



**File di microdati
per la ricerca**

Indagine sulle nuove attività imprenditoriali (FOBS)

Anno 2005

Aspetti metodologici dell'indagine

INDICE

1. Introduzione	3
2. La popolazione di studio	4
3. I domini di studio	4
4. Il disegno campionario	4
5. Metodo di stima	6
6. Presentazione sintetica degli errori di campionamento.....	8
7. La rilevazione e il trattamento dei dati	13
8. La diffusione dei risultati dell'indagine	15
9. Glossario	15
10. Riferimenti bibliografici.....	16
11. Contatti.....	17

1. Introduzione

Tra la metà di giugno 2005 e marzo 2006 l'Istat ha condotto, per la prima volta, l'indagine sulle "nuove attività imprenditoriali" allo scopo di approfondire la conoscenza dei fattori sociali, demografici ed economici che determinano il successo imprenditoriale. L'indagine è armonizzata con l'indagine europea "FOBS" (Factors of Business Success). A tale indagine, che ha lo scopo di ottenere dati confrontabili e armonizzati a livello Europeo sui vari fattori di successo imprenditoriale, partecipano, su base volontaria e utilizzando un questionario comune, 15 Paesi della Comunità Europea.

La popolazione target dell'indagine sulle "nuove attività imprenditoriali" è costituita – sulla base delle definizioni di natalità reale e di sopravvivenza utilizzate per la produzione dei dati sulla demografia d'impresa - dalle imprese "reali" nate nel 2002 e sopravvissute nel 2005. Le imprese rispondenti hanno permesso di acquisire informazioni utili allo studio delle determinanti della sopravvivenza, quindi del successo. I contenuti informativi sono di due tipologie: le caratteristiche demo-sociali, le motivazioni all'avviamento dell'impresa, le barriere e i rischi incontrati nel primo anno di attività sono riferite alle imprese nelle quali l'imprenditore attuale è lo stesso che ha fondato l'impresa, mentre le informazioni di natura economica relative all'impresa, sulla situazione corrente e i piani per il futuro, riguardano tutte le imprese.

L'indagine è campionaria. Il campione iniziale, pari a 12.778 unità, è stato selezionato in modo casuale stratificato per attività economica (22 aggregazioni di Ateco91), per dimensione (3 classi di dipendenti alla nascita) e regione; in particolare tenuto conto delle variabili di stratificazione, dei domini di stima, degli errori accettati per le stime, le imprese nate con oltre 10 dipendenti (1.177 unità) sono state censite. La strategia campionaria adottata, al fine di ottenere stime con livelli di precisione garantiti a livello regionale (dominio di stima non previsto dal progetto europeo), e a livello del dettaglio settoriale ha previsto, già in fase di progettazione dell'indagine, l'allargamento del campione iniziale. Il campione definitivo consiste quindi di un totale di 19.091 unità (circa il 9,2% della popolazione target).

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

2. La popolazione di studio

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 (al momento dell'indagine il 2003 era l'ultimo aggiornamento delle informazioni presenti nell'archivio ASIA utilizzate per l'applicazione della metodologia di BD).

È importante sottolineare che la lista di imprese così costituita non coincide con la popolazione oggetto di interesse che invece riguarda le imprese nate nel 2002 e sopravvissute a 3 anni. Tale disallineamento è stato ridotto sfruttando le informazioni che si sono rese disponibili nel corso del 2005 relative all'aggiornamento di ASIA2004 per la costruzione di opportuni 'stimatori di ponderazione vincolata'. Questi ultimi, come è noto, consentono di calcolare, per ciascuna unità di analisi, dei pesi correttivi che tengano conto sia della mancata risposta totale sia di eventuali errori di sottocopertura di lista. Tutto ciò ha permesso di stimare la popolazione di interesse che ammonta a 153.708 imprese, di cui 145.325 con l'imprenditore fondatore.

3. I domini di studio

Per questa indagine sono stati individuati 5 tipi di dominio:

DOM 1 – Regioni (20 domini);

DOM 2 – Settori di attività economica (22 domini);

DOM 3 – Classi di dipendenti (3 domini);

DOM 4 – Regioni x 3 macrosettori di attività economica (60 domini);

DOM 5 – 9 suddivisioni di attività economica x 3 classi di dipendenti (18 domini).

Le stime sono state calcolate anche per ripartizione geografica (4 domini ottenuti come aggregazione del primo dominio).

4. Il disegno campionario

Il disegno di campionamento adottato è ad uno stadio stratificato. Gli strati sono stati costruiti come concatenazione dei primi tre tipi di dominio (Regioni x Settori di attività economica x classi di dipendenti) e costituiscono pertanto la partizione minima della

popolazione che consente di ottenere tutti i domini di stima come aggregazione di strati elementari. Sono stati costruiti 1103 strati contenenti almeno un'impresa.

Per la dimensione del campione si è deciso di tenere in considerazione sia i vincoli di costo sia l'affidabilità delle stime. La procedura utilizzata è quella di allocazione ottima, che generalmente consiste nel fissare per la varianza di alcuni parametri di interesse $Var(Y_{kd})$,

un limite superiore $Var(Y_{kd})^*$ e nel minimizzare una funzione di costo $\sum_h^{L_d} C_h n_h = \min$ sotto i vincoli $Var(Y_{kd}) \leq Var(Y_{kd})^*$ (dove: $k=1, \dots, K$ numero di parametri scelti; $d=1, \dots, D$ numero di domini; $h=1, \dots, L_d$ numero di strati nel dominio d). Si tenga presente che si è preferito fissare i vincoli di precisione in termini di coefficienti di variazione, anziché in termini di varianze.

Tale procedura è alla base della metodologia adottata dall'ISTAT per la determinazione della dimensione del campione nella maggior parte delle indagini e fa riferimento alla soluzione proposta da Bethel¹. Naturalmente, dato lo slineamento tra lista e popolazione oggetto di indagine e le mancate risposte totali, il limite massimo teorico di errore fissato per ogni dominio, non coincide con l'effettivo errore commesso.

E' stato selezionato un campione di 12.778 unità a cui in seguito se ne è aggiunto un secondo pari a 6.313 unità, al fine di ridurre il più possibile il problema della mancata risposta totale, per un totale di 19.091 imprese. Fissato un limite massimo di unità da estrarre, il secondo campione è stato selezionato proporzionalmente ai tassi di risposta del primo campione ottenuti per ogni strato.

Di seguito si riporta, a titolo illustrativo, la distribuzione per ripartizione geografica dei vari insiemi a cui fino ad ora si è fatto riferimento.

¹Bethel J., (1989), Sample Allocation in Multivariate Survey, *Survey Methodology*, 15, pp. 47-57.

Tavola 1 - Popolazione oggetto di indagine, campione teorico ed effettivo, per un dominio di studio

DOMINIO	Popolazione		Campione teorico	Campione Di imprese eleggibili	Campione di imprenditori eleggibile	
	Popolazione (lista)	Eleggibile di imprese				
TOTALE	208459	153708	145325	19091	6457	5868
RIPARTIZIONE GEOGRAFICA						
Nord-ovest	54778	39871	37036	3812	1272	1138
Nord-est	41278	31666	29946	4098	1661	1495
Centro	45108	33084	31270	3967	1311	1180
Sud e Isole	67295	49087	47074	7214	2213	2055

5. Metodo di stima

Come è noto, al fine di stimare parametri sconosciuti della popolazione oggetto di studio, ci si basa su un campione che sia rappresentativo di tutte le unità della popolazione stessa. Più in particolare a ciascuna unità inclusa nel campione viene attribuito un peso, che indica il numero di elementi della popolazione rappresentati dall'unità campionaria.

Nella presente indagine il peso da attribuire a ciascuna unità è ottenuto in base alla seguente procedura:

- in primo luogo, viene calcolato un peso iniziale, definito peso diretto, determinato dal reciproco della probabilità di inclusione dell'unità campionata;
- successivamente, è stato calcolato un fattore correttivo che consente sia di correggere la distorsione dovuta alle mancate risposte totali, sia di soddisfare a livello di singolo dominio la condizione di uguaglianza tra totali noti delle variabili ausiliarie e le corrispondenti stime campionarie;
- infine, viene determinato un peso finale, espresso come prodotto del peso base per il fattore correttivo.

Per questa indagine sono stati considerati 45 totali noti che fanno riferimento alla distribuzione della popolazione aggiornata per i primi tre tipi di dominio (20 regioni + 22 settori di attività economica + 3 classi di dipendenti). Per quanto concerne la determinazione dei fattori correttivi suddetti, essi sono stati ottenuti risolvendo un problema di minimo vincolato: più precisamente è stata minimizzata la distanza logaritmica tra i pesi base e i pesi finali sotto il vincolo di uguaglianza tra totali noti e le corrispondenti stime campionarie.

Poiché in questa indagine sono state essenzialmente prodotte stime di frequenze assolute e relative, nel seguito si riporta un esempio di come tali stime siano state ottenute e il calcolo dei relativi errori campionari.

Si supponga di voler stimare il totale di una generica variabile y di interesse. La sua stima è formalmente espressa da:

$$\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} d_{hi} \gamma_{hi} y_{hi} = \sum_{h=1}^L \sum_{i=1}^{n_h} w_{hi} y_{hi}$$

dove:

$h=1, \dots, L$ è il generico strato;

y_{hi} è la variabile osservata sull'unità i -esima dello strato h -esimo;

$d_{hi} = \pi_{hi}^{-1}$ indica il peso diretto pari all'inverso della probabilità d'inclusione;

γ_{hi} fattore correttivo del peso diretto;

w_{hi} è il peso finale attribuito all'unità campionata i -esima e ottenuto dalla risoluzione del seguente problema di minimo vincolato:

$$\begin{cases} \text{Min} \left[\sum_{i \in S} c_i G_i(w_{hi}; d_{hi}) \right] \\ \sum_{i \in S} w_{hi} x_{hi} = X \end{cases}$$

dove con $G_{hi}(w_{hi}; d_{hi})$ è indicata la funzione di distanza logaritmica tra il peso diretto e il peso finale incognito; c_i è un peso indipendente da π_i attribuito a ciascuna unità del campione $\sum_{i \in S} w_{hi} x_{hi} = X$ è l'insieme dei vincoli di uguaglianza tra totali noti e le corrispondenti stime campionarie. La soluzione di tale problema è stata ottenuta adottando il metodo dei moltiplicatori di Lagrange.

Al fine di valutare l'attendibilità delle stime prodotte per ognuna di esse sono stati calcolati sia l'errore di campionamento assoluto che l'errore di campionamento relativo; il primo è

dato dall'errore standard: $\hat{\sigma}(\hat{Y}) = \sqrt{\text{Var}(\hat{Y})}$. Il secondo è dato dalla seguente espressione:

$$\hat{\varepsilon}(\hat{Y}) = \frac{\hat{\sigma}(\hat{Y})}{\hat{Y}}$$

Poiché lo stimatore utilizzato non è funzione lineare dei dati campionari, per la stima della varianza si è fatto riferimento al metodo proposto da Woodruff secondo il quale la varianza di ogni stimatore non lineare, che sia funzione di totali, è approssimabile con la varianza dello stimatore linearizzato corretto del totale della variabile. Più precisamente, poiché tutti gli stimatori a ponderazione vincolata convergono asintoticamente a quello di

regressione generalizzata, si è fatto riferimento a quest'ultimo al fine di ottenere un'espressione linearizzata dello stimatore per il quale la stima della varianza è più facile da calcolare.

6. Presentazione sintetica degli errori di campionamento

La presentazione per ciascuna stima del corrispondente errore campionario relativo non è di facile lettura. Per tale motivo si è deciso di ricorrere ad una presentazione sintetica degli errori, 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 2 sono riportati i valori a e b e l'indice di determinazione R^2 del modello per i domini di interesse, rispettivamente per l'insieme delle imprese con imprenditore fondatore e per tutte le imprese. 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:

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

Le Tavole 3 e 4 riportano un prospetto dei valori delle stime percentuali (in fiancata, stime dall'1 al 50 per cento) e dei corrispondenti errori relativi in percentuale per ogni dominio di stima. Errori superiori al 40 per cento sono segnalati dalla presenza dei due puntini (..).

Tavola 2 - Valori dei coefficienti a, b e R^2 (%) delle funzioni utilizzate per le interpolazioni degli errori campionari delle stime riferite alle imprese con imprenditore fondatore e a tutte le imprese, per ciascun dominio di stima

DOMINI DI STUDIO	IMPRESE CON IMPRENDITORE FONDATORE			TUTTE LE IMPRESE		
	a	b	R^2 (%)	a	b	R^2 (%)
ITALIA	7,0535	-1,34271	91,3	6,2453	-1,25820	94,5
SETTORI di ATTIVITA' ECONOMICA						
C e E	0,5197	-1,62467	87,8	-0,0817	-1,36065	87,6
DA	5,2744	-1,57994	89,0	3,9182	-1,36249	93,2
DB	5,8379	-1,60301	89,9	3,9999	-1,32133	94,5
DC	3,1729	-1,55810	90,3	0,6465	-1,06826	23,5
DD	4,2868	-1,60478	89,4	1,9221	-1,19229	82,7
DE	3,8409	-1,58048	91,0	2,5093	-1,33214	93,4
DFG	-0,6510	-1,36321	68,7	0,1797	-1,59441	87,3
DH	1,3730	-1,58476	89,8	0,6486	-1,36976	94,1
DI	3,8746	-1,62156	90,7	0,9450	-1,09677	26,7
DJ	5,7801	-1,57059	89,2	3,8969	-1,28219	93,2
DK	4,5455	-1,57806	91,4	3,1189	-1,33139	94,8
DL	4,6215	-1,58126	90,7	1,8659	-1,11740	78,0
DM	1,0442	-1,53821	87,9	0,4986	-1,37063	95,0
DN	4,8247	-1,62076	90,3	3,0110	-1,31595	93,5
F	9,0590	-1,57846	88,4	7,3375	-1,38462	92,7
50-51	8,8020	-1,58703	88,1	7,0094	-1,36951	92,6
52	8,0319	-1,58566	87,6	5,3230	-1,28658	86,5
H	7,0053	-1,54935	88,1	6,0122	-1,41536	91,9
I	6,5900	-1,57196	89,1	4,4515	-1,28333	90,0
J	5,4715	-1,60354	91,8	2,8474	-1,20606	90,0
70-71	7,1317	-1,57963	89,2	6,2671	-1,45192	93,1
72-73-74	9,0813	-1,59336	89,8	7,5650	-1,42281	94,1
RIPARTIZIONI GEOGRAFICHE						
Nord-ovest	6,8646	-1,33133	89,1	5,9372	-1,20768	94,4
Nord-est	6,1614	-1,33797	91,1	4,9600	-1,18046	95,7
Centro	6,1947	-1,31165	90,5	5,1993	-1,18491	96,0
Sud e Isole	6,5317	-1,35046	88,1	4,4800	-1,11144	95,6
REGIONI						
Piemonte e Valle d /	6,0110	-1,34465	88,2	4,8316	-1,1915	91,9
Lombardia	6,4647	-1,27027	80,2	5,1532	-1,1030	87,6
Bolzano	3,4460	-1,48348	93,2	2,7737	-1,3618	93,8
Trento	3,4056	-1,46766	76,2	2,5236	-1,2987	90,8
Veneto	5,0265	-1,21491	83,1	4,0537	-1,0829	86,4
Friuli Venezia Giulia	4,2777	-1,37542	91,5	2,8106	-1,1563	90,0
Liguria	5,1621	-1,41693	86,7	4,0728	-1,2503	93,4
Emilia Romagna	5,8641	-1,31848	72,7	4,9395	-1,1877	91,4
Toscana	5,7296	-1,31260	61,2	4,0417	-1,0852	87,4
Umbria	4,4790	-1,44155	76,9	3,3715	-1,2629	93,3
Marche	4,2786	-1,29892	86,8	3,1945	-1,1368	91,0
Lazio	5,8241	-1,26096	80,3	4,6606	-1,1093	88,5
Abruzzo	4,9253	-1,42214	65,3	3,6315	-1,2229	93,7
Molise	2,9954	-1,49103	90,9	2,8381	-1,4526	91,4
Campania	6,0123	-1,27771	83,6	3,8170	-1,0014	84,1
Puglia	5,9678	-1,36691	68,9	5,0511	-1,2441	86,3
Basilicata	3,5828	-1,41795	92,0	3,2339	-1,3611	94,0
Calabria	5,3959	-1,38538	71,2	4,4361	-1,2412	91,6
Sicilia	5,6945	-1,30191	62,1	4,0319	-1,0871	85,5
Sardegna	4,9009	-1,38435	72,8	4,5163	-1,3035	92,8

Tavola 3 - Valori di errori relativi percentuali delle stime percentuali riferiti alle imprese con imprenditore fondatore per ripartizione geografica, regione e settore di attività economica

STIME (%)	ITALIA	Nord-ovest	Nord-est	Centro	Sud e Isole
	errore (%)				
1	25,62
2	16,09	38,05	30,19	32,46	25,72
3	12,25	29,05	23,02	24,88	19,56
4	10,10	23,99	18,99	20,60	16,11
5	8,70	20,68	16,36	17,80	13,86
6	7,69	18,32	14,48	15,79	12,25
7	6,94	16,53	13,06	14,27	11,04
8	6,34	15,12	11,94	13,08	10,09
9	5,86	13,98	11,04	12,11	9,32
10	5,46	13,04	10,29	11,30	8,68
15	4,16	9,95	7,84	8,66	6,60
20	3,43	8,22	6,47	7,17	5,43
25	2,95	7,08	5,57	6,19	4,67
30	2,61	6,27	4,93	5,50	4,13
35	2,35	5,66	4,45	4,97	3,72
40	2,15	5,18	4,07	4,55	3,40
45	1,99	4,79	3,76	4,21	3,14
50	1,85	4,47	3,51	3,93	2,93

STIME (%)	Piemonte e Valle d'Aosta	Lombardia	Bolzano	Trento	Veneto	Friuli Venezia Giulia	Liguria	Emilia Romagna	Toscana	Umbria
	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)
1
2
3	38,0	32,9	39,1
4	35,9	33,3	35,5	30,8	27,6	33,1	35,6	32,4	33,5	37,4
5	30,9	28,9	30,1	26,1	24,1	28,4	30,4	28,0	28,9	31,8
6	27,4	25,7	26,3	22,9	21,6	25,0	26,7	24,8	25,7	27,9
7	24,7	23,3	23,4	20,4	19,6	22,5	23,9	22,4	23,2	25,0
8	22,6	21,4	21,2	18,5	18,1	20,5	21,8	20,5	21,2	22,7
9	20,8	19,9	19,5	17,0	16,9	18,9	20,0	19,0	19,7	20,8
10	19,4	18,6	18,0	15,7	15,8	17,6	18,6	17,7	18,3	19,3
15	14,8	14,4	13,3	11,7	12,4	13,3	14,0	13,5	14,1	14,4
20	12,2	12,0	10,8	9,4	10,4	10,9	11,4	11,2	11,6	11,7
25	10,5	10,4	9,1	8,0	9,1	9,4	9,7	9,7	10,1	10,0
30	9,3	9,3	8,0	7,0	8,1	8,3	8,5	8,6	8,9	8,8
35	8,4	8,4	7,1	6,3	7,4	7,4	7,7	7,8	8,1	7,8
40	7,6	7,7	6,4	5,7	6,8	6,8	7,0	7,1	7,4	7,1
45	7,1	7,2	5,9	5,2	6,3	6,3	6,4	6,6	6,8	6,5
50	6,6	6,7	5,5	4,8	6,0	5,8	5,9	6,1	6,4	6,1

STIME (%)	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
	errore (%)									
1
2
3	37,6	39,7
4	31,2	33,1	36,3	34,0	35,1	34,8	35,3	..	35,4	35,4
5	27,0	28,8	31,0	28,8	30,4	29,9	30,1	34,9	30,6	30,3
6	24,0	25,7	27,2	25,1	27,1	26,4	26,5	30,7	27,2	26,7
7	21,7	23,3	24,4	22,4	24,5	23,7	23,7	27,6	24,6	24,0
8	19,9	21,4	22,2	20,3	22,5	21,7	21,6	25,2	22,6	21,9
9	18,4	19,9	20,4	18,6	20,9	20,0	19,9	23,2	20,9	20,2
10	17,2	18,6	18,9	17,2	19,5	18,6	18,4	21,6	19,5	18,8
15	13,2	14,4	14,2	12,7	15,1	14,1	13,8	16,3	15,0	14,2
20	11,0	12,0	11,6	10,2	12,5	11,6	11,3	13,3	12,4	11,6
25	9,5	10,4	9,9	8,7	10,9	9,9	9,6	11,4	10,7	10,0
30	8,4	9,3	8,7	7,6	9,7	8,8	8,5	10,1	9,5	8,8
35	7,6	8,4	7,8	6,7	8,8	7,9	7,6	9,1	8,6	7,9
40	7,0	7,8	7,1	6,1	8,1	7,2	6,9	8,3	7,9	7,2
45	6,5	7,2	6,5	5,6	7,5	6,7	6,3	7,6	7,3	6,6
50	6,1	6,7	6,0	5,2	7,0	6,2	5,9	7,1	6,8	6,2

STIME (%)	Settori C e E	Settore DB	Settore DD	Settore DFG	Settore DI	Settore DK	Settore DFG	Settore DH	Settore DI	Settore DJ	Settore DK
	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)
1
2	39,23
3	29,75
4	38,26	24,46	37,44
5	31,92	38,81	..	36,45	35,74	..	21,01	31,37	38,58	39,60	37,20
6	27,52	33,60	37,22	31,63	30,87	34,77	18,55	27,15	33,27	34,32	32,22
7	24,28	29,75	32,90	28,05	27,28	30,78	16,70	24,03	29,37	30,41	28,53
8	21,79	26,77	29,56	25,28	24,51	27,70	15,25	21,61	26,35	27,38	25,68
9	19,80	24,39	26,90	23,06	22,30	25,24	14,07	19,69	23,95	24,96	23,40
10	18,18	22,45	24,72	21,24	20,49	23,22	13,10	18,11	21,99	22,98	21,53
15	13,07	16,29	17,86	15,49	14,80	16,86	9,93	13,13	15,83	16,71	15,64
20	10,35	12,98	14,18	12,38	11,75	13,43	8,17	10,46	12,54	13,33	12,46
25	8,63	10,88	11,86	10,40	9,82	11,26	7,01	8,76	10,46	11,19	10,45
30	7,45	9,42	10,25	9,03	8,49	9,75	6,19	7,58	9,02	9,70	9,05
35	6,57	8,34	9,06	8,01	7,50	8,63	5,58	6,71	7,96	8,59	8,01
40	5,89	7,51	8,14	7,21	6,74	7,76	5,09	6,04	7,15	7,74	7,21
45	5,36	6,84	7,40	6,58	6,13	7,07	4,70	5,50	6,50	7,05	6,57
50	4,92	6,29	6,80	6,06	5,63	6,51	4,37	5,06	5,96	6,49	6,05

STIME (%)	Settore DL	Settore DM	Settore DN	Settore F	Settori 50,51	Settore 52	Settore H	Settore I	Settore J	Settori 70,71	Settori 72,73,74
	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)				
1
2
3	..	38,25	34,57	33,80
4	..	30,66	..	33,58	..	27,52	37,92	37,93	38,06	..	26,87
5	38,18	25,82	38,81	28,16	36,86	23,06	31,90	31,82	31,82	34,77	22,50
6	33,06	22,44	33,48	24,39	31,89	19,95	27,70	27,58	27,50	30,11	19,46
7	29,26	19,93	29,54	21,59	28,22	17,66	24,58	24,43	24,30	26,66	17,21
8	26,33	17,99	26,51	19,43	25,38	15,88	22,16	22,00	21,83	23,99	15,47
9	23,99	16,43	24,10	17,71	23,12	14,47	20,23	20,05	19,86	21,86	14,08
10	22,07	15,15	22,13	16,29	21,27	13,31	18,65	18,46	18,26	20,11	12,95
15	16,02	11,09	15,93	11,83	15,42	9,65	13,62	13,42	13,19	14,60	9,38
20	12,76	8,89	12,62	9,43	12,27	7,68	10,90	10,70	10,47	11,63	7,46
25	10,70	7,49	10,53	7,91	10,28	6,44	9,17	8,98	8,76	9,75	6,24
30	9,26	6,51	9,08	6,85	8,89	5,57	7,96	7,78	7,57	8,45	5,40
35	8,20	5,78	8,02	6,06	7,87	4,93	7,06	6,89	6,69	7,48	4,77
40	7,38	5,22	7,20	5,46	7,08	4,43	6,37	6,21	6,01	6,73	4,29
45	6,72	4,76	6,54	4,97	6,45	4,04	5,81	5,66	5,47	6,13	3,91
50	6,18	4,39	6,01	4,57	5,93	3,72	5,36	5,21	5,02	5,64	3,59

Tavola 4 - Valori di errori relativi percentuali delle stime percentuali riferiti a tutte le imprese per ripartizione geografica, regione e settore di attività economica

STIME (%)	ITALIA errore (%)	Nord-ovest errore (%)	Nord-est errore (%)	Centro errore (%)	Sud e Isole errore (%)
1	22,46	..	39,91	..	30,02
2	14,52	34,44	26,51	28,70	20,42
3	11,25	26,96	20,87	22,57	16,30
4	9,39	22,66	17,61	19,04	13,89
5	8,16	19,81	15,44	16,68	12,27
6	7,28	17,74	13,86	14,97	11,09
7	6,60	16,16	12,66	13,66	10,18
8	6,07	14,91	11,70	12,62	9,45
9	5,64	13,89	10,91	11,77	8,85
10	5,28	13,03	10,25	11,06	8,35
15	4,09	10,20	8,07	8,70	6,67
20	3,41	8,58	6,81	7,34	5,68
25	2,96	7,49	5,97	6,43	5,02
30	2,64	6,71	5,36	5,77	4,53
35	2,40	6,12	4,90	5,27	4,16
40	2,21	5,64	4,52	4,87	3,86
45	2,05	5,26	4,22	4,54	3,62
50	1,92	4,93	3,97	4,26	3,41

STIME (%)	Piemonte e Valle d Aosta errore (%)	Lombardia errore (%)	Bolzano errore (%)	Trento errore (%)	Veneto errore (%)	Friuli Venezia Giulia errore (%)	Liguria errore (%)	Emilia Romagna errore (%)	Toscana errore (%)	Umbria errore (%)
1
2	36,3	38,8
3	36,5	34,0	36,9	31,9	29,2	30,7	37,1	34,9	32,7	37,4
4	30,7	29,0	30,4	26,5	25,0	26,0	31,0	29,4	27,9	31,2
5	26,9	25,7	26,1	22,9	22,1	22,8	26,9	25,7	24,8	27,1
6	24,1	23,2	23,0	20,4	20,0	20,6	24,0	23,1	22,4	24,1
7	22,0	21,3	20,7	18,4	18,4	18,8	21,8	21,1	20,6	21,9
8	20,3	19,8	18,9	16,9	17,1	17,4	20,1	19,5	19,2	20,1
9	19,0	18,6	17,5	15,6	16,1	16,3	18,7	18,2	18,0	18,7
10	17,8	17,5	16,3	14,6	15,2	15,3	17,5	17,1	17,0	17,5
15	14,0	14,0	12,3	11,2	12,2	12,1	13,6	13,4	13,6	13,5
20	11,8	12,0	10,1	9,3	10,4	10,2	11,3	11,3	11,7	11,3
25	10,3	10,6	8,7	8,1	9,3	9,0	9,8	9,9	10,3	9,8
30	9,2	9,6	7,7	7,2	8,4	8,1	8,8	8,9	9,4	8,7
35	8,4	8,8	6,9	6,5	7,7	7,4	8,0	8,1	8,6	7,9
40	7,8	8,2	6,3	5,9	7,2	6,9	7,3	7,5	8,0	7,3
45	7,3	7,6	5,8	5,5	6,7	6,4	6,8	7,0	7,5	6,8
50	6,8	7,2	5,4	5,1	6,4	6,0	6,4	6,6	7,1	6,3

STIME (%)	Marche errore (%)	Lazio errore (%)	Abruzzo errore (%)	Molise errore (%)	Campania errore (%)	Puglia errore (%)	Basilicata errore (%)	Calabria errore (%)	Sicilia errore (%)	Sardegna errore (%)
1
2	39,7	39,1
3	31,5	33,8	36,2	..	31,9	36,4	39,1	..	33,9	..
4	26,8	28,8	30,3	32,6	27,6	30,4	32,1	35,3	29,0	34,3
5	23,6	25,5	26,5	27,7	24,7	26,5	27,6	30,8	25,7	29,7
6	21,3	23,0	23,7	24,3	22,6	23,7	24,4	27,5	23,2	26,3
7	19,5	21,1	21,5	21,7	20,9	21,5	21,9	25,0	21,4	23,8
8	18,1	19,6	19,9	19,7	19,5	19,8	20,0	23,0	19,9	21,8
9	16,9	18,4	18,5	18,1	18,4	18,4	18,5	21,4	18,6	20,2
10	15,9	17,3	17,3	16,8	17,5	17,2	17,2	20,0	17,6	18,9
15	12,6	13,9	13,5	12,5	14,3	13,4	13,1	15,6	14,1	14,5
20	10,7	11,8	11,3	10,1	12,4	11,2	10,7	13,0	12,1	12,0
25	9,4	10,4	9,9	8,6	11,0	9,7	9,2	11,3	10,7	10,4
30	8,5	9,4	8,8	7,6	10,1	8,7	8,1	10,1	9,7	9,2
35	7,8	8,7	8,1	6,8	9,3	7,9	7,3	9,2	8,9	8,3
40	7,2	8,0	7,4	6,1	8,7	7,3	6,7	8,5	8,3	7,7
45	6,8	7,5	6,9	5,6	8,2	6,8	6,2	7,9	7,8	7,1
50	6,4	7,1	6,5	5,2	7,8	6,3	5,8	7,4	7,3	6,6

STIME (%)	Settori C e E	Settore DA	Settore DB	Settore DC	Settore DD	Settore DE	Settore DFG	Settore DH	Settore DI	Settore DJ	Settore DK
	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)
1	35,53	17,89	38,47	33,77	35,47
2	38,92	38,09	28,27	15,74	28,51	38,00	..	24,88	26,58	..	31,53
3	29,71	31,04	24,73	14,60	23,92	30,83	33,81	20,81	22,46	..	26,39
4	24,53	26,84	22,49	13,85	21,12	26,58	27,24	18,34	19,92	34,23	23,26
5	21,15	23,98	20,89	13,29	19,18	23,69	23,03	16,62	18,16	29,67	21,09
6	18,73	21,87	19,67	12,85	17,72	21,56	20,08	15,34	16,83	26,40	19,47
7	16,90	20,23	18,69	12,48	16,58	19,91	17,88	14,33	15,78	23,91	18,19
8	15,47	18,91	17,89	12,18	15,65	18,59	16,17	13,51	14,93	21,95	17,16
9	14,30	17,82	17,20	11,92	14,87	17,49	14,80	12,83	14,22	20,35	16,29
10	13,33	16,90	16,62	11,69	14,21	16,57	13,68	12,25	13,61	19,02	15,56
15	10,18	13,77	14,54	10,84	11,92	13,44	10,08	10,24	11,50	14,67	13,02
20	8,40	11,91	13,22	10,28	10,53	11,59	8,12	9,03	10,20	12,20	11,48
25	7,24	10,64	12,28	9,87	9,56	10,33	6,87	8,18	9,29	10,57	10,41
30	6,42	9,70	11,56	9,54	8,83	9,40	5,99	7,55	8,62	9,41	9,61
35	5,79	8,97	10,99	9,27	8,26	8,68	5,33	7,05	8,08	8,52	8,98
40	5,30	8,39	10,51	9,05	7,80	8,10	4,82	6,65	7,64	7,82	8,47
45	4,90	7,90	10,11	8,85	7,41	7,62	4,41	6,31	7,28	7,25	8,04
50	4,57	7,50	9,77	8,68	7,08	7,22	4,08	6,03	6,97	6,78	7,68

STIME (%)	Settore DL	Settore DM	Settore DN	Settore F	Settori 50,51	Settore 52	Settore H	Settore I	Settore J	Settori 70,71	Settori 72,73,74
	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)	errore (%)				
1	31,15	..	35,02	25,14	35,22	37,41	36,10	..	38,60
2	24,62	36,56	27,02	28,61	37,60	18,63	26,65	27,45	26,56	32,62	26,17
3	21,46	28,34	23,22	23,34	30,54	15,64	22,63	22,90	22,19	26,96	20,84
4	19,47	23,66	20,85	20,20	26,36	13,81	20,16	20,14	19,54	23,55	17,74
5	18,05	20,56	19,18	18,06	23,51	12,54	18,43	18,23	17,70	21,21	15,65
6	16,97	18,34	17,91	16,48	21,41	11,59	17,12	16,81	16,33	19,46	14,13
7	16,10	16,64	16,91	15,25	19,78	10,84	16,09	15,69	15,25	18,10	12,96
8	15,39	15,31	16,08	14,26	18,47	10,23	15,25	14,78	14,37	17,00	12,02
9	14,79	14,21	15,39	13,44	17,39	9,73	14,54	14,02	13,64	16,09	11,26
10	14,27	13,30	14,80	12,75	16,48	9,29	13,94	13,38	13,02	15,31	10,61
15	12,44	10,31	12,71	10,40	13,38	7,80	11,84	11,16	10,88	12,65	8,45
20	11,28	8,61	11,42	9,00	11,55	6,89	10,55	9,82	9,58	11,05	7,19
25	10,46	7,48	10,50	8,05	10,30	6,25	9,64	8,89	8,68	9,95	6,35
30	9,83	6,67	9,81	7,34	9,38	5,78	8,96	8,19	8,00	9,13	5,73
35	9,33	6,06	9,26	6,80	8,67	5,41	8,42	7,65	7,47	8,50	5,26
40	8,92	5,57	8,81	6,35	8,10	5,10	7,98	7,20	7,05	7,98	4,88
45	8,57	5,17	8,43	5,99	7,62	4,85	7,61	6,83	6,69	7,55	4,56
50	8,27	4,84	8,10	5,68	7,22	4,64	7,29	6,52	6,38	7,18	4,30

7. La rilevazione e il trattamento dei dati

La rilevazione delle informazioni è stata effettuata mediante un'indagine postale e tramite auto compilazione del questionario disponibile su web in formato elettronico.

Il questionario è costituito da 4 sezioni. La prima comprende le informazioni anagrafiche dell'impresa quali denominazione, indirizzo, attività economica principale. I campi sono pre-stampati con le informazioni dell'archivio Asia e ai rispondenti è richiesto di modificare le informazioni laddove risultino non corrette. In questa sezione è presente la prima domanda filtro: solo se l'imprenditore è il fondatore dell'impresa la sezione 2 deve essere compilata altrimenti dovranno essere compilate le altre due sezioni. La sezione 2 contiene informazioni sul profilo dell'imprenditore (età, grado d'istruzione, condizione professionale precedente, etc.) e sulle condizioni all'avvio dell'impresa (motivazioni alla nascita, difficoltà

incontrate, etc). Anche in questa sezione esiste una domanda filtro: solo le imprese nate nuove (non collegate ad altre imprese esistenti) sono di interesse per l'indagine. La sezione 3, comune a tutti i rispondenti, si focalizza sulla situazione dell'impresa alla data dell'indagine (occupazione, fatturato, cooperazione, innovazione e ostacoli allo sviluppo). Infine la sezione 4 chiude il questionario con alcune domande sull'evoluzione dell'impresa e le aspettative per il futuro.

La rilevazione è stata articolata in più fasi. La prima, iniziata a fine Giugno 2005, ha previsto il contatto dell'impresa tramite una lettera informativa. L'obiettivo è stato quello di identificare l'unità di rilevazione in modo più mirato per limitare le mancate risposte (causate ad esempio dalla tendenza da parte delle piccole imprese a consegnare il questionario Istat al commercialista). In seguito si sono articolate le successive spedizioni dei questionari e dei solleciti postali, affiancati dai solleciti telefonici (concentrati principalmente sulle unità da censire e su campioni di unità non rispondenti). La fase di raccolta delle informazioni si è conclusa a fine Marzo 2006.

Le informazioni rilevate vengono registrate all'interno di un database costruito ad hoc per tale indagine. La gestione dei non rispondenti, dei respinti, degli errori è stata svolta attraverso processi di tipo *batch* che fa uso di programmi SAS.

In particolare, per il trattamento delle mancate risposte parziali (MRP), il processo di controllo e correzione ha fatto uso di metodi deterministici - regole che se soddisfatte determinano l'imputazione di uguali valori per l'unità trattata - e metodi stocastici *hot deck* di tipo *nearest-neighbour*, in cui all'unità con dato mancante è assegnato il valore di un'unità rispondente, estratta da un sottoinsieme di unità rispondenti considerate "vicine" all'unità con dato mancante rispetto a prefissate variabili ausiliarie. Per le variabili quantitative (il numero di addetti e il fatturato) si è fatto uso delle variabili registrate nell'archivio Asia (metodo *cold deck*). Per la misura dell'occupazione (in particolare per la suddivisione tra occupati dipendenti e indipendenti) l'archivio è stato utilizzato nel 97% dei casi soprattutto a causa di elevati tassi di mancata risposta parziale, tra il 15 e il 39%. Per il fatturato 2004 si è registrato solo l'8% di MRP. Anche in questo caso l'archivio Asia è stato usato principalmente come *benchmark*, per correggere i dati in presenza di outlier. Per la variabile "fatturato previsto nel 2005" il dato mancante (o anomalo) è stato imputato mediante rapporti di variazione che fanno uso delle variabili addetti, fatturato 2004 e la modalità di risposta a) del quesito Q4_2.

8. La diffusione dei risultati dell'indagine

I risultati dell'indagine sono stati analizzati e pubblicati in forma sintetica attraverso la *Statistica in breve* del 13 Luglio 2006. Tale pubblicazione è stata inoltre spedita via e-mail a tutti i rispondenti che hanno manifestato espressamente nel questionario, compilando un apposito campo, il desiderio di ricevere i risultati dell'indagine.

Oltre al file di microdati per la ricerca scientifica, sono disponibili altri prodotti di informazione statistica.

I dati e la documentazione dell'indagine sono disponibili sul sito www.istat.it al seguente indirizzo:

1. http://www.istat.it/salastampa/comunicati/non_calendario/20060713_00/
2. http://www.istat.it/dati/dataset/20070227_00/

9. Glossario

Ripartizioni geografiche

Nord-ovest: Piemonte, Valle d'Aosta, Lombardia, Liguria (Italia nord-occidentale)

Nord-est: Trento e Bolzano, Veneto, Friuli-Venezia Giulia, Emilia-Romagna (Italia nord-orientale)

Centro: Toscana, Umbria, Marche, Lazio

Sud e Isole: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria (Italia meridionale); Sicilia, Sardegna (Italia insulare)

Imprenditore fondatore: è la persona che originariamente ha costituito l'impresa. Partecipa attivamente nella gestione dell'impresa, ne è il proprietario da solo o in associazione con altri imprenditori. L'amministratore di una società di capitali è considerato imprenditore se ne è stato il fondatore. Nelle società di persone è imprenditore colui/colei che ha avuto l'idea di costituire l'impresa o ne è il principale amministratore.

Impresa: *una unità giuridico-economica organizzata per la produzione di beni e servizi destinati alla vendita, dotata di autonomia decisionale, in particolare per quanto attiene alla destinazione delle sue risorse correnti. Una impresa esercita una o più attività in una o più unità locali.* Le definizione precedente deve essere intesa in senso statistico e si riferisce a qualsiasi soggetto fisico o giuridico che svolge una attività con rilevanza economica, pertanto la presente rilevazione interessa anche tutte le attività svolte, in forma individuale o **associata**, da liberi professionisti, lavoratori

Gruppo di imprese: “un’associazione di imprese retta da legami di tipo finanziario e non avente diversi centri decisionali, in particolare per quel che concerne la politica della produzione, delle vendite, degli utili” e in grado di “unificare alcuni aspetti della gestione finanziaria e della fiscalità”. Il gruppo si caratterizza come “entità economica che può effettuare scelte con particolare riguardo alle unità alleate che lo compongono”.

Indipendenti: i titolari (lo stesso imprenditore), soci e amministratori delle imprese, i familiari coadiuvanti a condizione che effettivamente lavorino nell’impresa, non siano iscritti nei libri paga, non siano remunerati con fattura, non abbiano un contratto di collaborazione coordinata e continuativa.

Dipendenti: la persone iscritte nei libri paga (anche se responsabili della gestione dell’impresa) e in particolare i dirigenti, i quadri, gli impiegati e gli operai, a tempo pieno o part time, gli apprendisti, i lavoratori stagionali, con contratto di formazione e lavoro, con contratto a termine, in Cassa Integrazione Guadagni, i lavoratori a domicilio, gli studenti che hanno un impegno formale per contribuire al processo produttivo in cambio di una remunerazione e/o formazione. Sono da considerare anche i lavoratori in sciopero, in maternità, in malattia o in congedo di breve durata.

Fatturato (in Euro): le vendite di prodotti fabbricati dall’impresa, le vendite di merci acquistate in nome proprio e rivendute senza trasformazione, le lavorazioni per conto terzi su materie prime e semilavorati di terzi, le lavorazioni e i servizi industriali su ordinazioni di terzi, le attività di intermediazione (commissioni, provvigioni, ecc.), gli introiti lordi del traffico (per le sole imprese di trasporto), le prestazione di servizi a terzi. I ricavi derivanti dall’attività dell’impresa devono essere indicati al lordo di tutte le spese addebitate ai clienti (trasporti, imballaggi, assicurazioni, ecc.) e di tutte le imposte indirette (fabbricazione, consumo, ecc.) ad eccezione dell’I.V.A. fatturata ai clienti. Tutte le voci di ricavo devono essere indicate al netto delle merci rese, degli sconti, degli abbuoni e premi.

10. Riferimenti bibliografici

Bethel, J., (1989), Sample Allocation in Multivariate Survey, *Survey Methodology*, 15, pp. 47-57.

Cicchitelli, G., Herzel, A., Montanari, G.E.. *Il campionamento statistico*, 1992.

Falorsi, P.D., Falorsi, S. (1995). Un metodo di stima generalizzato per le indagini sulle famiglie e sulle imprese. *Rapporto di ricerca*, n. 13. Quaderni CON PRIN dell’università di Bologna, 1995.

ISTAT (2005). Conti economici delle imprese, anno 2000. *Collana Informazioni*, n.6, 2005.

ISTAT (2005). GENESEES v.3.0, Funzione Stime ed Errori. *Manuale utente e aspetti metodologici*, n.3, 2005

11. Contatti

Per informazioni riguardanti l'indagine è possibile contattare: Caterina Viviano, viviano@istat.it

Curatori dei capitoli

I paragrafi 1, 2, 7, 8 e 9 sono stati redatti da Caterina Viviano, mentre i paragrafi 3, 4, 5 e 6 da Patrizia Cella.