Corso di Statistica per Ingegneria Fisica
Anno accademico 2020/2021

Esercitazione 7 - Intervalli di confidenza per media, varianza e proporzione

Esercizio 1. La resa di un impianto chimico viene sottoposta a uno studio. In base all'esperienza precedente, si sa che la resa giornaliera dell'impianto è una variabile aleatoria gaussiana con media incognita μ e varianza uguale a 2. Inoltre, gli ultimi cinque giorni di operatività hanno dato le rese seguenti (in %):

$$91.60 \qquad 88.75 \qquad 90.80 \qquad 89.95 \qquad 91.30\,.$$

(a) Costruire un intervallo di confidenza bilatero per il parametro μ al 95%. Quanto vale l'ampiezza dell'intervallo di confidenza? *[$IC_\mu(0.95) = (89.240\,,\,91.720)$; ampiezza = 2.480]*

(b) Al livello di confidenza del 95%, qual è l'errore massimo che si commette stimando μ con la media campionaria delle cinque osservazioni precedenti? *[1.240]*

(c) Dal momento che più è ampio l'intervallo di confidenza per μ, maggiore è il costo che deve sostenere l'azienda, il capo reparto dice agli statistici che l'ampiezza dell'intervallo non deve essere superiore a 2.2. A questo scopo, è disposto a fornire loro ulteriori osservazioni, purché però si mantenga invariato il livello di confidenza del 95%. Quante nuove misure deve fornire il capo reparto affinché la sua richiesta venga rispettata? *[Almeno altre 2 misure.]*

Esercizio 2. Il peso in Kg dei sacchetti di farina prodotti dal mulino del Signor Sabbioso può essere descritto con una variabile aleatoria normale di media μ e deviazione standard pari a 0.5. Vengono pesati 100 sacchetti trovando un peso medio campionario $\bar{x}_n = 15.3$ Kg.

(a) Si costruisca un intervallo di confidenza bilatero per μ al livello del 90% e si indichi l'errore massimo commesso stimando μ con $\bar{x}_n$. *[(15.3 ± 0.08225) Kg; 0.08225 Kg]*

(b) Se, mantenendo lo stesso livello di confidenza del 90%, al punto (a) avessimo voluto trovare un intervallo di ampiezza inferiore a 0.1 Kg, come avremmo dovuto scegliere la numerosità del campione? *[$n \geq 271$]*

(c) Quanto può essere confidente il Signor Sabbioso che il peso medio dei suoi sacchetti di farina sia maggiore di 15.2 Kg? *[97.725%]*

(d) In tutte le risposte precedenti, è essenziale l'ipotesi di normalità? *[No]*

Esercizio 3. Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione normale di media μ incognita e varianza $\sigma^2 = 16$.

(a) Sapendo che $n = 9$ e che $(2.387, 7.613)$ è un intervallo di confidenza simmetrico per μ, se ne determini il livello di confidenza. *[$\gamma = 95\%$]*

(b) Si determini la minima numerosità n del campione tale che un intervallo di confidenza per μ di livello 0.95 non risulti più lungo di 4. *[16]*

(c) Nelle situazioni concrete, aumentare la numerosità del campione ha un costo; d'altra parte, costa anche avere stime imprecise. Supponiamo che il costo di ogni misurazione sia di 25 €, e che il costo di ogni unità di ampiezza dell'intervallo sia di 5200 €. Si determini n in modo da rendere minimo il costo totale di un intervallo di confidenza per μ di livello 0.95 *[$n = 139$]*.

Esercizio 4. Sulle scatole di puntine da disegno prodotte dalla PUNT è dichiarato che il contenuto della scatola è di 100 puntine per ogni scatola. Un ispettore di qualità, che deve certificare la PUNT secondo lo standard ISO9002, allo scopo di verificare la veridicità dell'affermazione precedente esamina il contenuto di 100 scatole; ne ottiene una media campionaria di 95.3 puntine per scatola. Si supponga che la quantità di puntine nelle scatole segua una densità incognita, di cui si sa solo che la deviazione standard è $\sigma = 18.25$.

(a) Si determini un intervallo di confidenza di livello 95% per il numero medio di puntine in ciascuna scatola. *[*95.3 ± 3.577*]*

(b) Quante scatole avrebbe dovuto esaminare l'ispettore se avesse voluto un intervallo di confidenza di livello 95% di lunghezza non superiore a 4? *[*$n \geq 320$*]*

Esercizio 5. Si consideri la densità di probabilità

$$f(t) = \frac{3}{4}\Big[1 - (t + 1 - \theta)^2\Big] 1_{(\theta-2,\theta)}(t),$$

dove θ è un numero reale incognito.

(a) Si disegni il grafico di f.

(b) Si calcoli il valore atteso μ e la varianza σ^2 di una variabile aleatoria con densità f. *[*$\mu = \theta - 1$, $\sigma^2 = 1/5$*]*

(c) Dato un campione $X_1, \dots, X_n$ di variabili i.i.d. con densità f, si costruisca uno stimatore T_n non distorto per θ. *[*$T_n = \overline{X}_n + 1$*]*

(d) Si calcoli l'errore quadratico medio dello stimatore T_n del punto precedente. *[*$\text{mse}(T_n; \theta) = \frac{1}{5n}$*]*

(e) Lo stimatore T_n è consistente in media quadratica? *[Sì]*

(f) Dato un campione di numerosità 50 la cui media vale 15.13, si forniscano una stima puntuale per θ e un intervallo di confidenza per μ di livello 90%. *[*$\hat{\theta} = 16.13$*;* $\mu \in (15.13 \pm 0.10)$*]*

Esercizio 6. Si considerino due campioni casuali, indipendenti fra di loro, il primo dei quali, $X_1, \dots, X_n$, viene estratto da una $N(\mu, 1)$, mentre il secondo, $Y_1, \dots, Y_m$, viene estratto da una $N(\mu, 3)$. Le due popolazioni hanno quindi lo stesso valore medio μ, che però è incognito. Si consideri come stimatore di μ la seguente variabile aleatoria:

$$T = \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{n+m}.$$

(a) T è uno stimatore corretto di μ? *[Sì]*

(b) Si determini la densità di T. *[*$T \sim N\Big(\mu, \dfrac{n+3m}{(n+m)^2}\Big)$*]*

(c) Per ogni $1/2 < \alpha < 1$, si calcoli il livello del seguente intervallo di confidenza per μ:

$$\mu \in \left(t - z_\alpha \sqrt{\frac{n+3m}{(n+m)^2}}\,,\, t + z_\alpha \sqrt{\frac{n+3m}{(n+m)^2}} \right),$$

dove t è la realizzazione della variabile aleatoria T sul risultato delle $n+m$ misure. *[*$2\alpha - 1$*]*

Esercizio 7. Supponiamo che quando un segnale elettrico di valore μ viene trasmesso dalla sorgente A, il ricevente registri un valore $\mu + N$, dove N, che denota il rumore, è distribuito come una normale di media 0 e varianza 4. Immaginiamo che per ridurre l'errore il segnale sia trasmesso 9volte. I valori registrati in ricezione sono stati

$$5 \quad 8.5 \quad 12 \quad 15 \quad 7 \quad 9 \quad 7.5 \quad 6.5 \quad 10.5\,.$$

(a) Calcolare un intervallo di confidenza simmetrico al 99% per μ, e i due intervalli di confidenza unilaterali dello stesso livello. *[*$\bar{x} = 9$*;* $(7.28; 10.72)$*;* $(7.45; +\infty)$*;* $(-\infty; 10.55)$*.]*

Si supponga ora di non conoscere la varianza di N.

(b) Determinare gli intervalli di confidenza bilaterale e unilaterali al livello 99% per μ. *[*$s^2 = 9.5$*;* $(5.55; 12.45)$*;* $(6.02; \infty)$*;* $(-\infty; 11.98)$*.]*

Esercizio 8. Vengono effettuate 20 misurazioni della concentrazione di un certo enzima nel sangue di diversi individui e si osservano una media campionaria $\bar{x} = 1.23$ ed una varianza campionaria $s^2 = 0.4$.

(a) Supponendo che i valori di questa concentrazione seguano una distribuzione normale (con entrambi i parametri incogniti!), qual è un intervallo di confidenza bilatero al livello 95% per la media della concentrazione? *[*$IC_\mu(95\%) = (0.934\,,\,1.526)$

(b) Quale sarebbe l'intervallo di cui sopra se la concentrazione a cui si è interessati fosse distribuita come una normale di varianza nota $\sigma^2 = 0.4$? *[IC= [0.953;1.507]]*

(c) Quale tra i due intervalli trovati in (a) e in (b) è più ampio? Il risultato ottenuto sembra ragionevole? *[L'intervallo in (a). Sì; però perchè?]*

Esercizio 9. Avendo osservato il campione, proveniente da una legge normale,

$$2.28 \quad 3.45 \quad 5.62 \quad 6.12 \quad 4.26 \quad 6.30 \quad 7.09 \quad 1.97 \quad 5.77 \quad 7.68 \quad 8.16$$

(a) si fornisca una stima puntuale della media e della varianza, utilizzando stimatori non distorti; *[*$\bar{x}_n = 5.336$*;* $s^2 = 4.374$*]*

(b) si calcoli l'intervallo di confidenza per la media al 90%. *[*(5.336 ± 1.143)*]*

Esercizio 10. Il Centro di Ricerche Spaziali (CRS) vuole determinare la distanza d della stella Vega dalla Terra. Le sue misure X_1 sono affette da errore casuale $\epsilon \sim N(0, \sigma^2)$, dove la deviazione standard σ è incognita. Il CRS sulla base di $n = 15$ misure indipendenti stima d con $(25.09\,,\,25.35)$, intervallo di confidenza bilatero simmetrico di livello 95%.

Quanto valgono le stime puntuali di d e σ fornite dal campione? *[*$\bar{x}_{10} = \frac{25.09+25.35}{2} = 25.22$ *e* $s_{10} = \dfrac{(25.35 - 25.09)\sqrt{15}}{2 \cdot 2.1448} = 0.2347$*]*

Esercizio 11. In uno studio sull'inquinamento atmosferico effettuato da una certa stazione sperimentale, si sono registrati, su 8 diversi campioni di aria, i seguenti valori di una certa sostanza tossica (in microgrammi per metro cubo):

$$2.2\,;\ 1.8\,;\ 3.1\,;\ 2.0\,;\ 2.4\,;\ 2.0\,;\ 2.1\,;\ 1.2$$

(a) Assumendo che la popolazione campionaria sia normale, si stabilisca un intervallo di confidenza per la media, al livello del 95%. *[*$IC_\mu(0.95) = 2.1 \pm 0.45$*]*

(b) L'ipotesi di normalità della popolazione è essenziale per giustificare il procedimento seguito? *[Sì, perchè la numerosità del campione non è elevata, dunque la densità della sua media campionaria non può essere approssimata con una normale.]*

Esercizio 12. In seguito ad un controllo medico in una scuola elementare durante un'epidemia di morbillo, 9 bambini su 75 risultano aver contratto il virus. Detta p la probabilità che un bambino della stessa scuola sia malato, determinare:

(a) una stima puntuale $\hat{p}$ per p; *[$\hat{p} = 0.12$]*

(b) un intervallo di confidenza per p di livello 0.95 simmetrico rispetto a $\hat{p}$; *[$IC_p(95\%) = (0.046, 0.194)$]*

(c) un intervallo di confidenza per p di livello 0.9 del tipo $(0, a)$. *[$IC_p(90\%) = (0\,,\, 0.168)$]*

Esercizio 13. Il Sindaco Will Piedebianco vuole conoscere la proporzione p di abitanti della Contea contrari al consumo di erba-pipa. Fredegario Bolgeri riceve pertanto l'incarico di intervistare un campione casuale di n abitanti per stimare p sulla base della proporzione $\bar{x}_n$ di contrari al consumo di erba-pipa fra gli intervistati.

(a) Si trovi la minima numerosità n_0 del campione grazie alla quale un intervallo di confidenza di livello 90% simmetrico rispetto a $\bar{x}_n$ ha ampiezza sicuramente inferiore a 0.05. *[$n_0 = 1083$]*

Dopo aver chiesto il parere di n_0 abitanti della Contea scelti a caso, Fredegario Bolgeri trova ben 268 contrari al consumo di erba-pipa! Si stimi p mediante:

(b) uno stimatore puntuale non distorto; *[$\hat{p} = 0.2475$]*

(c) un intervallo di confidenza bilatero di livello 90%; *[$IC_p(90\%) = (0.2475 \pm 0.0216)$]*

(d) un intervallo di confidenza di livello 90% di tipo $(0, a)$. *[$IC_p(90\%) = (0\,,\, 0.2642)$]*

Esercizio 14. L'Assessorato all'Ecologia di Pietraforata vuole conoscere la proporzione p di abitanti favorevoli all'iniziativa delle "domeniche ecologiche". Per questo motivo si decide di intervistare un campione di n abitanti, scelti a caso e giudicati rappresentativi della popolazione, per stimare p sulla base della proporzione $\bar{x}_n$ di risposte favorevoli.

(a) Si stabilisca quale deve essere la numerosità minima n_0 del campione affinchè un intervallo di confidenza bilatero per p di livello 95% abbia ampiezza sicuramente inferiore a 0.05. *[$n_0 = 1537$]*

(b) Eseguita l'indagine, 882 degli n_0 abitanti intervistati si sono dichiarati favorevoli alle "domeniche ecologiche"; si stimi p sia con una stima puntuale sia con un intervallo di confidenza bilatero di livello 95%. *[$\hat{p} = 0.5738; IC_p(95\%) = (0.5491\,,\, 0.5986)$]*

(c) Si supponga ora $p = 0.65$. Con quale probabilità un'indagine condotta su n_0 abitanti fornisce un errore di stima puntuale inferiore a 0.05? *[$\mathbb{P}\left(\left|\bar{X}_{n_0} - p\right| < 0.05\right) = 2\Phi(4.11) - 1 = 99.996\%$]*

Esercizio 15. Per una proiezione dei risultati elettorali, sono state esaminate 3000 schede, estratte a caso tra le schede da scrutinare. Il partito A ha ottenuto 600 voti. Determinare un intervallo di confidenza al 95% (proiezione) per la percentuale di voti ottenuti dal partito A. *[$(20 \pm 1.4)\%$]*

Esercizio 16. Una ditta riceve un lotto molto numeroso di lampade. Per giudicarne la qualità, ne esamina 60 scelte a caso, trovandone 8 difettose.

(a) Fornire un intervallo di confidenza bilatero di livello 90% per la proporzione p di lampade difettose nel lotto. *[$IC_p(90\%) = (0.06114\,,\, 0.20552)$]*

(b) Per una successiva indagine, quanto deve valere la numerosità n del campione affinché l'intervallo di confidenza per p (sempre di livello $0,9$) risulti avere ampiezza inferiore a 0.1? *[$n \geq 271$]*

Esercizio 17. La Società degli Acquedotti del comune di Pietraforata vuole determinare il pH dell'acqua che circola nei suoi impianti. Per questo motivo, incarica un chimico di misurare il pH di 10 campioni provenienti da punti diversi della città, e, in base al risultato delle misure, gli chiede di stimare la media e la varianza della variabile aleatoria

$$X = \text{pH dell'acqua in un punto a caso dell'acquedotto}\,.$$

Il chimico trova nei campioni i seguenti valori di pH:

$$7.94 \quad 8.18 \quad 6.76 \quad 9.43 \quad 8.23 \quad 7.82 \quad 8.01 \quad 8.41 \quad 8.27 \quad 7.16\,.$$

(a) Calcolare un intervallo di confidenza bilatero di livello 95% per la media della variabile aleatoria X, specificando le ipotesi statistiche che si stanno usando. *[Ipotesi: $X \sim N(\mu, \sigma^2)$; $IC_\mu(95\%) = (7.5070\,,\, 8.5350)$]*

(b) Specificando anche qui le ipotesi statistiche che si stanno usando, calcolare un intervallo di confidenza per la varianza della variabile aleatoria X al livello 95% e del tipo seguente:

(i) bilatero; *[Stessa ipotesi del punto (a); $IC_{\sigma^2}(95\%) = (0.2442\,,\, 1.7205)$]*

(ii) unilatero del tipo $(0, c)$; *[Stessa ipotesi del punto (a); $IC_{\sigma^2}(95\%) = (0\,,\, 1.3973)$]*

(iii) unilatero del tipo $(c, +\infty)$. *[Stessa ipotesi del punto (a); $IC_{\sigma^2}(95\%) = (0.2746\,,\, +\infty)$]*

Esercizio 18. Una ditta produce termostati per scaldabagni; i termostati sono costruiti per aprire un circuito a una temperatura t_0 incognita. Un addetto al controllo di qualità esamina uno di questi termostati nel modo seguente: immerge il termostato in un bagno termico a temperatura controllata da un termometro di precisione e fa aumentare la temperatura fino a quando il termostato apre il circuito, poi annota la temperatura alla quale il termostato ha aperto il circuito. Ripetendo questa procedura per 250 volte, l'addetto ottiene i seguenti valori per la media e la varianza campionarie delle 250 misure (espressi in °C):

$$\bar{x}_{250} = 53.375 \qquad\qquad s^2_{250} = 6.908\,.$$

(a) Calcolare un intervallo di confidenza bilatero di livello 95% per la media della variabile aleatoria $X =$ temperatura a cui il termostato apre il circuito, specificando le ipotesi statistiche sulla densità di X che si stanno assumendo. *[Non è necessario fare alcuna ipotesi sulla densità di X. $IC_{\mathbb{E}[X]}(95\%) = (53.049\,,\, 53.701)$]*

(b) Calcolare un intervallo di confidenza bilatero, di livello 95% per la varianza della variabile $X =$ temperatura a cui il termostato apre il circuito, specificando le ipotesi statistiche sulla densità di X che si stanno assumendo. *[Ipotesi: $X \sim N(\mu, \sigma^2)$; $IC_{\mathrm{Var}(X_i)}(95\%) = (5.848\,,\, 8.321)$]*

Esercizio 19. Una ditta produce cinghie di trasmissione per automobili. È noto che la durata di tali cinghie, calcolata in *migliaia* di chilometri percorsi, ha distribuzione normale con media 50 e deviazione standard 5. Denotiamo con X la durata di una cinghia. Successivamente la ditta progetta un congegno che – applicato al sistema di trasmissione – consente di allungare la 'vita' di una cinghia di una quantità $Y \sim N(\mu, \sigma^2)$, con parametri μ e σ^2 entrambi incogniti. In altre parole, la durata delle cinghie in presenza del congegno è $Z = X + Y$ e assumiamo che X e Y siano indipendenti. Un test su 30 automobili provviste del nuovo congegno ha permesso di raccogliere dati relativi ad un campione $Z_1, \ldots, Z_{30}$ che ha fornito $\bar{z}_{30} = 57$, $s_{30} = 6.5$, dove $\bar{Z}_{30}$ indica la media campionaria e S^2_{30} è la varianza campionaria di $Z_1, \ldots, Z_{30}$.

(a) Qual è la distribuzione di Z? *[*$Z \sim N(50 + \mu, 25 + \sigma^2)$*]*

(b) Sulla base del campione $Z_1, \dots, Z_{30}$, fornite due stimatori non distorti per μ e σ^2, rispettivamente. Calcolate inoltre le stime corrispondenti. *[*$\hat{\mu} = \bar{Z} - 50 = 7$; $\hat{\sigma}^2 = S_Z^2 - 25 = 17.25$*]*

(c) Determinare un intervallo di confidenza per μ di livello 0.95. *[*$IC_\mu(95\%) = (4.573, 9.427)$*]*

(d) Determinare un intervallo di confidenza bilatero per la deviazione standard σ di livello 0.95. *[*$IC_\sigma(95\%) = (1.341, 7.166)$*]*

Soluzione 1.

Introduciamo la v.a.

$$X_i = \text{resa dell'impianto nel giorno } i\text{-esimo (in \%)}$$

Allora possiamo supporre che le 5 v.a. $X_1, \ldots, X_5$ costituiscano un campione aleatorio (= siano i.i.d.). Il testo del problema ci dice inoltre che

$$X_i \sim N(\mu, 2) \qquad \forall i = 1, \ldots, 5$$

con μ parametro incognito.

(a) Nel formulario, troviamo che un $IC_\mu(\gamma)$ bilatero per un campione gaussiano a varianza nota è

$$\mu \in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \, , \, \bar{x}_n + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Nel nostro caso,

$$\gamma = 0.95 \qquad z_{\frac{1+\gamma}{2}} = z_{0.975} = 1.96 \qquad \sigma = \sqrt{2} \qquad n = 5 \, ,$$

e dobbiamo calcolare la realizzazione della media campionaria sul risultato delle 5 misure:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{5}(91.60 + 88.75 + 90.80 + 89.95 + 91.30) = 90.48 \, .$$

In conclusione, l'intervallo è

$$\mu \in \left(90.48 - 1.96\sqrt{\frac{2}{5}} \, , \, 90.48 + 1.96\sqrt{\frac{2}{5}} \right) = (89.240 \, , \, 91.720) \, .$$

L'ampiezza dell'intervallo è

$$2 \, z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} = 91.720 - 89.240 = 2.480 \, .$$

(b) Al livello di confidenza del 95%, l'errore massimo che si commette stimando μ con $\bar{x}_5$ è la metà dell'ampiezza precedente. Infatti, l'errore massimo è $|\mu - \bar{x}_5|$. Dall'$IC_\mu(\gamma)$ del punto (a) troviamo

$$\begin{aligned}
\mu \in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \, , \, \bar{x}_n + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad &\Rightarrow \quad \mu - \bar{x}_n \in \left(-z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \, , \, +z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right) \\
&\Rightarrow \quad |\mu - \bar{x}_n| \in \left(0 \, , \, z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right) .
\end{aligned}$$

Perciò, con confidenza $\gamma = 95\%$, l'errore massimo è

$$z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} = \frac{2.480}{2} = 1.240 \, .$$

(c) Il capo reparto vuole che, mantenendo inalterata la confidenza $\gamma = 95\%$ e la deviazione standard $\sigma = \sqrt{2}$, si abbia

$$2 \, z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \leq 2.2 \, ,$$

cioè

$$2\,z_{0.975}\sqrt{\frac{2}{n}} \leq 2.2 \quad \Rightarrow \quad n \geq \left(\frac{2\,z_{0.975}\sqrt{2}}{2.2}\right)^2 = \left(\frac{2\cdot 1.96\cdot\sqrt{2}}{2.2}\right)^2 = 6.350\,.$$

Poiché n deve essere intero, ne ricaviamo

$$n \geq 7\,,$$

e pertanto servono almeno altre $7-5=2$ misure.

Soluzione 2.

Abbiamo un campione aleatorio $X_1,\ldots,X_{100}$, con

$$X_i = \text{peso dell'}i\text{-esimo sacco di farina} \sim N(\mu,\,\sigma^2)\,,$$

dove μ è incognita e $\sigma = 0.5$ è nota.

(a) Un $IC_\mu(\gamma)$ bilatero per un campione gaussiano a varianza nota è

$$\mu \in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}}\frac{\sigma}{\sqrt{n}}\,,\,\bar{x}_n + z_{\frac{1+\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Nel nostro caso,

$$\gamma = 0.90 \qquad z_{\frac{1+\gamma}{2}} = z_{0.95} = 1.645 \qquad \sigma = 0.5 \qquad n = 100 \qquad \bar{x}_{100} = 15.3\,,$$

e quindi

$$\mu \in \left(15.3 - 1.645\frac{0.5}{\sqrt{100}}\,,\,15.3 + 1.645\frac{0.5}{\sqrt{100}}\right) = (15.3 \pm 0.08225)$$

(modo alternativo per scrivere un intervallo simmetrico intorno a $\bar{x}_n = 15.3$).
L'errore massimo $|\mu - \bar{x}_n|$ è la semiampiezza dell'intervallo precedente, cioè

$$z_{\frac{1+\gamma}{2}}\frac{\sigma}{\sqrt{n}} = 1.645\frac{0.5}{\sqrt{100}} = 0.08225\,.$$

(b) Ora vogliamo scegliere n in modo che l'ampiezza dell'$IC_\mu(90\%)$ bilatero precedente sia minore di 0.1. L'ampiezza è $2\,z_{\frac{1+\gamma}{2}}\frac{\sigma}{\sqrt{n}}$, quindi deve essere

$$2\,z_{\frac{1+\gamma}{2}}\frac{\sigma}{\sqrt{n}} = 2\cdot 1.645\cdot\frac{0.5}{\sqrt{n}} < 0.1 \quad \Rightarrow \quad n > \left(\frac{2\cdot 1.645\cdot 0.5}{0.1}\right)^2 = 270.603\,.$$

Poiché n deve essere intero, occorrono $n \geq 271$ misure.

(c) Un $IC_\mu(\gamma)$ unilatero sinistro per un campione gaussiano a varianza nota è

$$\mu \in \left(\bar{x}_n - z_\gamma\frac{\sigma}{\sqrt{n}}\,,\,+\infty\right)$$

In altre parole, il Signor Sabbioso può essere confidente al livello γ che il parametro μ dei suoi sacchi di farina sia maggiore di $\bar{x}_n - z_\gamma\frac{\sigma}{\sqrt{n}}$. Se vogliamo trovare a quale livello γ egli può essere confidente che $\mu > 15.2$, dobbiamo quindi risolvere l'equazione

$$\bar{x}_n - z_\gamma\frac{\sigma}{\sqrt{n}} \equiv 15.2$$

rispetto a γ. Con $\bar{x}_n = 15.3$, $\sigma = 0.5$ e $n = 100$, l'equazione precedente diventa

$$15.3 - z_\gamma \frac{0.5}{\sqrt{100}} = 15.2 \qquad \Rightarrow \qquad z_\gamma = \frac{(15.3 - 15.2)\sqrt{100}}{0.5} = 2 \qquad \Rightarrow \qquad \gamma = \Phi(2) = 0.97725\,.$$

In conclusione, il Signor Sabbioso può essere confidente al livello 97.725% che $\mu > 15.2$.

(d) L'ipotesi di gaussianità non è necessaria, in quanto, essendo il campione numeroso, tutti gli intervalli precedenti sono $IC_{\mathbb{E}[X_i]}(\gamma)$ approssimati qualunque sia la densità delle X_i (importa solo che, essendo $n = 100$ grande, si ha $\bar{X}_{100} \approx N\left(\mathbb{E}[X_i]\,,\,\frac{0.5^2}{100}\right)$ indipendentemente dalla densità delle X_i).

Soluzione 6.

(a) Per verificare che T sia uno stimatore corretto (= non distorto) del parametro μ, dobbiamo controllare che $\mathbb{E}[T]$ coincida con μ. Quindi, calcoliamo

$$\begin{aligned}\mathbb{E}[T] = \mathbb{E}\left[\frac{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}{n+m}\right] &\underset{\substack{\text{linearità}\\ \text{di } \mathbb{E}}}{=} \frac{1}{n+m}\left(\sum_{i=1}^n \mathbb{E}[X_i] + \sum_{j=1}^m \mathbb{E}[Y_j]\right) \\ &\underset{\substack{X_i \sim N(\mu,1)\\ Y_j \sim N(\mu,3)}}{=} \frac{1}{n+m}\left(\sum_{i=1}^n \mu + \sum_{j=1}^m \mu\right) = \frac{1}{n+m}(n\mu + m\mu) = \mu\,.\end{aligned}$$

Quindi, T è effettivamente corretto.

(b) T è normale perchè combinazione lineare di normali indipendenti. In altre parole, $T \sim N(\mu_T, \sigma_T^2)$, dove ci restano solo da calcolare i parametri $\mu_T = \mathbb{E}[T]$ e $\sigma_T^2 = \mathrm{Var}(T)$. Per quanto riguarda μ_T, sappiamo dal punto (a) che

$$\mu_T = \mu\,. \tag{*}$$

Per quanto riguarda invece σ_T^2,

$$\begin{aligned}\sigma_T^2 = \mathrm{Var}(T) = \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}{n+m}\right) &\underset{\substack{\text{quadraticità}\\ \text{di Var}}}{=} \left(\frac{1}{n+m}\right)^2 \mathrm{Var}\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right) \\ &\underset{\substack{X_i,\, Y_j \text{ tutte}\\ \text{indipendenti}}}{=} \left(\frac{1}{n+m}\right)^2 \left(\sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{j=1}^m \mathrm{Var}(Y_j)\right) \\ &\underset{\substack{X_i \sim N(\mu,1)\\ Y_j \sim N(\mu,3)}}{=} \left(\frac{1}{n+m}\right)^2 \left(\sum_{i=1}^n 1 + \sum_{j=1}^m 3\right) = \left(\frac{1}{n+m}\right)^2 (n+3m) \\ &= \frac{n+3m}{(n+m)^2}\,. \end{aligned} \tag{**}$$

In conclusione,

$$T \sim N\left(\mu\,,\, \frac{n+3m}{(n+m)^2}\right)\,.$$

(c) Il livello di confidenza dell'intervallo assegnato è per definizione la probabilità

$$\mathbb{P}\left(T - z_\alpha\sqrt{\frac{n+3m}{(n+m)^2}} < \mu < T + z_\alpha\sqrt{\frac{n+3m}{(n+m)^2}}\right) \underset{\substack{\text{per (*)} \\ \text{e (**)}}}{=} \mathbb{P}\left(T - z_\alpha\sqrt{\sigma_T} < \mu_T < T + z_\alpha\sqrt{\sigma_T}\right)$$

$$= \mathbb{P}\left(-z_\alpha\sqrt{\sigma_T} < \mu_T - T < z_\alpha\sqrt{\sigma_T}\right) = \mathbb{P}\Bigg(z_\alpha > \underbrace{\frac{T-\mu_T}{\sqrt{\sigma_T}}}_{\sim N(0,1)} > -z_\alpha\Bigg)$$

$$= \Phi(z_\alpha) - \Phi(-z_\alpha) = \alpha - [1-\Phi(z_\alpha)] = \alpha - (1-\alpha)$$

$$= 2\alpha - 1\,.$$

Soluzione 8.

(a) *Prima* di misurare effettivamente la concentrazione dell'enzima nel sangue dei 20 individui, abbiamo a che fare con un campione aleatorio $X_1, \ldots, X_{20}$, in cui

$$X_i = \text{concentrazione dell'enzima nel sangue dell'}i\text{-esimo individuo}\,.$$

Il punto (a) ci dice che possiamo supporre

$$X_i \sim N(\mu, \sigma^2)\,,$$

e ci chiede di determinare un $IC_\mu(95\%)$ senza darci né μ (naturalmente!) né σ^2.
Dopo aver misurato la concentrazione dell'enzima nei 20 pazienti, abbiamo trovato le realizzazioni

$$\bar{x}_{20} = 1.23 \qquad\qquad s_{20}^2 = 0.4\,.$$

Dobbiamo quindi dare un intervallo di confidenza per la media di una popolazione gaussiana a varianza incognita; sul formulario, troviamo che un tale intervallo al livello γ è

$$\mu \in \left(\bar{x}_n - t_{\frac{1+\gamma}{2}}(n-1)\,\frac{s_n}{\sqrt{n}}\,,\ \bar{x}_n + t_{\frac{1+\gamma}{2}}(n-1)\,\frac{s_n}{\sqrt{n}}\right)$$

con

$$n = 20 \qquad \gamma = 0.95 \qquad t_{\frac{1+\gamma}{2}}(n-1) = t_{\frac{1+0.95}{2}}(20-1) = t_{0.975}(19) = 2.0930$$

e $\bar{x}_n$ e s_n^2 come sopra. Quindi,

$$\mu \in \left(1.23 - 2.0930\cdot\frac{\sqrt{0.4}}{\sqrt{20}}\,,\ 1.23 + 2.0930\cdot\frac{\sqrt{0.4}}{\sqrt{20}}\right) = (0.934\,,\ 1.526)\,.$$

(b) Ora sappiamo in più rispetto al punto (a) che il vero valore della varianza è

$$\sigma^2 = 0.4\,.$$

Di conseguenza, possiamo usare un $IC_\mu(95\%)$ per una popolazione gaussiana a varianza nota, cioè (sempre dal formulario)

$$\mu \in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}}\,\frac{\sigma}{\sqrt{n}}\,,\ \bar{x}_n + z_{\frac{1+\gamma}{2}}\,\frac{\sigma}{\sqrt{n}}\right) \qquad (*)$$

$$= \left(1.23 - 1.96\cdot\frac{\sqrt{0.4}}{\sqrt{20}}\,,\ 1.23 + 1.96\cdot\frac{\sqrt{0.4}}{\sqrt{20}}\right) = (0.953\,,\ 1.507)\,,$$

dove questa volta abbiamo usato i quantili della normale standard

$$z_{\frac{1+\gamma}{2}} = z_{\frac{1+0.95}{2}} = z_{0.975} = 1.96\,.$$

Da notare che in questo caso il valore della varianza campionaria s_{20}^2 non ci interessa più. Sia che dalle misure si ottenga $s_{20}^2 = 0.4$ come al punto (a), sia che si ottenga un altro valore (cosa in realtà molto più probabile!), l'intervallo (*) si calcola sempre usando la varianza nota a priori $\sigma^2 = 0.4$.

(c) È più ampio l'intervallo a varianza incognita del punto (a). Il risultato è ragionevole in quanto, se varianza e varianza campionaria sono uguali come in questo caso, abbiamo più incertezza nel punto (a), dove la varianza è incognita e va quindi ricavata dai dati, rispetto al punto (b), dove invece la varianza è nota a priori. Confrontando le formule esplicite dei due intervalli, ciò si traduce nel fatto che per i quantili dello stesso ordine della t di Student e della normale standard abbiamo

$$t_{0.975}(19) = 2.0930 > z_{0.975} = 1.96\,.$$

Soluzione 14.

Abbiamo un campione aleatorio $X_1, \ldots, X_n$, in cui

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esimo intervistato è favorevole} \\ 0 & \text{se l'}i\text{-esimo intervistato è contrario}\,. \end{cases}$$

Chiaramente, $X_i \sim B(1,p)$, in cui il parametro incognito p è la probabilità che un abitante a caso di Pietraforata sia favorevole alle domeniche ecologiche. Il nostro obiettivo è stimare il parametro incognito p.

(a) Dal formulario, sappiamo che un $IC_p(\gamma)$ bilatero approssimato è

$$p \in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\ ,\ \bar{x}_n + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \right) \qquad (*)$$

L'ampiezza di tale intervallo è

$$2\, z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\,.$$

Poiché la frequenza empirica $\bar{x}_n$ è sempre compresa nell'intervallo $[0,1]$, abbiamo di conseguenza $\bar{x}_n(1-\bar{x}_n) \in \left[0\,,\,\frac{1}{4}\right]$, e quindi la maggiorazione

$$2\, z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \leq 2\, z_{\frac{1+\gamma}{2}} \sqrt{\frac{\frac{1}{4}}{n}} = \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\,.$$

Se vogliamo essere sicuri che al livello di confidenza $\gamma = 95\%$ l'ampiezza dell'$IC_p(\gamma)$ sia inferiore a 0.05, dobbiamo quindi imporre

$$\frac{1}{\sqrt{n}} z_{\frac{1+0.95}{2}} < 0.05\,.$$

Svolgendo i calcoli, la disequazione precedente diventa

$$\frac{1}{\sqrt{n}} z_{0.975} < 0.05 \quad \Rightarrow \quad n > \left(\frac{z_{0.975}}{0.05}\right)^2 = \left(\frac{1.96}{0.05}\right)^2 = 1536.64\,.$$

Serve quindi un campione di almeno $n_0 = 1537$ intervistati.

(b) Uno stimatore non distorto e consistente del parametro p è la frequenza campionaria $\bar{X}_{n_0}$. La sua realizzazione coi risultati nella nostra intervista è

$$\bar{x}_{n_0} = \frac{882}{1537} = 0.5738\,,$$

e questa è la nostra stima di p.
Se invece voglia dare un $IC_p(95\%)$ bilatero, dobbiamo calcolare l'intervallo (*) con

$$\gamma = 0.95 \qquad z_{\frac{1+\gamma}{2}} = z_{0.975} = 1.96 \qquad n = 1537 \qquad \bar{x}_n = 0.5738\,.$$

Troviamo così

$$p \in \left(0.5738 \pm 1.96\sqrt{\frac{0.5738\cdot(1-0.5738)}{1537}}\right) = (0.5738 \pm 0.0247) = (0.5491\,,\,0.5986)\,.$$

(c) Abbiamo l'uguaglianza di eventi

$$\text{“l'errore della nostra stima puntuale di } p \text{ è inferiore a } 0.05\text{”} = \text{“ } \left|\bar{X}_{n_0} - p\right| < 0.05 \text{ ”}\,.$$

Quindi, si richiede di calcolare la probabilità

$$\begin{aligned}
&\mathbb{P}\left(\left|\bar{X}_{n_0} - p\right| < 0.05\right) = \mathbb{P}\left(-0.05 < \bar{X}_{n_0} - p < +0.05\right) \\
&\quad = \mathbb{P}\left(\frac{-0.05}{\sqrt{\frac{p(1-p)}{n_0}}} < \underbrace{\frac{\bar{X}_{n_0} - p}{\sqrt{\frac{p(1-p)}{n_0}}}}_{\approx N(0,1)} < \frac{+0.05}{\sqrt{\frac{p(1-p)}{n_0}}}\right) \qquad \text{standardizzazione di } \bar{X}_{n_0} \approx N\left(p\,,\,\frac{p(1-p)}{n_0}\right) \\
&\quad = \Phi\left(\frac{0.05}{\sqrt{\frac{p(1-p)}{n_0}}}\right) - \Phi\left(-\frac{0.05}{\sqrt{\frac{p(1-p)}{n_0}}}\right) = 2\Phi\left(\frac{0.05}{\sqrt{p(1-p)}}\sqrt{n_0}\right) - 1 \\
&\quad = 2\Phi\left(\frac{0.05}{\sqrt{0.65\cdot(1-0.65)}}\sqrt{1537}\right) - 1 = 2\Phi(4.11) - 1 = 2\cdot 0.99998 - 1 \\
&\quad = 99.996\%\,.
\end{aligned}$$

Soluzione 17.

Le due ipotesi statistiche che faremo in tutti i punti seguenti sono:

- $X \sim N(\mu, \sigma^2)$, in cui $\mu = \mathbb{E}\,[X]$ e $\sigma^2 = \mathrm{Var}\,(X)$ sono i due parametri incogniti da stimare;
- le 10 misure costituiscono un campione gaussiano $X_1, \ldots, X_{10}$, con $X_i \sim N(\mu, \sigma^2)$, cioè con la stessa densità di X.

Questo perché:

- Il campione è troppo poco numeroso; quindi non si può dire che $T := \frac{\bar{X}_n - \mu}{S_n}\sqrt{n} \approx N(0,1)$. Se non ipotizzassimo $X_i \sim N(\mu, \sigma^2)$, non potremmo sapere la densità della quantità pivotale T, e non potremmo quindi determinare un IC per la media μ. Supponendo invece $X_i \sim N(\mu, \sigma^2)$, abbiamo $T \sim t(n-1)$, e perciò possiamo costruirci un IC per μ usando i quantili (tabulati) della densità $t(n-1)$.

- Se il campione non è gaussiano, non sappiamo qual è la densità della quantità pivotale $X^2 := \frac{(n-1)S_n^2}{\sigma^2}$, e quindi non possiamo neppure determinare un IC per la varianza σ^2. Supponendo invece $X_i \sim N(\mu, \sigma^2)$, sappiamo dalla teoria che $X^2 \sim \chi^2(n-1)$, e possiamo costruirci un IC per σ^2 usando i quantili (anch'essi tabulati) di $\chi^2(n-1)$.

Calcoliamo anzitutto la media e la varianza campionarie delle 10 misure:

$$\bar{x}_{10} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(7.94 + 8.18 + \ldots + 7.16) = 8.0210$$
$$s_{10}^2 = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}_n^2\right] = \frac{1}{10-1}\left(7.94^2 + 8.18^2 + \ldots + 7.16^2 - 10 \cdot 8.0210^2\right) = 0.5162\,.$$

(a) Si tratta di un $IC_\mu(95\%)$ bilatero per la media μ di una popolazione gaussiana a media e varianza incognite. Esso è dato da

$$\begin{aligned}\mu \in &\left(\bar{x}_n - t_{\frac{1+\gamma}{2}}(n-1)\frac{s_n}{\sqrt{n}}\,,\ \bar{x}_n + t_{\frac{1+\gamma}{2}}(n-1)\frac{s_n}{\sqrt{n}}\right)\\ &= \left(8.0210 - 2.2622\sqrt{\frac{0.5162}{10}}\,,\ 8.0210 + 2.2622\sqrt{\frac{0.5162}{10}}\right)\\ &= (7.5070\,,\ 8.5350)\,.\end{aligned}$$

dove

$$t_{\frac{1+0.95}{2}}(10-1) = t_{0.975}(9) = 2.2622\,.$$

(b.i) Si tratta di un $IC_{\sigma^2}(95\%)$ bilatero per la varianza σ^2 di una popolazione gaussiana a media e varianza incognite. Dal formulario,

$$\sigma^2 \in \left(\frac{(n-1)s_n^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}\,,\ \frac{(n-1)s_n^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}\right) = \left(\frac{(10-1)\cdot 0.5162}{19.0228}\,,\ \frac{(10-1)\cdot 0.5162}{2.7004}\right) = (0.2442\,,\ 1.7205)$$

dove

$$\chi^2_{\frac{1+0.95}{2}}(10-1) = \chi^2_{0.975}(9) = 19.0228 \qquad \chi^2_{\frac{1-0.95}{2}}(10-1) = \chi^2_{0.025}(9) = 2.7004\,.$$

(b.ii) Sempre dal formulario,

$$\sigma^2 \in \left(0\,,\ \frac{(n-1)s_n^2}{\chi^2_{1-\gamma}(n-1)}\right) = \left(0\,,\ \frac{(10-1)\cdot 0.5162}{3.3251}\right) = (0\,,\ 1.3973)$$

dove

$$\chi^2_{1-0.95}(10-1) = \chi^2_{0.05}(9) = 3.3251\,.$$

(b.iii) Infine,

$$\sigma^2 \in \left(\frac{(n-1)s_n^2}{\chi^2_\gamma(n-1)}\,,\ +\infty\right) = \left(\frac{(10-1)\cdot 0.5162}{16.9190}\,,\ +\infty\right) = (0.2746\,,\ +\infty)$$

dove

$$\chi^2_{0.95}(10-1) = 16.9190\,.$$

Soluzione 18.

Per il momento, non facciamo nessuna ipotesi sulla densità della v.a.

$$X = \text{temperatura a cui il termostato apre il circuito}\,.$$

Sappiamo solo che abbiamo un campione di 250 misure $X_1, \ldots, X_{250}$, in cui tutte le X_i hanno la stessa densità di X, e dunque $\mathbb{E}[X_i] = \mathbb{E}[X] =: \mu$ e $\mathrm{Var}(X_i) = \mathrm{Var}(X) =: \sigma^2$ per ogni i.

(a) Per calcolare un $IC_\mu(95\%)$, non occorre fare nessuna ipotesi statistica sulla densità di X (e delle X_i di conseguenza). Infatti, possiamo usare un $IC_\mu(95\%)$ per una popolazione numerosa qualsiasi a media e varianza incognite. Quello bilatero è dato da

$$\begin{aligned}\mu &\in \left(\bar{x}_n - z_{\frac{1+\gamma}{2}} \frac{s_n}{\sqrt{n}}\,,\, \bar{x}_n + z_{\frac{1+\gamma}{2}} \frac{s_n}{\sqrt{n}}\right)\\ &= \left(53.375 - 1.96\sqrt{\frac{6.908}{250}}\,,\, 53.375 + 1.96\sqrt{\frac{6.908}{250}}\right)\\ &= (53.049\,,\, 53.701)\,,\end{aligned}$$

dove

$$z_{\frac{1+0.95}{2}} = z_{0.975} = 1.96\,.$$

(b) Per calcolare un $IC_{\sigma^2}(95\%)$, bisogna invece supporre che $X, X_i \sim N(\mu, \sigma^2)$. Infatti, per densità diverse non conosciamo nessun intervallo di confidenza per la varianza. Al contrario, con questa ipotesi, sul formulario troviamo

$$\sigma^2 \in \left(\frac{(n-1)s_n^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}\,,\, \frac{(n-1)s_n^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}\right) = \left(\frac{(250-1)\cdot 6.908}{294.116}\,,\, \frac{(250-1)\cdot 6.908}{206.726}\right) = (5.848\,,\, 8.321)\,,$$

dove per calcolare i quantili $\chi^2_\alpha(k)$ con k grande abbiamo usato la formula approssimata

$$\chi^2_\alpha(k) \simeq \frac{(z_\alpha + \sqrt{2k-1})^2}{2}$$

che dà i valori

$$\begin{aligned}\chi^2_{\frac{1+0.95}{2}}(250-1) &= \chi^2_{0.975}(249) \simeq \frac{(z_{0.975} + \sqrt{2\cdot 249 - 1})^2}{2} = \frac{(1.96 + \sqrt{497})^2}{2} = 294.116\\ \chi^2_{\frac{1-0.95}{2}}(250-1) &= \chi^2_{0.025}(249) \simeq \frac{(z_{0.025} + \sqrt{2\cdot 249 - 1})^2}{2} = \frac{(-1.96 + \sqrt{497})^2}{2} = 206.726\,.\end{aligned}$$

Soluzione 19.

(a) Dai dati, sappiamo che

$$X \sim N(50, 25)\,, \qquad Y \sim N(\mu, \sigma^2)\,, \qquad X \text{ e } Y \text{ sono indipendenti}\,, \qquad Z = X + Y\,.$$

Per la proprietà di riproducibilità della densità gaussiana, segue che

$$\begin{aligned}Z \sim N\left(\mathbb{E}[X+Y]\,,\, \mathrm{Var}(X+Y)\right) &= N\left(\mathbb{E}[X] + \mathbb{E}[Y]\,,\, \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\cancel{\mathrm{Cov}(X,Y)}\right)\\ &= N(50 + \mu,\, 25 + \sigma^2)\,.\end{aligned}$$

(b) Sappiamo che la media campionaria $\bar{Z}$ del campione $Z_1, \ldots, Z_{30}$ è uno stimatore non distorto del valore atteso $\mathbb{E}[Z_i] = 50 + \mu$, mentre la varianza campionaria S_Z^2 è uno stimatore non distorto della varianza vera $\mathrm{Var}(Z_i) = 25 + \sigma^2$. Per la linearità del valore atteso, $\bar{Z} - 50$ è dunque uno stimatore non distorto del parametro μ, mentre $S_Z^2 - 25$ è uno stimatore non distorto del parametro σ^2. Infatti,

$$\mathrm{bias}\left(\bar{Z} - 50;\, \mu\right) = \mathbb{E}\left[\bar{Z} - 50\right] - \mu = \mathbb{E}\left[\bar{Z}\right] - 50 - \mu = 50 + \mu - 50 - \mu = 0\,,$$
$$\mathrm{bias}\left(S_Z^2 - 25;\, \sigma^2\right) = \mathbb{E}\left[S_Z^2 - 25\right] - \sigma^2 = \mathbb{E}\left[S_Z^2\right] - 25 - \sigma^2 = 25 + \sigma^2 - 25 - \sigma^2 = 0\,.$$

(c) A partire dal campione gaussiano a varianza incognita $Z_1, \ldots, Z_{30}$, possiamo costruire il seguente $IC_{\mathbb{E}[Z]}(95\%)$:

$$\mathbb{E}[Z] \in \left(\bar{z} - t_{\frac{1+\gamma}{2}}(n-1)\,\frac{s_Z}{\sqrt{n}}\,,\, \bar{z} + t_{\frac{1+\gamma}{2}}(n-1)\,\frac{s_Z}{\sqrt{n}}\right) = \left(57 - 2.0452 \cdot \frac{6.5}{\sqrt{30}}\,,\, 57 + 2.0452 \cdot \frac{6.5}{\sqrt{30}}\right)$$
$$= (54.573\,,\, 59.427)\ .$$

dove

$$t_{\frac{1+\gamma}{2}}(n-1) = t_{\frac{1+0.95}{2}}(30-1) = t_{0.975}(29) = 2.0452\,.$$

D'altra parte, a noi interessa trovare un $IC_\mu(95\%)$. Dunque, poiché i parametri $\mathbb{E}[Z]$ e μ sono legati dalla relazione $\mathbb{E}[Z] = 50 + \mu$, o, equivalentemente, $\mu = \mathbb{E}[Z] - 50$, abbiamo

$$\mu = \mathbb{E}[Z] - 50 \in (54.573\,,\, 59.427) - 50 = (54.573 - 50\,,\, 59.427 - 50) = (4.573\,,\, 9.427)\ .$$

(d) Anche qui, a partire dal campione gaussiano $Z_1, \ldots, Z_{30}$, possiamo costruire un $IC_{\mathrm{Var}(Z)}(95\%)$ usando relazioni del formulario:

$$\mathrm{Var}(Z) \in \left(\frac{(n-1)\,s_Z^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}\,,\, \frac{(n-1)\,s_Z^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}\right) = \left(\frac{(30-1)\cdot 6.5^2}{45.7223}\,,\, \frac{(30-1)\cdot 6.5^2}{16.0471}\right)$$
$$= (26.798\,,\, 76.353)\ .$$

dove

$$\chi^2_{\frac{1+\gamma}{2}}(n-1) = \chi^2_{\frac{1+0.95}{2}}(30-1) = \chi^2_{0.975}(29) = 45.7223\,,$$
$$\chi^2_{\frac{1-\gamma}{2}}(n-1) = \chi^2_{\frac{1-0.95}{2}}(30-1) = \chi^2_{0.025}(29) = 16.0471\,.$$

L'esercizio richiede però di determinare un $IC_\sigma(95\%)$, dunque usiamo la relazione che lega $\mathrm{Var}(Z)$ a σ:

$$\mathrm{Var}(Z) = 25 + \sigma^2 \quad \Leftrightarrow \quad \sigma = \sqrt{\mathrm{Var}(Z) - 25}$$

per trovare l'intervallo di confidenza per il parametro σ a partire da quello per il parametro $\mathrm{Var}(Z)$:

$$\sigma = \sqrt{\mathrm{Var}(Z) - 25} \in \sqrt{(26.798\,,\, 76.353) - 25} = \left(\sqrt{26.798 - 25}\,,\, \sqrt{76.353 - 25}\right)$$
$$= (1.341\,,\, 7.166)\ .$$