

ELABORAZIONE STATISTICA DEI DATI DELL'INVENTARIO FORESTALE DELLA REGIONE EMILIA ROMAGNA. MODALITÀ DI CALCOLO.

***PROF.SSA ELISABETTA CARFAGNA
DIPARTIMENTO SCIENZE STATISTICHE
UNIVERSITÀ DI BOLOGNA***

Nella parte preponderante della regione, il piano di campionamento prevede la selezione sistematica dei cluster di aree di saggio. Si ritiene di poter considerare la tecnica di campionamento seguita come un campionamento a due stadi. Si indicano come unità di primo stadio (UPS) i cluster di aree di saggio. Le unità di secondo stadio (USS) sono le aree di saggio.

La stima del totale (così come della corrispondente varianza) viene effettuata singolarmente per ciascuna variabile.

Le unità di primo stadio vengono estratte tramite un campionamento sistematico con inizializzazione non casuale e sono posizionate sui nodi del reticolo chilometrico UTM. Due bacini sono posizionati a cavallo del cambio di fuso; pertanto il grigliato generato dal reticolo chilometrico UTM cambia orientamento nel passare da un fuso all'altro. Ai fini della presente elaborazione, si ignora l'effetto del cambiamento di fuso, in considerazione del fatto che tale cambio di fuso non genera variazioni nel tasso di campionamento delle unità primarie.

In corrispondenza di ciascuna unità primaria, è stata effettuata una fotointerpretazione di foto aeree per superfici di mezzo ettaro. Nel caso in cui il mezzo ettaro corrispondente all'unità primaria è risultato non boscato alla fotointerpretazione, il rilievo a terra non è stato effettuato. Ciononostante, le unità primarie classificate come non boscate dalla fotointerpretazione vanno considerate nel computo, attribuendo loro un valore pari a zero per ciascuna variabile.

Le aree di saggio (le unità di secondo stadio) vengono considerate come punti privi di dimensione estratti all'interno di ciascuna unità primaria.

Il numero delle aree di saggio per unità primaria varia da 4 a 7, per via del fatto che in assenza di bosco sull'area di saggio selezionata, si effettua la rilevazione sull'area di saggio successiva, secondo lo schema di rotazione, nel caso in cui quest'ultima sia boscata; in caso contrario, si continua la rotazione fino alla settima area di saggio. Le aree di saggio periferiche sono poste ad una distanza di 40 metri dall'area di saggio centrale; pertanto, possiamo considerare le unità primarie di campionamento come dei cerchi di raggio 40 metri.

1. Stima del totale e della varianza del totale per i caratteri quantitativi nell'area con selezione sistematica delle unità primarie di campionamento

Sia h l'indice di bacino; $h = 1, 2, \dots, H$,

N_h = numero di unità primarie (cluster) nel bacino h , dato dalla superficie del bacino divisa per l'area di ciascuna unità primaria ($40 \times 40 \times \pi : 10000 =$ circa 0,50265 ettari);

n_h = numero di unità primarie (cluster) selezionate nel bacino h , includendo anche i nodi del grigliato UTM sui quali non sono state effettuate rilevazioni perché non boscati;

M_{hi} = superficie dell' i -esima unità primaria = circa 0,50265 ettari;

m_{hi} = numero di unità secondarie estratte nell' i -esima unità primaria = 4 o più, fino a 7, nel caso venga effettuata la rotazione;

y_{hij} = valore della variabile y ottenuto per la j -esima unità secondaria (area di saggio) della i -esima unità primaria;

Lo stimatore del totale della variabile y quantitativa per l' i -esima unità primaria del bacino h è dato da:

$$y_{hi} = M_{hi} \sum_{j=1}^{m_{hi}} \frac{y_{hij}}{m_{hi}} .$$

Tutti i nodi del reticolo UTM presenti sull'area nella quale il campionamento è sistematico sono stati campionati. Nei casi in cui la fotointerpretazione ha evidenziato assenza di bosco, il rilievo a terra non è stato effettuato. Ritenendo che la fotointerpretazione non abbia effettuato errori di omissione per il bosco, i nodi del reticolo UTM classificati come privi di bosco dalla fotointerpretazione vengono considerati come delle UPC per le quali tutte le variabili riguardanti il bosco assumono valore zero con varianza interna nulla. Naturalmente, sarebbe stato preferibile effettuare la rilevazione a terra almeno su un numero molto esiguo di nodi del reticolo UTM classificati come non boscati, al fine di ottenere una stima del possibile errore di omissione.

Sotto le assunzioni fatte, lo stimatore del totale della variabile y a livello del bacino h è dato da:

$$y_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

Lo stimatore della media per unità primaria della variabile y a livello del bacino h è dato da:

$$y_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

Lo stimatore del totale della variabile y a livello della parte della regione ove il campionamento è sistematico è dato da:

$$y_S = \sum_{h=1}^H y_h = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} =$$

$$= \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi} \sum_{j=1}^{m_{hi}} \frac{y_{hij}}{m_{hi}}$$

Lo stimatore della media per unità primaria della variabile y_S a livello della parte della regione dove il campionamento è sistematico è dato da:

$$\bar{y}_S = \frac{y_S}{\sum_{h=1}^H N_h} = \frac{1}{\sum_{h=1}^H N_h} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi} \sum_{j=1}^{m_{hi}} \frac{y_{hij}}{m_{hi}}$$

Il campionamento sistematico in base ad un grigliato regolare al primo stadio di campionamento dà luogo ad un varianza minore rispetto a quella del campionamento casuale semplice in presenza di autocorrelazione spaziale positiva di ciascuna variabile oggetto d'indagine, supposto che ciascuna variabile non presenti una periodicità coincidente con la distanza tra le unità secondarie estratte.

Si impiega pertanto lo stimatore della varianza per il campionamento casuale senza reintroduzione come sovrastima della varianza per il campionamento sistematico.

Supposto che sussista l'indipendenza tra la localizzazione delle unità di secondo stadio selezionate e le variabili osservate, è possibile evitare l'estrazione casuale della prima unità del campione sistematico.

Uno stimatore della varianza dello stimatore del totale a livello del bacino h è dunque il seguente:

$$v(y_h) = \frac{N_h^2}{n_h} \frac{N_h - n_h}{N_h} \sum_{i=1}^{n_h} \frac{\left(y_{hi} - \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \right)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} \frac{\left(y_{hij} M_{hi} - \frac{\sum_{j=1}^{m_{hi}} y_{hij} M_{hi}}{m_{hi}} \right)^2}{m_{hi} - 1}$$

Uno stimatore dei dati quantitativi ad ettaro a livello del bacino h è dato dal rapporto tra il totale del dato quantitativo nel bacino e il totale della superficie corrispondente:

$$R_{ettaroh} = \frac{Y_h}{\text{sup}_h};$$

la cui varianza è data da:

$$v(R_{ettaroh}) = \frac{1}{\text{sup}_h^2 n_h (n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - R_{ettaroh} y_{suhi})^2$$

Ove sup_h è la superficie corrispondente al dato quantitativo in oggetto, riferita allo strato h .

2. Stima del totale e della varianza del totale per i caratteri rilevati come presenza - assenza (bosco, tipologie forestali, ecc.) nell'area con selezione sistematica delle unità primarie di campionamento

Sia h l'indice di bacino; $h = 1, 2, \dots, H$

N_h, n_h, M_{hi} sono definite come nel caso della stima per i caratteri quantitativi;

P_i = proporzione delle aree di saggio in un cluster ove il carattere è presente = numero delle aree di saggio di un cluster nelle quali è presente la caratteristica diviso il numero totale delle aree di saggio del cluster.

Lo stimatore del totale delle caratteristiche rilevate come presenza - assenza per l'i-esima unità primaria del bacino h è dato da:

$$y_{hi} = M_{hi} p_i$$

Lo stimatore del totale della variabile y a livello del bacino h è dato da:

$$y_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

Lo stimatore della media per unità primaria della variabile y a livello del bacino h è dato da:

$$\bar{y}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$$

Uno stimatore della varianza dello stimatore del totale a livello del bacino h è il seguente:

$$v(y_h) = \frac{N_h^2}{n_h} \frac{N_h - n_h}{N_h} \sum_{i=1}^{n_h} \frac{\left(y_{hi} - \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \right)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 \frac{1}{m_{hi} - 1} p_i q_i$$

Dove q_i è la proporzione di aree di saggio di un cluster sulle quali la caratteristica non viene riscontrata (è pari a $1 - p_i$)

3. Stima del totale e della varianza del totale per i caratteri quantitativi per il bacino del Panaro, dove la selezione delle unità di primo stadio è casuale stratificata

Nel bacino del Panaro è stata operata una suddivisione dei fotopunti in tre strati, uno dei quali è costituito dai fotopunti interpretati come non boscati e su di esso nessun rilievo a terra è stato effettuato. Pertanto, l'elaborazione delle aree di saggio può essere effettuata

soltanto sugli altri due strati. Lo strato considerato non boscato in base alla fotointerpretazione sarà da considerare completamente privo di bosco e quindi anche di tutte le sue caratteristiche. Purtroppo, non essendo stato effettuato alcun rilievo a terra in tale strato, non è possibile stimare in alcun modo un errore a tale dato di assenza di bosco. Per quanto riguarda i rimanenti due strati, siano:

l l'indice di strato all'interno del bacino del Panaro; $l = 1, 2, \dots, L$

N_l = numero di unità primarie (cluster) nello strato l , dato dalla superficie dello strato divisa per l'area di ciascuna unità primaria ($40 \times 40 \times \pi : 10000 =$ circa 0,50265 ettari). In assenza delle superfici degli strati, si farà ricorso al numero di fotopunti attribuiti allo strato l moltiplicato per 4 ettari;

n_l = numero di unità primarie (cluster) selezionate nello strato l ;

M_{li} = superficie dell' i -esima unità primaria = circa 0,50265 ettari;

m_{li} = numero di unità secondarie estratte nell' i -esima unità primaria = 4 o più, fino a 7, nel caso venga effettuata la rotazione;

y_{lij} = valore della variabile y ottenuto per la j -esima unità secondaria (area di saggio) della i -esima unità primaria;

Lo stimatore del totale della variabile y quantitativa per l' i -esima unità primaria dello strato l è dato da:

$$y_{li} = M_{li} \sum_{j=1}^{m_{li}} \frac{y_{lij}}{m_{li}} .$$

Lo stimatore del totale della variabile y a livello dello strato l è dato da:

$$y_l = \frac{N_l}{n_l} \sum_{i=1}^{n_l} y_{li}$$

Lo stimatore della media per unità primaria della variabile y a livello dello strato l è dato da:

$$\bar{y}_l = \sum_{i=1}^{n_l} \frac{y_{li}}{n_l}$$

Lo stimatore del totale della variabile y a livello della parte del bacino del Panaro ove sono stati effettuati rilievi a terra è dato da:

$$y_{pa} = \sum_{l=1}^L y_l = \sum_{l=1}^L \sum_{i=1}^{n_l} y_{li} \frac{N_l}{n_l} =$$

$$= \sum_{l=1}^L \frac{N_l}{n_l} \sum_{i=1}^{n_l} M_{li} \sum_{j=1}^{m_{li}} \frac{y_{lij}}{m_{li}}$$

Lo stimatore della media per unità primaria della variabile y_{pa} a livello della parte del bacino del Panaro dove sono stati effettuati rilievi a terra è dato da:

$$\bar{y}_{pa} = \frac{y_R}{\sum_{l=1}^L N_l} = \frac{1}{\sum_{l=1}^L N_l} \sum_{l=1}^L \frac{N_l}{n_l} \sum_{i=1}^{n_l} M_{li} \sum_{j=1}^{m_{li}} \frac{y_{lij}}{m_{li}}$$

Uno stimatore della varianza dello stimatore del totale a livello dello strato l è dunque il seguente:

$$v(y_l) = \frac{N_l^2}{n_l} \frac{N_l - n_l}{N_l} \sum_{i=1}^{n_l} \frac{\left(y_{li} - \frac{\sum_{i=1}^{n_l} y_{li}}{n_l} \right)^2}{n_h - 1} +$$

$$\frac{N_l}{n_l} \sum_{i=1}^{n_l} \frac{1}{m_{li}} \sum_{j=1}^{m_{li}} \frac{\left(y_{lij} M_{li} - \sum_{j=1}^{m_{li}} \frac{y_{lij} M_{li}}{m_{li}} \right)^2}{m_{li} - 1}$$

Uno stimatore della varianza del totale a livello della parte del bacino del Panaro sulla quale sono stati fatti rilievi a terra è dato da:

$$v(y_{pa}) = \sum_{l=1}^L V(y_l)$$

Nel bacino del Panaro, uno stimatore dei dati quantitativi ad ettaro a livello di strato l , è dato dal rapporto tra il totale del dato quantitativo nello strato e il totale della superficie corrispondente:

$$R_{ettarol} = \frac{Y_l}{\text{sup}_l};$$

la cui varianza è data da:

$$v(R_{ettarol}) = \frac{1}{\text{sup}_l^2 n_l(n_l - 1)} \sum_{i=1}^{n_l} (y_{li} - R_{ettarol} y_{suli})^2$$

Ove sup_l è la superficie corrispondente al dato quantitativo in oggetto, riferita allo strato l . La stima dei dati ad ettaro per l'intero bacino del Panaro è data da:

$$R_{ettaropa} = \sum_{l=1}^L \frac{N_l}{N_{pa}} \frac{Y_l}{\text{sup}_l}$$

Ove N_{pa} è la somma delle superfici degli strati.

$$v(R_{ettaropa}) = \sum_{l=1}^L \frac{N_l^2}{N_{pa}^2} \frac{1}{\text{sup}_l^2 n_l(n_l - 1)} \sum_{i=1}^{n_l} (y_{li} - R_{ettarol} y_{suli})^2$$

4. Stima del totale e della varianza del totale per i caratteri rilevati come presenza - assenza (bosco, tipologie forestali, ecc.) nel bacino del Panaro

Sia l l'indice di strato; $l = 1, 2, \dots, L$

N_l, n_l, M_{li}, m_{li} sono definite come nel caso della stima per le variabili;

P_i = proporzione delle aree di saggio in un cluster ove il carattere è presente = numero delle aree di saggio di un cluster nelle quali è presente la caratteristica diviso il numero totale delle aree di saggio del cluster.

Lo stimatore del totale delle caratteristiche rilevate come presenza - assenza per l' i -esima unità primaria dello strato l è dato da:

$$y_{li} = M_{li} P_i$$

Lo stimatore del totale della variabile y a livello dello strato l è dato da:

$$y_l = \sum_{i=1}^{n_l} y_{li} \frac{N_l}{n_l}$$

Lo stimatore della media per unità primaria della variabile y a livello dello strato l è dato da:

$$\bar{y}_l = \sum_{i=1}^{n_l} \frac{y_{li}}{n_l}$$

Lo stimatore del totale della variabile y a livello della parte del bacino del Panaro dove sono stati effettuati rilievi a terra è dato da:

$$\begin{aligned} y_{pa} &= \sum_{l=1}^L y_l = \sum_{l=1}^L \sum_{i=1}^{n_l} y_{li} \frac{N_l}{n_l} = \\ &= \sum_{l=1}^L \frac{N_l}{n_l} \sum_{i=1}^{n_l} M_{li} P_i \end{aligned}$$

Lo stimatore della media per unità primaria della variabile y a livello della parte del bacino del Panaro dove il sono stati effettuati rilievi a terra è dato da:

$$\bar{y}_{pa} = \frac{y}{\sum_{l=1}^L N_l} = \frac{1}{\sum_{l=1}^L N_l} \sum_{l=1}^L \frac{N_l}{n_l} \sum_{i=1}^{n_l} M_{li} p_i$$

Uno stimatore della varianza dello stimatore del totale a livello dello strato l è il seguente:

$$v(y_l) = \frac{N_l^2}{n_l} \frac{N_l - n_l}{N_l} \sum_{i=1}^{n_l} \frac{\left(y_{li} - \frac{\sum_{i=1}^{n_l} y_{li}}{n_l} \right)^2}{n_l - 1} + \frac{N_l}{n_l} \sum_{i=1}^{n_l} M_{li}^2 \frac{1}{m_{li} - 1} p_i q_i$$

Dove q_i è la proporzione di aree di saggio di un cluster sulle quali la caratteristica non viene riscontrata.

Uno stimatore della varianza del totale a livello della parte del bacino del Panaro dove sono stati effettuati rilievi a terra è dato da:

$$v(y_{pa}) = \sum_{l=1}^L V(y_l)$$

5. Stima del totale e della varianza del totale a livello dell'intera regione Emilia Romagna

Il totale per l'intera regione Emilia Romagna, per i caratteri rilevati come presenza-assenza, viene stimato da:

$$y = y_s + y_{pa}$$

Mentre la varianza dello stimatore del totale per l'intera regione è dato da:

$$V(y) = v(y_s) + v(y_{pa})$$

Per l'intera regione Emilia Romagna, le stime per i dati quantitativi ad ettaro sono ottenute nel modo seguente:

$$R_{ettaro} = \frac{N_s}{N} R_{ettaros} + \frac{N_{pa}}{N} R_{ettaropa}$$

Ove $R_{ettaros}$ è il dato ad ettaro per l'intera area di allocazione sistematica-

$$V(R_{ettaro}) = \frac{N_s^2}{N^2} V(R_{ettaros}) + \frac{N_{pa}^2}{N^2} V(R_{ettaropa})$$

6. Stima del totale e della varianza del totale a livello di aree che intersecano più bacini (ad esempio comunità montane)

Nel caso in cui l'area a al livello della quale vanno effettuate delle stime risulta essere un'aggregazione di bacini o parti di bacini ove il campionamento è sistematico, lo stimatore del totale è dato da:

$$y_a = \frac{N_a}{n_a} \sum_{i=1}^{n_a} y_{ai} = \frac{N_a}{n_a} \sum_{i=1}^{n_a} M_{ai} \sum_{j=1}^{m_{ai}} \frac{y_{aij}}{m_{ai}}$$

dove

N_a = numero di unità primarie (cluster) nell'area a , dato dalla superficie dell'area divisa per la superficie di ciascuna unità primaria ($40 \times 40 \times \pi : 10000 =$ circa 0,50265 ettari);

n_a = numero di unità primarie (cluster) ricadenti nell'area a , includendo anche i nodi del grigliato UTM sui quali non sono state effettuate rilevazioni;

M_{ai} = superficie dell' i -esima unità primaria = circa 0,50265 ettari;

m_{ai} = numero di unità secondarie estratte nell' i -esima unità primaria = 4 o più, fino a 7, nel caso venga effettuata la rotazione;

y_{aij} = valore della variabile y ottenuto per la j -esima unità secondaria (area di saggio) della i -esima unità primaria.

Lo stimatore della varianza della stima a livello dell'area a è dato da:

$$v(y_a) = \frac{N_a^2}{n_a} \frac{N_a - n_a}{N_a} \sum_{i=1}^{n_a} \frac{\left(y_{ai} - \frac{\sum_{i=1}^{n_h} y_{ai}}{n_a} \right)^2}{n_a - 1} + \frac{N_a}{n_a} \sum_{i=1}^{n_a} \frac{1}{m_{ai}} \sum_{j=1}^{m_{ai}} \frac{\left(y_{aij} M_{ai} - \sum_{j=1}^{m_{ai}} \frac{y_{aij} M_{ai}}{m_{ai}} \right)^2}{m_{ai} - 1}$$

Nel caso in cui l'area a al livello della quale vanno effettuate delle stime risulta essere un'aggregazione anche del bacino del Panaro o parte di esso, le stime vanno effettuate separatamente per la parte sistematica e per la porzione del bacino del Panaro. Per la parte sistematica gli stimatori del totale e della varianza del totale coincidono con quelli appena descritti. Per la parte del Panaro, si seguono formulazioni del tutto analoghe a quelle riportate nei paragrafi 3 e 4, considerando però le sole unità primarie ricadenti nella porzione di bacino del Panaro appartenente all'area a per la quale si vogliono fornire le stime.

7. Stima del totale e della varianza del totale tramite i soli fotopunti

Sui vertici di un grigliato regolare che distano tra loro 200 metri, dei punti sono stati allocati in modo sistematico e fotointerpretati (in verità cerchi di mezzo ettaro di superficie). Tali punti sono detti fotopunti. Il piano di campionamento prevede dunque una selezione sistematica di unità areali circolari di mezzo ettaro di superficie. Come detto precedentemente, nella regione Emilia Romagna avviene il cambio di fuso, con il conseguente cambiamento di orientamento del grigliato; tuttavia, si è deciso di ignorare il problema a causa della impossibilità di calcolare con esattezza l'area ricadente in ciascuno dei due fusi.

7.1 Stima del totale e della varianza del totale per le variabili quantitative non riferite all'ettaro, tramite fotointerpretazione dei fotopunti

Sia h l'indice di bacino; $h = 1, 2, \dots, H$,

N_h = numero di unità areali nella popolazione (l'intero bacino), dato dalla superficie del bacino divisa per l'area di ciascun fotopunto (mezzo ettaro);

n_h = numero di fotopunti fotointerpretati nel bacino h ;

Lo stimatore del totale della variabile y a livello del bacino h è dato da:

$$y_{fh} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

dove y_{hi} è il valore presentato dal generico fotopunto del bacino h ;

Lo stimatore della media per unità primaria della variabile y a livello del bacino h è dato da:

$$\bar{y}_{fh} = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

Uno stimatore della varianza dello stimatore del totale a livello del bacino h è il seguente:

$$v(y_{fh}) = \frac{N_h^2}{n_h} \frac{N_h - n_h}{N_h} \sum_{i=1}^{n_h} \frac{\left(y_{hi} - \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \right)^2}{n_h - 1}$$

Lo stimatore del totale della variabile y a livello della regione è dato da:

$$y_f = \frac{N}{n} \sum_{i=1}^n y_i$$

dove y_i è il valore presentato dal generico fotopunto della regione;

Lo stimatore della media per unità primaria della variabile y a livello della regione è dato da:

$$\bar{y}_f = \frac{\sum_{i=1}^n y_i}{n}$$

Uno stimatore della varianza dello stimatore del totale a livello dell'intera regione è il seguente:

$$v(y_f) = \frac{N^2}{n} \frac{N-n}{N} \sum_{i=1}^n \frac{\left(y_i - \frac{\sum_{i=1}^n y_i}{n} \right)^2}{n-1}$$

Nel caso in cui si desideri effettuare la stima a livello di una qualsiasi area all'interno della regione, le formulazioni vengono modificate soltanto prendendo in considerazione la superficie ed il numero dei fotopunti fotointerpretati dell'area considerata.

7.2 Stima del totale e della varianza del totale per le variabili rilevate come presenza - assenza tramite fotointerpretazione dei fotopunti

Uno stimatore del totale per il bacino h per una variabile rilevata tramite fotointerpretazione dei fotopunti come presenza-assenza, è dato da:

$$y_{fh} = N_h P_h$$

dove:

N_h = numero di unità areali nella popolazione (l'intero bacino), dato dalla superficie del bacino divisa per l'area di ciascun fotopunto (mezzo ettaro);

P_h = proporzione dei fotopunti del bacino che presentano la caratteristica = numero di fotopunti del bacino che presentano la caratteristica diviso il numero dei fotopunti fotointerpretati nel bacino.

Lo stimatore della media per fotopunto della variabile y a livello del bacino è dato da

$$P_h .$$

Uno stimatore della varianza del totale per il bacino h , sulla base dei soli fotopunti è dato da:

$$v(y_{fh}) = \frac{N_h(N_h - n_h)}{n_h} P_h Q_h$$

dove:

n_h = numero di fotopunti fotointerpretati nel bacino h ;

Q_h = proporzione dei fotopunti del bacino che non presentano la caratteristica = $1 - P_h$

Uno stimatore del totale per l'intera regione per una variabile rilevata tramite fotointerpretazione dei fotopunti come presenza-assenza, è dato da:

$$y_f = N p$$

dove:

N = numero di unità areali nella popolazione (l'intera regione), dato dalla superficie della regione divisa per l'area di ciascun fotopunto (mezzo ettaro);

p = proporzione dei fotopunti della regione che presentano la caratteristica = numero di fotopunti della regione che presentano la caratteristica diviso il numero dei fotopunti fotointerpretati nella regione.

Lo stimatore della media per fotopunto della variabile y a livello della regione è dato da P .

Uno stimatore della varianza del totale per la regione, sulla base dei soli fotopunti è dato da:

$$v(y_f) = \frac{N(N-n)}{n} p q$$

dove:

n = numero di fotopunti fotointerpretati nella regione;

q = proporzione dei fotopunti della regione che non presentano la caratteristica = $1 - p$.

Nel caso in cui si desideri effettuare la stima a livello di una qualsiasi area all'interno della regione, le formulazioni vengono modificate soltanto prendendo in considerazione la superficie, il numero dei fotopunti fotointerpretati ed il numero di fotopunti che presentano la caratteristica dell'area considerata.

8. Dati mancanti

Alcuni fotopunti classificati a bosco e selezionati hanno generato delle UPC sulle quali non è stato poi effettuato il rilievo. Per tali UPC abbiamo dei dati completamente mancanti; pertanto, per la stima dei totali, non entrano nel computo, sia per quanto riguarda le sommatorie, sia per la divisione per n_h , n_l o n_a a seconda delle formule.

In alcuni casi (ad esempio per le aree di saggio inaccessibili) mancano i dati soltanto per alcune variabili, quindi il numero di dati mancanti varia al variare della variabile per la quale effettuare la stima. Naturalmente, per i fotopunti fotointerpretati come non boscati, non si verificano mai dati mancanti; pertanto, per evitare distorsioni nelle stime, è opportuno sottrarre al numero di fotopunti corrispondenti alle UPC non boscate un numero di UPC proporzionale al numero di UPC boscate per le quali mancano i dati.

Contrariamente a quanto avviene nel calcolo dei totali, nel calcolo delle varianze non si sottrae al numero di fotopunti corrispondenti alle UPC non boscate un numero di UPC proporzionale al numero di UPC boscate per le quali mancano i dati.

Se per una variabile mancano i dati su un'area di saggio all'interno di un'UPC, tale area di saggio non viene considerata nel novero delle aree di saggio.

Naturalmente, il procedimento di trattamento di dati mancanti descritto presuppone che non ci sia relazione tra la mancanza del dato ed il valore o l'attributo che tale dato presenterebbe se fosse rilevato. Tale ipotesi è scarsamente realistica; tuttavia, indagini successive rivolte allo studio delle caratteristiche dei dati mancanti non sono state effettuate; pertanto non abbiamo elementi per formulare ipotesi alternative a quella adottata.

9. Stima della superficie delle diverse tipologie forestali

Le superfici stimate tramite la fotointerpretazione dei fotopunti risultano inevitabilmente diverse rispetto a quelle stimate tramite i rilievi a terra, in quanto la fotointerpretazione viene condotta ad una scala diversa rispetto al rilievo a terra. La metodologia seguita nell'effettuazione dei rilievi a terra ha consentito di riscontrare casi in cui le UPC erano state fotointerpretate come boscate e tuttavia non risultavano essere tali al rilievo a terra. Inoltre, durante il rilievo a terra si tiene conto delle aree non a bosco all'interno di cluster classificati come boscati. Purtroppo, nessun punto chilometrico, classificato come non boscato, è stato rilevato a terra e dunque non è stato possibile riscontrare l'eventuale presenza di bosco su UPC fotointerpretate come non boscate; pertanto la procedura di stima tramite i soli dati a terra dà luogo inevitabilmente ad una sottostima della superficie boscata e delle superfici delle varie tipologie boschive. Si ritiene dunque opportuno utilizzare i dati a terra al fine di individuare la frazione delle varie tipologie boschive sul totale della superficie boscata; tale frazione viene moltiplicata per la stima della superficie totale boscata derivante dalla fotointerpretazione.

In sostanza, il dato di superficie dell'area boscata si ritiene più attendibile di quello derivante dai rilievi a terra, il cui ruolo diviene quello di fornire dati ad una scala di maggiore dettaglio. In particolare, per quanto riguarda la stima delle superfici delle varie tipologie forestali, si stima il rapporto tra ciascuna tipologia forestale e l'area boscata (stimata tramite i rilievi a terra) e si moltiplica tale rapporto per la superficie boscata stimata tramite i fotopunti.

Sia

$$y_{ht} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hti}$$

lo stimatore della superficie della tipologia forestale t stimata tramite i rilievi a terra per il bacino h ;

$$y_{hb} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hbi}$$

lo stimatore della superficie boscata tramite i rilievi a terra per il bacino h ,

$$R_{ht} = \frac{\frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hti}}{\frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hbi}}$$

il rapporto tra y_{ht} e y_{hb} .

$$y_{htRf} = R_{ht} N_h p_h$$

è lo stimatore della superficie della tipologia t tramite rilievi a terra e fotointerpretazione,

nello strato h ; $N_h p_h$ corrisponde alla superficie boscata stimata tramite fotointerpretazione dei fotopunti.

Lo stimatore della varianza di y_{htRf} è dato da:

$$V(y_{htRf}) = \frac{N_h^2}{n_h(n_h - 1)} \frac{N_h - n_h}{N_h} \sum_{i=1}^{n_h} (y_{hti} - R_{ht} y_{hbi})^2$$

Per il bacino del Panaro, le formulazioni per la stima della superficie della tipologia t e della sua varianza seguono le analoghe formulazioni; pertanto, la stima della superficie per la tipologia t a livello della regione è data da:

$$y_{tRf} = y_{stRf} + y_{patRf}$$

Ove y_{stRf} è lo stimatore del totale per tipologia forestale tramite dati a terra e fotointerpretazione per la parte della Regione per la quale il campionamento è sistematico.

Mentre la varianza dello stimatore della superficie della tipologia t per l'intera regione è data da:

$$V(y_{tRf}) = v(y_{stRf}) + v(y_{patRf}).$$