

Strategia di campionamento e valutazione degli errori campionari¹

1. Introduzione

La *popolazione di interesse* dell'indagine è costituita dalle famiglie residenti in Italia e dagli individui di 14 anni e più che le compongono. Sono esclusi gli individui che sono membri permanenti delle convivenze. Per famiglia si intende la *famiglia di fatto*, ovvero un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

L'indagine viene svolta mediante intervista telefonica e utilizza come lista di selezione l'archivio degli abbonati Telecom al telefono; le *unità di campionamento* sono, pertanto, i numeri telefonici appartenenti a detto archivio.

L'indagine è di tipo trasversale a cadenza quinquennale e ha la finalità di fornire stime di parametri di diversa natura (totali, medie, rapporti, frequenze assolute e relative) con diversi riferimenti territoriali:

- l'intero territorio nazionale;
- le cinque ripartizioni geografiche (Nord-Est, Nord-Ovest, Centro, Sud e Isole);
- le regioni geografiche;
- sei aree basate sulla tipologia socio-demografica dei comuni, così definite:

A, *area metropolitana* suddivisa in :

A₁ , comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;

A₂ , comuni che gravitano intorno al centro dell'area metropolitana;

B, *area non metropolitana* suddivisa in :

B₁ , comuni con oltre 50.000 abitanti;

B₂ , comuni con 10.001-50.000 abitanti

B₃ , comuni con 2.001-10.000 abitanti;

B₄ , comuni aventi fino a 2.000 abitanti

2. Caratteristiche della lista di selezione

La *base di campionamento* adottata, ovvero la lista di selezione delle unità campionarie, è l'*archivio informatizzato ufficiale delle famiglie abbonate alla Telecom*. Tale scelta è motivata dal fatto che le informazioni dell'archivio in oggetto sono contenute in un *file* che viene costantemente aggiornato sulle variazioni degli intestatari e degli indirizzi telefonici; esso è, inoltre, di agevole utilizzo per la selezione delle unità campionarie in quanto si presta facilmente alla scelta di diversi criteri di ordinamento.

Le informazioni relative a ciascun indirizzo, utilizzabili per la stratificazione delle unità della popolazione di riferimento, sono essenzialmente di tipo territoriale; esse sono la provincia, il comune, la sezione di censimento, la via, il numero civico, l'ampiezza del comune di appartenenza, in termini demografici e in termini di numero di indirizzi.

I principali problemi che derivano dall'utilizzo della lista telefonica come base di campionamento per le indagini sulle famiglie, sono legati alla non perfetta coincidenza tra la popolazione oggetto d'indagine e l'insieme delle unità contenute nella lista; indicando con L ed U rispettivamente la lista e la popolazione d'interesse, le possibili situazioni in cui gli insiemi L e U differiscono sono in generale le seguenti:

- i) *sovracopertura*, quando alcuni elementi dell'insieme L non appartengono all'insieme U;

¹ Di Claudia De Vitiis e Stefano Falorsi

- ii) *sottocopertura*, nel caso in cui alcuni elementi di U non sono contenuti in L ;
- iii) *duplicazione di alcune unità*, se alcuni elementi di U sono presenti più volte in L ;
- iv) *grappoli di unità*, quando alcuni elementi dell'insieme L contengono grappoli di elementi dell'insieme U .

La lista telefonica utilizzata per l'indagine presenta in varia misura tutte le situazioni elencate; tuttavia, i fenomeni che appaiono più rilevanti sono quelli della *sovracopertura* e della *sottocopertura*. Rientrano, infatti, nel caso della *sovracopertura* tutti i numeri telefonici che corrispondono a seconde case o ad attività professionali oppure a pubblici esercizi, ecc.; mentre rientrano nel caso della *sottocopertura* tutte le famiglie non abbonate al telefono o intestatarie di numeri riservati. Rientrano, poi, nel caso iii) le famiglie intestatarie di più di un numero riferito all'abitazione principale e nel caso iv) i numeri telefonici a cui corrispondono più famiglie.

L'effetto di ciascuna delle situazioni appena elencate è quella di modificare in modo non controllato le *probabilità di inclusione* delle unità della popolazione rispetto alle probabilità di inclusione *teoriche* assegnate alle stesse unità dal disegno di campionamento adottato. Ciò può essere causa di distorsioni nelle stime prodotte dall'indagine.

In particolare la *sottocopertura* determina una distorsione delle stime prodotte tanto maggiore quanto più le unità di U che appartengono ad L sono differenti dalle unità di U non presenti in L relativamente alle variabili d'interesse dell'indagine. E' possibile in parte attenuare tale effetto distorsivo ricorrendo a opportune procedure di stima, note in letteratura come stimatori di *ponderazione vincolata*, che tengono conto di totali noti (sulla base di fonti esterne all'indagine) sulla popolazione relativi a variabili ausiliarie correlate a quelle d'interesse.

Un ulteriore effetto dovuto alla *sovracopertura*, è il fatto che la numerosità effettiva del campione sia minore di quella definita nella fase di progettazione del disegno e ciò comporta un aumento della variabilità delle stime. Si tratta allora di decidere se accettare una dimensione campionaria inferiore a quella prevista o, alternativamente, ricorrere a un sovra-dimensionamento del campione o a una lista di unità sostitutive. Nelle indagini di tipo telefonico si ricorre in genere alla sostituzione in quanto è necessario garantire che venga raggiunto il numero prefissato di interviste utili.

3. Descrizione del disegno di campionamento

Il disegno di campionamento è a *due stadi* con stratificazione delle unità di primo stadio. Le unità di primo stadio sono gli indirizzi telefonici dell'archivio di selezione, e quindi le famiglie ad essi corrispondenti. Le unità di secondo stadio sono gli individui: per ciascuna famiglia selezionata al primo stadio si seleziona un individuo campione tra i componenti della famiglia aventi età pari o superiore a 14 anni.

Gli indirizzi telefonici sono stati stratificati per regione geografica e per tipologia di comune. La tipologia di comune non è, ovviamente, una variabile presente sull'Archivio Telefonico Informatizzato delle Famiglie, in quanto viene definita dall'ISTAT sulla base delle caratteristiche socio-demografiche dei comuni italiani rilevate al Censimento della Popolazione. La stratificazione mediante tale variabile ha comportato, quindi, la necessità di integrare l'archivio con tali informazioni.

Il disegno campionario ora descritto presenta caratteristiche molto differenti rispetto ai disegni che vengono correntemente utilizzati per le indagini sulle famiglie condotte mediante intervista diretta. In primo luogo, l'indagine in esame prevede la selezione diretta degli indirizzi campione senza estrarre al primo stadio un campione di comuni; ciò come è noto determina un aumento dell'efficienza delle stime prodotte e tale guadagno è tanto più elevato quanto più è alto il coefficiente di correlazione intra-classe, ovvero il grado di omogeneità delle variabili di interesse all'interno dei comuni. Inoltre, l'indagine prevede la selezione di un solo individuo campione da

ciascuna famiglia; anche questa scelta conduce a stime tanto più efficienti quanto più è alto il coefficiente di correlazione intra-classe delle famiglie.

La determinazione del numero totale di unità campionarie e la sua allocazione tra gli strati è in genere, per un'indagine ad obiettivi plurimi come quella in esame, un'operazione complessa. E' poco realistico, infatti, pensare di poter definire un campione che assicuri prefissati livelli di precisione a tutte le stime d'interesse, considerando anche il fatto che le stime vengono prodotte con diversi riferimenti territoriali. L'allocazione ottimale delle unità del campione con riferimento ad un dato tipo di dominio può risultare contrastante con l'allocazione ottimale con riferimento ad un altro tipo di dominio. In particolare, per quanto riguarda le stime riferite all'intero territorio nazionale l'allocazione ottimale risulta vicina a quella proporzionale tra le diverse regioni; per quanto riguarda, invece, le stime riferite alle regioni, l'allocazione ottimale risulta prossima a quella che assegna a tutte le regioni un campione di uguale numerosità. E' necessario quindi di un procedimento complesso articolato in più fasi.

Dapprima, mediando tra esigenze operative e di costo ed esigenze relative all'attendibilità delle principali stime di interesse, si è definita la numerosità n complessiva del campione. Successivamente, sulla base di valutazioni dell'errore di campionamento atteso delle principali stime a livello regionale e nazionale, è stata determinata l'allocazione del campione tra le regioni; si è ottenuta in tal modo un'allocazione di compromesso tra l'allocazione *uguale* e quella *proporzionale*. Infine, le numerosità campionarie regionali sono state ripartite tra le diverse tipologie di comune in modo proporzionale alla popolazione residente.

Nel prospetto 1 sono riportate le numerosità campionarie per regione.

Prospetto 1. Distribuzione regionale del campione

Regioni	Indirizzi campione
Piemonte	3.796
Valle d'Aosta	1.328
Lombardia	4.622
Bolzano	1.393
Trento	1.518
Veneto	2.854
Friuli V. G.	2.013
Liguria	2.216
Emilia Romagna	2.800
Toscana	3.024
Umbria	1.526
Marche	1.995
Lazio	3.138
Abruzzo	2.068
Molise	1.368
Campania	3.036
Puglia	2.610
Basilicata	1.407
Calabria	2.197
Sicilia	2.986
Sardegna	2.105
ITALIA	50.001

4. Selezione del campione

L'estrazione degli indirizzi campione da ciascuno strato è stata effettuata con probabilità uguali e senza reimmissione mediante tecnica di selezione sistematica.

La selezione del campione di indirizzi è avvenuta nel seguente modo:

- per ciascuna regione, con riferimento ad una data tipologia comunale, sono stati ordinati gli indirizzi dell'Archivio Telefonico delle Famiglie per ampiezza demografica del comune, strada, numero civico e piano;
- definito il passo di estrazione regionale, sono stati selezionati 6 punti di partenza casuali, uno per ciascuna tipologia comunale. Si è proceduto, poi, alla selezione degli indirizzi campione sulla base del passo di estrazione regionale a partire dal punto di partenza casuale;
- ogni indirizzo selezionato dall'Archivio è entrato a far parte dell'elenco base, e per ognuno di essi si sono selezionati i tre indirizzi successivi, che sono entrati a far parte dell'elenco sostitutivo; in tal modo le unità sostitutive sono le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, sono anche simili dal punto di vista delle caratteristiche socio-economiche.

La selezione casuale del componente da intervistare è stata effettuata nel seguente modo:

- a) è stato chiesto alla persona rispondente al telefono l'elenco dei componenti della famiglia, e per ciascuno di essi il sesso, l'età e la relazione con il capo famiglia;
- b) è stata formata la graduatoria per età dei componenti, assegnando un numero progressivo ai componenti risultati *intervistabili* (età maggiore di 14 anni);
- c) è stato estratto un numero casuale compreso tra 1 ed il numero totale dei componenti *intervistabili*;
- d) è stato intervistato il componente il cui numero in graduatoria coincide con il numero casuale estratto.

Si fa presente che ai fini dell'individuazione del componente da intervistare l'ISTAT ha fornito alla ditta che effettua la rilevazione un file costituito da 50.000 record in cui ciascun record è così formato: codice di regione, tipologia di comune, numero progressivo da 1 al numero di indirizzi campione nel dominio; 9 campi successivi così formati: il primo campo contiene un numero casuale pari a 1 o a 2; il secondo campo contiene un numero casuale pari a 1, 2 o 3 e così via fino all'ultimo campo che contiene un numero casuale pari a 1, 2, 3..., 10. Per la selezione dell'individuo da intervistare verrà utilizzato il primo campo per le famiglie di due componenti, il secondo per le famiglie di tre componenti e così via. Ciascun numero casuale è estratto indipendentemente dagli altri e pertanto, nell'ambito di ciascun campo k ($k=1, \dots, 9$), ogni numero i ($i=1, \dots, k+1$) compare lo stesso numero di volte nei 50.000 record. In tal modo, ciascun componente delle famiglie di ampiezza $k+1$ ($k=1, \dots, 9$) ha la stessa probabilità di essere selezionato.

5. Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono stime del numero di eventi delittuosi subiti dagli individui e del numero di eventi delittuosi subiti dalle famiglie, nonché le caratteristiche strutturali degli individui e delle famiglie che hanno subito tali fatti delittuosi.

Le stime sono ottenute mediante uno stimatore di ponderazione vincolata, che è il metodo di stima standard per la maggior parte delle indagini ISTAT sulle imprese e sulle famiglie.

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione.

Questo principio viene realizzato attribuendo ad ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentate dall'unità medesima. Se, ad esempio, ad un'unità campionaria viene attribuito un peso pari a 100, vuol dire che questa unità rappresenta se stessa ed altre 99 unità della popolazione che non sono state incluse nel campione.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia: d , indice di livello territoriale di riferimento delle stime; h , indice di strato; j , indice di famiglia; q indice di individuo all'interno della famiglia j ; y , generica variabile oggetto di indagine; Y_{hjp} valore

di y osservato sull'individuo p della famiglia j dello strato h ; Q_{hj} , numero di individui di 14 anni e oltre appartenenti alla famiglia j dello strato h ; M_h , numero di famiglie residenti nello strato h ; m_h , campione di famiglie nello strato h ; p_h , numero di individui campione nello strato h (dal momento che si intervista un unico individuo in ciascuna famiglia campione si ha $p_h = m_h$); H_d , numero di strati nel dominio d .

Ipotizziamo di voler stimare, con riferimento ad un generico dominio d (ad esempio una regione geografica) il totale della variabile y oggetto di indagine (ad esempio il numero totale di scippi subiti dagli individui di 14 anni e oltre), espresso dalla seguente relazione:

$${}_d Y = \sum_{h=1}^{H_d} \sum_{j=1}^{M_h} \sum_{q=1}^{Q_{hj}} Y_{hj} \quad (1)$$

Una stima del totale (1) è data dalla seguente espressione:

$${}_d \hat{Y} = \sum_{h=1}^{H_d} \hat{Y}_h = \sum_{h=1}^{H_d} \sum_{j=1}^{m_h} Y_{hj} \cdot W_{hj}, \quad (2)$$

in cui Y_{hj} e W_{hj} rappresentano rispettivamente il valore assunto dalla variabile y e il peso finale da attribuire all'individuo campione della famiglia j dello strato h .

Dalla precedente relazione si desume, quindi, che per ottenere la stima del totale (1) occorre moltiplicare il peso finale associato a ciascuna unità campionaria per il valore della variabile y assunto da tale unità ed effettuare, a livello del dominio di interesse, la somma dei prodotti così ottenuti.

Il peso da attribuire alle unità campionarie è ottenuto per mezzo di una procedura complessa che ha le seguenti finalità:

- correggere l'effetto distorsivo dovuto agli errori di lista e al fenomeno della mancata risposta totale;
- tenere conto della conoscenza di alcuni totali noti sulla popolazione oggetto di studio, nel senso che le stime campionarie di tali totali devono coincidere con i rispettivi valori noti; nell'indagine in oggetto i totali noti, che si riferiscono alla popolazione di 14 anni ed oltre, sono: a) la distribuzione regionale della popolazione per sesso e classi di età; b) la distribuzione della popolazione nelle aree $A_1, A_2, B_1, B_2, B_3, B_4$ definite nel primo paragrafo; c) la distribuzione delle famiglie per ampiezza².

La procedura per la costruzione dei pesi finali da attribuire alle unità campionarie, è articolata nelle seguenti fasi:

- 1) viene dapprima calcolato il peso base (o peso diretto), ottenuto come reciproco della probabilità di inclusione di ogni unità campionaria;
- 2) si calcola quindi il fattore correttivo che consente di soddisfare la condizione di uguaglianza tra i totali noti della popolazione e le corrispondenti stime campionarie;
- 3) il peso finale è dato dal prodotto del peso base per i fattori correttivi sopra indicati.

Il fattore correttivo del punto 3) è ottenuto mediante la risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è la distanza tra i pesi base ed i pesi finali, ed i vincoli sono definiti dalla condizione che le stime campionarie dei totali di popolazione sopra definiti coincidano con i valori noti degli stessi.

² Le distribuzioni di cui ai punti b) e c) non sono ottenute da archivi di popolazione bensì utilizzando i dati dell'indagine annuale multiscopo *aspetti della vita quotidiana*.

6. Valutazione del livello di precisione delle stime

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte dall'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo.

Indicando con $\hat{V}ar({}_d\hat{Y})$ la varianza della stima ${}_d\hat{Y}$, riferita al dominio d, la stima dell'errore di campionamento assoluto di ${}_d\hat{Y}$ si può ottenere mediante la seguente espressione:

$$\hat{\sigma}({}_d\hat{Y}) = \sqrt{\hat{V}ar({}_d\hat{Y})} \quad (3)$$

La stima dell'errore di campionamento relativo di ${}_d\hat{Y}$, è invece definita dall'espressione:

$$\hat{\varepsilon}({}_d\hat{Y}) = \frac{\sqrt{\hat{V}ar({}_d\hat{Y})}}{{}_d\hat{Y}} \quad (4)$$

La stima della varianza, $\hat{V}ar({}_d\hat{Y})$, viene calcolata come somma della stima della varianza dei singoli strati appartenenti al dominio d; in simboli:

$$\hat{V}ar({}_d\hat{Y}) = \sum_{h=1}^{H_d} \hat{V}ar(\hat{Y}_h) = \sum_{h=1}^{H_d} \frac{m_h}{m_h - 1} \sum_{j=1}^{m_h} \frac{(\hat{Y}_{hj} - \hat{Y}_h)^2}{m_h - 1} \quad (5)$$

dove

$$\hat{Y}_{hj} = Y_{hj} W_{hj} \quad e \quad \hat{Y}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} \hat{Y}_{hj}.$$

Gli errori campionari delle espressioni (3) e (4), consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire l'intervallo di confidenza, che, con una certa probabilità, contiene il parametro d'interesse. Con riferimento alla generica stima \hat{Y} tale intervallo assume la seguente forma:

$$\Pr\{\hat{Y} - k \hat{\varepsilon}(\hat{Y}) \leq Y \leq \hat{Y} + k \hat{\varepsilon}(\hat{Y})\} = P \quad (6)$$

Nella (6) il valore di k dipende dal valore fissato per la probabilità P; ad esempio, per P=0,95 si ha k=2.

7. Presentazione sintetica degli errori campionari

Ad ogni stima ${}_d\hat{Y}$ corrisponde un errore campionario relativo $\hat{\varepsilon}({}_d\hat{Y})$; ciò significa che per consentire un uso corretto delle stime sarebbe necessario pubblicare per ogni stima il corrispondente errore di campionamento relativo. Questo tuttavia non è possibile sia per limiti di tempo e di costi di elaborazione, sia perché le tavole di pubblicazione risulterebbero appesantite e di non facile consultazione per l'utente finale.

Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Per tali motivi si ricorre, in genere, ad una presentazione sintetica degli errori relativi basata sul *metodo dei modelli regressivi* fondata sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento. L'approccio utilizzato per la costruzione di questi modelli è diverso a seconda che si tratti di variabili qualitative o quantitative. Infatti, per quanto riguarda le stime di frequenze assolute (o relative) riferite alle modalità di

variabili qualitative, è possibile utilizzare modelli che hanno un fondamento teorico e secondo cui gli errori relativi delle stime di frequenze assolute sono funzione decrescente dei valori delle stime stesse. Per quanto riguarda, invece, le stime di totali di variabili quantitative, si tratta di un problema di notevole complessità, per il fatto che non è stata elaborata un'adeguata base teorica per l'interpolazione degli errori campionari delle stime in questione. L'approccio qui seguito per risolvere il problema in oggetto è di tipo empirico nel senso che si adattano diversi modelli regressivi che legano gli errori assoluti o relativi alle corrispondenti stime; tra i modelli stimati si sceglie quello che conduce ad un R^2 maggiore.

7.1. Presentazione sintetica degli errori campionari per stime di frequenze

Il modello utilizzato per le stime di frequenze assolute, con riferimento al generico dominio d , è del tipo seguente:

$$\log \left(\hat{\varepsilon}^2({}_d\hat{Y}) \right) = a + b \log({}_d\hat{Y}) \quad (7)$$

dove i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati.

Il prospetto 2 riporta i valori dei coefficienti a e b e dell'indice di determinazione R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze riferite alle famiglie e agli individui, per area territoriale. Le informazioni contenute nel prospetto 2 permettono di calcolare l'errore relativo di una qualsiasi stima di frequenza assoluta (e relativa).

Allo scopo di facilitare il calcolo degli errori campionari, nel prospetto 3 sono riportati i valori interpolati degli errori campionari relativi percentuali di alcuni valori tipici assunti dalle stime di frequenze assolute per area territoriale.

Il prospetto 3 è costituito da due colonne, la prima riporta K valori crescenti delle stime ${}_d\hat{Y}^k$ ($k=1, \dots, K$) e la seconda riporta i rispettivi errori relativi interpolati $\hat{\varepsilon}({}_d\hat{Y}^k)$. Le informazioni contenute in tale prospetto permettono, quindi, di calcolare l'errore relativo di un generica stima di frequenza assoluta mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (7).

Il primo metodo consiste nel cercare nella prima colonna del prospetto il livello di stima che più si avvicina alla stima di interesse ${}_d\hat{Y}$; l'errore relativo corrispondente $\hat{\varepsilon}({}_d\hat{Y})$ si trova sulla stessa riga della seconda colonna.

Nel secondo metodo, l'errore campionario della stima ${}_d\hat{Y}$, si ricava mediante la seguente espressione:

$$\hat{\varepsilon}({}_d\hat{Y}) = \hat{\varepsilon}({}_d\hat{Y}^{k-1}) + \frac{\hat{\varepsilon}({}_d\hat{Y}^k) - \hat{\varepsilon}({}_d\hat{Y}^{k-1})}{{}_d\hat{Y}^k - {}_d\hat{Y}^{k-1}} ({}_d\hat{Y} - {}_d\hat{Y}^{k-1}) \quad (8)$$

dove: ${}_d\hat{Y}^{k-1}$ e ${}_d\hat{Y}^k$ sono i valori delle stime entro i quali è compresa la stima ${}_d\hat{Y}$ e $\hat{\varepsilon}({}_d\hat{Y}^{k-1})$ e $\hat{\varepsilon}({}_d\hat{Y}^k)$ i corrispondenti errori relativi letti sul prospetto.

7.2. Presentazione sintetica degli errori campionari per stime di totali

Il modello utilizzato per le stime di totali di variabili quantitative, con riferimento al generico dominio d , è del tipo seguente:

$$\log \left(\hat{\varepsilon}^2({}_d\hat{Y}) \right) = a + b \log({}_d\hat{Y}) \quad (9)$$

dove i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (9) ad una nuvola di punti costituita da un consistente numero di coppie $(\hat{\varepsilon}({}_d\hat{Y}), {}_d\hat{Y})$. In considerazione del fatto che il modello (9) è di tipo empirico, l'insieme delle stime ${}_d\hat{Y}$ utilizzate per interpolare il modello è stato determinato in modo da includere la maggiorparte delle stime pubblicate.

Il prospetto 4 riporta i valori dei coefficienti a , b e dell'indice di determinazione R2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per area territoriale per famiglie e individui.

Nel prospetto 5 sono riportati i valori delle stime ${}_d\hat{Y}^*$ ottenuti in corrispondenza di valori tipici prefissati, indicati come $\varepsilon^* = \varepsilon({}_d\hat{Y}^*)$, dell'errore relativo percentuale ottenuti in base al modello (9); tutte le stime con valori superiori a ${}_d\hat{Y}^*$ presentano valori dell'errore relativo inferiori a ε^* , mentre le stime che assumono valori inferiori a ${}_d\hat{Y}^*$ presentano valori dell'errore relativo superiori a ε^* . I valori prefissati di ε^* sono: 5, 10, 15, 20, 25, 30 , 35 e 40%.