

NOTA METODOLOGICA

Caratteristiche e contenuti informativi della rilevazione

Nel quadro del Sistema europeo delle statistiche su retribuzioni e costo del lavoro, la rilevazione sulla struttura delle retribuzioni (SES) è una indagine quadriennale il cui obiettivo è di monitorare, per *policy-making* e per scopi di ricerca, la struttura e la distribuzione delle retribuzioni negli stati membri dell'Unione Europea.

La SES del 2010 è disciplinata dal Regolamento del Consiglio 530/99¹, dal Regolamento attuativo della Commissione 1738/2005 e dal Regolamento della Commissione 698/06 sulla valutazione della qualità. La SES è stata anche inserita nel Regolamento della Commissione 831/02 sul rilascio dei microdati per scopi scientifici².

Gli indicatori prodotti dalla rilevazione, in particolare quelli retributivi e sull'orario di lavoro, incorporano gli effetti indotti all'interno di ogni Paese dalle condizioni del mercato del lavoro nazionale (grado di flessibilità, assetti istituzionali e del *welfare*), come pure una diversa composizione dell'occupazione (per settore di attività economica e posizione di lavoro) e un accesso diversificato, ad esempio, per titolo di studio.

In particolare, la SES fornisce l'indicatore del differenziale salariale per genere (*gender pay gap*)³. Sulla base delle stime del 2010, a livello europeo si è stabilito di elaborare annualmente tale indicatore, anche negli anni non coperti dalla rilevazione, secondo una metodologia stabilita all'interno del gruppo di lavoro *Labour Market Statistics* (LAMAS) dell'Eurostat. Ad ogni Paese è stata lasciata facoltà di selezionare le fonti nazionali da utilizzare per aggiornare l'indicatore negli anni non coperti dalla rilevazione. L'Istat utilizza i dati annuali della rilevazione annuale sulle famiglie EuSilc per stimare il *gender pay gap* per gli anni non coperti dalla rilevazione SES.

La struttura dell'indagine sulle retribuzioni si riferisce all'anno 2010, per le variabili annuali, e ad Ottobre 2010 per i dati mensili⁴.

La SES relativa all'anno 2010 ha come universo di riferimento le imprese a totale controllo privato o con prevalente controllo pubblico e gli enti e istituzioni pubbliche con almeno 10 dipendenti attive nell'industria, nei servizi orientati al mercato e nei servizi sociali e personali (le sezioni da B a S della classificazione Ateco 2007) mentre sono esclusi i settori dell'agricoltura, silvicoltura e pesca e della pubblica amministrazione (sezioni A e O).

Indipendentemente dal tipo di controllo economico, le imprese con almeno 250 dipendenti vengono rilevate in modo censuario, mentre per quelle con una dimensione inferiore e con almeno 10 dipendenti viene selezionato un campione.

Gli enti e istituzioni pubbliche sono invece rilevate in maniera censuaria.

Per il settore della scuola pubblica l'universo di riferimento viene definito attraverso un processo di integrazione tra archivi fiscali, amministrativi e da indagini (quelle sulle famiglie sulle Forze Lavoro ed EuSilc), escludendo le unità locali con meno di 10 dipendenti.

¹ Il Regolamento della Commissione n° 530/1999 prevede l'alternarsi ogni due anni delle due rilevazioni sulla struttura delle retribuzioni e sulla struttura del costo del lavoro (*Labour Costs Survey*) prevista sull'anno 2012.

² L'Eurostat acquisisce direttamente i microdati che sono diffusi attraverso il proprio sito web (all'interno dell'Eurostat Dissemination Database) in forma aggregata secondo modalità concordate con i Paesi prima dell'avvio della rilevazione. I microdati vengono messi a disposizione dei ricercatori, previa l'anonimizzazione di quei record per i quali non è garantita la confidenzialità del dato individuale, come File per la Ricerca (MFR) sia dall'Istat, sia attraverso il Safe Center dell'Eurostat.

³ La retribuzione oraria mensile media per il calcolo del *gender pay gap* viene ottenuta come media ponderata delle retribuzioni orarie mensili osservate, con pesi relativi dati dai fattori di ponderazione di indagine. Nelle diffusioni Istat la retribuzione mensile media viene calcolata come rapporto tra i monti delle due quantità relative all'intera popolazione, rispettivamente retribuzioni mensili e numero di ore osservate nel mese moltiplicate per i pesi relativi di indagine. La differenza tra le due modalità di calcolo è rappresentata da un coefficiente moltiplicativo, dato dal peso delle ore retribuite delle singole unità sulle ore medie di aggregato. Al crescere della distanza media (la dispersione) dei valori osservati dal valore medio di aggregato, aumenta la differenza tra i due indicatori.

⁴ L'Eurostat lascia facoltà ai Paesi della UE di selezionare un mese indicativo per la rilevazione, caratterizzato da minori tassi di assenza per malattia, bassa stagionalità della produzione e in cui non siano presenti feste nazionali che riducono il numero dei giorni lavorati. Nella generalità dei Paesi tra cui l'Italia, i dati mensili fanno riferimento al mese di ottobre.

Basate su criteri armonizzati, le indicazioni fornite dalla rilevazione permettono di documentare come la struttura occupazionale e l'orario di lavoro influenzino direttamente l'esistenza e la persistenza dei differenziali retributivi nazionali (territoriali, settoriali, di genere, etc.) in funzione di alcune caratteristiche personali e professionali dei singoli dipendenti. Le stime derivate dalla SES consentono di analizzare sia gli aspetti collegati al "capitale umano" in termini di abilità e conoscenze (come l'età, il livello di istruzione e formazione acquisito e l'anzianità aziendale) sia quelli strettamente correlati all'inserimento nel posto di lavoro (come la professione, il tipo e la durata del contratto di lavoro).

In particolare, le informazioni obbligatorie a livello di unità locali riguardano:

- l'attività economica, a livello di due cifre della classificazione Nace Rev.2;
- la localizzazione geografica; secondo la classificazione NUTS a una cifra;
- la dimensione dell'unità locale e dell'impresa (ente o istituzione);
- il tipo di controllo economico e finanziario;
- l'esistenza o meno di contratti collettivi di lavoro.

Quelle relative ai dipendenti sono :

- genere ed età;
- professione, secondo la classificazione ISCO08 a due cifre;
- livello più elevato di istruzione e formazione personale raggiunto, secondo la classificazione ISCED;
- anzianità di servizio nella struttura dove lavora;
- durata del tempo di lavoro (a tempo pieno o a tempo parziale) e tipo di contratto (a tempo indeterminato, determinato o apprendistato);
- numero di ore retribuite e di straordinario nel mese di ottobre e nell'anno;
- numero di settimane a cui si riferisce la retribuzione annuale, numero di giorni di assenza;
- retribuzione lorda, per straordinari e per indennità di turno nel mese di ottobre;
- retribuzioni lorde annuali.

Dal punto di vista metodologico, la SES 2010 incorpora, migliorandole, le innovazioni nel processo di produzione statistica sperimentate nel ciclo 2006, anno in cui è stato realizzato l'ampliamento della rilevazione alle istituzioni pubbliche. Rispetto al ciclo precedente, nel 2010 si è effettuata una indagine diretta ed esaustiva anche per la sanità e le università a controllo pubblico. Per la scuola è stata migliorata la procedura utilizzata nel 2006

Quindi, come di seguito specificato, nel 2010 è stata effettuata una indagine diretta sulle imprese (indipendentemente dal tipo di controllo economico) e sugli enti e istituzioni pubbliche, mentre le stime nel settore della scuola pubblica sono derivate dallo stesso processo di integrazione di archivi già effettuato nel 2006, migliorato da informazioni aggiuntive che si sono rese disponibili.

Sono state rilevate 7.912 imprese a controllo privato o prevalente controllo pubblico (a cui corrispondono a 24.540 unità locali e 214.542 dipendenti) e 7.718 unità locali degli enti e istituzioni pubbliche comprese le scuole pubbliche (a cui corrispondono 50.190 dipendenti).

Stime derivanti da indagine diretta

L'indagine diretta ha riguardato le imprese che hanno come universo di riferimento l'archivio ASIA delle imprese con almeno 10 dipendenti e gli enti e istituzioni osservate nella lista delle unità istituzionali che fanno parte del settore delle Amministrazioni Pubbliche (Settore S13)⁵, escludendo, come previsto dal Regolamento della U.E quelle appartenenti alla Pubblica Amministrazione (Comuni, Province, Comunità montane, Ministeri, etc.)

Fanno quindi parte della lista sottoposta a indagine diretta:

⁵ Per il Settore S13 non sono disponibili i dati occupazionali ufficiali. Quindi non è stato possibile effettuare una preselezione delle unità con meno di 10 dipendenti, da escludere dalla rilevazione, che nel corso dell'indagine diretta sono state individuate ed esentate dalla rilevazione, a seguito di autocertificazione"

- tutte le imprese che indipendentemente dal tipo di controllo economico hanno almeno 250 dipendenti;
- una selezione di quelle imprese che hanno da 10 a 249 dipendenti, risultante dall'applicazione di tecniche di allocazione multivariata e sovradimensionata sulla base dei tassi di caduta di risposta osservati nella rilevazione SES del 2006;
- tutti gli enti e istituzioni della lista S13.

L'insieme delle unità statistiche così individuate rappresenta il primo stadio del campione della SES 2010, costituito dalle unità stratificate per settori di attività economica (a due cifre della classificazione Nace Rev.2), per sette classi di dipendenti (10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000 e oltre) e per le cinque ripartizioni territoriali (secondo la classificazione NUTS1)⁶. Il secondo stadio è rappresentato da un campione di dipendenti la cui numerosità dipende dalla dimensione dell'impresa, ente o istituzione in cui è occupato.

Il numero di dipendenti campionato all'interno di ciascuna unità statistica è stato prefissato adottando le indicazioni suggerite da Eurostat⁷ e sulla base dell'esperienza maturata nei cicli precedenti. Il Prospetto 1 riporta la numerosità campionaria di secondo stadio secondo la dimensione dell'impresa, ente o istituzione.

Complessivamente il campione iniziale è risultato pari a 19.534 imprese corrispondenti a circa 480.000 dipendenti attesi.

PROSPETTO 1 - NUMEROSITÀ CAMPIONARIA DI SECONDO STADIO PER CLASSI DI DIPENDENTI DELL'UNITÀ CAMPIONATA.

Classe di dipendenti	Dipendenti da campionare
da 10 a 19	tutti
da 20 a 49	20
da 50 a 99	25
da 100 a 249	35
da 250 a 499	40
da 500 a 999	50
da 1000 a 1999	60
da 2000 a 3999	65
da 4000 a 7499	75
da 7500 a 9999	100
10000 e oltre	200

Il campione finale è costituito da 8.870 imprese, enti o istituzioni e da 251.702 dipendenti ed è risultato sufficiente ad assicurare la significatività statistica dei risultati, dato l'iniziale sovra-campionamento del 30%.

Per le unità di maggiore dimensione, sono stati utilizzati⁸ i dati ricavati dall'archivio delle dichiarazioni fiscali dei modelli 770 dell'anno 2010 (riferiti all'anno d'imposta 2009) che hanno consentito sia la stratificazione del campione di secondo stadio, sia la selezione guidata dei dipendenti da campionare. In particolare, le informazioni delle dichiarazioni 770 sono state utilizzate per le 5.244 delle 5.367 delle imprese con almeno 100

⁶ I domini di stima sono definiti a livello degli strati campionari, con la sola eccezione per la distinzione secondo la classe di dipendenti che nella fase di definizione dei domini di stima vede aggregate le due classi di maggiore dimensione (cioè si ha 500 dipendenti ed oltre).

⁷ Doc. Eurostat/E1/Sal/5.2/02 (Room document) PHARE 1999 Multi-Country Programme: Pilot Project on Labour Cost and Earnings Statistics- Model SES Questionnaires : One for Local Units and One for Enterprises -prepared by Christophe Demunter- January 2002

⁸ Già nel 2006 si era proceduto a tale selezione del campione di dipendenti effettuata per genere e per classe di età, sulla base di quanto sviluppato nel 2005 nell'ambito dell'ambito del Gruppo di Lavoro Istat MEF/Dipartimento delle politiche fiscali per l'utilizzo a fini statistici dell'archivio dei Modelli 770.

dipendenti, cioè per quelle in cui la partita IVA (o codice fiscale) ha trovato l'aggancio con il corrispettivo dei sostituti d'imposta presente nell'archivio dei modelli fiscali 770.

Attraverso i codici fiscali (o partita IVA) dei percipienti, i dipendenti sono stati stratificati per sei classi di età espresse in anni (minore di 20, 20-30, 30-40, 40-50, 50-60, oltre 60 anni) e per genere, variabili che presentano un elevato livello di correlazione con i livelli retributivi percepiti dai dipendenti. Ai rispondenti sono stati forniti i codici fiscali dei dipendenti per i quali fornire le informazioni richieste⁹.

La rilevazione è stata condotta mediante auto compilazione di questionari, con solleciti telefonici e postali volti a garantire una buona copertura del campione base. Nel 2010, per la prima volta, gli intervistati hanno avuto l'opportunità di compilare un questionario elettronico sul Web con accesso tramite password fornita individualmente. Il questionario è paginato in diverse sezioni, la maggior parte delle quali afferenti tematicamente ai dati dell'impresa, ente o istituzione nel suo complesso, mentre l'ultima sezione è tesa a rilevare le caratteristiche personali, professionali ed economiche di ciascun dipendente campionato.

I quesiti presentano le variabili secondo varie configurazioni topologiche: riga, colonna, tabella (ogni colonna è una variabile), matrice (ogni cella è una variabile) e sono corredati da suggerimenti, note e *tooltip*. Il suddetto questionario web è stato sviluppato attraverso l'applicazione GINO++ (Gathering Information Online), un software generalizzato per l'acquisizione di dati via web progettato come strumento *general purpose*. L'applicazione ha sviluppato diverse funzioni. Si è proceduto alla costruzione grafica del questionario e all'implementazione del piano di controllo, sia guidando la compilazione attraverso l'abilitazione/disabilitazione contestuale di parti del questionario, sia attraverso l'inserimento di controlli bloccanti o meno e di coerenza.

Il passaggio alla compilazione web, coadiuvata per un ristretto numero di casi con il cartaceo (326 imprese), così come l'opzione di un inserimento massivo di dati attraverso un tracciato record per le unità complesse (210 imprese) ha consentito di raggiungere una più efficace acquisizione dei dati con un significativo ritorno sia in termini quantitativi (numero dei rispondenti e dei dipendenti ad essi collegate), sia in termini qualitativi (ossia qualità e coerenza del dato raccolto).

Il piano di controllo e correzione dei record-dipendente si compone di due parti principali. La prima riguarda le variabili anagrafiche (sex, età, titolo di studio, qualifica professionale, etc.) per le quali i valori delle mancate risposte sono stati imputati utilizzando la metodologia del donatore di minima distanza con il software RIDA (Ricostruzione informazioni con donatore automatica). La seconda parte è relativa al gruppo di variabili numeriche riguardanti la sezione dell'orario di lavoro e delle retribuzioni. Per queste variabili, un primo passo è stato quello di individuare le unità "pericolose" (con il software SeleMix), in quanto influenzano in modo significativo le stime. Dopo una loro revisione interattiva, è stato usato il software Banff¹⁰, che applica il principio del minimo cambiamento per localizzare le variabili errate e imputarle con il metodo del donatore.

Il riporto dei dati all'universo è stato effettuato con l'ausilio della procedura solitamente utilizzata per le indagini campionarie sulle imprese, definendo i pesi a livello di imprese, unità locali e dipendenti. La metodologia adottata si basa sullo stimatore di ponderazione vincolata e si avvale delle informazioni strutturali ausiliarie fornite dall'archivio delle imprese di riferimento ASIA 2010 e ASIA unità locali 2009 per la stima dei pesi finali elaborati a livello di imprese, unità locali e dipendenti.¹¹

⁹ Prima di avviare il processo di controllo e correzione, sono state effettuate valutazioni e correzioni di coerenza dei più significativi indicatori. I microdati EMens di fonte INPS sono stati utilizzati per la correzione delle mancate risposte parziali e totali, laddove possibile l'aggancio per codice fiscale tra unità rilevate e archivio INPS. Questi dati non sono comunque esaustivi delle richieste SES e forniscono dati per "cassa" e non di "competenza" come richiesto per la SES.

¹⁰ BANFF è un software generalizzato per il controllo e la correzione dei dati (imputazione) per le variabili numeriche, continue e non-negative, sviluppato dall'Istituto di Statistica Canadese (Statistics Canada).

¹¹ Si assume che le variabili oggetto di indagine siano strettamente correlate con quelle strutturali ausiliarie. Inoltre, è garantito il rispetto di uguaglianza tra i "totali noti" delle variabili ausiliarie e le stime delle stesse, introducendo fattori correttivi dei pesi così calcolati. Tali fattori correttivi tengono conto della mancata risposta totale dovuta all'impossibilità di intervistare alcune imprese selezionate per rifiuto esplicito o per irreperibilità.

Per quanto concerne il peso dell'unità locale, il modello di rilevazione ha registrato, a livello di singolo dipendente, le seguenti informazioni relative all'unità locale dove effettivamente lavora: numero di dipendenti medi, attività economica, ripartizione geografica e provincia di appartenenza. In base a queste informazioni si è effettuato l'aggancio con l'archivio ASIA unità locali del 2009, risultando 14.200 unità locali di riferimento.

Stime derivanti da integrazione di archivi

Nel 2006, l'estensione della rilevazione agli enti e istituzioni pubblici (resa obbligatoria dal Regolamento attuativo della Commissione Europea n. 1738/2005) ha presupposto uno studio di fattibilità al fine di valutare la possibilità di fornire stime per i settori dell'istruzione e della sanità che ne rappresentano la quota prevalente attraverso l'utilizzo di fonti amministrative, fiscali e statistiche.

Questo processo ha avuto l'obiettivo di assicurare i livelli di qualità delle stime richieste da Eurostat, su informazioni personali e professionali, richieste per la prima volta a livello di singolo dipendente, rispetto alle quali non esistono riferimenti statistici di controllo desumibili da altre fonti ufficiali.

Attraverso l'analisi dei contenuti informativi degli archivi amministrativi e statistici disponibili a livello nazionale, è stato possibile predisporre un complesso lavoro di armonizzazione delle classificazioni e definizioni presenti nelle diverse fonti che garantiscono l'integrazione delle informazioni obbligatorie per Eurostat e non presenti negli archivi amministrativi e fiscali considerati.

Per la scuola pubblica, le fonti amministrative e fiscali utilizzate sono rappresentate da archivi amministrativi provenienti da tre diversi Enti: Archivio dei modelli fiscali 770 anno di riferimento 2010 (fonte: Ministero Economia e Finanze); Cedolini mensili stipendiali del 2010 (fonte: MEF- Sistema SPT); Liste dei dirigenti, del personale educativo di ruolo e non di ruolo, docenti dell'infanzia, scuola primaria, scuole medie e superiori per l'anno scolastico 2005-2010 (fonte: Ministero della Istruzione dell'Università e della Ricerca) e liste del personale docente e non docente universitario del 2010 (fonte: Consorzio interuniversitario Cineca). Nel 2010 è stato anche possibile utilizzare un archivio direttamente prodotto dal MIUR relativo alle ore effettivamente lavorate (al netto delle non lavorate ma retribuite), al tipo di contratto (a tempo indeterminato e determinato) e durata del tempo di lavoro (a tempo pieno e a tempo parziale).

I principali problemi metodologici affrontati sono: l'identificazione ovvero riconoscimento delle unità locali e l'integrazione di archivi statistici e amministrativi. La prima questione è formalizzata tramite opportuni grafi che modellando le scuole come una gerarchia di sedi portano alla individuazione dell'unità locale di riferimento¹². Le problematiche relative all'integrazione tra archivi fiscali e amministrativi hanno richiesto la ricostruzione dei codici fiscali di persone fisiche per gli archivi delle indagini Eusilc e Forze di Lavoro, ottenendo un algoritmo di record linkage con gli archivi dei dipendenti della scuola¹³.

La numerosità del campione di secondo stadio del campione è stata definita tenendo conto della proporzione fissa dell'1% negli strati definiti, nell'Universo di riferimento, dall'incrocio di sei classi di retribuzioni¹⁴, sei classi di età¹⁵, genere e cinque classi della classificazione delle professioni ISCO08 a una cifra. Queste variabili obbligatorie sono state validate a livello dell'universo, al fine di garantire una selezione efficiente delle unità campionarie. In ciascuno strato sono stati selezionati tutti quei dipendenti i cui codici fiscali erano presenti contemporaneamente nell'universo di riferimento e nelle Rilevazioni sulle Forze di lavoro o EuSilc (campionamento ragionato); la restante numerosità campionaria, in ogni strato, è stata raggiunta estraendo casualmente altri codici fiscali.

¹² S. Cardinaleschi, V. Spinelli "Archivi amministrativi per le statistiche Eurostat", presentazione Convegno "Prendere decisioni: il ruolo della statistica per la conoscenza e la governance", Università Europea di Roma, SIS VSP 19-20 Aprile 2012, Roma.

¹³ La ricostruzione del codice fiscale è il risultato di un processo di normalizzazione applicato alla Sezione Dati Anagrafici del modello di rilevazione sulle Forze Lavoro. Per maggiori specifiche si veda la Nota definita per la diffusione dei dati del 2006 al seguente link. http://www3.istat.it/dati/dataset/20100614_01/nota_metodologica.pdf.

¹⁴ Minore di 6000 euro; da 6.000 a 19.999 euro; da 20.000 a 31.999 euro; da 32.000 a 39.999 euro; da 40.000 a 49.999 euro; 50.000 euro ed oltre.

¹⁵ Da 16 a 19 anni, da 20 a 29 anni; da 30 a 39 anni; da 40 a 49 anni; da 50 a 59 anni; 60 anni ed oltre.

Per la valutazione e la correzione delle informazioni campionarie si sono utilizzate le stesse tecniche e software indicati in precedenza nella sezione relativa alla indagine diretta, con alcune differenze nell'impostazione del lavoro dovute al fatto che alcune delle variabili obbligatorie vengono validate a livello dell'universo. La stratificazione per la successiva estrazione di un campione è stata effettuata per genere, classi di età, classi di retribuzione e professione ISCO8 a una cifra, tale da assicurare una selezione significativa del campione.

La procedura di riporto all'universo della scuola pubblica si differenzia invece da quanto precedentemente illustrato e da quanto predisposto nel 2006. Il calcolo dei pesi a livello di unità locale e dipendente tiene infatti conto delle caratteristiche campionarie e relativi pesi dei record selezionati dalle rilevazioni Forze di Lavoro ed EuSilc. Sono state ottenute ponderazioni distinte per le unità locali (unità di primo stadio) e per i dipendenti (unità di secondo stadio).