Stima Parametrica Intervallare

Università Mediterranea di Reggio Calabria
Decisions Lab

Uno stimatore per intervallo per un parametro di una popolazione P, è una funzione delle variabili campionarie (campione causale): determina gli estremi di un intervallo di valori che *verosimilmente* contiene il parametro da stimare. La stima corrispondente viene chiamata stima per intervallo.

Problemi stima puntuale:

- Lo stimatore puntuale è una singola statistica che però varia da campione a campione, quindi dipende dagli elementi selezionati;
- Non rispetta il teorema del limite centrale, infatti la stima puntuale della proporzione di pezzi difettosi in un carico, è la stessa sia osservando 1 pezzo in proporzione a 10 pezzi, sia osservando 100 pezzi in proporzione di 1000 pezzi.

Con la stima intervallare

- si considerano deviazioni positive e negative dal vero valore, fornendo un *intervallo di valori* per la stima del parametro incognito;
- Rispetta il teorema del limite centrale: a parità di condizioni, campioni più grandi determinano stime per intervallo più precise

Si supponga venga estratto un campione casuale da una popolazione P, sulla base delle informazioni campionarie si supponga sia possibile determinare due v.a.

$$L_1 = L_1(X_1, \ldots, X_n); \quad L_2 = L_2(X_1, \ldots, X_n)$$

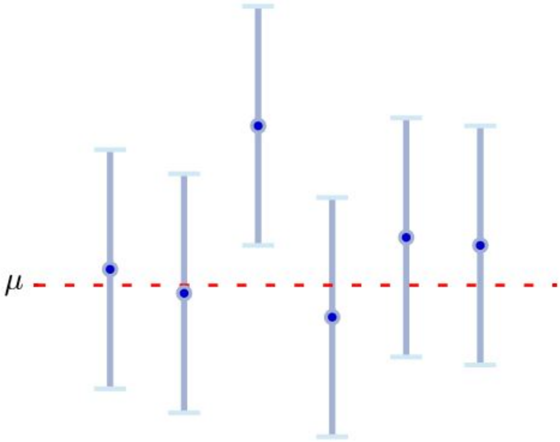
con $L_1 < L_2$.
Siano l_1 ed l_2 i valori assunti da L_1 ed L_2: l'intervallo $[l_1, l_2]$ contiene o non contiene il parametro stimato.

Si estraggano ripetutamente campioni casuali dalla popolazione determinando più intervalli corrispondenti a L_1 ed L_2, allora **solo una certa percentuale di questi intervalli (**95%**,** 99% **etc) conterrà il valore incognito**.

Se si estraggono successivamente più campioni indipendenti dalla stessa popolazione e si determina gli intervalli (di confidenza), il k% di intervalli conterrà il vero valore del parametro incognito.

L'intervallo $[L_1, L_2]$ è definito uno stimatore per intervallo a *livello di confidenza* k% per il parametro.

95% degli intervalli contiene il parametro μ

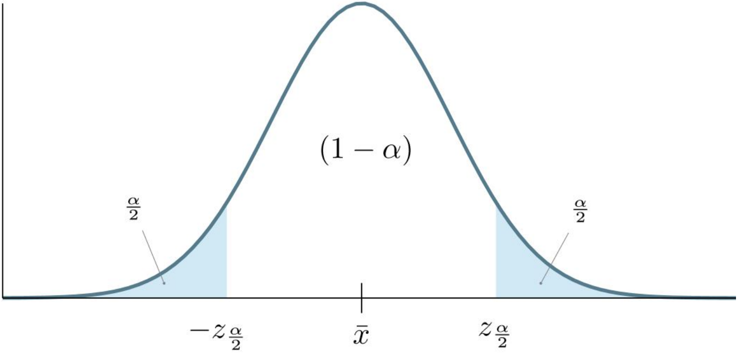


Stimatore per intervallo e intervallo di confidenza

Uno **stimatore per intervallo** per un parametro di una popolazione è una funzione delle variabili campionarie: determina gli estremi di un intervallo di valori che verosimilmente contiene il parametro da stimare. La stima corrispondente viene chiamata **stima per intervallo.**

Sia θ un parametro incognito. Si ottiene lo stimatore per intervallo a livello $1-\alpha$ per θ se, sulla base delle informazioni campionarie, si possono determinare due variabili aleatorie A e B, $A < B$, tali che $P(A < \theta < B) = 1 - \alpha$, con $0 \leq \alpha \leq 1$. Se a e b sono i valori osservati per A e B in corrispondenza al campione considerato, l'intervallo da a a b, la stima per intervallo, viene definito **intervallo di confidenza** a livello $100(1-\alpha)\%$ per il parametro θ. La quantità $100(1-\alpha)\%$ viene definita **livello di confidenza** dell'intervallo.

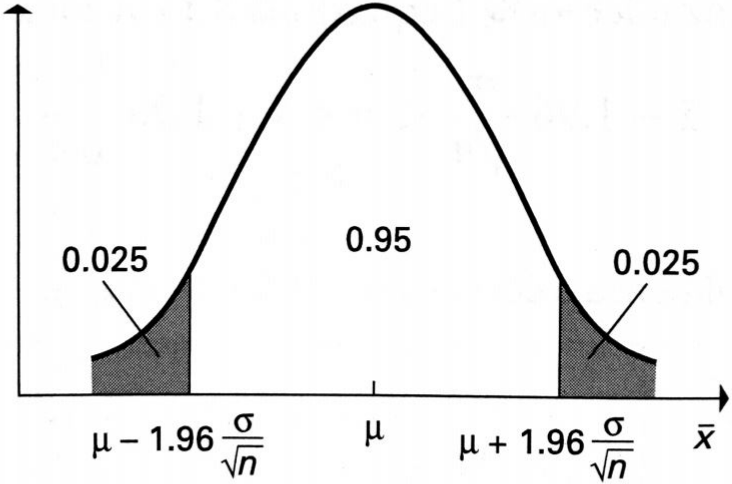
Se si estrae ripetutamente e in modo indipendente un numero molto elevato di campioni, il vero valore del parametro θ sarà contenuto nel $100(1-\alpha)\%$ degli intervalli determinati in questo modo. L'intervallo di confidenza così ottenuto viene indicato come $a < \theta < b$ con livello di confidenza $100(1-\alpha)\%$.



$(1-\alpha)$
$\frac{\alpha}{2}$
$\frac{\alpha}{2}$
$-z_{\frac{\alpha}{2}}$
$\bar{x}$
$z_{\frac{\alpha}{2}}$

Considerando una normale standard $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).
\end{aligned}
$$



0.025
0.95
0.025
$\mu - 1.96\frac{\sigma}{\sqrt{n}}$
μ
$\mu + 1.96\frac{\sigma}{\sqrt{n}}$
$\bar{x}$

Intervalli di confidenza per la media di una popolazione distribuita normalmente: varianza nota

Si consideri un campione casuale di n osservazioni estratto da una popolazione normale con media μ e varianza σ^2. Se $\overline{x}$ è il valore osservato della media campionaria, un **intervallo di confidenza per la media della popolazione con varianza nota**, a livello $100(1-\alpha)\%$, è dato da:

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.1}$$

o, in modo equivalente:

$$\overline{x} \pm ME$$

dove ME, **margine di errore** (anche chiamato *errore di campionamento*) è dato da:

$$ME = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.2}$$

L'**ampiezza**, *w*, è uguale al doppio del margine di errore:

$$w = 2(ME) \tag{8.3}$$

Il **limite superiore dell'intervallo di confidenza** (**UCL**, *upper confidence limit*) è dato da:

$$UCL = \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.4}$$

Il **limite inferiore dell'intervallo di confidenza** (**LCL**, *lower confidence limit*) è dato da:

$$LCL = \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.5}$$

Livello di confidenza	90%	95%	98%	99%
α	0.10	0.05	0.02	0.01
$Z_{\alpha/2}$	1.645	1.96	2.33	2.58

Zucchero raffinato (Intervallo di confidenza)

Un processo produce sacchi di zucchero raffinato. Il loro peso è distribuito normalmente, con deviazione standard pari a 0.6 Kg. Il contenuto medio di un campione casuale di 25 sacchi è 10 Kg. Trovare il limite superiore e il limite inferiore dell'intervallo di confidenza, a livello 99%, per il peso medio di tutti i sacchi prodotti dal processo.

Soluzione

Per un intervallo di confidenza a livello 99%, il fattore di affidabilità è:

$$z_{0.005} = 2.58$$

e, con media campionaria $\bar{x} = 10$, $n = 25$ e $\sigma = 0.6$, i limiti dell'intervallo di confidenza sono:

Limite superiore: $$\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 + 2.58\frac{0.6}{\sqrt{25}} = 10.31$$

Limite inferiore: $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 - 2.58\frac{0.6}{\sqrt{25}} = 9.69$$

Sia X una variabile casuale $X \sim N(\mu, \sigma^2)$ dove σ^2 assumiamo sia nota.

Esempio.

L'intervallo di confidenza per μ al 95% con ampiezza campionaria pari a $n = 64$, è

$$\left(\bar{x} - 1.96\sqrt{\frac{\sigma}{64}}, \bar{x} + 1.96\sqrt{\frac{\sigma}{64}}\right).$$

L'ampiezza dell'intervallo è

$$2 \cdot 1.96\frac{\sigma}{8}$$

Passiamo adesso ad un intervallo con livello di confidenza 99%, dove $z_{0.995} = 2.58$. In generale avremo

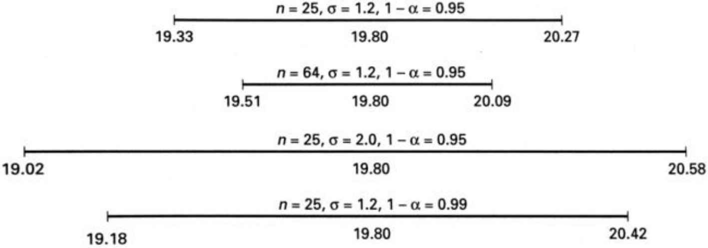
$$2 \cdot 2.58\frac{\sigma}{\sqrt{n}}.$$

Il problema è mantenere la stessa ampiezza dei due intervalli al variare del livello di confidenza

$$2 \cdot 1.96\frac{\sigma}{8} = 2 \cdot 2.58\frac{\sigma}{\sqrt{n}},$$

che, risolvendo, si può ottenere aumentando l'ampiezza campionaria a $n = 111$.

Effetti prodotti sulla ampiezza dell'intervallo di confidenza da: n, σ e livello di confidenza



	Stima puntuale	Stima intervallare
Campione casuale	$X_1, \ldots, X_n$	$X_1, \ldots, X_n$
Obiettivo	Stima puntuale del parametro θ	Stima per intervallo del parametro θ
Strumento	Stimatore puntuale: $T = T(X_1, \ldots, X_n)$	Stimatore intervallo di confidenza: $[L_1, L_2] = [L_1(X_1, \ldots, X_n), L_2(X_1, \ldots, X_n)]$
Accuratezza	Errore quadratico medio $MSE(T) = E[(T - \theta)^2]$	Livello di confidenza $P(L_1 \leq \theta \leq L_2) = 1 - \alpha$
Dati campionari	$x_1, \ldots, x_n$	$x_1, \ldots, x_n$
Risultato	Stima puntuale: $t = T(x_1, \ldots, x_n)$	Intervallo di confidenza stimato: $[l_1, l_2] = [L_1(x_1, \ldots, x_n), L_2(x_1, \ldots, x_n)]$