Alcuni compiti di precedenti appelli

G. Marchetti

2016 ver. 1

Indice

1 Compito 2 gennaio 2012

12. Sia X una variabile casuale distribuita come una normale con media μ e deviazione standard σ. Siano $\bar{X}_1$ e $\bar{X}_2$ le medie di due campioni di dimensione rispettivamente pari a 16 e 25 unità selezionati casualmente ed indipendentemente dalla popolazione descritta dalla v.c. X. Si considerino le seguenti espressioni:

$$(1) \quad P(\bar{X}_1 < \mu), \qquad (2) P(\bar{X}_2 < \mu)$$

Quale delle seguenti affermazioni è vera?

A) Non è possibile stabilire una relazione tra le due espressioni.
B) Il valore dell espressione 1 è uguale al valore dell espressione 2 .
C) Il valore dell espressione 1 è maggiore del valore dell espressione 2.
D) Il valore dell espressione 2 è maggiore del valore dell espressione 1.

Soluzione

$\bar{X}_1 \sim N(\mu, \sigma^2/16)$ e $\bar{X}_2 \sim N(\mu, \sigma^2/25)$. Quindi

$$P(\bar{X}_1 < \mu) = P(Z < 0) = 0.5 \qquad P(\bar{X}_2 < \mu) = P(Z < 0) = 0.5$$

Perciò B).

13. Se X è una variabile casuale binomiale con $n = 20$, e $p = 0.5$, allora P(X = 20) = 1.0.$ Vero o Falso?

Soluzione

$$P(X = 20) = \binom{20}{20} 0.5^{20} \cdot 0.5^0 = 0.5^{20} \neq 1$$

quindi Falso.

15. In una piccola università ci sono 728 matricole e, tra loro, 211 usano Skype. Si consideri un campione di 65 matricole. La probabilità che la proporzione campionaria sia inferiore a 0.3 è (usare 4 cifre decimali)

A) 0.4249
B) 0.5714
C) 0.4388
D) 0.5866

Soluzione Si considera una popolazione dicotomica con una proporzione di successi (chi usa Skype) $p = 211/728$. Dato che il campione di 65 matricole è abbastanza grande (e $np(1-p) = 13.379 > 9$) la proporzione $\hat{P}$ di successi nel campione è approssimativamente normale con media $p = 211/728$ e con deviazione standard $\sigma = \sqrt{p(1-p)/n} = 0.05627279$.

La probabilità richiesta è che la proporzione campionaria sia < 0.3 cioè

$$P(\hat{P} < 0.30) = P[(0.30 - p)/\sigma)] = P(Z < 0.18) = 0.5714.$$

Quindi la risposta è B).

2 Compito 13 febbraio 2012

16. Negli ultimi 10 anni, i rendimenti dello Stock A sono stati in media dell'8.4% con deviazione standard 2.1% e quelli dello Stock B sono stati in media del 3.6% con deviazione standard 0.9%. Quale delle seguenti affermazioni è vera?

A) Lo Stock B ha una variabilità relativa più piccola dello Stock A.
B) I due stock presentano la stessa variabilità relativa.
C) Lo Stock A ha una variabilità relativa più piccola dello Stock B.
D) Non c è sufficiente informazione per dire quale dei due stock presenta la variabilità relativa più piccola.

Soluzione La variabilità relativa si misura con il coefficiente di variazione $CV = \sigma/\mu$. Quindi

$$CV_A = 2.1/8.4 = 0.25, \qquad CV_B = 0.9/3.6 = 0.25$$

e quindi A e B hanno la stessa variabiità relativa. B).

3 Compito del 2 settembre 2015

1. Un campione casuale di dimensione 15 estratto da una popolazione distribuita normalmente fornisce una media campionaria di 75 ed una varianza campionaria di 25. Il limite superiore di un intervallo di confidenza al 95% per la media campionaria sarà pari a:

A) 72.727
B) 77.769
C) 72.231
D) 77.273

Soluzione

Limite superiore $= \bar{x} + t_{14,\alpha/2}s/\sqrt{n} = 75 + 2.145\sqrt{(}25/15) = 77.769$. B)

2. Sia $Z \sim N(\mu = 0, \sigma^2 = 1)$. Calcolare $P(-0.44 < Z < 1.2)$.

Soluzione Posto $F(z) = P(Z \leq z)$ (tavola 1 della normale, funzione di ripartizione)

$$(-0.44 < Z < 1.2) = F(1.2) - [1 - F(0.44)] = 0.5549.$$

3. L'azienda produttrice di un nuovo macchinario asserisce che questo produce in media 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. Fissato $\alpha = 0.05$, quale dovrebbe essere la regola di decisione?

A) Si rifiuta H0 se la statistica t è < -2.145.
B) Si rifiuta H0 se la statistica t è > 2.145.
C) Si rifiuta H0 se la statistica t è < -1.761.

D) Si rifiuta H0 se la statistica t è > 1.761.

Soluzione

$H_0 : \mu \geq 29$ contro $H_1 : \mu < 29$. $n = 15$ macchinari con $\bar{x} = 26$ e $s = 6.2$. Statistica test

$$t = (\bar{x} - 29)/(6.2/\sqrt{15})$$

Si rifiuta H_0 se $ t < - t_{0.05, 14} = 1.761$.

4. Supponi di voler effettuare con un livello di significatività $\alpha = 0.10$ il seguente test sulla media di una popolazione normale : $H_0 : \mu = 277$ vs. $H_1 : \mu \neq 277$. Supponi inoltre di sapere che la deviazione standard della popolazione è $\sigma = 13.5$. Se selezioni un campione casuale di 20 osservazioni, per quali valori della media campionaria rifiuterai l'ipotesi nulla?

5. Siano X_1, X_2, X_3, e X_4 le osservazioni di un campione casuale semplice estratto da una popolazione con media μ e varianza σ^2. Si consideri il seguente stimatore di μ

$$T = 0.15X_1 + 0.35X_2 + 0.20X_3 + 0.30X_4.$$

Qual è la varianza di T?

A) 0.55
B) $0.55\sigma^2$
C) $0.275\sigma^2$
D) 0.275

Soluzione

$$\begin{aligned}
\text{var}(T) &= \text{var}(0.15X_1 + 0.35X_2 + 0.20X_3 + 0.30X_4) \\
&= 0.15^2\sigma^2(X_1) + 0.35^2\sigma^2(X_2) + 0.20^2\sigma^2(X_3) + 0.30^2\sigma^2(X_4) \\
&= 0.15^2\sigma^2 + 0.35^2\sigma^2 + 0.20^2\sigma^2 + 0.30^2\sigma^2 \\
&= \sigma^2(0.15^2 + 0.35^2 + 0.20^2 + 0.30^2) \\
&= \sigma^2 0.275.
\end{aligned}$$

6. Un professore ha raccolto i dati sul numero di assenze ad un corso introduttivo di Giornalismo della durata di un semestre in una classe di 100 studenti. I dati sono raccolti nella tabella seguente.

# di assenze	0	1	2	3	4	5	6
# di Studenti	5	13	24	23	17	11	7

Qual è la media del numero di assenze nel semestre?

A) 3.07
B) 2.95
C) 3.14
D) 2.00

Soluzione

Media = Totale assenze/totale studenti = $\frac{\sum_i x_i n_i}{\sum_i n_i}$ cioè

$$Media = \frac{0 \cdot 5 + 1 \cdot 13 + 2 \cdot 24 + 3 \cdot 23 + 4 \cdot 17 + 5 \cdot 11 + 6 \cdot 7}{5 + 13 + 24 + 23 + 17 + 11 + 7)} = 2.95.$$

7. Quando si seleziona un campione casuale di dimensione n da una popolazione normalmente distribuita la distribuzione della media campionaria sarà normale indipendentemente dalla dimensione campionaria. Vero (T) o Falso (F)?

Soluzione Se $X \sim N(\mu, \sigma^2)$ allora la media nell'universo dei campioni è $\bar{X} \sim N(\mu, \sigma^2/n)$. Quindi qualunque sia n $\bar{X}$ e normale. Vero.

8. Data una variabile casuale normale X con media 70 e deviazione standard 12, il valore della variabile casuale normale standardizzata Z corrispondente a X = 82 è più grande di zero. Vero (T) o Falso (F)?

Soluzione Z si ottiene da X con la trasformazione

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 70}{12}$$

quindi se $X = 82$

$$Z = \frac{82 - 70}{12} = 12/12 = 1 > 0.$$

Quindi è vero.

9. Una recente indagine ha studiato la condizione lavorativa delle coppie che costituiscono una famiglia. I dati mostrano che nell 88% delle coppie almeno uno dei membri lavora. Nel 20% delle coppie in cui la donna non lavora, nemmeno l'uomo lavora. Inoltre, nel 40% delle coppie in cui l uomo non lavora, nemmeno la donna lavora. Qual è la probabilità che né l'uomo né la donna lavorino?

Soluzione

Siano M ed F gli eventi M = l'uomo lavora e F = la donna lavora. Dunque $P(M \cup F) = 0.88$. Inoltre $P(\bar{M} \mid \bar{F}) = 0.20$ e $P(\bar{F} \mid \bar{M}) = 0.4$.

Allora

$$P(\bar{M} \cap \bar{F}) = 1 - P(M \cup F) = 1 - 0.88 = 0.12.$$

Gli altri dati non servono.

10. Le vendite da parte di un grossista di impianti idraulici avvengono tramite spedizione oppure a pronta consegna. In un mese le vendite pronta consegna hanno una media di $,102972 con una deviazione standard di $,13523. Durante lo stesso mese le spedizioni hanno una media di $,242354 con una deviazione standard di $,24956. Supponiamo che le vendite tramite spedizione e a pronta consegna siano indipendenti. Calcolare la media e la deviazione standard del totale delle vendite mensili.

Soluzione

Sia X = vendite a pronta consegna e Y = vendite per spedizione. Sappiamo che

$$\mu_X = 102972 \quad \mu_Y = 242354; \qquad \sigma_X = 13523, \quad \sigma_Y = 24956.$$

La media e la deviazione standard delle vendite totali $T = X + Y$ sono

$$\mu_T = \mu_X + \mu_Y = 102972 + 242354 = 345326.$$

e, poiché X e Y sono indipendenti e quindi $\sigma_{XY} = 0$,

$$\sigma_T = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{805673465} = 28384.$$

11. Considera la seguente distribuzione del numero di studenti secondo il numero di assenze scolastiche. Quanti studenti hanno fatto una sola assenza?

Mod.	Frequenza	Percentuale Cumulata
0	200	20.00
1	*	50.00
2	350	85.00
3	100	95.00
4	50	100.00

A) 250

B) 300
C) 50
D) 150

Soluzione

Basta trovare le frequenze relative percentuali e fare una proporzione per le prime due righe.

Mod.	Frequenza	Percentuale
0	200	20.00
1	x	30.00

Perciò

$$200\colon 20.00 = x\colon 30.00$$

cioè $x = 200 \cdot 30/20 = 300$.

12. La probabilità che una persona acquisti un costume da bagno durante le vacanze a Forte dei Marmi è 0.4. Si selezionano a caso 10 persone. Qual è la deviazione standard del numero di persone che compreranno un costume?

A) 1.125
B) 1.245
C) 1.549
D) 1.265

Soluzione Se S = numero di persone che compreranno il costume su 10 selezionati casualmente con ripetizione $S = X_1 + \cdots + X_{10}$ ha distribuzione binomiale con media $\mu_S = np = 10 \cdot 0.4 = 4$ e deviazione standard

$$\sigma = \sqrt{np(1-p)} = \sqrt{4 \cdot 0.6} = 1.549.$$

13. Una statistica campionaria tale che la media dei valori che può assumere nell universo dei campioni è uguale al valore del parametro della popolazione che vuole stimare è uno stimatore corretto. Vero (T) o Falso (F)?

Soluzione

Sì è la defінzione esatta.

14. Se la covarianza tra due variabili casuali X e Y è positiva, allora $\text{var}(X-Y)$ è più grande di $\text{var}(X+Y)$. Vero (T) o Falso (F)?

Soluzione Poiché $\sigma_{XY} \geq 0$

$$\text{var}(X-Y) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} \leq \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \text{var}(X+Y)$$

Quindi Falso.

15. Da un'indagine condotta su 472 operai, è risultato che il 63% pensa di aderire ad uno sciopero sindacale nei prossimi tre mesi. Quali dei seguenti intervalli rappresenta un intervallo di confidenza al 98% per la proporzione degli operai che hanno pianificato di aderire ad uno sciopero nei prossimi tre mesi?

A) 0.63 ± 0.042
B) 0.63 ± 0.052
C) 0.63 ± 0.047
D) 0.63 ± 0.057

Soluzione Il margine di errore (per grandi campioni) è

$$ME = z_{0.01}\sqrt{\hat{p}(1-\hat{p})/n} = 2.326\sqrt{0.63(1-0.63)/472} = 0.052$$

Nota che il livello di confidenza è $1 - \alpha = 0.98$ da cui $\alpha/2 = 0.01$

16. Un dirigente esamina i vestiti da lavoro presenti nel suo armadio. Ha cinque abiti (giacca e pantalone), sei camicie e tre paia di scarpe. Poiché è in partenza per un viaggio, decide di prenderne due di ogni tipo (due abiti, due camicie, due paia di scarpe). Quante sono le possibili combinazioni di completi (abito+camicia+scarpe)?

A) 680
B) 320
C) 224
D) 450

Soluzione

Sono

$$\binom{5}{2}\binom{6}{2}\binom{3}{2} = (10)(15)(3) = 450.$$

17. Il tempo di attesa ad un ufficio postale è approssimativamente distribuito come una variabile casuale uniforme tra 0 e 120 secondi. Calcolare la probabilità di dover aspettare il vostro turno tra 15 e 45 secondi.

A) 0.25
B) 0.45
C) 0.35
D) 0.15

Soluzione

La probabilità è proporzionale alla lunghezza dell'intervallo

$$p = \frac{45-15}{120-0} = 0.25.$$

18. Supponiamo di lanciare due dadi. Si consideri la somma dei due dadi: sia E l'evento: "si osserva un numero pari" e G l'evento: "si osserva un numero maggiore di 7". Qual è l'evento $\bar{E} \cup \bar{G}$?

A) {2, 3, 4, 5, 6, 7, 9, 11}
B) {2, 3, 4, 5, 6, 7}
C) {2, 4, 6}
D) {3, 5, 7}

Soluzione

$E = \{2, 4, 6, 8, 10, 12\}$, $G = \{8, 9, 10, 11, 12\}$. Quindi

$$\bar{E} = \{3, 5, 7, 9, 11\}, \qquad \bar{G} = \{2, 3, 4, 5, 6, 7\}$$

e dunque

$$E \cup G = \{2, 3, 4, 5, 6, 7, 9, 11\}.$$

19. Un'ipotesi nulla è rifiutata a livello di significatività 0.025, ma non ad un livello di 0.01. Ciò significa che il p-value del test è compreso tra 0.01 e 0.025. Vero (T) o Falso (F)?

Soluzione

Vuol dire il p-value è minore di 0.025 e maggiore di 0.01. Vero.

20. Il tempo impiegato per evadere un'ordinazione in una pizzeria è distribuito normalmente con media 4.1 minuti e deviazione standard 1.3 minuti. Qual è la probabilità che il tempo di attesa medio di un campione casuale di 10 clienti sia compreso tra 4.0 e 4.2 minuti?

Soluzione

$\bar{X} \sim N(4.1, 1.3/\sqrt{10})$ quindi

$$P(4.0 \leq X \leq 4.2) = P\left(\frac{4-4.1}{0.4110961} \leq Z \leq \frac{4.2-4.1}{0.4110961} =\right) = P(-0.24 \leq Z \leq 0.24) = F(0.24) - (1-F(0.24)) = 0.1896.$$

4 Compito F del 7 giugno 2012

1. L'azienda produttrice di sacchi di cemento afferma che ciascun sacco contiene almeno 50.1 kg di cemento. La deviazione standard della quantità di cemento contenuta in ciascun sacco è 1.2 kg. La regola di decisione adottata dall'azienda è di mettere in manutenzione una macchina riempitrice se la media campionaria della quantità di cemento in un campione di 40 sacchi è inferiore a 49.7. Qual è la probabilità di commettere un errore del primo tipo?

A) 0.024
B) 0.017
C) 0.030
D) 0.036

Soluzione

Affermazione: $\mu \geq 50.1$ dove $\mu =$ quantità media di cemento. Questa è l'ipotesi H_0 perché contiene il segno di uguaglianza. La deviazione standard $\sigma = 1.2$ è per la popolazione perché ancora non è stato estratto nessun campione. La regola di decisione è: rifiuta (e metti in manutenzione la macchina) se

$$\bar{X} < 49.7 \text{ in un campione di 40 sacchi.}$$

La probabilità di commettere errore del I tipo è il p-value associato alla media 49.7.

La statistica test per $H0 : \mu \geq 50.1$ contro $H_1 : \mu < 50.1$ è

$$Z = \frac{\bar{X} - 50.1}{1.2/\sqrt{40}} \sim N(0,1)$$

Ora se $\bar{X} = 49.7$

$$Z = (49.7 - 50.1)/(1.2/\sqrt{40}) = -2.11$$

Il p-value è

$$P(Z \leq -2.11) = 1 - P(Z < 2.11) = 1 - 0.9826 = 0.0174.$$

2. La probabilità che una persona prenda il raffreddore durante l'inverno è 0.4. Si selezionano a caso 10 persone. Qual è la probabilità che esattamente 4 di loro prenderanno il raffreddore?

A) 0.242
B) 0.251
C) 0.502
D) 0.751

Soluzione Sono 10 prove di Bernoulli indipendenti ciascuna con probabilità $p = 0.4$ e $q = 0.6$. Se X è il numero di "successi"

$$P(X = 4) = \binom{10}{4} p^4 q^6 = 210 \cdot (0.4^4)(0.6^6) = 0.251.$$

3. Il responsabile della lotteria nazionale ha dichiarato che il reddito medio familiare annuo di coloro che giocano alla lotteria è superiore a 37000 euro. Si suppone che la distribuzione del reddito di tale categoria di individui si distribuisca normalmente con deviazione standard di 5756 euro. Si supponga che in un campione di 25 famiglie che giocano alla lotteria si sia rilevato un reddito medio annuo pari a 36243 euro.Qual è la statistica test osservata?

A) Z $= -0.66$

B) Z = 1.92
C) t = 0.66
D) t = 1.92

Soluzione Affermazione: $\mu > 37000$. È l'ipotesi alternativa perché non contiene l'uguaglianza. Quindi $H_0 : \mu \leq 37000$. La popolazione è $N(\mu, \sigma = 5756)$. La media campionaria è $\bar{X} = 36243$ su 25 osservazioni. La statistica test è

$$Z = (36243 - 37000)/(57/\sqrt{25}) = -0.66.$$

4. Se si rifiuta l'ipotesi nulla contro l'ipotesi alternativa ad un livello di significatività del 5% , allora, con gli stessi dati deve essere rifiutata anche ad un livello di significatività dell'1%. Vero o Falso?

Soluzione. Falso perché se si rifiuta al livello del 5% vuol dire che possiamo commettere errore del primo tipo anche 5 volte su 100. In tal caso violeremmo il criterio di un test che richiede un livello di errore minore dell'1%.

5. L'U.S. Postal Service afferma che almeno il 63.4% di un certo pubblicità postale viene letta dai destinatari. Un gruppo ambientalista vuole verificare tale affermazione. Intervistano un campione casuale di 220 famiglie ed osservano che solo il 58.7% legge la pubblicità. Dovremmo credere all'affermazione dell'U.S. Postal Service ad un livello di significatività del 10%?

Soluzione

Affermazione: la proporzione di successi è $p \geq 63.4\%$. È l'ipotesi nulla perché contiene l'uguaglianza. Dunque il sistema di ipotesi è $H_0 : p > 0.634$ contro $H_1 : p < 0.634$.

La statistica test usa l'approssimazione normale

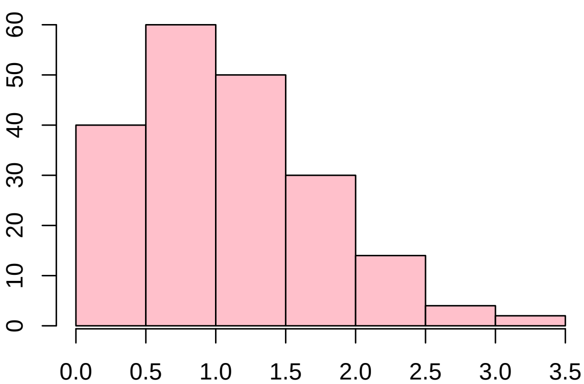
$$Z = \frac{0.587 - 0.634}{\sqrt{\frac{0.634(1 - 0.634)}{220}}} = -1.447$$

La regione critica è $Z < -z_{0.1} = -1.282$. Poiché il valore osservato -1.447 cade nella regione critica si rifiuta l'affermazione.

6. Data la seguente distribuzione delle famiglie per classi di spesa mensile (in migliaia di euro), disegnare un istogramma.

Classi di spesa	$0 \dashv 0.5$	$0.5 \dashv 1$	$1 \dashv 1.5$	$1.5 \dashv 2$	$2 \dashv 2.5$	$2.5 \dashv 3$	$3 \dashv 3.5$
% famiglie	20	30	25	15	7	2	1

Soluzione



7. La covarianza tra due variabili X e Y:

A) è compresa tra -1 e $+1$.
B) può essere positiva o negativa.
C) è sempre positiva.

D) è sempre minore di 1.

Soluzione

A) Falso: i limiti sono $\pm\sigma_X\sigma_Y$
B) Vero
C) Falso
D) Falso

8. Il numero di televisioni che escono ogni giorno da una certa linea di produzione si distribuisce come una variabile casuale con deviazione standard di 17.4. La media giornaliera della linea di produzione determinata su un campione di 20 giorni è 452.3. Quale dei seguenti intervalli rappresenta un intervallo di confidenza del 95% per la media della produzione in un giorno?

A) 453 ± 9.4
B) 452.3 ± 13.8
C) 452.3 ± 11.3
D) 452.3 ± 7.63

Soluzione $X \sim ?(\mu = ?, \sigma = 17.4)$. In un campione con $n = 20$ $\bar{X} = 452.3$. Se X fosse normale l'intervallo di confidenza sarebbe

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 452.3 \pm 1.9617.4/\sqrt{20} = 452.3 \pm 7.63.$$

9. Un analista finanziario fornisce le stime dell'utile di un'azienda nel prossimo anno, considerando anche il tasso di interesse. Qual è la probabilità che l'azienda realizzi un utile di almeno l'8%?

Tasso di interesse	Utile:	$< 8\%$	da 8 a 12%	$> 12\%$
$< 3\%$		0.09	0.15	0.16
da 3% a 5%		0.14	0.17	0.05
$< 5\%$		0.16	0.07	0.01

La somma delle probabilità è 1 e quindi è una tavola congiunta di probabilità.

$$P(\text{Utile} \geq 8\%) = 1 - (0.09 + 0.14 + 0.16) = 0.61.$$

10. Un professore ha raccolto dati sul numero di assenze ad un corso introduttivo di Statistica della durata di un semestre in una classe di 100 studenti. I dati sono raccolti nella tabella seguente. Qual è la media del numero di assenze per semestre?

# di assenze	0	1	2	3	4	5	6
# di studenti	5	13	24	23	17	11	7

A) 3.07
B) 2.95
C) 3.14
D) 2.00

Soluzione La media è = numero totale assenze / numero totale di studenti. Quindi

$$\bar{x} = \frac{1}{100}(0 \cdot 5 + 1 \cdot 13 + 2 \cdot 24 + 3 \cdot 23 + 4 \cdot 17 + 5 \cdot 11 + 6 \cdot 7) = 295/100 = 2.95.$$

11. Se A e B sono mutuamente esclusivi allora $\bar{A}$ e $\bar{B}$ devono essere mutuamente esclusivi. Vero o falso?

Soluzione. Falso perché per esempio essere divorziato e essere sposato sono mutuamente esclusivi ma non essere divorziato e non essere sposato non sono mutuamente esclusivi perché uno potrebbe essere single.

12. Sia X una variabile casuale normale con media μ e deviazione standard σ. Sia $\bar{X}$ la media di un campione casuale selezionato dalla popolazione descritta dalla variabile casuale X. Quale delle seguenti affermazioni è vera?

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

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

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

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

Soluzione

Siccome $Z = \bar{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0,1)$ le affermazioni si possono tradurre come segue

A) $P(Z < 1) < P(Z < -1)$: Falsa
B) $P(Z < \sqrt{n}/(2n) < P(Z < \sqrt{n}/n)$: Vera
C) $P(Z < 1) = P(Z < -1)$: Falsa
D) $P(Z < -\sqrt{n}/(2n)) < P(Z < -\sqrt{n}/n)$: Falsa.

Esercizio difficile!

13. L'errore di stima è la differenza tra il valore di una statistica determinata su un campione ed il corrispondente valore del parametro determinato nella popolazione. Vero o Falso?

Soluzione. Vero, è $T - \theta$. Per esempio per la media è $\bar{X} - \mu$.

14. Un test a risposta multipla ha 5 domande, ognuna con 5 possibili risposte da A a E. Supponi di tirare a caso per tutte le domande. Qual è la probabilità di rispondere correttamente a tutte le domande?

A) 0.20
B) 0.50
C) 0.00032
D) 0.03125

Soluzione

È la probabilità di avere 5 successi su 5 prove ciascuna delle quali ha probabilità $p = 1/5 = 0.2$. Quindi dato che le prove sono indipendenti si ha $p^5 = 0.2^5 = 0.00032$.

15. Il grafico della variabile casuale binomiale è sempre simmetrico. Vero o Falso?

Soluzione Falso. Il diagramma a barre è simmetrico solo se la probabilità di successo è $p = 5$.

16. Un portafoglio comprende 20 azioni ALFA e 30 azioni BETA. Il prezzo delle azioni ALFA è una variabile casuale con media 10 e varianza 9, il prezzo delle azioni BETA è una variabile casuale con media 25 e varianza 16. I prezzi delle due azioni sono correlati negativamente con un coefficiente di correlazione lineare pari a -0.4. Calcolare il valore atteso e la varianza del valore del portafoglio.

Soluzione

Il valore del portafoglio è $X = 20A + 30B$. Il valore atteso è

$$E(X) = 20E(A) + 30E(B) = 20 \cdot 10 + 30 \cdot 25 = 200 + 750 = 950.$$

La varianza è

$$\sigma^2[20E(A) + 30E(B)] = 20^2\sigma^2(A) + 30^2\sigma^2(B) + 2 \cdot 20 \cdot 30\sigma(A, B)$$

poiché la covarianza tra A e B è

$$\sigma(A, B) = \rho(A, B)\sigma(A)\sigma(B) = -0.4 \cdot 3 \cdot 4 = -4.8$$

avremo

$$\sigma^2[20E(A) + 30E(B)] = 400 \cdot 9 + 900 \cdot 16 + 1200 \cdot (-4.8) = 12240.$$

17. Una squadra di manutenzione è responsabile di un certo tratto di oleodotto lungo 4 chilometri. La distanza (in chilometri) alla quale può verificarsi un guasto può essere rappresentato da una variabile casuale continua uniforme. Specificare la funzione di densità e calcolare la probabilità che si verifichi un guasto tra 0 e 1 km.

Soluzione La densità uniforme è $f(x) = 1/4$ per $x \in [0, 4]$. La probabilità è

$$\Pr(0 < X < 1) = 1/4$$

18. Sia $X \sim N(2, 1)$. La differenza interquartile è pari a

A) 0.2266
B) 2.78
C) -4
D) 1.34

Soluzione La differenza interquartile di una $N(0, 1)$ è $Q_3 - Q_1$ dove $P(Z < Q_3) = 0.75$ e $P(Z < Q_1) = 0.25$. Dalle tavole si vede che $P(Z < 0.67) = 0.75$. Quindi $P(Z < -0.67) = 0.25$. Fate un disegno!

Allora la differenza interquartile è $IQR = 0.67 - (-0.67) = 2 \cdot 0.67 = 1.34$.

La differenza interquartile della normale $N(2, 1)$ è la dispersione di una normale ottenuta spostando la normale standard a destra di 2. Quindi la dispersione resta la stessa ed è uguale a 1.34. Volendo verificare precisamente, basta notare che se $X \sim N(2, 1)$ $P(X < 2 + 0.67) = 0.75$ e $P(X < 2 - 0.67) = 0.25$. Quindi lo scarto interquartile è $IQR = (2 + 0.67) - (2 - 0.67) = 1.34$.

19. 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 4ore è

A) 0.2987
B) 0.3080
C) 0.3174
D) 0.3085

Soluzione Se $X \sim N(\mu, \sigma = 8)$ la media in un campione di $n = 4$ ha distribuzione $\bar{X} \sim N(\mu, 8/2 = 4)$. Perciò la probabilità che la media campionaria differisca da μ per più di 4 è

$$P(|\bar{X} - \mu| > 4) = P(|\bar{X} - \mu|/4 > 1) = P(|Z| > 1)$$

dove Z $N(0, 1)$.

Quindi si ha $2(1 - P(Z < 1)) = 2(1 - 0.8413) = 0.3174$. Risposta C.

20. La media è una misura di tendenza centrale migliore della mediana quando ci sono valori anomali. Vero o Falso?

Soluzione È esattamente il contrario. Falso.

5 Compito H del 7 giugno 2012

1. Una variabile è classificata come ordinale se:

A) I dati si ottengono con un processo di misurazione continuo
B) Esiste un ordinamento naturale delle sue modalità
C) Osserviamo la variabile per un periodo di tempo
D) Non esiste un ordinamento naturale delle sue modalità

Soluzione

Una variabile si dice ordinale se esiste un ordinamento naturale delle modalità. Per esempio la religione non è ordinale, mentre il livello di istruzione sì. Quindi la risposta è B).

2. Quale delle seguenti affermazioni è vera?

A) La media è sempre più piccola della mediana.

B) La deviazione standard è sempre più piccola della varianza.
C) La somma degli scarti dalla media è sempre zero.
D) La somma degli scarti al quadrato dalla media è sempre zero.

Soluzione

A) FALSO, può capitare tranquillamente il contrario. Per esempio $\{1, 2, 30\}$ ha media 11 e mediana 2.

B) FALSO, poiché la dev. st. è la radice della varianza. Se la varianza è per esempio 0.25 la deviazione standard è 0.5 che è più grande.

C) VERO, perché
$$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = n\bar{x} - n\bar{x} = 0.$$

D) FALSO. È evidente che in generale gli scarti al quadrato sono positivi e quindi la loro somma non può essere zero a meno che tutti i dati non siano uguali.

3. Siete incaricati di organizzare il catering per un convegno. La vostra società è responsabile di tutti gli sprechi e quindi non volete ordinare troppi pasti che potrebbero non essere consumati. Supponiamo che il numero di persone partecipanti al convegno segua una distribuzione normale con valore atteso 84 e varianza 16. Qual è il numero di pasti che dovreste ordinare in modo tale che la probabilità di avere più persone che pasti sia del 5%?

Soluzione Numero di persone che vanno al convegno $X \sim N(84, \sigma = 4)$. La domanda è ingannevole: i pasti sono la stessa cosa delle persone. Se si fissano $k = 84$ pasti per esempio la probabilità che le persone siano più dei pasti è $P(X > 84) = 0.5$ perché 84 è la media. Qui invece si chiede trovare il numero di pasti k tale che $P(X > k) = 0.05$. Quantile 0.95.

Si standardizza ottenendo
$$P(Z > (k - 84)/4) = 0.05$$
Allora dalle tavole sappiamo che $(k-84)/4 = 1.645 = 90.58$. Qundi, se si ordinano 91 pasti c'è la probabilità di solo il 5% di avere più persone che pasti.

4. Si supponga che il punteggio medio ad un esame sia 73 con una deviazione standard di 2. In accordo con il teorema di Chebychev, al più il 60% dei punteggi sono compresi nell'intervallo tra 70 e 76. Vero o falso?

Soluzione Per la diseguaglianza di Chebyshev

$$P(70 < X < 76) = P(73 - 2\sigma < X < 73 + 2\sigma) \geq 0.75$$

Perciò l'affermazione "al più il 60% dei punteggi sono compresi nell'intervallo tra 70 e 76" è FALSA.

5. Un commercialista afferma di poter completare una dichiarazione dei redditi standard in meno di un ora. Per un campione di 24 dichiarazioni, il commercialista impiega una media di 63.2 minuti con una deviazione standard di 7.7 minuti. Qual è la statistica test osservata?

A) t = 2.04
B) Z = 1.79
C) Z = 2.04
D) t = 1.79

Soluzione L'ipotesi nulla è $\mu \geq 60$ min contro $H_1 : \mu < 60$. Supponendo che il tempo per completare dichiarazione sia distribuito normalmente con media e varianza incognita, si usa un test t di Student. La statistica test è
$$t = (63.2 - 60)/(7.7/\sqrt{24}) = 2.04.$$
La soluzione perciò è A) $t = 2.04$ e non C) $z = 2.04$ (test z) perché 7.7 è la deviazione standard stimata e non quella della popolazione normale.

6. Da un indagine condotta su 472 direttori del personale, è risultato che il 63% pensa di assumere nuovo personale nei prossimi tre mesi. Quali dei seguenti intervalli rappresenta un intervallo di confidenza del 98% per la proporzione dei direttori del personale che hanno pianificato di assumere nuovo personale nei prossimi tre mesi?

A) 0.63 ± 0.042
B) 0.63 ± 0.047
C) 0.63 ± 0.057
D) 0.63 ± 0.052

Soluzione Se il livello è $1-\alpha = 0.98$ $\alpha = 0.02$ e $\alpha/2 = 0.01$.

Il margine di errore è

$$ME = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = z_{0.01}\sqrt{0.63 \cdot 0.37/472} = 2.326 \cdot 0.02222287 = 0.052.$$

7. In una recente indagine su 600 adulti, il 16.4% ha dichiarato di essersi addormentato almeno una volta di fronte alla televisione nel mese scorso. Qual è il livello di confidenza associato all intervallo [12.88%, 19.92%]?

Soluzione

La proporzione stimata è il punto centrale dell'intervallo cioè

$$\hat{p} = (12.88 + 19.92)/2 = 16.4\% = 0.164.$$

L'errore standard della stima è $ES = \sqrt{\hat{p}(1-\hat{})/600} = 0.0151164$.

$ Il margine di errore è

$$ME = 1/2 Ampiezza = (19.92 - 12.88)/2 = 3.52\%.$$

Poiché $ME = z_{\alpha/2}ES$ risulta che

$$z_{\alpha/2} = ME/ES = 0.0352/0.0151164 = 2.33.$$

Quindi $P(Z > 2.33) = \alpha/2$. Dalle tavole della normale $P(Z > 2.33) = 0.01$ e quindi $\alpha = 0.02$. Siccome il livello di confidenza è $1-\alpha$ questo risulta $1 - 0.02 = 0.98$.

8. Un imprenditore ha installato una nuova fornace energeticamente più efficiente. Si stima che in un anno la nuova fornace ridurrà il costo dell energia per un importo che può essere visto come una variabile casuale con media Euro 265 e deviazione standard di Euro 55. Sulla base di ipotesi che ritenete necessarie formulare, trovare la media e la deviazione standard della riduzione del costo totale dell energia in un periodo di 5 anni.

Soluzione

Riduzione del costo $= X$ con $E(X) = 265$ e $\sigma(X) = 55$. In 5 anni la riduzione del costo è $5X$ dove si suppone che la riduzione di costo sia costante ogni anno.

Quindi la media della riduzione del costo totale è $E(5X) = 5E(X) = 5 \cdot 265 = 1325$.

La varianza è $\sigma^2(5X) = 25\sigma^2(X) = 25 \cdot 55^2$, mentre la deviazione standard è $5 \cdot 55 = 275$.

9. Un gruppo locale di pianificazione dei trasporti è preoccupato per la mancanza di car-sharing per i pendolari. Temono che la proporzione di utilizzatori locali di car-sharing sia inferiore alla media nazionale, pari al 20%. Un indagine su 356 guidatori locali rileva che il 18.7% di loro usa il car-sharing. Quali sono le tue conclusioni, considerando un livello di significatività del 5%?

A) Non c'è evidenza che la proporzione di utilizzatori locali di car-sharing non sia inferiore alla media nazionale.
B) C'è evidenza che la proporzione di utilizzatori locali di car-sharing sia inferiore alla media nazionale.
C) Non c'è evidenza che la proporzione di utilizzatori locali di car-sharing sia inferiore alla media nazionale.

D) C'è evidenza che la proporzione di utilizzatori locali di car-sharing non sia inferiore alla media nazionale.

Soluzione

L'affermazione riguarda la proporzione p di utilizzatori e si può scrivere come $p < 0.2$. Questa ipotesi va contrapposta a $p \geq 0.2$ che avendo il segno di uguaglianza è l'ipotesi H_0:

$$H_0 : p \geq 0.2 \text{ contro } H_1 : p < 0.2$$

La statistica test è

$$z = (0.187 - 0.2)/\sqrt{0.2 \cdot 0.8/356} = -0.61.$$

La regione critica al livello del 5% è $Z < -1.645$ quindi non si rifiuta H_0 cioè non c'è evidenza sufficiente per rifiutare H_0.

Alcune domande contegono errori concettuali.

A) Non c'è evidenza che p non sia < 0.2. L'evidenza può essere solo contro H_0 non a favore di H_0.

B) C'è evidenza che $p < 0.2$, cioè contro H_0. Falso.

C) Non c'è evidenza che p < 0.2. Questo e' VERO perché infatti abbiamo accettato H_0.

D) C'è evidenza che p non sia < 0.2. Vedi A).

La risposta corretta è C).

10. Nella seguente distribuzione di frequenza, qual è la frequenza cumulata mancante nella posizione occupata dall'asterisco?

Modalità	Frequenza	Perc. cumulata
≤ 12.8	1	5
$12.8 \dashv 41.6$	5	30
$41.6 \dashv 70.4$	6	60
$70.4 \dashv 99.2$	6	$*$
> 99.2	2	100

A) 100%
B) 60%
C) 80%
D) 90%

Soluzione Se x è la percentuale cumulata che manca è chiaro che

$$x = (6 + 6 + 5 + 1)/(2 + 6 + 6 + 5 + 1) = 90\%$$

cioè D).

11. All'aumentare della dimensione del campione l'errore standard della media campionaria non cambia. Vero o falso?

Soluzione L'errore standard della media è $\sigma/\sqrt{n}$ quindi è chiaro che se aumenta la dimensione campionaria n l'ES diminuisce. Perciò l'affermazione è FALSA.

12. Supponiamo di lanciare due dadi. Si consideri la somma dei due dadi: sia A l' evento "si osserva un numero pari". Qual è il complementare dell'evento A?

A) $\{1, 3, 5, 7, 9, 11\}$
B) $\{3, 5, 7, 9, 11\}$
C) $\{8, 10, 11, 12\}$
D) $\{8, 9, 10, 11, 12\}$.

Soluzione La somma di due dadi è $\{2,3,4,5,6,7,8,9,10,11,12\}$. Se A è l'evento "esce un numero pari" il complementare è "esce un numero dispari". Quindi $\bar{A} = \{3,5,7,9,11\}$. Risposta B).

La risposta A) è sbagliata perché contiene 1 che è dispari ma non può capitare.

13. Il 30% delle famiglie ha un lettore DVD. Supponi di selezionare 20 famiglie casualmente. In media quante tra le 20 famiglie selezionate ti aspetti abbiano il lettore DVD? Qual è la deviazione standard del numero di famiglie con lettore DVD?

Soluzione $p = 0.3$, \$X = \$ numero di famiglie su 20 che hanno il DVD. X è Binomiale con indice $n = 20$ e $p = 0.3$. Perciò la media è $E(X) = np = 20 \cdot 0.3 = 6$ famiglie. La deviazione standard è $\sqrt{npq} = \sqrt{}\{6 \cdot 0.7\} =$ 2.04939.

14. Aumentando il livello di significatività di un test, la probabilità dell'errore del II tipo aumenta. Vero o falso?

Soluzione Il livello è $\alpha = P(I)$. è noto che all'aumentare di α $\beta = P(II)$ dimnuisce. Vero.

15. Il processo di rifinitura di mobili nuovi lascia piccole imperfezioni. La tabella seguente mostra la distribuzione di probabilità del numero di difetti stimata dall'azienda produttrice.

N. di difetti	0	1	2	3	4	5
Probabilità	0.34	0.25	0.19	0.11	0.07	

In media quanti difetti ci sono su un mobile?

A) 0.28
B) 0.85
C) 1.44
D) 0.77

Soluzioe La media di una variabile aleatoria si calcola come una media ponderata:

$$0 \cdot 0.34 + 1 \cdot 0.25 + 2 \cdot 0.19 + 3 \cdot 0.11 + 4 \cdot 0.07 + 5 \cdot 0.04 = 1.44,$$

risposta C).

16. Quale delle seguenti affermazioni è vera?

A) Il campo di variazione (range) è la misura di variabilità più utile
B) Le misure di variabilità sono numeri che descrivono il grado di dispersione dei valori osservati.
C) La media ponderata è la misura di variabilità più utile.
D) Tutte le affermazioni precedenti sono vere.

Soluzione A) FALSO, altre misure come la deviazione standard o lo scarto interquartile sono più usate B) VERO. Dispersione è sinonimo di variabilità. C) FALSO. La media non è una misura di variabilità, ma di posizione. D) Ovviamente è FALSO perché B) è VERA.

17. Un Internet café ha 4 computer. La probabilità che in un dato momento un computer sia libero è 0.4. Si assuma indipendenza. Qual è la probabilità che tutti i computer siano occupati?

Soluzione È come calcolare la probabilità di 4 successi su 4 prove indipendenti ciascuna delle quali ha probabilità di successo p = 0.6 (probabilità di essere occupato). Questa probabilità è

$$P(X = 4) = \binom{4}{4} 0.6^4 \cdot 0.4^0 = 1 \cdot 0.6^4 \cdot 1 = 0.1296.$$

18. Uno stimatore è una variabile casuale calcolata su un campione casuale che fornisce la stima puntuale per il parametro della popolazione. Vero o falso?

Soluzione È esattamente la definizione di uno stimatore che è una sintesi dei dati campionari che fornisce la stima puntuale di un parametro della popolazione. VERO.

19. Se X e Y sono due variabili casuali tali che $Y = a + bX$, allora $\mu_Y = a + b\mu_X$. Vero o falso?

Soluzione Se $Y = a + bXE(Y) = E(a + bX) = a + bE(X)$. È una proprietà del valore atteso. VERO.

20. Se $Z\tilde{N}(0, 1)$, allora $P(-1.25 < Z < -0.75)$ è:

A) 0.6678
B) 0.1210
C) 0.2266
D) 0.1056.

Soluzione

$P(-1.25 < Z < -0.75) = P(0.75 < Z < 1.25) = P(Z < 1.25) - P(Z < 0.75) = 0.89435 - 0.773373 = 0.1210.$

Risposta B).