

Indagine MET

2009

Nota metodologica



Indice

1. Piano di campionamento.....	3
Premessa	3
Stratificazione del campione	4
Fase di stima	10
2. Il quadro delle attività	12
Modalità operative.....	12
Personale.....	14
Sperimentazione	15
3. Qualità	16
Aspetti di carattere generale.....	16
Azioni preventive	16
Azioni nel corso della rilevazione	17
Azioni di valutazione	17
4. Appendice	19
Procedura per l’allocazione del campione nei domini pianificati.....	19
Allocazione del campione nei domini pianificati e livelli minimi di stime attendibili.....	22
Effetto del disegno	23
Riferimenti bibliografici.....	24

L'indagine campionaria sulle imprese per il 2009 si è avvalsa della supervisione di un comitato scientifico (G. Alleva, G. Barbieri, A. Zuliani) che ha contribuito all'impostazione metodologica generale e al disegno del campione, cui è stato affiancato un comitato tecnico (M. Centra e P. Falorsi) che ha seguito le diverse fasi operative e le problematiche di definizione dei dati e di riporto all'universo.

Il lavoro complessivo è stato coordinato da R. Brancati, con un gruppo di lavoro costituito da P. Ascani, C. A. Bollino, C. Borzaga, R. Brancati, D. Ciferri, A. Maresca, P. Sabatino, A. Silvani. Assistenza di A. Olivieri Moretti.

L'indagine è stata svolta da FORMAT srl nel periodo settembre--novembre 2009

1. Piano di campionamento

Premessa

L'indagine MET è riferita alla popolazione di imprese appartenenti ai settori dell'industria in senso stretto e dei servizi alla produzione per tutte le classi dimensionali e per tutte le regioni italiane. Le interviste sono state effettuate nel periodo compreso tra settembre e novembre 2009.

Dalla popolazione di riferimento è stato selezionato un campione casuale stratificato con selezione delle imprese negli strati con probabilità uguale e senza reimmissione. Il disegno del campione ha tenuto conto del fatto che in alcune regioni è stato richiesto un maggior livello di accuratezza delle stime prodotte rispetto ad altre e un conseguente sovracampionamento. Per tenere conto di tale vincolo le regioni sono state divise in 3 gruppi, caratterizzati da diversi livelli di attendibilità delle stime.

Il disegno prevede la pianificazione *ex-ante* dei domini di studio e la predeterminazione dell'attendibilità delle stime.

L'indagine 2009 contiene una componente longitudinale, relativa ad un sottocampione di imprese intervistate anche nella rilevazione dell'anno precedente. Tale fattore rappresenta un'importante fonte informativa, dal momento che la componente longitudinale del campione permette di produrre stime di flusso, limitatamente alla popolazione di imprese comprese tra il 2008 e il 2009, in grado di far luce su aspetti importanti delle reazioni del tessuto imprenditoriale alla crisi economica. Il disegno parzialmente longitudinale dell'indagine ha rappresentato un vincolo nella messa a punto del campione per la rilevazione 2009, riducendo i margini per l'allocatione del campione, dal momento che buona parte della distribuzione del campione negli strati risulta determinata dalle imprese presenti nella rilevazione 2008. Un ulteriore vincolo fissato riguarda la numerosità campionaria totale, stabilita in circa 22.000 unità.

Per la fase di stima è stato approntato un coefficiente di riporto in grado di riprodurre la popolazione di riferimento dell'indagine e di ridurre, tramite il ricorso a tecniche di calibrazione, gli effetti distorsivi dovuti al fenomeno della mancata risposta e dell'autoselezione del campione. Lo stimatore è stato costruito separatamente sia per l'intero campione dell'indagine 2009 che per la componente longitudinale 2008-2009.

Stratificazione del campione

Il campione è stratificato rispetto alla regione, alla dimensione e al settore di attività economica delle imprese, ripartiti come indicato nelle tabelle seguenti. Sia per la classe dimensionale che per il settore di attività economica si sono adottate le medesime suddivisioni in classi utilizzate nell'indagine 2008.

Tabella 1: Elenco dei settori economici (ATECO 2002)

Sottosezione ateco 2002	Descrizione del settore di attività economica
DA	Industrie alimentari, delle bevande e del tabacco
DB- DC	Industrie tessili e dell'abbigliamento - industrie conciarie
DD -D36.1	Industria del legno e dei prodotti in legno - fabbricazione di mobili
DE	Fabbricazione di pasta-carta, carta e prodotti di carta; stampa ed editoria
DH - DG -DF	Fabbricazione di articoli in gomma e materie plastiche -
DJ	Produzione di metallo e fabbricazione di prodotti in metallo
DM	Fabbricazione di mezzi di trasporto
DK	Fabbricazione macchine ed apparecchi meccanici; installazione e riparazione
DL	Fabbricazione macchine elettriche e apparecchiature elettriche ed ottiche
E -C DI - DN (escluso 36.1)	Produzione e distribuzione di energia elettrica, gas e acqua - estrazione di minerali - fabbricazione di prodotti della lavorazione di minerali
I – (solo 60 -61 - 62 -63 – 64)	Trasporti terrestri; trasporti mediante condotte -trasporti marittimi e per vie d'acqua -trasporti aerei - attività di supporto ed ausiliarie dei trasporti;
K – (solo 71 -72 - 73 -74)	Noleggio macchinari, attrezz. senza operatore, beni per uso pers. e domestico -informatica e attività connesse -ricerca e sviluppo - altre attività

Tabella 2: Elenco delle classi dimensionali

Dimensione impresa (addetti)
Micro 1-9
Piccole 10-49
Medie 50-249
Grandi 250 e oltre

Domini pianificati e allocazione del campione

Il campione è stato allocato in modo da imporre un errore massimo delle stime prodotte nei domini di interesse, riportati più avanti, i quali sono definiti come aggregazione di strati elementari. Il parametro oggetto di interesse è rappresentato da una generica proporzione P nella popolazione, ad esempio la quota di imprese che rispondono ad una specifica modalità di una domanda del questionario.

Al fine di definire i domini di studio la variabile regione, come già accennato, è stata ripartita in tre gruppi, che assicurano diversi livelli di attendibilità delle stime. In particolare il primo gruppo (regioni sovra campionate) comprende le regioni per le quali è garantita l'attendibilità delle stime di caratteristiche *rare* nella popolazione (in tali regioni si è effettuato il sovra campionamento), vale a dire per valori di una generica proporzione P non inferiore a 0.02; il secondo gruppo garantisce stime attendibili per caratteristiche *diffuse*, corrispondenti ad un valore di P non inferiore a 0.06; il terzo gruppo infine comprende regioni per le quali è assicurata una sufficiente attendibilità delle stime di caratteristiche *frequenti* nella popolazione, con P non inferiore a 0.09.

Il livello di attendibilità delle stime prodotte è stato determinato tramite una misura della variabilità campionaria, espressa dal coefficiente di variazione: una stima è giudicata attendibile se il suo coefficiente di variazione è inferiore a 0.20.

Tabella 3: Livelli di attendibilità delle stime per regione

Attendibilità per stime di caratteristiche rare (Regioni sovracampionate)	Attendibilità per stime di caratteristiche diffuse	Attendibilità per stime di caratteristiche frequenti
$P \geq 0,02$	$P \geq 0,06$	$P \geq 0,09$
Lombardia	Emilia Romagna	Piemonte
Veneto	Umbria	Valle D Aosta
Toscana	Marche	Trentino A. A
Lazio	Puglia	Friuli V.G.
Campania	Calabria	Liguria
Sicilia	Sardegna	Abruzzo
		Molise
		Basilicata

Nel piano di vincoli che l'allocazione è chiamata a soddisfare è stato introdotto un dominio unico nazionale, allo scopo di controllare l'accuratezza delle stime nazionali.

Oltre a garantire l'attendibilità delle stime regionali, il criterio di allocazione utilizzato ha permesso di controllare l'accuratezza della stime a livello di classe dimensionale delle imprese e di settore di attività economica, ma, in alcuni casi, anche rispetto alla distribuzione congiunta di classe dimensionale e settore di attività economica: i criteri applicati hanno garantito infatti che le sole regioni sovracampionate abbiano un livello di attendibilità accettabile per le stime relative ai domini definiti dalla distribuzione congiunta della regione e del settore di attività economica.

La pianificazione dei domini è riportata di seguito in termini di stima minima attendibile e di numerosità campionaria.

Tabella 4: Pianificazione dei domini, stima minima attendibile e numerosità campionaria

Dominio di studio		Popolazione	Campione	Livello minimo di una proporzione % di cui è garantita l'affidabilità	
Regione	Attendibilità per caratteristiche rare	Lombardia	319.824	1.838	2,6%
		Veneto	147.142	2.492	1,5%
		Toscana	121.675	2.492	1,4%
		Lazio	144.482	2.538	1,5%
		Campania	109.860	1.645	1,8%
		Sicilia	84.853	1.266	2,4%
	Attendibilità per caratteristiche diffuse	Emilia Romagna	141.174	1.221	3,3%
		Umbria	24.773	603	6,3%
		Marche	49.042	845	4,0%
		Puglia	79.735	1.155	2,8%
		Calabria	35.009	966	3,1%
		Sardegna	35.783	503	7,8%
	Attendibilità per caratteristiche frequenti	Piemonte	117.346	1.651	6,7%
		Valle D Aosta	3.624	203	16,4%
		Trentino A. A	26.404	508	8,5%
		Friuli V.G.	30.628	418	12,5%
		Liguria	42.019	415	12,3%
		Abruzzo	32.801	406	11,9%
		Molise	6.716	403	8,4%
Basilicata	12.246	399	9,9%		
Classe dimensionale (addetti)	Micro 1-9	1.441.976	14.245	0,3%	
	Piccole 10-49	105.966	3.325	1,5%	
	Media 50-249	14.828	2.183	1,8%	
	Grandi 250 e oltre	2.366	2.214	0,1%	
Settore di attività economica	DA	72.449	1.429	3,8%	
	DB- DC	80.410	1.464	3,7%	
	DD -D36.1	61.839	1.189	5,0%	
	DE	41.865	903	6,2%	
	DH - DG -DF	18.703	976	12,0%	
	DJ	101.792	1.644	3,6%	
	DM	6.730	723	10,6%	
	DK	41.860	1.158	7,3%	
	DL	48.270	1.062	6,6%	
	E -C DI - DN (escluso 36.1)	52.362	1.268	4,7%	
	I - 60 -61 - 62 -63 - 64	158.149	2.166	2,4%	
	K - 71 -72 - 73 -74	880.707	7.985	0,6%	
Totale Italia		1.565.136	21.967	0,3%	

Per i domini definiti dalla distribuzione congiunta dei caratteri di stratificazione il livello minimo di stima attendibile è di seguito riportato:

Tabella 5: Livelli minimi di stima attendibile

Dominio di studio	Livello minimo di una proporzione % di cui è garantita l'affidabilità
Regione x dimensione	0,117
Regione (sovracampionate) x settore	0,150

Componente longitudinale del campione

La componente longitudinale del campione è utilizzata per stimare le caratteristiche sulla popolazione costituita dalle imprese attive ed eleggibili sia nel 2008 che nel 2009. Per massimizzare il livello di accuratezza delle stime di flusso sono state incluse nel campione tutte le imprese intervistate nel 2008, ancora attive nel 2009 e disposte ad essere intervistate nuovamente. L'attendibilità delle stime di flusso, ripartita secondo i domini di studio pianificati nell'allocatione messa a punto per l'indagine 2009, è riportata nella tabella seguente.

Tabella 6: Domini di studio - componente longitudinale

Dominio di studio		Campione longitudinale	Livello minimo di una proporzione % di cui è garantita l'affidabilità	
Regione	Attendibilità per caratteristiche rare	Lombardia	1.204	8.0%
		Veneto	1.711	4.2%
		Toscana	1.600	4.0%
		Lazio	1.810	3.2%
		Campania	649	6,7%
		Sicilia	335	14.0%
	Attendibilità per caratteristiche diffuse	Emilia Romagna	1.063	3.9%
		Umbria	287	16.2%
		Marche	355	17.5%
		Puglia	934	3.0%
		Calabria	125	36.4%
		Sardegna	127	40.2%
	Attendibilità per caratteristiche frequenti	Piemonte	1.176	6.3%
		Valle D Aosta	110	36.0%
		Trentino A.A	335	20.3%
		Friuli V.G.	101	55.0%
		Liguria	254	16.4%
		Abruzzo	230	8.1%
		Molise	123	36.0%
Basilicata	234	13.0%		
Classe dimensionale (addetti)	Micro 1-9	6.316	0.9%	
	Piccole 10-49	4.137	0.4%	
	Media 50-249	1.811	0.5%	
	Grandi 250 e oltre	499	2.7%	
Settore di attività economica	DA	905	2.9%	
	DB- DC	945	2.7%	
	DD -D36.1	929	3.5%	
	DE	494	7.6%	
	DH - DG -DF	643	5.5%	
	DJ	1.184	3.3%	
	DM	451	8.9%	
	DK	779	5.8%	
	DL	726	7.3%	
	E -C DI - DN (escluso 36.1)	720	4.2%	
	I - 60 -61 - 62 -63 - 64	1.580	3.8%	
	K - 71 -72 - 73 -74	3.407	2.1%	
Totale Italia		12.763	0.7%	

Fase di stima

Le stime prodotte dai dati dell'indagine sono ottenute tramite una combinazione lineare delle osservazioni campionarie, dove i coefficienti, espressi come fattore moltiplicativo applicato ad ogni impresa intervistata, rappresentano il numero di imprese della popolazione che ogni impresa intervistata è chiamata a rappresentare.

Lo stimatore è stato costruito tramite due passi successivi:

- definizione del peso base;
- costruzione dello stimatore calibrato.

Il peso base, secondo la definizione di Horvitz-Thompson, è stato calcolato come reciproco della probabilità di inclusione nel campione. Tale coefficiente è stato poi modificato mediante un correttore moltiplicativo in grado di eliminare parte delle distorsioni dovute alla mancata risposta totale e all'autoselezione del campione dei rispondenti. Il correttore è ottenuto tramite l'applicazione di tecniche di calibrazione ed è stato formalizzato come soluzione di un sistema di minimo vincolato, in modo da imporre alle stime prodotte di riprodurre la popolazione cui è riferita l'indagine, sia in termini di totale di imprese, sia in termini di composizione secondo la regione, la dimensione e il settore di attività economica.

Nella fase di stima sono stati inoltre recuperati anche i piccoli scostamenti rilevati nei singoli strati tra il campione teorico e le numerosità effettive delle imprese rispondenti, la cui distribuzione nei domini di studio è riportata nella tabella seguente.

Tabella 7 Distribuzione del campione effettivo (annuale e longitudinale) secondo i domini di studio

Dominio di studio		Campione effettivo	Campione effettivo longitudinale	
Regione	Attendibilità per caratteristiche rare	Lombardia	1.838	1.204
		Veneto	2500	1.711
		Toscana	2500	1.600
		Lazio	2.538	1.810
		Campania	1.645	649
		Sicilia	1.266	335
	Attendibilità per caratteristiche diffuse	Emilia Romagna	1301	1.063
		Umbria	603	287
		Marche	845	355
		Puglia	1.156	934
		Calabria	964	125
		Sardegna	503	127
	Attendibilità per caratteristiche frequenti	Piemonte	1.651	1.176
		Valle D Aosta	203	110
		Trentino A.A	565	335
		Friuli V.G.	418	101
		Liguria	415	254
		Abruzzo	644	230
		Molise	403	123
		Basilicata	400	234
	Classe dimensionale e (addetti)	Micro 1-9	13.413	6.316
Piccole 10-49		5.823	4.137	
Media 50-249		2.329	1.811	
Grandi 250 e oltre		793	499	
Settore di attività economica	DA	2.094	905	
	DB- DC	1.668	945	
	DD -D36.1	1.680	929	
	DE	886	494	
	DH - DG -DF	1.069	643	
	DJ	2.315	1.184	
	DM	573	451	
	DK	1.304	779	
	DL	1.204	726	
	E -C DI - DN (escluso 36.1)	1.632	720	
	I - 60 -61 - 62 -63 - 64	2.538	1.580	
	K - 71 -72 - 73 -74	5.395	3.407	
Totale Italia		22.358	12.763	

2. Il quadro delle attività

Modalità operative

Sotto il profilo metodologico la raccolta dei dati si configura come un'indagine quantitativa realizzata con la tecnica delle interviste telefoniche, che sono state effettuate con il Sistema Cati (Computer assisted telephone interview), somministrando alle imprese un questionario strutturato, fatta eccezione per le grandi imprese che sono state contattate ricorrendo anche a tecniche Web (sistema CAWI¹).

Interviste alle grandi imprese

Tenuto conto dell'importanza delle grandi imprese nell'ambito del disegno del campione, tali aziende sono state rilevate non solo per mezzo dell'intervista telefonica Cati.

La rilevazione sul campione delle grandi imprese si è articolata nelle seguenti fasi operative:

- ✓ costruzione di un sito web ad accesso riservato con password (Sistema Cawi) con il questionario della rilevazione messo a disposizione dei soggetti da intervistare;
- ✓ effettuazione di un primo contatto telefonico per identificare presso ciascuna impresa la persona di riferimento alla quale effettuare l'intervista. A seconda dell'esito della telefonata sono stati comunicati alle imprese (direttamente al telefono, o via mail, o via fax) il login e la password per accedere al sito riservato e compilare il questionario messo a disposizione on line, oppure è stata effettuata l'intervista telefonicamente in tutti i casi in cui l'interlocutore avesse preferito

¹ CAWI – computer assisted web interview.

procedere direttamente in questo senso, o in tutti in cui l'impresa non disponesse di una connessione Internet;

- ✓ effettuazione di un'azione di sollecito telefonico presso tutti coloro che entro un mese circa dalla data prevista di chiusura della rilevazione non avevano compilato il questionario via web, nonostante avessero precedentemente manifestato la propria volontà di procedere in tal senso. In tutti i casi in cui si è reso necessario le interviste mancanti sono state effettuate telefonicamente utilizzando il Sistema Cati.

L'attività di rilevazione si è conclusa al 30 novembre 2009 con un tasso di rifiuto del 17,0% e un tasso di successo del 34,7%.

Andamento attività Rapporto Met

Riepilogo delle attività di indagine	Numero casi	%
Concesse	22.358	34,7
Rifiuti	10.932	17,0
Appuntamenti non andati a buon fine (solleciti e-mail/fax inviati)	9.770	15,2
Liberi/fax/segreterie	14.402	22,4
Errati	6.906	10,7
Tot lavorate	64.368	100,0

Personale

Per lo svolgimento delle interviste è stato impiegato un gruppo di intervistatori (ricercatori assistenti) esperto nello svolgimento delle interviste con il sistema Cati, sotto la direzione e il controllo di due ricercatori coordinatori. L'addestramento del personale è stato sia di tipo teorico, sia di tipo pratico. Complessivamente la formazione di ogni intervistatore non è stata inferiore alle sei ore.

Dal punto di vista teorico

Il personale è stato formato per mezzo di una riunione di briefing finalizzata a illustrare l'attività di ricerca nella sua completezza e a impartire le opportune istruzioni in merito alla gestione operativa del Cati progettato per l'indagine in oggetto.

Dal punto di vista pratico

Tutto il personale è stato formato con l'effettuazione di interviste di prova simulate (l'intervistato è stato uno dei ricercatori dell'istituto) e in condizioni reali (interviste dirette a soggetti "fuori campione"), di modo da standardizzare il più possibile il comportamento degli intervistatori. Tutto il personale coinvolto nella rilevazione è stato informato sull'intero processo di lavoro e sulle finalità della ricerca, e non soltanto quindi sulle modalità di somministrazione del questionario. Gli intervistatori sono stati resi consapevoli dell'importanza del loro ruolo: sull'importanza, per esempio, di evitare il più possibile le "cadute" delle anagrafiche, sull'importanza di avere questionari completi, senza mancate risposte, sull'importanza del corretto atteggiamento da tenere nel corso del colloquio di intervista, per non influenzare le risposte degli intervistati.

Tra gli obiettivi del coordinatore del gruppo di ricerca è rientrato il controllo dell'operato degli intervistatori con l'effettuazione di riunioni periodiche di debriefing (non meno di una riunione di debriefing ogni settimana) con lo scopo

di verificare l'andamento della rilevazione e risolvere gli eventuali problemi non previsti.

Sperimentazione

L'attività di sperimentazione ha risposto alla necessità di verificare la tecnica dell'intervista telefonica e l'applicazione software del Cati per la gestione delle interviste, verificare la procedura per il rispetto dei criteri di sostituzione delle unità campione, verificare i parametri campionari e le misure di attendibilità delle stime (es. propensione alle mancate risposte da parte delle aziende campione, la tendenza alla concentrazione spuria di modalità; etc.).

La sperimentazione è stata effettuata mediante una indagine pilota che ha permesso di verificare sul campo che l'insieme delle soluzioni previste per lo svolgimento della ricerca fosse adeguato al conseguimento degli obiettivi previsti.

L'indagine pilota è stata realizzata con le stesse modalità dell'indagine vera e propria, e ha comportato l'effettuazione di 30 interviste presso altrettante aziende "fuori campione", di modo da non alterare il campione di indagine vero e proprio.

3. Qualità

Aspetti di carattere generale

Il sistema dei controlli di qualità ha lo scopo di contenere il cosiddetto errore non-campionario con l'obiettivo ultimo di ottenere il massimo livello di affidabilità dell'indagine. Tale sistema è basato su tre sottosistemi principali, a loro volta articolati su una serie di azioni ben identificate. I tre sottosistemi sono il sottosistema delle azioni preventive, il sottosistema delle azioni da svolgere nel corso della rilevazione, e il sottosistema delle azioni di valutazione.

Azioni preventive

L'obiettivo delle azioni preventive è quello di creare un clima favorevole allo svolgimento delle interviste facilitando il più possibile il compito degli intervistatori presso le aziende campione.

Nell'ambito della presente rilevazione sono state applicate tre azioni di tipo preventivo: la *“doppia telefonata” alle imprese*, l'istituzione di un *sito web*, ad accesso riservato con password, a disposizione delle imprese, e l'istituzione di un *servizio di informazione*.

Doppia telefonata

Questa tecnica ha lo scopo di presentare l'indagine all'impresa con una prima telefonata, identificare la *“persona di riferimento”* al suo interno, spiegarli i motivi che ci hanno indotto a cercare la sua collaborazione, informarla dei suoi diritti con riferimento alla privacy, e concordare l'effettuazione dell'intervista vera e propria (seconda telefonata).

Sito web

Il sito Web, ad accesso riservato con password, è stato destinato a tutte le aziende che hanno avuto o il bisogno di avere maggiori informazioni sull'indagine alla quale sono state chiamate a partecipare. Sul sito web è stata predisposta una copia del questionario di rilevazione, che nel caso le *“persone di*

riferimento” in azienda, hanno potuto scaricare e visionare prima dell’effettuazione dell’intervista vera e propria.

Azioni nel corso della rilevazione

Lo scopo delle azioni in corso d’opera è stato quello di diminuire il più possibile le possibilità di errore degli intervistatori in itinere. Tali azioni sono basate essenzialmente sui controlli automatici previsti dal Sistema Cati e dai report periodici dell’attività svolta compilati dal coordinatore della ricerca.

Controlli automatici

I controlli automatici consistono nella registrazione controllata dei dati inseriti e nell’identificazione automatica degli errori. Tali controlli riguardano in particolare le violazioni delle regole logico-formali (regole di compatibilità) relative ai limiti imposti dal dominio di variazione delle singole variabili.

Report periodici

I report periodici descrivono in corso d’opera lo stato di avanzamento della rilevazione dati, basandosi su indicatori quali il numero delle interviste effettuate per ciascuno strato, il numero dei contatti telefonici effettuato per ottenere una intervista, il tasso di ricorso alle sostituzioni, il tasso dei rifiuti e dei numeri errati, la produttività per singolo intervistatore, e così via.

Azioni di valutazione

Lo scopo delle azioni di valutazione è quello di identificare gli errori effettuati nel corso della rilevazione, nell’ambito delle singole interviste, nonostante il sistema dei controlli automatici del Cati. L’obiettivo ultimo delle azioni di valutazione è quello di limitarne l’impatto sulla coerenza dell’indagine nel suo insieme.

Tali azioni sono basate sull’effettuazione di una indagine di controllo, che consiste nella ripetizione di alcune interviste presso un campione ridotto di aziende tra quelle intervistate per verificare il livello dello scostamento tra i

risultati ottenuti per mezzo delle interviste originali e i risultati ottenuti per mezzo delle interviste di controllo realizzate *ex novo*. L'indagine di controllo ha comportato l'effettuazione di 150 interviste presso altrettante aziende, clusterizzate per intervistatore, con lo scopo di individuare eventuali scostamenti rilevanti tra le due dimensioni considerate (interviste originali e interviste di controllo).

Sono state effettuate infine, delle verifiche sulla coerenza dei valori ricavati dall'indagine e riportati all'universo (in particolare sul numero degli addetti) rispetto alle informazioni disponibili per aggregazioni regionali e settoriali.

4. Appendice

Procedura per l'allocazione del campione nei domini pianificati

Il procedimento seguito per allocare il campione nei domini di studio pianificati utilizza una tecnica specifica per le indagini multiobiettivo, dove l'interesse si concentra oltre che sulla stima di un generico parametro riferito all'intera popolazione oggetto di studio, anche sul valore del medesimo parametro in specifiche sottopopolazioni, chiamate domini di studio. In genere, in assenza di ulteriori informazioni sulla popolazione di interesse, l'allocazione ottimale prevede un disegno proporzionale. Tuttavia l'allocazione proporzionale, pur massimizzando l'efficienza delle stime riferite all'intera popolazione, non assicura il controllo dell'accuratezza delle stime riferite ai singoli domini di studio. La tecnica utilizzata (Cochran, 1977) permette di predeterminare il livello di accuratezza, espresso in termini di varianza campionaria delle stime prodotte, nei domini di studio, controllando parallelamente l'attendibilità delle stime riferite all'intera popolazione.

Sia data una popolazione stratificata in H strati ($h=1,2,\dots,H$) di numerosità N_h ; siano definiti D domini, ottenuti tramite una determinata aggregazione di strati, di numerosità N_d , ($d=1,2,\dots,D$):

$$N_d = \sum_{h=1}^H N_h \cdot I_{h,d}$$

dove $I_{h,d}$ indica se lo strato h contribuisce alla definizione del dominio d :

$$I_{h,d} = \begin{cases} 1 & \text{se } h \in d \\ 0 & \text{se } h \notin d \end{cases}$$

La varianza campionaria della stima di una frazione P della popolazione nel dominio d è data da:

$$V_d = \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{(N_h - 1) \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d}$$

approssimando la quantità $(N_h - 1) \cong N_h$ per popolazioni sufficientemente grandi, si ottiene:

$$V_d \cong \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{N_h \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d}$$

La quantità V_d può essere scomposta in due addendi, uno dei quali dipende dalle quantità n_h , cioè dalla allocazione del campione negli strati, mentre l'altro è indipendente dall'allocazione del campione ed è funzione della partizione della popolazione nei domini:

$$V_d \cong \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{N_h \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d} = \sum_{h=1}^H \frac{S_h^2 \cdot N_h^2}{N_d^2 \cdot n_h} \cdot I_{h,d} - \sum_{h=1}^H \frac{S_h^2 \cdot N_h}{N_d^2} \cdot I_{h,d}$$

Posto:

$$V_{d0} = - \sum_{h=1}^H \frac{S_h^2 \cdot N_h}{N_d^2} \cdot I_{h,d} ; \quad V_{d1} = \sum_{h=1}^H \frac{S_h^2 \cdot N_h^2}{N_d^2 \cdot n_h} \cdot I_{h,d} ; \quad V_{dh}^2 = \frac{S_h^2 \cdot N_h^2}{N_d^2} \cdot I_{h,d}$$

si ottiene:

$$V_d = V_{d0} + V_{d1} = V_{d0} + \sum_{h=1}^H \frac{V_{dh}^2}{n_h}$$

La scomposizione della varianza così ottenuta permette di controllare l'allocazione del campione negli strati h in modo da assicurare un livello predeterminato del valore della varianza V_d della stima riferita al dominio d .

Il problema si sostanzia quindi nel definire un'allocazione del campione, di numerosità non fissata, in grado di rispettare una serie di vincoli posti alla varianza delle stime per ciascun dominio di studio.

A tale scopo è utile il teorema di Kuhn-Tucker che risolve problemi di minimo vincolato nella programmazione non lineare.

Applicato al problema in questione il teorema di Kuhn-Tucker può essere formulato nel modo seguente (Bethel, 1989):

dato un vettore di vincoli $V^*, \{V_1^*, V_2^*, \dots, V_D^*\}$ (upper bound) delle varianze V_d esiste un vettore di moltiplicatori $\lambda_D, \{\lambda_1, \lambda_2, \dots, \lambda_D\}$ tali che l'allocazione

$$n_h = \sqrt{\sum_{d=1}^D \lambda_d \cdot V_{dh}^2} \text{ soddisfa i vincoli } V_d \leq V_d^*, \quad d = 1, 2, \dots, D.$$

La definizione dei moltiplicatori λ_d non è immediata e richiede un algoritmo iterativo (Chromy, 1987) di seguito esposto.

Sia fissato un valore iniziale del vettore dei moltiplicatori λ_d , pari a 1 per ciascun dominio di studio:

$${}_0\lambda_d = 1 \quad \forall d$$

sia definita l'allocazione del campione al passo k in funzione dei moltiplicatori λ_d :

$${}_k n_h = \sqrt{\sum_{d=1}^D {}_k \lambda_d \cdot V_{dh}^2}$$

siano definiti i valori dei moltiplicatori λ_d al generico passo $k+1$ come funzione dei corrispondenti valori al passo precedente:

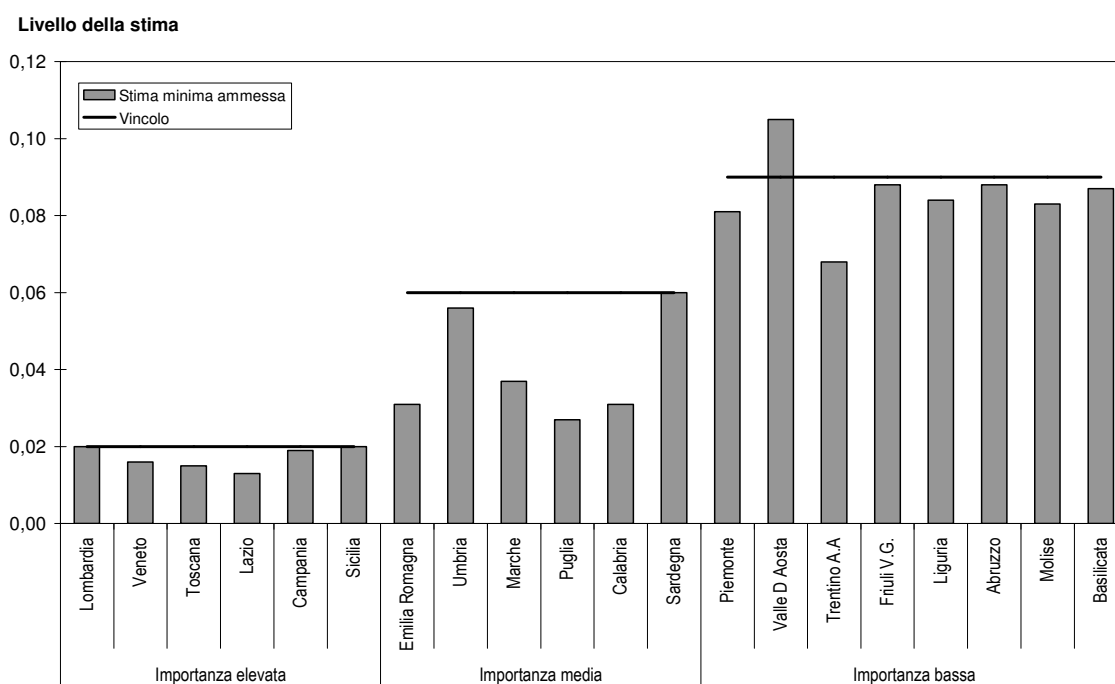
$${}_{k+1} \lambda_d = {}_k \lambda_d \cdot \left(\frac{\sum_{h=1}^H \frac{V_{dh}^2}{{}_k n_h}}{V_d^* - V_{d0}} \right)^2$$

Iterando la procedura l'intero vettore dei moltiplicatori converge verso un valore che soddisfa il vincolo imposto.

Allocazione del campione nei domini pianificati e livelli minimi di stime attendibili

Il procedimento utilizzato ha prodotto l'allocazione del campione nei 960 strati², definiti dalla distribuzione congiunta delle imprese nella popolazione secondo la regione, il settore di attività economica e la classe dimensionale, come riportati in precedenza. La procedura applicata impone una serie di vincoli espressi in termini di varianza massima ammessa su distribuzioni marginali rispetto alla distribuzione congiunta dei caratteri di stratificazione. Una volta ottenuta l'allocazione del campione i livelli della varianza possono risultare più bassi del vincolo imposto, utilizzato come limite massimo, permettendo la produzione di stime più attendibili per alcuni domini di studio. Di seguito è riportato il livello minimo effettivo di una stima attendibile nei singoli domini (regione, settore di attività economica e classe dimensionale).

Figura 1: Livello della stima



² La numerosità degli strati riportata è solo teorica: in realtà gli strati non vuoti nella popolazione sono risultati essere 837.

La procedura utilizzata non permette di determinare a priori la numerosità campionaria totale, che è stata pertanto raggiunta con approssimazioni successive. Questo elemento non ha permesso di rispettare il vincolo imposto nella regione Valle d'Aosta, che, a fronte di un livello della stima minima ammessa pari al 9% presenta un livello ammesso non inferiore al 10.5%.

Effetto del disegno

L'effetto del disegno, vale a dire la perdita in termini di accuratezza delle stime riferite all'intera popolazione dovuta al particolare piano di campionamento utilizzato, è pari a 2.46. Ciò significa che, a livello nazionale, la varianza della stima di una generica proporzione della popolazione prodotta dal disegno utilizzato è pari a circa due volte e mezzo la varianza del campione casuale semplice di pari numerosità. Il livello dell'effetto del disegno non è giudicato particolarmente alto e va messo in relazione con gli indubbi benefici dovuti alla possibilità di predeterminare i livelli di accuratezza delle stime nei domini di studio.

Riferimenti bibliografici

M.Ballin, P.Falorsi, G.Scepi, P.Vicard, “New Tools for Planning Business Sampling Surveys”, Istat- Servizio Studi Metodologici, 2004.

BETHEL J. (1989), “Sample Allocation in Multivariate Surveys”, Survey Methodology, 15.

COCHRAN W.G. (1977), “Sampling Techniques”, 3rd ed., Wiley, New York.