Compito scritto dell'esame di Statistica e analisi dei dati 15 luglio 2021	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 e appunti.
- 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. Una recente indagine ha rivelato che il 14% delle segretarie ha dolore al polso. Inoltre, il 6% delle segretarie intervistate ha dolore al polso e al tempo stesso assume regolarmente un farmaco antinfiammatorio.

(a) Qual é la probabilitá che una segretaria che ha dolore al polso assuma regolarmente un farmaco antinfiammatorio?

Soluzione: Definiamo gli eventi seguenti:

$$A = \text{"la segretaria ha il dolore al polso"};$$

$$B = \text{"la segretaria assume il farmaco"}.$$

$$P(A) = 0.14, P(A,B) = 0.06.$$

La probabilitá richiesta é la condizionata:

$$P(B \mid A) = \frac{P(B,A)}{P(A)} = 0.428$$

Esercizio 2.

Un gioco consiste nel lanciare una moneta e nel generare un numero casuale Y tra zero e uno. La vincita é $2Y$ se compare testa e 1 se compare croce.

(a) Calcolare la vincita attesa.

Soluzione: (Valore atteso) Sia T l'evento: "compare testa" definito su un classico spazio di probabilitá $(S, \mathcal{F}, P)$, $S = \{T,C\}$ e sia V la vincita

Usiamo la classica variabile aleatoria indicatrice

$$X_1(\omega) = I_T(\omega) = \begin{cases} 1 & \text{se } \omega \in T \\ \\ 0 & \text{se } \omega \notin T (\omega \in \sim T) \end{cases}. \tag{1}$$

e la VA indicatrice speculare

$$X_2(\omega) = I_{\sim T}(\omega) = \begin{cases} 0 & \text{se } \omega \in T \\ 1 & \text{se } \omega \notin T (\omega \in \sim T) \end{cases}. \tag{2}$$

Allora:

$$V = 2Y I_T - I_{\sim T}$$

Poiché X e I_T sono v.a. indipendenti

$$E[V] = E[2Y I_T - I_{\sim T}] = E[2Y I_T] - E[I_{\sim T}] = 2E[Y]E[I_T] - E[I_{\sim T}]$$

Il valore atteso della VA indicatrice X_1 si calcola facilmente:

$$E[X_1] = 0 \cdot P_X(X_1 = 0) + 1 \cdot P_X(X_1 = 1) = 0 \cdot (1 - P(T)) + 1 \cdot P(T) = P(T) = \frac{1}{2},$$

ovvero, il valor medio della funzione indicatrice é la probabilitá di successo. Analogamente, per $I_{\sim A}$.
Sappiamo che il valor atteso di una VA distribuita uniformemente, $Y \sim Unif(a,b)$ é

$$E[Y] = \frac{b+a}{2}$$

dunque se $Y \sim Unif(0,1)$, $E[Y] = \frac{1}{2}$
Quindi:

$$E[V] = 2 \times \frac{1}{2} \times \frac{1}{2} - \frac{1}{2} = 0$$

Esercizio 3. Il responsabile dell'illuminazione di un palazzo ha a disposizione 3 lampadine di riserva. Ciascuna delle lampadine utilizzate dura in media 200 ore e il responsabile deve aspettare 24 ore perché gli vengano consegnate nuove lampadine di riserva.

(a) Qual é la probabilitá che il responsabile non sia in grado di sostituire le lampadine fulminate?

Soluzione: Il problema sorge se si fulminano piú di tre lampadine entro le 24 ore: in tal caso il responsabile non sará in grado di sostituirle. Dobbiamo dunque determinare la probabilitá congiunta

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}).$$

dove X_1,X_2,X_3 sono i tempi di attesa della rottura della prima, seconda e terza lampadina.
Si assuma una distribuzione esponenziale negativa dei tempi di attesa e si ragioni, per semplicitá di calcolo in termini di centinaia di ore.
Il tempo medio é pari a $E[X] = 2$ (in centinaia di ore). Il parametro della distribuzione esponenziale é dunque

$$\lambda = \frac{1}{2} = 0.5$$

La probabilitá del tempo di rottura di una singola lampadina entro le 24 ore é

$$P(X \leq 0.24) = 1 - e^{-\lambda \cdot x} = 1 - e^{-0.5 \cdot 0.24} = 1 - 0.887 = 0.113$$

Poiché la rottura di ciascuna lampadina é indipendente dalle altre:

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}) = P(\{X_1 \leq 0.24\}) \cdot P(\{X_2 \leq 0.24\}) \cdot P(\{X_3 \leq 0.24\}) = (0.113)^3 = 0.0014$$

Esercizio 4. Una variabile aleatoria X ha una distribuzione con media 250 e deviazione standard 20.

(a) Determinare le probabilitá :

- $P(210 < X < 290)$
- $P(220 < X < 280)$

Soluzione: La distribuzione di X é ignota e si usa la disuguaglianza di Chebychev nella forma

$$P_X(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

con $\mu = 250, \sigma = 20$.
Entrambi gli intervalli hanno come punto centrale la media della distribuzione. Per esempio, nel primo caso il punto centrale vale $(210 + 290)/2 = 250$
Si trova il raggio dell'intorno di 250 che é $(290 - 210)/2 = 40$. Quindi l'intervallo é 250 ± 40
Si esprime il raggio come un multiplo della deviazione standard:

$$40 = k \times 20$$

da cui $k = 2$ (il raggio 40 é il doppio della deviazione standard).
Usando Chebychev:

$$P(210 < X < 290) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{4} = 0.75$$

Analogamente:

$$P(220 < X < 280) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{1.5^2} = 0.555$$

(b) Determinare le stesse probabilitá sapendo che X segue una legge normale $X \sim \mathcal{N}(\mu, \sigma)$

Soluzione: Usando le tavole della Gaussiana standard:

$$P(210 < X < 290) = P(-2 < Z < 2) = 2(0.9772) - 1 = 0.9544$$

$$P(220 < X < 280) = P(-1.5 < Z < 1.5) = 2(0.9332) - 1 = 0.8664$$

Esercizio 5. 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

(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.