

Concetti di teoria dei campioni ad uso degli studenti
di Statistica Economica e Finanziaria, A.A. 2018/2019

Giovanni Lafratta

Indice

1	Spazi, Disegni e Strategie Campionarie	1
2	Campionamento casuale semplice	7
2.1	Vettori e sottoinsiemi da un insieme finito	7
2.2	Disegno casuale semplice di ampiezza data	9
2.2.1	Probabilità di inclusione	10
3	Statistiche di Horvitz-Thompson	13
3.1	Stimatore del totale di una popolazione	13
3.2	Stimatore della media di una popolazione	14
4	Materiali per l'esame	17
4.1	Esercitazioni	17

Capitolo 1

Spazi, Disegni e Strategie Campionarie

Si introduce il concetto di popolazione finita.

Definizione 1 (*Popolazione finita $\mathcal{P}_M$*)

1. $\mathcal{P}_1 = \{1\}$ è una popolazione finita;
2. per ogni $M \in \mathbb{N} \setminus \{0\}$ se $\mathcal{P}_M$ è una popolazione finita, allora

$$\mathcal{P}_{M+1} = \mathcal{P}_M \cup \{M+1\}$$

è una popolazione finita;

3. le uniche popolazioni finite sono quelle formate sulla base delle clausole 1 e 2.

Esempio 1 Per $M = 4$, si ha

$$\mathcal{P}_4 = \{1, 2, 3, 4\}.$$

Spazi e disegni di campionamento costituiscono strumenti utili nello studio della variabilità statisticamente derivante dal non aver osservato in modo esaustivo un dato fenomeno su una popolazione $\mathcal{P}_M$:

Definizione 2 (*Spazio campionario $S_{\mathcal{P}_M}$*) Sia $\mathcal{P}_M$ una popolazione finita, e, per ogni $n \in \mathcal{P}_M$, sia S_n la classe dei sottoinsiemi di $\mathcal{P}_M$ aventi cardinalità pari a n . Per spazio campionario su $\mathcal{P}_M$ si intenderà la classe

$$S_{\mathcal{P}_M} = \bigcup_{n \in \mathcal{P}_M} S_n.$$

Per **campione** s in $\mathcal{P}_M$ si intenderà ogni elemento $s \in S_{\mathcal{P}_M}$.

Esempio 2 Per $M = 3$, si ha

$$S_{\mathcal{P}_M} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \mathcal{P}_3\}.$$

Definizione 3 (*Disegno campionario* p) Sia $S_{\mathcal{P}_M}$ uno spazio di campionamento. Una funzione $p : S_{\mathcal{P}_M} \rightarrow [0, +\infty[$ è un disegno campionario su $S_{\mathcal{P}_M}$ se e solo se risulta

$$\sum_{s \in S_{\mathcal{P}_M}} p(s) = 1.$$

Un disegno p si dirà **intenzionale** se e solo se

$$\exists s \in S_{\mathcal{P}_M} : p(s) = 1.$$

Un disegno p può quindi essere interpretato come la funzione di probabilità di una variabile casuale S a valori in $S_{\mathcal{P}_M}$.

Esempio 3 Si consideri la seguente funzione $p : S_{\mathcal{P}_3} \rightarrow [0, +\infty[$:

$s \in S_{\mathcal{P}_3}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	$\mathcal{P}_3$
$p(s)$	0	0	1/4	1/3	5/12	0	0

Essa è un disegno, perché $1/4 + 1/3 + 5/12 = 1$.

Definizione 4 (*Probabilità d'inclusione I*) Per $i \in \mathcal{P}_M$, si definiscano gli insiemi

$$A_i = \{s \in S_{\mathcal{P}_M} : i \in s\}.$$

La probabilità π_i di inclusione **del primo ordine** dell'unità i -esima è definita come segue:

$$\pi_i = \sum_{s \in A_i} p(s).$$

Esempio 4 (*Continua l'esempio 3*) L'unità 2 è inclusa nei seguenti elementi (campioni) dello spazio campionario $S_{\mathcal{P}_3}$:

$$A_2 = \{\{2\}, \{1, 2\}, \{2, 3\}, \mathcal{P}_3\}$$

La probabilità d'inclusione di ordine 1 per l'unità 2 è quindi

$$\begin{aligned} \pi_2 &= p(\{2\}) + p(\{1, 2\}) + p(\{2, 3\}) + p(\mathcal{P}_3) \\ &= 0 + \frac{1}{3} + 0 + 0 \\ &= \frac{1}{3}. \end{aligned}$$

L'unità 1 è inclusa nei seguenti elementi (campioni) dello spazio campionario $S_{\mathcal{P}_3}$:

$$A_1 = \{\{1\}, \{1, 2\}, \{1, 3\}, \mathcal{P}_3\}.$$

La probabilità d'inclusione di ordine 1 per l'unità 1 è quindi

$$\begin{aligned}\pi_1 &= p(\{1\}) + p(\{1, 2\}) + p(\{1, 3\}) + p(\mathcal{P}_3) \\ &= 0 + \frac{1}{3} + \frac{5}{12} + 0 \\ &= \frac{9}{12}.\end{aligned}$$

Per ogni $i \in \mathcal{P}_M$, si definisca ora la funzione $\delta_i : S_{\mathcal{P}_M} \longrightarrow \{0, 1\}$ come segue:

$$\delta_i(s) = \begin{cases} 1 & \text{se } i \in s, \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione δ_i prende il nome di *funzione indicatrice dell'unità i -esima*, grazie alla quale si può definire la variabile casuale

$$\Delta_i = \delta_i(S)$$

e si ha

$$\begin{aligned}\pi_i &= \sum_{s \in S_{\mathcal{P}_M}} \delta_i(s) p(s) \\ &= E_p[\Delta_i].\end{aligned}$$

Definizione 5 (*Cardinalità campionaria*) Dato il campione casuale S , si definisce come *cardinalità campionaria* la variabile

$$|S| = \sum_{i \in \mathcal{P}_M} \delta_i(S).$$

La cardinalità attesa è

$$\begin{aligned}E_p[|S|] &= \sum_{s \in S_{\mathcal{P}_M}} \sum_{i \in \mathcal{P}_M} \delta_i(s) p(s) \\ &= \sum_{i \in \mathcal{P}_M} \sum_{s \in A_i} p(s) \\ &= \sum_{i \in \mathcal{P}_M} \pi_i.\end{aligned}$$

Inoltre, se il disegno campionario p ammette l'estrazione solo di campioni aventi una data cardinalità $n \in \mathcal{P}_M$, ovvero se

$$\forall s \in S_{\mathcal{P}_M} : |s| \neq n \Rightarrow p(s) = 0,$$

allora

$$E_p[|S|] = \sum_{s \in S_n} |s| p(s) = n,$$

perché $\sum_{s \in S_n} p(s) = 1$ e $|s| = n$ quando $s \in S_n$.

Definizione 6 (*Probabilità d'inclusione II*) Per $i, j \in \mathcal{P}_M$, si definiscano gli insiemi

$$A_{i,j} = \{s \in S_{\mathcal{P}_M} : \{i, j\} \subset s\}.$$

La probabilità $\pi_{i,j}$ di inclusione **del secondo ordine** delle unità $\{i, j\}$ è definita come segue:

$$\pi_{i,j} = \sum_{s \in A_{i,j}} p(s).$$

Esempio 5 (*Continua l'esempio 3*) Le unità 1, 3 sono incluse nei seguenti elementi (campioni) dello spazio campionario $S_{\mathcal{P}_M}$:

$$A_{1,3} = \{\{1, 3\}, \mathcal{P}_3\}$$

La loro probabilità d'inclusione di ordine 2 è quindi

$$\begin{aligned} \pi_{1,3} &= p(\{1, 3\}) + p(\mathcal{P}_3) \\ &= \frac{5}{12} + 0 \\ &= \frac{5}{12}. \end{aligned}$$

Per ogni $i, j \in \mathcal{P}_M$, si definisca ora la funzione $\delta_{i,j} : S_{\mathcal{P}_M} \longrightarrow \{0, 1\}$ come segue:

$$\delta_{i,j}(s) = \begin{cases} 1 & \text{se } \{i, j\} \subset s, \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione $\delta_{i,j}$ è detta *funzione indicatrice della coppia di unità i e j* e si ha

$$\delta_{i,j}(s) = \delta_i(s) \delta_j(s).$$

Definita la variabile

$$\Delta_{i,j} = \delta_{i,j}(S)$$

risulta inoltre

$$\begin{aligned} \pi_{i,j} &= \sum_{s \in S_{\mathcal{P}_M}} \delta_{i,j}(s) p(s) \\ &= E_p[\Delta_{i,j}]. \end{aligned}$$

Associando ad ogni elemento di $\mathcal{P}_M$ un numero reale Y_i , $i = 1, \dots, N$, si ottiene la funzione Y come segue:

$$i \in \mathcal{P}_M \mapsto Y(i) = Y_i,$$

alla quale si fa riferimento parlando di **fenomeno su** $\mathcal{P}_M$. Dato il campione s , sia D_s la funzione definita su s e a valori in $\mathbb{R}$ tale che

$$i \in s \mapsto D_s(i) = Y_i.$$

Per definizione, D_s è quindi l'insieme delle coppie (i, Y_i) per $i \in s$:

$$D_s = \{(i, Y_i) : i \in s\}.$$

Ci si riferisce a D_s come all'insieme dei **dati statistici** relativi al campione s .

Esempio 6 (*Continua l'esempio 3*) *Posto*

$$\begin{array}{cccc} i & 1 & 2 & 3 \\ Y_i & 5 & 7 & 5 \end{array}$$

risultano, ad esempio,

$$D_{\{1\}} = \{(1, 5)\},$$

$$D_{\{1,3\}} = \{(1, 5), (3, 5)\}$$

e

$$D_{\{1,2,3\}} = \{(1, 5), (2, 7), (3, 5)\}.$$

Definizione 7 (*Statistica campionaria*) Sia Y un fenomeno su $\mathcal{P}_M$ e sia D l'insieme dei possibili dati statistici disponibili:

$$D = \{D_s : s \in S_{\mathcal{P}_M}\}.$$

Data una funzione reale t definita su D :

$$D_s \in D \mapsto t(D_s) \in \mathbb{R},$$

si definisce *statistica campionaria* la variabile casuale

$$T = t(D_S).$$

Si noti che i valori assunti da una statistica T dipendono dal campione osservato. Di conseguenza, $t(D_S)$ dipende dalla funzione di probabilità p , il disegno campionario applicato.

Esempio 7 La variabile $T = t(D_S) = \sum_{i \in S} Y_i$ è una statistica.

Una statistica rispondente alla definizione 7 è destinata a fungere da stimatore di quantità deterministiche dipendenti da Y , note come **parametri della popolazione**, delle quali sono esempi la variabile media della popolazione $\bar{Y} = M^{-1} \sum_{i \in \mathcal{P}_M} Y_i$, o la variabile $\max(Y) = \sup_{i \in \mathcal{P}_M} Y_i$.

Definizione 8 (*Distribuzione di una statistica campionaria*) Per un fissato piano di campionamento $p(s)$, la distribuzione campionaria dello stimatore $T = t(D_S)$ è definita dalla seguente funzione di probabilità p_T . Sia C_T l'insieme di tutti i possibili valori che T può assumere presso i vari campioni. Per $x \in C_T$, si definisca l'insieme

$$B_x = \{s \in S_{\mathcal{P}_M} : t(D_s) = x\}.$$

Allora

$$p_T(x) \equiv \Pr\{T = x\} = \sum_{s \in B_x} p(s).$$

Esempio 8 (*Continua l'esempio 3*) Si vuole la distribuzione campionaria di $T = t(D_S) \equiv \sum_{i \in S} Y_i$. I valori che T può assumere presso i relativi campioni sono dati come segue:

s	D_s	$t(D_s)$
$\{1\}$	$\{(1, 5)\}$	5
$\{2\}$	$\{(2, 7)\}$	7
$\{3\}$	$\{(3, 5)\}$	5
$\{1, 2\}$	$\{(1, 5), (2, 7)\}$	12
$\{1, 3\}$	$\{(1, 5), (3, 5)\}$	10
$\{2, 3\}$	$\{(2, 7), (3, 5)\}$	12
$\mathcal{P}_3$	$\{(1, 5), (2, 7), (3, 5)\}$	17

Al singolo valore di T (si ha $C_T = \{5, 7, 10, 12, 17\}$) va associata una probabilità pari alla somma delle probabilità dei campioni in corrispondenza dei quali tale valore viene generato:

x	B_x	$\Pr\{T = x\}$
5	$\{\{1\}, \{3\}\}$	$0 + 1/4$
7	$\{\{2\}\}$	0
10	$\{\{1, 3\}\}$	$5/12$
12	$\{\{1, 2\}, \{2, 3\}\}$	$1/3 + 0$
17	$\{\{1, 2, 3\}\}$	0

In conclusione, gli strumenti con cui raccogliere informazioni su un fenomeno Y presso una popolazione $\mathcal{P}_M$ sono sostanzialmente due: un disegno di campionamento e una statistica. Ciò giustifica la seguente:

Definizione 9 (*Strategia campionaria per Y su $\mathcal{P}_M$*) Sia Y un fenomeno su $\mathcal{P}_M$. Una strategia per Y su $\mathcal{P}_M$ è una qualunque coppia (p, T) , con p disegno campionario su $S_{\mathcal{P}_M}$ e T statistica per Y .

La realizzazione di una indagine richiede, quindi, la scelta di una strategia.

Capitolo 2

Campionamento casuale semplice

2.1 Vettori e sottoinsiemi da un insieme finito

Cardinalità di un prodotto cartesiano

Siano $L_1, \dots, L_k$ insiemi di cardinalità finita $|L_i|$. Il loro prodotto cartesiano è definito come segue:

$$\times_{i \in \{1, \dots, k\}} L_i = \{(\lambda_1, \dots, \lambda_k) : \forall i \in \{1, \dots, k\} : \lambda_i \in L_i\}.$$

La sua cardinalità è:

$$\left| \times_{i \in \{1, \dots, k\}} L_i \right| = \prod_{i \in \{1, \dots, k\}} |L_i|. \quad (2.1)$$

Ci si riferisce agli elementi di un prodotto cartesiano parlando di *vettori* o di *sequenze*.

Vettori da un insieme finito

Sia L un insieme finito non vuoto e si indichi con l la sua cardinalità $|L|$. Il prodotto cartesiano di L moltiplicato per se stesso k volte ($k > 0$) ha cardinalità pari a l^k . Il prodotto ha per elementi tutte le sequenze (vettori) $(\lambda_1, \dots, \lambda_k)$ che si possono formare scegliendo per λ_i una qualunque unità in L . Questo significa che, all'interno di una sequenza, la stessa unità può ripetersi ed essere quindi presente in più di una delle k posizioni disponibili. Quale sarebbe il numero delle sequenze se si imponesse che ogni unità non può occupare più di una posizione? Per rispondere, si osservi che l'insieme delle sequenze di k elementi distinti (*senza ripetizioni*) da L ammette la seguente rappresentazione:

$$\bigcup_{\lambda_1 \in L} \bigcup_{\lambda_2 \in L \setminus \{\lambda_1\}} \bigcup_{\lambda_3 \in L \setminus \{\lambda_1, \lambda_2\}} \dots \bigcup_{\lambda_k \in L \setminus \{\lambda_1, \dots, \lambda_{k-1}\}} \{(\lambda_1, \dots, \lambda_k)\}.$$

Di conseguenza, esso può essere costruito come segue. Quando devo selezionare la prima componente del vettore, ($i = 1$), posso scegliere su tutto L : ho quindi l alternative disponibili. Fissato il primo valore λ_1 , posso scegliere la seconda componente ($i = 2$) da $L \setminus \{\lambda_1\}$, avendo $l - 1$ alternative. Per la terza componente, ($i = 3$), scelgo da $L \setminus \{\lambda_1, \lambda_2\}$,

che contiene $l - 2$ alternative. La k -esima volta scelgo da $L \setminus \{\lambda_1, \dots, \lambda_{k-1}\}$, un insieme avente cardinalità $l - (k - 1)$. In tal modo, tutte le posizioni disponibili vengono occupate avendo cura che le unità selezionate siano distinte tra loro. Si hanno quindi

$$l(l-1)(l-2) \cdots (l-(k-2))(l-(k-1))$$

vettori di k elementi distinti da L .

In generale, vale il seguente risultato.

Teorema 1 *Sia L un insieme finito non vuoto, con $l = |L| \in \mathbb{N} \setminus \{0\}$, e sia k un intero positivo che soddisfa $k \leq l$. Sia $\mathcal{V}_{L,k}$ l'insieme dei vettori lunghi k senza ripetizioni che si possono estrarre da L . Allora*

$$|\mathcal{V}_{L,k}| = \frac{l!}{(l-k)!}$$

dove, per $x \in \mathbb{N} \setminus \{0\}$, $x! = x(x-1) \cdots (2)(1)$ e $0! = 1$.

Infatti si ha

$$\begin{aligned} \frac{l!}{(l-k)!} &= \frac{l(l-1) \cdots (l-(k-1))(l-k) \cdots (1)}{(l-k) \cdots (1)} \\ &= l(l-1) \cdots (l-(k-1)). \end{aligned}$$

Sottoinsiemi di un insieme finito

Un sottoinsieme di ampiezza k dall'insieme finito L , con $l \equiv |L|$, è una collezione di k unità, ($k \leq l$), disposte senza alcun ordine particolare. Sia $\mathcal{F}_{L,k}$ la collezione i cui elementi sono i sottoinsiemi di L aventi cardinalità k :

$$\mathcal{F}_{L,k} = \{E \subset L : |E| = k\}.$$

Sia ora $E \in \mathcal{F}_{L,k}$. Quante sequenze lunghe k senza ripetizioni posso ottenere da E ? Tante quanti sono gli elementi di $\mathcal{V}_{E,k}$ e, per il teorema 1, si ha

$$|\mathcal{V}_{E,k}| = \frac{k!}{(k-k)!} = k!.$$

Si noti che, se tali sequenze sono estratte da E , si possono considerare estratte da L , perché $E \subset L$. Inoltre, $\mathcal{V}_{L,k}$, la collezione di **tutti** i vettori senza ripetizioni di k elementi da L , ha cardinalità

$$|\mathcal{V}_{L,k}| = \frac{l!}{(l-k)!}.$$

Di questi vettori, dunque, $k!$ si possono costruire a partire da E , al variare di E in $\mathcal{F}_{L,k}$.

Sia ora $\tilde{E}$ un secondo elemento di $\mathcal{F}_{L,k}$ diverso da E . Allora i $k!$ vettori che si possono costruire a partire da $\tilde{E}$ sono tutti diversi da quelli che si possono costruire utilizzando gli elementi di E :

$$\mathcal{V}_{E,k} \cap \mathcal{V}_{\tilde{E},k} = \emptyset.$$

Inoltre, risulta

$$\mathcal{V}_{L,k} = \bigcup_{E \in \mathcal{F}_{L,k}} \mathcal{V}_{E,k},$$

da cui gli insiemi $\mathcal{V}_{E,k}$, per $E \in \mathcal{F}_{L,k}$, costituiscono una partizione dell'insieme $\mathcal{V}_{L,k}$. Di conseguenza, deve risultare:

$$\begin{aligned} |\mathcal{V}_{L,k}| &= \sum_{E \in \mathcal{F}_{L,k}} |\mathcal{V}_{E,k}| \\ &= \sum_{E \in \mathcal{F}_{L,k}} k! \\ &= |\mathcal{F}_{L,k}| k!, \end{aligned}$$

da cui

$$|\mathcal{F}_{L,k}| k! = \frac{l!}{(l-k)!}.$$

Vale, quindi, il seguente risultato.

Teorema 2 *Sia L un insieme finito non vuoto, con $l = |L| \in \mathbb{N} \setminus \{0\}$, e sia k un intero positivo che soddisfa $k \leq l$. Sia $\mathcal{F}_{L,k}$ la classe dei sottoinsiemi di L aventi cardinalità k . Allora*

$$|\mathcal{F}_{L,k}| = \binom{l}{k},$$

dove il coefficiente binomiale

$$\binom{l}{k} = \frac{l!}{k! (l-k)!}$$

rappresenta il numero di combinazioni di l elementi presi a k a k .

2.2 Disegno casuale semplice di ampiezza data

I risultati discussi nella sezione precedente sono utili per determinare il numero di campioni di cardinalità n che si possono estrarre da una popolazione finita $\mathcal{P}_M$, con $0 < n \leq M$.

Si consideri $S_n = \{s \in \mathcal{P}_M : |s| = n\}$. Si vuole calcolare $|S_n|$: quanti sono i sottoinsiemi di cardinalità n di un insieme di cardinalità M ? Risponde il teorema 2, ponendo $L = \mathcal{P}_M$ e $k = n$:

$$|S_n| = |\mathcal{F}_{\mathcal{P}_M,n}| = \binom{M}{n}.$$

Data una popolazione $\mathcal{P}_M$, esistono M disegni casuali semplici senza ripetizione, uno per ogni $n \in \mathcal{P}_M$, ciascuno di essi rende equiprobabili i campioni di ampiezza n ed assegna ai rimanenti probabilità zero:

$$p_{C,M,n}(s) = \begin{cases} \frac{1}{\binom{M}{n}} & \text{se } s \in S_n \\ 0 & \text{se } s \notin S_n \end{cases}$$

Esempio 9 Per $\mathcal{P}_4 = \{1, 2, 3, 4\}$, si consideri $p_{C,4,3}$. I campioni di ampiezza 3 da $\mathcal{P}_4$ sono $\binom{4}{3} = \frac{4!}{(4-3)!3!} = 4$:

$$\{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 4\}, \{1, 3, 4\}.$$

Si ha

$$\begin{aligned} p_{C,4,3}(\{1, 2, 3\}) &= p_{C,4,3}(\{1, 2, 4\}) \\ &= p_{C,4,3}(\{2, 3, 4\}) \\ &= p_{C,4,3}(\{1, 3, 4\}) \\ &= \frac{1}{4} \end{aligned}$$

Per tutti gli altri elementi s di $S_{\mathcal{P}_4}$, ad esempio per $\{1, 3\}, \{1\}, \{3, 4\}$, si ha $p_{C,4,3}(s) = 0$.

2.2.1 Probabilità di inclusione

Primo ordine

Si consideri la generica unità $i \in \mathcal{P}_M$. Per definizione,

$$\pi_i = \sum_{s \in A_i} p_{C,M,n}(s),$$

dove

$$A_i = \{s \in S_{\mathcal{P}_M} : i \in s\}.$$

Si osservi che, se $s \notin S_n$, allora $p_{C,M,n}(s) = 0$, così possiamo limitare la somma agli s in $A_i \cap S_n$:

$$\pi_i = \sum_{s \in A_i \cap S_n} p_{C,M,n}(s)$$

Inoltre, se $s \in S_n$ si ha $p_{C,M,n}(s) = \binom{M}{n}^{-1}$, da cui

$$\pi_i = \binom{M}{n}^{-1} \sum_{s \in A_i \cap S_n} 1$$

Bisogna quindi contare i campioni di ampiezza n che includono i . A tal fine, si costruiscono tutti i campioni di $n - 1$ unità da $\mathcal{P}_M$ che **non** contengono i : essi sono in numero di

$$\binom{M-1}{n-1}.$$

Per dimostrarlo, basta applicare il teorema 2 per $L = \mathcal{P}_M \setminus \{i\}$ e porre $k = n - 1$. Se ad ognuno di tali campioni aggiungiamo l'unità i -esima, otteniamo tutti i campioni di ampiezza n che contengono i .

Di conseguenza,

$$\begin{aligned}\pi_i &= \binom{M}{n}^{-1} \sum_{s \in A_i \cap S_n} 1 \\ &= \binom{M}{n}^{-1} |A_i \cap S_n| \\ &= \frac{\binom{M-1}{n-1}}{\binom{M}{n}}\end{aligned}$$

ovvero

$$\begin{aligned}\pi_i &= \frac{(M-1)!}{(M-n)!(n-1)!} \frac{n!(M-n)!}{M!} \\ &= \frac{n}{M}.\end{aligned}$$

Secondo ordine

Si consideri la generica coppia $(i, j) \in \mathcal{P}_M \times \mathcal{P}_M$, con $i \neq j$. Si ha

$$\pi_{i,j} = \sum_{s \in A_{i,j}} p_{C,M,n}(s),$$

dove

$$A_{i,j} = \{s \in S_{\mathcal{P}_M} : \{i, j\} \subset s\}.$$

Dato che, per $s \notin S_n$, risulta $p_{C,M,n}(s) = 0$, possiamo limitare la somma agli s in $A_{i,j} \cap S_n$:

$$\pi_i = \sum_{s \in A_{i,j} \cap S_n} p_{C,M,n}(s).$$

Inoltre, se $s \in S_n$ si ha $p_{C,M,n}(s) = \binom{M}{n}^{-1}$, da cui

$$\pi_i = \binom{M}{n}^{-1} \sum_{s \in A_{i,j} \cap S_n} 1$$

Bisogna quindi contare i campioni di ampiezza n che includono sia i sia j . A tal fine, si costruiscano tutti i campioni di $n - 2$ unità da $\mathcal{P}_M$ che contengono né i né j : sempre per il teorema 2, per $L = \mathcal{P}_M \setminus \{i, j\}$ e $k = n - 2$, essi sono in numero di

$$\binom{M-2}{n-2}.$$

Se ad ognuno di tali campioni aggiungiamo le unità i e j , otteniamo tutti i campioni di ampiezza n che contengono i e j . Di conseguenza,

$$\begin{aligned}\pi_{i,j} &= \binom{M}{n}^{-1} \sum_{s \in A_{i,j} \cap S_n} 1 \\ &= \binom{M}{n}^{-1} |A_{i,j} \cap S_n| \\ &= \frac{\binom{M-2}{n-2}}{\binom{M}{n}},\end{aligned}$$

ovvero

$$\begin{aligned}\pi_{i,j} &= \frac{(M-2)!}{(M-n)!(n-2)!} \frac{n!(M-n)!}{M!} \\ &= \frac{n(n-1)}{M(M-1)}.\end{aligned}$$

Capitolo 3

Statistiche di Horvitz-Thompson

3.1 Stimatore del totale di una popolazione

Definizione 10 Sia $\mathcal{P}_M$ una popolazione finita sulla quale è rilevabile un fenomeno Y e p un disegno di campionamento su $\mathcal{P}_M$. Per $i \in \mathcal{P}_M$ sia π_i la probabilità di inclusione dell'unità i -esima secondo il piano p . Si definisce **stimatore di Horvitz-Thompson** per il totale di Y la seguente statistica

$$\hat{Y}_{Total,HT}(S) = \sum_{i \in S} \frac{Y_i}{\pi_i}.$$

Esempio 10 Per $s = \{2, 1\}$ da $\mathcal{P}_3$ nell'esempio 6, si ha $D(s) = \{(2, 5), (1, 7)\}$, da cui

$$\begin{aligned}\hat{Y}_{Total,HT}(\{2, 1\}) &= \sum_{i \in \{2, 1\}} \frac{Y_i}{\pi_i} \\ &= \frac{Y_2}{\pi_2} + \frac{Y_1}{\pi_1} \\ &= \frac{7}{\pi_2} + \frac{5}{\pi_1}\end{aligned}$$

Se si applica un piano casuale semplice (di ampiezza 2) risulta, per ogni i ,

$$\pi_i = \frac{n}{M} = \frac{2}{3}$$

così

$$\begin{aligned}\hat{Y}_{Total,HT}(\{2, 1\}) &= \frac{3}{2}(7 + 5) \\ &= 18.\end{aligned}$$

Teorema 3 Sia $\mathcal{P}_M$ una popolazione finita sulla quale è rilevabile un fenomeno Y e p un disegno di campionamento su $\mathcal{P}_M$. Per $i, j \in \mathcal{P}_M$ siano π_i la probabilità di inclusione del

primo ordine dell'unità i -esima e $\pi_{i,j}$ quella di secondo ordine per la coppia (i, j) secondo il piano p . Vale allora la seguente proprietà di correttezza:

$$E_p \left[\widehat{Y}_{Total, HT} \right] = \sum_{i \in \mathcal{P}_M} Y_i$$

Inoltre,

$$Var_p \left[\widehat{Y}_{Total, HT} \right] = \sum_{i \in \mathcal{P}_M} \frac{1 - \pi_i}{\pi_i} Y_i^2 + \sum_{i \in \mathcal{P}_M} \sum_{j \in \mathcal{P}_M \setminus \{i\}} \left(\frac{\pi_{i,j}}{\pi_i \pi_j} - 1 \right) Y_i Y_j.$$

Dimostrazione. Dimostriamo solo la correttezza. Possiamo esprimere $\widehat{Y}_{Total, HT}$ come segue:

$$\widehat{Y}_{Total, HT}(S) = \sum_{i \in \mathcal{P}_M} \frac{Y_i}{\pi_i} \delta_i(S),$$

$\widehat{Y}_{Total, HT}$ è quindi una combinazione lineare delle variabili casuali $\Delta_1, \dots, \Delta_M$ i cui coefficienti sono dati da $\frac{Y_1}{\pi_1}, \dots, \frac{Y_M}{\pi_M}$. Il suo valore atteso rispetto a p è quindi pari alla combinazione lineare dei valori attesi delle δ_i :

$$E_p \left[\widehat{Y}_{Total, HT} \right] = E_p \left[\sum_{i \in \mathcal{P}_M} \frac{Y_i}{\pi_i} \Delta_i \right]$$

da cui, ricordando che

$$\pi_i = \sum_{s \in S_{\mathcal{P}_M}} \delta_i(s) p(s) \equiv E_p [\Delta_i],$$

si ha

$$\begin{aligned} E_p \left[\widehat{Y}_{Total, HT} \right] &= \sum_{i \in \mathcal{P}_M} \frac{Y_i}{\pi_i} \pi_i \\ &= \sum_{i \in \mathcal{P}_M} Y_i. \end{aligned}$$

■

3.2 Stimatore della media di una popolazione

La media del fenomeno Y nella popolazione $\mathcal{P}_M$ può essere stimata a partire dallo stimatore Horvitz-Thompson del totale, essendo i due parametri legati dalla relazione

$$Y_{Total} = M(Y_{Mean}), \quad Y_{Total} = \sum_{i \in \mathcal{P}_M} Y_i.$$

Basta definire il seguente stimatore per il parametro media:

$$\widehat{Y}_{Mean,HT}(S) = \frac{1}{M} \widehat{Y}_{Total,HT}(S).$$

Essendo $\widehat{Y}_{Total,HT}$ corretto per il totale, $\widehat{Y}_{Mean,HT}$ risulterà corretto per la corrispondente media. Per la varianza, invece, vale la relazione:

$$Var_p \left[\widehat{Y}_{Mean,HT} \right] = \frac{1}{M^2} Var_p \left[\widehat{Y}_{Total,HT} \right].$$

Capitolo 4

Materiali per l'esame

4.1 Esercitazioni

Esercizio 4.1.1 È data la popolazione finita $\mathcal{P}_3$. Sul corrispondente spazio campionario è definito il seguente piano di campionamento:

$$p(s) = \begin{cases} \frac{1}{6} & \text{se } s = \{1, 2\} \\ \frac{1}{3} & \text{se } s = \{1, 3\} \\ \frac{1}{2} & \text{se } s = \{2, 3\} \\ 0 & \text{altrimenti} \end{cases}$$

Determinare quanto segue.

1. Le probabilità d'inclusione del primo ordine.
2. La varianza della variabile Δ_2 .
3. Le probabilità d'inclusione del secondo ordine.
4. La covarianza tra le variabili Δ_1 e Δ_3 .
5. È stato estratto il campione $s = \{2, 3\}$ e si è rilevato

$$D_s = \{(2, 17), (3, 8)\}.$$

Stimare con la statistica di Horvitz-Thompson il totale di Y su $\mathcal{P}_3$.

6. È stato estratto il campione $s = \{1, 3\}$ e si è rilevato

$$D_s = \{(1, 6), (3, 12)\}.$$

Sapendo che il valore atteso della statistica di Horvitz-Thompson per il totale di Y è pari a 50, determinare Y_2 .

Esercizio 4.1.2 *E' data la popolazione finita $\mathcal{P}_8$. Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice di ampiezza 3.*

Determinare quanto segue.

1. *Le probabilità $\pi_{1,4}$ e $\pi_{7,7}$.*
2. *Il valore atteso della variabile $\Delta_{2,8}$;*
3. *La varianza della variabile Δ_3 .*
4. *La correlazione tra le variabili Δ_1 e Δ_7 .*
5. *È stato estratto il campione $s = \{4, 6, 8\}$ e si è rilevato*

$$D_s = \{(4, 10), (6, 20), (8, 15)\}.$$

Stimare con la statistica di Horvitz-Thompson il totale di Y su $\mathcal{P}_8$.

Esercizio 4.1.3 *E' data la popolazione finita $\mathcal{P}_8$. Sul corrispondente spazio campionario è definito un piano di campionamento p , rispetto al quale risultano noti i seguenti momenti:*

$$\text{Cov}[\Delta_3, \Delta_5] = -1/16, \text{ Var}[\Delta_3] = 3/16, E[\Delta_5] = 1/4.$$

Determinare quanto segue.

1. La probabilità π_5 .
2. La probabilità π_3 sapendo che $\pi_3 > 1/2$.
3. Il valore atteso della variabile $\Delta_{3,5}$.

Esercizio 4.1.4 *E' data la popolazione finita $\mathcal{P}_{20}$. Sul corrispondente spazio campionario è definito un piano di campionamento p . Rispetto a p , delle variabili $X = 3\Delta_7 + 1$ e $Y = -4\Delta_{11} - 5$ sono noti i seguenti momenti:*

$$\text{Cov}[X, Y] = -2/9, E[X] = 2, E[Y] = -6.$$

Determinare quanto segue.

1. Le probabilità π_7 , π_{11} e $\pi_{7,11}$.

Esercizio 4.1.5 *E' data la popolazione finita $\mathcal{P}_9$. Sul corrispondente spazio campionario è definito un piano di campionamento p , rispetto al quale sono noti i seguenti momenti:*

$$\text{Corr}[\Delta_1, \Delta_3] = -1/6, E[\Delta_1] = 4/25, E[\Delta_3] = 1/5.$$

Determinare quanto segue.

1. Il valore atteso della variabile $\Delta_{1,3}$.

Esercizio 4.1.6 *E' data la popolazione finita $\mathcal{P}_M$, con $M > 1$. Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice $p_{C,M,n}$, con $n < M$, rispetto al quale sono note le seguenti probabilità di inclusione:*

$$\forall i, j \in \mathcal{P}_M : \pi_{i,j} = 2/9, \text{ se } i \neq j.$$

$$\forall i \in \mathcal{P}_M : \pi_i = 1/2.$$

Determinare quanto segue.

1. I valori di M e di n .

Esercizio 4.1.7 (A) È data la popolazione finita $\mathcal{P}_M$, con $M > 1$. Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice $p_{C,M,n}$, con $0 < n < M$, rispetto al quale sono note le seguenti probabilità di inclusione:

$$\forall i, j \in \mathcal{P}_M : \pi_{i,j} = 1/2, \text{ se } i \neq j.$$

$$\forall i \in \mathcal{P}_M : \pi_i = 6/8.$$

Determinare quanto segue.

1. I valori di M e di n .
2. La varianza della variabile $\Delta_{1,2}$.

Esercizio 4.1.8 (B) È data la popolazione finita $\mathcal{P}_M$, con $M > 1$. Sul corrispondente spazio campionario è definito un piano di campionamento p . Rispetto a p , definite le variabili $X = 2\Delta_1$ e $Y = (\Delta_2)^2$, sono noti i seguenti momenti:

$$\text{Cov}[-X, Y] = -2/3, E[X] = 1/3, E[Y] = 4/9.$$

Determinare quanto segue.

1. La probabilità di inclusione del primo ordine dell'unità 1 e quella dell'unità 2.
2. La probabilità di inclusione del secondo ordine per la coppia di unità 1 e 2.

Esercizio 4.1.9 (C) Data la popolazione finita $\mathcal{P}_5$, sul corrispondente spazio campionario è definito il seguente piano di campionamento:

$s \in S_{\mathcal{P}_5}$	$\{1, 2, 4\}$	$\{2, 3\}$	$\{1, 3\}$	$\{1, 2\}$	$\{2, 3, 4\}$	$\{2, 4, 5\}$	$\{4, 5\}$	altri
$p(s)$	3/20	3/20	2/20	1/20	6/20	4/20	1/20	0

Siano definiti il fenomeno $Y_i = 1$, per $i \in \mathcal{P}_5$, e la statistica campionaria $T(D_S) = \sum_{i \in S} Y_i$, dove $D_S = \{(i, Y_i) : i \in S\}$.

Determinare quanto segue.

1. L'insieme dei valori che T può assumere, sia C_T .
2. La probabilità dell'evento $T < 1$.
3. Il valore atteso di T .

Esercizio 4.1.10 Sulla popolazione finita $\mathcal{P}_3$ è definito un piano di campionamento tale che:

- (A) $\pi_1 = 1/2$;
 (B) $E[|S|] = 3/2$;
 (C) $E[\Delta_2] = 1/8$.

Determinare quanto segue.

1. La varianza di Δ_3 .

Esercizio 4.1.11 (R) Data la popolazione finita $\mathcal{P}_7$, sul corrispondente spazio campionario è definito il seguente piano di campionamento:

$s \in S_{\mathcal{P}_7}$	$\{1, 2\}$	$\{3, 5, 7\}$	$\{2, 3, 4\}$	$\{4, 5, 7\}$	$\{1, 3, 5\}$	$\{5, 7\}$	$\{1, 2, 5\}$	altri
$p(s)$	1/30	2/30	5/30	3/30	4/30	6/30	9/30	0

Siano definiti il fenomeno $Y_i = i$, per $i \in \mathcal{P}_7$, e la statistica campionaria $T(D_S) = \max_{i \in S} Y_i$.

Determinare quanto segue.

1. L'insieme dei valori che T può assumere, sia C_T .
2. La funzione di probabilità di T .
3. Il valore atteso e la varianza di T .

Esercizio 4.1.12 (α) Sulla popolazione finita $\mathcal{P}_M$ è definito un piano di campionamento tale che T è una variabile casuale così caratterizzata:

$x \in C_T$	-1	1	0	-2	2
$p_T(x)$	2/16	3/16	4/16	1/16	6/16

Determinare, per $W = T^2$, quanto segue.

1. L'insieme dei valori che W può assumere, sia C_W .
2. La funzione di probabilità di W .
3. Il valore atteso di W .
4. La varianza di W .

Esercizio 4.1.13 (β) È data la popolazione finita $\mathcal{P}_M$, con $M > 1$, sulla quale è osservabile un fenomeno Y . Sul corrispondente spazio campionario è definito un piano di campionamento casuale semplice $p_{C,M,n}$, con $0 < n < M$, rispetto al quale sono valide le seguenti condizioni:

$$(A) \text{VAR}[Y_{TOTAL_Y,HT}] = 3/4.$$

$$(B) \text{VAR}[Y_{MEAN_Y,HT}] = 1/48.$$

$$(C) \forall i \in \mathcal{P}_M : \pi_i = 1/2.$$

Determinare quanto segue.

1. I valori di M e di n .

Esercizio 4.1.14 È data la popolazione finita $\mathcal{P}_M$. Sul corrispondente spazio campionario è definito un piano di campionamento p , rispetto al quale sono valide le seguenti condizioni:

$$(A) p \equiv p_{C,M,n}, \text{ con } 0 < n \leq M.$$

$$(B) \Pr(\{1, 3, 15\}) > 0.$$

$$(C) E[\Delta_3] \geq 1/5.$$

Determinare quanto segue.

1. I valori di M e di n .

Esercizio 4.1.15 Sulla popolazione finita $\mathcal{P}_M$ è definito un piano di campionamento tale che T è una variabile casuale così caratterizzata:

$x \in C_T$	-7	0	1
$p_T(x)$	p_{-7}	$2/6$	p_1

Sia inoltre valida la seguente condizione:

$$(A) E[T] = -4/6.$$

Determinare quanto segue.

1. La funzione di probabilità di T .