Esercitazione 6 del corso di Statistica 2
Prof. Domenico Vistocco

Alfonso Iodice D'Enza

June 9, 2007

1 Esercizio

Si consideri di estrarre un campione di 100 elementi ($N = 100$) tra gli studenti iscritti all'università Alpha, che in totale sono 1546 ($N_p = 100$). Si supponga, avendo osservato il peso degli studenti, che i parametri media e scarto quadratico medio calcolati sul campione siano rispettivamente $\bar{X} = 67.45kg$ e $S = \sqrt{8.5275} = 2.92kg$.

- Determinare la stima corretta ed efficiente della media della popolazione;
- determinare la stima corretta ed efficiente della varianza della popolazione.

1.1 Svolgimento

- Determinare la stima corretta ed efficiente della media della popolazione;

Si consideri la distribuzione della media campionaria $\bar{X}$: poichè la media della distribuzione media campionaria coincide con la media della popolazione, ovvero $\mu_{\bar{X}} = \mu$, allora lo stimatore media campionaria $barX$ è corretto, dal momento che il suo valore atteso coincide con il parametro della popolazione, infatti $E(\bar{X}) = \mu_{\bar{X}} = \mu$. Inoltre lo stimatore è efficiente perchè, tra gli stimatori corretti esso è caratterizzato da varianza minima.

- Determinare la stima corretta ed efficiente della varianza della popolazione;

La media della distribuzione della varianza campionaria è

$$\mu_{s^2} = \frac{N-1}{N}\sigma^2$$

con σ^2 varianza della popolazione. Si può pertanto concludere che S^2 è uno stimatore distorto della varianza. Il fattore di correzione da applicare è dato da $\frac{N}{N-1}$, pertanto lo stimatore della varianza corretto è

$$\hat{S}^2 = \frac{N}{N-1} S^2$$

Con $\hat{S}$ corretto poichè risulta $E(\hat{S}^2) = \mu_{\hat{S}^2} = \sigma^2$.
Sulla base di quanto detto, sapendo che $N = 100$ e che $S^2 = 8,5275$. Pertanto la stima corretta ed efficiente è data da

$$\hat{S}^2 = \frac{N}{N-1} S^2 = \frac{100}{99} \times 8.5275 = 8.6136$$

Con $\hat{s}^2 = \sqrt{8.6136} = 2.93$, laddove $s^2 = \sqrt{8.5275} = 2.92$: il valore quasi coincidente della stima corretta e di quella distorta è da ricondurre al fatto che il fattore di correzione, quando la taglia del campione è elevata, tende ad 1.

2 Esercizio

Con riferimento al campione di 100 studenti universitari sui quali la distribuzione del peso è caratterizzata dai parametri $\bar{X} = 67.45kg$ e $\hat{S}^2 = 2.93$, che abbiamo trovato essere stime corrette ed efficienti della media e dello scarto quadratico medio della popolazione.

- Calcolare i limiti degli intervalli di confidenza al 90%
- Calcolare i limiti degli intervalli di confidenza al 95%
- Calcolare i limiti degli intervalli di confidenza al 99%

2.1 Svolgimento

La forma generale per il calcolo dei limiti di un intervallo di confidenza su una statistica S distribuita normalmente con media μ_S e scarto quadratico medio σ_S, ad un dato livello di confidenza, è dato da

$$S \pm z_c \sigma_S.$$

A seconda della statistica utilizzata e del livello di confidenza voluto, si calcoleranno i limiti dell'intervallo di confidenza.

- Calcolare i limiti degli intervalli di confidenza al 90%

In questo caso la statistica considerata è la media campionaria $\hat{X}$; lo scarto quadratico medio della media campionaria, in caso di popolazioni finite $\sigma_{\hat{X}} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$: tuttavia, il fattore di correzione $\sqrt{\frac{N_p - N}{N_p - 1}}$ è circa 1, e si può omettere: questo perchè la taglia del campione $(N = 100) \ll (N_p = 1546)$. Alla luce di quanto trovato, i limiti dell'intervallo sulla media sono dati da

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

Sapendo che $\bar{X} = 67.45$, che la stima corretta di σ è $\hat{S}^2$ e che ad un livello di confidenza del 90% il valore critico $z_c = 1.645$, possiamo costruire i limiti dell'intervallo

$$67.45 - 1.645\frac{2.93}{\sqrt{100}} = 66.968 \text{ e } 67.45 + 1.645\frac{2.93}{\sqrt{100}} = 67.932$$

Ovvero la media del peso della popolazione degli studenti iscritti all'università Alpha cadrà, con una confidenza del 99%, nell'intervallo di valori $[66.968, 67.932]$. Formalmente si può scrivere $P(66.968 < \mu < 67.932)$. Per i diversi livelli di confidenza, si procede analogamente, trovando i diversi valori critici z_c.

- Calcolare i limiti degli intervalli di confidenza al 95%

In questo caso il solo valore a cambiare è z_c, in particolare ad un livello di confidenza del 95% risulta essere $z_c = 1.96$. Quindi i valori dell'intervallo vanno aggiornati nel seguente modo:

$$67.45 - 1.96\frac{2.93}{\sqrt{100}} = 66.876 \text{ e } 67.45 + 1.96\frac{2.93}{\sqrt{100}} = 68.024$$

- Calcolare i limiti degli intervalli di confidenza al 99%

Anche in questo caso il solo valore a cambiare è z_c, in particolare ad un livello di confidenza del 99% risulta essere $z_c = 2.58$. Quindi i valori dell'intervallo vanno aggiornati nel seguente modo:

$$67.45 - 2.58\frac{2.93}{\sqrt{100}} = 66.69 \text{ e } 67.45 + 2.58\frac{2.93}{\sqrt{100}} = 68.206$$

3 Esercizio

Si consideri la popolazione dei voti riportati da 200 studenti ad una prova di matematica. Si consideri inoltre un campione casuale di 50 studenti, caratterizzato da una media-voto $\bar{X} = 75$ e scarto quadratico medio pari a 10.

- Si calcolino i limiti dell'intervallo di confidenza al 95%
- Quale grado di confidenza è attribuibile all'affermazione che la media della popolazione cada nell'intervallo 75 ± 1?

3.1 Svolgimento

In questo caso bisogna tenere conto del fatto che il campionamento è senza ripetizione e applicare il fattore di correzione $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}}$ allo scarto quadratico medio della distribuzione della media campionaria: questo si rende necessario considerato che $N_p = 200$ ed $N = 50$.

- Si calcolino i limiti dell'intervallo di confidenza al 95%

Pertanto i limiti dell'intervallo di confidenza sulla media saranno dati da

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

Ad un livello di confidenza del 99%, i limiti dell'intervallo saranno dati da

$$75 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{150}{199}} = 75 \pm 2.4$$

Pertanto i limiti sono 72.6 e 77.4.

- Quale grado di confidenza è attribuibile all'affermazione che la media della popolazione cada nell'intervallo 75 ± 1?

Per trovare il grado di confidenza si considerino i limiti dell'intervallo sulla media

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}.$$

In particolare, essendo interessati all'intervallo 75 ± 1 deve risultare che

$$z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} = 1 \text{ da cui } z_c = \frac{1}{1.23} = 0.81$$

Il livello di confidenza cercato è dunque dato da $\alpha = 2 \times P(0 < Z < 0.81) = 2 \times 0.291 = 0.582$

4 Esercizio

Dai risultati di un sondaggio effettuato su 100 votanti è emerso che il 55% di essi ha scelto il candidato A.

- Si calcolino i limiti di confidenza al 95%.
- Qual'è l'ampiezza del campione da scegliere per essere confidenti al 99% che il candidato A vinca le elezioni?

4.1 Svolgimento

- Si calcolino i limiti di confidenza al 95%.

Per stimare la proporzione p di elettori di A relativa alla popolazione si utilizza la proporzione campionaria $P = 0.55$. I limiti dell'intervallo di confidenza sono

$$P \pm z_c \sigma_p = P \pm z_c \sqrt{\frac{p(1-p)}{N}} = 0.55 \pm 1.96 \sqrt{\frac{0.55 \times 0.45}{100}} = 0.55 \pm 0.1$$

Pertanto si può essere confidenti al 95% che la proporzione di votanti il candidato A nella popolazione sia compresa tra 0.45 e 0.65.

- Qual'è l'ampiezza del campione da scegliere per essere confidenti al 99% che il candidato A vinca le elezioni?

Si assuma che, per essere eletto, il candidato A debba avere più del 50% dei voti; questo vuol dire che bisogna cercare la taglia del campione N a cui corrisponde, con un livello di confidenza del 99%, un intervallo di confidenza $P \pm 0.05 = 0.55 \pm 0.05$. Un raggio maggiore di 0.05 includerebbe valori della proporzione inferiori al 50%, che non consentirebbero al candidato A di vincere le elezioni. Formalmente deve risultare che

$$z_c \sqrt{\frac{p(1-p)}{N}} < 0.05 \rightarrow 2.575 \frac{0.5}{\sqrt{N}} < 0.05 \rightarrow N > 663.1$$

Pertanto la taglia del campione deve essere di almeno 664 elementi.