

Esercizio 0

Sia C una variabile aleatoria distribuita secondo un modello binomiale di parametri n_C e $p_C \in [0, 1]$. Nel resto del tema d'esame C modellerà il numero di bottiglie d'acqua minerale di una certa marca che un nucleo familiare acquista nell'arco di una settimana.

1. Quali valori può assumere il parametro n_C ? e quali valori può assumere C ? Giustificate la risposta
2. Indichiamo con f_C la funzione di massa di probabilità della variabile aleatoria C . Scrivete la forma analitica di f_C in dipendenza dai parametri indicati al punto precedente e di una generica specificazione x della variabile aleatoria C .
3. Fissiamo, **solo in questo punto**, $n_C = 20$ e $p_C = 0.8$. Utilizzando il computer visualizzate graficamente l'andamento di $f_C(x)$ al variare di x , **nella zona in cui questa funzione assume valori sensibilmente diversi da zero**. Come avete scelto i valori minimo e massimo per x ?

Esercizio 1

A ogni acquisto, il distributore effettua uno sconto di x euro per ogni bottiglia vuota resa. Ipotizziamo che il numero di bottiglie rese da un nucleo familiare in una settimana sia descritto da una variabile aleatoria R distribuita secondo un modello binomiale di parametri $n_R \in N$ e $p_R \in [0, 1]$. In questo esercizio supporremo che n_C e n_R **assumano valori piccoli**.

1. Indichiamo con p il prezzo di acquisto di una bottiglia. Che cosa modella la variabile aleatoria $T = pC - sR$?
2. La distribuzione esatta di T è nota? Se sì, indicate a quale modello si riferisce e specificate i relativi parametri. Cambia qualcosa nel caso fossimo interessati a una distribuzione approssimata?
3. Indichiamo con μ_T il valore atteso di T . Esprimete μ_T in funzione dei parametri sopra introdotti, giustificando i passaggi matematici intermedi.
4. Indichiamo con σ_T la deviazione standard di T . Esprimete σ_T in funzione dei parametri sopra introdotti, giustificando anche in questo caso i passaggi matematici intermedi.

Esercizio 3

Indichiamo con $n \in N$ il numero totale di clienti del distributore, e con C_1, \dots, C_n le variabili aleatorie che indicano il numero di bottiglie acquistate in una settimana da ogni cliente. Analogamente, siano R_1, \dots, R_n le variabili aleatorie che indicano il numero di bottiglie rese in una settimana. Assumiamo, infine, che vi sia indipendenza tra le abitudini di acquisto dei vari clienti. In questo esercizio supporremo che n assuma **un valore elevato**.

1. Esiste un modello che descrive in modo esatto la distribuzione della variabile aleatoria $C_{tot} = \sum_{i=1}^n C_i$? In caso affermativo, indicare quali sono i suoi parametri, giustificando la vostra risposta.
2. Esiste un modello che approssima ragionevolmente bene la distribuzione di C_{tot} ? In caso affermativo, indicate quali sono i suoi parametri, giustificando la vostra risposta.
3. Se nelle domande 1 e 2 di questo esercizio sostituiamo C_{tot} con $R_{tot} = \sum_{i=1}^n R_i$, le risposte cambiano significativamente? Perché?
4. Che cosa modella la variabile aleatoria $T_{tot} = p C_{tot} - s R_{tot}$?
5. Esiste un modello che descrive la distribuzione di T_{tot} ? In caso affermativo, indicate quali sono i suoi parametri, e in caso negativo indicare un modello che approssima ragionevolmente la distribuzione, sempre specificandone i parametri.

Esercizio 4

Il file "acquisti.csv" contiene le seguenti informazioni raccolte settimanalmente dal distributore dell'acqua della marca considerata nell'arco di più anni (il carattere , separa le colonne):

- *bottiglie_acquistate*: numero totale di bottiglie acquistate dei clienti dal distributore in una settimana;
- *bottiglie_rese*: numero totale di bottiglie rese dai clienti dal distributore in una settimana;
- *temperatura*: temperatura media nella settimana di riferimento;

1. Scrivete ed eseguite del codice che calcoli la percentuale di casi del dataset che contengono almeno un valore mancante
2. Descrivete l'attributo *bottiglie_acquistate* utilizzando la rappresentazione grafica che ritenete più adeguata, motivando la scelta fatta e commentando i risultati ottenuti.
3. L'ipotesi fatta all'inizio dell'esercizio 3 dovrebbe consentire di utilizzare un istogramma in alternativa alla rappresentazione del punto precedente; perchè? Visualizzate tale istogramma, scegliendo opportunamente il numero di intervalli da considerare.
4. Valutate l'ipotesi che vi sia una relazione di tipo diretto tra gli attributi *bottiglie_acquistate* e *temperatura*, utilizzando in modo opportuno sia un metodo grafico sia un indice numerico. Commentate i risultati ottenuti.
5. Una bottiglia d'acqua costa 1€, e il distributore sconta 10 centesimi per ogni bottiglia resa. Aggiungete al dataset una colonna dal nome *ricavo* che contenga il ricavo settimanale per ogni caso.

6. Valutate l'ipotesi che il ricavo settimanale di ogni acquisto sia ben descritto da un modello normale, commentando i risultati ottenuti.
7. Alla luce dei risultati ottenuti al punto precedente, è opportuno rivedere le ipotesi fatte all'inizio dell'esercizio 3? Perché? E quali ipotesi devono essere riesaminate?
8. Valutate l'ipotesi che il numero di bottiglie acquistate sia ben descritto da un modello normale, commentando i risultati ottenuti.

Esercizio 5

In questo esercizio interpreteremo i valori dell'attributo *bottiglie_acquistate* come un campione casuale estratto da una popolazione distribuita come la variabile aleatoria C_{tot} definita nell'esercizio 3, e i valori dell'attributo *ricavo* come un campione casuale estratto da una popolazione T .

1. Stimate il valore atteso della popolazione descritta dalla variabile aleatoria C_{tot} , indicando la dimensione del campione utilizzato e specificando eventuali proprietà dello stimatore utilizzato.
2. Stimate la deviazione standard della popolazione descritta dalla variabile aleatoria C_{tot} , specificando eventuali proprietà dello stimatore utilizzato. La dimensione del campione utilizzato è la stessa del punto precedente? Perché?
3. Il magazzino del distributore riceve ogni settimana 3600 bottiglie. Come potete indicare, in funzione di un'appropriata quantità precedentemente utilizzata, l'evento E che si verifica se queste bottiglie non sono sufficienti a soddisfare la domanda dei clienti?
4. Sulla base del risultato del punto 8 dell'esercizio 4, calcolate la probabilità dell'evento E descritto al punto precedente.
5. Stimate il valore atteso della popolazione T , indicando quale stimatore avete utilizzato.
6. Siete in grado di calcolare la probabilità che la stima fatta al punto precedente disti, per eccesso o per difetto, più di 10€ rispetto al valore sconosciuto?