Esercitazione 14

Statistica

Alfonso Iodice D'Enza
iodicede@unina.it

Università degli studi di Cassino

Outline

Outline

Outline

Outline

Ex.1: IC 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.1: IC 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.2: IC 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.2: IC 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.2: IC 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.3: IC 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.3: IC 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.4: IC 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 $\sigma_{R1+R2+R3+R4} = \sqrt{4 \times \sigma^2_{Ri}} = \sqrt{4 \times 2^2} = 4$.

Ex.4: IC su somme tra medie

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 $[3977.3, 3982.7]$

Ex.5: IC sullo scarto quadratico medio

La durata di vita di un campione di 200 motori di auto da corsa è stato stimato avere uno scarto quadratico medio di 100h a pieno regime.

- calcolare i limiti dell'intervallo di confidenza al 95% e al 99%.
- stabilire la numerosità campionaria necessaria a stabilire, con un livello di confidenza del 99.73%, che lo scarto quadratico medio della popolazione non differisca da quello stimato di più del 5% o del 10%.

Svolgimento

Sia s lo scarto quadratico medio campionario. I limiti dell'intervallo di confidenza su σ sono dati da

$$s \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}} = s \pm Z_{\alpha/2}\frac{s}{\sqrt{2n}}$$

σ viene stimato da s.

Ex.5: IC sullo scarto quadratico medio

Svolgimento

- intervallo di confidenza al 95%

$$s \pm Z_{\alpha/2}\frac{s}{\sqrt{2n}} = 100 \pm 1.96\frac{100}{\sqrt{200}} = 100 \pm 9.8$$

gli estremi dell'intervallo sono $[90.2, 109.8]$

- intervallo di confidenza al 99%

$$s \pm Z_{\alpha/2}\frac{s}{\sqrt{2n}} = 100 \pm 2.58\frac{100}{\sqrt{200}} = 100 \pm 12.9$$

gli estremi dell'intervallo sono $[87.1, 112.9]$

Ex.5: IC sullo scarto quadratico medio

Svolgimento

- stabilire la numerosità campionaria necessaria a stabilire, con un livello di confidenza del 99.73%, che lo scarto quadratico medio della popolazione non differisca da quello stimato di più del 5% o del 10%.

Per $(1-\alpha)=0.9973$ il corrispondente valore $Z_{\alpha/2}=3$. Inoltre si vuole che la quanità $Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}$ non si discosti dal valore dello scarto quadratico di più del 5%. Risulta pertanto

$$Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}:s=5:100$$

da cui, poichè $s=100$ e σ è stimato tramite s, la precedente si può riscrivere

$$Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}=5\longrightarrow 3\frac{100}{\sqrt{2n}}=5$$

$$3\frac{100}{5}=\sqrt{2n}\longrightarrow\left(\frac{300}{5}\right)^2=2n\longrightarrow n=\frac{(60)^2}{2}=1800$$

il calcolo rispetto al 10% è poco dissimile. $Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}:s=10:100$; da cui

$$3\frac{300}{10}=\sqrt{2n}\longrightarrow\left(\frac{300}{10}\right)^2=2n\longrightarrow n=\frac{(30)^2}{2}=450$$