

Esercizi riepilogativi “Teoria dei Campioni”

Esercizio 1:

La popolazione della città X ammonta a 50.000 abitanti. Si vuole condurre un'indagine campionaria al fine di stimare la proporzione di cittadini favorevoli all'allestimento di un nuovo servizio di trasporto. Un'indagine precedente afferma che ben il 70% dei cittadini è favorevole all'allestimento del nuovo servizio.

Si decide di estrarre, con campionamento casuale semplice senza reintroduzione, un campione di 1.000 cittadini. L'amministrazione comunale decide che il servizio sarà allestito solo se i risultati dell'indagine confermeranno che almeno il 70% dei cittadini è favorevole. Si concorda che il livello di fiducia sia del 95%.

- Calcolare le probabilità di inclusione di primo e secondo ordine degli elementi della popolazione.
- Calcolare l'errore assoluto associato alla strategia campionaria, senza tenere conto dell'informazione derivante dall'indagine precedente. Che ipotesi è necessario fare sulla varianza di popolazione?
- Calcolare l'errore assoluto associato alla strategia campionaria, tenendo conto dell'informazione derivante dall'indagine precedente.
- Calcolare la numerosità campionaria necessaria ad ottenere un errore assoluto pari a 0,02, senza tenere conto dell'informazione derivante dall'indagine precedente.

In seguito all'estrazione del campione ($n=1.000$), 620 cittadini dichiarano di essere favorevoli all'allestimento del nuovo servizio.

- Stimare la proporzione di cittadini favorevoli. Stimare il totale di cittadini favorevoli.
- Mediante la costruzione di un intervallo di confidenza, quale decisione prenderà l'amministrazione sulla base dell'esito campionario, tenendo conto di un 5% di probabilità di errore?

Esercizio 2

Si vuole valutare la spesa totale che le imprese edili italiane sopportano per il trasporto su strada e per il trasporto ferroviario. A tale scopo viene estratto un campione casuale semplice senza reintroduzione di 1.000 imprese su un totale di 50.000 imprese operanti sul territorio. Per ciascuna impresa si rileva: spesa per il trasporto su strada (y), spesa per il trasporto ferroviario (y_1), spesa per carburante nell'ultimo anno (z). Di seguito vengono riportate alcune statistiche campionarie:

$$\begin{array}{lll} \sum_{i=1}^{1000} y_i = 7600 & \sum_{i=1}^{1000} y_i^2 = 79840 & \sum_{i=1}^{1000} y_{1i} = 8000 \\ \sum_{i=1}^{1000} z_i = 3980 & \sum_{i=1}^{1000} z_i^2 = 43255 & \sum_{i=1}^{1000} y_i z_i = 34200 \end{array}$$

- Può essere utilizzata l'informazione relativa alla variabile z per ottenere stimatori più efficienti di quello utilizzato nel ccssr?
- Stimare la spesa totale per trasporto su strada e la spesa totale per il trasporto ferroviario, senza tenere conto della variabile ausiliaria.
- Stimare la varianza degli stimatori impiegati.
- Le due stime sono significativamente diverse? (livello di confidenza del 95%).

L'ufficio del registro delle imprese comunica che la spesa totale per carburante delle imprese edili italiane è pari a 250.000 euro.

- Stimare la spesa totale per trasporto su strada utilizzando lo stimatore per quoziente.
- Stimare la varianza dello stimatore impiegato.
- Anche se lo stimatore per quoziente è distorto, in questo caso è possibile confrontare solo la varianza dello stimatore per quoziente e quella dello stimatore nel ccssr? Commentare l'affermazione.
- Di quanto è migliorata la precisione della stima utilizzando lo stimatore per quoziente?

Esercizio 3

Nel comune X operano 5 diverse aziende nel comparto chimico. Si vuole valutare la disponibilità degli operai delle aziende ad adottare un nuovo tipo di maschera protettiva. Nella tabella successiva sono riportati i seguenti dati: numero operai (N_k), proporzione di favorevoli ad adottare il nuovo equipaggiamento secondo un'indagine precedente ($p_h = \pi_k$).

Azienda	N_k	π_k
1	1600	0.65
2	300	0.75
3	200	0.80
4	700	0.60
5	400	0.70

Si decide di effettuare un campionamento stratificato, campionando in totale 400 operai.

- Determinare la numerosità di strato con allocazione ottimale.
- Determinare la numerosità di strato con allocazione proporzionale.

Si estrae un campione con le numerosità di strato ottenute al punto b). Nelle 5 aziende si dichiarano favorevoli rispettivamente 150,19,20,73 e 32 operai.

- Stimare la proporzione dei favorevoli in popolazione. Stimare la varianza dello stimatore impiegato.
- La varianza dello stimatore sarebbe stata inferiore se si fosse estratto un campione con le numerosità di strato ottenute al punto a)?
- Costruire un intervallo di confidenza al 95% per la proporzione di favorevoli. Alla luce di questo risultato si può dire che gli operai abbiano cambiato idea rispetto all'indagine precedente?

Per evitare investimenti per la ricerca troppo elevati, si vuole stimare la proporzione di favorevoli solo nella prima azienda essendo la più rappresentativa delle altre, utilizzando lo stesso campione estratto al punto b). Dopo l'estrazione del campione si viene a conoscenza del fatto che 150 operai si sono dichiarati favorevoli ad utilizzare la nuova maschera.

- Stimare la proporzione di favorevoli nell'azienda
- Stimare la varianza dello stimatore proposto al punto precedente
- calcolare il Deff tra i due stimatori per capire quale strategia migliore perseguire.

Esercizio 4

Si consideri una popolazione composta da 4 unità. Di tale popolazione sono note le distribuzioni di una variabile di interesse (η) e di due variabili ausiliarie (ζ_1 e ζ_2). I dati di popolazione sono riportati nella tabella successiva.

	η	ζ_1	ζ_2
A	4	1	10
B	8	8	8
C	12	6	6
D	16	10	2

Si vuole estrarre, con un campionamento a probabilità variabile senza reintroduzione, un campione di ampiezza $n=2$.

- Lo statistico incaricato di effettuare il campionamento decide di utilizzare la variabile ζ_1 per assegnare le probabilità di estrazione, perché?
- Calcolare le probabilità di estrazione associate a ciascun elemento della popolazione, utilizzando la variabile ausiliaria ζ_1
- Calcolare la probabilità di inclusione del primo ordine per l'elemento A della popolazione.

Si supponga di estrarre il campione di etichette (B,D).

- Qual è la probabilità associata a tale campione?
- Stimare il totale sulla base del campione estratto
- Sempre sulla base dei dati campionari, stimare la varianza dello stimatore impiegato al punto precedente. Costruire un intervallo di confidenza al 95% per il totale di popolazione. L'intervallo di confidenza coglie adeguatamente ciò che si desidera sapere della popolazione

Esercizio 5

Si vuole valutare il valore medio di colesterolo su una popolazione di individui adulti che presentano una certa malattia. In base a precedenti studi si sa che i valori dello scarto quadratico medio della variabile colesterolo sono quelli riportati in tabella.

Fascia età	31-50	51-70	>71
Scarto quadratico medio	25	38	41

La popolazione di interesse è così suddivisa per fasce di età:

Fascia età	31-50	51-70	>71
Numero individui popolazione	5000	2000	3000

Si effettua un campionamento di 250 soggetti stratificato rispetto alle fasce di età con allocazione ottima.

- Da quanti individui di ogni fascia di età deve essere composto il campione?
- Determinare la stima della media di colesterolo nella popolazione se i valori medi di colesterolo delle tre fasce sono quelli riportati in tabella.

Fascia età	31-50	51-70	>71
Valore medio colesterolo	254	275	215

- Determinare un intervallo di confidenza per la media di colesterolo nella popolazione, dopo aver stimato la varianza dello stimatore della media