STATISTICA: esercizi svolti sulla STIMA INTERVALLARE

1 STIMA INTERVALLARE

1.1 Esercizi

1. Una partita di bulloni presenta un diametro medio μ incognito; la varianza del diametro dei bulloni è invece nota e pari a $0,01$ cm. Si estrae un campione di $n = 1000$ bulloni, sui quali si osserva un diametro medio pari a $1,2$ cm.

   a) Si determini l'intervallo di confidenza per μ avendo fissato un livello di confidenza del 99%.

   b) Si determini l'ampiezza di tale intervallo.

Svolgimento

a) Per determinare l'intervallo di confidenza (I.C.) per μ a livello di confidenza pari al 99%, bisogna innanzitutto ricavare il quantile di ordine $1 - \dfrac{\alpha}{2}$ della distribuzione normale standard.
Quindi, poichè
$$1 - \alpha = 0.99$$
si ha che
$$\alpha = 1 - 0.99 = 0.01.$$
Perciò
$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$
e di conseguenza
$$1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995.$$
Consultando le tavole della distribuzione normale standard e interpolando tra i punti di coordinate (2.57 ; 0.99492) e (2.58 ; 0.99506), si ricava che
$$\Phi(2.576) = 0.995002 \cong 0.995$$
pertanto
$$z_{1-\frac{\alpha}{2}} = z_{0.995} = z_{0.995} = 2.576.$$
Si ricorda che l'intervallo di confidenza per μ a livello $1 - \alpha$ è dato da
$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}} \; ; \; \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\right]$$
dove

- $\bar{X}$ è lo stimatore media campionaria;
- $z_{1-\frac{\alpha}{2}}$ è il quantile di ordine $1 - \dfrac{\alpha}{2}$ della distribuzione normale standard;
- σ^2 è la varianza della popolazione di riferimento;

– n è l'ampiezza campionaria.

Sostituendo quindi i valori forniti dal testo e il quantile calcolato precedentemente, si ricava che:

$$\left[1.2 - 2.576 \cdot \sqrt{\frac{0.01}{1000}}\ ;\ 1.2 + 2.576 \cdot \sqrt{\frac{0.01}{1000}}\right]$$

è l'I.C. per μ a livello 0.99.
Si ottiene pertanto:

$$[1.1918\ ;\ 1.2081] \qquad \text{I.C. per } \mu \text{ a livello } 0.99.$$

b) L'ampiezza di tale intervallo è data da

$$\begin{aligned} Amp &= (estremo\ sup) - (estremo\ inf) \\ &= 1.2081 - 1.1918 = 0.01629 \qquad \text{(cm)}. \end{aligned}$$

Si ricorda che in generale l'ampiezza dell'intervallo di confidenza a livello $1-\alpha$ per la media μ nel caso di varianza σ^2 nota è pari a:

$$Amp = 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}.$$

Nel caso in esame:

$$\begin{aligned} Amp &= 2 \cdot 2.576 \cdot \sqrt{\frac{0.01}{1000}} \\ &= 0.01629. \end{aligned}$$

2. Tra i pasticcini prodotti artigianalmente in una pasticceria se ne prelevano $n = 100$; risulta che il loro peso medio è pari a 35 g. Si sa che lo scarto quadratico medio del peso di tutti i pasticcini prodotti dalla pasticceria è pari a 4 g.

   a) Si trovi l'intervallo di confidenza per il peso medio di tutti i pasticcini prodotti a livello di confidenza del 98%.

   b) Di quanto deve aumentare la numerosità campionaria se si vuole che l'ampiezza dell'intervallo si dimezzi?

   c) Si determini quanti pasticcini occorre ancora estrarre se si vuole che lo stimatore del peso medio si discosti dal vero peso medio per meno di un grammo con probabilità del 96%.

Svolgimento

a) Per determinare l'intervallo di confidenza (I.C.) per μ a livello di confidenza pari al 98%, bisogna innanzitutto ricavare il quantile di ordine $1 - \frac{\alpha}{2}$ della distribuzione normale standard.
Quindi, poichè

$$1 - \alpha = 0.98$$

si ha che

$$\alpha = 1 - 0.98 = 0.02.$$

Perciò

$$\frac{\alpha}{2} = \frac{0.02}{2} = 0.01$$

e di conseguenza

$$1 - \frac{\alpha}{2} = 1 - 0.01 = 0.99.$$

Consultando le tavole della distribuzione normale standard e interpolando tra i punti di coordinate $(2.32\ ;\ 0.98983)$ e $(2.33\ ;\ 0.99010)$, si ricava che

$$\Phi(2.326) = 0.989991 \cong 0.99,$$

pertanto

$$z_{1-\frac{\alpha}{2}} = z_{0.99} = 2.326.$$

Ricordando che l'intervallo di confidenza per μ a livello $1 - \alpha$ è dato da

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\ ;\ \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\right]$$

sostituendo, otteniamo che

$$\left[35 - 2.326 \cdot \sqrt{\frac{4^2}{100}}\ ;\ 35 + 2.326 \cdot \sqrt{\frac{4^2}{100}}\right]$$

è l'I.C. per μ a livello 0.98.

Si ottiene pertanto:

$$[34.0696\ ;\ 35.9304] \qquad \text{I.C. per } \mu \text{ a livello } 0.98.$$

b) Si ricorda che l'ampiezza dell'intervallo di confidenza determinato nel punto precedente è pari a:

$$\begin{aligned} Amp &= 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{100}} \\ &= 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{10}. \end{aligned}$$

A questo punto si vuole determinare un'ampiezza campionaria $\tilde{n}$ che dimezzi l'ampiezza dell'I.C. a livello di confidenza 0.98. In altre parole, si vuole determinare $\tilde{n}$ tale che:

$$2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{\tilde{n}}} = \frac{Amp}{2} = \frac{1}{2}\left[2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{10}\right].$$

Quindi, dalla relazione

$$2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{\tilde{n}}} = \frac{1}{2}\left[2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{10}\right]$$

si ottiene che

$$2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{\tilde{n}}} = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{10}$$

e pertanto:

$$2 \cdot \frac{1}{\sqrt{\tilde{n}}} = \frac{1}{10}.$$

Risolvendo l'equazione rispetto a $\tilde{n}$, si ricava:

$$\tilde{n} = (2 \cdot 10)^2 = 400.$$

É possibile quindi affermare che l'ampiezza dell'intervallo di confidenza si dimezza se si aumenta l'ampiezza campionaria di 400-100=300 unità (cioè l'ampiezza campionaria deve quadruplicare).

c) Bisogna determinare n tale che:

$$P\{|\bar{X} - \mu| < 1\} = 0.96.$$

Riscriviamo la relazione precedente come segue:

$$P\{-1 < \bar{X} - \mu < 1\} = 0.96$$

e riconosciamo che tale probabilità è uguale a

$$P\left\{-\frac{1}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{1}{\sigma/\sqrt{n}}\right\} = 0.96. \tag{1}$$

A questo punto, riconoscendo che

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \sim N(0,1)$$

la relazione (1) diventa

$$P\left\{-\frac{\sqrt{n}}{\sigma} < Z < \frac{\sqrt{n}}{\sigma}\right\} = 0.96$$

e, ricordando che $\sigma = 4$:

$$\begin{aligned}
&P\left\{-\frac{\sqrt{n}}{4} < Z < \frac{\sqrt{n}}{4}\right\} = 0.96\\
&\Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 0.96\\
&\Phi\left(\frac{\sqrt{n}}{4}\right) - \left[1 - \Phi\left(\frac{\sqrt{n}}{4}\right)\right] = 0.96\\
&2 \cdot \Phi\left(\frac{\sqrt{n}}{4}\right) = 1.96\\
&\Phi\left(\frac{\sqrt{n}}{4}\right) = 0.98.
\end{aligned}$$

Consultando le tavole della distribuzione normale standard, si ricava che

$$\Phi(2.05) = 0.97982 \cong 0.98.$$

Quindi

$$\frac{\sqrt{n}}{4} = 2.05$$

e pertanto:

$$n = (2.05 \cdot 4)^2 = 67.24 \qquad \text{(arrotondando } n = 68\text{)}.$$

Tale relazione ci informa che affinchè lo stimatore del peso medio dei pasticcini si discosti dal vero peso medio per meno di un grammo con probabilità pari a 0.96 è necessaria un'ampiezza campionaria pari a 68.

É possibile quindi concludere che l'ampiezza campionaria è già sufficientemente grande: non è necessario estrarre alcun pasticcino in più affichè lo stimatore del peso medio si discosti dal vero peso medio per meno di un grammo con probabilità pari a 0.96.

3. In un vivaio ci sono 1000 alberi tra i quali una proporzione incognita p ha contratto una malattia.

   a) Stabilire quanti alberi occorre controllare affinché l'intervallo di confidenza a livello $1 - \alpha = 0.99$ per la proporzione incognita di alberi malati risulti ampio meno di 0.1.

   b) Si decide di selezionare 100 alberi con riposizione al fine di stimare la proporzione p di alberi che hanno contratto la malattia. Di questi 100 risulta che 40 hanno contratto la malattia. Si calcoli l'intervallo di confidenza per la proporzione p al livello di confidenza del 98%.

c) Tenendo conto del risultato campionario di cui al punto b), si determini la numerosità campionaria che assicura che la varianza dello stimatore della proporzione di alberi ammalati sia pari a 0.001.

Svolgimento

a) Come è noto, l'ampiezza dell'I.C. (a livello $1 - \alpha = 0.99$) per la proporzione incognita di alberi malati è data da:

$$Amp = 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}.$$

Calcoliamo quindi innanzitutto $z_{1-\frac{\alpha}{2}}$. Dalla relazione

$$1 - \alpha = 0.99$$

ricaviamo che

$$\alpha = 0.01$$

cioè

$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$

e pertanto

$$1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995.$$

Dalle tavole della distribuzione normale standard, interpolando tra i punti di coordinate (2.57 : 0.99492) e (2.58 ; 0.99506), si ricava che:

$$\Phi(2.576) = 0.995002 \cong 0.995,$$

quindi si pone:

$$z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.576.$$

Non avendo informazioni su p e q, consideriamo il caso più sfavorevole, cioè quello in cui

$$p = q = 0.5.$$

É questo il caso di massima incertezza per il quale il prodotto pq è massimo.

L'ampiezza dell'I.C. è di conseguenza uguale a:

$$Amp = 2 \cdot 2.576 \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}}.$$

Imponendo che

$$Amp < 0.1$$

otteniamo che

$$2 \cdot 2.576 \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} < 0.1$$

$$\frac{2 \cdot 2.576 \cdot 0.5}{\sqrt{n}} < 0.1$$
$$\sqrt{n} > \frac{2.576}{0.1}$$

da cui

$$n > 663.5776 \qquad \text{(arrotondando } n \geq 664\text{)}.$$

Controllando 664 alberi si ha che l'ampiezza dell'I.C. a livello 0.99 per la proporzione ignota di alberi malati è minore di 0.1.

b) Per determinare l'I.C. a livello 0.98, come al solito calcoliamo

$$\begin{aligned} 1-\alpha &= 0.98 \\ \alpha &= 0.02 \\ \frac{\alpha}{2} &= \frac{0.02}{2} = 0.01 \\ 1-\frac{\alpha}{2} &= 1-0.01 = 0.99. \end{aligned}$$

Dalle tavole della distribuzione normale standard, interpolando tra i punti di coordinate (2.32 ; 0.98983) e (2.58 ; 0.99010), si ricava che:

$$\Phi(2.326) = 0.989991 \cong 0.99$$

e quindi

$$z_{1-\frac{\alpha}{2}} = z_{0.99} = 2.326.$$

La stima per la proporzione di alberi malati sulla base delle $n = 100$ osservazioni campionarie è

$$\hat{p} = \frac{40}{100} = 0.4$$

quindi

$$\hat{q} = 1 - \hat{p} = 0.6$$

possiamo scrivere l'I.C.:

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}} \ ; \ \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}\right].$$

Non conoscendo ovviamente i veri valori di p e q, al posto di essi, usiamo le loro stime ($\hat{p}$ e $\hat{q}$), ricavate dal campione e abbiamo quindi l'I.C.

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \ ; \ \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

cioè

$$\left[0.4 - 2.326 \cdot \sqrt{\frac{0.4 \cdot 0.6}{100}} \ ; \ 0.4 + 2.326 \cdot \sqrt{\frac{0.4 \cdot 0.6}{100}}\right]$$

vale a dire:

$$[0.286 \ ; \ 0.514] \qquad \text{I.C. per } p \text{ a livello 0.98.}$$

c) É noto che la varianza dello stimatore $\hat{P}$ per l'ignota proporzione p è

$$var(\hat{P}) = \frac{pq}{n}$$

e, utilizzando l'informazione campionaria:

$$var(\hat{P}) = \frac{\hat{p}\hat{q}}{n}.$$

Quindi, affinchè $var(\hat{P}) = 0.001$, ricordando che $\hat{p} = 0.4$ e $\hat{q} = 0.6$, si deve avere che

$$\frac{0.4 \cdot 0.6}{n} = 0.001$$

cioè

$$n = \frac{0.4 \cdot 0.6}{0.001}$$

vale a dire:

$$n = 240.$$

4. Da un lotto di gelati se ne estraggono $n = 100$ e si stima che il peso medio è pari a 82 g. Sapendo che $\sigma^2 = 25$:

   a) si determini l'intervallo di confidenza per il peso medio μ dei gelati al livello di confidenza del 97%;

   b) si determini la probabilità che la differenza in valore assoluto fra la media campionaria e il peso medio μ dei gelati sia inferiore a 3 g.

Svolgimento

a) Per determinare l'I.C. a livello 0.97, come al solito calcoliamo

$$\begin{aligned} 1-\alpha &= 0.97 \\ \alpha &= 0.03 \\ \frac{\alpha}{2} &= \frac{0.03}{2} = 0.015 \\ 1-\frac{\alpha}{2} &= 1-0.015 = 0.985. \end{aligned}$$

Dalle tavole della distribuzione normale standard, si ricava che:

$$\Phi(2.17) = 0.984997 \cong 0.985$$

e quindi

$$z_{1-\frac{\alpha}{2}} = z_{0.985} = 2.17.$$

La stima per il peso medio dei gelati sulla base delle $n = 100$ osservazioni campionarie è

$$\bar{x} = 82$$

e ricordando che l'intervallo di confidenza per μ a livello $1 - \alpha$ è dato da

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\ ;\ \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\right]$$

sostituendo, otteniamo che

$$\left[82 - 2.17 \cdot \sqrt{\frac{25}{100}}\ ;\ 82 + 2.17 \cdot \sqrt{\frac{25}{100}}\right]$$

è l'I.C. per μ a livello 0.97.

Si ottiene pertanto:

$$[80.915\ ;\ 83.085] \qquad \text{I.C. per } \mu \text{ a livello } 0.97.$$

b) La probabilità richiesta è la seguente:

$$P\{|\bar{X} - \mu| < 3\}.$$

Tale probabilità è però uguale a

$$\begin{aligned} P\{|\bar{X} - \mu| < 3\} &= P\{-3 < \bar{X} - \mu < 3\} \\ &= P\left\{-\frac{3}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{3}{\sigma/\sqrt{n}}\right\}. \end{aligned}$$

A questo punto, ricordando che

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \sim N(0, 1),$$

si ha che la probabilità cercata è pari a:

$$\begin{aligned} P\{|\bar{X} - \mu| < 3\} &= P\left\{-\frac{3}{\sigma/\sqrt{n}} < Z < \frac{3}{\sigma/\sqrt{n}}\right\} \\ &= P\left\{-\frac{3}{\sqrt{25}/\sqrt{100}} < Z < \frac{3}{\sqrt{25}/\sqrt{100}}\right\} \\ &= P\left\{-\frac{3 \cdot 10}{5} < Z < \frac{3 \cdot 10}{5}\right\} \\ &= P\left\{-6 < Z < 6\right\} \\ &= \Phi(6) - \Phi(-6) \\ &= \Phi(6) - [1 - \Phi(6)] \\ &= 2 \cdot \Phi(6) - 1 \\ &= 2 \cdot 1 - 1 \\ &= 1. \end{aligned}$$

La probabilità che la differenza in valore assoluto tra la media campionaria e il peso medio μ dei gelati sia inferiore a 3 g è pari pertanto a 1.

5. Sia X la variabile casuale che descrive il peso dei pacchetti di caffè di un lotto. Dal lotto si estraggono $n = 100$ pezzi e si ottiene:

$$\sum_{i=1}^{100} x_i = 24800 \qquad \sum_{i=1}^{100} x_i^2 = 6152900.$$

Si costruisca l'intervallo di confidenza per il peso medio al livello di confidenza del 97%.

Svolgimento
L'esercizio richiede il calcolo dell'intervallo di confidenza per l'incognita media μ dato da:

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\ ;\ \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\right].$$

Al solito, calcoliamo

$$\begin{aligned} 1-\alpha &= 0.97 \\ \alpha &= 0.03 \\ \frac{\alpha}{2} &= \frac{0.03}{2} = 0.015 \\ 1-\frac{\alpha}{2} &= 1 - 0.015 = 0.985. \end{aligned}$$

Dalle tavole della distribuzione normale standard, si ricava che:

$$\Phi(2.17) = 0.984997 \cong 0.985$$

e quindi

$$z_{1-\frac{\alpha}{2}} = z_{0.985} = 2.17.$$

É facile anche calcolare il valore che la variabile casuale media campionaria $\bar{X}$ assume per il campione estratto:

$$\bar{x} = \frac{1}{100} \cdot \sum_{i=1}^{100} x_i = \frac{24800}{100} = 248.$$

Per costruire l'I.C. sarebbe necessario conoscere la varianza σ^2 della popolazione; non essendo σ^2 nota si impiega lo stimatore "varianza campionaria corretta" (S_c^2) data da:

$$S_c^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Per prima cosa, calcoliamo quindi il valore assunto dalla devianza campionaria in corrispondenza del campione estratto:

$$\begin{aligned} Dev.\, Campionaria &= \sum_{i=1}^{100}(x_i^2 - \bar{x})^2 \\ &= \sum_{i=1}^{100} x_i^2 - 100 \cdot \bar{x}^2 \\ &= 6152900 - 100 \cdot (248)^2 \\ &= 12303300. \end{aligned}$$

Per calcolare il valore che lo stimatore "varianza campionaria corretta" (S_c^2) assume in corrispondenza del campione estratto, dividiamo la devianza campionaria per

$$n - 1 = 100 - 1 = 99,$$

ottenendo

$$s_c^2 = \frac{12303300}{99} = 25.\overline{25}.$$

A questo punto, possiamo stimare la varianza della popolazione di riferimento con il valore $25.\overline{25}$ e quindi la varianza dello stimatore "media campionaria" è data da:

$$var(\bar{X}) = \frac{25.\overline{25}}{100} = 0.\overline{25}.$$

Si è in grado ora di ottenere l'I.C. cercato:

$$\left[248 - 2.17 \cdot \sqrt{0.\overline{25}}\ ;\ 248 - 2.17 \cdot \sqrt{0.\overline{25}}\right]$$

da cui si ricava che:

$$[246.91\ ;\ 249.09] \qquad \text{I.C. per } \mu \text{ a livello } 0.97.$$

6. Sia p la proporzione di individui che preferiscono il prodotto A ad altri prodotti simili. Intervistati 250 consumatori è emerso che 130 di essi dichiarano di preferire il prodotto A rispetto ad altri prodotti simili.

   a) si determini la numerosità campionaria affinchè il valore assoluto della differenza tra lo stimatore e la vera proporzione p sia inferiore a 0.05 con probabilità del 98%;

   b) si costruisca l'intervallo di confidenza per p al livello di confidenza del 97%.

Svolgimento

a) Bisogna determinare n tale che:

$$P\{|\hat{P} - p| < 0.05\} = 0.98.$$

Riscriviamo la relazione precedente come segue:

$$P\{-0.05 < \hat{P} - p < 0.05\} = 0.98$$

$$P\left\{-\frac{0.05}{\sqrt{pq/n}} < \frac{\hat{P} - p}{\sqrt{pq/n}} < \frac{0.05}{\sqrt{pq/n}}\right\} = 0.98. \tag{2}$$

A questo punto, ricordando che

$$\frac{\hat{P} - p}{\sqrt{pq/n}} = Z \sim N(0,1)$$

la relazione (2) diventa

$$P\left\{-\frac{0.05}{\sqrt{pq/n}} < Z < -\frac{0.05}{\sqrt{pq/n}}\right\} = 0.98.$$

Stimando ora p con $\hat{p}$ ($\hat{p} = \dfrac{130}{250} = 0.52$) e q con $\hat{q}$ ($\hat{q} = 1 - 0.52 = 0.48$), si ottiene:

$$P\left\{-\frac{0.05 \cdot \sqrt{n}}{\sqrt{0.52 \cdot 0.48}} < Z < \frac{0.05 \cdot \sqrt{n}}{\sqrt{0.52 \cdot 0.48}}\right\} = 0.98$$

$$P\left\{-\frac{0.05 \cdot \sqrt{n}}{0.5} < Z < \frac{0.05 \cdot \sqrt{n}}{0.5}\right\} = 0.98$$

$$P\left\{-0.1 \cdot \sqrt{n} < Z < 0.1 \cdot \sqrt{n}\right\} = 0.98$$

$$\Phi\left(0.1\sqrt{n}\right) - \Phi\left(-0.1\sqrt{n}\right) = 0.98$$

$$2 \cdot \Phi\left(0.1\sqrt{n}\right) = 1.98$$

$$\Phi\left(0.1\sqrt{n}\right) = 0.99.$$

Consultando le tavole della distribuzione normale standard, si ricava che

$$\Phi(2.326) = 0.989991 \cong 0.99.$$

Quindi

$$0.1\sqrt{n} = 2.326$$

e pertanto:

$$n = \left(\frac{2.326}{0.1}\right)^2 = 541.0276 \qquad \text{(arrotondando } n = 542\text{)}.$$

Tale valore ci informa che affinchè la differenza tra lo stimatore della proporzione p e la proporzione stessa p abbia modulo minore di 0.05 con probabilità pari a 0.98, è necessario intervistare ulteriori $542 - 250 = 292$ individui.

b) Per determinare l'I.C. a livello 0.97, come al solito calcoliamo

$$\begin{aligned} 1-\alpha &= 0.97 \\ \alpha &= 0.03 \\ \frac{\alpha}{2} &= \frac{0.03}{2} = 0.015 \\ 1-\frac{\alpha}{2} &= 1-0.015 = 0.985. \end{aligned}$$

Dalle tavole della distribuzione normale standard, si ricava che:

$$\Phi(2.17) = 0.984997 \cong 0.985$$

e quindi

$$z_{1-\frac{\alpha}{2}} = z_{0.985} = 2.17.$$

La stima per la proporzione di individui che preferiscono il prodotto A ad altri prodotti simili sulla base delle $n = 250$ osservazioni campionarie è

$$\hat{p} = \frac{130}{250} = 0.52$$

e quindi

$$\hat{q} = 1 - \hat{p} = 0.48,$$

possiamo scrivere l'I.C.:

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}\ ;\ \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}\right].$$

Non conoscendo ovviamente i veri valori di p e q, al posto di essi, usiamo delle loro stime, ricavate dal campione e abbiamo quindi l'I.C.

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\ ;\ \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

cioè

$$\left[0.52 - 2.17 \cdot \sqrt{\frac{0.52 \cdot 0.48}{250}}\ ;\ 0.52 + 2.17 \cdot \sqrt{\frac{0.52 \cdot 0.48}{250}}\right]$$

vale a dire:

$$[0.4514\ ;\ 0.5886] \qquad \text{I.C. per } p \text{ a livello } 0.97.$$

7. Si è svolta un'indagine su 100 persone per saggiare l'opinione su una proposta politica. Avendo ottenuto 48 risposte favorevoli:

   a) si determini l'intervallo di confidenza per la proporzione di risposte favorevoli nella popolazione con un livello di confidenza del 97%;

b) si determini quanto deve essere l'ampiezza campionaria se si vuole che la varianza dello stimatore della suddetta proporzione non sia superiore a 0.001, tenendo conto dei risultati ottenuti dall'indagine sulle 100 persone.

Svolgimento

a) Per determinare l'I.C. a livello 0.97, come al solito calcoliamo

$$\begin{aligned} 1-\alpha &= 0.97 \\ \alpha &= 0.03 \\ \frac{\alpha}{2} &= \frac{0.03}{2} = 0.015 \\ 1-\frac{\alpha}{2} &= 1-0.015 = 0.985. \end{aligned}$$

Dalle tavole della distribuzione normale standard, si ricava che:

$$\Phi(2.17) = 0.984997 \cong 0.985$$

e quindi

$$z_{1-\frac{\alpha}{2}} = z_{0.985} = 2.17.$$

La stima per la proporzione di persone che hanno un'opinione favorevole alla proposta politica sulla base delle $n = 100$ osservazioni campionarie è

$$\hat{p} = \frac{48}{100} = 0.48$$

e quindi

$$\hat{q} = 1 - \hat{p} = 0.52$$

possiamo scrivere l'I.C.:

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}} \; ; \; \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}\right].$$

Non conoscendo ovviamente i veri valori di p e q, al posto di essi, essendo l'ampiezza campionaria n sufficientemente elevata, usiamo delle loro stime, ricavate dal campione e abbiamo quindi l'I.C.

$$\left[\hat{P} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \; ; \; \hat{P} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

cioè

$$\left[0.48 - 2.17 \cdot \sqrt{\frac{0.48 \cdot 0.52}{100}} \; ; \; 0.48 + 2.17 \cdot \sqrt{\frac{0.48 \cdot 0.52}{100}}\right]$$

vale a dire:

$$[0.3715 \; ; \; 0.5885] \qquad \text{I.C. per } p \text{ a livello } 0.97.$$

b) Ricordando che
$$var(\hat{P}) = \frac{pq}{n}$$
basta imporre che
$$\frac{pq}{n} < 0.001.$$
Tenendo quindi conto delle informazioni campionarie, ponendo cioè
$$p = \hat{p} = 0.48$$
e di conseguenza
$$q = \hat{q} = 1 - p = 0.52,$$
si ottiene
$$\frac{0.48 \cdot 0.52}{n} < 0.001.$$
Dalla precedente relazione si ricava che
$$n > \frac{0.48 \cdot 0.52}{0.001}$$
vale a dire
$$n > 249.6 \qquad \text{(arrotondando } n \geq 250\text{).}$$
Intervistando 250 individui si ha che la varianza dello stimatore della ignota proporzione p di persone favorevoli alla proposta politica è inferiore a 0.001.