



Rilevazione Statistica sull'Innovazione nelle Imprese (CIS3)

Anni 1998-2000

Aspetti metodologici dell'indagine

INDICE

1. Introduzione	3
2. La popolazione di riferimento.....	4
3. Il disegno campionario.....	5
4. La rilevazione e il trattamento dei dati	7
5. La metodologia di calcolo dei pesi campionari	9
6. La diffusione dei risultati dell'indagine	10
7. Glossario	11
8. Riferimenti bibliografici.....	13
9. Contatti.....	13

1. Introduzione

La Cis, sviluppata congiuntamente dall'Eurostat e dagli Istituti statistici dei Paesi Ue (in collaborazione con la Commissione europea), è finalizzata a raccogliere dati sugli input e sugli output dei processi di innovazione delle imprese europee dell'industria e dei servizi. Ad esempio, come indicatore di input viene considerata la spesa in R&S e in altre attività innovative. Tra gli indicatori di output sono inclusi: l'impatto dell'innovazione sul fatturato dell'impresa, gli effetti sull'attività aziendale e sul livello occupazionale, la propensione a brevettare o a utilizzare altre modalità di protezione della proprietà intellettuale. La rilevazione Cis considera anche altri aspetti del processo innovativo come: i fattori che ostacolano o limitano l'attività innovativa delle imprese, il supporto pubblico all'innovazione, gli accordi di cooperazione e le fonti di informazione sull'innovazione. Con riferimento al triennio 1998-2000 sono stati, inoltre, prodotti per la prima volta indicatori sulle attività di innovazione non strettamente tecnologica svolte dalle imprese italiane.

L'indagine è condotta sulla base di criteri definitivi e metodologie di rilevazione comuni a tutti i Paesi dell'Unione Europea ed è inserita nel quadro concettuale del cosiddetto "Manuale di Oslo", che dal 1992 rappresenta la base concettuale e metodologica per la misurazione dell'innovazione tecnologica svolta dalle imprese (Ocse/Eurostat, 1997).

Oltre alla diffusione di dati aggregati a livello settoriale o dimensionale, l'Istat mette a disposizione di un'utenza specializzata - analisti e ricercatori appartenenti ad organizzazioni pubbliche o private - un file di dati elementari resi anonimi nel pieno rispetto delle norme sulla tutela della riservatezza dei dati individuali.

Scopo del presente documento è fornire alcuni elementi informativi sui principali aspetti metodologici della Cis³. In particolare, nel Capitolo 2 è descritta la popolazione di riferimento. Nel Capitolo 3 è illustrata il disegno di campionamento. Nel Capitolo 4 sono presentate le modalità di acquisizione e trattamento dei dati, mentre nel Capitolo 5 è indicata la procedura di calcolo delle stime campionarie. Le ultime sezioni, infine, contengono informazioni sui principali canali di diffusione dati, sulla terminologia ricorrente e i riferimenti bibliografici fondamentali per lo svolgimento dell'indagine.

2. La popolazione di riferimento

La popolazione oggetto della rilevazione è costituita da 164.593 imprese dell'industria, delle costruzioni e dei servizi con almeno dieci addetti attive nel 2000.

2.1 Attività economica (NACE)

L'indagine europea prevede l'obbligo di copertura dei seguenti settori-chiave:

- Attività estrattive (NACE 10-14);
- Attività manifatturiere (NACE 15-37);
- Produzione e distribuzione di energia elettrica, gas e acqua (NACE 40-41);
- Commercio all'ingrosso (NACE 51);
- Trasporti, magazzinaggio e comunicazioni (NACE 60-64);
- Attività finanziarie (NACE 65-67);
- Informatica e attività connesse (NACE 72);
- Attività degli studi di architettura, ingegneria ed altri studi tecnici (NACE 74.2);
- Collaudi ed analisi tecniche (NACE 74.3).

Altri settori coperti dall'indagine italiana sono stati:

- Commercio di autoveicoli (NACE 50);
- Commercio al dettaglio (NACE 52);
- Alberghi e ristoranti (NACE 55);
- Attività immobiliari (NACE 70);
- Noleggio di macchinari e attrezzature senza operatore (NACE 71);
- Ricerca e sviluppo (NACE 73);
- Altre attività di servizi alle imprese: Attività legali, contabilità, consulenza fiscale e societaria; studi di mercato e sondaggi di opinione; consulenza commerciale e di gestione (NACE 74.1); Pubblicità (NACE 74.4); Servizi di ricerca, selezione e fornitura di personale (NACE 74.5); Servizi di investigazione e vigilanza (NACE 74.6); Servizi di pulizia e disinfestazione (NACE 74.7); Altre attività di servizi alle imprese (NACE 74.8).

2.2 Dimensione aziendale

La popolazione di riferimento è composta da imprese con almeno dieci addetti.

La variabile utilizzata per misurare la dimensione aziendale è il numero medio annuo di addetti nel 2000.

2.3 Unità statistica di rilevazione

L'unità di rilevazione assunta per l'indagine è l'impresa, definita come organizzazione di un'attività economica esercitata con carattere professionale al fine della produzione di beni o per la prestazione di servizi destinabili alla vendita¹.

Tra le imprese sono comprese: le imprese individuali, le società di persone, le società di capitali, le società cooperative, le aziende speciali di comuni o province o regioni. Sono considerate imprese anche i lavoratori autonomi e i liberi professionisti.

3. Il disegno campionario

La procedura con cui è stato costruito il campione di partenza è di tipo casuale.

Il disegno di campionamento adottato è ad uno stadio stratificato, con selezione delle unità con probabilità uguali. Almeno tre imprese in ciascuno strato sono state selezionate. Nel caso di un numero totale di imprese inferiore presenti nello strato, tutte le unità sono state coinvolte.

La popolazione è stata suddivisa in strati (ossia, sottoinsiemi tra loro non sovrapposti definiti sulla base di alcune caratteristiche strutturali delle unità statistiche e all'interno dei quali le unità sono fra loro omogenee riguardo alle variabili oggetto di studio).

Gli strati sono stati definiti dalla concatenazione delle modalità identificative dei settori di attività economica, delle classi di addetti e delle regioni di localizzazione delle imprese. In particolare, le variabili di stratificazione sono state:

- l'attività economica (*divisione Ateco 1991*², ad eccezione dei gruppi '244', '353', '742' e '743');
- classe di addetti (10-19 addetti, 20-49 addetti, 50-249 addetti, 250 addetti e oltre);
- ripartizione geografica (livello 1 della classificazione europea Nuts).

La stratificazione adottata costituisce la partizione minima della popolazione che permette di ottenere i domini di stima come aggregazione di strati elementari.

La numerosità teorica del campione è stata ottenuta impiegando una metodologia sviluppata dall'Istat per la determinazione della dimensione campionaria ottima in presenza di molteplici vincoli e obiettivi. La metodologia impiegata adotta una generalizzazione dell'algoritmo proposto da Bethel (1989), in cui viene definito un problema di minimo vincolato con funzione obiettivo, rappresentata dal costo complessivo

¹ Regolamento (Cee) n. 696/93 del Consiglio, del 15 marzo 1993, relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità (Gazzetta ufficiale n. L 76 del 30 marzo 1993).

² Istat. *Classificazione delle attività economiche*. (Metodi e norme, serie C-n.11). Roma, 1991.

dell'indagine, e vincoli di tipo lineare. I vincoli sono rappresentati dai livelli massimi di errore ammessi nella stima delle principali variabili di interesse, espressi in termini di coefficiente di variazione e con riferimento al grado di dettaglio richiesto per i domini di stima finali. La ricerca dell'allocazione ottima è avvenuta assumendo numero di addetti, fatturato e spesa totale per l'innovazione come variabili d'interesse e impiegando le informazioni disponibili - in termini di medie e varianze - relative alla precedente occasione d'indagine Cis2 (periodo di riferimento: 1994-1996). Il dimensionamento del campione è stato, dunque, ottenuto imponendo alle stime delle tre variabili (numero di addetti, fatturato e spesa totale per innovazione) per dominio di studio un errore massimo pari al 10 per cento.

Allo scopo di limitare il carico statistico gravante sulle imprese, è stata adottata una tecnica di selezione coordinata delle unità campionarie (tecnica JALES) che consente di ridurre la probabilità di selezionare unità già estratte nell'ambito di altre indagini condotte in precedenza (Ohlsson, 1995)

3.1 Lista di riferimento (archivio di base)

La lista di riferimento per l'estrazione del campione e delle informazioni strutturali e anagrafiche delle unità campionate è l'Archivio statistico delle imprese attive (Asia).

L'Archivio raccoglie le informazioni identificative (denominazione, localizzazione), strutturali (addetti, attività economica prevalente e secondaria, natura giuridica, volume degli affari) e demografiche (data inizio attività, data cessazione, stato di attività, presenza di procedure concorsuali) di tutte le imprese (e relative unità locali) attive in tutti i settori di attività economica (ad eccezione delle sezioni A, B, L, P, Q e dei soggetti privati non profit) della classificazione Ateco (versione 2002 per gli anni dal 2000 in poi)³.

L'Archivio Statistico delle Imprese Attive (Asia) rappresenta la base informativa di tutte le indagini Istat sulle imprese poiché individua la popolazione di riferimento per i piani di campionamento e per il loro riporto all'universo. Costituisce inoltre la base per l'estrazione delle liste di imprese cui inviare i questionari. E' aggiornato annualmente attraverso un processo di integrazione di informazioni provenienti da fonti di diversa natura (amministrative e statistiche).

³ L'Archivio è stato creato in ottemperanza al regolamento Cee n. 2186/93 del Consiglio del 22 luglio 1993 relativo al coordinamento comunitario dello sviluppo dei registri di imprese utilizzati a fini statistici e in attuazione del decreto del Presidente del Consiglio dei ministri del 26 febbraio 1994.

Con riferimento alla CIS3, la popolazione di riferimento è costituita dalle imprese attive nel 2000.

3.2 La dimensione del campione di partenza e il tasso di risposta

La dimensione del campione teorico così definito è stata pari a 29.157 imprese.

Nel campione totale sono, inoltre, sono state incluse tutte le 2.499 imprese con 250 addetti e oltre operanti all'interno del campo di osservazione definito e attive nell'ultimo anno del periodo di riferimento dell'indagine (2000).

Tenendo conto delle imprese cessate, inattive, fallite, in liquidazione e fuori campo di osservazione, i questionari ricevuti sono stati circa 18.000. Tuttavia, ai fini della produzione delle stime campionarie, dopo aver verificato lo stato di eleggibilità dei rispondenti⁴, sono stati utilizzati 15.512, pari al 53 per cento delle imprese del campione di partenza.

4. La rilevazione e il trattamento dei dati

L'indagine ha carattere campionario per le imprese con meno di 250 addetti. È invece censuaria per le imprese con 250 addetti.

La rilevazione è stata condotta mediante auto-compilazione di un questionario cartaceo. Il questionario è stato inviato per via postale alle imprese del campione di partenza nel dicembre del 2001. A causa degli elevati tassi di non risposta, al primo invio dei questionari sono seguiti due solleciti: il primo, di tipo postale; il secondo, svolto in parte con modalità postale e in parte tramite contatti telefonici.

A conclusione della raccolta dati, secondo quanto raccomandato da Eurostat, si è proceduto anche ad una rilevazione campionaria delle imprese non rispondenti. È stato selezionato un campione casuale stratificato di circa 1.500 imprese a cui è stato rivolto – con tecnica Cati⁵ – un questionario riguardante lo svolgimento o meno di attività di innovazione. Le informazioni provenienti dai rispondenti alla rilevazione Cati sono state poi utilizzate nel calcolo dei pesi finali al fine di attenuare l'eventuale distorsione dovuta all'auto-selezione dei rispondenti.

⁴ Un'unità si definisce eleggibile se appartiene alla popolazione oggetto di indagine.

⁵ Cati, *Computer-aided telephone interviewing* è la tecnica di intervista telefonica assistita da computer.

Per il controllo e correzione (C&C) dei dati sono state impiegate procedure standard definite da Eurostat al fine di implementare una metodologia per il trattamento dei dati armonizzata a livello europeo.

Il processo di C&C si è articolato in 5 fasi principali:

1. Controllo e correzione manuale dei dati. È un'attività di revisione che viene svolta prima della registrazione dei questionari e che prevede una serie di operazioni di individuazione e di correzione degli errori emersi in fase di compilazione del modello (ad esempio, a causa di un'errata interpretazione dei quesiti e delle regole di compilazione da parte del rispondente). Questa attività è eseguita manualmente da tecnici mediante revisione dei modelli, re-intervista, uso di informazioni ausiliarie e/o di conoscenze sul fenomeno investigato.
2. Localizzazione deterministica degli errori mediante procedure automatiche. L'individuazione deterministica delle situazioni di errore e delle variabili da imputare è condotta sulla base degli edit specificati da Eurostat definiti a partire delle regole interne del questionario e di conoscenze a priori del fenomeno oggetto di rilevazione. Di diversa natura sono gli errori non campionari che possono originarsi nella Cis (errori di dominio, mancate risposte parziali, valori anomali e incompatibilità fra risposte, errori di codifica e di percorso) e diverse sono le procedure e le tecniche utilizzate per il trattamento delle varie tipologie di errore.
3. Controllo e correzione interattivi dei valori anomali (*outliers*) di tipo micro. L'insieme dei record considerati non corretti comprende anche quelle unità che presentano, per qualcuna delle variabili quantitative rilevate, valori che si discostano in modo significativo dai valori che le stesse variabili assumono nel resto delle unità campionarie. Obiettivo di questa fase è l'individuazione di quei valori anomali (o sospetti) che incidono significativamente in termini di capacità distorsiva sulle stime finali e sull'imputazione delle altre variabili. I record/imprese anomali sono quelli che violano il range di accettazione, stabilito in base a funzioni che utilizzano i valori correnti delle variabili di interesse. Il programma individua diverse modalità di trattamento degli *outliers* applicate solo dopo una preliminare e scrupolosa analisi circa la loro natura e l'impatto sulle stime finali.
4. Imputazione dei valori mancanti o errati di tipo misto mediante l'attivazione di procedure automatiche di tipo sia deterministico che probabilistico. Questa fase prevede l'implementazione sequenziale di tecniche diverse individuate in funzione della tipologia di variabile (quantitativa o qualitativa) e dell'errore riscontrato

(incoerenze logiche, valori anomali, valori mancanti). In particolare, il processo di correzione si compone di tre passi:

- l'esecuzione iniziale delle procedure di imputazione logico-deduttiva che permette, sulla base di un sistema di vincoli e relazioni logiche tra le variabili, di eliminare tutte le incongruenze interne al singolo record;
- l'imputazione delle variabili quantitative, che viene effettuata utilizzando uno 'stimatore rapporto', basato su una serie di relazioni predefinite tra le variabili da imputare e alcune variabili ausiliarie ad esse altamente correlate;
- l'applicazione delle procedure di imputazione per le variabili qualitative (dicotomiche e categorico-ordinali) che utilizza il metodo del 'donatore hot deck' e consiste nell'individuare, per ogni record errato e rispetto a ciascuna specifica tipologia di variabile, il record donatore 'più vicino', i cui valori consentono al recipiente di soddisfare tutti gli edit. Il donatore è scelto in modo tale da minimizzare la distanza tra esso ed il ricevente.

5. Validazione del processo di C&C e analisi delle stime delle variabili principali del questionario per dominio di impresa mediante un confronto dei dati aggregati corretti e opportunamente ponderati con informazioni storiche o ausiliarie al fine di evidenziare eventuali situazioni 'sospette'.

Per un approfondimento degli aspetti concernenti il C&C della Cis, consultare il contributo di Valeria Mastrostefano "Un approccio armonizzato al controllo e correzione dei dati: il caso della Community Innovation Survey", pubblicato in *Metodi per il C&C dei dati nelle indagini sulle imprese: alcune esperienze nel settore delle statistiche strutturali*, Istat, Collana Contributi n. 7/2008.

5. La metodologia di calcolo dei pesi campionari

Data la tipologia campionaria dell'indagine, la stima dei totali delle variabili di interesse è stata calcolata attribuendo ad ogni unità rispondente un coefficiente di riporto o peso, indicante il numero di unità della popolazione rappresentate dall'impresa, inclusa se stessa.

I pesi finali da associare alle unità rispondenti sono stati calcolati in base alla teoria dello stimatore di *ponderazione vincolata* di Deville e Särndal (1992), utilizzabile quando sono noti i totali per dominio di alcune variabili – c.d. variabili *ausiliarie* – correlate con quelle di interesse; le stime finali sono state prodotte per *domini* definiti dalle modalità delle

seguenti variabili: settore di attività economica (divisione Nace Rev.1, ad eccezione dei gruppi '244', '353', '742' e '743'); classe dimensionale espressa in termini di numero di addetti (10-19, 20-49, 50-249, 250 ed oltre) .

Il peso finale è stato calcolato in base alla metodologia di *Dewille e Särndal* (1992) risolvendo un sistema di minimo vincolato. Sintetizzando, si può affermare che il vettore dei pesi finali è quello che, modificando il meno possibile, il vettore dei pesi diretti corretti per mancata risposta totale, rispetta per ciascun dominio di stima la condizione di uguaglianza tra le stime dei totali delle variabili ausiliarie ed i corrispondenti *totali noti* della popolazione.

Le variabili ausiliarie utilizzate per il calcolo dei pesi finali dell'indagine sull'innovazione 2002-2004 sono le due seguenti: *numero di imprese* e *numero di addetti*, i cui valori sono noti dall'archivio *Asia 2004* per tutte le unità della popolazione obiettivo. La convergenza delle stime delle variabili ausiliarie ai corrispondenti totali noti è stata realizzata congiuntamente per i tre domini di stima già definiti.

Per il calcolo dei pesi finali si è tenuto conto, inoltre, dell'informazione proveniente dai rispondenti alla rilevazione *Cati* effettuata su un sottocampione di non rispondenti. In particolare, si è fatto riferimento alla metodologia di Hansen e Hurwitz (1946) che, utilizzando l'informazione proveniente dalle imprese rispondenti sia all'indagine principale sia alla successiva, ha permesso la costruzione di uno stimatore non distorto, sotto determinate condizioni⁶, del totale della variabile di interesse⁷.

Per il calcolo delle stime campionarie si è utilizzato Genesees, un software generalizzato sviluppato dall'Istat in linguaggio Sas e disponibile al seguente indirizzo del sito Istat:

http://www.istat.it/strumenti/metodi/software/produzione_stime/genesees/index.html .

6. La diffusione dei risultati dell'indagine

Oltre alla predisposizione del file di dati elementari per la ricerca, l'Istat ha reso disponibili i risultati dell'indagine CIS3 attraverso altri canali di diffusione.

In particolare, i dati e la documentazione dell'indagine sono disponibili sul sito dell'Istat al seguente indirizzo: http://www.istat.it/dati/catalogo/20040621_02/ .

⁶ La condizione per la non distorsione dello stimatore è che il campione di non rispondenti sottoposto alla seconda rilevazione sia di adeguata numerosità e non si verifichino casi di mancata risposta totale. Per maggiori dettagli si confronti Hansen, M. H., and Hurwitz, W. N., 1946.

⁷ Nel caso di specie, il numero di imprese innovatrici.

Le statistiche europee sull'innovazione sono, invece, disponibili sul sito di Eurostat:
<http://epp.eurostat.ec.europa.eu>

Infine, un'alternativa al file di dati elementari per la ricerca, per l'accesso ai microdati è il Laboratorio Adele dell'Istat che consente agli analisti e ai ricercatori appartenenti ad organizzazioni pubbliche o private di condurre analisi micro-economiche per esclusivi motivi di ricerca. L'accesso ai microdati avviene presso una struttura sotto costante controllo per garantire il rispetto della riservatezza dei dati individuali (per informazioni, contattare: adele@istat.it).

7. Glossario

Il questionario dell'indagine CIS3, contenuto nel file ISTAT_MFR_CIS_Questionario_2000.pdf, contiene tutte le definizioni necessarie a interpretare correttamente i fenomeni rilevati. Le definizioni principali sono riportate anche di seguito e servono a distinguere le imprese sulla base delle attività innovative svolte e della tipologia di innovazione introdotta.

Innovazione tecnologica

Un prodotto, servizio o processo introdotti dall'impresa che può essere considerato nuovo o significativamente migliorato, rispetto a quelli precedentemente disponibili, in termini di caratteristiche tecniche e funzionali, prestazioni, facilità d'uso, ecc..

Un'innovazione tecnologica si realizza nel momento della sua introduzione sul mercato (innovazione di prodotto o servizio) o del suo utilizzo in un processo produttivo (innovazione di processo).

Le innovazioni di prodotto e di processo non devono necessariamente consistere in prodotti, servizi o processi totalmente nuovi; è infatti sufficiente che risultino nuovi per l'impresa che li introduce.

Innovazione di processo

Ha luogo quando i processi di produzione o di distribuzione:

- vengono automatizzati o integrati in misura crescente;
- sono resi più flessibili;
- migliorano in termini di qualità;
- sono resi più sicuri e compatibili con l'ambiente.

Sempre nell'ambito dell'innovazione di processo, le funzioni logistiche e di controllo a livello d'impresa possono beneficiare dell'introduzione di tecnologie che:

- rendono più efficiente la pianificazione e la gestione del trasporto merci;
- incrementano la flessibilità nella distribuzione;
- migliorano il controllo degli stock di merce in entrata e in uscita.

Innovazione di prodotto

È rilevabile quando:

- l'adozione di una nuova tecnologia determina miglioramenti significativi nelle caratteristiche di utilità dei beni e dei servizi;
- aumenta la gamma di beni o servizi offerti da un'impresa sul mercato.

Attività innovative

Si definiscono innovative le attività che si rendono necessarie per sviluppare e introdurre prodotti, servizi o processi produttivi tecnologicamente nuovi (o significativamente migliorati). Comprendono la ricerca e sviluppo (R&S), l'acquisto di macchinari, attrezzature, software e licenze, la progettazione industriale e le altre attività preliminari alla produzione e alla fornitura di servizi, la formazione del personale connessa all'introduzione di prodotti o servizi o processi tecnologicamente nuovi o significativamente migliorati e il marketing di prodotti e servizi innovativi.

Impresa con attività innovative

Impresa che ha dichiarato di aver svolto nel triennio di riferimento dell'indagine attività finalizzate allo sviluppo o all'introduzione di innovazioni di prodotto, servizio o processo. Sono incluse in questa categoria: le imprese innovatrici (per la definizione, consultare la voce successiva); le imprese con attività ancora in corso e non concluse alla fine del 2000 o avviate nel triennio 1998-2000 ma poi abbandonate nello stesso periodo.

Impresa innovatrice

Impresa che ha introdotto con successo sul mercato o nel proprio processo produttivo innovazioni nel triennio 1998-2000.

8. Riferimenti bibliografici

Bethel, J. (1989) "Sample allocation in multivariate surveys". *Survey Methodology*, 15: 47-57.

Deville, J.C. and Särndal, C.E. (1992) "Calibration estimators in survey sampling", *Journal of the American Statistical Association* 87, 367.382.

Eurostat (2004) "Innovation in Europe: Results for the EU, Iceland and Norway", Panorama of the European Union, Theme 9 Sciences and technologies, European Communities.

Hansen, M. H., and Hurwitz, W. N., (1946) "The Problem of Nonresponse in Sample Surveys", *Journal of the American Statistical Association*, 41, 516-529.

ISTAT, (2004) "Statistiche sull'innovazione delle imprese", Collana *Informazioni*, N. 1.

OECD (1997) "Oslo-Manual, Proposed Guidelines for Collecting and Interpreting Technological Innovation Data", Organisation for Economic Co-Operation and Development, Paris.

Ohlsson, E., B. G. Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Kott e P. S. Colledge (eds.). "Coordination of samples using permanent random numbers". In *Business Survey Methods*. Wiley, New York, 1995.

9. Contatti

Per informazioni riguardanti l'indagine si prega di contattare Valeria Mastrostefano: mastrost@istat.it.

Per informazioni relative alla metodologia di anonimizzazione dei dati elementari, si prega di contattare Daniela Ichim: ichim@istat.it

Curatori dei capitoli

Il documento è stato redatto da Valeria Mastrostefano.
