

ISTAT

# ***Rilevazione Forze di Lavoro***

**I Trimestre**

**2016**

Codice SN163



## **UniData**

Biccocca Data Archive

[www.unidata.unimib.it](http://www.unidata.unimib.it)

E-mail: [unidata@unimib.it](mailto:unidata@unimib.it)

Tel.: 02 64487513

Fax: 02 64487561

La presente documentazione è distribuita da UniData.

La fonte che ha prodotto i dati e UniData che li ha distribuiti non rispondono per alcun utilizzo improprio dei dati e delle elaborazioni pubblicate.



Università degli Studi di Milano-Bicocca

**INDICE**  
***TABLE OF CONTENTS***

Note metodologiche  
*Methodological Notes*

1. Aspetti metodologici dell'indagine  
*Survey's methodological issues* ..... p. 3
  
2. Descrizione del file  
*File description* ..... p. 16

**ASPETTI METODOLOGICI  
DELL'INDAGINE**

*SURVEY'S METHODOLOGICAL  
ISSUES*



## <Rilevazione sulle Forze di Lavoro>

<dal primo trimestre 2004>

### **Aspetti metodologici dell'indagine**

<Edizione: luglio 2014>

## INDICE

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

## 1. Introduzione

L'indagine ha la finalità di ottenere informazioni sulla situazione lavorativa, sulla ricerca di lavoro e sugli atteggiamenti verso il mercato del lavoro della popolazione in età lavorativa. In questa nota sono riportati in sintesi gli aspetti salienti dell'indagine. Per una trattazione più ampia e dettagliata si faccia riferimento al volume: ISTAT, *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*, Metodi e Norme, Roma 2006

## 2. La popolazione di riferimento

La popolazione di interesse è costituita da tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente emigrati all'estero, mentre esclude i membri permanenti delle convivenze (ospizi, istituti religiosi, caserme, eccetera). La famiglia è intesa come famiglia di fatto, ossia come un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti e aventi dimora abituale nello stesso comune; nel caso in cui la famiglia selezionata coabiti con altre famiglie viene intervistata soltanto quella estratta.

I parametri di interesse dell'indagine sono espressi in termini di numero di individui che possiedono un dato attributo  $z$  e riguardano i seguenti aggregati di popolazione: forze di lavoro, occupati, persone in cerca di occupazione.

Relativamente a ciascuno di tali aggregati, i principali parametri oggetto di stima sono il numero di individui che presentano l'attributo  $z$  nel trimestre  $t$  e nell'anno  $a$  (ottenuto come media dei quattro trimestri dell'anno  $a$ ) e la differenza tra il numero di individui che presentano l'attributo  $z$  nel trimestre  $t$  e il numero di individui che presentano il medesimo attributo in un trimestre precedente  $t'$ .

I domini territoriali di studio sono l'intero territorio nazionale, le ripartizioni geografiche e le regioni; a partire dal 1993, vengono considerate anche le province unicamente per la costruzione di stime medie annue.

## 3. Il disegno campionario

Il disegno di campionamento adottato in ciascun trimestre è un disegno a due stadi con stratificazione delle unità di primo stadio; le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie. All'interno di ciascuna provincia i comuni sono suddivisi in

due sottoinsiemi: i comuni la cui dimensione demografica è superiore a una prefissata soglia sono detti comuni auto rappresentativi (Ar); i rimanenti comuni vengono denominati non auto rappresentativi (Nar). Ciascun comune Ar costituisce strato a sé stante e viene incluso con certezza nel campione; i comuni Nar, invece, vengono stratificati sulla base della dimensione demografica e da ogni strato così definito viene estratto un comune con probabilità proporzionale alla dimensione demografica. Dalla lista anagrafica di ogni comune campione viene selezionato, mediante scelta sistematica, un campione di famiglie; tutti gli individui appartenenti alle famiglie estratte vengono intervistati.

Con riferimento a ciascun trimestre la rilevazione è di tipo continuo e viene condotta tutte le settimane. Ogni famiglia campione viene intervistata una sola volta in una specifica settimana.

I campioni relativi a trimestri differenti sono parzialmente sovrapposti in base a uno schema di rotazione secondo cui una famiglia è inclusa nel campione per due rilevazioni successive e, dopo una pausa di due trimestri, viene reinserita nel campione per altre due rilevazioni. Ciò comporta una sovrapposizione del campione teorico del 50 per cento a un trimestre di distanza, del 25% a tre trimestri, del 50 per cento a quattro trimestri e del 25 per cento a cinque trimestri. Viene utilizzato, quindi, uno schema di rotazione trimestrale in cui le famiglie vengono intervistate per due trimestri consecutivi, escluse per due trimestri e successivamente re-intervistate per altri due trimestri<sup>1</sup>. Nei quattro trimestri di rilevazione in cui ciascuna famiglia campione viene intervistata, la settimana di rilevazione è sempre la medesima, ad esempio se una famiglia viene intervistata per la prima volta nella seconda settimana del trimestre, anche nei successivi trimestri in cui è coinvolta verrà intervistata nella seconda settimana.

La sovrapposizione dei campioni trimestrali risponde all'esigenza di ridurre le fluttuazioni delle stime di livello riferite a trimestri differenti. Inoltre, la specifica tecnica di rotazione permette di ridurre la stima delle variazioni nette tra trimestri consecutivi e tra trimestri a un anno di distanza.

Il disegno di campionamento adottato per ciascuna rilevazione trimestrale si basa su una stratificazione temporale di tipo mensile; infatti, il campione trimestrale di unità finali di campionamento (le famiglie) viene suddiviso in tre gruppi distinti ciascuno dei quali viene assegnato casualmente a un mese del trimestre in modo tale che ciascuno dei tre gruppi

---

<sup>1</sup> Tenendo conto che lo schema trimestrale di rotazione delle famiglie è del tipo (2-2-2) e del fatto che nei quattro trimestri di rilevazione in cui viene intervistata ciascuna famiglia campione la settimana di rilevazione è sempre la medesima (ad esempio la famiglia viene intervistata sempre nella seconda settimana di ogni trimestre), le distanze effettive tra la prima intervista e le successive sono rispettivamente di 1, 4 e 5 trimestri (intendendo per trimestre un periodo di 13 settimane consecutive).

costituisca un campione rappresentativo della popolazione di riferimento nel mese considerato.

A differenza delle famiglie, che ruotano secondo lo schema sopra introdotto, i comuni campione rimangono sempre gli stessi nel tempo.

Complessivamente, in ciascuna rilevazione trimestrale vengono coinvolti circa 1.400 comuni per un totale di circa 70 mila famiglie (circa 78 mila fino al 2011).

## **4. La rilevazione e il trattamento dei dati**

La prima intervista di ogni famiglia presente nel campione è di tipo diretto, condotta mediante tecnica Capi (intervista faccia a faccia assistita da computer) mentre le tre successive sono da effettuarsi, in linea teorica, mediante intervista telefonica con tecnica Cati (con l'eccezione delle famiglie straniere e di quelle che non dispongono di un telefono fisso).

L'attività di rilevazione si interrompe in alcune settimane dell'anno in cui la reperibilità delle famiglie è piuttosto scarsa. In particolare, l'interruzione delle interviste è prevista per le prime tre settimane di agosto e per il periodo tra Natale e Capodanno<sup>2</sup>. Le interviste da svolgersi durante l'interruzione vengono recuperate in un periodo successivo. In tale frangente, per evitare il sovraccarico della rete di rilevazione Capi, che altrimenti si troverebbe a gestire un numero troppo elevato di interviste con il rischio di un aumento del tasso di caduta, tutte le famiglie in prima intervista che dispongono di telefono vengono recuperate con tecnica Cati.

Le innovazioni introdotte in termini di tecnica di rilevazione e l'importante contenuto informatico a supporto comportano, una strategia di controllo e correzione dati diversa rispetto a quella usata per le indagini tradizionali. In particolare, il dato "grezzo" ha un livello qualitativo estremamente più elevato rispetto alle indagini eseguite con tecniche non assistite da computer. Il dato grezzo presenta, infatti, un numero estremamente ridotto di incongruenze tra le informazioni raccolte, in quanto queste ultime vengono sottoposte ad un piano di verifica implementato nel questionario elettronico che agisce sin dal momento della rilevazione. Tale piano si basa sul controllo dei domini delle variabili, dei percorsi del questionario e delle incongruenze logiche tra le informazioni raccolte.

La strategia adottata tratta essenzialmente due tipologie di errori: gli errori strutturali e quelli logico-formali; i primi derivano dai percorsi strutturali del questionario mentre i secondi dalle incongruenze logico formali delle informazioni raccolte. Nel questionario

---

<sup>2</sup> Nei periodi indicati i rilevatori Capi generalmente fruiscono delle ferie.



elettronico, ovviamente, sono implementate tutte le regole (denominate hard) che determinano i percorsi e il flusso del questionario stesso e gran parte di quelle logico-formali (denominate soft)<sup>3</sup>. In assenza di errori strutturali del questionario, quindi, le sole tipologie di correzione che vengono effettuate riguardano le incongruenze logico-formali derivanti da forzature di regole soft del questionario o di incongruenze non rilevate.

Infine i dati sottoposti ai trattamenti appena descritti passano al vaglio della “procedura famiglie”. Tale procedura permette, dopo aver controllato e corretto le notizie anagrafiche dei componenti, di costruire i nuclei presenti nella famiglia di fatto intervistata, e classificarla secondo un set di tipologie familiari.

## 5. La metodologia di calcolo dei pesi campionari

Lo stimatore scelto per la Rcfl è uno stimatore di ponderazione vincolata<sup>4</sup>, cioè uno stimatore in cui i pesi finali consentono di ottenere nell’ambito di diversi domini territoriali (regioni, province autonome di Trento e Bolzano, province, grandi comuni) stime di popolazione residente (per sesso e classi di età) 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 venga 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.

La realizzazione del coefficiente finale di riporto all’universo si articola in tre fasi distinte. La prima fase prevede il calcolo del coefficiente base ottenuto come reciproco della probabilità che ogni famiglia ha di far parte del campione. Per la determinazione del coefficiente base, si tiene conto della probabilità di estrarre un generico Comune dallo strato di appartenenza, e della probabilità di estrarre le famiglie campione dalla lista anagrafica di tale Comune. Ogni famiglia del campione, appartenente allo stesso Comune, ha pertanto lo stesso coefficiente base.

Successivamente, per tenere conto delle interviste non realizzate in ciascun Comune, si determina un elemento correttivo del coefficiente base detto coefficiente per mancata risposta totale. In genere, il numero di famiglie intervistate per ciascun Comune del campione è inferiore al numero teorico previsto dal disegno campionario. Al fine di calcolare l’effettiva proporzione delle correzioni per mancata risposta totale. Attraverso un processo iterativo di ottimizzazione matematica che tiene conto dei totali noti, si determina

---

<sup>3</sup> A differenza della regola hard, quella soft può essere disattivata dal rilevatore in maniera forzata qualora non riesca a rilevare informazioni coerenti.

<sup>4</sup> Nella letteratura in lingua inglese tale stimatore è noto con il nome di *calibration estimator*; per ulteriori informazioni si veda Deville & Sarndal (1992).

infine il coefficiente finale di riporto all'universo L'ultima fase riguarda il vincolo delle stime ottenute con il coefficiente base corretto per mancata risposta a "totali noti" desunti da fonti anagrafiche. La Rilevazione sulle Forze di Lavoro utilizza come totali noti: la popolazione residente per regione, 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), e la popolazione residente per provincia, sesso e cinque classi d'età (0-14, 15-29, 30-49, 50-64, 65 e più).

## 6. La diffusione dei risultati dell'indagine

La diffusione dei dati dell'indagine tiene conto delle molteplici esigenze degli utenti e si articola pertanto su diversi piani. Il primo è associato alla diffusione trimestrale del Comunicato stampa; il secondo al rilascio dei dati elementari; il terzo alle elaborazioni ad hoc; il quarto alla diffusione dei risultati dettagliati in media d'anno.

- Diffusione trimestrale

Dal I trimestre 2011 insieme all'uscita, in base al calendario prestabilito, del Comunicato stampa i risultati dell'indagine vengono diffusi attraverso il data-warehouse dell'Istituto I.Stat, raggiungibile all'indirizzo web: <http://dati.istat.it/>. Nella struttura gerarchica organizzata per aree tematiche del sistema I.Stat i risultati dell'indagine sulle forze di lavoro sono consultabili nella sezione dedicata al Lavoro. I dati sono presentati in forma aggregata in tavole multidimensionali, con la possibilità di comporre tabelle e grafici personalizzati, agendo sulle variabili, i periodi di riferimento e la disposizione di testate e fiancate. Il sistema è interrogabile secondo le seguenti modalità: per parola chiave, per tema e per territorio.

Fino al IV trimestre del 2010 i risultati venivano diffusi attraverso una serie di tavole in formato excel. Tali tavole, ancora disponibili nell'archivio dei comunicati stampa o nella sezione del data-warehouse I.Stat "*Esportazioni – File pronti all'uso*", riportano le serie storiche dei principali indicatori del mercato del lavoro. Si tratta dei file denominati:

1. popregio: che contiene i dati della popolazione distinti tra le forze di lavoro (occupati, persone in cerca di occupazione) e le non forze di lavoro (per grado di attachment al mercato del lavoro, con meno di 15 anni, con più di 64 anni). I dati, distinti per genere, sono articolati a livello regionale e riportano anche i principali indicatori (tasso di attività, di occupazione, di disoccupazione).
2. setregio: contenente i dati relativi agli occupati -dipendenti, indipendenti, totale per macro settore di attività (agricoltura, industria in senso stretto, costruzioni, servizi). I dati, distinti per genere, sono articolati a livello regionale.

3. ripartiz: con i dati articolati per ripartizione geografica (Nord-ovest, Nord-est, Centro, Mezzogiorno) della popolazione, delle forze di lavoro, dell'occupazione, dei tassi di occupazione, delle persone in cerca di occupazione, dei tassi di disoccupazione, delle non forze di lavoro. Diverse tavole sono articolate oltre che per ripartizione e genere anche per classe di età, ovvero per settore di attività, per carattere dell'occupazione, per titolo di studio.
4. ripartiz-stranieri: con le principali tavole del file ripartiz elaborate con riguardo esclusivamente ai cittadini stranieri. Per alcune stime viene segnalata la presenza di un elevato errore campionario tale da sconsigliarne l'utilizzo.

Sempre nell'ambito della diffusione trimestrale viene fornito un file che riporta una breve serie storica dei principali aggregati dell'indagine e un apposito cartogramma che presenta una disaggregazione dei dati dei principali aggregati di riferimento dell'indagine: occupati, persone in cerca di occupazione, inattivi.

- Dati elementari

Il rilascio dei dati elementari avviene attraverso file distinti, ognuno orientato ad uno specifico gruppo di utenti:

1. Il file Sistan, in virtù di una specifica regolamentazione tra Istat e soggetti Sistan, contiene un'informazione particolarmente dettagliata. In particolare:
  - permette elaborazioni per famiglia mediante l'uso di un codice familiare;
  - adotta la classificazione a 4 digit delle attività economiche;
  - adotta la classificazione a 4 digit delle professioni;
  - i dati per singolo anno di età;
  - rilascia i codici sia delle regioni sia delle province e dei comuni.

Solo i dati relativi ad alcuni quesiti sono riservati e pertanto non vengono rilasciati.

È disponibile a partire dal I trimestre 2004.

2. Il file standard (ottenuto dopo il trattamento del contenuto informativo del file originale per la riduzione del rischio di violazione della riservatezza) contiene:
  - alcune variabili ricostruite al fine di facilitare le elaborazioni;
  - i codici delle regioni ma non delle province e dei comuni;
  - la classificazione a 2 digit delle attività economiche;
  - la classificazione a 3 digit delle professioni;
  - i dati per classi di età;
  - l'oscuramento di alcune variabili familiari per motivi di riservatezza.

I dati relativi ad alcuni quesiti sono riservati e pertanto non vengono diffusi.

La produzione del file Standard è iniziata con il I trimestre 2004 ed è cessata con l'edizione relativa al IV trimestre 2013.

3. Il file per la ricerca che contiene:

- i codici delle regioni, delle province e dei grandi comuni;
- la classificazione a 4 digit delle attività economiche;
- la classificazione a 4 digit delle professioni (con un'aggregazione del primo grande gruppo);
- i dati per singolo anno di età;
- i dati per tutte le elaborazioni familiari.

Solo i dati relativi ad alcuni quesiti sono riservati e pertanto non vengono diffusi.

È disponibile a partire dal I trimestre 2009.

4. Il file mlcro.STAT, disponibile sul sito Istat, che presenta le seguenti caratteristiche:

- grazie ad una appropriata metodologia statistica volta a tutelare la riservatezza dei rispondenti, i mlcro.STAT possono essere scaricati liberamente e direttamente dal sito Istat;
- è ottenuto attraverso tecniche di campionamento applicate al file per la ricerca, di cui quindi eredita struttura, livello di dettaglio e trattamento delle variabili;
- a seguito del campionamento vengono calcolati i pesi di riporto all'universo da utilizzare per le analisi dei dati;
- precisione e accuratezza delle stime risultano inferiori rispetto a quelle ottenute utilizzando i microdati originali oppure il corrispondente file per la ricerca. Quindi è possibile che alcuni dati ottenuti elaborando il mlcro.STAT non coincidano con quanto già pubblicato dall'Istat.

È disponibile a partire dal I trimestre 2014.

Peraltro, il mondo della ricerca scientifica può avvalersi, anche con riguardo alla Rilevazione sulle forze di lavoro, dei servizi offerti dal Laboratorio per l'Analisi dei Dati ELEmentari (Laboratorio ADELE). In tale laboratorio i ricercatori possono, in un ambiente controllato, effettuare le proprie elaborazioni dei microdati dell'indagine. In questo caso i controlli relativi alla tutela della riservatezza sono preliminari al rilascio dei risultati prodotti.

- Elaborazioni ad hoc

Nonostante l'elevato numero delle elaborazioni effettuate, e rese disponibili nelle tavole diffuse correntemente, la vastità delle informazioni raccolte dall'indagine stimola l'ampia richiesta di elaborazioni ad hoc da parte degli utenti, singoli o istituzionali. Si tratta nella

gran parte dei casi della richiesta di elaborazioni riferite a specifici livelli di dettaglio settoriale, territoriale, lavorativo, ecc. Il rilascio dei risultati delle elaborazioni ad hoc è comunque soggetto alla verifica della validità statistica dei risultati stessi, con riguardo in particolare al rilievo dell'errore campionario.

- Dati di media

A circa un mese dall'uscita dei risultati del quarto trimestre (marzo di ogni anno) vengono diffusi in rete, accompagnati da un comunicato stampa, i dati della media relativi all'anno precedente. Si diffondono i risultati dettagliati per la popolazione, le forze di lavoro, gli occupati, le persone in cerca di occupazione e le non forze di lavoro. Vi sono poi informazioni relative alla frequenza di attività di istruzione e/o formazione professionale, disaggregate in base alla condizione professionale, al sesso, all'età e al territorio. Infine, nei risultati di media sono presenti anche quelli riferiti alle province.

## 7. 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.

**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.

**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.

**Dato destagionalizzato:** dato depurato dalla stagionalità.

**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.

## 8. Riferimenti bibliografici

- ISTAT, “*La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione.*” Collana Metodi e norme n. 32 - 2006 – Roma
- Deville J. C., Särndal C. E., (1992), “*Calibration Estimator in Survey Sampling*”, Journal of the American Statistical Association, vol. 87, pp.376-382.
- ISTAT, “*La progettazione e lo sviluppo informatico del sistema Capi sulle forze di lavoro.*” Collana Metodi e norme n. 36 - 2008 – Roma

## 9. Contatti

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

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

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

**DESCRIZIONE  
DEL FILE**

*FILE  
DESCRIPTION*





**File ad uso pubblico  
micro.STAT**

**Rilevazione Continua sulle Forze di Lavoro**  
dal primo trimestre 2014

**Descrizione del file**

Giugno 2014

# INDICE

1. Introduzione .....	3
2. Descrizione delle variabili .....	4
3. Metodologia statistica per la tutela della riservatezza .....	4
4. Riferimenti bibliografici.....	6
5. Contatti .....	6

## 1. Introduzione

I file di microdati ad uso pubblico (mlcro.STAT) sono collezioni campionarie di dati elementari relative ad alcune indagini svolte dall'Istat, per le quali siano già stati sviluppati i corrispondenti file di microdati per la ricerca (MFR).

Grazie ad una appropriata metodologia statistica volta a tutelare la riservatezza dei rispondenti, i mlcro.STAT possono essere scaricati liberamente e direttamente dal sito Istat.

Il file ad uso pubblico mlcro.STAT presenta le seguenti caratteristiche:

- è prodotto dal corrispondente file della ricerca attraverso tecniche di campionamento,
- la struttura, il livello di dettaglio ed il trattamento delle variabili sono “ereditate” dal corrispondente file per ricerca,
- a seguito del campionamento vengono calcolati i pesi di riporto all'universo da utilizzare per le analisi dei dati,
- precisione e accuratezza delle stime risultano inferiori rispetto a quelle ottenute utilizzando i microdati originali oppure il corrispondente file per la ricerca. Quindi è possibile che alcuni dati ottenuti elaborando il mlcro.STAT non coincidano con quanto già pubblicato dall'Istat.

Questo documento illustra le misure di protezione adottate al fine di ridurre il rischio di violazione della riservatezza.

La Sezione 2 fornisce alcune informazioni sintetiche sullo status di un insieme di variabili presenti nel file per la ricerca e nel mlcro.STAT, mentre la Sezione 3 contiene brevi cenni su ulteriori aspetti rilevanti per la riservatezza.

## 2. Descrizione delle variabili

Per il significato delle variabili e delle rispettive modalità si rinvia al questionario d'indagine. Si sottolinea che in alcuni casi il contenuto delle categorie può cambiare rispetto a quello originale, a parità di etichetta: nel prospetto seguente viene fornita indicazione delle variazioni che il mlcro.STAT “eredita” dal file per la ricerca.

**Tabella 1. Trattamento delle variabili a fini di tutela della riservatezza nel file mlcro.STAT.**

Variabile	Descrizione	Modifica per la tutela della riservatezza
GRACOM	codice Comune	In ciascuna provincia, codifica binaria in grande comune e comuni residuali
MFRFAM	codice famiglia	Ricodifica con metodo casuale
MFRIND	identificativo dell'individuo all'interno della famiglia	Ricodifica con metodo casuale
ETAM	età in anni compiuti	anni compiuti, fatti salvi i raggruppamenti: 00 fino a 2 anni, 03 da 3 a 5 anni, 06 da 6 a 10 anni, 11 da 11 a 14 anni, 75 75 anni e oltre.
STACIM	stato civile	aggregazione delle modalità 3, 4 e 5 nella 3 (separati e divorziati): 1 celibe/nubile 2 coniugato/a 3 separato/a o divorziato/a 6 vedovo/a
TFM	tipologia familiare	aggregazioni effettuate per coerenza con quelle relative allo stato civile: le modalità 12 e 13 sono accomunate nella 12, le 17 e 18 nella 18, 26 e 27 nella 26, 31 e 32 nella 31.
RETRIC	retribuzione in unità di euro del mese precedente escluse 13me, 14me ed emolumenti non percepiti tutti i mesi	raggruppamento in classi delle intensità fino a € 250 e oltre € 3000 e per le rimanenti arrotondamento ai 10 euro.
PROFM, PROFSM, PROFEM	posizione professionale, posizione professionale secondaria, posizione professionale precedente, espresse a 4 digit	limitazione al terzo digit per le posizioni di legislatori e dirigenti

## 3. Metodologia statistica per la tutela della riservatezza

### 3.1. Apprezzamento del rischio

Sono variabili identificative indirette (o chiave) quelle che da sole, o in combinazione con altre, possono portare alla re-identificazione di uno o più record. In particolare, sono considerate l'identificazione da archivio esterno e l'identificazione spontanea sulla base di

conoscenze a priori (“circolo delle conoscenze”). L’esistenza di un archivio esterno pubblico e nominativo, contenente informazioni associabili a quelle contenute nel file rilasciato, permetterebbe all’intruso di identificare l’unità statistica ed apprendere informazioni su di essa. Diversa è la fattispecie dell’identificazione riconducibile al “circolo delle conoscenze” ossia ascrivibile ad informazioni sufficienti alla re-identificazione e note all’utente senza che venga posto in essere alcun tentativo di intrusione.

### 3.2. Protezione dei dati

Assimilando il rischio di violazione della riservatezza al costo di inclusione di una unità statistica nel campione (Casciano, Ichim, Corallo, 2011), il campionamento è stato orientato alla minimizzazione del rischio e all’approssimativa riproduzione, nel tipo di dominio  $d=REG$ , degli insiemi di totali  ${}_1Y, {}_2Y, {}_3Y, {}_4Y$ , riferiti, rispettivamente, alle combinazioni di variabili:

- $SG11 \cap COND3 \cap DETIND \cap PIEPAR \cap DIPIND$ ,
- $SG11 \cap$ cinque classi di età (0-14; 15-29; 30-49; 50-64; 65 e più),
- $SG11 \cap CITTAD$ ,
- $EDULEV$ .

A seguito del campionamento sono stati aggiornati i pesi di riporto all’universo da utilizzare per le analisi dei dati.

### 3.3. Analisi del contenuto informativo

Ponendo

$p=1, \dots, P$  l’indice della variabile d’interesse,

$d$  il tipo di dominio,

$j_d=1, \dots, J_d$  il dominio di tipo  $d$ ,

$\hat{Y}_{j_d}^p$  Il totale pesato del file per la ricerca, nel dominio  $j_d$ , della  $p^{ma}$  variabile,

$\hat{Y}_{j_d}^{\hat{p}}$  Il totale pesato del mlcro.STAT, nel dominio  $j_d$ , della  $p^{ma}$  variabile,

si possono definire gli errori medi assoluti relativi :

$$emar_{p,d} \equiv \frac{1}{J_d} \sum_{j_d} \left| 1 - \frac{\hat{Y}_{j_d}^{\hat{p}}}{\hat{Y}_{j_d}^p} \right|$$

Ai fini interpretativi, occorre sottolineare che in presenza di caratteri qualitativi si debbono intendere quali variabili gli indicatori di presenza/assenza riferiti a ciascuna modalità. Ciò

precisato, per i totali negli insiemi  ${}_1Y, {}_2Y, {}_3Y, {}_4Y$  gli errori medi assoluti relativi sono risultati inferiori a 0.05.

## 4. Riferimenti bibliografici

1. AA.VV. (2010). *Handbook on Statistical Disclosure Control*. ESSnet on Statistical Disclosure Control. <http://neon.vb.cbs.nl/casc/handbook.htm>. 18/08/2011.
2. Casciano C., Ichim D., Corallo L. (2011). *Sampling as a way to reduce risk and create a Public Use File maintaining weighted totals*. Unece/Eurostat Work Session on Statistical Data Confidentiality. 26 - 28 Ottobre 2011, Tarragona (Spagna).
3. Falorsi, P.D., Ballin M., De Vitiis C., Scepi G. (1998). Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat. *Statistica applicata*. 10(2). 235-257.

## 5. Contatti

Per eventuali informazioni sulla metodologia di tutela della riservatezza, rivolgersi a:  
[rilascio.microdati@istat.it](mailto:rilascio.microdati@istat.it).

## Curatori dei paragrafi

Flavio Foschi, Alessandro Martini