Esami di Probabilità e Statistica

Fabio Ferrario
@fefabo

2021/2022

Indice

Capitolo 1

Esame di Settembre 2023

1.1 Domande chiuse

1 QUARTILI

Il terzo quartile dell'insieme di dati $\{5, -1, 4, 0, \frac{\sqrt{3}}{2}\}$ è:

$$= 4$$

Spiegazione: Per calcolare il quartile di un insieme lo devo ordinare (crescente) e poi usare la formula del formulario per il calcolo dei quartili.

2 COMBINAZIONI

Ho un'associazione con 50 soci. Devo scegliere 5 membri che compongano il comitato direttivo. Quante sono le possibili scelte?

$$\binom{50}{5} = \frac{50!}{45!5!}$$

Spiegazione: Questo è il perfetto esempio di **combinazione**: Una collezione *non ordinata* di k elementi distinti scelti tra n possibili. Il numero di combinazioni semplici è sul formulario.

3 P. EVENTI INDIPENDENTI

Siano A, B eventi indipendenti tali che $P(A) = \frac{1}{2}$ e $P(B) = \frac{1}{2}$. . Quale delle seguenti affermazioni è corretta?

(a) $P(A|B) = \frac{1}{4}$
(b) $P(A \cup B) = \frac{2}{3}$
(c) $P(A \cap B) = \frac{1}{2}$
(d) $P(A \cup B) = \frac{3}{4}$

d: $P(A \cup B) = \frac{3}{4}$

Spiegazione: Essendo indipendenti, sappiamo che $P(A|B) = P(A) = \frac{1}{2}$ e viceversa, quindi $P(B|A) = P(B) = \frac{1}{2}$.
Dalla **Regola del Prodotto**, sappiamo che $P(A \cap B) = P(A) \cdot P(B|A) = \frac{1}{4}$.
Dalle proprietà della probabilità sappiamo che $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{4}$, quindi la risposta è la **d**.

4 V.A.

Sia X una v.a. uniforme continua su $(-z, z)$, dove z è un numero reale. Quanto vale $E[X]$?

$$E[X] = 0$$

Spiegazione: Essendo X una v.a. uniforme continua su $(-z, z)$, dal formulario sappiamo che $E[X] = \frac{a+b}{2} = \frac{-z+z}{2} = 0$.

5

Siano $X_1, X_2, ..., X_{100}$ v.a. i.i.d. Bernoulliane di parametro $\frac{1}{2}$. Quale delle seguenti è la migliore approssimazione di

$$P(\frac{X_1 + X_2 + ... + X_{100} - 50}{5} < 1.12)$$

(Usare il Teorema Limite Centrale e le tavole delle distribuzioni notevoli)

(a) 0.5000
(b) 0.5886
(c) 0.8686
(d) 0.9998

0.8686

Spiegazione: Da capire. dalle tavole,$\phi(1.12) = 0.8686$.

6
In un test statistico per la verifica dell'ipotesi nulla $H_0 : \mu \geq 4$ contro l'alternativa $H_1 : \mu < 3$, si rifiuta H_0 a livello di significativià 5%. Allora posso sicuramente concludere che:

(a) Rifiuto l'ipotesi H_0 a livello di significatività del 10%
(b) Il $p-$value del test è pari a 0.05
(c) Rifiuto l'ipotesi H_0 a livello di significatività del 3%
(d) L'errore di seconda specie del test è pari a 0.05.

a: Rifiuto l'ipotesi H_0 a livello di significatività del 10%

Spiegazione: Non so

7
Quanto vale l'*ampiezza* di un intervallo bilatero di confidenza al 98% per la media μ di un campione $x_1, x_2, ..., x_n$ estratto da una popolazione normale con varianza campionaria s_n pari a 1? (Si ricordi che $P(t(n) > t_{n,\beta} = \beta$, dove $t(n)$ è una v.a. t-student a n gradi di libertà.))

$$\tfrac{2}{\sqrt{n}} t_{n-1,0.01}$$

Spiegazione: In questo caso stiamo stimando la media di una popolazione normale con la varianza incognita, perchè quella fornita è s_n. Se l'IC è al 98%, allora $\alpha = 0.02$. Dal formulario devo prendere i due estremi dell'intervallo di confidenza e per trovarne l'ampiezza devo sottrarre uno all'altro.

8
La statistica del test chi-quadrato di indipendenza per due variabili aleatorie X (che assume r valori) e Y (che assume s valori) ha legge:

Chi-quadrato a (r-1)(s-1) gradi di libertà

Spiegazione: Guardando dal formulario il test chi-quadrato di indipendenza, la regione critica avrà una legge: $\chi^2_{(r-1)(s-1),\alpha}$.

1.2 Esercizi

1 Un'urna contiene 20 palline colorate: 17 rosse e 3 bianche. Vengono estratte a caso tre palline, *con reimmissione*. Determinare la probabilità che:

1. La prima pallina estratta sia bianca;
2. Le prime due palline estratte siano entrambe bianche;
3. Almeno una delle tre palline estratte sia bianca.

Risposta: Definisco prima gli eventi: per $i = 1, 2, 3$, $B_i = \{$ Pallina bianca all'estrazione $i\}$. quindi:

1. Visto che ci sono 20 palline in totale, di cui 3 bianche e la probabilità è Uniforme:
$$P(B_1) = \frac{3}{20} = 15\%$$
2. Vogliamo calcolare l'intersezione ("and") di B_1 e B_2 che, siccome c'è reimmissione, sono indipendenti.
Quando due eventi sono indipendenti sappiamo che la probabilità dell'intersezione è il prodotto delle loro probabilità. quindi:
$$P(B_1 \cap B_2) = P(B_1)P(B_2) = (\frac{3}{20})^2 = 2.23\%$$
3. Siccome vogliamo calcolare un "or" vogliamo calcolare l'unione degli eventi 1,2,3. Andando a logica, la propbabilità che si verifichi almeno un B_i è il contrario della propbabilità che si verifichi l'evento $B_1^c \cap B_2^c \cap B_3^c$, ovvero il caso in cui non esca neanche una pallina bianca nelle tre estrazioni.
Calcoliamo quindi la probabilità di questo caso:
$$P(B_1^c \cap B_2^c \cap B_3^c) = P(B_1^c)P(B_2^c)P(B_3^c) = (\frac{17}{20})^3 = 61.41\%$$
E quindi:
$$P(B_1 \cup B_2 \cup B_3) = 1 - (\frac{17}{20})^3 = 38.59\%$$

2 Una moneta truccata restituisce testa con probabilità $\frac{1}{3}$. Si Lancia un dado equilibrato a 6 facce, se esce un numero pari si lancia la moneta 1 volta, altrimenti si lancia la moneta per 2 volte consecutive. Sia X la variabile aleatoria che indica il numero di teste osservate. (Si osservi che X può assumere i valori $0, 1, 2$.)

1. Se lanciando il dado esce un numero pari, qual'è la probabilità (condizionale) che non si osservi nessuna testa? Se invece lanciando il dado esce un numero dispari, qual'è la probabilità (condizionale) che non si osservi nessuna testa?
2. Calcolare $P(X = 0)$,
3. Determinare la densità discreta di X.

Risposta: Per covenienza Indico con D l'esito del lancio del dado. Quindi:

$$P(D \text{ è pari }) = \frac{1}{2}$$

1. Se lanciando il dado esce un numero pari, si lancia una sola volta la moneta, quindi la probabilità condizionale che non si osservi nessuna testa (evento $X = 0$) vale $\frac{2}{3}$.
   Se invece esce un numero dispari, si lancia due volte la moneta, quindi la probabilità condizionale che non si osservi nessuna testa ($X = 0$) vale: $\frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$. Riassumendo:

$$P(X = 0| \text{ D è pari }) = \frac{2}{3} \qquad P(X = 0| \text{ D è dispari }) = (\frac{2}{3})^2 = \frac{4}{9}$$

3

Esercizio 3. Il manager che si occupa del controllo di qualità di un'azienda che produce lampadine intende stimare la durata media delle lampadine facenti parte di un determinato lotto di produzione. Si sa che la deviazione standard della durata delle lampadine prodotte è pari a $\sigma = 100$ ore. Viene estratto un campione di $n = 64$ lampadine, che forniscono una durata media pari a $\bar{x}_n = 350$ ore. Rispondere alle seguenti domande:

(a) calcolare un intervallo di confidenza *sinistro* al livello di fiducia del 95% per la durata media delle lampadine;

(b) effettuare un opportuno test statistico al livello di significatività del 5% per verificare l'ipotesi che la durata media delle lampadine sia superiore a 400 ore;

(c) ripetere il test effettuato al punto (b) con un livello di significatività del 1%, stimare quindi il p-value $\bar{\alpha}$ del test (ossia trovare p tale che $\bar{\alpha} \leq p$) e commentare il risultato.

Risposta: Dobbiamo stimare la media con la deviazione standard (e quindi la varianza) note. Dal formulario l'intervallo di confidenza sinistro è dato da:

$$(-\infty, \bar{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}})$$

Nel nostro caso abbiamo che: $n = 64$, $\bar{x}_n = 350$, $\sigma = 100$. Dobbiamo trovare z_α:
Per trovare alpha, sappiamo che $100(1 - \alpha) = 95\% \implies \alpha = 5\% = 0.05$.

trovare $z_{0.05}$ significa trovare il valore di z t.c. $\phi(z) = 1 - \alpha = 0.95$. Dalle tavole dell distribuzione normale troviamo $z_{0.05}$ per interpolazione, infatti:

$\phi(1.64) = 0.9495$ e $\phi(1.65) = 0.9505$, quindi $z_{0.05} = 1.645$, cioè a metà trai due.

Sostituiamo nella fomula e troviamo che un intervallo di confidenza sinistro al 95% per μ è:

$$(-\infty, 350 + 1.645\frac{100}{8}) \simeq (-\infty, 370.56)$$

Risposta: Dobbiamo impostare un test con media incognita e varianza nota. Siccome ci chiediamo se $\mu > 400$, allora $H_0 : \mu \geq 400$ e $H_1 : \mu < 400$.
Dal formulario, la regione critica è $\frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n} < -z_\alpha$. Che significa che i dati consentono di rifiutare H_0 a livello di significatività α se questa disequazione è rispettata. In questo caso: $\mu_0 = 400$, $n = 64$ $\bar{x}_n = 350, \sigma = 100, \alpha = 0.05$. Sostituendo troviamo:

$$q < z_\alpha \rightarrow -4 < -1.645$$

Siccome questa disequazione è **vera** rifiuto $H_0 : \mu \geq 400$ a livello di significatività del 5%.

Osservazione: Il valore 400 non appartiene all'intervallo sinistro di confidenza al 95% per μ che abbiamo trovato nel punto precedente, quindi avremmo potuto rifiutare H_0 automaticamente.

Risposta: Ripetendo il test effettuato al punto precedente ma al 1%, troviamo $z_0.01 \simeq 2.325$ (z tale che $\phi(z) = 0.9900$). Dato che q non cambia abbiamo ancora che $-4 < -2.325$ e quindi rifiutiamo H_0 al livello 1%.

Per il p-value del test possiamo concludere che $\bar{\alpha} \geq 0.01$ e che i dati sono in constrasto significativo con H_0, ovvero c'è forte evidenza statistica che la durata media delle lampadine è < 400 ore.

Non stiamo a calcolare il p-value perchè dovremmo trovare un valore α tale per cui $z_\alpha = 4$, che sarebbe un valore estremamente piccolo, tanto che le nostre tavole vanno fino a $z_{0.0002} = 3.49$.