Statistica

A. Iodice

I. C. sulla media
Esercizio 1
Esercizio 2
Esercizio 3

I. C. sulla proporzione
Esercizio 4

I. C. su somme e differenze tra medie
Esercizio 5
Esercizio 6
Esercizio 7

Statistica

Esercitazione 13

Alfonso Iodice D'Enza
iodicede@unicas.it

Università degli studi di Cassino

Ex.1: I. C. sulla media

Si suppponga di dover comunicare un segnale da una sorgente A ad una destinazione B. Il segnale emesso ha intensità μ. L'intensità del segnale viene percepito in B secondo una distribuzione Normale con media μ e deviazione standard $\sigma = 3$. Quindi, per effetto dei disturbi alla tramissione, l'intensità del segnale in B differisce da quella in A con media $\mu = 0$ e scarto quadratico medio $\sigma = 3$. Si consideri di aver effettuato $n = 10$ trasmissioni tra A e B registrando l'intensità a destinazione.

$$\{17, 21, 20, 18, 19, 22, 20, 21, 16, 19\}$$

- costruire l'intervallo di confidenza all 95%
- costruire l'intervallo di confidenza all 90%
- costruire l'intervallo di confidenza all 99%

Ex.1: I. C. sulla media

Svolgimento

Si consideri lo stimatore media campionaria $\bar{X}$, la cui media è $\mu_{\bar{X}} = \mu$ e il cui scarto quadratico medio è $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$. Il valore standardizzato di $\bar{X}$ è

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\bar{X} - \mu}{\sigma}$$

dunque, il livello di confidenza $(1 - \alpha)$ corrisponde a

$$P(\frac{\sqrt{n}}{\sigma} \mid \bar{X} - \mu \mid \leq Z_{\alpha/2}) = 1 - \alpha$$

equivalentemente

$$P(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Ex.1: I. C. sulla media

Svolgimento

- costruire l'intervallo di confidenza all 95%

il livello di confidenza $(1-\alpha) = 0.95$ corrisponde a

$$P(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 0.95$$

da cui, gli estremi dell'intervallo di confidenza sono $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$$\bar{X} = \frac{17+21+20+18+19+22+20+21+16+19}{10} = 19.3$$

Inoltre $\sigma = 3$, $n = 10$ e $Z_{\alpha/2} = Z_{0.025} = 1.96$. E' ora possibile calcolare gli estremi

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.3 \pm 1.96\frac{3}{\sqrt{10}} = 19.3 \pm 1.86$$

La stima intervallare è dunque $[17.44, 21.16]$

Ex.1: I. C. sulla media

Svolgimento

- costruire l'intervallo di confidenza all 90%

il livello di confidenza $(1-\alpha) = 0.90$ corrisponde a
Gli elementi per costruire l'intervallo sono: $\bar{X} = 19.3$, $\sigma = 3$, $n = 10$ e $Z_{\alpha/2} = Z_{0.05} = 1.645$. E' ora possibile calcolare gli estremi

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.3 \pm 1.645\frac{3}{\sqrt{10}} = 19.3 \pm 1.56$$

La stima intervallare è dunque $[17.74, 20.86]$

- costruire l'intervallo di confidenza all 99%

In questo caso $Z_{\alpha/2} = Z_{0.005} = 2.576$

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.3 \pm 2.576\frac{3}{\sqrt{10}} = 19.3 \pm 2.44$$

La stima intervallare è dunque $[16.86, 21.74]$

Ex.2: I. C. sulla media; determinazione numerosità campionaria

Dato un intervallo di confidenza, fissata l'ampiezza dell'intervallo a b, si determini la numerosità n del campione. Ricordando che gli estremi dell'intervallo sono dati da $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ pertanto l'ampiezza dell'intervallo è $2 \times Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$$2 \times Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = b$$

dunque la numerosità è data da

$$2 \times Z_{\alpha/2}\frac{\sigma}{b} = \sqrt{n} \Longrightarrow \left(2 \times Z_{\alpha/2}\frac{\sigma}{b}\right)^2 = n$$

Quale deve essere la numerosità per ottenere un intervallo di confidenza sulla media μ al 95% di ampiezza $b = 0.01$, tenuto conto che $\sigma = 2$?

Svolgimento

$$n = \left(2 \times 1.96\frac{2}{0.01}\right)^2 = (784)^2 = 614656$$

Ex.3: I. C. sulla media; varianza non nota

...prima di cominciare

Se lo scarto quadratico medio σ della popolazione non è noto occorre stimarlo attraverso lo scarto quadratico medio s della media campionaria $\bar{X}$. Dunque,

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

La quantità standardizzata è

$$T_{n-1} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

e si distribuisce secondo una distribuzione t di Student con $n-1$ gradi di libertà.

Ex.3: I. C. sulla media; varianza non nota

Un'agenzia deve valutare il grado di concentrazione di una sostanza tossica, il PCB, nel latte materno. Per fare questo, viene considerato un campione di 20 madri e studiato il grado di concentrazione della sostanza nel latte. I valori risultanti dalle analisi sono, in parti per milione,

$$\{16, 0, 0, 2, 3, 6, 8, 2, 5, 0, 12, 10, 5, 7, 2, 3, 8, 17, 9, 1\}$$

media e scarto quadratico medio campionari sono

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 5.8 \text{ e } s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} = 5.085$$

costruire intervalli di confidenza al 95% e al 99%.

Ex.3: IC sulla media; varianza non nota

Svolgimento

Essendo la varianza non nota

$$T_{n-1} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Dunque, valgono le seguenti relazioni

$$P(\frac{s}{\sqrt{n}} \mid \bar{X} - \mu \mid \leq t_{\alpha/2}) = 1 - \alpha$$

equivalentemente

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

da cui, gli estremi dell'intervallo di confidenza sono $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$

Ex.3: I. C. sulla media; varianza non nota

- gli estremi dell'intervallo di confidenza al 95%

Poichè risulta $\bar{X} = 5.8$,$s = 5.085$,$n = 20$ e $t_{\alpha/2} = 2.093$

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 5.8 \pm 2.093\frac{5.085}{\sqrt{20}} = 5.8 \pm 2.38$$

le stime ad intervallo sono dunque $[3.42, 8.18]$

- gli estremi dell'intervallo di confidenza al 99%

Poichè risulta $\bar{X} = 5.8$,$s = 5.085$,$n = 20$ e $t_{\alpha/2} = 2.861$

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 5.8 \pm 2.861\frac{5.085}{\sqrt{20}} = 5.8 \pm 3.25$$

le stime ad intervallo sono dunque $[2.55, 9.05]$

Ex.4: I. C. sulla proporzione

Sia p la proporzione di unità statistiche della popolazione presentano una certa caratteristica. La statistica campionaria corrispondente è la proporzione campionaria $\hat{p} = \frac{x}{n}$, dove x rappresenta il numero di unità nel campione che presentano una determinata caratteristica. $\hat{p}$ è tale che $\mu_{\hat{p}} = p$ e $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
Ricorrendo all'approssimazione della binomiale alla normale, gli estremi dell'intervallo sono dati da

$$\hat{p} \pm Z_{\alpha/2}\sigma_{\hat{p}}$$

essendo p incognito si stima lo scarto quadratico medio $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$; dunque

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Ex.4: I. C. sulla proporzione

Su un campione di 100 studenti universitari, 82 hanno terminato gli esami del primo anno negli appelli previsti.

- Si costruisca un intervallo di confidenza per p al 99%.

Svolgimento

In base ai dati del problema, $\hat{p} = \frac{82}{100}$

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.82 \pm 2.576\sqrt{\frac{0.82(1-0.82)}{100}} = 0.82 \pm 0.099$$

dunque le stime intervallari sono $[0.721, 0.919]$.

Ex.5: I. C. sulla proporzione; determinazione di n

Su un certo quotiniano viene riportato il risultato di un sondaggio secondo il quale il 46% della popolazione condivide le scelte di politica economica del governo. Sapendo che il margine di errore riportato è del 3%, e che il livello di confidenza utilizzato è $(1-\alpha) = 0.95$.

- Quante persone sono state intervistate?.

Svolgimento

In base ai dati del problema, $\hat{p} = 0.46$, $Z_{\alpha/2} = Z_{0.025} = 1.96$,poichè

$$Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.46 \times 0.54}{n}} = 0.03$$

esplicitando, otteniamo n, numero di intervistati

$$(1.96)^2\frac{0.46 \times 0.54}{n} = (0.03)^2 \longrightarrow n = (1.96)^2\frac{0.46 \times 0.54}{(0.03)^2}$$

quindi

$$n = (1.96)^2\frac{0.46 \times 0.54}{(0.03)^2} = 1060.3$$

dunque le persone intervistate sono state 1060

Ex.6: I. C. su somme e differenze tra medie

Si consideri di avere due popolazioni da cui si estraggono due campioni di numerosità rispettivamente n_1 e n_2. Siano S_1 e S_2 due generiche statistiche campionarie, la cui media e scarto quadratico medio sono date rispettivamente da μ_{S_1}, σ_{S_1},μ_{S_2},σ_{S_2}. Sulla base di tali informazioni si può costruire la distribuzione campionaria delle differenze tra le due statistiche $S_1 - S_2$. Media e scarto quadratico medio sono

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2}$$

$$\sigma_{S_1 - S_2} = \sqrt{\sigma^2_{S_1} + \sigma^2_{S_2}}$$

assumendo che i campioni siano indipendenti.

Se $S_1 = \bar{X}_1$ e $S_2 = \bar{X}_2$, allora risulta $\mu_{\bar{X}_1} = \mu_1$ e $\mu_{\bar{X}_2} = \mu_2$, poichè la media delle medie campionarie corrisponde alla media della popolazione. Inoltre $\sigma^2_{\bar{X}_1} = \frac{\sigma^2_1}{n_1}$ e $\sigma^2_{\bar{X}_2} = \frac{\sigma^2_2}{n_2}$; dunque

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}$$

Ex.5: I. C. su somme e differenze tra medie

Un campione di 150 lampadine di marca A ha un tempo di vita medio di 1400h, con uni scarto quadratico medio pari a 120h. Un campione di 100 lampadine di marca B ha un tempo di vita medio di 1200h, con uni scarto quadratico medio pari a 80h. Costruire un intervallo di confidenza al 95% e 99% sulla differenza media dei tempi di durata delle lampadine di marca A e B. Poichè

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

allora gli estremi dell'intervallo di confidenza sono

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Ex.5: I. C. su somme e differenze tra medie

- costruire l'intervallo di confidenza al 95%

In base ai dati del problema
$\bar{X}_A = 1400, \bar{X}_B = 1200, \sigma_A = 120, \sigma_B = 80, n_1 = 150, n_2 = 100, Z_{\alpha/2} = 1.96$ gli estremi dell'intervallo di confidenza sono

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1400 - 1200 \pm 1.96\sqrt{\frac{120^2}{150} + \frac{80}{100}} = 200 \pm 24.8$$

Gli estremi dell'intervallo sono $[175, 225]$.

- costruire l'intervallo di confidenza al 99%

In base ai dati del problema
$\bar{X}_A = 1400, \bar{X}_B = 1200, \sigma_A = 120, \sigma_B = 80, n_1 = 150, n_2 = 100, Z_{\alpha/2} = 2.58$ gli estremi dell'intervallo di confidenza sono

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1400 - 1200 \pm 2.58\sqrt{\frac{120^2}{150} + \frac{80}{100}} = 200 \pm 32.6$$

Gli estremi dell'intervallo sono $[167, 233]$.

Ex.6: I. C. su somme e differenze tra proporzioni

In un sondaggio sul gradimento di una certa trasmissione televisiva sono stati intervistati due campioni, uno di adulti (400) ed uno di adolescenti (600). Gli adolescenti che hanno espresso apprezzamento sono stati 300, gli adulti sono invece stati 100. Calcolare i limiti di confidenza al 95% e 99% sulla differenza tra la proporzione di adulti ed adolescenti favorevoli.

Svolgimento

Se $S_1 = \hat{p}_1$ e $S_2 = \hat{p}_2$, allora gli estremi dell'intervallo di confidenza sono

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Pertanto, in base ai dati del problema, $\hat{p}_1 = \frac{300}{600} = 0.5$ e $\hat{p}_1 = \frac{100}{400} = 0.25$

Ex.6: I. C. su somme e differenze tra proporzioni

Svolgimento

Essendo i dati del problema, $\hat{p}_1 = \frac{300}{600} = 0.5$, $\hat{p}_1 = \frac{100}{400} = 0.25$,$n_1 = 600$ e $n_2 = 400$.

- intervallo di confidenza al 95%

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$$

$$0.5 - 0.25 \pm 1.96\sqrt{\frac{0.5 \times 0.5}{600} + \frac{0.25 \times 0.75}{400}} = 0.25 \pm 0.06$$

gli estremi dell'intervallo sono $[0.19, 0.31]$

- intervallo di confidenza al 95%

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$$

$$0.5 - 0.25 \pm 2.58\sqrt{\frac{0.5 \times 0.5}{600} + \frac{0.25 \times 0.75}{400}} = 0.25 \pm 0.08$$

gli estremi dell'intervallo sono $[0.17, 0.33]$

Ex.7: I. C. su somme tra medie

La capacità media delle memorie RAM prodotte è di 995 megabyte. Lo scarto quadratico medio è invece 2 megabyte. Si supponga di aver montato quattro schede RAM su una scheda madre. Quali sono gli intervalli di confidenza al 95%, 99% e al 50% della capacità di memoria RAM installata in totale?

Svolgimento

Si consideri

$$\mu_{R1+R2+R3+R4} = \mu_{R1} + \mu_{R2} + \mu_{R3} + \mu_{R4} = 4 \times \mu_{Ri}$$

$$\sigma_{R1+R2+R3+R4} = \sqrt{\sigma^2_{R1} + \sigma^2_{R2} + \sigma^2_{R3} + \sigma^2_{R4}} = \sqrt{4 \times \sigma^2_{Ri}}$$

In base ai dati del problema, $\mu_{R1+R2+R3+R4} = 4 \times 995 = 3980$ e $\sigma_{R1+R2+R3+R4} = \sqrt{4 \times \sigma^2_{Ri}} = \sqrt{4 \times 2^2} = 4$.

Statistica
A. Iodice
I. C. sulla media
Esercizio 1
Esercizio 2
Esercizio 3
I. C. sulla proporzione
Esercizio 4
I. C. su somme e differenze tra medie
Esercizio 5
Esercizio 6
Esercizio 7

Svolgimento

- intervallo di confidenza al 95%

$$3980 \pm 1.96 \times 4 = 3980 \pm 7.84$$

gli estremi dell'intervallo sono $[3972.16, 3987.4]$

- intervallo di confidenza al 99%

$$3980 \pm 2.58 \times 4 = 3980 \pm 10.32$$

gli estremi dell'intervallo sono $[3969.68, 3990.32]$

- intervallo di confidenza al 50%

$$3980 \pm 0.6745 \times 4 = 3980 \pm 2.698$$

gli estremi dell'intervallo sono $[3977.3, 3982.7]$