

# **Rilevazione sulle Forze di Lavoro**

Dati longitudinali a 12 mesi di distanza

## **Aspetti metodologici dell'indagine**

Edizione Aprile 2013

## INDICE

1. Introduzione .....	3
2. Il disegno campionario e la componente longitudinale .....	4
3. L'abbinamento dei record individuali.....	5
4. La tecnica di Indagine .....	6
5. Il trattamento dei dati .....	6
6. La popolazione di riferimento.....	9
7. La metodologia di calcolo dei pesi campionari .....	10
8. Le matrici di transizione .....	12
9. La diffusione dei dati longitudinali .....	15
10. Glossario .....	17
11. Riferimenti bibliografici.....	19
12. Contatti .....	20

# 1. Introduzione

La Rilevazione sulle forze di lavoro costituisce dal 1959 la principale fonte informativa sul mercato del lavoro; essa fornisce, con periodicità trimestrale, da un lato la stima del numero degli occupati e dei non occupati (persone in cerca di occupazione e non forze di lavoro) e dall'altro le variazioni tendenziali e congiunturali dell'occupazione e della disoccupazione. Tali stime si riferiscono a tutti i componenti delle famiglie residenti in Italia al netto dei membri permanenti delle convivenze (ospizi, brefotrofi, istituti religiosi, caserme, ecc.) .

La Rilevazione incorpora una componente longitudinale derivante dal sistema di rotazione delle famiglie campione<sup>1</sup>. In particolare, la metà delle famiglie intervistate in un trimestre viene re-intervistata a distanza di 3 e 12 mesi, un quarto a distanza di 15 mesi. Archivi di microdati longitudinali possono essere ottenuti abbinando le informazioni raccolte sugli stessi individui / famiglie in diversi trimestri. Le matrici di transizione desunte dai file in questione, forniscono una stima del numero di permanenze e di transizioni in entrata e in uscita dalle diverse condizioni occupazionali, e consentono di comprendere con un maggiore dettaglio le dinamiche del mercato del lavoro e le caratteristiche degli individui coinvolti.

I file di microdati longitudinali e le relative matrici di transizione costituiscono un "sottoprodotto" della rilevazione stessa, ed è bene sottolineare che non si tratta di un vero e proprio panel relativo a tutta la popolazione. Difatti, un individuo, intervistato la prima volta in uno dei comuni campione, non viene re-intervistato se nell'arco di tempo tra la prima e la successiva intervista ha cambiato residenza o si è trasferito all'estero. Ne consegue che, in un definito arco temporale, la componente longitudinale non rappresenta tutta la popolazione, ma solo quella residente in uno stesso comune sia all'inizio sia alla fine del periodo considerato, ossia la popolazione longitudinale.

Questa nota riporta gli aspetti più importanti relativi alla costruzione dei file di microdati longitudinali a 12 mesi e delle matrici di transizione.

## 2. Il disegno campionario e la componente longitudinale

A partire dal primo trimestre 2004, la Rilevazione sulle Forze di lavoro (d'ora in poi RFL) ha subito una profonda trasformazione, ma il disegno campionario è rimasto sostanzialmente invariato. Si tratta, difatti, di campionamento a due stadi con stratificazione delle unità di primo stadio e rotazione delle unità di secondo stadio; le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie.

Il disegno campionario prevede la sostituzione di una parte delle famiglie campione nelle varie occasioni di indagine. I campioni relativi a trimestri differenti, quindi, risultano parzialmente sovrapposti in base a uno schema di rotazione di tipo 2-2-2 (Figura 1), secondo cui una famiglia viene inclusa nel campione per due rilevazioni successive e, dopo una pausa di due trimestri, viene reinserita nel campione per altre due rilevazioni. A differenza delle famiglie, che ruotano secondo lo schema sotto riportato, i comuni campione rimangono sempre gli stessi nel tempo<sup>2</sup>.

Figura 1. Schema di rotazione delle famiglie campione

PERIODO DI RIFERIMENTO		GRUPPO DI ROTAZIONE									
ANNO	TRIMESTRE	I2	L1								
2006	1		L2	M1							
	2			M2	N1						
	3	I3			N2	O1					
	4	I4	L3			O2	P1				
2007	1		L4	M3			P2	Q1			
	2			M4	N3			Q2	R1		
	3				N4	O3			R2	S1	
	4					O4	P3			S2	T1
2008	1						P4	Q3			T2
	2							Q4	R3		
	3								R4	S3	
	4										

La Figura 1 riporta la ripartizione dei gruppi di rotazione rispetto alle occasioni di indagine. Il campione di famiglie intervistate nel primo trimestre 2007 è composto da un gruppo di famiglie che entrano per la prima volta nel campione (P1), un gruppo di famiglie intervistate per la seconda volta (O2), un gruppo di famiglie intervistate per la terza volta (L3) ed un gruppo di famiglie intervistate per la quarta ed ultima volta (I4).

In particolare, il campione teorico di famiglie relative a ciascuna occasione di indagine (rilevazione trimestrale) era costituito fino al 2011 da circa 76mila famiglie<sup>3</sup>, suddivise equamente nei quattro gruppi di rotazione costituiti, quindi, ciascuno da circa 19mila famiglie; a partire dal 2012 tale numerosità è stata ridotta a circa 71mila famiglie.

La struttura longitudinale così congegnata comporta una sovrapposizione del campione teorico del 50% a tre ed a dodici mesi di distanza, del 25% a nove e quindici mesi. Tale sovrapposizione consente di costruire file di microdati longitudinali a 3, a 12 e a 15 mesi di distanza. Ad esempio, il file longitudinale a 12 mesi di distanza, che ha come riferimento il primo trimestre 2007 e il primo trimestre 2008 contiene le informazioni delle famiglie presenti contemporaneamente nei gruppi P1 – P3, e di quelle nei gruppi O2 – O4. Ne consegue che, inevitabilmente, il livello di precisione delle stime longitudinali risulta ridotto rispetto alla rilevazione trimestrale, e quindi eventuali risultati ottenuti per sottogruppi di popolazione e/o domini territoriali molto piccoli potrebbero avere un elevato grado di incertezza.

### **3. L'abbinamento dei record individuali**

L'abbinamento delle informazioni relative alle famiglie intervistate e dei suoi componenti non pone problemi particolari nella RFL, diversamente da quanto accadeva nella vecchia Rilevazione trimestrale (RTFL) effettuata fino al 2003, basata esclusivamente su questionari cartacei. La RFL prevede l'uso di codici identificativi familiari e individuali univoci per tutto il periodo in cui la famiglia e i suoi componenti partecipano all'indagine (15 mesi). L'uso della tecnica di indagine CAPI e CATI, inoltre, garantisce che, nella quasi totalità dei casi, questi codici vengono registrati in maniera corretta, permettendo un abbinamento deterministico.

In realtà, mentre il codice identificativo della famiglia risulta sempre corretto, un limitato numero di errori può essere presente sui codici identificativi dei componenti della famiglia. Tra i problemi che possono generare questi errori si possono menzionare i seguenti:

- errori nel funzionamento dei software di acquisizione dei dati;
- errori compiuti dai rilevatori durante la somministrazione del questionario elettronico;
- errori di compilazione e registrazione del questionario cartaceo quando quello elettronico non può essere usato.

Per risolvere completamente il problema dei pochi casi di errati abbinamenti, è stata implementata una procedura di verifica deterministica che, usando variabili di controllo quali sesso, data di nascita e nome, permette l'individuazione sia dei falsi positivi (componenti diversi che hanno lo stesso codice), sia dei falsi negativi (lo stesso componente che in due diverse occasioni ha due diversi codici individuali). Il numero degli

errati abbinamenti individuati (corretti con procedure ad hoc) è risultato sempre molto basso e tale da far ritenere che il numero di ulteriori errati abbinamenti rimasti nei file sia prossimo a zero.

## **4. La tecnica di Indagine**

A partire dal primo trimestre 2004, la RFL si avvale di una tecnica di indagine mista CAPI-CATI. In linea generale, la prima intervista (prima wave) di ogni famiglia presente nel campione è condotta mediante tecnica Capi (intervista faccia a faccia assistita da computer) mentre le tre successive sono da effettuarsi mediante tecnica Cati (intervista telefonica assistita da computer). Alcune eccezioni a queste regole riguardano: le famiglie con capofamiglia straniero, che vengono sempre intervistate con tecnica CAPI<sup>4</sup>; le famiglie che non dispongono o non vogliono fornire un numero di telefono che devono essere re-intervistate con tecnica CAPI; le interviste in prima wave, per le famiglie che dispongono di telefono, nei periodi delle vacanze di Ferragosto e Natale - Capodanno, vengono effettuate con tecnica Cati.

L'uso di una tecnica di indagine assistita da computer, permette di ottenere un dato "grezzo" trasversale qualitativamente elevato rispetto alle indagini eseguite con tecniche non assistite da computer. Grazie all'implementazione diretta sul questionario elettronico di un piano di controllo che agisce al momento della somministrazione dell'intervista (controllo dei domini delle variabili, dei percorsi del questionario e delle incongruenze logiche) il numero di errori dal punto di vista trasversale risulta estremamente ridotto.

Le tecniche CAPI e CATI consentono di usare, inoltre, per le re-interviste, e solo per alcuni specifici quesiti, la somministrazione dell'intervista di domande con modalità a conferma. Questo metodo consente di ridurre i tempi di somministrazione, la molestia statistica e garantisce una maggiore coerenza delle informazioni rilevate nei diversi trimestri sullo stesso individuo campione.

## **5. Il trattamento dei dati**

Nonostante le regole di controllo implementate nei questionari elettronici, un certo numero di incoerenze longitudinali, relativo alle informazioni rilevate in diversi trimestri, risulta comunque presente nei dati "grezzi". Tali incoerenze possono derivare da una molteplicità di cause, tra le quali per esempio: gli errori di risposta, che possono avvenire per incomprensione della domanda, oppure per la errata collocazione nel tempo di un evento, oppure se l'intervistato mente consapevolmente, o ancora, se il rispondente proxy

non è a conoscenza della reale situazione del familiare per conto del quale fornisce la risposta; errori di registrazione, che possono avvenire per incomprensione della risposta o per errore dell'intervistatore.

Nei file longitudinali rilasciati dall'Istituto, le incoerenze longitudinali vengono comunque risolte mediante un piano di controllo e correzione longitudinale, imputando le informazioni che risultano incoerenti o mancanti. L'aver rilevato una molteplicità di informazioni su più istanti di tempo, consentirà di ricostruire le storie lavorative degli individui campione in maniera più precisa e dettagliata.

L'imputazione di un record longitudinale, in cui figurano i valori delle variabili rilevate in differenti occasioni d'indagine su una stessa unità, può essere effettuata secondo due diverse strategie: imputare tutte le variabili che lo richiedono a prescindere dall'occasione a cui si riferiscono, oppure limitare l'imputazione soltanto alle variabili di una data occasione, mantenendo inalterate quelle delle occasioni precedenti.

Seguendo la prima strategia, i dati rilevati sarebbero considerati sempre provvisori, poiché con l'aggiunta delle nuove informazioni è possibile che alcune di quelle precedentemente acquisite vengano corrette. In tal caso i dati longitudinali diventerebbero definitivi solo dopo l'ultima occasione d'indagine, quando tutte le occasioni vengono corrette contemporaneamente.

Seguendo la seconda strategia, man mano che si acquisiscono nuove informazioni, la correzione si effettua tenendo fissa la storia passata. Sicuramente si pongono ulteriori restrizioni alla metodologia di imputazione (vincoli di "fissità" sulle variabili non imputabili), ma si riescono a produrre tempestivamente, per ogni occasione d'indagine, risultati definitivi. Tale strategia è stata usata nella procedura di correzione dei microdati longitudinali della RFL, a partire dal primo trimestre 2004, nell'intento di poter procedere ad una produzione a regime con cadenza trimestrale, di file di microdati abbinati a 12 mesi e delle relative matrici di transizione.

La correzione dei dati longitudinali della RFL ha lo scopo principale di garantire la coerenza longitudinale delle informazioni e di conservarne, in secondo luogo, quella trasversale. Essa viene effettuata, per la maggior parte delle variabili, con un metodo deterministico; un metodo misto deterministico-probabilistico viene usato, invece, per le variabili relative alle date di inizio o fine occupazione.

Le regole generali di imputazione, da usare per la correzione a regime dei dati longitudinali, sono state definite studiando l'intera storia lavorativa, dalla prima occasione di indagine alla quarta occasione (o wave), per ciascun periodo considerato, dal primo trimestre 2004 al primo trimestre 2008. Le regole generali di imputazione sono state

definite in modo da minimizzare il rischio di errata imputazione nella situazione “a regime” quando, cioè, la correzione deve essere effettuata trimestralmente, e quindi l'intera storia lavorativa non è ancora nota.

A questo scopo la procedura di imputazione longitudinale, usata quando si conoscono solo le prime due occasioni (wave) di indagine, è diversa da quella usata quando viene acquisito il dato della terza e della quarta occasione.

Nel caso di incoerenze riscontrate sulle prime due wave, la procedura corregge le incoerenze imputando il valore della variabile, nella prima o nella seconda wave, nel rispetto delle seguenti regole gerarchiche:

- quando l'intervista nella seconda wave è stata eseguita con modalità a conferma, la correzione viene effettuata sulla variabile della prima wave se il rispondente, nella seconda wave, non conferma il dato rilevato nella prima wave;
- quando una delle due intervista è stata rilasciata dal diretto interessato, mentre l'altra è stata rilasciata da un rispondente proxy, la variabile che è mancante o incoerente viene imputata sempre sulla base della risposta fornita dal diretto interessato, nella prima o nella seconda wave;
- nei casi diversi da quelli precedenti la risposta data nella prima wave viene imputata usando la risposta fornita nella seconda wave.

Quando su di un certo individuo campione viene acquisito il dato relativo alla sua terza o quarta occasione di indagine, occorre tener conto che i dati precedentemente rilevati sono già stati corretti (ed eventualmente diffusi), per cui non è possibile procedere ad una nuova correzione contemporanea su tutte le occasioni. In questo caso la correzione (imputazione dei dati mancanti o incoerenti) riguarda solo i nuovi dati (la nuova wave) e viene effettuata condizionatamente alla storia lavorativa passata (le wave già corrette precedentemente).

È evidente che questa strategia inserisce nuovi vincoli al metodo di correzione, ma i risultati delle analisi finora condotte evidenziano che le differenze nei risultati, ottenuti con il metodo prescelto e quelli ottenuti con l'uso di un metodo di correzione contemporanea, sono da ritenersi trascurabili. D'altra parte, il metodo scelto assicura che le stime prodotte non debbano essere revisionate ogni volta che una nuova wave è disponibile, ossia ogni trimestre.



## 6. La popolazione di riferimento

Come in ogni altro tipo di indagine, anche per la componente longitudinale della RFL è fondamentale definire la popolazione di riferimento, cioè la popolazione che può essere correttamente rappresentata dal campione longitudinale degli individui abbinati.

Uno dei punti fondamentali da tenere in considerazione è che la popolazione si modifica nell'arco di un determinato periodo a causa di entrate (nascite e iscrizioni anagrafiche dovute a cambi di residenza o immigrazione) e uscite (morti e cancellazioni anagrafiche dovute a cambi di residenza o emigrazione).

In un determinato arco temporale, il campione longitudinale della RFL, che scaturisce dall'abbinamento di due trimestri, non può rappresentare tutta la popolazione, ma è in grado di rappresentare correttamente solo quella residente in uno stesso comune sia all'inizio che alla fine del periodo considerato<sup>5</sup>.

Tale popolazione, che per comodità espositiva chiameremo popolazione longitudinale, è definita come la popolazione residente a inizio periodo (esclusi gli individui che fanno parte di convivenze), al netto delle morti e dei cambi di residenza verso altri comuni e/o verso l'estero verificatisi nel periodo. Grazie alla disponibilità all'interno dell'Istituto, di dati demografici - affidabili e tempestivi circa la popolazione residente, le iscrizioni e le cancellazioni - prodotti sulla base delle comunicazioni provenienti dai registri anagrafici, la popolazione longitudinale è stata calcolata a livello di singola provincia con disaggregazione per sesso e classe di età, e a livello di regione con disaggregazione per sesso e cittadinanza (italiani/stranieri). Il metodo di costruzione della popolazione longitudinale garantisce la coerenza tra le popolazioni di riferimento delle stime trimestrali e la popolazione di riferimento delle stime longitudinali.

E' chiaro che la scelta della popolazione di riferimento condiziona il tipo di matrice di transizione che si può costruire. Alla diffusa esigenza e richiesta di dati di flusso riguardanti l'intera popolazione si contrappone, però, l'esigenza di conservare e garantire un elevato rigore metodologico. La componente longitudinale, che è un sottoprodotto della RFL, non può essere considerata come un vero e proprio panel; non fornisce, infatti, informazioni sulla condizione occupazionale, a inizio e fine periodo, di tutta la popolazione di partenza, ma solo relativamente ad una sua parte, seppur considerevole. Questo limite è dovuto al fatto che il disegno campionario della RFL non prevede di seguire sul territorio, per le interviste successive, né gli individui che escono dalla famiglia campione, né le famiglie intere che si trasferiscono verso altri comuni o verso l'estero.

## 7. La metodologia di calcolo dei pesi campionari

Lo stimatore scelto per il riporto all'universo dei dati longitudinali della RFL, come per i dati trasversali, appartiene alla classe degli stimatori di ponderazione vincolata<sup>6</sup>. Si tratta di uno stimatore in cui i pesi finali consentono di ottenere, nell'ambito di diversi domini territoriali (ripartizioni, regioni, province), stime di popolazione (per sesso, classi di età e/o cittadinanza) uguali ai corrispondenti totali noti di fonte anagrafica. I pesi sono calcolati a livello familiare e ciò implica che a ciascun individuo appartenente alla stessa famiglia sia attribuito un medesimo peso di riporto all'universo; l'unicità del peso a livello familiare assicura la coerenza delle stime familiari con le stime individuali.

Il peso di riporto all'universo per un campione longitudinale viene determinato mediante diverse fasi, distinte e consecutive, che permettono da un lato il trattamento delle mancate risposte totali, dall'altro la ponderazione delle stime longitudinali sia su totali noti di popolazione di fonte anagrafica, sia sulle stime trasversali della stessa RFL riferite ai trimestri iniziale e finale.

La procedura di riporto all'universo dei file di dati longitudinali a 12 mesi di distanza, prevede le seguenti fasi:

- nel trimestre iniziale (rispetto al periodo di riferimento dei dati longitudinali) tutti gli individui che dovrebbero essere re-intervistati dopo 12 mesi (sottocampione degli abbinabili) vengono selezionati e il loro peso finale trasversale viene utilizzato come peso base longitudinale.
- il peso base longitudinale viene corretto mediante l'uso di una procedura di ponderazione vincolata, in modo che il sottocampione degli abbinabili riproduca esattamente un certo numero di stime trasversali ottenute con il campione trasversale completo (popolazione per sesso, classe di età, territorio e cittadinanza, classificata per condizione professionale e altre sue caratteristiche rilevanti). Al termine di questa fase si ottiene il peso intermedio longitudinale per tutti gli individui abbinabili.
- nella terza fase, vengono selezionati solo gli individui effettivamente abbinati (che formano quindi il dataset longitudinale) e, partendo dai loro pesi intermedi longitudinali, si usa ancora la procedura di ponderazione vincolata per ottenere i pesi finali longitudinali. In questa fase i vincoli della procedura sono costituiti dalla popolazione longitudinale, relativa al periodo in questione, con una suddivisione per cittadinanza, sesso e classi di età, diversa per i diversi domini territoriali. I vincoli suddetti possono essere così riassunti:

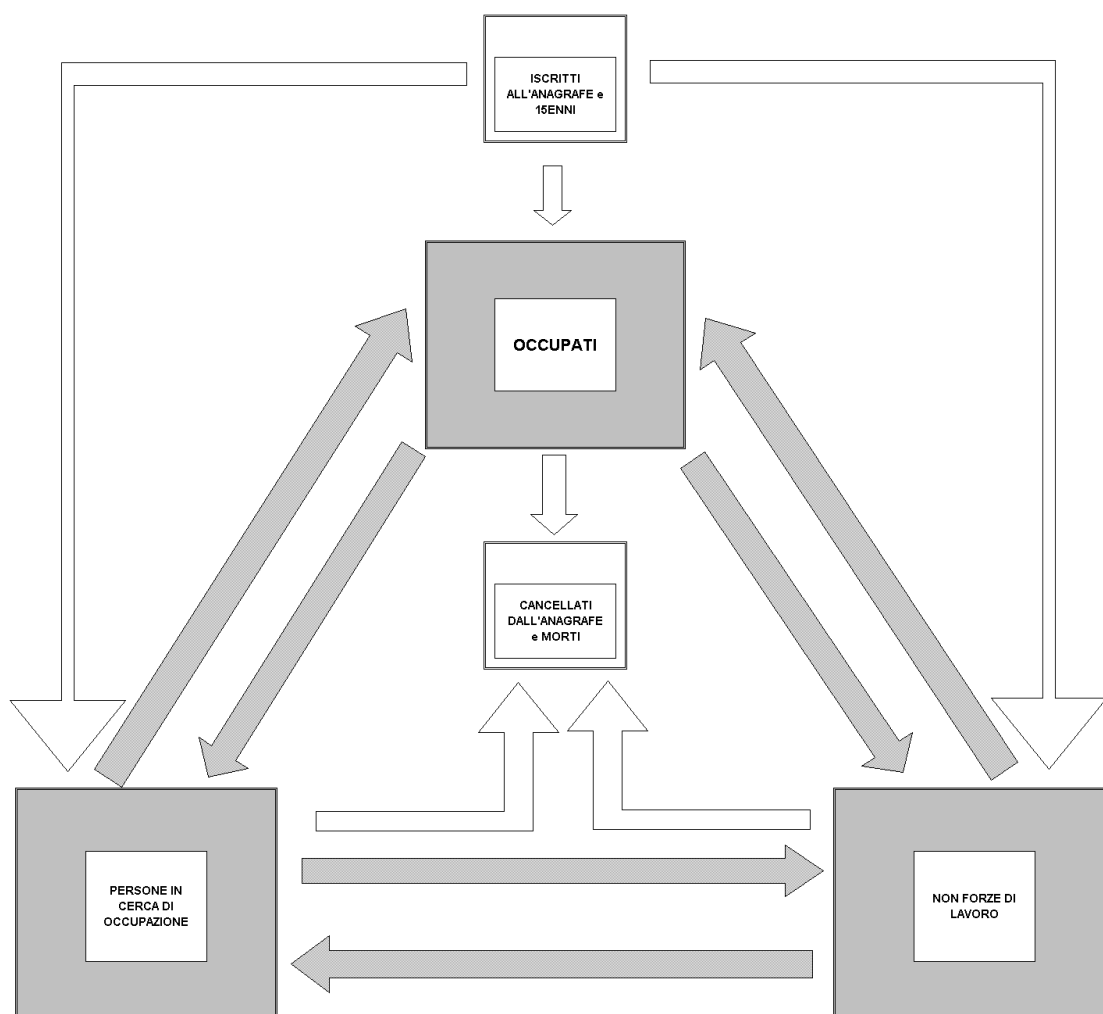
- popolazione longitudinale per ripartizione geografica (Nord-Ovest, Nord-Est, Centro e Mezzogiorno), sesso e quattordici classi d'età (0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e più);
- popolazione longitudinale per regione, sesso e 5 classi d'età (0-14, 15-24, 25-44, 45-64, 65-e più);
- popolazione longitudinale per regione, sesso e cittadinanza (italiani e stranieri),
- popolazione longitudinale per provincia, sesso e 3 classi d'età (0-14, 15-64, 65-e più).

## 8. Le matrici di transizione

Come già affermato, la costruzione degli archivi longitudinali della RFL e la conseguente possibilità di fornire matrici di transizione per la stima dei flussi tra condizioni nel mercato del lavoro è subordinata, e soprattutto limitata, dalla particolare natura del disegno dell'indagine che ha come obiettivo fondamentale quello di fornire stime trimestrali trasversali dei principali indicatori strutturali del mercato del lavoro in perfetta aderenza alle definizioni internazionali di occupato e di persona in cerca di occupazione (Regolamenti Comunitari n° 1575/2000 e n° 1897/2000).

La sottostante Figura 2 rappresenta i flussi della popolazione complessiva per un dato intervallo di tempo (un trimestre o un anno), che possono essere rappresentati dai campioni trasversali e longitudinali della RFL.

Figura 2. Diagramma dei flussi della popolazione complessiva

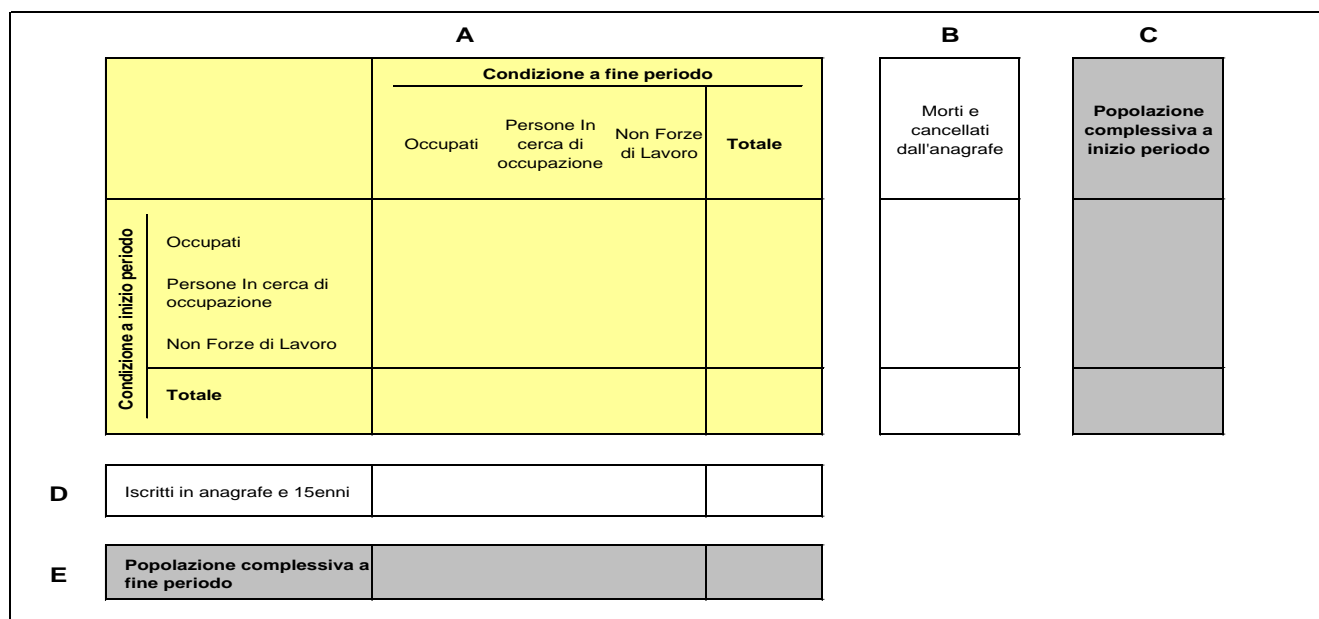


I campioni trasversali della RFL forniscono una stima della distribuzione per condizione professionale, sia della popolazione iniziale sia di quella finale. Come detto, però, parte della popolazione iniziale può cambiare residenza, emigrare o morire. Ne consegue che di essa se ne conosce solo la condizione a inizio periodo. Allo stesso modo, di quella parte della popolazione che si è iscritta in anagrafe o ha compiuto 15 anni nel periodo sotto osservazione, sempre non rappresentabile dalla componente longitudinale, è nota solo la condizione a fine periodo.

La Figura 3 mostra la matrice completa degli stock e dei flussi della popolazione complessiva; fornisce quindi, con diverso grado di dettaglio, una serie di informazioni riferite agli aggregati di popolazione: iniziale, finale, longitudinale, uscita e entrata.

La componente longitudinale, quindi, descrive solo i flussi tra le diverse condizioni (le frecce colorate nel diagramma della Figura 2) intervenuti per la popolazione longitudinale, sintetizzati nella matrice di transizione (indicata con A nella Figura 3).

Figura 3. Schema della matrice completa degli stock e dei flussi della popolazione complessiva



Considerando solo la popolazione in età lavorativa in entrambe le occasioni (con almeno 15 anni a inizio periodo), la matrice completa contiene:

- la matrice di transizione (indicata con la lettera A) con la distribuzione congiunta secondo la condizione a inizio e fine periodo, desunta dalla componente longitudinale della RFL, riferita esclusivamente alla popolazione longitudinale;
- due vettori con la distribuzione per condizione della popolazione complessiva, sia a inizio che a fine periodo, risultante dai relativi campioni trasversali (indicati con C e E);

- due vettori di raccordo tra stime trasversali e stime longitudinali (ottenuti per differenza) che riportano la condizione a inizio periodo per coloro che risultano morti o cancellati dalle anagrafi nel periodo (indicato con B) e la condizione a fine periodo per coloro che compiono 15 anni e si iscrivono in anagrafe nello stesso periodo (indicato con D).

Tenendo conto dello schema sopra riportato, le stime contenute nelle matrici di transizione, prodotte a partire dai dati longitudinali mediante l'uso del peso finale longitudinale descritto in precedenza, risultano coerenti con le stime trasversali ufficiali e relative alla popolazione complessiva della RFL<sup>7</sup>.

Le matrici di transizione consentono di comprendere con un maggiore dettaglio l'evoluzione degli stock di occupati, disoccupati e inattivi, arricchendo il panorama delle informazioni statistiche sulle dinamiche del mercato del lavoro. Difatti, esse forniscono una stima del numero di permanenze e di transizioni in entrata e in uscita dalle diverse condizioni occupazionali. Consentono, inoltre, di stimare le probabilità di permanenza e/o di passaggio da una condizione di origine a una di arrivo, e di analizzare le caratteristiche degli individui che ne sono coinvolti.

I dati longitudinali e le matrici di transizione qui descritte fanno parte di una nuova serie di microdati che inizia dal primo trimestre 2004, momento in cui la Rilevazione Trimestrale sulle Forze di Lavoro è stata ristrutturata ed è divenuta continua (cfr Istat - 2006). Così come avviene per le stime trasversali, anche le stime di flusso prodotte con questi file, non sono confrontabili con quelle prodotte con i file di microdati longitudinali relativi al periodo 1993-2003.

## 9. La diffusione dei dati longitudinali

La diffusione dei dati longitudinali dell'indagine RFL, così come già avviene per i dati trasversali, terrà conto delle differenziate esigenze degli utenti e si articolerà su diversi piani:

- rilascio dei microdati longitudinali abbinati a 12 mesi, con cadenza trimestrale;
- diffusione dei principali risultati e delle relative matrici di transizione, con cadenza trimestrale, per l'approfondimento della dinamica tendenziale del mercato del lavoro già messa in luce dai dati trasversali;
- elaborazioni ad hoc su richiesta degli utenti;
- diffusione dei risultati sulla base di media d'anno (media dei 4 abbinamenti a 12 mesi) con approfondimenti sia a livello territoriale (regione e ripartizione) sia sulla popolazione straniera.

La diffusione dei dati longitudinali (della nuova serie che parte dal primo trimestre 2004) inizia con il rilascio di file di microdati con abbinamenti a 12 mesi di distanza riferiti ai seguenti periodi:

- Primo trimestre 2004 – Primo trimestre 2005
- Primo trimestre 2005 – Primo trimestre 2006
- Primo trimestre 2006 – Primo trimestre 2007
- Primo trimestre 2007 – Primo trimestre 2008
- Primo trimestre 2008 – Primo trimestre 2009
- Primo trimestre 2009 – Primo trimestre 2010
- Primo trimestre 2010 – Primo trimestre 2011
- Primo trimestre 2011 – Primo trimestre 2012

In questa prima fase sono state previste due versioni di file di microdati: SISTAN e MFR. Attualmente, alcuni quesiti sono ancora in corso di validazione in un'ottica longitudinale, e verranno resi disponibili agli utenti in un successivo rilascio. Si tratta, in particolare, di quesiti relativi allo stato civile, e alla tipologia familiare; aspetti dell'eventuale occupazione precedente (per i non occupati); informazioni inerenti la formazione professionale. Sono inoltre, in corso di validazione, le informazioni circa la durata dell'occupazione, della disoccupazione e della non occupazione. In merito a queste ultime occorre precisare che nel file è già presente l'informazione relativa alla durata della ricerca di occupazione, ma essa non può essere usata per la stima della durata della

disoccupazione in quanto quest'ultima viene attribuita sulla base del valore tra i mesi di ricerca di lavoro e mesi trascorsi dalla fine dell'occupazione (per coloro che ne avevano una).

A partire dal file relativo all'abbinamento I° trimestre 2010 - I° trimestre 2011 si tiene conto dell'introduzione delle nuove classificazioni del settore di attività economica ATECO2007, entrata a regime dopo un periodo di sovrapposizione di tre anni con la precedente ATECO2002, e delle professioni, con l'adozione della CP2011.

#### ▪ **File di microdati longitudinali in versione SISTAN**

Destinato a utenti appartenenti al SISTAN, viene usato anche per lo scambio dei microdati tramite il Protocollo Interno ISTAT. Sebbene alcuni quesiti siano riservati, questo tipo di file contiene informazioni particolarmente dettagliate che permettono analisi su molteplici aspetti. Per facilitare le elaborazioni da parte degli utenti, il file contiene un certo numero di variabili ricostruite riguardanti la condizione professionale e altre sue importanti caratteristiche. In particolare il file di microdati SISTAN:

- contiene informazioni sull'età per singolo anno;
- contiene i codici delle regioni e delle province
- contiene il codice dello stato estero di nascita e di cittadinanza.
- contiene la classificazione a 4 digit delle attività economiche<sup>8</sup>;
- contiene la classificazione a 4 digit delle professioni;
- permette elaborazioni per famiglia mediante l'uso di un codice familiare.

Il file di microdati in versione SISTAN può essere usato anche dal mondo della ricerca scientifica nell'ambito dei servizi offerti dal Laboratorio per l'Analisi dei Dati ELEMENTARI (Laboratorio ADELE). In tale laboratorio, presente nei locali dell'Istituto, i ricercatori possono effettuare le proprie elaborazioni dei microdati dell'indagine. In questo caso il controllo del rischio di violazione della riservatezza viene effettuato prima del rilascio degli output delle elaborazioni.

#### ▪ **File di microdati in versione MFR (Microdata for research).**

Si tratta di un file di microdati con contenuto informativo ridotto rispetto al file SISTAN, ottenuto dopo il trattamento per la riduzione del rischio di violazione della riservatezza. Anche in questo file, per facilitare le elaborazioni da parte degli utenti, sono state inserite un certo numero di variabili ricostruite riguardanti la condizione professionale e altre sue importanti caratteristiche. In particolare il file di microdati MFR:



- contiene l'età dettagliata al singolo anno per le età comprese tra 15 e 74 anni, e l'età in classi per le altre;
- contiene i codici delle regioni e delle province.
- contiene il codice dello stato estero di nascita e di cittadinanza.
- contiene la classificazione a 4 digit delle attività economiche;
- contiene la classificazione a 4 digit delle professioni (con un'aggregazione a 3 digit nel primo grande gruppo);
- permette elaborazioni per famiglia mediante l'uso di un codice familiare.

## 10. Glossario

**Forze di lavoro:** comprendono gli occupati e le persone in cerca di occupazione.

**Occupati:** comprendono le persone di 15 anni e più che nella settimana di riferimento:

- hanno svolto almeno un'ora di lavoro in una qualsiasi attività che preveda un corrispettivo monetario o in natura;
- hanno svolto almeno un'ora di lavoro non retribuito nella ditta di un familiare nella quale collaborano abitualmente;
- sono assenti dal lavoro (ad esempio, per ferie o malattia). I dipendenti assenti dal lavoro sono considerati occupati se l'assenza non supera tre mesi, oppure se durante l'assenza continuano a percepire almeno il 50% della retribuzione. Gli indipendenti assenti dal lavoro, ad eccezione dei coadiuvanti familiari, sono considerati occupati se, durante il periodo di assenza, mantengono l'attività. I coadiuvanti familiari sono considerati occupati se l'assenza non supera tre mesi.

**Persone in cerca di occupazione:** comprendono le persone non occupate tra 15 e 74 anni che:

- hanno effettuato almeno un'azione attiva di ricerca di lavoro nei trenta giorni che precedono l'intervista e sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive all'intervista;
- oppure, inizieranno un lavoro entro tre mesi dalla data dell'intervista e sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive all'intervista, qualora fosse possibile anticipare l'inizio del lavoro.

**Inattivi:** comprendono le persone che non fanno parte delle forze di lavoro, ovvero quelle non classificate come occupate o in cerca di occupazione. Gli inattivi vengono classificati in “Inattivi minori di 15 anni”, “Inattivi di 65 anni o più” Inattivi tra 15 e 64 anni”. Questi ultimi vengono ulteriormente distinti, secondo il grado di attachment al mercato del lavoro, nelle seguenti quattro categorie:

- Inattivi che “cercano lavoro non attivamente ma disponibili a lavorare” o “forze lavoro potenziali”;
- Inattivi che “cercano lavoro ma non disponibili a lavorare”;
- Inattivi che “non cercano ma disponibili a lavorare”;
- Inattivi che “non cercano e non disponibili a lavorare”.

**Tasso di attività:** rapporto tra le persone appartenenti alle forze di lavoro e la corrispondente popolazione di riferimento. La somma del tasso di attività e del tasso di inattività è pari al 100 per cento.

**Tasso di occupazione:** rapporto tra gli occupati e la corrispondente popolazione di riferimento.

**Tasso di disoccupazione:** rapporto tra le persone in cerca di occupazione e le corrispondenti forze di lavoro.

**Tasso di disoccupazione di lunga durata:** rapporto tra le persone in cerca di occupazione da dodici mesi e oltre e le forze di lavoro. Occorre precisare che la durata della disoccupazione non sempre coincide con la durata della ricerca di un’occupazione; difatti quest’ultima risulta uguale al minimo tra i mesi di ricerca di lavoro e i mesi trascorsi dalla fine dell’ultima occupazione (per coloro che hanno avuto precedenti esperienze lavorative).

**Tasso di inattività:** rapporto tra le persone non appartenenti alle forze di lavoro e la corrispondente popolazione di riferimento. La somma del tasso di inattività e del tasso di attività è pari al 100 per cento.

**Variazione congiunturale:** variazione rispetto al trimestre precedente.

**Variazione tendenziale:** variazione rispetto allo stesso periodo dell’anno precedente.

**Settimana di riferimento:** settimana a cui fanno riferimento le informazioni raccolte.

## 11. Riferimenti bibliografici

- “La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione.”  
a cura di S. Gazzelloni  
ISTAT - Metodi e Norme ,n. 32 - 2006 - Roma
- “Gli stranieri nella rilevazione sulle forze di lavoro.”  
a cura di M. Albisinni e L. Quattrociochi  
ISTAT - Metodi e norme, n. 27 - 2006 – Roma
- “Le matrici di transizione della Rilevazione trimestrale sulle forze di lavoro. Nota metodologica”  
a cura di C. Ceccarelli, A. R. Discenza, S. Rosati , A. Paggiaro e N. Torelli.  
ISTAT - Roma, 12 dicembre 2002.
- “La mobilità nel mercato del lavoro: principali risultati aprile 1998-aprile 2002” (con Appendice statistica )  
a cura di M. Albisinni e A. R. Discenza.  
ISTAT - Approfondimenti - Settore Lavoro. Roma, 12 dicembre 2002.
- “Weighting Procedure for Longitudinal Data of the Italian Labour Force Survey”.  
di A. R. Discenza  
In Atti della XLII Riunione Scientifica della SIS – Sessioni spontanee  
Università di Bari, 9-11 giugno 2004 - pp. 29-32 – CLEUP – 2004
- Deville J. C., Särndal C. E., (1992), Calibration Estimator, in Survey Sampling, Journal of the American Statistical Association, vol. 87, pp.376-382.

## 12. Contatti

Servizio Formazione Lavoro – e.mail: [infolav@istat.it](mailto:infolav@istat.it)

Antonio R. Discenza – FOL/B – e.mail: [discenza@istat.it](mailto:discenza@istat.it)

Barbara Boschetto – FOL/B – e.mail: [boschett@istat.it](mailto:boschett@istat.it)

Antonella Iorio – FOL/B – e.mail: [iorio@istat.it](mailto:iorio@istat.it)

Alessandro Martini – FOL/B – e.mail: [alemartini@istat.it](mailto:alemartini@istat.it)

- 
- 1 Per approfondimenti circa la rilevazione trimestrale che ne costituisce il presupposto, si rimanda ai volumi: “La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione.”, a cura di S. Gazzelloni, ISTAT - Metodi e Norme ,n. 32 - 2006 – Roma e “Gli stranieri nella rilevazione sulle forze di lavoro.”, a cura di M. Albisinni e L. Quattrociochi, ISTAT - Metodi e norme, n. 27 - 2006 – Roma
  - 2 In realtà alcuni comuni di ridotta dimensione in termini demografici vengono periodicamente sostituiti con altri con le stesse caratteristiche a causa dell'esaurimento delle liste anagrafiche
  - 3 Si tratta di un campione teorico di circa 300mila famiglie in un anno.
  - 4 Per approfondimenti vedi il volume: Metodi e norme ISTAT 2006 “Gli stranieri nella rilevazione sulle forze di lavoro.”
  - 5 La componente longitudinale non può correttamente rappresentare tutta la popolazione che risiede sul territorio nazionale in quanto il comportamento degli individui che cambiano residenza differisce fortemente da quello degli individui che non la cambiano. L'estensione dei risultati relativi alla popolazione longitudinale alla popolazione complessiva comporterebbe l'introduzione di una distorsione nelle stime di flusso.
  - 6 Si tratta di uno stimatore usato in ISTAT per tutte le indagini sulle famiglie. Nella letteratura in lingua inglese tale stimatore è noto con il nome di calibration estimator; per ulteriori informazioni si veda Deville & Särndal (1992).
  - 7 Le matrici di transizione prodotte fanno riferimento soltanto alla popolazione longitudinale (che è una parte di quella iniziale e finale), quindi, tutte le analisi possono essere condotte solo su matrici “al netto” dei flussi realizzatisi per la popolazione iscritta e cancellata nel periodo. D'altro canto questo tipo di matrici consentono di raggiungere un elevato dettaglio informativo, con una minima distorsione, soltanto sulla popolazione longitudinale.
  - 8 Per gli abbinamenti I° trimestre 2010- I° trimestre 2011 e I° trimestre 2011- I° trimestre 2012 al fine di garantire la coerenza longitudinale dei dati alla luce dell'introduzione della nuova classificazione ATECO2007 si è reso necessario ridurre il dettaglio a due digit per le informazioni del I° trimestre 2010 e per alcuni casi del I° trimestre 2011.