ESERCIZIO 1

La somma dei numeri di passeggeri di un autoservizio suburbano rilevati in 7 giorni scelti a caso è pari a 385. Si assuma un modello gaussiano per descrivere il numero giornaliero di passeggeri.
a) Si fornisca, sulla base della realizzazione campionaria osservata, una stima puntuale del numero medio di passeggeri giornalieri. Si dica qual è la distribuzione dello stimatore che ha fornito la stima richiesta.
b) Sapendo che la deviazione standard campionaria è pari a 7.64, si determini un intervallo di confidenza al 99% per il numero medio giornaliero di passeggeri.

Soluzione

X =numero di passeggeri giornaliero

a. Stima puntuale
$$\overline{x} = \frac{385}{7} = 55.$$
Lo stimatore è la media campionaria, che ha distribuzione $N(\mu, \sigma^2/7)$, con μ, σ incogniti.

b. L'intervallo al 99% è
$$(\overline{x} - t_{\alpha/2}^{n-1}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}^{n-1}\frac{s}{\sqrt{n}})$$
dove

$\overline{x} = 55$ $\qquad n = 7$ $\qquad t_{\alpha/2}^{n-1} = t_{0,005}^{6} = 3,71$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n-1}} = 7,64$$

Quindi l'intervallo è
$$(55 - 3,71\frac{7,64}{\sqrt{7}}; 55 + 3,71\frac{7,64}{\sqrt{7}}) = (55 - 10,71; 55 + 10,71) = (44,29; 65,71)$$

ESERCIZIO 2

Per stimare il tempo medio con cui viene svolto un certo test attitudinale a tempo, viene estratto un campione di 60 studenti e, per essi, viene rilevato il tempo di esecuzione del test. I dati rilevati forniscono un tempo medio pari a 12 minuti ed una varianza pari a 15. Si determini un intervallo di confidenza per il tempo medio di esecuzione del test, prendendo come livello di confidenza 0.95.

Soluzione

X = tempo esecuzione test,

$X_1, ..., X_{60}$ campione ($n = 60$ è sufficientemente grande, quindi $\overline{X} \approx N(\mu, \frac{\sigma^2}{n})$)

L'intervallo è

$$(\overline{x} \mp z_{\alpha/2}\frac{s}{\sqrt{n}}) = (12 - 1,96\sqrt{\frac{15}{60}}; 12 + 1,96\sqrt{\frac{15}{60}}) = (12 - 0,98; 12 + 0,98) = (11,02; 12,98)$$

ESERCIZIO 3

Per stimare la durata media delle telefonate effettuate in una certa fascia oraria, una compagnia telefonica sceglie a caso un campione di 1200 telefonate, rilevando una durata media di queste pari a 4.7 minuti, con una varianza delle stesse pari a 4.5. Si determinino intervalli di confidenza per la durata media di tutte le telefonate nella fascia oraria considerata di livelli 0.9, 0.95 e 0.99. Quale intervallo è il più preciso?

Soluzione

X = durata telef. nella fascia considerata

$X_1, ..., X_{1200}$ (n quindi grande, quindi $\overline{X}$ si può approssimare alla normale)

Realizzazone campionaria: $\bar{x} = 4,7$, $s^2 = 4,5$

Intervallo al 90%:

$(4,7 - 1,645\sqrt{\frac{4,5}{1200}}; 4,7 + 1,645\sqrt{\frac{4,5}{1200}}) = (4,7 - 0,1; 4,7 + 0,1) = (4,6; 4,8)$

Intervallo al 95%:

$(4,7 - 1,96\sqrt{\frac{4,5}{1200}}; 4,7 + 1,96\sqrt{\frac{4,5}{1200}}) = (4,7 - 0,12; 4,7 + 0,12) = (4,58; 4,82)$

Intervallo al 99%:

$(4,7 - 2.576\sqrt{\frac{4,5}{1200}}; 4,7 + 2,576\sqrt{\frac{4,5}{1200}}) = (4,54; 4,86)$

L'intervallo più preciso é quello di ampiezza inferiore, quindi con un minor livello di confidenza. Pertanto è l'intervallo al 90%: (4,6;4,8).

ESERCIZIO 4

Un'indagine sui possessori di carta di credito, effettuata su un campione di ampiezza 20 , ha fornito una spesa mensile media con la carta di credito pari a 820 euro, con uno scarto quadratico medio pari a 156. Si determini un intervallo di confidenza al 95% per la spesa media mensile nell'ambito di tutti i possessori di carta di credito. E' necessaria qualche assunzione per la determinazione dell'intervallo?

Soluzione

$X =$ spesa mensile.

$X_1, ..., X_{20}$ campione.

$\overline{x} = 820, \qquad s = 156. \qquad$ Essendo il campione piccolo, è necessario assumere $X \sim N(\mu, \sigma^2)$

Intervallo al 95%:

$(\overline{x} - t^{n-1}_{\alpha/2} \frac{s}{\sqrt{n}}; \overline{x} + t^{n-1}_{\alpha/2} \frac{s}{\sqrt{n}}) = (820 - 2,09 \frac{156}{\sqrt{20}}, 820 + 2,09 \frac{156}{\sqrt{20}}) = (820 - 72,9; 820 + 72,9) = (747,1; 892,9)$

ESERCIZIO 5

Allo scopo di stimare la proporzione di giovani (di una certa fascia di età) che vanno a teatro almeno una volta al mese, si intervistano 200 giovani, dei quali 18 dichiarano di andare a teatro almeno una volta al mese.
a) Sulla base di questi dati, si determini una stima puntuale per la proporzione cercata e si dica qual è la distribuzione dello stimatore che ha fornito tale stima.
b) Si determini un intervallo di confidenza al 90% per la proporzione cercata e se ne fornisca una interpretazione.

Soluzione

$X \sim Bern(p)$, $p =$ proporzione di giovani che vanno a teatro almeno una volta al mese

$X_1, ..., X_{200}$ campione da distribuzione Bernoulliana di parametro p

a) La stima puntuale è $\overline{x} = \widehat{p} = \frac{18}{200} = 0,09$
Lo stimatore corrispondente è $\hat{P}$, che ha (approssimativamente, per il teorema centrale del limite) distribuzione normale con media p e varianza $p(1-p)/200$.
b) L'intervallo al 90% è

$(\hat{p} \mp z_{\alpha/2} \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}) = (0,09 \mp 1,645 \sqrt{\frac{0,09 \cdot 0,91}{200}}) = (0,057; 0,123)$
Possiamo dire che, con fiducia del 90%, la proporzione di giovani della fascia d'età considerata che vanno a teatro almeno una volta al mese è compresa tra 0.057 e 0.123.

ESERCIZIO 6

Per stimare la spesa media μ per il pasto di mezzogiorno sostenuta dai lavoratori pendolari in una città si rilevano le spese di un campione di 20 pendolari. Si assume una distribuzione gaussiana per descrivere la spesa (in euro), con una deviazione standard pari a 0.8.
a) Se la spesa media rilevata sul campione di 20 pendolari è stata pari a 5.3 euro, qual è l'intervallo di confidenza al 90% per la spesa media di tutti i pendolari della città?
b) Immaginiamo ora, in una simulazione, di ripetere 800000 volte l'operazione di campionamento, estraendo cioè 800000 campioni casuali, ciascuno costituito da 20 pendolari. In corrispondenza a ciascuno degli 800000 campioni, calcoliamo la spesa media dei 20 pendolari e costruiamo l'intervallo di confidenza al 90% per μ. Possono esserci, tra questi 800000, intervalli che non contengono μ? Se sì, quanti ci aspettiamo che siano? Giustificare opportunamente le risposte.

Soluzione

$X =$ spesa pasto; $X \sim N(\mu, 0,8^2)$

$X_1, ..., X_{20}$ campione da $N(\mu, 0,8^2)$

a) $\overline{x} = 5,3$; l'intervallo al 90% è:

$$(\overline{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = (5,3 \mp 1,645\frac{0,8}{\sqrt{20}}) = (5,01; 5,59)$$

b) Sì, possono esserci intervalli che non contengono μ, in percentuale sono il 10% circa, cioè circa 80000, in quanto l'intervallo è al 90% e, quindi, la probabilità di estrarre un campione per il quale l'intervallo corrispondente non contiene μ è $0,1$.

ESERCIZIO 7

Un processo che produce CD vergini è ritenuto sotto controllo se la probabilità che un CD prodotto sia difettoso non supera 0.03 (in altri termini, se la percentuale complessiva di CD difettosi prodotti non è superiore al 3%). Si rileva un campione di 300 CD prodotti, dei quali 12 sono difettosi.
a) Si fornisca una stima puntuale della proporzione di CD difettosi prodotti dal processo.
b) L'addetto al controllo del processo, sulla base, **esclusivamente**, della stima al punto precedente, decide di bloccare l'impianto, ritenendolo fuori controllo in quanto la stima ottenuta sul campione supera la soglia di accettabilità stabilita (il 3%). E' ragionevole un comportamento di questo tipo? Si giustifichi la risposta.
c) Si determini un intervallo di confidenza al 99% per la proporzione di CD difettosi prodotti dal processo. Qual è il margine di errore di tale intervallo? Questo margine di errore varia al variare della realizzazione del campione?

Soluzione

$X \sim Bern(p)$, $p =$ prop. CD difettosi

a) $\overline{x} = \widehat{p} = \frac{12}{300} = 0,04$

b) No, non si può basare una decisione riguardante il parametro incognito p basandosi esclusivamente sulla stima puntuale, che dipende dal campione osservato; osservando un diverso campione, la realizzazione di $\hat{P}$ avrebbe potuto essere inferiore a 0.3. Per effettuare valutazioni ragionevoli, occorre tener conto della variabilità campionaria e, quindi, della distribuzione della statistica che ha fornito la stima.

c) L'intervallo è $(0,011; 0,069)$. Il margine di errore è $z_{\alpha/2}SE = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/300} = 2.576 * 0.011 = 0.029$. Tale margine di errore varia al variare della realizzazione campionaria, in quanto varia lo standard error (che dipende dalla realizzazione $\hat{p}$).