Compito scritto dell'esame di Statistica e analisi dei dati 30 giugno 2023	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri, appunti, cellulari.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Un segnale binario X emesso nella forma "1" con probabilitá $P(X = 1) = 0.75$, é inviato su un canale di trasmissione non simmetrico nel quale la probabilitá di errore nella trasmissione del segnale $X = 1$ é dell'8%. Il segnale X é ricevuto nella forma $Y = 1$ con probabilitá $P(Y = 1) = 0.70$. Calcolare:

(a) la probabilitá che il segnale "0" trasmesso sia ricevuto nella forma "1";

Soluzione: Dobbiamo calcolare

$$P(Y = 1 \mid X = 0)$$

Dalle condizioni specificate sappiamo che

$$P(X = 1) = 0.75 \implies P(X = 0) = 1 - P(X = 1) = 0.25$$
$$P(Y = 1) = 0.70 \implies P(Y = 0) = 1 - P(Y = 1) = 0.30$$

$$P(Y = 0 \mid X = 1) = 0.08 \implies P(Y = 1 \mid X = 1) = 1 - P(Y = 0 \mid X = 1) = 0.92$$

Per la legge delle probabilitá totali:

$$\begin{aligned} P(Y = 1) &= P(Y = 1 \mid X = 0)P(X = 0) + P(Y = 1 \mid X = 1)P(X = 1) \\ \implies 0.70 &= P(Y = 1 \mid X = 0) \times 0.25 + 0.92 \times 0.75 \\ \implies P(Y = 1 \mid X = 0) &= \frac{0.70 - 0.92 \times 0.75}{0.25} = 0.04 \end{aligned}$$

(b) la probabilitá totale di errore nella ricezione del segnale.

Soluzione: Dobbiamo calcolare

$$\begin{aligned} P(\texttt{errore}) &= P(Y = 0, X = 1) + P(Y = 1, X = 0) = P(Y = 0 \mid X = 1)P(X = 1) + P(Y = 1 \mid X = 0)P(X = 0) \\ &= 0.08 \times 0.75 + 0.04 \times 0.25 = 0.07 \end{aligned}$$

Esercizio 2. Sia data la funzione

$$f_X(x) = \begin{cases} |1 - x|, & per \quad 0 \leq x \leq a \\ 0 & altrove \end{cases}$$

dove a é un parametro incognito.

(a) Si determini il valore di a che rende $f_X(\cdot)$ la funzione di densitá di una VA X.

Soluzione: $|1-x| \geq 0$ sempre.
Si impone la condizione di normalizzazione

$$\int_0^a f_X(x) = 1 \tag{1}$$

Per $0 \leq x \leq 1$, $|1-x| = 1-x$, pertanto

$$\int_0^1 f_X(x)dx = \int_0^1 |1-x|\,dx = \int_0^1 (1-x)dx = \frac{1}{2} \tag{2}$$

Quindi (1) e (2) $\implies a > 1 \implies |1-x| = x-1$, per $1 \leq x \leq a$.
Allora

$$\int_0^a f_X(x) = \int_0^1 (1-x)dx + \int_1^a (x-1)dx = \frac{1}{2} + \left[\frac{x^2}{2}\right]_1^a = \frac{1}{2} + \frac{a^2}{2} - a - \frac{1}{2} + 1 = 1 \implies \frac{a^2}{2} - a = 0 \tag{3}$$

Le soluzioni sono $a = 0$ e $a = 2$, ma $a = 0$ rende nullo l'integrale (1).
In conclusione

$$a = 2$$

e la densitá si scrive come

$$f_X(x) = \begin{cases} |1-x|, & per \quad 0 \leq x \leq 2 \\ 0 & \qquad altrove \end{cases}$$

(b) Si calcoli il valore atteso di X

Soluzione:

$$E[X] \int_0^2 x f_X(x)dx = \int_0^2 x|1-x|\,dx = \int_0^1 x(1-x)dx + \int_1^2 x(x-1)dx = 1 \tag{4}$$

(c) Si ricavi l'espressione della funzione di ripartizione $F_X(x)$

Soluzione:

$$F_X(x) = \int_{-\infty}^x |1-t|\,dt$$

Per $0 \leq x < 1$:

$$F_X(x) = \int_0^x (1-t)dt = x - \frac{x^2}{2}$$

Per $1 \leq x < 2$

$$F_X(x) = \int_0^1 (1-t)dt + \int_1^x (t-1)dt = 1 - x + \frac{x^2}{2}$$

In conclusione:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x - \frac{x^2}{2} & 0 \leq x < 1 \\ 1 - x + \frac{x^2}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Esercizio 3. Sia $X \sim Poiss(x;\mu)$.

(a) Si derivino per esteso $E_{Pois}[X], var_{Pois}(X)$ nel limite della distribuzione Binomiale.

Soluzione: Vedere appunti di teoria

Esercizio 4.

Si sa che la probabilitá di errore in ricezione di una sequenza di 150 segnali trasmessi con modalitá statisticamente indipendenti é $p = 0.01$.

(a) Determinare la probabilitá che due dei segnali ricevuti siano errati.

Soluzione: Il numero di componenti difettosi segue una distribuzione binomiale. Ponendo i parametri $p = 0.003$ (probabilitá di errore) e $n = 150$

$$P(X = x) = Bin(x; n = 150, p = 0.01)$$

Per $X = 2$ errori:

$$Bin(2; n = 150, p = 0.01) = \binom{150}{2}(0.01)^2(0.99)^{148}$$

Poiché $p = 0.01$ é piccolo e $n = 150$ abbastanza grande, tali che $np = 1.5 \approx 1$, conviene utilizzare un'approssimazione con distribuzione di Poisson, $Poiss(x; \mu)$:

$$\mu = np = 1.5 \implies Poiss(2; 1.5) = \frac{(1.5)^2}{2} e^{-1.5} \simeq 0.251$$

Esercizio 5. Si vuole studiare la durata di un processo produttivo che dal materiale grezzo porta al prodotto finito. Il venditore del meccanismo di produzione sostiene che la durata del processo si distribuisce normalmente con media pari a 11 ore e deviazione standard pari a 4 ore. L'acquirente, sulla base delle valutazioni di un esperto, sospetta invece che, pur distribuendosi normalmente e con deviazione standard 4, la durata media del processo sia 14 ore. Si mettono allora in produzione 16 pezzi e si decide che il meccanismo di produzione verrá acquistato soltanto se la durata media della produzione nel campione é inferiore a 13.

(a) Si calcoli la probabilità di rischio dell'acquirente ovvero dell'errore del I tipo associato al criterio di decisione sopra riportato

Soluzione: La durata T é distribuita con legge normale $T \sim \mathcal{N}(\mu_0 = 11, \sigma^2 = 4^2)$.
Siccome, fino a prova contraria, non si puó sostenere che il venditore stia barando, il problema si puó mettere nella forma di un test di ipotesi unilaterale

$$\begin{cases} H_0 & : \mu = 11 = \mu_0 \\ H_1 & : \mu > 11 (\mu_1 = 14). \end{cases} \tag{5}$$

Dalla teoria dei test di ipotesi sappiamo che il rischio ovvero la probabilitá di errore di tipo I associato al criterio di decisione é

$$P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera}) = \alpha.$$

Considerato il criterio di decisione $\bar{T} < 13$, la regione di rifiuto di H_0 é $\bar{T} \geq 13$. Nell'ipotesi che H_0 sia vera, $\mu = \mu_0 = 11$; dunque:

$$\alpha = P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera}) = P(\bar{T} \geq 13 \mid \mu = \mu_0)$$

Standardizzando e usando le tavole della distribuzione normale standard:

$$\alpha = P(\bar{T} \geq 13 \mid \mu = \mu_0) = P(\frac{\bar{T} - \mu}{\sigma/\sqrt{n}} \geq \frac{13 - \mu}{\sigma/\sqrt{n}} \mid \mu = \mu_0) = P(\frac{\bar{T} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{13 - \mu_0}{\sigma/\sqrt{n}}) = P(Z \geq \frac{13 - 11}{4/\sqrt{16}}) = P(Z \geq 2) \simeq 0.0228.$$

(Analogamente, si poteva ragionare sul valore critico z_α che determina la regione critica $P(Z \geq z_\alpha)$ con

$$Z = \frac{\bar{T} - 11}{4/\sqrt{16}} \geq z_\alpha \quad \implies \quad \bar{T} \geq 11 + z_\alpha \frac{4}{\sqrt{16}} \tag{6}$$

sotto la condizione che l'acquirente accetta un rischio a patto che $\bar{T} < 13$, ovvero con valore critico determinato dall'equazione

$$11 + z_\alpha \frac{4}{\sqrt{16}} = 13; \tag{7}$$

dunque $z_\alpha = \frac{13-11}{4/\sqrt{16}} = 2$ come sopra)

(b) Si valuti la potenza del test associato al criterio di decisione sopra riportato

Soluzione:
La potenza del test é definita nella teoria dei test come complemento della probabilitá β dell'errore di tipo II

$$P = 1 - \beta = 1 - P(\texttt{non rifiuto } H_0 \mid H_0 \texttt{ falsa}) = P(\texttt{rifiuto } H_0 \mid H_1 \texttt{vera}).$$

Nell'ipotesi che H_1 sia vera, $\mu = \mu_1 = 14$. Pertanto

$$\beta = P(\bar{T} \leq 13 \mid \mu = \mu_1) = P(\frac{\bar{T} - \mu}{\sigma/\sqrt{n}} \leq \frac{13 - \mu}{\sigma/\sqrt{n}} \mid \mu = \mu_1) = P(\frac{\bar{T} - \mu_1}{\sigma/\sqrt{n}} \leq \frac{13 - \mu_1}{\sigma/\sqrt{n}}) = P(Z \leq \frac{13 - 14}{4/\sqrt{16}}) = P(Z \leq -1) \simeq 0.1587.$$

Dunque la potenza é notevole, pari a $P = 1 - \beta \simeq 84{,}13\%$

Esercizio 6. Un campione di 20 alberghi italiani presenta un numero di letti pari in media a 70, con deviazione standard pari a 25. Sapendo che gli alberghi contano in media 61 letti e assumendo una popolazione approssimativamente normale:

(a) verificare l'ipotesi che la media sia variata rispetto a quanto dichiarato, al livello di significativitá del 99%.

Soluzione:
Usiamo un test di ipotesi bilaterale

$$\begin{cases} H_0 & : \mu = \mu_0 = 61 \\ H_1 & : \mu \neq 61. \end{cases} \tag{8}$$

E' un test effettuato sulla media: distribuzione approssimativamente normale, varianza non nota, campione piccolo. Dunque si deve utilizzare la statistica test t:

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{70 - 61}{25/\sqrt{20}} = 1.61$$

La regola di decisione é: rifiuto H_0 se

$$|T| > t_{\frac{\alpha}{2}, n-1}.$$

Con $\nu = n - 1 = 19$gdl e $\alpha = 0.01$gdl

$$|T| > t_{0.005,19} = 2.861$$

Vediamo che $-2.861 < T = 1.61 < 2.861$
Non possiamo rigettare H_0, ovvero l'ipotesi che la media sia rimasta sostanzialmente invariata