Compito scritto dell'esame di Statistica e analisi dei dati 12 luglio 2022	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.

Un sistema é formato da tre componenti disposti in parallelo. Il sistema funziona se almeno due dei tre componenti funzionano correttamente. Indicati con C_1, C_2, C_3 gli eventi "funzionamento corretto" del primo, secondo e terzo componente, rispettivamente, siano $P(C_1) = P(C_2) = P(C_3) = p$ con $p = 0.7$. Si assuma che gli eventi C_1, C_2, C_3 siano indipendenti tra di loro.

(a) Si mostri che la probabilitá che il sistema funzioni correttamente é maggiore della probabilitá di funzionamento dei singoli componenti.

Soluzione:
Definiamo l'evento $C =$ `"il sistema funziona correttamente"`. Il problema si risolve calcolando la probabilitá $P(C)$ e mostrando che $P(C) > p$.
Il sistema funziona se almeno due componenti funzionano correttamente. I casi possibili che si possono presentare sono 8, di questi, quelli in cui almeno due generatori sono attivi sono 4: il caso in cui funzionano tutti e tre, e i tre casi in cui si ha il malfunzionamente di uno solo. Dunque

$$C = \{C_1 \cap C_2 \cap C_3\} \cup \{\sim C_1 \cap C_2 \cap C_3\} \cup \{C_1 \cap \sim C_2 \cap C_3\} \cup \{C_1 \cap C_2 \cap \sim C_3\}.$$

Applicando il Principio di additivitá generalizzata, la probabilitá di corretto funzionamento del sistema é

$$P(C) = P(\{C_1 \cap C_2 \cap C_3\}) + P(\{\sim C_1 \cap C_2 \cap C_3\}) + P(\{C_1 \cap \sim C_2 \cap C_3\}) + P(\{C_1 \cap C_2 \cap \sim C_3\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = p^3 + 3p^2(1-p) = 3p^2 - 2p^3 = 1.47 - 0.686 = 0.784 > p$$

Esercizio 2.

In un ufficio le pratiche relative ad una certa procedura amministrativa vengono affidate casualmente a tre impiegati che indicheremo con A, B e C. Statisticamente, la percentuale dei casi in cui la pratica viene completata entro una settimana da ciascun impiegato é riportata nella seguente tabella:

Impiegato	A	B	C
Perc.	40%	80%	30%

(a) Avendo ricevuto una pratica espletata entro una settimana, qual é secondo voi l'impiegato a cui era stata affidata e qual é la probabilitá della vostra conclusione?

Soluzione: Definiamo i seguenti eventi

- $S =$ "pratica completata entro una settimana"
- $A =$ "pratica affidata all'impiegato A"
- $B =$ "pratica affidata all'impiegato B"
- $C =$ "pratica affidata all'impiegato C"

L'affidamento pratiche é casuale dunque:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Dalla tabella ricaviamo che:

$$P(S \mid A) = 0.4, P(S \mid B) = 0.8, P(S \mid C) = 0.3$$

Per rispondere al quesito si tratta di stabilire qual é la maggiore fra le probabilitá a posteriori

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

Il denominatore, cioé la probabilitá che una pratica venga completata entro una settimana indipendentemente da quale impegato l'ha seguita é:

$$P(S) = P(S \mid A)P(A) + P(S \mid B)P(B) + P(S \mid C)P(C) = \frac{1}{3}(0.4 + 0.8 + 0.3) = \frac{1}{3} \times 1.5$$

Sostituendo tutti i dati:

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)} = 0.267$$

$$P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)} = 0.533$$

$$P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)} = 0.2$$

Pertanto, sará stato l'impiegato B a seguire la pratica con probabilitá 0.533.

Esercizio 3. L'altezza di un uomo segue una distribuzione normale di media 178 cm con una deviazione dalla media di 8 cm. L'altezza di una donna ha invece una media di 165 cm con una deviazione pari a 7 cm. .

(a) Qual é, nella stessa popolazione, la proporzione di uomini la cui altezza é superiore a 185 cm?

Soluzione: Sia U la variabile aleatoria che denota l'altezza degli uomini.
Vogliamo determinare $P(U > 185)$, dove la distribuziione é Normale con $\mu = 178, \sigma = 8$. Standardizzando le variabili e usando le tabelle della distribuzione normale

$$P(U > 185) = P(\frac{U - 178}{8} > \frac{185 - 178}{8}) = P(Z < 0.875) \approx 0.19$$

(b) Qual é la proporzione di donne che sono piú alte della metá degli uomini?

Soluzione: Sia D la variabile aleatoria che denota l'altezza delle donne.
Sappiamo che media e mediana nella distribuzione normale coincidono dunque

$$P(U > u) = 0.5$$

quando $u = \mu = 178$
Pertanto, si tratta di determinare $P(D > 178)$:

$$P(D > 178) = P(Z > \frac{178 - 165}{7}) = P(Z > 1.857) \approx 0.032$$

.
Solo il 3.2% delle donne é piú alto di metá degli uomini.

Esercizio 4.
Un vostro amico asserisce di aver ottenuto una media di 3.25 punti per lancio su 1000 lanci di un dado non truccato.

(a) Qual é la probabilitá che stia mentendo?

Soluzione: (Teorema limite centrale)
Sia X_i la variabile aleatoria (v.a.) che denota l'esito dell'i-simo lancio di dado.
Sappiamo che le v.a. $X_1, X_2, \cdots, X_{1000}$ sono indipendenti e identicamente distribuite con distribuzione uniforme sull'insieme di interi $\{1,2,\cdots,6\}$ da cui é facile ricavare il valore atteso e la deviazione standard

$$E[X_i] = 3.5$$
$$\sigma_{X_i} = \sqrt{E[X_i^2] - E[X_i]^2} \approx 1.7078$$

In breve, il Teorema del limite centrale ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite di valore atteso μ e deviazione standard σ vale

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \tag{1}$$

dove $S_n = \sum_{i=1}^{n} X_i$. Nella fattispecie $n = 1000$ e S_n é il numero di punti totalizzati in 1000 lanci.
Sotto tali ipotesi, la probabilitá che il vostro amico affermi il vero si riduce a calcolare la probabilitá definita nel T.L.C riscritta come

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x) \tag{2}$$

dove $\frac{S_n}{n} = 3.25 = \overline{X}$ é la media campionaria del campione di taglia $n = 1000$ dichiarata dal vostro amico. In buona sostanza il T.L.C. il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ. Pertanto,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{3.25 - 3.5}{1.7078/\sqrt{1000}} = -4.629$$

La probabilitá che il vostro amico dichiari il vero é dunque pari a $\Phi(-4.629)$
Anche senza consultare le tabelle e tenendo presente che per la normale standard $\sigma = 1$ e dunque $z = z\sigma$, significa che stiamo valutando la probabilitá di un valore di media dichiarata che é di 4.629 deviazioni standard al di sotto della media $\mu = 3.5$.
Considerando la legge 3σ, é immediatamente evidente che la probabilitá $\Phi(-4.629)$ é molto bassa e pertanto la probabilitá che che il vostro amico dica il falso molto elevata $(1 - \Phi(-4.629))$.
Se poi, per scrupolo, si consultasse la tabella della normale ridotta, si leggerebbe che $\Phi(-4.629) = 1.84 \times 10^{-7}$

Esercizio 5.
La quantitá di stoffa per produrre poltrone segue una legge normale. Su un campione casuale di 15 poltrone, si é riscontrato che l'ammontare medio del materiale é $912cm^2$, con una deviazione standard di $64cm^2$.

(a) Qual é l'intervallo di confidenza al 99% per la media della quantitá di materiale?

Soluzione:
Sappiamo che $n = 15, \bar{x} = 912cm^2, s = 64cm^2$. Il campione di taglia $n = 15$ é piccolo e l'intervallo é basato sulla t di Student con $\nu = n - 1 = 14$ gradi di libertá.
Con $\alpha = 0.01$, si ricava dalla tabella della t che il quantile é $t_\alpha = t_{.005} = 2.977$.
L'errore standard é $SE = \frac{\bar{x}}{\sqrt{n}} = \frac{912}{\sqrt{15}} = 16.52473$
L' intervallo di confidenza al 99% per la media é dunque

$$\bar{x} - 2.977 \cdot SE < \mu < \bar{x} + 2.977 \cdot SE,$$

ovvero

$$\mu = 912 \pm 49.2cm^2$$

Esercizio 6. Il termostato di un condizionatore d'aria é supposto essere tarato sul valore di soglia $\mu = 25°C$ (cioé l'apparecchio entra in funzione quando la temperatura dell'ambiente supera i $25°C$). Un tecnico misura in 8 occasioni diverse la temperatura X alla quale il termostato scatta, ottenendo i valori seguenti ($°C$):

24.6 24.8 25.2 25.4 25.5 24.0 24.7 25.3

(a) Formulando un adeguato modello per X, stabilire se, al 95% di confidenza, il valore di μ sia plausibile statisticamente

Soluzione:
Assumendo una distribuzione normale della temperatura, $X \sim \mathcal{N}(\mu, \sigma^2)$, poiché la varianza σ^2 é incognita, la stima intervallare di μ al 95% di confidenza si ottiene considerando

$$P\left(\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 1 - \alpha = 0.95$$

Poiché al 95% di confidenza $\alpha = 0.05$, allora $t_{\frac{\alpha}{2}} = t_{0.025}$, da determinarsi dalla distribuzione di Student con $\nu = n - 1 = 7$ gdl:

$$t_{0.025} = 2.365$$

Dai dati conosciuti si ottengono media e deviazione standard campionaria:

$$\bar{x} \approx 24.93$$

$$s \approx 0.83$$

Pertanto

$$24.93 - 2.365 \times \frac{0.83}{2.64} < \mu < 24.93 + 2.365 \times \frac{0.83}{2.64}$$

ovvero $24.2 < \mu < 25.66$ e dunque il valore $\mu = 25°C$ puó essere ritenuto plausibile.

(b) Stabilire quanto vale al massimo la precisione p (ovvero il reciproco della deviazione standard) del termostato al 95% di confidenza

Soluzione:
Assumendo $\mu = 25°C$ per l'analisi precedente, $X \sim \mathcal{N}(25, \sigma^2)$.
Sappiamo che $p = \frac{1}{\sigma}$, dunque il valore di precisione massima é quello di deviazione standard minima, $p_{max} = \frac{1}{\sigma_{min}}$.
Il problema richiede di determinare

$$P(p < p_{max}) = 1 - \alpha = 0.95$$

Dalla definizione di p questo equivale alla stima unilaterale sul limite inferiore della varianza (σ^2_{min}) :

$$P(p < p_{max}) = P\left(\sqrt{\frac{(n-1)s^2}{\chi^2_\alpha}} < \sigma\right) = 1 - \alpha$$

Ne deriva che al 95% di confidenza il limite superiore di precisione vale $p_{max} = \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}}$, dunque, leggendo dalla tabella della distribuzione Chi-quadro, per $\nu = 7$, $\chi^2_{0.05}(7) = 14.067$ e usando s^2 calcolato precedentemente.

$$p < \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}} \approx 1.69°C$$