

Tema d'esame di Statistica e analisi dei dati

Prova scritta del 16 settembre 2019

Esercizio 0

Dato un evento C , indichiamo con $P(C)$ la probabilità che esso si verifichi.

- 1. Dati due eventi A e B , esprimete la probabilità condizionata $P(A|B)$ in funzione delle opportune probabilità marginale e congiunta.
- 2. Esprimete la probabilità di A in funzione delle probabilità condizionate $P(A|B)$, $P(A|\bar{B})$ e di opportune probabilità marginali.

Esercizio 1

Sia X una variabile casuale geometrica di parametro p , i cui punti di massa sono $\{0, 1, 2, \dots\}$.

- 1. Esprimete, in funzione di p :
 - 1.1. la massa di probabilità f_X ,
 - 1.2. il valore atteso di X ,
 - 1.3. la varianza di X .
- 2. Fissato, solo in questo punto, $p = 0.6$, tracciate sul foglio il grafico della massa di probabilità di X , evidenziando su tale grafico tutte le informazioni che ritenete importanti.

Esercizio 2

Sia Y una variabile casuale normale di parametri μ e σ^2 .

- 1. Indicate con F_Y la funzione di ripartizione di Y . Quanto vale $F_Y(\mu)$? Giustificate la risposta.
- 2. Determinate un valore $r > 0$ tale che $P(|Y - \mu| \leq r \cdot \sigma) = 0.5$.
- 3. Fissati, solo in questo punto, $\mu = 26$ e $\sigma^2 = 25$, tracciate sul foglio i grafici della densità di probabilità f_Y e della funzione di ripartizione F_Y nel modo più preciso possibile, evidenziando in entrambi i grafici il valore μ .

Esercizio 3

Sia $\bar{Y}_{(n)}$ la media campionaria di un campione casuale Y_1, \dots, Y_n estratto da una popolazione normale Y di valore atteso μ e varianza $\sigma^2 = 22$.

1. Fissati i valori $\epsilon = 0.5$ e $\alpha = 0.9$, determinate una condizione sufficiente per n affinché sia maggiore di α la probabilità $P(|\bar{Y}_{(n)} - \mu| \leq \epsilon)$.

- ✓ Proponete uno stimatore, chiamiamolo T_n , di μ .
- ✓ Lo stimatore che avete proposto al punto precedente è non distorto? Giustificate la risposta.

Esercizio 4

Collegatevi al sito `upload.di.unimi.it`, selezionate l'esame di *Statistica e analisi dei dati* per l'appello odierno e scaricate il file `acquisti.csv`. Questo file contiene, tra le altre, le seguenti informazioni raccolte da un sito di *e-commerce* relativamente ai propri clienti:

- *ID*: identificatore univoco;
- *Sesso*: 0 se il cliente è un uomo, 1 se è una donna;
- *Anni*: età del cliente, espressa in anni;
- *Altezza*: altezza del cliente, espressa in metri;
- *Peso*: peso del cliente, espresso in chilogrammi;
- *BMI*: *Body Mass Index*, indicatore biometrico di massa corporea, espresso in kg/m^2 ;
- *eBook*: 0 se il cliente non possiede un lettore di eBook, 1 se lo possiede;
- *Libri*: numero di libri acquistati dal cliente nel corso del 2018;
- *Film*: numero di film acquistati dal cliente nel corso del 2018;
- *Serie*: numero di serie TV acquistate dal cliente nel corso del 2018.

In questo file il carattere "," separa le colonne e i numeri reali sono stati registrati usando il carattere "." come separatore dei decimali.

- ✓ Quanti casi contiene il dataset?
- ✓ Il carattere *BMI* è scalare, ordinale o categorico? Giustificate la risposta.
- ✓ Visualizzate la tabella delle frequenze congiunte dei caratteri *Sesso* e *eBook*.
- ✓ Utilizzate la tabella delle frequenze congiunte per rispondere alle seguenti domande:
 - 4.1. Quanti uomini posseggono un lettore di eBook?
 - 4.2. Quante donne posseggono un lettore di eBook?
 - 4.3. Qual è la percentuale di uomini che possiede un lettore di eBook?
 - 4.4. Qual è la percentuale di donne che possiede un lettore di eBook?

Esercizio 5

- ✓ Tracciate un grafico opportuno per controllare se esiste una relazione tra i caratteri *Film* e *Serie*.
- ✓ Caratterizzate la relazione tra *Film* e *Serie*, eventualmente con l'ausilio di un indice numerico.

Esercizio 6

1. Selezionate tutti i casi in cui il carattere *Sesso* è uguale a 0 e memorizzateli in una variabile chiamata *uomini*.

In questo esercizio e anche nel prossimo si farà riferimento a questo nuovo insieme di dati costituito da soli uomini.

2. Tracciate un grafico opportuno per descrivere il carattere *Libri* negli uomini.
3. Compilate la Tabella 1, in cui dovete inserire il nome e il valore degli indici più appropriati, e sul foglio giustificate le vostre scelte.

Tabella 1: Indici descrittivi del carattere *Libri* negli uomini.

nome dell'indice scelto	Indice di centralità	Indice di dispersione	primo quartile	terzo quartile
	media	deviazione standard		
valore	0.67	0.96	0	1

4. Quale modello probabilistico potrebbe essere considerato per descrivere il carattere *Libri* negli uomini? Giustificate la risposta.
5. La Tabella 2 mostra le frequenze cumulate assolute del carattere *Libri* negli uomini. Completate la tabella calcolando le corrispondenti frequenze cumulate relative.

Tabella 2: Frequenze cumulate del carattere *Libri* negli uomini.

valori	0	1	2	3	4
frequenze cumulate assolute	45	61	74	76	77
frequenze cumulate relative	0.58	0.79	0.96	0.99	1

Esercizio 7

In questo esercizio consideriamo ancora l'insieme di osservazioni costituite da soli uomini.

1. Tracciate il grafico (diverso dal *boxplot*) che ritenete più opportuno per descrivere il carattere *BMI* negli uomini.
2. L'ipotesi che il *BMI* negli uomini segua una legge normale è plausibile? Giustificate risposta attraverso un eventuale grafico.

- ✓ 2. Calcolate la media del carattere BMI negli uomini; nel seguito indicheremo con μ tale valore.
- ✓ 3. Calcolate la deviazione standard del carattere BMI negli uomini; nel seguito indicheremo con σ tale valore.
- ✓ 4. Calcolate la frequenza relativa degli uomini che hanno il BMI compreso tra $\mu - 0.67 \cdot \sigma$ e $\mu + 0.67 \cdot \sigma$.
- ✓ 5. Calcolate la distanza interquartile del BMI negli uomini.
- ✓ 6. In Figura 1 è mostrato il *boxplot* del carattere BMI negli uomini. Su tale figura evidenziate la distanza interquartile calcolata nel punto precedente.
- ✓ 7. Fornite una stima del valore atteso del carattere BMI negli uomini.
- ✓ 8. Qual è la taglia del campione che avete utilizzato per la stima?
- ✓ 9. La taglia del campione che avete utilizzato è sufficiente affinché si abbia una probabilità almeno uguale a 0.9 di compiere nella stima del BMI atteso negli uomini un errore al più uguale a 0.5, per eccesso o per difetto? Giustificate la risposta.

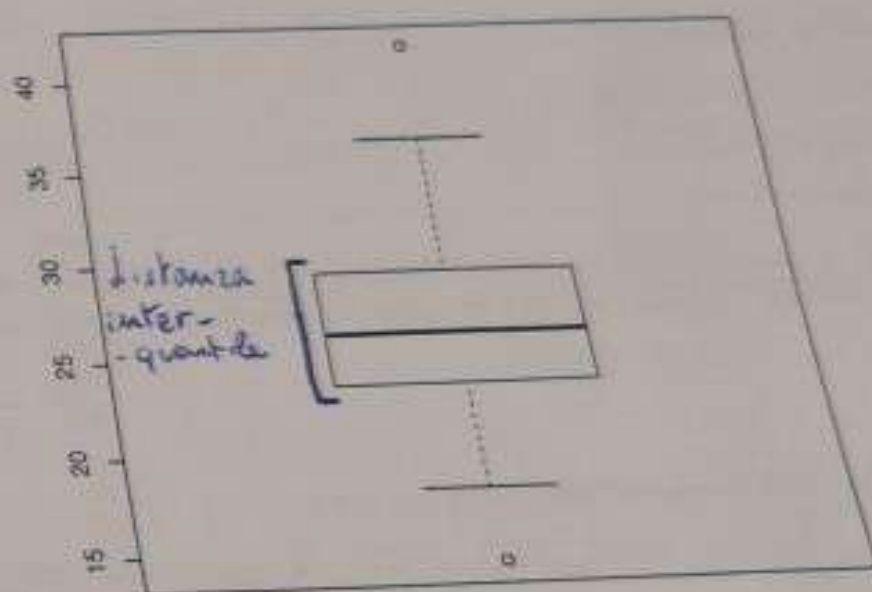


Figura 1: Boxplot del BMI negli uomini.