{0,005} = \pm 2,58$

$$IC_{0,99} = 101 \pm 2,58 \frac{6,5}{\sqrt{200}} = \left[101 - 2,58 \frac{6,5}{\sqrt{200}} ; 101 + 2,58 \frac{6,5}{\sqrt{200}}\right] =$$
$$= [99,81;102,19]$$

quindi:

$$P(99,8 \leq \mu \leq 102,2) = 0,99$$

ALLEGATO

Schema per la costruzione di un intervallo di confidenza per la media, al livello di significatività 1 - α

Varianza σ^2 nota	Varianza σ^2 non nota	

Popolazione normale

$IC = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	s.q.m. campionario corretto s **n < 30:** $IC = \bar{x} \pm t_{\alpha/2;n-1} \frac{s}{\sqrt{n}}$ **n > 30:** $IC = \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$	s.q.m. campionario non corretto $\hat{\sigma}$ **n < 30:** $IC = \bar{x} \pm t_{\alpha/2;n-1} \frac{\hat{\sigma}}{\sqrt{n-1}}$ **n > 30:** $IC = \bar{x} \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$

Popolazione non nota, n > 30

$IC = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	s.q.m. campionario corretto s $IC = \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$	s.q.m. campionario non corretto $\hat{\sigma}$ $IC = \bar{x} \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$

Popolazione non nota, n < 30

$P\left(\left\lvert\bar{x} - \mu\right\rvert \leq k\sigma\right) \geq 1 - \frac{1}{k^2}$	Non determinabile