

Nota metodologica

Introduzione e quadro normativo

La rilevazione sulla formazione del personale è una indagine campionaria condotta mediante interviste alle imprese, il cui obiettivo primario è la stima dei principali aggregati sull'offerta di formazione professionale. Le principali caratteristiche della rilevazione, dagli aspetti metodologici alle definizioni delle variabili e degli indicatori, sono armonizzate a livello europeo. Il regolamento che istituisce una indagine campionaria armonizzata sulla formazione nei Paesi dell'Unione europea è il Regolamento Ue n. 1153/2014 della Commissione Europea. L'indagine è inserita nel Piano statistico nazionale (edizione in vigore: Psn 2017-2019 - aggiornamento 2019), pubblicato nella Gazzetta Ufficiale - serie generale - n. 35 del 11 febbraio 2021.

Nuova definizione dell'unità statistica di analisi dall'anno 2020

L'Istituto nazionale di statistica è stato impegnato negli ultimi anni nella ricerca di metodologie e nello sviluppo di tecniche volte all'implementazione nel sistema dei registri e dei conti economici delle imprese di una nuova unità statistica 'impresa'. La definizione di tale nuova unità statistica tiene conto delle relazioni che intercorrono tra unità giuridiche appartenenti allo stesso gruppo di imprese. Il 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, definisce l'impresa come "la più piccola combinazione di unità giuridiche che costituisce un'unità organizzativa per la produzione di beni e servizi che fruisce d'una certa autonomia decisionale". La completa applicazione del Regolamento prevede quindi l'aggregazione di più unità giuridiche, qualora queste non abbiano sufficiente autonomia nel processo decisionale. Ne consegue che l'impresa può corrispondere a una sola unità giuridica o ad un gruppo di unità giuridiche sottoposte a comune controllo. In particolare, le principali innovazioni introdotte hanno avuto un impatto sulle seguenti variabili, per le sole imprese appartenenti a gruppi d'impresa: - numero di unità (imprese) - livelli di fatturato e di costo per beni e servizi - distribuzione per classi dimensionali e settori di attività economica delle variabili economiche e di struttura, in particolare del valore aggiunto.

A seguito dell'implementazione, il nuovo Registro Asia-Imprese o Asia Ent (Ent=enterprise) è composto per la maggior parte da imprese indipendenti dove 1 impresa = 1 unità giuridica e da imprese complesse, formate da più unità giuridiche appartenenti a uno stesso gruppo.

Popolazione di riferimento, unità di rilevazione e di analisi

La popolazione di riferimento dell'indagine è costituita dalle Ent attive e residenti in Italia che, in base all'archivio Asia Ent 2019, hanno almeno 10 addetti complessivi ed esercitano attività economica principale in uno dei seguenti settori della classificazione Ateco 2007: estrazione di minerali da cave e miniere (B); attività manifatturiere (C); fornitura di energia elettrica, gas, vapore e aria condizionata (D); fornitura di acqua, reti fognarie, attività di gestione dei rifiuti e risanamento (E); costruzioni (F); commercio all'ingrosso e al dettaglio e riparazione di autoveicoli e motocicli (G); trasporto e magazzinaggio (H); servizi di alloggio e di ristorazione (I); servizi di informazione e comunicazione (J); attività finanziarie e assicurative (K); attività immobiliari (L); attività professionali, scientifiche e tecniche (M); noleggio, agenzie di viaggio e servizi di supporto alle imprese (N); attività artistiche, sportive, di intrattenimento e divertimento (R); altre attività di servizi (S).

L'unità di rilevazione è l'Ug – Unità Giuridica, mentre l'unità di analisi è l'impresa (Ent) così come definita nella popolazione di riferimento.

Il disegno di campionamento

Il piano di campionamento è stato implementato in analogia alle precedenti edizioni dell'indagine, considerando l'unità Ent in luogo dell'Unità giuridica (Ug). La rilevazione è campionaria per le Ent aventi un numero di addetti compreso tra 10 e 249 e censuaria per il restante sottoinsieme di imprese della popolazione ed il piano di campionamento è di tipo probabilistico: casuale, stratificato, a uno stadio, con probabilità di selezione costante per tutte le Ent appartenenti allo stesso strato. Gli strati sono stati definiti concatenando le modalità delle seguenti variabili: 20 settori di attività economica¹, 6 classi di addetti (10-19, 20-49, 50-249, 250-499, 500-999 e 1.000 o più addetti medi) e 21 ripartizioni territoriali corrispondenti al livello 2 della classificazione europea Nuts². La stratificazione adottata costituisce la partizione minima della popolazione che permette di ottenere i domini di stima pianificati come aggregazione di strati elementari. La numerosità campionaria è stata determinata tenendo conto della richiesta di replicare la dimensione in termini di Ug della scorsa edizione (anno di riferimento 2015, in cui l'unità di rilevazione e di analisi era l'Ug. Nella precedente edizione, il campione era risultato pari a circa 35.000 unità, con un tasso di sondaggio del 19%. Poiché a differenza dell'edizione 2015, il piano di campionamento e la selezione casuale delle unità sono stati effettuati rispetto all'unità di analisi Ent, l'insieme delle Ug costituenti le unità di rilevazione dell'indagine CVTS 2020 è stato ricavato “esplosando” il file del campione teorico di Ent ricavato in seguito alla selezione casuale.

Allo scopo di limitare l'onere statistico gravante sulle imprese, è stata adottata una tecnica di selezione coordinata delle unità (tecnica di Jales³) che ha consentito di ridurre la probabilità di selezionare unità già estratte nell'ambito di altre indagini sulle imprese svoltesi negli ultimi tre anni e della precedente edizione della stessa indagine (anno di riferimento 2015).

La raccolta delle informazioni

L'intervista all'impresa è stata condotta nel periodo settembre 2021 – marzo 2022 attraverso la compilazione online di un questionario elettronico. Il modulo di compilazione è stato strutturato nelle seguenti 6 sezioni: Sezione A - Informazioni generali sull'impresa e organizzazione della formazione: - numero di dipendenti distinti per sesso e qualifica professionale, ore lavorate totali, costo del lavoro, introduzione di innovazioni nei prodotti e/o processi di produzione dell'impresa, responsabile della formazione, programmi e/o budget per attività formative, ruolo delle strutture sindacali nelle decisioni relative alla formazione, competenze professionali e modalità di acquisizione delle stesse. Sezione B - Attività di formazione aziendale: - corsi a gestione interna/esterna, altre tipologie di formazione e percentuali di partecipanti, contributi versati e finanziamenti ricevuti da soggetti esterni per lo svolgimento dei corsi di formazione. Sezione C - Corsi di formazione aziendale: - partecipanti a corsi di formazione distinti per sesso e qualifica professionale, ore di corso a gestione interna/esterna, per materia/oggetto della formazione e per organismo erogatore, costi sostenuti per i corsi di formazione. Sezione D - Qualità e valutazione della formazione: - iniziative assunte dall'impresa per la qualità dell'attività di formazione aziendale, risultati e modalità di valutazione delle attività di formazione, fattori limitativi all'attività di formazione. Sezione

¹ 1 B, C10-C12, C13-C15, C17-C18, C19-C23, C24-C25, C26-C28+C33, C29-C30, C16+C31-C32, D-E, F, G45, G46, G47, H, I, J, K64-K65, K66, L+M+N+R+S.

² Ai sensi del Regolamento Ce n.1059/2003, si tratta delle regioni amministrative con le province autonome di Trento e Bolzano considerate separatamente. La regione attribuita all'impresa è quella della sede legale o amministrativa come risulta dall'archivio Asia di riferimento.

³ C.d coordinamento negative. Ohlsson, E. (1995), Coordination of samples using permanent random numbers. In: Business Survey Methods (eds. Cox, B. G., Binder, D. A., Chinnapa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S.), Wiley, New York, Chapter 9, 153–169.

E - Fattori di ostacolo alla formazione aziendale: - problemi che hanno impedito l'attività di formazione. Sezione F - Formazione professionale iniziale: - numero di apprendisti nel 2020, motivi della loro assunzione presso l'impresa.

L'elaborazione dei dati: processo, strumenti e tecniche

Le stime finali delle variabili di interesse sono state calcolate a partire da 17.617 Ent del campione teorico, corrispondenti alle 28.510 Ug distinte che hanno fornito all'indagine risposte valide. Tale insieme è stato ottenuto dall'insieme delle Ent con almeno una Ug rispondente che, in base alla fonte Asia Ent 2020, presentano le seguenti caratteristiche: sono attive, appartengono al campo di osservazione dell'indagine e hanno Ug rappresentativa attiva per almeno sei mesi nel 2020. Le stime finali sono state calcolate secondo la teoria dello stimatore di ponderazione vincolata⁴, utilizzato nella maggior parte delle indagini campionarie Istat sulle imprese. Utilizzando questo metodo, ad ogni Ent rispondente è stato associato un peso finale, ottenuto risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti dk e i pesi finali wk ; i vincoli sono rappresentati da condizioni di uguaglianza delle stime campionarie di variabili correlate a quelle di interesse (c.d. variabili ausiliarie) ai corrispondenti totali della popolazione, i c.d. totali noti, entro determinati domini di interesse.

L'archivio di riporto da cui sono stati calcolati i totali noti è stato desunto da Asia Ent 2020 ed è costituito dalle 198.683 Ent attive almeno sei mesi nell'anno e appartenenti al campo di osservazione dell'indagine. Le variabili ausiliarie considerate sono: numero di Ent e relativo numero di addetti, disponibili per ciascuna Ent dell'archivio di riporto; la funzione di distanza $G(wk, dk)$ scelta è di tipo euclideo e i domini entro cui i suddetti vincoli risultano soddisfatti sono i seguenti:

- 20 settori di attività economica
- 5 macro settori di attività economica x 6 classi di addetti
- 5 macro settori di attività economica x 21 ripartizioni territoriali dell'Ug rappresentativa.

Al suddetto problema di minimo vincolato è stato aggiunto il c.d. vincolo di range, del tipo seguente: $wk/dk \geq 0$ posto per eliminare il rischio che i pesi finali wk soluzione del problema presentino valori negativi.

I pesi calibrati sono stati calcolati mediante il software ReGenesees⁵, progettato in ambiente R, utilizzato anche per il calcolo degli errori campionari richiesti per la stesura del Quality Report. I pesi calibrati associati alle imprese rispondenti hanno realizzato la convergenza delle stime dei totali delle variabili ausiliarie (numero di imprese e numero di addetti) ai corrispondenti totali noti, entro i domini di stima sopra indicati. L'universo utilizzato per il calcolo dei totali noti, desunto da Asia Ent 2020 e dunque allineato temporalmente al periodo di riferimento dell'indagine, è costituito dalle 198.683 imprese attive almeno 6 mesi nell'anno e appartenenti al campo di osservazione dell'indagine.

L'output: principali misure di analisi

La rilevazione sulla formazione nelle imprese ha l'obiettivo di produrre stime riguardanti le imprese che effettuano

⁴ Deville, J.-C. and Särndal, C.-E. (1992), Calibration Estimators in Survey Sampling Journal of the American Statistical Association, Vol. 87, N°. 418, 376–382.

⁵ ReGenesees è un sistema software basato su R per l'analisi design-based e model-assisted di indagini campionarie complesse. Il software e la relativa documentazione sono disponibili sul sito dell'Istat, al link: <https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it-per-la-produzione-statistica/elaborazione/strumenti-di-elaborazione/regenesees>.

formazione, la tipologia delle attività formative, i costi sostenuti e i partecipanti coinvolti. La popolazione di riferimento viene ripartita in imprese formatrici e non. Inoltre un'importante distinzione della popolazione di riferimento è tra imprese che effettuano formazione attraverso corsi d'aula ed imprese che si avvalgono di altre tipologie, come ad es.: training on the job, partecipazione a convegni, rotazione delle mansioni lavorative, autoapprendimento mediante formazione a distanza.

La precisione delle stime

La dimensione del campione delle Ent negli strati è stata calcolata risolvendo un problema di allocazione multivariata e multi dominio, in cui la dimensione campionaria è quella minima che consente di ottenere simultaneamente predefiniti livelli di accuratezza delle stime di determinati parametri di interesse, entro domini di stima pianificati. Si tratta di una generalizzazione al caso di più domini di stima dell'algoritmo proposto da Bethel⁶. Le variabili per cui sono stati fissati gli errori campionari massimi (in termini di coefficienti di variazione percentuali) sono: il numero di addetti e i ricavi, desunti da Frame SBS Ent relativo al 2019, nei seguenti domini, considerati separatamente:

- 1) 20 settori di attività economica
- 2) 6 classi di addetti
- 3) 5 macro settori di attività economica x 6 classi di addetti
- 4) 5 macro settori di attività economica x 21 Nuts2 dell'UG rappresentativa
- 5) Totale.

Il calcolo dell'allocazione ottima, effettuato mediante il software generalizzato MAUSS-R implementato in Istat, ha dato luogo a una dimensione complessiva pari a 31.622 Ent e al seguente sistema di vincoli per gli errori campionari (cv%) massimi delle variabili di interesse riportato nella seguente tabella:

PROSPETTO 1. CV% PIANIFICATI PER I DOMINI DI INTERESSE

Tipo di dominio	Descrizione	N° addetti	Fatturato
DOM1	20 settori della classificazione Ateco 2007 (cfr. nota 2)	1,0	2,5
DOM2	6 classi di addetti	0,8	2,0
DOM3	5 macro settori di attività (estrattivo e manifattura; energia e acqua; costruzioni; servizi non finanziari; servizi finanziari) x 6 classi di addetti	1,5	3,0
DOM4	5 macro settori di attività (estrattivo e manifattura; energia e acqua; costruzioni; servizi non finanziari; servizi finanziari) x 6 classi di addetti	2,0	5,0
DOM5	Totale	0,2	0,5

L'esplosione delle Ent nelle Ug corrispondenti ha generato un insieme di 38.849 unità distinte che, a fronte di un universo di 223.809 unità, dà luogo ad un tasso di sondaggio del 17% in termini di Ug. Tale tasso, comunque inferiore a quello del 19% della precedente edizione di CVTS, sembra un compromesso adeguato tra l'esigenza di attestarsi intorno alle 35.000 Ug e quella di ottenere livelli di precisione pianificati delle variabili vicini a quelli delle scorse edizioni dell'indagine

⁶ Bethel J. (1989), "Sample Allocation in Multivariate Surveys", Survey Methodology, 15, pp 47-57.

I questionari utili ricevuti sono stati 17.617 (55,7% del campione teorico).

Informazioni sulla riservatezza dei dati

I dati raccolti dalla rilevazione sulla formazione del personale nelle imprese sono tutelati dal segreto statistico e sottoposti alla normativa sulla protezione dei dati personali. Questi possono essere utilizzati, anche per successivi trattamenti, esclusivamente per fini statistici dai soggetti del Sistema statistico nazionale e possono, altresì, essere comunicati per finalità di ricerca scientifica alle condizioni e secondo le modalità previste dall'art. 7 del Codice di deontologia per il trattamento di dati personali effettuato nell'ambito del Sistema statistico nazionale e dal regolamento comunitario n. 831/2002. Le stime diffuse in forma aggregata, sono tali da non poter risalire ai soggetti che hanno fornito i dati o a cui si riferiscono.

Diffusione

I dati sono disponibili nel formato "tavole di dati" e l'intero set informativo sarà disponibile nei prossimi mesi presso il laboratorio Adele. Il Laboratorio ADELE (per l'Analisi dei Dati ELEMENTARI) è un ambiente "sicuro" in cui ricercatori di università, istituti, enti di ricerca o organismi, cui si applica il Codice di deontologia per i trattamenti statistici effettuati al di fuori del Sistan (allegato A.4 del D.lgs. 30 giugno 2003, n. 196), possono condurre analisi statistiche che necessitano dell'utilizzo di dati elementari.