

Tema d'esame di Statistica e analisi dei dati

Prova scritta del 3 luglio 2019

Esercizio 0

Sia X una variabile casuale binomiale di parametri n e p .

1. Fissiamo, *solo in questo punto*, $n = 9$ e $p = 0.5$. Tracciate a mano il grafico della funzione di massa di probabilità di X . Siate attenti a evidenziare sul grafico tutte le informazioni importanti.
2. Esprimete, in funzione di n e p , il valore atteso $E(X)$.
3. Esprimete p in funzione di n e $E(X)$.
4. Esprimete, in funzione di n e p , la varianza $\text{var}(X)$.

Esercizio 1

Sia $\bar{X}_{(n)}$ la media campionaria di un campione casuale X_1, \dots, X_n estratto da una popolazione X binomiale di parametri n e p .

1. Proponete uno stimatore, chiamiamolo T_n , del valore atteso di X .
2. Lo stimatore che avete proposto al punto precedente è non distorto? Giustificate la risposta.
3. Proponete uno stimatore non distorto, chiamiamolo U_n , del parametro p . Giustificate la risposta.

Esercizio 3

Collegatevi al sito upload.di.unimi.it, selezionate l'esame di *Statistica e analisi dei dati* per l'appello odierno e scaricate il file `impianti.csv`. Questo file contiene le seguenti informazioni raccolte dal Comune di Milano riguardo agli impianti termici installati negli edifici della città:

- `IDENTIFICATIVO_IMPIANTO`: identificatore dell'impianto;
- `GENERATORI_NUMERO`: numero di generatori dell'impianto;
- `POTENZA_IMPIANTO_RISC`: potenza dell'impianto di riscaldamento, espressa in kW;
- `POTENZA_IMPIANTO_RAFF`: potenza dell'impianto di raffreddamento, espressa in kW;
- `EDIFICIO_CATEGORIA`: categoria catastale dell'edificio;
- `VOLUMETRIA_RISC`: volumetria da riscaldare, espressa in m^3 ;

1

- `VOLUMETRIA_RAFF`: volumetria da raffreddare, espressa in m^3 ;
- `GENERATORE_POTENZA`: potenza del generatore, espressa in kW;
- `GENERATORE_COMBUSTIBILE`: tipo di combustibile utilizzato;

- *VOLUMETRIA_RAFF*: volumetria da raffreddare, espressa in m³;
- *GENERATORE_POTENZA*: potenza del generatore, espressa in kW;
- *GENERATORE_COMBUSTIBILE*: tipo di combustibile utilizzato;
- *GENERATORE_DATA_INST*: anno di installazione;
- *RAPPORTO_DI_CONTROLLO_DATA*: anno in cui è stato fatto il controllo dell'impianto;
- *RAP_DI_CONTROLLO_ESITO*: esito del controllo;
- *ISPEZIONE_DATA*: anno in cui è stata effettuata l'ispezione dell'impianto;
- *ISPEZIONE_ESITO*: esito dell'ispezione.

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

1. Quanti casi contiene il dataset?
2. Tracciate un grafico opportuno per descrivere il carattere *POTENZA_IMPIANTO_RISC*.
3. Compilate la Tabella 1, in cui dovete inserire il nome e il valore degli indici più appropriati, e sul foglio giustificate la vostra scelta.

Tabella 1: Potenza dell'impianto di riscaldamento

| | Indice di centralità | Indice di dispersione | | |
|----------------------------|-------------------------|--------------------------|-------------------|-------------------|
| nome dell'indice scelto | | | primo quartile | terzo quartile |
| valore | | | | |

Considerate ora il carattere *GENERATORE_DATA_INST*.

4. Il carattere *GENERATORE_DATA_INST* è nominale, ordinale o scalare?
5. In quale anno sono stati installati i generatori più datati?
6. Tra quali anni è avvenuto il 50% centrale delle installazioni?
7. In quale anno si è registrato il maggior numero di installazioni?
8. Qual è l'anno dopo il quale si sono registrate la metà delle installazioni totali?
9. In Figura 1 è mostrato il boxplot del carattere *GENERATORE_DATA_INST*. Su tale grafico evidenziate a mano, nel modo che ritenete più efficace, la moda del carattere.

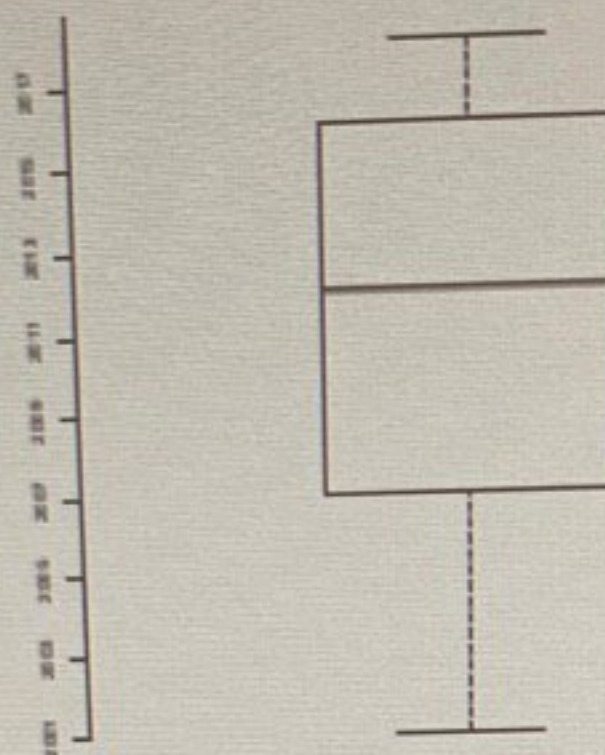


Figura 1: Boxplot dell'anno di installazione del generatore dell'impianto.

Esercizio 4

Consideriamo ora la volumetria degli ambienti da riscaldare d'inverno (tramite l'impianto di riscaldamento) e la volumetria degli ambienti da da raffreddare d'estate (tramite l'impianto di raffreddamento).

1. Tracciate un grafico che vi permetta di visualizzare se c'è una relazione tra la volumetria da riscaldare e quella da raffreddare.
2. Caratterizzate la relazione suggerita dal grafico del punto precedente e fornite il valore di un indice numerico che confermi la vostra ipotesi.
3. Quanti sono i casi in cui la volumetria da riscaldare e quella da raffreddare sono diverse?
Attenzione a non farvi ingannare dal grafico!

Esercizio 5

I dati raccolti riguardano impianti controllati in un particolare intervallo di anni, e il carattere `RAPPORTO_DI_CONTROLLO_DATA` corrisponde all'anno in cui è stato effettuato il controllo di un dato impianto.

La Figura 2 mostra la distribuzione delle frequenze degli anni nell'intervallo di osservazione.

1. Calcolate il minimo e il massimo del carattere `RAPPORTO_DI_CONTROLLO_DATA` per ottenere l'estremo inferiore e quello superiore di tale intervallo.
2. Il grafico di figura 2 riguarda le frequenze assolute oppure quelle relative?
3. Traslate tutti i valori di `RAPPORTO_DI_CONTROLLO_DATA` in modo da far partire i nuovi valori da 0, e salvate i dati così trasformati in una variabile chiamata `anni`.

Da questo punto studieremo il carattere `anni` dimenticandoci del fatto che i valori di questa variabile siano stati ottenuti a partire da `RAPPORTO_DI_CONTROLLO_DATA`. Possiamo considerare i valori del carattere `anni` come la realizzazione campionaria di un campione casuale estratto da una popolazione X .

sta variabile siano stati ottenuti a partire da *RAPPORTO_DI_CONTROLLO_DATA*. Possiamo considerare i valori del carattere *anni* come la realizzazione campionaria di un campione casuale estratto da una popolazione X .

3

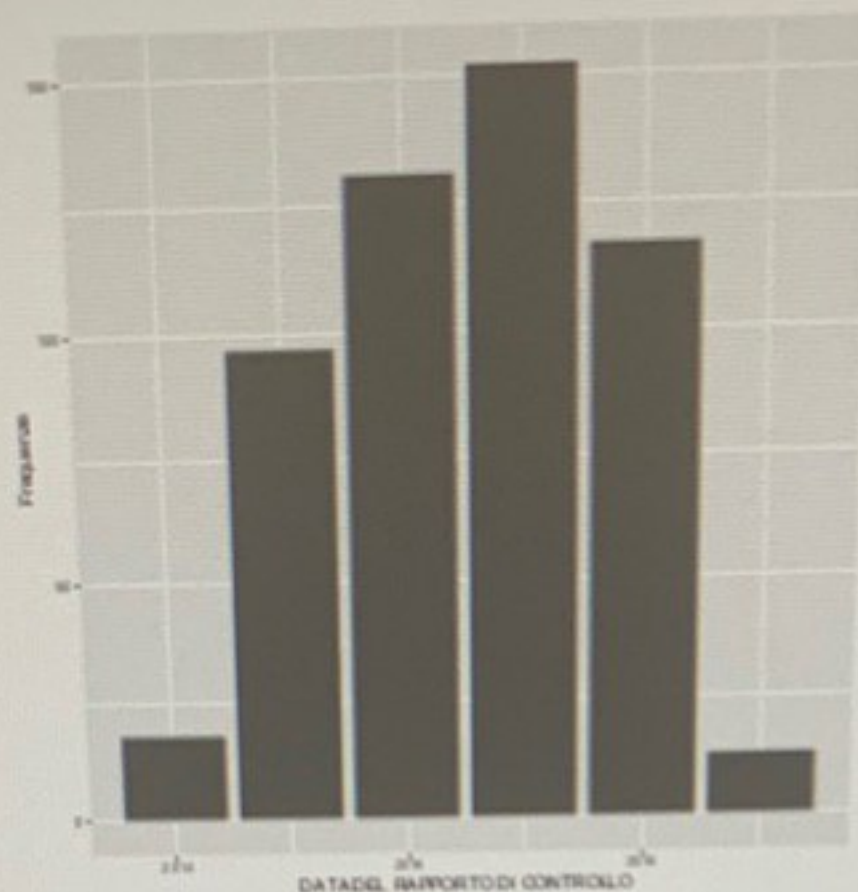


Figura 2: Distribuzione delle frequenze del carattere *RAPPORTO_DI_CONTROLLO_DATA*.

4. Stimare il valore atteso di X .
5. Stimare la varianza di X .
6. Calcolare il primo e il terzo quartile di X .
7. Tracciare un grafico a barre delle frequenze relative di X .
8. Il grafico ottenuto al punto precedente potrebbe avervi fatto pensare a una distribuzione binomiale. Nell'ipotesi che X segua un legge binomiale di parametri n e p , qual è una stima sensata per il valore del parametro n ?
9. Fornite una stima del parametro p .
10. Calcolare i quartili teorici di una distribuzione binomiale che ha lo stesso valore atteso e la stessa varianza di X , utilizzando la stima di n trovata al punto precedente per fornire una stima di p .
11. Confrontate i quartili empirici con quelli teorici: confermate l'ipotesi che i valori osservati di X siano la realizzazione campionaria di un campione estratto da una popolazione binomiale?