# Esercizi di Inferenza statistica

a cura di Giovanni M. Marchetti

14 febbraio 2019

*Questi esercizi hanno la struttura degli esercizi dell'esame scritto di Statistica, cioè aono di 3 tipi: vero/falso, scelta multipla, risposta aperta.*

**1.** Se un campione di dimensione $n$ viene selezionato casualmente da una popolazione normale, allora la distribuzione campionaria della media sarà:

A) normale solo per $n > 30$;

B) approssimativamente normale per qualsiasi valore di $n$;

C) normale per qualsiasi valore di $n$;

D) approssimativamente normale solo per $n > 30$.

**Soluzione**

La media di un campione è $\overline{X} = (X_1 + \cdots X_n)/n$ è una combinazione lineare di variabili normali e quindi è sempre esattamente normale. Soluzione C).

**2.** Secondo una regola generale, la distribuzione normale approssima bene la distribuzone della proporzione campionaria solo se:

A) la popolazione ha una distribuzione normale;

B) $nP(1 - P) > 9$;

C) la dimensione del campione $n$ è maggiore di 30;

D) la proporzione $P$ nella popolazione è prossima a 0.5.

**Soluzione**

A) È una sciocchezza. Si può calcolare una proporzione solo se la popolazione è fatta di 0 e 1 e non se è normale.

B) Corretta e suggerita dal libro.

C) Sensata ma la B) è meglio

D) No, l'approssimazione dipende essenzialemente da $n$.

**3.** Se da una popolazione che non ha distribuzione normale si estraggono tutti i possibili campioni di dimensione $n$ e per ciascuno di questi si calcola la media, cosa si può dire della sua distribuzione campionaria?

A) È approssimativamente normale a patto che $n$ sia abastanza grande.

B) È positivamente asimmetrica.

C) È negativamente asimmetrica.

D) Nessuna delle precedenti affermazioni è vera.

**Soluzione**

A) Corretta, per il teorema centrale del limite.

B) Può essere, ma non sempre.

C) Può essere, ma non sempre.

D) Falsa.

**4.** Se la deviazione standard della distribuzione campionaria della proporzione è pari a 0.02049 per campioni di dimensione 500, allora la proporzione nella popolazione deve essere pari a:

A) 0.5 o 0.5.

B) 0.6 o 0.4.

C) 0.3 o 0.7.

D) 0.2 o 0.8.

**Soluzione**

La distribuzione campionaria di una proporzione $\widehat{P}$ è Binomiale relativa cioè ha media $p$ e deviazione standard $\sigma_{\widehat{p}} = \sqrt{pq/n}$.

Se $\sigma_{\widehat{p}} = 0.02049$ e $n = 500$ avremo

$$p(1-p)/500 = (0.02049)^2$$

cioè

$$p(1-p) \approx 0.21$$

Dunque la risposta giusta è C) perché $0.3 \times 0.7 = 0.21$.

**5.** In un'indagine su studenti di scuola media superiore è stato rilevato che la spesa settimanale per svago di ogni studente è distribuita normalmente con media pari a Euro 52.30 e deviazione standard pari a Euro 18.23. Assumendo che questi valori siano rappresentativi di tutti gli studenti di scuola media superiore, qual è la probabilità che per un campione casuale di 25 studenti la spesa media settimanale degli studenti ecceda Euro 60 ?

A) 0.1628

B) 0.0174

C) 0.4826

D) 0.3372

**Soluzione**

Per ipotesi se $X$ è la spesa, $X \sim N(\mu = 52.3, \sigma = 18.23)$. In un campione casuale di dimensione $n = 25$ la spesa media è $\overline{X}$ e sappiamo che $\overline{X} \sim N(\mu, \sigma_{\overline{X}} = \sigma/\sqrt{n})$ cioè

$$\overline{X} \sim N(52.3, \sigma_{\overline{X}} = 18.23/5 = 3.646)$$

Perciò la probabilità che la spesa *media* ecceda 60 Euro è

$$P(\overline{X} > 60) = P(Z > (60 - 52.3)/3.646) = 1 - P(Z \leq 2.11) = 1 - 0.9826 = 0.0174.$$

Risposta B).

**6.** Se viene estratto un campione di dimensione 30, il valore di $c$ per la probabilità $P(T > c) = 0.01$, dove $T$ è la distribuzione t di Student, è

A) 2.462.

B) 2.750.

C) 2.756.

D) 2.247.

**Soluzione**

Il problema suppone che si voglia stimare la media di una popolazione normale facendo un intervallo di confidenza al livello 0.98. La distribuzione t di Student quindi deve essere scelta con $n - 1 = 29$ gradi di libertà. Sulle tavole dei quantili della t di Student si trova

$$c = 2.462.$$

Soluzione A).

**7.** Si abbia una popolazione dicotomica con una proporzione di successi pari a $p = 0.4$. Si estrae un campione casuale di dimensione 3 e si costruisce il seguente stimatore di $p$

$$T = \tfrac{1}{4}X_1 + \tfrac{1}{2}X_2 + \tfrac{1}{4}X_3.$$

Allora la varianza di T è A) 0.09 B) 0.06 C) 0.07 D) 0.08

**Soluzione**

La popolazione è dicotomica (cioè fatta di 1 e 0) e quindi è una variabile $X$ di Bernoulli con probabilità di successo $p$. La sua media è $p$ e la sua varianza è $\mathrm{var}(X) = pq$.

Si vuole stimare $p$ con un campione $(X_1, X_2, X_3)$ di $n = 3$ elementi. La varianza di $T$ è

$$\mathrm{var}(T) = \mathrm{var}(\frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3).$$

Siccome il campione è casuale, le variabili $X_1, X_2$ e $X_3$ sono indipendenti e quindi la varianza è la somma delle varianze:

$$\mathrm{var}(T) = \mathrm{var}(\frac{1}{4}X_1) + \mathrm{var}(\frac{1}{2}X_2) + \mathrm{var}(\frac{1}{4}X_3).$$

Quindi si applica le regola $\mathrm{var}(aX) = a^2\mathrm{var}(X)$ ottenendo

$$\mathrm{var}(T) = \frac{1}{16}\mathrm{var}(X_1) + \frac{1}{4}\mathrm{var}(X_2) + \frac{1}{16}\mathrm{var}(X_3).$$

A questo punto siccome il campione è casuale, le distribuzioni di $X_1, X_2$ e $X_3$ sono identiche alla distribuzione della popolazione $X$. Quindi hanno la stessa varianza di $X$ ossia $pq$. Pertanto

$$\mathrm{var}(T) = \frac{1}{16}pq + \frac{1}{4}pq + \frac{1}{16}pq = \frac{3}{8}pq.$$

NOTA: se lo stimatore fosse stato la proporzione campionaria

$$\widehat{P} = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$$

un calcolo analogo avrebbe portato a trovare la varianza

$$\mathrm{var}(\widehat{P}) = \frac{1}{3}pq$$

secondo la regola nota.

**8.** Quali dei seguenti fattori non influisce sul margine di errore di un intervallo di confidenza?

A) Il livello di confidenza.

B) La media campionaria.

C) L'ampiezza campionaria.

D) La varianza della popolazione.

**Soluzione**. Supponendo di voler fare un ingtervallo di confidenza per la media della popolazione, il margine di errore è

$$ME = z_{\alpha/2}\sigma/\sqrt{n}.$$

A) Falso, il livello influisce su $z_{\alpha/2}$

B) Vero perché non compare nella formula.

C) Falso, compare a denominatore sotto radice.

D) Falso, compare a numeratore sotto radice.

**9.** Si supponga di voler effettuare un test su $H_0 : \mu = 0.54$ contro $H_1 : \mu < 0.54$ basato su un campionedi $n = 25$ con deviazione standard stimata $s = 13.2$. Quale dovrebbe essere la statistica test?

A) $(\overline{X} - 0.54)/2.64$

B) $(\overline{X} - 0.54)/0.528$

C) $(\overline{X} - 0.54)/34.848$

D) $(\overline{X} - 0.54)/0.2789$

**Soluzione**

La statistica test è

$$T = \frac{\overline{X} - 0.54}{s/\sqrt{n}} = \frac{\overline{X} - 0.54}{2.64}.$$

Soluzione A).

**10.** L'azienda produttrice di un nuovo macchinario asserisce che questo produce almeno 29 unità al giorno in più del macchinario attualmente in uso in un'industria manufatturiera. Il manager decide di acquistare 15 nuovi macchinari e osserva che la produzione media giornaliera aumenta solo di 26 unità con una deviazione standard di 6.2.

Il valore osservato della statistica test è: A) 1.240. B) -1.240. C) -1.874. D) 1.874.

**Soluzione**. La statistica test è

$$T = \frac{\overline{X} - 29}{s/\sqrt{n}} = \frac{26 - 29}{6.2/\sqrt{15}} = -1.874$$

Risposta C).

**11.** La Commissione Daytona Beach Tourism vuole verificare se la media della spesa giornaliera media di uno studente del college durante le vacanze primaverili sia superiore a $70. Sulla base di studi precedenti, è noto che la deviazione standard della spesa giornaliera è $17.32. La Commissione intervista un campione di 35 studenti ottenendo una media campionaria di $72.43. Quale delle seguenti affermazioni è vera?

A) Si rifiuta l'ipotesi nulla con $\alpha = 0.10$.

B) Si rifiuta l'ipotesi nulla con $\alpha = 0.05$.

C) Si rifiuta l'ipotesi nulla con $\alpha = 0.01$.

D) Non è possibile rifiutare l'ipotesi nulla con $\alpha \leq 0.10$.

**Soluzione**. Ipotesi alternativa $H_1 : \mu > 70$. Nel testo dice superiore a 70 e quindi non uguale a 70. Per convenzione le affermazioni che contengono la relazione di uguaglanza si traducono in un'ipotesi $H_0$. Quindi $H_0 : \mu \leq 70$. Il valore 17.32 è la deviazione standard *nella popolazione* assunta nota perché deriva da studi precedenti. La statistica test è

$$Z = (\overline{X} - 70)/\sigma/\sqrt{n} = (72.43 - 70)/(17.32/\sqrt{(35)}) = 0.83.$$

Il p-value è $P(Z > 0.83) = 1 - P(Z \leq 0.83) = 0.2033$. Quindi

A) Falsa

B) Falsa

C) Falsa

D) Vera.

**12.** Un professore asserisce che il punteggio medio conseguito ad un recente esame è stato 83. Si assuma che la variabile punteggio conseguito si distribuisca normalmente. Tu chiedi ad alcuni in classe quale punteggio abbiano conseguito ed ottieni le seguenti risposte:

$$82, 77, 85, 76, 81, 91, 70, 82.$$

Supponi di voler verificare se l'affermazione del professore è corretta. Quale affermazione tra le seguenti è più appropriata per il p-value?

A) p-value $< 0.10$

B) p-value $< 0.01$

C) p-value $< 0.05$

D) p-value $> 0.10$

**Soluzione**.
Si ha $n = 8$, la media è

$$\bar{x} = (82 + 77 + 85 + 76 + 81 + 91 + 70 + 82)/8 = 80.5.$$

La varianza campionaria corretta è

$$s^2 = ((82-80.5)^2+(77-80.5)^2+(85-80.5)^2+(76-80.5)^2+(81-80.5)^2+(91-80.5)^2+(70-80.5)^2+(82-80.5)^2)/(8-1) = 39.7$$

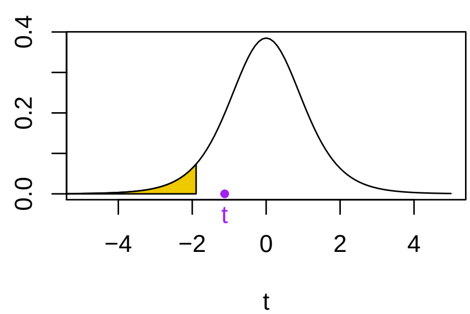
Quindi la deviazione standard è $s = \sqrt{39.71429} = 6.3$.
Il sistema di ipotesi è

$$H_0 : \mu \geq 83 \text{ contro } H_1 : \mu < 83$$

La regione critica unilaterale al livello del 5% è $t < -1.895$ (quantile inferiore t di Student con 7 gradi di libertà al 5%)

Statistica test $= (80.5 - 83)/(6.3/\sqrt{8}) = -1.122$.

Quindi non si rifiuta al livello del 5% perché $-1.122 > -1.895$.




Conclusione. La risposta D) è corretta.

**13.** In una piccola università ci sono 728 matricole e 211 di loro usano Skype. Si consideri un campione di 65 matricole. La probabilità che la proporzione campionaria sia compresa tra 0.27 e 0.3 è (usare 4 cifre decimali) A) 0.2082 B) 0.2592 C) 0.1669 D) 0.1537

**Soluzione** Si considera una popolazione con una proporzione di successi (chi usa Skype) $p = 211/728 = 0.2898352$. Dato un campione di 65 matricole il numero di successi nel campione senza ripetizione ha distribuzione ipergeometrica e in un campione con ripetizione è Binomiale. *Qui supporremo che si tratti di un campione con ripetizione*. La probabilità richiesta è che la proporzione campionaria sia compresa tra 0.27 e 0.3 cioè

$$P(0.27 < \widehat{P} < 0.23) = P(0.27(65) < X < 0.3(65) = P(17.55 < X < 19.5)$$

Si approssima la distribuzione di $\widehat{P}$ con una normale con media $p = 211/728$ e con varianza $pq/n = 0.003166627$. Perciò

$$P(0.27 < \widehat{P} < 0.3) \approx P(0.27 - 0.2898352)/\sqrt{0.003166627} < (0.3 - 0.2898352)/\sqrt{0.003166627})$$

Cioè

$$P(-0.35 < Z < 0.18) = P(Z < 0.18) - [1 - P(Z < 0.35)] = 0.5714 - (1 - 0.6368) = 0.2082.$$

Risposta A) .

Notare che tuttavia la probabilità cercata si potrebbe calcolare esattamente perché

$$P(17.55 < X < 19.5) = P(X = 18) + P(X - 19)$$

e si potrebbe usare la Binomiale con $(n = 65, p = 211/728)$ ottenendo

$$\begin{aligned} P(X = 18) + P(X = 19) &= \binom{65}{18} p^{18} q^{65-18} + \binom{65}{19} p^{19} q^{65-19} \\ &= 0.1071326 + 0.1081578 = 0.2152904. \end{aligned}$$

**14.** Il limite inferiore di un intervallo di confidenza al 95% per la proporzione della popolazione $p$ data una numerosità campionaria $n = 100$ e una proporzione campionaria $\widehat{p} = 0.62$, è pari a: A) 0.525. B) 0.715. C) 0.699. D) 0.540.

**Soluzione**. Il limite inferiore è

$$\widehat{p} - 1.96\sqrt{\widehat{p}(1 - \widehat{p})/n} = 0.62 - 1.96\sqrt{(0.62)(0.38)/100} = 0.525.$$

Risposta A).

**15.** Nel costruire un intervallo di confidenza per la media della popolazione è stato utilizzato un campione di 40 osservazioni. La stima intervallare risultante è stata $28.76 \pm 1.48$. Se la numerosità campionaria fosse stata 160 invece che 40, la stima intervallare sarebbe stata:

A) $28.76 \pm 0.37$.

B) $7.19 \pm 0.37$.

C) $7.19 \pm 1.48$.

D) $28.76 \pm 0.74$.

**Soluzione**

All'aumentare della dimensione campionaria tenendo fermo il livello di confidenza il margine di errore cambia solo in ragione di $\sqrt{n}$. Quindi se $n = 160$ la numerosità campionaria è 4 volte quella di prima e il margine di errore sarà la metà del precedente cioè $1.48/2 = 0.74$. L'esercizio non dice niente (MA DOVREBBE) sulla stima della media nel campione di 160 elementi, ma si suppone che voglia dire che la stima resta la stessa. Quindi la risposta è $28.76 \pm 0.74$. Le altre risposte sono sbagliate per i motivi seguenti: A) ha diviso per 4 il margine di errore B) ha diviso per quattro la media e il margine di errore C) Ha diviso per 4 la media.

**16.** All'aumentare della dimensione del campione l'errore standard della media non cambia. Vero o Falso?

**Soluzione**. L'errore standard della media è $\sigma/\sqrt{n}$. Falso

**17.** Si è selezionato un campione casuale quando ogni elemento nel campione ha la stessa probabilità di essere selezionato e ogni elemento è selezionato in modo indipendente.

**Soluzione**. La risposta fa confusione: il campione è casuale se ogni elemento nella popolazione ha la stessa probabilità di essere scelto e se le osservazioni sono indipendenti.

**18.** Il Teorema del Limite Centrale afferma che la distribuzione della media campionaria sarà sempre più prossima ad una distribuzione normale indipendentemente dalla dimensione dei campioni. Vero o Falso?

**Soluzione**. Falso. Nel teorema è cruciale che il campione abbia una numerosità alta per poter fare l'approssimazione.

**19.** Un intervallo di confidenza al 95% per la proporzione nella popolazione $p$ è dato dall'intervallo da 0.214 a 0.336. Basandosi su tale informazione, la proporzione campionaria che ha determinato tale intervallo è 0.234.

**Soluzione**

L'intervallo è $\widehat{p} \pm ME$. Siccome il punto medio è $\widehat{p}$ risulta

$$\widehat{p} = (0.336 + 0.214)/2 = 0.275.$$

Quindi è Falso.

**20.** Il livello di significatività di un test è la probabilità che l'ipotesi nulla sia falsa. Vero o Falso?

**Soluzione**. Il livello del test è la massima probabilità di commettere errore del primo tipo cioè di rifiutare $H_0$ quando è vera. Quindi È FALSO.

**21.** L'errore di stima è la differenza tra una stima ed il parametro. Vero o Falso?

**Soluzione**. Vero.

**22.** Si verifica un errore del I tipo quando viene rifiutata un'ipotesi nulla vera. Vero o Falso?

Vero.

**23.** Il valore atteso della media della popolazione è uguale alla media campionaria. Vero o Falso?

**Soluzione**. Falso. La frase giusta è: il valore atteso della media campionaria è uguale alla media della popolazione, qualunque essa sia.

**24.** Lo stimatore $\widehat{\theta}$ è uno stimatore corretto per $\theta$ se $\mathrm{var}(\widehat{theta}) = \mathrm{var}(\theta)$. Vero o Falso?

**Soluzione**. Falso, è corretto se $E(\widehat{\theta}) = \theta$.

**25.** Un modo per ridurre il margine di errore in un intervallo di confidenza è diminuire il livello di confidenza. Vero o Falso?

**Soluzione**. Risulta: $ME = z_{\alpha/2}\sigma/sqrtn$. Quindi se si diminuisce il livello di confidenza per esempio da 0.99 a 0.95, $z_{\alpha/2}$ si riduce da 2.58 a 1.96. Quindi il ME si riduce. Tuttavia la domanda è molto ingannevole perché in realtà l'ampiezza dell'intervallo si riduce, ma anche il livello di confidenza.

## 0.1 Esercizi aggiuntivi

**26.** Il tempo che gli studenti dedicano allo studio segue una distribuzione normale con deviazione standard di 8 ore. Si estrae un campione casuale di 4 studenti. La probabilità che la media campionaria differisca dalla media della popolazione per più di 4 ore è

A) 0.2987

B) 0.3080

C) 0.3174

D) 0.3085

**Soluzione**. Tempo $= X \sim N(\mu =?, \sigma = 8)$. Se ho un campione di dimensione $n = 4$, la media campionaria $\overline{X} = (X_1 + X_2 + X_3 + X_4)/4$ ha distribuzione

$$\overline{X} \sim N(\mu =?, 8/2 = 4).$$

La probabilità che $\overline{X}$ differisca da $\mu$ per più di 4 ore è

$$P(|\overline{X} - \mu| > 4) = 1 - P(-4 < \overline{X} - \mu < 4) = 1 - P(\mu - 4 < \overline{X} < \mu + 4)$$

(fate attenzione a queste disuguaglianze, studiatele con calma)

Quindi se si standardizza $\overline{X}$ rispetto alla sua media $\mu$ e alla sua deviazione standard $= 4$ si ha

$$P(|^{-}\mu| > 4) = 1 - P(-1 < Z < 1) = 1 - (0.8413 - (1 - 0.8413)) = 0.3174$$

e quindi la risposta corretta è la C).

**27.** Sia $X$ una variabile casuale normale con media $\mu$ e deviazione standard $\sigma$. Sia $\overline{X}$ la media di un campione casuale selezionato dalla popolazione descritta dalla variabile casuale $X$. Quale delle seguenti affermazioni è vera?

A) $P(\overline{X} < \mu + \sigma/\sqrt{n}) < P(\overline{X} < \mu - \sigma/\sqrt{n})$

B) $P(\overline{X} < \mu + \sigma/(2n)) < P(\overline{X} < \mu + \sigma/n)$

C) $P(\overline{X} < \mu + \sigma/\sqrt{n}) = P(\overline{X} < \mu - \sigma/\sqrt{n})$

D) $P(\overline{X} < \mu - \sigma/(2n)) < P(\overline{X} < \mu - \sigma/n)$

**Soluzione**

Le disuguaglianze si possono scrivere così dopo aver standardizzato.

A) $P(Z < 1) < P(Z < -1)$, Falsa.

B) $P(X < 1/(2\sqrt{n})) < P(Z < 1/\sqrt{n})$, Vera

C) $P(Z < 1) = P(Z < -1)$, Falsa

D) $P(Z < 1/(2\sqrt{n})) < P(Z < -1/\sqrt{n})$, Falsa.

Le risposte accanto si deducono facendo il grafico della normale standard e notando che $1/(2\sqrt{n}) < 1/\sqrt{n}$. Esercizio complicato.