Marco Minozzo e Annamaria Guolo
Laurea in Economia del Commercio Internazionale
Laurea in Economia e Amministrazione delle Imprese
Università degli Studi di Verona (sede di Vicenza)

Statistica Inferenziale
Soluzioni 2. Intervalli di confidenza

Introduzione.
Siano $X_1, \ldots, X_n$ v.a. indipendenti e con identica distribuzione. Sia $f(x_1, \ldots, x_n; \theta)$ la densità o la distribuzione di probabilità di $X_1, \ldots, X_n$, con parametro incognito θ. Siano date due statistiche $L_1 = \ell_1(X_1, \ldots, X_n)$ e $L_2 = \ell_2(X_1, \ldots, X_n)$ tali che $L_1 \leq L_2$ e

$$P(L_1 < \theta < L_2) = 1 - \alpha,$$

dove $1 - \alpha$ è noto come il *coefficiente di confidenza* (non dipendente da θ). Allora, l'intervallo aleatorio L_1, L_2 è detto intervallo di confidenza per θ di livello $(1 - \alpha)\%$, il quale conterrà il valore vero (ignoto) di θ con probabilità pari a $(1 - \alpha)\%$. Una volta ottenuto un campione di osservazioni da $X_1, \ldots, X_n$, l'intervallo calcolato non sarà più aleatorio, ma sarà un intervallo numerico che conterrà o meno il valore vero (ignoto) di θ.

Esemplificazioni.
Sia X una v.a. normale con media μ e varianza σ^2. Sia dato un campione casuale di dimensione n di osservazioni da X. I seguenti casi sono di particolare interesse.

- *Intervallo di confidenza per μ con σ^2 noto.* Si fa riferimento alla statistica

$$Z = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{\sigma^2}} \sim N(0, 1).$$

  Considerando che

$$P\left(-z_{\alpha/2} < \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{\sigma^2}} < z_{\alpha/2}\right) = 1 - \alpha$$

  dove $z_{\alpha/2}$ è il quantile di ordine $\alpha/2$ di Z, si ha che l'intervallo di confidenza per μ al livello $(1 - \alpha)\%$ è dato da

$$\left(\overline{X}_n - z_{\alpha/2}\sqrt{\sigma^2/n}, \overline{X}_n + z_{\alpha/2}\sqrt{\sigma^2/n}\right).$$

- *Intervallo di confidenza per μ con σ^2 incognito.* Si fa riferimento alla statistica

$$T = \frac{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{\sigma^2}}}{\sqrt{S^2(n-1)/\sigma^2}} \sim t_{n-1}.$$

  Considerando che

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

  dove $t_{\alpha/2}$ è il quantile di ordine $\alpha/2$ di T, si ha che l'intervallo di confidenza per μ al livello $(1 - \alpha)\%$ è dato da

$$\left(\overline{X}_n - t_{\alpha/2}\sqrt{S^2/n}, \overline{X}_n + t_{\alpha/2}\sqrt{S^2/n}\right).$$

- *Intervallo di confidenza per σ^2 con μ incognito.* Si fa riferimento alla statistica

  $$T = \frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}.$$

  Considerando che

  $$P\left(\chi^2_{n-1,1-\alpha/2} < \frac{S^2(n-1)}{\sigma^2} < \chi^2_{n-1,\alpha/2}\right) = 1-\alpha,$$

  dove $\chi^2_{n-1,1-\alpha/2}$ è il quantile di una χ^2_{n-1} che lascia alla propria sinistra una probabilità pari a $\alpha/2$ e $\chi^2_{\alpha/2}$ è il quantile che lascia alla propria destra una probabilità pari a $\alpha/2$, si ha che l'intervallo di confidenza per σ^2 al livello $(1-\alpha)\%$ è dato da

  $$\left(\frac{S^2(n-1)}{\chi^2_{n-1,\alpha/2}}, \frac{S^2(n-1)}{\chi^2_{n-1,1-\alpha/2}}\right).$$

Inoltre consideriamo i seguenti risultati.

- *Intervallo di confidenza per la probabilità di successo in una popolazione dicotomica (grandi campioni).* Siano $X_1, \ldots, X_n$ v.a. indipendenti ed identicamente distribuite di tipo Bernoulliano con probabilità di successo p. Consideriamo lo stimatore di p fornito da $\sum_{i=1}^n X_i/n$ e la sua distribuzione per n grande ottenuta sulla base del teorema del limite centrale

  $$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \approx N(0,1).$$

  Un'estensione del teorema del limite centrale ci porta a

  $$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx N(0,1).$$

  Allora, un intervallo di confidenza per p al livello $(1-\alpha)\%$ è dato da

  $$\left(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\right),$$

  dove $z_{\alpha/2}$ è il quantile di ordine $\alpha/2$ di una $N(0,1)$.

- *Intervallo di confidenza per la media di una popolazione qualsiasi (grandi campioni).* Siano $X_1, \ldots, X_n$ v.a. indipendenti ed identicamente distribuite con valore atteso $E(X_i) = \mu$ e varianza $Var(X_i) = \sigma^2$, $i = 1, \ldots, n$. Per n sufficientemente grande, dall'applicazione del teorema del limite centrale si ha che

  $$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \approx N(0,1)$$

  e, per un'estensione del teorema del limite centrale,

  $$\frac{\overline{X}_n - \mu}{\sqrt{S^2/n}} \approx N(0,1).$$

  Allora un intervallo di confidenza approssimato al livello $(1-\alpha)\%$ per μ è dato da

  $$\overline{X}_n - z_{\alpha/2}\sqrt{S^2/n}; \overline{X}_n + z_{\alpha/2}\sqrt{S^2/n},$$

  dove $z_{\alpha/2}$ è il quantile di ordine $\alpha/2$ di una $N(0,1)$.

Esercizio A.

a) Il carattere in considerazione nella popolazione di riferimento non è distribuito normalmente, per cui l'intervallo di confidenza richiesto non si costruisce usando i quantili della distribuzione t di Student. Si può invece, sfruttando l'estensione del teorema del limite centrale, ricorrere al fatto che per grandi campioni la grandezza $(\bar{X} - \mu_x)/\sqrt{S^2/n}$ si distribuisce approssimativamente come una normale standardizzata. Di conseguenza, se il campione è grande, si può determinare l'intervallo di confidenza utilizzando i quantili di una normale standardizzata. Essendo $\alpha = 0,01$, si cerca il quantile $z_{0,005}$, ovvero quel valore tale che $P(Z \geq z_{0,005}) = 0,005$; esso è pari a 1,96. Quindi, l'intervallo cercato è dato da $[53,00 - 1,96 \cdot 20/\sqrt{50};\ 53,00 + 1,96 \cdot 20/\sqrt{50}]$, ovvero da $[47,46;\ 58,54]$.

b) L'ampiezza dell'intervallo precedentemente calcolato è pari a 11,08. L'intervallo di livello 99%, per una generica numerosità campionaria n, ha ampiezza pari a $2 \cdot 2,576 \cdot 20/\sqrt{n}$. Da qui, ponendo $2 \cdot 2,576 \cdot 20/\sqrt{n} = 11,08$, si ottiene $n = 87$ (arrotondando il risultato al fine di garantire n valore intero).

Esercizio B.

a) Sia $\bar{R}$ la v. a. che esprime la media del campione. Essendo l'alternativa $H_1 : \mu_R < 0,5$ composta da valori della media più piccoli che sotto H_0, si rifiuterà l'ipotesi nulla H_0 per piccoli valori di $\bar{R}$. Essendo $\alpha = 0,05$, si rifiuterà H_0 per valori di $\bar{R}$ inferiori a R_α, dove R_α è tale che $P(\bar{R} \leq R_\alpha; H_0) = 0,05$. Per trovare il valore R_α si consideri che le R_i sono distribuite normalmente e che se è vera H_0 queste hanno media pari a $0,5$. Quindi, potendo scrivere $P(\bar{R} \leq R_\alpha; H_0) = 0,05 = P\left(\frac{\bar{R}-0,5}{\frac{S}{\sqrt{n}}} \leq \frac{R_\alpha-0,5}{\frac{S}{\sqrt{n}}}; H_0\right)$, ed essendo $(\bar{R} - 0,5)/\sqrt{S^2/n}$ distribuita come una t di Student con g.d.l. pari a 29, il valore $(R_\alpha - 0,5)/\sqrt{S^2/n}$ deve essere pari a $-1,699$, ovvero $R_\alpha = -1,699\frac{S}{\sqrt{n}} + 0,5 = -1,113$. La zona di rifiuto di H_0 è pertanto costituita dai valori osservati di $\bar{R}$ inferiori a $-1,113$, ed essendo il valore osservato $\bar{R} = 0,3$ superiore a $-1,113$ si accetta H_0. Un procedimento alternativo consiste nel confrontare $(0,3 - 0,5)\sqrt{n}/(5,2)$ con $-1,699$ e rifiutare H_0 se tale valore è inferiore a $-1,699$, dal momento che se la disuguaglianza di sopra risulta verificata anche questa disuguaglianza risulta verificata, e viceversa.

b) Per trovare l'intervallo di confidenza richiesto si consideri che $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{29}$, per cui l'intervallo di confidenza al 95% è dato da $\left[\frac{29\times(5,2)^2}{45,722}; \frac{29\times(5,2)^2}{16,047}\right] = [17,150;\ 48,866]$.

c) Il "Value at Risk" T soddisfa alla relazione $P(R_i \leq T) = 0,01 = P(Z \leq \frac{T-0,5}{6,99})$, dove Z è una v. a. con distribuzione normale standardizzata. Essendo che il percentile della normale standardizzata che lascia alla sua sinistra un'area pari a 0,01 è pari a $-2,326$, si ricava che $T = -2,326 \cdot 6,99 + 0,5 = 15,759$, ovvero che la riserva patrimoniale a fronte di tale investimento deve essere pari a circa 15 milioni e 760 mila euro. Questa riserva fronteggia perdite che sono estremamente rare.

Esercizio C.
a) Per l'estensione del teorema del limite centrale, abbiamo che $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx N(0,1)$, per n grande. Quindi, l'intervallo di confidenza ha estremi $\hat{p} - 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ e $\hat{p} + 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, dove $n = 351$, ovvero, l'intervallo è dato da $[0,598;\ 0,698]$.
b) L'errore di stima è pari a $|\hat{p}-p|$. Si richiede di calcolare n tale per cui

$$P\left(|\hat{p}-p| < 0.02\right) = 0.90.$$

Considerando la distribuzione di $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ ottenuta tramite il teorema del limite centrale, n deve essere tale per cui

$$\Phi\left(\frac{0,02}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right) = \frac{1,90}{2}.$$

Da qui si ricava che $n = 1534$.