CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 4

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. Stimatore media campionaria

Il tempo in minuti necessario a un certo impiegato dell'anagrafe per sbrigare le pratiche ha una media aritmetica $\mu = 20$ e uno scarto quadratico medio $\sigma = 5$. Se si estrae un campione casuale di 50 pratiche d'ufficio:

1. qual è la probabilità che il tempo impiegato per ciascuna pratica sia, in media, non superiore a 18 min.?
2. qual è la probabilità che il tempo impiegato per ciascuna pratica sia, in media, superiore a 22 min.?
3. qual è il numero di minuti al di sotto del quale la media aritmetica del campione si colloca con probabilità 0.95?

Sol.

Indicando con X il tempo necessario all'impiegato,

Stimatore media campionaria:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Distribuzione dello stimatore media campionaria (risultato del TLC):

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Quindi:

$$\bar{X} \sim N\left(20, \frac{5}{\sqrt{50}}\right)$$

Pertanto:

1. $P(\bar{X} \leq 18) = P\left(Z \leq \frac{18-20}{5/\sqrt{50}}\right) = P(Z \leq -2.83) = 0.0023$

2. $P(\bar{X} > 22) = P\left(Z > \frac{22-20}{5/\sqrt{50}}\right) = P(Z > 2.83) = 0.0023$

3. si tratta di trovare quel valore x tale che: $P(\bar{X} < x) = 0.95$ ovvero tale che:

$$P\left(Z \leq \frac{x-20}{5/\sqrt{50}}\right) = 0.95$$

Consultando la tavola della normale standard troviamo che:

$$\frac{x-20}{5/\sqrt{50}} = 1.645$$

Per cui: $x = 20 + 1.645(5/\sqrt{50}) = 21\ minuti\ (arrotondando)$

Esercizio 2. Stimatore proporzione campionaria

Il 9% delle lenti da vista prodotte da una certa azienda presenta delle imperfezioni. Estraendo dalla produzione campioni casuali di 300 lenti, in quale proporzione i campioni presenteranno una frequenza di pezzi difettosi minore o uguale al 7%?

Sol.

Ogni singola estrazione X_i rappresenta una v.c. bernoulliana di parametro p. La somma di n v.c. bernoulliane i.i.d. e con lo stesso parametro p è una v.c. binomiale $Y \sim Bin(n, p)$

dove si è indicato con n il numero di lenti da vista estratte e con p la probabilità di estrarre una lente che presenta delle imperfezioni. Inoltre sappiamo che:

$E(X) = np \qquad Var(X) = np(1-p)$

A partire della variabile casuale X, possiamo ricavare lo stimatore:

$$\hat{P} = \frac{Y}{n}$$

Siccome n è sufficientemente grande (per il TLC) troviamo:

$$\hat{P} = \frac{Y}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Infatti:

$$E(\hat{P}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{np}{n} = p = 0.09$$

$$Var(\hat{P}) = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2}Var(Y) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} = 0.00027$$

$$P(\hat{P} \leq 0.07) = \left(Z \leq \frac{0.07 - 0.09}{\sqrt{\frac{0.09(1-0.91)}{300}}} \right) = P(Z \leq -1.21) = 0.1131$$

Esercizio 3. IC per la media incognita (varianza nota), determinazione ampiezza campionaria

Siete i responsabili delle analisi statistiche dell'Ufficio Immatricolazioni dell' Università di Roma e state lavorando alla stesura di un report da pubblicare sul sito ufficiale della facoltà per l'anno accademico 2012/2013. Da indagini precedenti è noto che l'altezza delle matricole universitarie di sesso maschile può essere considerata una variabile con distribuzione Normale, con media incognita e varianza pari a 10.66. Siete interessati ad avere informazioni sull'altezza media. A tale proposito estraete un campione casuale di 58 matricole e misurate l'altezza media, che risulta pari a 175.4 cm.

a) Costruite un intervallo di confidenza al 95% per la media incognita.

Supponiamo ora che in Ufficio il capo vi chieda maggiore precisione nel fornire i risultati e a tale proposito stabilisca che, per l'intervallo di confidenza al 95% per la media stimato, il margine di errore massimo tollerabile sia di 1.2 cm.

b) Quale numerosità deve avere il campione per soddisfare tale richiesta?

A questo punto, per avere maggiore fiducia che la media incognita sia compresa nell'intervallo, il capo vi chiede di costruire un intervallo di confidenza al 99%. Volendo mantenere lo stesso margine di errore tollerabile,

c) quanti altri studenti in più dovete considerare nel campione?

Sol.

a) Per determinare l'intervallo di confidenza (I.C.) per μ a livello di confidenza pari al 95%, bisogna innanzitutto ricavare il quantile di ordine $1 - \frac{\alpha}{2}$ della distribuzione normale standard.

Quindi, poiché:

$$1 - \alpha = 0.95$$

si ha che:

$$\alpha = 1 - 0.95 = 0.05$$

perciò:

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

e di conseguenza:

$$1 - \frac{\alpha}{2} = 1 - 0.0025 = 0.975$$

Consultando le tavole della distribuzione normale standard si ricava che $z_{0.975} = 1.96$

Indicando con α l'area nelle code e con 1-α il livello di confidenza desiderato, allora:

$$P\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

IC per la media incognita

$$\left[175.4 - 1.96\frac{\sqrt{10.66}}{\sqrt{58}}; 175.4 + 1.96\frac{\sqrt{10.66}}{\sqrt{58}}\right]$$

$$[175.4 - 0.840; 175.4 + 0.840]$$

$$[174.56; 176.24]$$

Interpretazione: c'è una fiducia del 95% che il valore medio della popolazione si trovi tra 174.56 cm e 176.24 cm.

b) In generale, la lunghezza (ampiezza) dell'intervallo di confidenza è data da:

$$b = 2 * z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Occorre determinare la numerosità campionaria affinchè l'intervallo di confidenza per la media al 95% non sia più lungo di un valore fissato b=1.2

$$\sqrt{n} \geq 2 * z_{\frac{\alpha}{2}}\frac{\sigma}{b}$$

Ossia n deve essere almeno pari a $4z_{\alpha/2}^2\frac{\sigma^2}{b^2}$

Sostituendo nel nostro caso $z_{\frac{\alpha}{2}} = 1.96$, $\sigma^2 = 10.66$, b=1.2 si ottiene

$$n > 113.75 \approx 114$$

Per avere maggiore precisione dell'IC è necessario aumentare la numerosità campionaria a parità di livello di confidenza.

c) IC al 99% per la media incognita

Consultando le tavole della distribuzione normale standard si ricava che $z_{0.995} = 2.58$

Indicando con α l'area nelle code e con 1-α il livello di confidenza desiderato, allora:

$$P\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 0.99$$

IC per la media incognita

$$\left[175.4 - 2.58\frac{\sqrt{10.66}}{\sqrt{58}}; 175.4 + 2.58\frac{\sqrt{10.66}}{\sqrt{58}}\right]$$

$$[175.4 - 1.106; 175.4 + 1.106]$$

$$[174.294; 176.506]$$

Interpretazione: c'è una fiducia del 99% che il valore medio della popolazione si trovi tra 174.29 cm e 176.51 cm.

Osservazione: un aumento del livello (1-α) si ottiene solo al prezzo di un ampliamento dell'IC; in altre parole, l'IC diventa più ampio con l'aumentare della precisione richiesta

Se si vuole mantenere lo stesso margine d'errore di cui al punto b) allora la numerosità campionaria sarà pari a:

$$\sqrt{n} \geq 2 * z_{\frac{\alpha}{2}}\frac{\sigma}{b}$$

Ossia n deve essere almeno pari a $4z_{\alpha/2}^2\frac{\sigma^2}{b^2}$

Sostituendo nel nostro caso $z_{\frac{\alpha}{2}} = 2.58$, $\sigma^2 = 10.66$, b=1.2 si ottiene

$$n > 197.10 \approx 197$$

E' necessario selezionare 53 studenti maschi in più rispetto al numero di studenti considerati nel punto b).

Osservazione: aumentando il livello di confidenza e volendo mantenere invariata l'ampiezza dell'intervallo è necessario aumentare la numerosità campionaria.

Esercizio 4. IC per la media incognita (varianza ignota)

Il regolamento delle Ferrovie dello Stato stabilisce che per alcuni treni regionali l'orario effettivo di partenza non debba superare di oltre 5 minuti quello ufficiale. Durante l'anno è stato osservato un campione casuale semplice di n = 20 orari effettivi di partenza. Per ognuno di essi sono stati calcolati i minuti di ritardo rispetto all'orario ufficiale e si suppone ragionevolmente che seguano una distribuzione normale. La media campionaria dei ritardi osservati è pari a 4.75 minuti, mentre la varianza campionaria corretta è di 2.848 minuti.

Sol.

$(X_1, X_2, \ldots, X_n)$ è un campione casuale estratto da una v.c. $X \sim N(\mu, \sigma^2)$ con varianza σ^2 incognita, inoltre la media campionaria è

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

e la varianza campionaria corretta è

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

Da ciò segue che:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}} \sim t_{n-1}$$

$t_{n-1,\alpha}$ indica il quantile α della distribuzione t con n-1 g.d.l.

Intervallo casuale

$$P\left(t_{n-1,\alpha/2} \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S^2}} \le t_{n-1,\alpha/2}\right) = 1 - \alpha$$ ovvero

$$P(\left(\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Intervallo di confidenza:

sapendo che:

- $\bar{X} = 4.75$
- $\sqrt{S^2} = \sqrt{2.848} = 1.688$
- $t_{n-1,\alpha/2} = 2.093$ con α=0.05, n=19

$$\left[4.75 - 2.093\frac{1.688}{\sqrt{20}}; 4.75 + 2.093\frac{1.688}{\sqrt{20}}\right]$$

IC al 95% per la media incognita è quindi pari a:

$$[3.961; 5.539]$$