Stimatori e Intervalli di Confidenza

1 Esercizi sugli stimatori

1.1

Da una v.c. X di media μ e varianza $\sigma^2 = 1$ viene estratto un campione di ampiezza n. Siano T_1 e T_2 due stimatori per μ tali che:

$$T_1 = X_1 + X_2 - X_n$$
$$T_2 = \frac{X_1 + X_2}{2} - X_n$$

1. Calcolare la distorsione di entrambi gli stimatori

2. Determinare la varianza di entrambi gli stimatori

3. Indicare per quali valori della media μ T_2 risulta più efficiente di T_1.

Soluzione

1. Si definisce distorsione di uno stimatore:

$$bias(\hat{\theta}) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

Ne consegue:

$$\begin{aligned}
bias(T_1) &= E(T_1) - \mu \\
&= E(X_1 + X_2 - X_n) - \mu \\
&= E(X_1) + E(X_2) - E(X_n) - \mu \\
&= \mu + \mu - \mu - \mu = 0
\end{aligned}$$

$$
\begin{aligned}
bias(T_2) &= E(T_2) - \mu \\
&= E\left(\frac{X_1 + X_2}{2} - X_n\right) - \mu \\
&= \frac{1}{2}\left[E(X_1) + E(X_2) - 2E(X_n)\right] - \mu \\
&= \frac{1}{2}(\mu + \mu - 2\mu) - \mu = -\mu
\end{aligned}
$$

2. Varianza

$$
\begin{aligned}
V(T_1) &= V(X_1 - X_2 - X_n) \\
&= V(X_1) + V(X_2) + V(X_n) \\
&= 1 + 1 + 1 = 3
\end{aligned}
$$

$$
\begin{aligned}
V(T_2) &= V\left(\frac{X_1 + X_2}{2} - X_n\right) \\
&= V\left(\frac{X_1 + X_2}{2}\right) + V(X_n) \\
&= \frac{1}{4}\left[V(X_1) + V(X_2)\right] + V(X_n) \\
&= \frac{2}{4} + 1 = \frac{3}{2}
\end{aligned}
$$

3. Si stabilisce se uno stimatore è più o meno efficiente rispetto ad un altro stimatore in base al rapporto:

$$e(T_1, T_2) = \frac{MSE(T_2)}{MSE(T_1)} \quad \text{con} \quad MSE(T_i) = E\left[(T_i - \theta)^2\right] = V(T_i) + bias(T_i)^2$$

Ne consegue:

$$
\begin{aligned}
MSE(T_1) &= V(T_1) + bias(T_1)^2 \\
&= 3 + 0 = 3
\end{aligned}
$$

$$
\begin{aligned}
MSE(T_2) &= V(T_2) + bias(T_2)^2 \\
&= \frac{3}{2} + \mu^2
\end{aligned}
$$

$$
\begin{aligned}
e(T_1, T_2) &= \frac{MSE(T_2)}{MSE(T_1)} \\
&= \frac{\frac{3}{2} + \mu^2}{3} \\
&= \frac{3 + 2\mu^2}{6}
\end{aligned}
$$

Ora, se T_2 è più efficinete di T_1 allora $e(T_1, T_2) < 1$. Pertanto:

$$
\begin{aligned}
&\frac{3 + 2\mu^2}{6} < 1 \\
&\mu^2 < \frac{3}{2} \\
&\text{ovvero per } -\sqrt{\frac{3}{2}} < \mu < \sqrt{\frac{3}{2}}
\end{aligned}
$$

1.2

Si supponga di voler stimare il valore atteso di una variabile X. Si considerino tre stimatori T_1, T_2 e T_3 con le seguenti distribuzioni di probabilità:

$$
\begin{aligned}
&P_{T_1}(t_1) = \frac{1}{3} \quad \text{per} \quad t_1 = 1, 3, 5 \\
&P_{T_2}(t_2) = 1 \quad \text{per} \quad t_2 = 3 \\
&P_{T_3}(t_3) = \frac{t_3}{10} \quad \text{per} \quad t_3 = 1, 2, 3, 4
\end{aligned}
$$

Sapendo che $E(X) = 4$, calcolare la distorsione dei tre stimatori.

Soluzione

Si definisce distorsione di uno stimatore:

$$
bias(\hat{\theta}) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta
$$

Occorre dunque calcolare il valore atteso dei tre stimatori. Riportando le informazioni a disposizione in forma tabellare, si ha:

$T_1 = t_1$	$P(T_1 = t_1)$
1	$\frac{1}{3}$
3	$\frac{1}{3}$
5	$\frac{1}{3}$
	1

$T_2 = t_2$	$P(T_2 = t_2)$
3	1
	1

$T_3 = t_3$	$P(T_3 = t_3)$
1	$\frac{1}{10}$
2	$\frac{2}{10}$
3	$\frac{3}{10}$
4	$\frac{4}{10}$
	1

Passando al calcolo del valore atteso di ogni stimatore si ottiene:

$$E(T_1) = 1 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} = 3$$
$$E(T_2) = 3 \cdot 1 = 3$$
$$E(T_3) = 1 \cdot \frac{1}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{3}{10} + 4 \cdot \frac{4}{10} = 3$$

Ne consegue che, essendo $\mu = E(X) = 4(= \theta)$:

$$bias(T_1) = E(T_1) - \theta = 3 - 4 = -1$$
$$bias(T_2) = E(T_2) - \theta = 3 - 4 = -1$$
$$bias(T_3) = E(T_3) - \theta = 3 - 4 = -1$$

2 Esercizi sugli Intervalli di Confidenza

2.1

Una partita di tondini presenta un diametro X medio μ incognito e una deviazione standard pari a $\sigma = 0.1$. Si estrae un campione di ampiezza $n = 1000$ sui quali si osserva $\bar{x} = 1.2$.

1. Determinare l' IC per μ al 99%.

2. Determinare l'ampiezza dell'intervallo.

Soluzione

1. Dato che è nota la deviazione standard e ricordando che $\bar{X} \sim N(\mu, \sqrt{\frac{\sigma^2}{n}})$ si ha che:

$$P\left(z_1 < \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < z_2\right) = 1 - \alpha$$

Ovvero:

$$P\left(z_1 \cdot \sqrt{\frac{\sigma^2}{n}} < \bar{x} - \mu < z_2 \cdot \sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha,$$

$$P\left(\bar{x} - z \cdot \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{x} + z \cdot \sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$$

dove, per la simmetria della v.c. Normale si è posto $z = |z_1| = z_2$. Ora, dato che nel caso in esame $0.99 = 1 - \alpha$ si ricava $\alpha = 0.01$ Occorre ora trovare il quantile $z_{\frac{\alpha}{2}} = z_1$ della v.c. Normale (0,1):

$$P(Z < z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} = 0.005$$

Utilizzando la funzione MatLab norminv(0.005) o la funzione in Excel INV.NORM(0,005;0;1) si ricava:

$$z = |-2.5758 = z_1| = 2.5758 = z_2$$

Pertanto:

$$IC^{-} = \bar{x} - z \cdot \sqrt{\frac{\sigma^2}{n}} = 1.2 - 2.5758 \cdot \frac{0.1}{\sqrt{1000}} = 1.1919$$
$$IC^{+} = \bar{x} + z \cdot \sqrt{\frac{\sigma^2}{n}} = 1.2 + 2.5758 \cdot \frac{0.1}{\sqrt{1000}} = 1.2081$$

In virtù di tali risultati, la lunghezza dell'IC, $\ell(IC_\alpha)$, è data da dalla differenza tra gli estremi dell'intervallo: $\ell(IC_\alpha) = 1.2081 - 1.1919 = 0.016$.

2.2

Il tempo X impiegato per percorrere la distanza tra un punto A ed un punto B segue una distribuzione Normale di parametri μ e σ^2 ignoti. Da un campione casuale di $n = 8$ osservazioni, si sono registrati i seguenti tempi in secondi:

$$298.0, 295.0, 306.0, 306.0, 283.0, 299.0, 297.0, 313.0$$

Determinare un intervallo di confidenza per la media μ al livello di fiducia $1 - \alpha = 0.95$.

Soluzione

1. Dato che non è nota la varianza si usa la statistica

$$\frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \quad \text{con} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

dove è noto che $\frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}} \sim T - di\ Student(gdl = n-1)$ e s^2 è lo stimatore non distorto della varianza. Segue pertanto che:

$$P\left(t_1^{(n-1)} < \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} < t_2^{(n-1)}\right) = 1 - \alpha$$

Ovvero:

$$P\left(t_1^{(n-1)} \cdot \sqrt{\frac{s^2}{n}} < \bar{x} - \mu < t_2^{(n-1)} \cdot \sqrt{\frac{s^2}{n}}\right) = 1 - \alpha,$$

$$P\left(\bar{x} - t \cdot \sqrt{\frac{s^2}{n}} < \mu < \bar{x} + t \cdot \sqrt{\frac{s^2}{n}}\right) = 1 - \alpha$$

dove, per la simmetria della v.c. T-di-Student si è posto $t = |t_1^{(n-1)}| = t_2^{(n-1)}$. Ora, dato che nel caso in esame $0.95 = 1 - \alpha$ si ricava $\alpha = 0.05$. In analogia all'esercizio precedente, occorre trovare il quantile della v.c. T-di Student tale che;

$$P\left(T < t_{\frac{\alpha}{2}}^{(n-1)} = t_1^{(n-1)}\right) = \frac{\alpha}{2}$$

Utilizzando la funzione MatLab tinv(0.025,gdl) o la funzione in Excel INV.T.2T(0,025*2;gdl) si ricava:

$$t = |t_1^{(n-1)}| = t_2^{(n-1)} = 2.3646$$

Dai dati a disposizione si ricava:

$$\bar{x} = 299.6250 \quad \text{e} \quad s^2 = 81.1242$$

Pertanto:

$$IC^- = \bar{x} - t \cdot \sqrt{\frac{s^2}{n}} = 299.6250 - 2.3646 \cdot \sqrt{\frac{81.1242}{8}} = 292.0951$$

$$IC^+ = \bar{x} + t \cdot \sqrt{\frac{s^2}{n}} = 299.6250 + 2.3646 \cdot \sqrt{\frac{81.1242}{8}} = 307.1549$$

2.3

L'allenatore di una squadra di salto in lungo, per valutare i propri allievi, rileva, per ognuno di loro, il salto misurato nell'ultima gara (espresso in metri). Di seguito si riportano i valori registrati da un campione di 8 allievi estratti a caso:

$$4.72, 3.42, 2.30, 3.70, 3.17, 4.63, 5.26, 3.22$$

Supponendo che i dati campionari provengano da una distribuzione normale, si determini l'intervallo di confidenza al 95% per l'ignota varianza.

Soluzione

Per determinare l'intervallo di confidenza per la varianza occorre ricordare che:

$$\frac{S^2 \cdot (n-1)}{\sigma^2} \sim \chi^2(gdl = n-1)$$

Si ha pertanto che:

$$P\left(\chi_1^{(n-1)} < \frac{s^2 \cdot (n-1)}{\sigma^2} < \chi_2^{(n-1)}\right) = 1 - \alpha$$

Dato che la distribuzione χ^2 non è simmetrica occorre calcolare entrambi i quantili $\chi_1^{(n-1)}$ e $\chi_2^{(n-1)}$ dove, più precisamente:

$$\chi_1^{(n-1)} = \chi_{\frac{\alpha}{2}}^{(n-1)} \quad \text{e} \quad \chi_2^{(n-1)} = \chi_{1-\frac{\alpha}{2}}^{(n-1)}$$

Anche in questo caso, detti quantili si ricavano utilizzando la funzione MatLab chi2inv(p,gdl) o la funzione Excel INV.CHI.QUAD(p;gdl). Nel caso in esame:

$$\chi_1^{(n-1)} = 1.6899 \quad \text{e } \chi_2^{(n-1)} = 16.0128$$

L'IC che si viene così a configurare è dunque:

$$IC = \left[\frac{s^2 \cdot (n-1)}{\chi_{1-\frac{\alpha}{2}}^{(n-1)}}, \frac{s^2 \cdot (n-1)}{\chi_{\frac{\alpha}{2}}^{(n-1)}}\right]$$

Dai dati a disposizione si ricava:

$$\bar{x} = 3.8025 \quad \text{e} \quad s^2 = 0.9721$$

e pertanto IC=(0.4249,4.0267).

2.4

Il responsabile del personale di un'azienda vuole determinare il numero medio di giorni di assenza dal lavoro da parte degli impiegati di uno stabilimento, mediante un'indagine campionaria che ammette un errore di 1 giorno, a livello di confidenza del 95%. Da precedenti indagini la deviazione standard risulta essere pari a 10 giorni. Determinare la dimensione del campione da esaminare.

Soluzione

Dai dati in esame e dalla teoria è noto che

$$n > \left(\frac{z \cdot \sigma}{\epsilon}\right)^2$$

dove $z = |z_1 = z_{\frac{\alpha}{2}}| = z_2 = z_{1-\frac{\alpha}{2}}$ ed ϵ che indica il margine di errore. Tramite la funzione MatLab norminv(p) o la funzione Excel INV.NORM(p;0;1) si ricava $z = 1.96$ e pertanto:

$$n > \left(\frac{1.96 \cdot 10}{1}\right)^2 = 384.16 \simeq 385$$