



STATISTICA · UNIVERSITÀ

Intervalli di confidenza

Raccolta di 12 esercizi svolti: stima di media e proporzione, con σ nota e ignota, e dimensionamento del campione. Ogni risultato è verificato numericamente.

Andrea Paone

INFOMAAT · Centro di ripetizioni · matematica e fisica dal 2011 12 esercizi · media & proporzione

FORMULE DI RIFERIMENTO

IC media · σ noto

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

IC media · σ ignoto

$$\bar{x} \pm t_{\alpha/2; n-1} \sqrt{\frac{S^2}{n}}$$

IC proporzione

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

Varianza corretta

$$S^2 = \frac{n}{n-1} S^{*2}$$

Numerosità · media

$$n \geq \left(\frac{z_{\alpha/2} \sigma}{E} \right)^2$$

Numerosità · proporzione

$$n \geq \frac{z_{\alpha/2}^2 p(1-p)}{E^2}$$

QUANTILI USATI

$$z_{0.025} = 1.96$$

$$z_{0.05} = 1.645$$

$$z_{0.04} = 1.75$$

$$z_{0.075} = 1.44$$

$$t_{19; 0.025} = 2.093$$

$$t_{24; 0.05} = 1.711$$

Convenzione: la varianza campionaria "corretta" S^2 usa il divisore $n - 1$; con σ ignoto e campione piccolo si usa la t di Student, con campione grande $t \approx z$.

ESERCIZIO 1 ★★

Media e varianza incognite · effetto della numerosità

Considera i seguenti 20 dati come osservazioni indipendenti da una variabile casuale Normale con media e varianza incognite:

18, 25, 30, 33, 15, 9, 14, 19, 21, 20, 22, 18, 16, 25, 28, 28, 15, 20, 30, 25

- Determina stime puntuali per media e varianza della popolazione.
- Calcola un intervallo di confidenza al 95% per la media incognita.
- Supponendo ora di aver rilevato un campione di ampiezza 100 con la stessa media campionaria del punto a), ricalcola l'intervallo al 95% per la media e confrontalo con quello del punto b).

SVOLGIMENTO

a) Stime puntuali. Per la media usiamo la media campionaria; per la varianza lo stimatore corretto $S^2 = \frac{n}{n-1} S^{*2}$.

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{431}{20} = 21.55$$

$$S^{*2} = \frac{1}{20} \sum x_i^2 - \bar{x}^2 = 38.0475 \implies S^2 = \frac{20}{19} \cdot 38.0475 = 40.05$$

b) IC 95% per la media, σ^2 ignoto. Con $n = 20$ (campione piccolo) usiamo la distribuzione t di Student con $n - 1 = 19$ gradi di libertà.

$$1 - \alpha = 0.95 \implies \frac{\alpha}{2} = 0.025, \quad t_{19; 0.025} = 2.093$$

$$\bar{x} \pm t_{19; 0.025} \sqrt{\frac{S^2}{n}} = 21.55 \pm 2.093 \sqrt{\frac{40.05}{20}} = 21.55 \pm 2.093 \cdot 1.4151 = 21.55 \pm 2.962$$

RISULTATO

$$18.588 \leq \mu \leq 24.512$$

ESERCIZIO 1 · SEGUITO

Effetto della numerosità campionaria (punto c)

SVOLGIMENTO

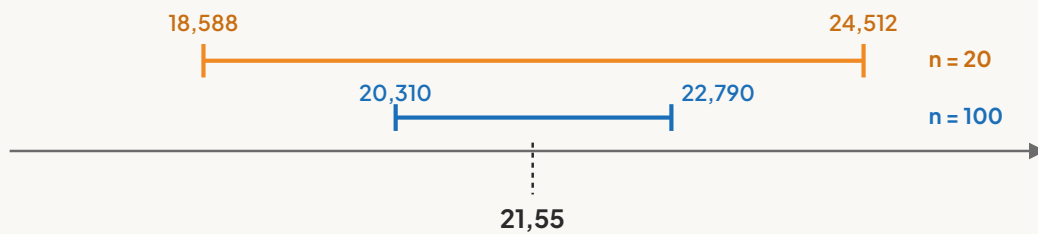
c) **Campione di ampiezza 100.** Con $n = 100$ (campione grande) il quantile $t_{99} \approx z$, quindi usiamo $z_{0,025} = 1.96$; media e varianza stimata restano le stesse.

$$21.55 \pm 1.96 \sqrt{\frac{40.05}{100}} = 21.55 \pm 1.96 \cdot 0.6329 = 21.55 \pm 1.240$$

RISULTATO

$$20.310 \leq \mu \leq 22.790$$

Confronto. Il centro non cambia (21,55): cambia l'ampiezza, che passa da circa 5,92 a circa 2,48. Moltiplicando la numerosità per 5, il semi-errore si riduce di un fattore $\sqrt{5} \approx 2.24$.



Stesso centro, ampiezza dimezzata: passando da $n = 20$ a $n = 100$ l'intervallo si stringe.

ESERCIZIO 2 ★☆☆

IC per una proporzione al 92%

In un campione casuale di 200 persone il 22% risulta consumatore abituale di un certo prodotto. Individua l'intervallo di confidenza al livello $1-\alpha = 0,92$ per l'ignota proporzione π di consumatori abituali nella popolazione.

SVOLGIMENTO

Stima puntuale della proporzione: $p = 0.22$, con $n = 200$. Per una proporzione l'intervallo è

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

$$1 - \alpha = 0.92 \Rightarrow \frac{\alpha}{2} = 0.04, \quad z_{0.04} = 1.75$$

$$0.22 \pm 1.75 \sqrt{\frac{0.22 \cdot 0.78}{200}} = 0.22 \pm 1.75 \cdot 0.02929 = 0.22 \pm 0.0513$$

RISULTATO

$$0.1687 \leq \pi \leq 0.2713 \quad (\text{circa } 16.9\% - 27.1\%)$$

ESERCIZIO 3 ★★☆☆

IC per la media con varianza corretta

Il tempo di reazione a uno stimolo sensoriale segue una Normale con media e varianza incognite. Su un campione di 25 soggetti si osserva media pari a 8,5 e varianza campionaria pari a 3,8. Usando lo stimatore corretto, calcola un intervallo di confidenza al 90% per la media incognita.

SVOLGIMENTO

La varianza fornita (3,8) è quella distorta S^{*2} ; la correggiamo:

$$S^2 = \frac{n}{n-1} S^{*2} = \frac{25}{24} \cdot 3.8 = 3.958$$

Con σ^2 ignoto e $n = 25$ usiamo la t di Student con $n - 1 = 24$ g.d.l.

$$1 - \alpha = 0.90 \Rightarrow \frac{\alpha}{2} = 0.05, \quad t_{24; 0.05} = 1.711$$

$$8.5 \pm 1.711 \sqrt{\frac{3.958}{25}} = 8.5 \pm 1.711 \cdot 0.3979 = 8.5 \pm 0.681$$

RISULTATO

$$7.819 \leq \mu \leq 9.181$$

ESERCIZIO 4 ★★☆☆

IC per una proporzione al 92%

In una città il 65% di un campione di 200 individui utilizza i mezzi di trasporto pubblici. Individua l'intervallo di confidenza al livello $1 - \alpha = 0,92$ per l'ignota frequenza relativa π di utilizzatori nella popolazione.

SVOLGIMENTO

Stessa struttura dell'Esercizio 2 (IC per una proporzione), con $p = 0.65$, $n = 200$, $z_{0.04} = 1.75$.

$$0.65 \pm 1.75 \sqrt{\frac{0.65 \cdot 0.35}{200}} = 0.65 \pm 1.75 \cdot 0.03373 = 0.65 \pm 0.0590$$

RISULTATO

$$0.591 \leq \pi \leq 0.709 \quad (\text{circa } 59.1\% - 70.9\%)$$

ESERCIZIO 5 ★☆☆

IC per la media con σ noto

Da una popolazione con media incognita e varianza pari a 25 estrai un campione di numerosità 400, ottenendo media campionaria 9,7. Calcola l'intervallo di confidenza al livello $1-\alpha = 0,92$ per la media.

SVOLGIMENTO

Qui la varianza (e quindi $\sigma = 5$) è **nota**: usiamo il quantile Normale z .

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad z_{0.04} = 1.75$$

$$9.7 \pm 1.75 \cdot \frac{5}{\sqrt{400}} = 9.7 \pm 1.75 \cdot 0.25 = 9.7 \pm 0.4375$$

RISULTATO

$$9.263 \leq \mu \leq 10.138$$

ESERCIZIO 6 ★★☆☆

Dimensionamento: ampiezza dell'intervallo

Da una popolazione Normale $N(\mu, 1)$ estrai un campione di ampiezza n . Determina la numerosità campionaria n affinché l'ampiezza dell'intervallo di confidenza al 95% per la media non superi 0,5.

SVOLGIMENTO

La varianza è nota ($\sigma = 1$). L'**ampiezza** dell'intervallo (lunghezza totale) è il doppio del semi-errore:

$$L = 2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 0.5, \quad z_{0.025} = 1.96$$

$$2 \cdot 1.96 \cdot \frac{1}{\sqrt{n}} \leq 0.5 \implies \sqrt{n} \geq \frac{2 \cdot 1.96}{0.5} = 7.84 \implies n \geq 61.47$$

Poiché n è intero, arrotondiamo per eccesso.

RISULTATO

$$n \geq 62$$

ESERCIZIO 7 ★★ ★

Copertura dell'intervallo di confidenza

Da una popolazione di maschi adulti con statura Normale di media $\mu = 170$ e varianza 16 estrai un campione di 25 unità, in cui la statura media risulta 173.

- a) Mostra che l'intervallo di confidenza al 95% costruito con questi valori non contiene il vero valore $\mu = 170$.
- b) Trova l'insieme dei valori della media campionaria per cui $\mu = 170$ risulta effettivamente contenuto nell'intervallo al 95% costruito sui dati.

SVOLGIMENTO

- a) Varianza nota ($\sigma = 4$), quindi quantile Normale $z_{0,025} = 1.96$.

$$173 \pm 1.96 \cdot \frac{4}{\sqrt{25}} = 173 \pm 1.568 \Rightarrow [171.432, 174.568]$$

Il valore vero $\mu = 170$ è **minore** dell'estremo inferiore 171,432: l'intervallo **non** contiene 170. Questo campione rientra tra il circa 5% di campioni per cui l'intervallo al 95% "manca" il parametro. b) Il parametro $\mu = 170$ cade nell'intervallo $\bar{x} \pm 1.568$ se e solo se

$$170 - 1.568 \leq \bar{x} \leq 170 + 1.568$$

RISULTATO

$$168.432 \leq \bar{x} \leq 171.568$$

ESERCIZIO 8 ★☆☆**IC per una proporzione all'85%**

Per stimare la frequenza di insuccessi di una terapia vengono esaminati 120 casi, di cui 25 non hanno tratto giovamento. Calcola l'intervallo di confidenza al livello $1-\alpha = 0,85$ per l'ignota frequenza relativa di insuccessi.

SVOLGIMENTO

Proporzione osservata: $p = \frac{25}{120} = 0.2083$, con $n = 120$.

$$1 - \alpha = 0.85 \Rightarrow \frac{\alpha}{2} = 0.075, \quad z_{0.075} = 1.44$$

$$0.2083 \pm 1.44 \sqrt{\frac{0.2083 \cdot 0.7917}{120}} = 0.2083 \pm 1.44 \cdot 0.03707 = 0.2083 \pm 0.0534$$

RISULTATO

$$0.155 \leq \pi \leq 0.262$$

ESERCIZIO 9 ★☆☆**IC per una proporzione al 95%**

Per conoscere la percentuale di possessori di telefono cellulare in un comune vengono intervistate 200 persone, di cui 125 dichiarano di possederne uno. Individua l'intervallo di confidenza al 95% per l'ignota percentuale di possessori.

SVOLGIMENTO

Proporzione osservata: $p = \frac{125}{200} = 0.625$, con $n = 200$, $z_{0.025} = 1.96$.

$$0.625 \pm 1.96 \sqrt{\frac{0.625 \cdot 0.375}{200}} = 0.625 \pm 1.96 \cdot 0.03423 = 0.625 \pm 0.0671$$

RISULTATO

$$0.558 \leq \pi \leq 0.692$$

Dati raggruppati: stime e IC per una proporzione

Un campione di 150 studenti fornisce le ore settimanali dedicate allo sport, riassunte nella distribuzione di frequenze:

ore	0	2	4	6	8
frequenze	10	20	60	30	30

- a) Determina le stime puntuali di media e varianza della popolazione.
 b) Considerando la percentuale di studenti che dedicano *meno* di 4 ore allo sport, calcola l'intervallo di confidenza al 90% per tale proporzione ignota.

SVOLGIMENTO

- a) Media e varianza da dati raggruppati ($n = 150$):

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{0 \cdot 10 + 2 \cdot 20 + 4 \cdot 60 + 6 \cdot 30 + 8 \cdot 30}{150} = \frac{700}{150} = 4.667$$

$$S^{*2} = \frac{\sum x_i^2 f_i}{n} - \bar{x}^2 = \frac{4040}{150} - 4.667^2 = 26.933 - 21.778 = 5.156$$

$$S^2 = \frac{n}{n-1} S^{*2} = \frac{150}{149} \cdot 5.156 = 5.191$$

- b) "Meno di 4 ore" significa 0 oppure 2 ore: $p = \frac{10+20}{150} = 0.2$.

$$1 - \alpha = 0.90 \Rightarrow z_{0.05} = 1.645$$

$$0.2 \pm 1.645 \sqrt{\frac{0.2 \cdot 0.8}{150}} = 0.2 \pm 1.645 \cdot 0.03266 = 0.2 \pm 0.0537$$

RISULTATO

$$0.146 \leq \pi \leq 0.254$$

ESERCIZIO 11 ★★☆☆

Dimensionamento: errore sulla media

Le uova di un'azienda avicola hanno peso Normale con media incognita e varianza 49. Determina la dimensione del campione che consente di stimare la media con un errore non superiore, in valore assoluto, a 4 g, con probabilità del 95%.

SVOLGIMENTO

Varianza nota ($\sigma = 7$). L'errore di stima (semi-ampiezza) deve restare entro 4 g:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 4, \quad z_{0.025} = 1.96$$

$$1.96 \cdot \frac{7}{\sqrt{n}} \leq 4 \implies \sqrt{n} \geq \frac{1.96 \cdot 7}{4} = 3.43 \implies n \geq 11.76$$

RISULTATO

$$n \geq 12$$

ESERCIZIO 12 ★★☆☆

Dimensionamento: proporzione nel caso peggiore

In una sperimentazione clinica si vuole stimare il tasso di guarigione p di un nuovo farmaco. Calcola il numero di unità da includere affinché, con probabilità del 95%, l'errore di stima di p sia inferiore a 0,05.

SVOLGIMENTO

Errore per una proporzione: $E = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq 0.05$, con $z_{0.025} = 1.96$. Poiché p è ignoto, usiamo il **caso peggiore** $p = 0.5$, che massimizza $p(1-p) = 0.25$.

$$n \geq \frac{z_{\alpha/2}^2 p(1-p)}{E^2} = \frac{1.96^2 \cdot 0.25}{0.05^2} = \frac{0.9604}{0.0025} = 384.16$$

Arrotondando per eccesso all'intero successivo.

RISULTATO

$$n \geq 385$$