Compito scritto dell'esame di Statistica e analisi dei dati 25 gennaio 2024	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 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 con un test d'ipotesi che la media sia variata rispetto a quanto dichiarato, al livello di significatività del 99% .

Soluzione:
Usiamo un test di ipotesi bilaterale

$$\begin{cases} H_0 & : \mu = \mu_0 = 61 \\ \\ H_1 & : \mu \neq 61. \end{cases} \tag{1}$$

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

Esercizio 2. Sia data la funzione

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

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

$$\int_0^a f_X(x) = 1 \tag{2}$$

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

$$\int_0^1 f_X(x)dx = \int_0^1 |1-x|\, dx = \int_0^1 (1-x)dx = \frac{1}{2} \tag{3}$$

Quindi (2) e (3) $\implies a > 1 \implies |1-x| = x-1$, per $1 \leq x \leq a$.
Allora

$$\int_0^a f_X(x) = \int_0^1 (1-x)dx + \int_1^a (x-1)dx = \frac{1}{2} + \left[\frac{x^2}{2}\right]_1^a = \frac{1}{2} + \frac{a^2}{2} - a - \frac{1}{2} + 1 = 1 \implies \frac{a^2}{2} - a = 0 \tag{4}$$

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

$$a = 2$$

e la densitá si scrive come

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

(b) Si calcoli il valore atteso di X

Soluzione:

$$E[X] \int_0^2 x f_X(x)dx = \int_0^2 x|1-x|\, dx = \int_0^1 x(1-x)dx + \int_1^2 x(x-1)dx = 1 \tag{5}$$

(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.

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 4.
Dovete partecipare ad un torneo di scacchi in cui giocate contro altri tre giocatori (una sola partita con ciascuno). Conoscete le probabilitá di vincere con ognuno di loro e potete scegliere l'ordine in cui incontrarli. Si vince il torneo se si vincono due partite consecutive.

(a) Volendo massimizzare la probabilitá di vittoria, si mostri che l'ordinamento ottimale é quello che vi fa incontrare alla seconda partita il giocatore piú debole dei tre, mentre l'ordine in cui incontrate i restanti due giocatori non ha rilevanza.

Soluzione:
Denotiamo con

$$P(\texttt{"vittoria contro il giocatore incontrato alla partita i -esima"}) = P(i) = p_i$$

con $i = 1{,}2{,}3$.
La descrizione del problema ci dice che si vince la partita se si gioca la seconda partita con il giocatore piú debole - dunque con $p_2 \geq p_1, p_2 \geq p_3$ - e, **congiuntamente**, si vince con l'uno **oppure** l'altro giocatore incontrati alla partita 1 o 3. Quindi, formalmente, la probabilitá $P(V)$ di vincere il torneo é

$$P(V) = P(2 \cap (1 \cup 3)) = P(2)P(1 \cup 3) = P(2)\left[P(1) + P(3) - P(1) \cap P(3)\right] =$$
$$= p_2(p_1 + p_3 - p_1p_3).$$

L'ordinamento é ottimale se tale probabilitá non é inferiore alle probabilitá calcolate utilizzando i due ordinamenti alternativi $P(1 \cap (2 \cup 3)) = p_1(p_2 + p_3 - p_2p_3)$ e $P(3 \cap (2 \cup 1)) = p_3(p_2 + p_1 - p_2p_1)$, ovvero:

$$p_2(p_1 + p_3 - p_1p_3) \geq p_1(p_2 + p_3 - p_2p_3), \tag{6}$$
$$p_2(p_1 + p_3 - p_1p_3) \geq p_3(p_2 + p_1 - p_2p_1). \tag{7}$$

Infatti, la prima disuguaglianza é equivalente alla condizione

$$p_2 \geq p_1,$$

mentre la seconda disuguaglianza si riduce a

$$p_2 \geq p_3.$$

Le due condizioni (6) e (7) ci dicono che la probabilitá di vincere alla seconda partita deve essere massima, ovvero: per massimizzare la probabilitá di vincere il torneo il secondo giocatore da incontrare deve essere il piú debole dei tre.

Esercizio 5. Un apparecchio elettronico é composto da due elementi in parallelo, l'uno indipendente dall'altro e ciascuno con un tempo di vita di media 8 giorni.

(a) Con quale probabilitá l'apparecchio durerá un tempo non superiore a 12 giorni, supposto che esso funzioni se una almeno delle due componenti funziona e sotto l'ipotesi che i guasti seguano legge poissoniana?

Soluzione: Il problema concerne un calcolo di tempo d'attesa ($T \leq 12$) La teoria dei processi poissoniani ci dice che se i guasti seguono la legge di Poisson, la durata T (in giorni), cioé il tempo di attesa al primo guasto, segue una legge esponenziale negativa $Exp(T; \lambda)$. Poiché una variabile aleatoria esponenziale ha media uguale all'inverso del parametro λ, nel nostro caso si ha

$$\lambda = \frac{1}{8};$$

Di conseguenza ciascuna componente ha un tempo di vita $T_i, i = 1{,}2$ avente densitá.

$$f_{T_i} = \begin{cases} 0, & t < 0 \\ \frac{1}{8}e^{-\frac{1}{8}t} & t \geq 0. \end{cases} \tag{8}$$

Indicato quindi con T il tempo di vita dell'apparecchio, si ha $T = \max\{T_1, T_2\}$. Sapendo poi che T_1 e T_2 sono indipendenti, si ha semplicemente che

$$P(T \leq t) = P(T_1 \leq t, T_2 \leq t) = P(T_1 \leq t)P(T_2 \leq t) = (1 - e^{-\frac{1}{8}t})^2$$

dunque

$$P(T \leq 12) = (1 - e^{-\frac{1}{8}\cdot 12})^2 \simeq (1 - 0.223)^2 \simeq 0.6035$$

Esercizio 6. Due tipi di soluzioni chimiche sono state provate per misurarne il pH. L'analisi di 6 campioni della prima soluzione ha mostrato un pH medio di 7.52 con deviazione standard campionaria di 0.032; su 5 campioni della seconda si sono misurati un pH medio di 7.49 e deviazione standard campionaria di 0.024. Mediante l'uso di intervalli di confidenza:

(a) Si verifichi se sia ragionevole assumere che le varianze delle due popolazioni sono uguali al livello di significatività 0.05

Soluzione:

Sotto l'ipotesi di distribuzione normale, consideriamo l'intervallo di confidenza

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2}}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)}$$

Questo è l'intervallo per cui $P(f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) < F < f_{\frac{\alpha}{2}}(\nu_1,\nu_2) = 1-\alpha$ con $\alpha = 0.05$ e dove F è la statistica $F = \frac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2}$

Semplicemente, nel caso in cui $\sigma_1^2 = \sigma_2^2$ è sufficiente verificare che

$$f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) < \frac{s_1^2}{s_2^2} < f_{\frac{\alpha}{2}}(\nu_1,\nu_2)$$

Sappiamo che $n_1 = 6, s_1 = 0.032, n_2 = 5, s_2 = 0.024$

Il rapporto delle varianze campionarie è

$$\frac{s_1^2}{s_2^2} = \frac{0.032^2}{0.024^2} = 1.78$$

I g.d.l sono $\nu_1 = n_1 - 1 = 5$, $\nu_2 = n_2 - 1 = 4$, dunque, dalle tavole

$$f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) = f_{0.975}(5{,}4) = \frac{1}{f_{0.025}(4{,}5)} = \frac{1}{7.39} = 0.14$$

$$f_{\frac{\alpha}{2}}(\nu_1,\nu_2) = f_{0.025}(5{,}4) = 9.36$$

Non possiamo dunque rigettare l'ipotesi che $\sigma_1^2 = \sigma_2^2$.

(b) Si stabilisca se i valori medi di pH delle due soluzioni possano essere ritenuti uguali con livello di significatività 0.05

Soluzione:

Abbiamo che $\bar{x}_1 = 7.51$, $\bar{x}_2 = 7.49$.

Sempre nell'ipotesi normale, le varianze della popolazione sono incognite, ma al punto precedente abbiamo inferito che plausibilmente sono uguali.

L'intervallo di confidenza per la differenza fra le medie della popolazione é dunque

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Nell'ipotesi che $\mu_1 = \mu_2$ avremo che per la statistica

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

deve valere

$$-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}$$

I gdl della distribuzione sono

$$\nu = n_1 + n_2 - 2 = 9$$

La varianza pooled é

$$s_P^2 = \frac{5 \cdot 0.032^2 + 4 \cdot 0.024^2}{6+5-2} = 0.000825$$

Quindi

$$T = \frac{7.52 - 7.49}{\sqrt{0.000825(\frac{1}{6}+\frac{1}{5})}} = 1.72$$

Per $\alpha = 0.05$, $t_{\frac{\alpha}{2}} = t_{0.025} = 2.262$, dunque

$$-2.262 < T < 2.262$$

e concludiamo che la differenza fra le due medie non é statisticamente significativa al livello di significatività 0.05.