

NOTA METODOLOGICA

La popolazione di studio

Il disegno di campionamento adottato è ad uno stadio stratificato. Ciascuno strato è ottenuto come concatenazione delle seguenti variabili: regione, settore di attività economica e classe di dipendenti. La scelta della dimensione campionaria ha tenuto conto di vincoli di costo e di affidabilità delle stime con l'obiettivo di poter ottenere informazioni fino al dettaglio di regione oltre che di sottosezioni di attività economica e di classe dimensionale.

Le informazioni sono state raccolte attraverso un'indagine postale e tramite auto compilazione del questionario, disponibile anche in formato elettronico (su web) in alternativa alla tradizionale compilazione manuale del questionario cartaceo.

La popolazione oggetto di studio è costituita dalle imprese reali nate nel 2002 e sopravvissute nel 2005. L'unità di rilevazione è l'impresa, mentre quelle di analisi sono sia l'impresa che l'imprenditore.

La lista delle unità da cui è stato estratto il campione è costituita da 208.459 imprese reali nate nel 2002 e sopravvissute ad un anno (il 2003 è l'ultimo aggiornamento delle informazioni presenti nell'archivio ASIA utilizzate nell'applicazione della metodologia di BD).

Si tenga presente che tale lista di imprese non coincide esattamente con la popolazione oggetto di interesse che invece riguarda le imprese nate nel 2002 e sopravvissute a tre anni.

Tale disallineamento è stato ridotto sfruttando le informazioni che si sono rese disponibili durante il 2005 e relative all'aggiornamento di ASIA2004. Ciò ha permesso di stimare la popolazione di interesse che ammonta a 168.514 imprese, di cui 160.027 con l'imprenditore fondatore.

Gli errori di campionamento

Per una lettura agevole dell'errore campionario corrispondente a ciascuna stima, si riporta di seguito una presentazione sintetica di tali errori, ottenuta applicando un modello regressivo che mette in relazione ciascuna stima campionaria con il proprio errore di campionamento. Poiché le stime effettuate sono frequenze (relative) riferite a modalità di variabili qualitative, l'approccio del modello regressivo trova un fondamento teorico secondo cui gli errori relativi delle stime delle frequenze sono espressi da una funzione decrescente al crescere dei valori delle stime stesse.

Per questa indagine il modello utilizzato è stato:

$$\log(\varepsilon^2(\hat{Y})) = a + b \log(\hat{Y}) \quad [1]$$

La stima dei parametri a e b del modello [1] è stata ottenuta mediante il metodo dei minimi quadrati. Nella Tavola A1 sono riportati i valori a e b e l'indice di determinazione R^2 del modello per i domini di interesse. Per calcolare l'errore di una qualsiasi stima di un qualsiasi dominio basta considerare i parametri a e b relativi al dominio di interesse e applicare la seguente formula:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp(a + b \log(\hat{Y}))}$$

Viceversa sempre a partire dal modello [1] è possibile calcolare il valore soglia di una stima in corrispondenza di uno specifico errore relativo di campionamento applicando la formula:

$$\hat{Y}^* = \exp\left[\frac{\log(\hat{\varepsilon}^2(\hat{Y}) - a)}{b}\right] \quad [2]$$

La [2] consente di determinare valori soglia \hat{Y}^* per i quali tutte le stime $> \hat{Y}^*$ avranno errori relativi più bassi; al contrario, tutte le stime $< \hat{Y}^*$ avranno errori relativamente più alti.

La Tavola A2 riporta un prospetto dei valori delle stime percentuali corrispondenti ad errori relativi rispettivamente del 5 per cento, del 10 per cento, del 15 per cento, del 20 per cento e del 25 per cento

per ogni dominio. Tale prospetto è di immediata lettura; ad esempio per il dominio “Industria” stime di valori percentuali inferiori al 3,2% avranno errori relativi più alti del 25%; sempre per l’ “Industria”, stime di valori percentuali al di sopra del 22,6% sono più precise perché presentano errori relativi inferiori al 5%.

Tavola A1 - Valori dei coefficienti a, b e R^2 (%) delle funzioni utilizzate per le interpolazioni degli errori campionari per dominio di stima

DOMINI DI STIMA	COEFFICIENTI		R^2 (%)
	a	b	
ITALIA	7,3691	-1,31819	94,2
RIPARTIZIONI GEOGRAFICHE			
Nord-ovest	10,8068	-1,68993	91,7
Nord-est	9,6745	-1,65219	92,8
Centro	9,9282	-1,66343	90,8
Sud e Isole	9,896	-1,62488	92,8
MACROSETTORI ECONOMICI			
Industria	7,451	-1,64485	92,8
Costruzioni	10,8883	-1,69029	93,7
Commercio	10,0563	-1,64517	91,7
Servizi	10,2032	-1,64893	91,5

Tavola A2 - Stime percentuali corrispondenti ad alcuni valori di errori relativi campionari per tipo di dominio

Errori rel. %	STIME %								
	ITALIA	Nord_Ovest	Nord-Est	Centro	Sud e Isole	Industria	Costruzioni	Commercio	Servizi
5	15,8	51,7	40,7	42,0	33,0	22,6	59,7	39,5	26,2
10	5,5	22,8	17,6	18,2	14,1	9,7	26,3	17,0	11,3
15	3,0	14,1	10,8	11,2	8,5	5,9	16,3	10,4	6,9
20	1,9	10,0	7,6	7,9	6,0	4,2	11,6	7,3	4,9
25	1,4	7,7	5,8	6,1	4,5	3,2	8,9	5,6	3,7