Compito scritto dell'esame di Statistica e analisi dei dati 14 giugno 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. Di due eventi A e B si sa che $P(A \cup B) = \alpha$ e che $P(A \mid B) = P(B \mid A) = p > 0$.

(a) Calcolare $P(A), P(B), P(A \cap B)$

Soluzione: Si ha che

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \implies pP(A) = pP(B) \implies P(A) = P(B)$$

Dunque

$$\alpha = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2P(A) - P(A \cap B) = 2P(A) - P(B \mid A)P(A) = (2-p)P(A)$$

$$\implies P(A) = \frac{\alpha}{2-p} = P(B)$$

Infine

$$P(A \cap B) = P(B \mid A)P(A) = \frac{\alpha p}{2-p}$$

Esercizio 2. Una variabile aleatoria discreta X puó assumere solo i valori $x = 1, 2, 3$. Inoltre, si ha che $E[X] = 1.5, var(X) = 0.5$.

(a) Calcolare la distribuzione di probabilitá di X.

Soluzione:
La la distribuzione di probabilitá discreta é

$$P(X=1) = p_1, P(X=2) = p_2, P(X=3) = p_3$$

con condizione di normalizzazione

$$p_1 + p_2 + p_3 = 1 \tag{1}$$

Sappiamo che:

$$E[X] = \sum_{x=1}^{3} xP(X=x) = 1p_1 + 2p_2 + 3p_3 = 1.5 \tag{2}$$

Inoltre $var(X) = E\left[X^2\right] - (E[X])^2 = E\left[X^2\right] - (1.5)^2 = 0.5$, con $E\left[X^2\right] = \sum_{x=1}^{3} x^2 P(X=x) = 1p_1 + 4p_2 + 9p_3$, dunque

$$1p_1 + 4p_2 + 9p_3 - (1.5)^2 = 0.5 \tag{3}$$

Risolvendo il sistema lineare di equazioni 1, 2, 3, si ottiene

$$p_1 = 0.625, p_2 = 0.250, p_3 = 0.125$$

ESERCIZIO 3.
Sparando a un bersaglio ho il 20% di probabilitá di fare centro.

(a) Se sparo due volte, che probabilitá ho di fare almeno un centro?

Soluzione: (Distribuzione binomiale) Il numero X_n di centri fatti con n colpi segue una legge binomiale, $X_n \sim Bin(X_n; n,0.2)$.
Quindi
$$P(X_2 \geq 1) = 1 - P(X_2 = 0) = 1 - Bin(0; 2,0.2) = 1 - (0.8)^2 = 0.36$$

(b) Che numero minimo di colpi devo sparare per aver piú del 50% di probabilitá di fare almeno un centro?

Soluzione: Si deve determinare
$$P(X_n \geq 1) > 0.5$$
Procedendo come prima
$$P(X_n \geq 1) = 1 - P(X_n = 0) = 1 - Bin(0; n,0.2) = 1 - (0.8)^n$$
dunque
$$1 - (0.8)^n > 5 \implies (0.8)^n < 0.5 \implies n > \frac{\ln 0.5}{\ln 0.8} \approx 3.1$$

ESERCIZIO 4. Un transistor ha una vita media nominale di 1000 giorni.

(a) Se in un test di durata il transistore ha giá funzionato ininterrottamente per 500 giorni, qual é la probabilit'a che si guasti nei successivi 200 giorni?

Soluzione: Assumendo che i guasti seguano un processo poissoniano, la durata T (in giorni) segue una legge esponenziale $Exp(T; \lambda)$ di parametro
$$\lambda = \frac{1}{10^3};$$
Per l'assenza di memoria che caratterizza tale distribuzione, la probabilitá che un transistor vecchio di 500 giorni duri ancora al massimo 200 giorni é uguale alla probabilitá che un transistor nuovo duri al massimo 200 giorni, Pertanto
$$P(T \leq 200) = 1 - e^{-\frac{200}{10^3}} \approx 18.13\%$$

(b) Se 5 transistor sono sottoposti al test di durata, qual é la probabilitá che almeno uno di essi funzioni piú di 2000 giorni? (Si assuma che i transistor si guastino indipendentemente.)

Soluzione: Sia p la probabilit'a che un transistore funzioni al massimo 2000 giorni
$$p = P(T \leq 200) = 1 - e^{-\frac{2000}{10^3}} = 1 - e^{-2}$$
Per indipendenza, la probabilitá che tutti i 5 transistor funzionino al massimo 2000 giorni é uguale a p^5, e quindi la probabilitá dell'evento contrario (almeno un transistor funziona piú di 2000 giorni), vale
$$1 - p^5 \approx 51.67\%$$

ESERCIZIO 5. Un costruttore sta considerando l'acquisto di speciali barre metalliche da due diversi fornitori. Un campione di 12 barre di lunghezza dichiarata pari a $127mm$ viene acquistato da ciascuno dei due fornitori e poi misurato. La deviazione standard della lunghezza delle barre del primo fornitore risulta essere $s_1 = 0.13mm$, mentre quella delle barre del secondo fornitore é di $s_2 = 0.17mm$.

(a) Questi dati indicano che la lunghezza di una barra del primo fornitore é soggetta a maggior variabilitá rispetto a quella del secondo fornitore? (Si assuma normalitá e un livello di significativitá 0.05)

Soluzione:
I dati del problema ci dicono che: $n_1 = 12$, $n_2 = 12$, $s_1^2 = (0.13)^2$, $s_2^2 = (0.17)^2$.
I gdl sono $\nu_1 = \nu_2 = 11$
Considerando la statistica $F = \frac{s_1^2}{s_2^2} = 0.585$ (nell'ipotesi $\frac{\sigma_1^2}{\sigma_2^2} = 1$, notiamo immediatamente che é largamente inferiore al valore critico $f_{0.025}(11, 11) = 3.53$ valore oltre il quale la probabilitá che i campioni siano stati generati da distribuzioni con $\sigma_1 = \sigma_2$ (ipotesi nulla) sarebbe molto bassa.
Verifichiamo con il calcolo dell'intervallo di confidenza (IC)
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:
$$\frac{s_1^2}{s_2^2}\frac{1}{f_{0.025}(\nu_1, \nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2, \nu_1)$$

dove $f_{0.025}(\nu_2, \nu_1) = f_{0.025}(\nu_1, \nu_2) = f_{0.025}(11, 11) = 3.53$
Sostituendo:

$$0.16565 < \frac{\sigma_1^2}{\sigma_2^2} < 2.0642$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1}{\sigma_2} = 1$, dunque i dati indicano che statisticamente non si puó rigettare l'ipotesi che le lunghezze delle barre dei due fornitori abbiano la stessa variabilitá.

Esercizio 6.

Vengono sottoposti a confronto i consumi delle autovetture A e B alla velocitá costante di $120Km/h$. Si ritiene che i consumi dei due tipi di autovetture possa essere descritto da variabili aleatorie con distribuzione normale con la stessa varianza. L'auto di tipo A in 20 prove consuma mediamente $6.5 litri/100Km$, quella di tipo B in 22 prove consuma mediamente $6.6 litri/100Km$. Le relative varianze campionarie sono rispettivamente di 0.30 e 0.28.

(a) Possiamo ritenere che le due autovetture abbiano lo stesso consumo medio al livello di significativitá del 5%?

Soluzione:
Abbiamo che:

- $\bar{x}_A = 6.5, s_A^2 = 0.30, n_A = 20$
- $\bar{x}_B = 6.6, s_B^2 = 0.28, n_B = 22$

Siamo nel caso di varianze ignote ma eguali, e la differenza fra medie é valutabile mediante una statistica t-Student, dove i gradi di libertá ν sono dati da $\nu = n_A + n_B - 2 = 40$, $\alpha = 0.05$, per cui, $t_{\nu,\frac{\alpha}{2}} = t_{40,0.025} \approx 2.021$
La deviazione standard pooled vale:

$$S_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} = 0.5381$$

L'intervallo di confidenza per la differenza fra le medie $\mu_A - \mu_B$ a livello $(100)(1 - \alpha) = (100)(1 - 0.05) = 95\%$ vale pertanto

$$\bar{x}_A - \bar{x}_B \pm t_{40,0.025} S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -0.1 \pm 0.33586$$

ovvero l'IC ha estremi $[-0.43596; 0.23596]$ per cui si puó ragionevolmente ritenere che le due autovetture abbiano differenza in consumo medio trascurabile.
Equivalentemente, si osserva che la statistica della differenza fra medie normalizzata

$$\left| \frac{\bar{x}_A - \bar{x}_B}{S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \right| = 0.612 < t_{40,0.025} = 2.021$$