



RESEARCH DATA CENTER

REX - Remote Execution / Sistema di elaborazione a distanza

Indagine sulle Imprese Industriali e dei Servizi (INVIND)

Documentazione per l'utilizzo dell'archivio

Periodo di riferimento dei dati: dal 1984 al 2022

Data di aggiornamento: luglio 2023

Parole chiave: Indagini sulle imprese, microdati campionari, Industria e servizi

A cura di: Divisione Indagini Campionarie del Servizio Analisi statistiche



BANCA D'ITALIA
EUROSISTEMA

Luglio 2023

Indice

Informazioni generali.....	3
Popolazione di riferimento	3
Periodo di disponibilità e frequenza dei dati.....	4
Rilevazione dei dati e riferimenti normativi	4
Statistiche aggregate e pubblicazioni di riferimento.....	4
Struttura e caratteristiche degli archivi.....	4
Aspetti metodologici ed eventuale imputazione di dati mancanti	6
Descrizione e caratteristiche delle variabili.....	7
Citazione dell'archivio.....	7
Appendice: Variabili presenti negli archivi e non incluse nella rilevazione.....	8

Informazioni generali

L'indagine sulle Imprese Industriali e dei Servizi (INVIND) raccoglie ogni anno informazioni su aspetti anagrafici e strutturali delle imprese italiane su un campione di circa 3.000 imprese appartenenti all'industria in senso stretto, 1.000 dei servizi privati di natura non finanziaria (questi ultimi comprendono commercio, alberghi e ristorazione, trasporti e comunicazioni, attività immobiliari, informatica e altri servizi privati) e 500 del settore delle costruzioni.

Alle imprese sono richieste informazioni di natura prevalentemente quantitativa sull'andamento delle principali variabili economiche. Il questionario (occupazione, investimenti, fatturato, risultato d'esercizio, capacità produttiva, indebitamento). Ogni anno sono inoltre introdotte sezioni monografiche dedicate a specifici temi di interesse.

Popolazione di riferimento

L'universo di riferimento è costituito dalle imprese con sede amministrativa in Italia appartenenti ai settori di attività economica dell'industria in senso stretto e dei servizi privati non finanziari (con almeno 20 addetti) e a quello delle costruzioni (con almeno 10 addetti).

Fino all'indagine sul 1998, la rilevazione si limitava alle imprese del settore manifatturiero con 50 addetti e oltre. A partire dall'indagine sul 1999, l'universo di riferimento è stato ampliato a tutta l'industria in senso stretto, integrando il campione con imprese della Sottosezione ATECO 2007 (derivata dalla Nace Rev. 2) B (estrazione di minerali da cave e miniere), D (fornitura di energia elettrica, gas, vapore e aria condizionata) ed E (fornitura di acqua, reti fognarie, attività di gestione dei rifiuti e risanamento).

Nel 2001 l'indagine è stata estesa (con un questionario ridotto) anche alle imprese con 20-49 addetti. Dal 2002 nella popolazione di riferimento sono state inserite anche le imprese dei servizi privati non finanziari con 20 addetti e oltre (escludendo dai servizi destinabili alla vendita le imprese del credito e le assicurazioni, i servizi pubblici e gli altri servizi sociali e personali). Dal 2006 la rilevazione si è ampliata anche al settore delle costruzioni con almeno 20 addetti. L'indagine sul 2013 ha esteso il campione delle costruzioni alle imprese con una fascia di addetti tra 10 e 19.

Periodo di disponibilità e frequenza dei dati

Gli archivi vengono aggiornati al termine di ogni rilevazione annuale, sulla base delle risposte fornite dalle imprese nel corso delle interviste.

Il dataset con i dati aggiornati al nuovo anno di riferimento viene messo a disposizione degli utenti in corrispondenza della pubblicazione del relativo fascicolo sulla pagina internet dell'Istituto (cfr. paragrafo Statistiche aggregate e pubblicazioni di riferimento).

Rilevazione dei dati e riferimenti normativi

La rilevazione viene svolta annualmente tra febbraio e maggio, con riferimento all'attività economica di fine anno precedente; le interviste sono condotte dalle Filiali della Banca. Dall'edizione relativa al 2010, il caricamento dei dati è interamente realizzato tramite un'applicazione accessibile via web.

Statistiche aggregate e pubblicazioni di riferimento

Le principali statistiche aggregate sono raccolte in [tavole](#) disponibili alla [pagina web dell'indagine](#) sul sito della Banca d'Italia.

I principali risultati dell'indagine vengono inoltre diffusi attraverso un apposito [fascicolo](#) della [Collana Statistiche](#) della Banca d'Italia e, corredati di altre informazioni, nella [Relazione Annuale](#), nel [Bollettino Economico](#) e nelle pubblicazioni della serie [Economie Regionali](#).

Struttura e caratteristiche degli archivi

Le informazioni raccolte nell'ambito dell'Indagine sulle imprese industriali e dei servizi sono disponibili a partire dal 1984 in un unico archivio (in formato csv); ogni rilevazione è identificabile a partire dalla variabile **annoril**, che indica l'anno oggetto di rilevazione, ovvero l'anno precedente quello di effettiva raccolta dei dati (ad esempio i dati relativi all'indagine sul 2019, condotta nei primi mesi del 2020, sono identificati dalla variabile **annoril=2019**).

Poiché i questionari rivolti alle imprese di costruzione differiscono significativamente da quelli indirizzati alle imprese dell'industria in senso stretto e dei servizi, anche i rispettivi archivi sono separati: il dataset con le sole imprese dell'industria in senso stretto e dei servizi viene denominato "indann_completo_csv", mentre quello con le sole imprese delle costruzioni viene denominato "costr".

In entrambi i dataset, ogni impresa ha un codice identificativo (la variabile **ident**), che consente l'aggancio delle informazioni nel caso in cui sia stata oggetto di rilevazione in più anni. Questo codice, generato in modo casuale, è del tutto incorrelato con le variabili identificative delle imprese e serve esclusivamente per le analisi longitudinali. La coppia (**annoril**, **ident**) identifica le informazioni riguardanti una data impresa in un dato anno.

Gli archivi contengono alcune variabili relative al riporto all'universo delle stime campionarie. L'adozione del peso campionario consente di allineare la struttura del campione a quella dell'universo secondo le variabili di stratificazione; se ne raccomanda l'uso nelle analisi per ottenere stime non distorte che riportino alla popolazione di riferimento.

Per le variabili che sono fornite dalle imprese sotto forma di rapporto, o comunque prive di fattore di scala, si dovrebbero in generale adottare pesi che tengano conto anche della dimensione del fenomeno. Gli archivi contengono anche alcune variabili relative alle classificazioni adottate per la stratificazione secondo il disegno campionario. Va tenuto presente che l'area geografica è considerata nel disegno dell'indagine come variabile di post-stratificazione.

Le informazioni monetarie provenienti dal questionario sono espresse a prezzi correnti e in migliaia di euro (ad eccezione delle variabili relative alle retribuzioni, espresse in euro), anche per quelle riferite agli anni in cui non era entrato in vigore l'euro.

A partire dall'indagine sul 2010, alcune variabili monografiche specificamente indicate nel database, sono state rilevate solo su metà del campione; le sezioni dei questionari che le contengono sono chiaramente contrassegnate dalla lettera "A" o "B", per indicare a quale metà del campione esse siano rivolte.

La suddivisione del campione totale in due sotto-campioni è effettuata in base a un meccanismo casuale, tale da mantenere la rappresentatività delle due metà del campione originario rispetto alla popolazione di riferimento. Nelle analisi relative a variabili rilevate su un sotto-campione è

importante utilizzare i pesi di riporto appositamente creati a tal fine (si veda descrizione pesoa e pesob di seguito) per garantire il corretto riporto all'universo anche nel sottoinsieme in esame.

Aspetti metodologici ed eventuale imputazione di dati mancanti

L'indagine adotta una procedura di ponderazione stratificata che si compone di due fasi. In una prima fase si usano come strati le combinazioni di settore di attività economica e classe dimensionale. A ogni impresa si attribuisce un peso iniziale, dato dal rapporto tra numero d'impresе contenute nella cella di strato e numero d'impresе del campione. In una seconda fase si effettua una stratificazione a posteriori, utilizzando la tecnica del raking, per tenere conto anche dell'area geografica in cui è localizzata la sede amministrativa dell'impresa.

In caso di mancata risposta nelle variabili principali (ad esempio la spesa programmata per investimenti, il fatturato o l'occupazione previsti), si procede all'imputazione tramite stimatori per rapporto, utilizzando come denominatore il numero di addetti medi dell'impresa nell'anno di riferimento, per cogliere l'effetto di scala. In alcuni casi nella ricostruzione ci si avvale anche dei dati storici dell'impresa, sotto forma di effetti individuali. In questo modo si stima un livello per addetto ottenuto come combinazione tra una media generale di tipo cross-section e una media calcolata solo sui valori storici dell'impresa. I livelli ai tempi t e $t+1$ sono ricostruiti in sequenza, calcolando variazioni medie in opportune celle contenenti imprese omogenee. Contestualmente al processo di imputazione, viene creata una nuova variabile flag, che permette di riconoscere se il valore nel database è stato fornito direttamente dall'impresa o è stato imputato: il flag assume valore 1 se il dato della relativa variabile di riferimento è frutto del processo di imputazione, altrimenti sono vuoti. Il nome delle variabili flag è sempre del tipo $f''X''$, dove X indica il nome della variabile imputata (ad esempio il flag d'imputazione della variabile occupazione attesa, $v611m$, sarà $fv611m$). La percentuale di dati imputati è comunque molto contenuta per le variabili fondamentali a livello di consuntivo e preconsuntivo. Le domande relative alle previsioni, in particolare quelle inerenti agli investimenti, tendono ad avere un tasso di mancata risposta relativamente più elevato. Può risultare elevato, in funzione della difficoltà dei quesiti, anche il tasso di mancata risposta delle sezioni monografiche.

Le risposte sono sottoposte a un processo di controllo qualità che precede la formazione del dataset completo. Le stime pubblicate nel fascicolo della collana Statistiche fanno riferimento agli archivi come disponibili al momento delle elaborazioni e all'universo delle imprese più recente disponibile. Limitati scostamenti tra l'archivio utilizzato per le stime nel fascicolo e quello messo a disposizione per le elaborazioni a distanza possono essere dovuti sia a successive revisioni delle risposte fornite dalle imprese, sia a revisioni nel sistema di pesi dovute all'aggiornamento della popolazione di riferimento. La revisione del sistema di ponderazione ha carattere sistematico, mentre eventuali rettifiche operate sulle risposte degli anni precedenti sono da considerarsi del tutto episodiche.

Per maggiori dettagli sul disegno campionario, la costruzione dei pesi, dei deflatori e per tutti gli altri aspetti riguardanti la metodologia utilizzata, si rimanda alla [Nota metodologica](#) reperibile sul sito internet della Banca d'Italia.

Descrizione e caratteristiche delle variabili

La descrizione delle variabili e il dominio dei relativi attributi sono contenuti nel seguente file excel.

[Descrizione e caratteristiche delle variabili \(formati delle variabili e codifiche/domini\)](#)

Per motivi di riservatezza, negli archivi non sono disponibili le variabili del questionario che permetterebbero l'identificazione dell'impresa rispondente, solitamente presenti nelle prime pagine dei questionari, tra le quali: Codice Fiscale, Ragione sociale, Filiale di rilevazione e Gruppo di appartenenza. Non sono inoltre rese disponibili le risposte testuali di tipo "Altro specificare". Al contrario, gli archivi contengono alcune variabili non presenti nei questionari (es. anno di riferimento dell'indagine, variabile **annoril**), ma utili ai fini delle elaborazioni (cfr. l'elenco completo in Appendice).

Citazione dell'archivio

Banca d'Italia (2023): Indagine sulle Imprese Industriali e dei Servizi, luglio 2023 (1984-2022).

Appendice: Variabili presenti negli archivi e non incluse nella rilevazione*a) Variabili di classificazione rispetto al settore di attività economica¹*

Nome	Valori	Descrizione	ATECO 2002	ATECO 2007
	SS1	Industrie alimentari, bevande e tabacco	DA	10, 11, 12
	SS2	Industrie tessili, dell'abbigliamento, pelli cuoio e calzature	DB, DC	13, 14, 15
	SS3	Fabbricazione di coke, industria chimica, gomma e plastica	DF, DG, DH	19, 20, 21, 22
	SS4	Industria della lavorazione dei minerali non metalliferi	DI	23
	SS5	Industria metalmeccanica	DJ, DK, DL, DM	24, 25, 26, 27, 28, 29, 30, 33
	SS6	Altre industrie manifatturiere	DD, DE, DN	16, 17, 18, 31, 32
settor11				05, 06, 07, 08,
	SS7	Industria energetico – estrattiva	CA, CB, CE	09, 35, 36, 37, 38, 39
	SS8	Commercio ingrosso e dettaglio, riparazione di autoveicoli e motocicli	G	45, 46, 47
	SS9	Servizi di alloggio e ristorazione	H	55, 56
	SS10	Trasporto e magazzinaggio, servizi di informazione e comunicazione	I	49, 50, 51, 52, 53, 58, 59, 60, 61, 62, 63
	SS11	Attività immobiliari, professionali, scientifiche e tecniche, amministrative e di servizi di supporto	K	68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82
	1	Industria manifatturiera	D	C
indag3	2	Industria energetico – estrattiva	C, E	B, D, E
	3	Servizi	G, I, H, K	G, I, H, J, L, M, N
	1	Industria in senso stretto	C, D, E	C, B, D, E
indagine	2	Servizi	G, I, H, K	G, I, H, J, L, M, N

¹ Fino al 2009 tali variabili erano ottenute dall'aggregazione di alcune sottosezioni ATECO 2002; dal 2010 sono basate sulle prime due cifre della classificazione ATECO 2007. Le variabili di questa tabella sono presenti nel solo archivio "indann_completo_csv.csv", tutte le imprese dell'archivio costruzioni hanno un unico codice ateco (2002, 2007).

b) Variabili di classificazione rispetto alla classe dimensionale²

Nome	Valori	Descrizione
cldimet	0	20 - 49 addetti
	1	50 - 99 addetti
	2	100 - 199 addetti
	3	200 - 499 addetti
	4	500 - 999 addetti
	5	1.000 addetti e oltre
cc2	1	20 - 49 addetti
	2	50 addetti e oltre

c) Variabili di classificazione rispetto all'area geografica³

Nome	Valori	Descrizione
areag4	1	Nord ovest
	2	Nord est
	3	Centro
	4	Sud e Isole
areag2	1	Nord, Centro
	2	Sud e isole

d) Variabili relative al disegno campionario e al sistema di ponderazione

strato: Formato dalle 66 combinazioni di **settor11** e **cldimet** alle quali sono aggiunti due strati riferiti alle imprese con almeno 5.000 addetti, che hanno peso unitario e sono considerate, separatamente per industria e servizi (strati 67 e 68).

poststrato: Formato dalle 48 combinazioni di **areag4**, **cc2** e una riaggregazione dei settori di attività economica in 6 gruppi: 1) indag3=1; 2) indag3=2; 3) settor11=SS8; 4) settor11=SS9; 5) settor11=SS10; 6) settor11=SS11.

peso: peso di espansione all'universo: a livello di strato e post-strato, la somma dei pesi uguaglia la numerosità della popolazione di riferimento, separatamente per ciascun anno senza tener conto della dimensione longitudinale del campione. Alle imprese con più di 5000 addetti (c.d. auto-rappresentative) e ad un numero limitato di imprese che non si ritengono rappresentative dello strato a cui appartengono si attribuisce peso unitario.

pesoadd: Peso campionario (di espansione): a livello di strato e poststrato, la somma dei pesi equivale alla numerosità degli addetti della popolazione di riferimento separatamente per ciascun anno e non tiene conto della dimensione longitudinale del campione (disponibile dal 2007). Questo peso è particolarmente indicato per

² Fino all'anno di riferimento 2003 la classe dimensionale è riferita al numero di addetti a fine anno; dal 2004 in poi al numero di addetti medi nell'anno.

³ Per motivi di riservatezza non sono disponibili le classificazioni relative alle singole regioni e province, ma solo quelle relative alle macro-aree geografiche.

la ponderazione delle variabili di tipo categorico, perché consente di tener conto della diversa scala dimensionale delle imprese.

- pesoa:** Equivalente a **peso** per le imprese appartenenti alla prima metà del campione (presente dal 2010, da usare per i quesiti monografici rilevati sulla prima metà del campione).
- pesoadda:** Equivalente a **pesoadd** per le imprese appartenenti alla prima metà del campione (presente dal 2010, da usare per i quesiti monografici rilevati sulla prima metà del campione).
- pesob:** Equivalente a **peso** per le imprese appartenenti alla seconda metà del campione (presente dal 2010, da usare per i quesiti monografici rilevati sulla seconda metà del campione).
- pesoaddb:** Equivalente a **pesoadd** per le imprese appartenenti alla seconda metà del campione (presente dal 2010, da usare per i quesiti monografici rilevati sulla seconda metà del campione).
- pesorisc:** La variabile **pesorisc** è ottenuta come prodotto di **peso** e un opportuno fattore di scala, in modo che essa, anno per anno, sommi alla numerosità del campione.
- popstr:** Numerosità della popolazione a livello di strato.
- poppost:** Numerosità della popolazione a livello di post-strato.

e) *Variabili di classificazione della quota del fatturato esportato*

Nome	Valori	Descrizione
a6	0	impresa non esportatrice
	1	meno di 1/3 di fatturato esportato
	2	tra 1/3 e 2/3 di fatturato esportato
	3	oltre 2/3 di fatturato esportato
qexp	1	meno di 1/3 di fatturato esportato o nessuna esportazione
	2	tra 1/3 e 2/3 di fatturato esportato
	3	oltre 2/3 di fatturato esportato

f) *Variabili i cui livelli sono disponibili a prezzi correnti e costanti*

Negli archivi, limitatamente ai livelli di investimenti e fatturato, sono presenti variabili espresse sia a prezzi correnti che a prezzi costanti. I prezzi costanti sono riferiti sia all'anno di riferimento più recente disponibile sia all'anno di riferimento di ogni singola indagine. Questi ultimi consentono di calcolare variazioni a prezzi costanti nell'anno in cui sono stati rilevati anche dopo l'aggiunta di nuove edizioni dell'indagine. Per le imprese dell'Industria in senso stretto e dei servizi, i deflatori sono ottenuti partendo da quelli forniti dalle stesse imprese e utilizzando una metodologia di aggregazione a livello di sottosezione e classe dimensionale.

Descrizione	Var. a prezzi correnti	Var. a prezzi costanti dell'anno più recente disponibile	Var. a prezzi costanti dell'anno di riferimento della rilevazione
Investimenti fissi t-1	v200	v200cos ^(a)	v200k ^(a)
Investimenti fissi t	v202	v202cos	v202k
Investimenti fissi t+1	v203	v203cos	v203k
Fatturato t-1	v209	v209cos	v209k
Fatturato t	v210	v210cos	v210k
Fatturato t+1	v437	v437cos	v437k
Fatturato esportato t-1	v211	v211cos	v211k
Fatturato esportato t	v212	v212cos	v212k
Fatturato esportato t+1	v438	v438cos	v438k
Investimenti immateriali ^(b) t-1	v810	v810cos	v810k
Investimenti immateriali ^(b) t	v811	v811cos	v811k
Investimenti immateriali ^(b) t+1	v812	v812cos	v812k

Note: (a) Disponibile dal 1985; (b) Secondo il SEC2010, la voce "investimenti materiali" comprende la spesa per software, basi di dati e prospezioni minerarie, mentre sono esclusi i brevetti e i marchi.