

STIMA DEI PARAMETRI

STIMA PUNTUALE DEI PARAMETRI

Per STIMA PUNTUALE DEI PARAMETRI s'intende l'insieme dei metodi inferenziali che permettono di attribuire un valore ad un parametro della popolazione, utilizzando i dati di un campione casuale osservato (x_1, x_2, \dots, x_n) ed elaborandoli.

Sia (x_1, x_2, \dots, x_n) il campione casuale osservato e sia $t(\bullet)$ una funzione matematica a n variabili. Applicando $t(\bullet)$ al campione casuale osservato si ottiene un valore:

$$t(\bullet) = t(x_1, x_2, \dots, x_n)$$

Si definisce **STIMA** il risultato t delle elaborazioni dei dati di un campione casuale osservato, al fine di stimare il parametro di interesse θ della popolazione.

Si definisce **STIMATORE** del parametro θ della popolazione la *statistica campionaria* $T = t(X_1, X_2, \dots, X_n)$ utilizzata per stimare θ .

Si chiama *errore di stima* la differenza fra lo STIMATORE e il parametro: **$T - \theta$**

Proprietà degli STIMATORI

A) CORRETTEZZA o NON DISTORSIONE

Uno stimatore T si definisce CORRETTO o NON DISTORTO se la sua media coincide con il parametro da stimare:

$$M(T) = \theta$$

Se $M(T) \neq \theta$, allora T si definisce stimatore *non corretto* o *distorto*: la quantità $D = M(T) - \theta$ si dice *distorsione*.

B) EFFICIENZA

Il concetto di *efficienza* di uno stimatore si basa sulla sua variabilità: uno stimatore sarà tanto più efficiente, ovvero preciso, quanto più piccola risulta la sua Varianza, e quindi il suo S.q.m.

Uno stimatore è tanto più efficiente quanto più piccolo è il suo errore medio $\sigma_T = \sqrt{\text{Var}(T)}$.

EFFICIENZA ASINTOTTICA

Se la varianza dello stimatore tende a zero per n (dimensione campionaria) che tende all'infinito:

$$\lim_{n \rightarrow +\infty} \text{Var}(T) = 0$$

allora T si definisce stimatore asintoticamente efficiente.

C) CONSISTENZA

Si definisce T stimatore CONSISTENTE del parametro θ se e solo se vale la seguente relazione:

$$\lim_{n \rightarrow +\infty} P(|T - \theta| < \varepsilon) = 1$$

Teorema: Se uno stimatore T è corretto e asintoticamente efficiente, allora T è stimatore consistente di θ .

STIMA DEI PARAMETRI

STIMA PER INTERVALLO

La tecnica inferenziale della STIMA PER INTERVALLO ha lo scopo di fissare un intervallo entro il quale si trovi, con la probabilità assegnata ($1-\alpha$), il valore vero del parametro θ incognito della popolazione.

Se questa probabilità $1-\alpha$ è vicina ad *uno*, allora si ottiene un intervallo che contiene quasi certamente il parametro. Invece la probabilità di estrarre un campione il cui intervallo non contenga il parametro è α e quindi è piccola: α esprime la misura del rischio di errore, ovvero la probabilità di sbagliare.

INTERVALLO DI CONFIDENZA o *INTERVALLO FIDUCIARIO*

$$1 - \alpha = P\{\theta_1 < \theta < \theta_2\}$$

La quantità $(1-\alpha)$ è il Coefficiente di confidenza.

Gli estremi θ_1 e θ_2 sono i Limiti di confidenza.

INTERVALLO DI CONFIDENZA per la Media μ della popolazione

$$1 - \alpha = P\{\mu_1 < \mu < \mu_2\}$$

Se la Media campionaria \bar{X} si distribuisce NORMALMENTE, allora

$$1 - \alpha = P\left\{-z_{\frac{\alpha}{2}} < z < +z_{\frac{\alpha}{2}}\right\}$$

Dove $z = z_m = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ (nel caso di campione casuale **CON** ripetizione), quindi sostituendo si ottiene:

$$1 - \alpha = P\left\{-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < +z_{\frac{\alpha}{2}}\right\}$$

Da cui, effettuando alcuni passaggi algebrici ai fini di *isolare* μ , si ottiene la formula finale dell' **INTERVALLO DI CONFIDENZA** per la Media μ della popolazione:

$$1 - \alpha = P\left\{\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right\}$$

Invece, nel caso di campione casuale **SENZA** ripetizione e popolazione di dimensione N finita, nella formula della $Var(\bar{X})$ bisogna tener conto del fattore di correzione, quindi:

$$z = z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Da sostituire nella formula:

$$1 - \alpha = P\{\mu_1 < \mu < \mu_2\} = P\left\{-z_{\frac{\alpha}{2}} < z < +z_{\frac{\alpha}{2}}\right\} = P\left\{-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} < +z_{\frac{\alpha}{2}}\right\}$$

Da cui, effettuando alcuni passaggi algebrici ai fini di *isolare* μ , si ottiene la formula finale dell' *INTERVALLO DI CONFIDENZA* per la Media μ della popolazione:

$$1 - \alpha = P\left\{\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}\right\}$$

Se la Media campionaria \bar{X} standardizzata si distribuisce come una v.c. *t di Student*, allora

$$1 - \alpha = P\left\{-t_{\frac{\alpha}{2}} < t < +t_{\frac{\alpha}{2}}\right\}$$

Dove $t = t_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$, quindi sostituendo si ottiene:

$$1 - \alpha = P\left\{-t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} < +t_{\frac{\alpha}{2}}\right\}$$

Da cui, effettuando alcuni passaggi algebrici ai fini di *isolare* μ , si ottiene la formula finale dell' *INTERVALLO DI CONFIDENZA* per la Media μ della popolazione:

$$1 - \alpha = P\left\{\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right\}$$

Esercizio Da una popolazione con varianza $\sigma^2=49$ è stato estratto con reinserimento un campione casuale di $n=100$ unità; sia 50 il valore della *media campionaria*.
Determinare l'intervallo di confidenza al livello del 92% per la media della popolazione.

SOLUZIONI

Poiché $n=100>30$ allora la *media campionaria* si distribuisce normalmente.
Dato un livello di significatività pari al $92\%=1-\alpha$, segue $\alpha=8\%$ e $\alpha/2=4\%$; quindi sulle tavole della v.c. Normale standardizzata bisogna individuare i valori

$$\pm z_{\frac{\alpha}{2}} = \pm z_{4\%} = \pm 1,75$$

$$1-\alpha = P\left\{\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right\}$$

$$1-\alpha = P\left\{50 - 1,75 \cdot \frac{7}{\sqrt{100}} < \mu < 50 + 1,75 \cdot \frac{7}{\sqrt{100}}\right\}$$

$$0,92 = P\{48,775 < \mu < 51,225\}$$

Esercizio Da un campione casuale di 35 impiegati ($n=35$) risulta che la retribuzione mensile (in migliaia di euro) presenta *media campionaria* pari a 2 e *varianza campionaria corretta* pari a 0,25.

Determinare l'intervallo di confidenza al livello del 96% per la media della popolazione.

SOLUZIONI

Poiché $n=35>30$ allora la *media campionaria* si distribuisce normalmente.

$$1-\alpha = P\left\{\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right\}$$

$$1-\alpha = P\left\{2 - 2,05 \cdot \frac{0,5}{\sqrt{35}} < \mu < 2 + 2,05 \cdot \frac{0,5}{\sqrt{35}}\right\}$$

$$0,96 = P\{1,82674 < \mu < 2,17326\}$$

INTERVALLO DI CONFIDENZA per la Varianza σ^2 della popolazione, nel caso di popolazione Normale

$$1 - \alpha = P\{\sigma_1^2 < \sigma^2 < \sigma_2^2\}$$

Poiché, se la popolazione d'origine è Normale la Varianza campionaria S^2 si distribuisce come una v.c. *chi-quadrato*, a meno di una coefficiente di proporzionalità, allora si può scrivere:

$$1 - \alpha = P\left\{\chi^2_{\left(\frac{\alpha}{2}\right)} < \chi^2 < \chi^2_{\left(1-\frac{\alpha}{2}\right)}\right\}$$

Dove $\chi^2 = \frac{S^2 \cdot v}{\sigma^2}$, quindi sostituendo si ottiene:

$$1 - \alpha = P\left\{\chi^2_{\left(\frac{\alpha}{2}\right)} < \frac{S^2 \cdot v}{\sigma^2} < \chi^2_{\left(1-\frac{\alpha}{2}\right)}\right\}$$

Da cui, effettuando alcuni passaggi algebrici ai fini di *isolare* σ^2 , si ottiene la formula finale dell' **INTERVALLO DI CONFIDENZA** per la Varianza σ^2 della popolazione:

$$1 - \alpha = P\left\{\frac{S^2 \cdot v}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}} < \sigma^2 < \frac{S^2 \cdot v}{\chi^2_{\left(\frac{\alpha}{2}\right)}}\right\}$$

Esercizio Da una popolazione distribuita normalmente viene estratto un campione casuale di 10 unità (: n=10); la varianza calcolata sul campione risulta pari a 600.
Determinare l'intervallo di confidenza al livello del 95% per la varianza della popolazione.

SOLUZIONI

$$1 - \alpha = P \left\{ \frac{S^2 \cdot \nu}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}} < \sigma^2 < \frac{S^2 \cdot \nu}{\chi^2_{\left(\frac{\alpha}{2}\right)}} \right\}$$

Dove $95\% = 1 - \alpha$, da cui segue $\alpha = 5\%$ e $\alpha/2 = 2,5\%$; quindi sulle tavole della v.c. "chi-quadrato" bisogna individuare i seguenti valori, in corrispondenza a $\nu = n - 1 = 10 - 1 = 9$:

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{2,5\%} = 2,70$$

$$\chi^2_{\left(1-\frac{\alpha}{2}\right)} = \chi^2_{97,5\%} = 19,02$$

La varianza del campione è pari a 600, quindi bisogna effettuare la *correzione* per ottenere S^2 :

$$S^2 = 600 \cdot \frac{n}{n-1} = 600 \cdot \frac{10}{9} = 666,67$$

Da cui, sostituendo:

$$\begin{aligned} 1 - \alpha = 0,95 &= P \left\{ \frac{666,67 \cdot 9}{19,02} < \sigma^2 < \frac{666,67 \cdot 9}{2,70} \right\} = \\ &= P \{ 315,45 < \sigma^2 < 2222,22 \} \end{aligned}$$