

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**Facoltà di Scienze Statistiche**

**Corso di laurea in  
Statistica, Popolazione e Società**

Tesi di laurea

**Analisi e confronto fra censimenti di tutto il mondo:  
le potenzialità offerte da IPUMS-International**

Relatore Prof. Fiorenzo Rossi

Laureanda Elisa Fasani  
Matricola N. 572424 - SPT

Anno accademico 2009-2010



## SOMMARIO

INTRODUZIONE.....	7
1. INFORMAZIONI GENERALI SUL PROGETTO .....	9
1.1. COS'È IPUMS E COSA C'È NEL SUO FUTURO.....	9
1.2. COME I DATI DIFFERISCONO DAI CAMPIONI GIÀ IN DISTRIBUZIONE .....	10
1.3. IL VALORE AGGIUNTO AI DATI .....	10
2. I DATI .....	13
2.1. COSA SONO I MICRODATI.....	13
2.2. DOCUMENTAZIONE .....	14
2.3. COME OTTENERE I DATI.....	14
2.3.1. REGISTRAZIONE.....	14
2.3.2. RESTRIZIONI SULL'USO .....	15
2.3.3. USO DEL SISTEMA DI ESTRAZIONE DEI DATI .....	16
2.4. USARE I DATI IUPMS.....	19
2.4.1. ASPETTI DI CUI ESSERE CONSAPEVOLI .....	20
2.4.2. LIMITAZIONI SUI DATI.....	21
2.4.3. ULTERIORI INFORMAZIONI .....	21
3. VARIABILI .....	23
3.1. USARE LA PAGINA DELLE VARIABILI .....	23
3.2. VARIABILI INTEGRATE.....	24
3.3. VARIABILI "NON ARMONIZZATE".....	24
3.4. LE VARIABILI PUNTATORE.....	25
3.5. VERSIONE GENERALE E DETTAGLIATA DELLE VARIABILI .....	25
3.6. UNIVERSO DELLA VARIABILE.....	26

<b>4. ELABORAZIONI .....</b>	<b>27</b>
<b>4.1. STUDI SULLE MIGRAZIONI INTERNE .....</b>	<b>28</b>
<b>4.1.1. DETTAGLIO TERRITORIALE: CANADA, FRANCIA, REGNO UNITO E STATI UNITI .....</b>	<b>28</b>
<b>4.2. STUDI SULLA FECONDITÀ.....</b>	<b>29</b>
<b>4.2.1. IL METODO DEI FIGLI PROPRI .....</b>	<b>30</b>
<b>4.2.2. APPAIARE MADRE E FIGLIO CON IPUMS-INTERNATIONAL .....</b>	<b>30</b>
<b>4.2.3. ANALISI DELLA POSSIBILITÀ DI APPLICARE IL METODO DEI FIGLI PROPRI: CANADA, FRANCIA, REGNO UNITO E STATI UNITI .....</b>	<b>31</b>
<b>4.3. STUDI SULLA TIPOLOGIA E LA DIMENSIONE DELLE FAMIGLIE.....</b>	<b>31</b>
<b>4.3.1. DETTAGLIO FAMILIARE: CANADA, FRANCIA,ITALIA, REGNO UNITO E STATI UNITI ...</b>	<b>33</b>
<b>4.4 ANALISI DEI DATI.....</b>	<b>33</b>
<b>4.4.1 ANALISI DELLA FECONDITÀ .....</b>	<b>34</b>
<b>4.4.2 ANALISI SULLE FAMIGLIE.....</b>	<b>41</b>
<b>4.5 CONCLUSIONI.....</b>	<b>44</b>
<b>APPENDICE A - INFORMAZIONI SUI CAMPIONI IPUMS .....</b>	<b>45</b>
<b>BIBLIOGRAFIA .....</b>	<b>51</b>





## **INTRODUZIONE**

Questo lavoro si propone di essere una guida all'utilizzo del sito internet di un progetto chiamato IPUMS-International, e dei dati censuari che mette disposizione. L'obiettivo è quello di offrire una guida pratica all'utilizzo, tramite informazioni sulla richiesta dei dati e delle variabili di interesse, e sull'elaborazione degli stessi, mettendo in risalto le innumerevoli potenzialità che può offrire in termini di confronto tra dati censuari di uno stesso paese e rilevati in momenti storici differenti, e di confronto tra paesi diversi.

L'idea è nata dalla curiosità di valutare quali fossero i dati che l'ISTAT (Istituto Nazionale di Statistica Italiano) avesse reso disponibili per i fruitori di IPUMS-International, e se fossero gli stessi disponibili per gli utenti che ne facciano richiesta direttamente all'ISTAT.

Dopo di che, si sono volute analizzare velocemente le differenze con alcuni dati offerti da altri paesi.

Il sito di riferimento è il seguente:

<https://international.ipums.org/international/> ,

da cui si sono ricavate le informazioni, e cui accedere per la richiesta dei dati censuari dei paesi che hanno aderito.



## **1. INFORMAZIONI GENERALI SUL PROGETTO**

IPUMS-International (Integrated Public Use Microdata Series International) è un progetto dedicato alla raccolta e alla diffusione dei dati di censimento di tutto il mondo. I suoi obiettivi sono:

- raccogliere e conservare i dati e la documentazione;
- armonizzare i dati;
- diffondere i dati assolutamente gratis.

Il progetto ha raccolto il più grande archivio mondiale di campioni di censimento a disposizione del pubblico, codificando e documentando i dati in modo coerente tra i paesi per facilitare la ricerca comparata.

IPUMS-International rende questi dati disponibili gratuitamente a ricercatori qualificati, attraverso un sistema di diffusione via web.

Il progetto IPUMS è una collaborazione tra il Minnesota Population Center, gli istituti statistici nazionali, e gli archivi di dati internazionali.

Attualmente il finanziamento maggiore al progetto viene fornito da US National Science Foundation (fino al 2014) e da Demographic and Behavioral Sciences Branch of the National Institute of Child Health and Human Development; un ulteriore sostegno proviene dalla University of Minnesota, dal Minnesota Population Center, e da Sun Microsystems.

### **1.1. COS'È IPUMS E COSA C'È NEL SUO FUTURO**

IPUMS-International, progetto iniziato nel 1960, è la più grande collezione al mondo di dati di censimento disponibili a livello individuale. La versione attuale del database descrive circa 279 milioni di persone in 44 paesi per un totale di 130 censimenti. I censimenti provengono da: Argentina, Armenia, Austria, Bielorussia, Bolivia, Brasile, Cambogia, Canada, Cile, Cina, Colombia, Costa Rica, Ecuador, Egitto, Francia, Ghana, Guinea, Grecia, Ungheria, India, Iraq, Israele, Italia, Giordania, Kenya, Kirghizistan, Malaysia, Messico, Mongolia, Olanda, Palestina, Panama, Filippine, Portogallo, Romania, Ruanda, Slovenia, Sud Africa, Spagna, Uganda, Regno Unito, Stati Uniti, Venezuela, e Vietnam. (Si veda l'Appendice A per maggiori informazioni sui dati disponibili.)

IPUMS-International è composto di microdati. Ogni record è una persona con tutte le sue caratteristiche codificate numericamente, il cui nome e altre informazioni identificative vengono rimossi.

Le variabili sono state codificate in modo da poter consentire comparazioni tra nazioni e comparazioni temporali.

Nella maggior parte dei campioni le persone sono organizzate in "household", la qual cosa permette di studiare le caratteristiche delle persone nel contesto delle loro famiglie o di altri coresidenti.

Il progetto prevede di raccogliere circa 250 campioni di dati entro la fine del 2014. Sono infatti numerose le trattative in corso con paesi che hanno espresso interesse a partecipare a questa iniziativa.

L'aspettativa degli organizzatori è quella di poter continuare il progetto a tempo indeterminato, ma il problema è riuscire a garantirsi ulteriori finanziamenti. Nel sito viene ribadito più volte l'invito ad utilizzare i dati di IPUMS e a far conoscere le sue potenzialità, al fine di riunire una grande massa di utenti, e potenzialmente di pubblicazioni, che avvalorino il loro sforzo.

## **1.2. COME I DATI DIFFERISCONO DAI CAMPIONI GIÀ IN DISTRIBUZIONE**

Un numero limitato di campioni IPUMS sono distribuiti da alcuni uffici statistici nazionali attraverso mezzi alternativi al sito di IPUMS-International. I campioni IPUMS si differenziano in quanto sono sottoposti a una batteria di test per l'integrità strutturale, e talvolta i dati possono essere stati modificati per correggere gli errori nella struttura delle famiglie. IPUMS inoltre genera una ampia gamma di variabili tecniche e "costruite", integrandole e ricodificandole per rafforzare la ricerca comparativa. Tutta la documentazione è altresì tradotta in inglese e modificata a seconda dei vari casi.

## **1.3. IL VALORE AGGIUNTO AI DATI**

Il processo di integrazione delle variabili è di per sé un valore aggiunto ai dati. Vengono però effettuati molti altri passaggi.

- La maggior parte dei campioni di censimento hanno un certo numero di errori strutturali: una famiglia potrebbe essere incompleta (non tutti i membri vengono codificati), membri di famiglie differenti potrebbero essere uniti, record di persona potrebbero non avere i corrispondenti record di famiglia, e così via. Tutti i campioni IPUMS sono trattati con un insieme coerente di strumenti diagnostici per individuare tali problemi e risolverli tramite modifiche logiche o sostituzioni di intere famiglie.
- IPUMS-International crea un insieme coerente di variabili “costruite” per tutti i campioni. Le più importanti sono le “family interrelationship pointer variables” (variabili puntatore delle relazioni familiari), che indicano la posizione della madre, del padre o della sposa della persona considerata, presenti all'interno della famiglia.
- In futuro, IPUMS eseguirà modifiche di coerenza di dati mancanti sulle variabili importanti. Le modifiche di coerenza dei dati eseguiranno correzioni logiche quando possibile, o richiederanno di trovare un donatore di record che condivida le caratteristiche chiave con la persona in questione e, quindi, di inserirne la risposta del donatore come modalità di risposta alla variabile mancante.



## **2. I DATI**

La maggior parte dei dati di popolazione sono tradizionalmente disponibili solo in forma aggregata o sotto forma di tabella.

IPUMS-International è composto di microdati, il che significa che fornisce informazioni sulle singole persone e sulle famiglie di appartenenza.

Questo rende possibile per i ricercatori creare tabulati adeguati alle loro domande particolari. I dati sono infatti disponibili attraverso un sistema di estrazione che permette di selezionare le variabili e i campioni desiderati.

Le serie di dati contengono informazioni su una vasta gamma di caratteristiche di popolazione, tra cui la fecondità, la nuzialità, le migrazioni, la partecipazione lavorativa, l'istruzione, l'etnia e la composizione del nucleo familiare.

Le informazioni disponibili in ciascun campione variano a seconda delle domande poste in fase di censimento.

L'uso di questi dati è limitato a fini scolastici ed educativi, e i potenziali utenti devono, all'accesso, accettare di rispettare le norme indicate per proteggerne la riservatezza.

### **2.1. COSA SONO I MICRODATI**

I file di microdati sono collezioni campionarie di dati elementari, relative ad alcune indagini svolte. Possono essere rilasciati per fini di studio e di ricerca su richiesta motivata e previa autorizzazione del Presidente dell'Istituto che li fornisce, purché siano resi anonimi e privi di ogni riferimento che ne permetta il collegamento con singole persone fisiche e giuridiche.

I microdati di censimento sono composti di singole registrazioni contenenti informazioni raccolte sulle persone e sulle famiglie. L'unità di osservazione è l'individuo, e le risposte di ogni persona alle varie domande sono registrate in variabili separate.

## **2.2. DOCUMENTAZIONE**

Tutte le informazioni, relative ai 130 campioni di censimento e alle variabili in essi contenute, unitamente alla spiegazione delle codifiche utilizzate, sono reperibili in maniera dettagliata all'interno del sito.

È inoltre possibile reperire per ogni campione il questionario di censimento di provenienza in lingua originale e tradotto in inglese.

## **2.3. COME OTTENERE I DATI**

L'accesso alla documentazione presente sul sito è disponibile senza restrizioni, tuttavia, gli utenti devono registrarsi per poter accedere ai dati.

Per accedere ai dati, un potenziale utilizzatore deve presentare un modulo elettronico di autorizzazione e identificarsi tramite nome, indirizzo di posta elettronica, appartenenza istituzionale.

Sta agli utenti selezionare i censimenti e le variabili cui sono interessati, mentre sarà cura di IPUMS-International fornirli in un unico file di dati. I dati distribuiti non sono censimenti completi, ma campioni generati dal sistema di estrazione previa autorizzazione.

### **2.3.1. REGISTRAZIONE**

Una volta effettuato l'accesso alla pagina di registrazione si devono compilare informazioni identificative della persona, indicare il grado di affiliazione istituzionale dell'utente, e compilare, in inglese, una descrizione di almeno 75 parole dell'applicazione che si intende effettuare con i dati di IPUMS-International.

Nel caso in cui gli organizzatori fossero incerti in merito all'idoneità dell'applicazione indicata, potrebbero richiedere ulteriori informazioni tramite l'indirizzo e-mail che ogni utente deve fornire.

Se il progetto di ricerca coinvolge più persone si devono indicare nome e cognome di ogni componente e un indirizzo e-mail istituzionale al quale sia possibile contattarli, con l'obbligo di invitarli a registrarsi a loro volta come utenti in modo che ognuno abbia una sua licenza individuale per l'utilizzo dei dati.

Al termine della registrazione il candidato utilizzatore è tenuto ad accettare un certo numero di condizioni per l'utilizzo dei dati. L'accesso al sistema sarà consentito entro tre giorni lavorativi da IPUMS-International, e notificato via mail. Sarà possibile da questo momento in poi estrarre i dati relativi a qualsiasi paese presente nel database. La licenza è valida per un anno e può poi essere rinnovata.

### **2.3.2. RESTRIZIONI SULL'USO**

Gli accordi degli organizzatori con i vari uffici statistici nazionali richiedono che i dati IPUMS-International possano essere utilizzati solo a fini scolastici ed educativi, inclusa la ricerca di ordine pubblico.

Per ottenere l'accesso, i richiedenti sono tenuti ad accettare una serie di condizioni che fungono da contratto legale:

- I dati non devono essere ridistribuiti senza autorizzazione.
- I microdati sono destinati solo alla ricerca accademica e a scopi educativi.
- L'uso commerciale e la redistribuzione dei microdati sono severamente proibiti.
- L'uso dei microdati deve seguire rigide regole di riservatezza. Gli utenti devono mantenere la riservatezza delle persone e delle famiglie. Qualsiasi tentativo di risalire all'identità di persone o di famiglie a partire dai microdati è vietato. Risultati di analisi che possano rivelare l'identità delle persone o delle famiglie non possono essere segnalati o pubblicati in qualsiasi forma.
- Gli utenti devono attuare rigide misure di sicurezza per impedire l'accesso ai microdati acquisiti da IPUMS-International a personale non autorizzato.
- Sono consentite pubblicazioni scientifiche delle ricerche effettuate con i microdati IPUMS-International, purché sia citata in modo appropriato la fonte dei dati e sia esplicitato in modo chiaro che eventuali commenti ai risultati sono a cura dell'autore.

La citazione deve essere la seguente: **“Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2009.”**

- Infine, ogni utente accetta di notificare a [ipums@pop.umn.edu](mailto:ipums@pop.umn.edu) qualsiasi errore nei dati che dovesse identificare durante l’elaborazione.

### **2.3.3. USO DEL SISTEMA DI ESTRAZIONE DEI DATI**

Per effettuare un estratto dei dati è necessario essere registrati tramite la procedura descritta in 2.2.1 e aver ricevuto l’autorizzazione tramite una mail. L’estrazione dei dati è un processo che si articola nelle 5 fasi di seguito indicate.

0. Dopo aver selezionato “Create Extract”, identificarsi tramite il proprio indirizzo e-mail e la password definita in fase di registrazione (se non si è ancora ricevuta l’autorizzazione si può accedere come ospiti e intanto esplorare le funzionalità del sistema).

#### **1. Selezione dei campioni e delle variabili**

In questa fase, in una prima parte, è possibile selezionare i paesi, e relativi anni di censimento, per cui si desiderano ricevere i campioni. A lato di ogni anno di censimento selezionabile, vengono indicate, tramite un’icona informativa, caratteristiche particolari di quel campione, che riguardano la copertura dei dati o altri fattori di limitazione (ad esempio: persone non raggruppate in famiglie, età in classi, alcune provincie o parti di quello stato non rilevate), che potrebbero rendere problematica una futura analisi.

In una seconda schermata, è possibile selezionare le variabili di interesse per la propria analisi, che si desidera vengano inserite nel dataset. In questo contesto è possibile anche modificare i campioni precedentemente selezionati.

#### **2. Definire le opzioni per l’estratto**

È possibile in questa fase definire alcune caratteristiche generali del proprio estratto: la struttura e la lunghezza dei nomi delle variabili non armonizzate (per la definizione di variabile “non armonizzata” si veda paragrafo 3.3).

- Struttura preferita per il proprio estratto: rettangolare se si desidera che tutte le informazioni relative alla famiglia siano allegare ai rispettivi membri; gerarchico se si desidera il record di famiglia sia seguito dal record di persona (l'impostazione predefinita è quella di estratto rettangolare).
- Nomi per le variabili non armonizzate: nomi corti, sempre di 8 caratteri di lunghezza e mai descrittivi; nomi lunghi, più interpretabili rispetto ai primi, ma che possono raggiungere i 16 caratteri di lunghezza (nomi lunghi è l'impostazione di default).

### 3. Definire le opzioni delle variabili

In questa fase è possibile specificare le azioni da intraprendere con le variabili inserite nell'estratto. Queste azioni riguardano la possibilità di creare nuove variabili utilizzando le informazioni provenienti da record di persone diverse ("attach characteristics"), o la selezione di specifici tipi di casi da includere nell'estratto ("select cases").

La funzione "select cases" consente di limitare il dataset per contenere solo i record con valori specifici per le variabili selezionate (ad esempio persone con età superiore ad una certa soglia).

Più variabili possono essere utilizzate in combinazione per selezionare i casi, e la logica sottostante è di tipo "and".

È possibile scegliere di includere solo le persone che contengono i criteri di selezione, oppure di includere tutti coloro che risiedono in un nucleo familiare con una persona che soddisfa i criteri. Questa distinzione è davvero molto importante sul piano analitico.

La funzione "attach characteristics" permette di allegare una caratteristica della madre, del padre o del coniuge dell'unità di osservazione sul record della persona. È inoltre possibile allegare le caratteristiche del capo famiglia.

Un esempio di questa opzione è il seguente: utilizzando la variabile "Occupazione", si può creare una nuova variabile "Professione della madre". Tutte le persone dell'estratto che risiedono in un nucleo familiare all'interno del quale sia presente la madre ricevono un valore per questa nuova variabile. Per le persone prive di madre nel nucleo familiare, viene generato

un valore mancante. Il sistema di estrazione genera automaticamente un nome univoco per la nuova variabile.

Questa funzione “attach characteristics” utilizza le “family interrelationship pointer variables” (variabili puntatore delle relazioni familiari) costruite da IPUMS che identificano co-residenti madri, padri, e coniugi per ogni persona. Le variabili puntatore identificano madri e padri sociali, e non necessariamente i genitori biologici. Ovviamente queste caratteristiche non possono essere evidenziate per campioni che non siano organizzati in famiglie.

Queste azioni non sono possibili per tutte le variabili.

Se si è selezionata una variabile che è presente sia in versione generale che dettagliata, ogni versione viene interpretata come una variabile distinta; è possibile in questa schermata deselezionare la versione che non si desidera.

#### **4. Personalizzare le dimensioni del campione**

Questo passaggio è facoltativo. Nella parte superiore dello schermo del sistema di estrazione viene calcolata la dimensione prevista dei dati estratti. Se questa è troppo grande, o se si desidera gestire le dimensioni dei singoli campioni all'interno dell'estratto, si può aprire un link che rimanda alla pagina di ridimensionamento del campione.

La dimensione del campione non tiene conto di eventuali selezioni effettuate con l'opzione “select cases”, quindi le dimensioni effettive del campione potrebbero essere ridotte rispetto a quelle previste. Per personalizzare le dimensioni del campione basterà digitare uno a scelta tra il numero di famiglie, il numero di persone e la densità (% di campionamento) che si desiderano per il proprio estratto. Il sistema calcolerà da se gli altri due campi. Il numero minimo di casi selezionabile è di 10mila famiglie. In qualsiasi momento sarà possibile cancellare la selezione e tornare alla dimensione completa per ogni campione.

L'unità di campionamento per dimensionare l'estratto è la famiglia. Il sistema si baserà su un campionamento sistematico di estrazione delle famiglie, che partirà da un inizio casuale, fino a raggiungere la densità adeguata a produrre il numero di casi è richiesto.

La sintassi dei file creati dal sistema di estrazione contiene la programmazione utile per modificare i pesi per persona e familiari in modo da riflettere la densità del nuovo campione. L'estratto deve includere le variabili peso WTPER e WTHH e SAMPLE per poter calcolare i nuovi pesi tramite un pacchetto statistico.

#### **5. Revisione dell'estratto richiesto**

In questa pagina sono riassunte tutte le caratteristiche selezionate nelle fasi precedenti. È possibile scrivere una nota all'interno di una casella di testo per descrivere il contenuto dell'estratto, che potrà servire da riferimento per il futuro.

Una serie di link nella parte inferiore dello schermo permettono di tornare alle fasi precedenti per modificare alcune scelte. Nel momento in cui si è sicuri del proprio estratto basta inoltrare la richiesta.

Quando si richiede un estratto questo potrebbe essere disponibile in più o meno tempo a seconda delle dimensioni del campione. Una mail viene inviata dal sistema quando l'estratto è pronto.

### **2.4 USARE I DATI IUPMS**

Tutti i dati vengono prodotti in formato compresso gzip. Il sistema produce solo file ASCII a colonne fisse, ma con ogni estratto vengono generati anche i comandi per leggere i dati e assegnare le etichette alle variabili con uno dei seguenti pacchetti statistici: SAS, SPSS, STATA. Il sistema crea anche un file, che descrive il contenuto dell'estratto, e contenente inoltre la codifica della variabili presenti nel dataset.

I dati sono del tutto numerici, e di default non viene mantenuto l'ordinamento in famiglie, bensì vengono inserite le informazioni familiari all'interno dei record personali.

Ogni record viene identificato in modo univoco tramite tre variabili: SAMPLE, SERIAL, e PERNUM (rispettivamente identificativo del censimento, identificativo di famiglia e identificativo della persona all'interno della famiglia). La combinazione di SAMPLE e SERIAL sono l'identificativo univoco della famiglia.

#### 2.4.1 ASPETTI DI CUI ESSERE CONSAPEVOLI

Alcuni campioni sono ponderati: ogni individuo rappresenta un differente numero di persone nella popolazione. È quindi importante utilizzare la variabile peso WTPER per generare statistiche rappresentative quando si lavora con questi dati:

Argentina 1980-1991

Brasile 1970-2000

Canada 2001

Colombia 1985, 2005

Ecuador 1974

Egitto 1996

India 1983-1999

Messico 1995-2000

Mongolia 1989

Olanda 1960-2001

Palestina 1997

Panama 1980

Filippine 1990, 2000

Sud Africa 1996-2007

Spagna 1981-1991

Uganda 1991

Stati Uniti 1990-2005

Venezuela 1971, 1990

Vietnam 1989-1999

Per gli altri campioni l'uso dei pesi è opzionale.

Non tutti i campioni contengono l'intero universo della popolazione nazionale. Diverse sottopopolazioni, usualmente modeste, potrebbero essere mancanti o sottorappresentate, come ad esempio la popolazione istituzionalizzata, migranti, transitori, indigeni, o altri gruppi. In alcuni casi potrebbe anche essere che alcune sottosezioni del paese manchino del tutto.

È importante esaminare la documentazione relativa alle variabili che si utilizzano. I codici e le etichette categoriali non sono completamente esplicative del loro significato. Inoltre l'universo della variabile (la popolazione a rischio di risposta per quella variabile) può variare anche marcatamente tra i campioni.

Per impostazione predefinita i dati sono rettangolari, tutte le informazioni relative alla famiglia sono cioè allegate ai rispettivi membri. Questo può alterare cospicuamente le analisi familiari, poiché il numero di osservazioni viene gonfiato per il numero di record di persona.

#### **2.4.2 LIMITAZIONI SUI DATI**

I dati non forniscono informazioni macroeconomiche o commerciali né statistiche aggregate.

Di particolare importanza, per alcuni argomenti, è che la maggior parte dei censimenti non include domande sul reddito.

Alcuni campioni sono individuali e non raggruppati in famiglie; altri ancora non forniscono l'età in anni, bensì in classi.

La densità dei campioni può andare dall'1% al 10% della popolazione generale. Alcune sottopopolazioni potrebbero essere troppo piccole per essere studiate.

Per questioni relative alla riservatezza, nomi e altre informazioni identificative sono state soppresse. Per lo stesso motivo le informazioni geografiche sono solitamente limitate, a volte in modo grave. In molti campioni, i luoghi con meno di 20mila abitanti non sono individuati; in altri la soglia è addirittura più elevata, al punto che solo lo Stato o le regioni possono essere determinati.

#### **2.4.3 ULTERIORI INFORMAZIONI**

IPUMS-International è stato progettato per facilitare la ricerca tra diverse nazioni e tra diversi momenti storici, tuttavia non vi sono limitazioni sugli studi di un unico paese.

I dati estratti possono contenere record da campioni di tutte le serie di dati (cioè da censimenti effettuati in vari anni) o record provenienti da un singolo campione.

Alcuni paesi non hanno fornito le variabili per ogni domanda del censimento. Questo perché le domande possono non essere mai state trattate, o per questioni legate alla riservatezza. Potrebbe anche essere che i dati siano disponibili ma che non vi siano metadati sufficienti per renderli disponibili al momento.

Quando viene aggiunto un campione di dati, solo le variabili che gli organizzatori ritengono più utili per i ricercatori vengono armonizzate; le altre variabili possono essere lette attraverso le variabili non armonizzate, a patto che non siano troppo dettagliate da poter creare problemi di riservatezza.

IPUMS-International fornisce file di confine (shapefile) per facilitare la mappatura di dati nazionali e internazionali. Due livelli di file di confine sono disponibili: un file mondiale con confini a livello nazionale, e un file a livello nazionale con un alto livello di confini amministrativi (gli Stati, le province, ecc.) Gli utenti possono creare mappe utilizzando un programma statistico e ArcMap (un software di mappatura GIS).

### **3. VARIABILI**

Le variabili presenti in IPUMS-International sono suddivise principalmente in due grandi categorie: variabili integrate e variabili non armonizzate. Entrambi sono a loro volta suddivise in due sottogruppi: quelle relative alla "Household" (familiari) e quelle relative alla "Person" (individuali).

Il sottogruppo delle variabili di "household" si suddivide a sua volta in 9 gruppi che aiutano l'utente a ricercare la variabile di cui ha bisogno: variabili tecniche, variabili relative a abitazioni collettive, geografiche, economiche, relative ai consumi della casa, agli elettrodomestici e accessori presenti nella famiglia, caratteristiche dell'abitazione, altro e infine variabili costruite a posteriori sulle caratteristiche familiari.

Il sottogruppo delle variabili di "person" si suddivide, invece, in 12 gruppi: variabili tecniche, variabili delle relazioni familiari, demografiche, di fecondità e mortalità, relative al luogo di nascita, sull'etnia e il linguaggio, sull'educazione, il lavoro, il reddito, migrazioni, disabilità e altro.

#### **3.1. USARE LA PAGINA DELLE VARIABILI**

Il principale strumento per esplorare le potenzialità e il contenuto di IPUMS-International è la pagina delle variabili ("Variables").

Inizialmente la schermata è impostata per dare visione delle variabili armonizzate presenti nei campioni; è possibile però visualizzare anche le variabili non armonizzate, specifiche di un particolare campione.

Quando si selezionano specifici campioni, la pagina visualizza solo le variabili presenti per tali censimenti, e una "X" ne indica la disponibilità.

Cliccando sul nome di una variabile, si apre immediatamente un link che ne indica tutta la documentazione relativa. In particolare contiene una descrizione della variabile e le discussioni sui problemi di comparabilità internazionale. È possibile anche esplorare la pagina dei codici della variabile, che mostra la struttura dei codici e delle etichette e la disponibilità delle varie categorie tra i campioni. Queste categorie possono suggerire il tipo di ricerca possibile con un dato campione.

### **3.2. VARIABILI INTEGRATE**

L'integrazione - o "armonizzazione" - è il processo che rende comparabili dati provenienti da censimenti diversi. I censimenti possono infatti differire sia negli schemi di classificazione delle modalità di una variabile, sia nei codici numerici assegnati a ciascuna modalità. Per creare una variabile integrata si ricodifica la variabile in questione per ogni censimento con una codifica univoca.

Poiché alcuni censimenti forniscono maggiori dettagli rispetto ad altri, una codifica che riducesse le modalità fino a un minimo comune denominatore tra tutti i campioni sarebbe una perdita di informazioni importante. Per ovviare a questo inconveniente viene utilizzato un sistema di codifica composto: le prime cifre del codice forniscono informazioni disponibili in tutti i campioni; le successive, che riguardano un dettaglio informativo maggiore, sono disponibili per un ampio sottoinsieme dei campioni, e così via. In questo modo tutti i dettagli dell'enumerazione originale sono disponibili per i ricercatori. Questa struttura composta viene formalmente distinta in versione "generale" e "dettagliata" della variabile.

### **3.3. VARIABILI "NON ARMONIZZATE"**

Non tutte le variabili IPUMS sono integrate. Le variabili "non armonizzate", che sono uniche per ogni campione, in genere corrispondono a variabili originali dei dataset presentati dai vari paesi al progetto IPUMS.

Molto spesso sono le variabili "non armonizzate" a fungere da input per la realizzazione delle variabili integrate.

Le variabili non armonizzate possono essere incluse negli estratti, così che i ricercatori possano ottenere entrambi i moduli (variabili integrate e non armonizzate).

Alcune variabili non armonizzate non sono, in ogni caso, a disposizione degli utenti, a causa di preoccupazioni in merito alla riservatezza.

### **3.4. LE VARIABILI PUNTATORE**

Le “variabili puntatore delle relazioni familiari” (family interrelationship pointer variables) costruite da IPUMS identificano co-residenti madri, padri, e coniugi per ogni persona. Esse identificano madri e padri sociali, e non necessariamente i genitori biologici, e ovviamente queste caratteristiche non possono essere evidenziate per campioni che non siano organizzati in famiglie. Permettono di costruire variabili a livello individuale che rappresentano, ad esempio, l'occupazione del coniuge, l'età della madre, o il livello di istruzione del padre.

Per effettuare tali collegamenti, è stato sviluppato un complesso algoritmo di base personalizzato per tenere conto delle peculiarità dei campioni specifici.

Le variabili puntatore sono chiamate MOMLOC, POPLOC, and SPLOC nel sistema IPUMS. Le variabili PARRULE e SPRULE descrivono il criterio tramite cui MOMLOC, POPLOC e SPLOC sono stati linkati rispettivamente come possibili genitori e come coniuge della persona considerata.

### **3.5. VERSIONE GENERALE E DETTAGLIATA DELLE VARIABILI**

La maggior parte delle variabili presenti in IPUMS-International hanno una struttura composita di codifica, dove la prima cifra è in larga misura comparabile tra i campioni, mentre la seconda e le successive forniscono dettagli sempre maggiori, disponibili in alcuni campioni e non in altri.

La struttura di codifica composita è molto interessante per alcune variabili, e in IPUMS si distingue tra la versione "generale" di codifica, disponibile per quasi tutti i campioni e che permette comparazioni a livello internazionale, e "dettagliata", che permette distinzioni più sottili.

Le due serie di codici sono del tutto coerenti l'una con l'altra, e variano solo per il livello di dettaglio.

Quando si seleziona una variabile dal sistema di estrazione che possiede più versioni, queste sono entrambe selezionate per impostazione predefinita, ma è possibile deselezionarle a piacere.

### **3.6. UNIVERSO DELLA VARIABILE**

L'universo è la popolazione a rischio di avere una risposta per la variabile in questione. Nella maggior parte dei casi queste sono le famiglie o le persone alle quali vengono poste le domande di censimento. Per alcune domande però, come ad esempio quelle relative all'occupazione o alla fecondità, non tutti gli individui hanno la possibilità di rispondere: i bambini non risponderanno a domande sul lavoro, e gli uomini non risponderanno a domande in merito alla loro fecondità. IPUMS-International verifica empiricamente gli universi per ottenerne la descrizione più precisa possibile. In alcuni casi, non vi è però alcuna informazione indipendente che dia la possibilità di verificare un universo.

I casi che per una variabile sono al di fuori dell'universo, sono classificati come "NIU (not in universe)" nella pagina dei codici.

Nei primi dati che IPUMS-International comunicò, erano state eseguite alcune modifiche per ripulire gli universi, ma questo approccio non era stato abbastanza completo per garantire la corretta identificazione della variabile con informazioni difettose. Le modifiche per l'universo furono quindi rimosse nel dicembre 2006.

#### 4. ELABORAZIONI

L'analisi delle potenzialità offerte da IPUMS-International, è iniziata per la curiosità di capire quali fossero i dati di censimento che l'ISTAT aveva reso disponibili, e in particolare se fossero gli stessi che vengono forniti, previa richiesta, a privati che necessitano di fare ricerca.

Con i dati che vengono forniti ai privati, non si possono infatti eseguire analisi sulle migrazioni interne (a causa dei limiti dovuti al dettaglio territoriale) e analisi che utilizzino il metodo dei figli propri (a causa delle età espresse in classi).

Dopo una breve analisi delle informazioni offerte dal sito, e uno sguardo alle variabili, si può subito capire che i dati forniti, con riguardo al censimento italiano, sono gli stessi che l'Istat rende disponibili ai privati.

La particolarità, con l'utilizzo dei dati forniti da IPUMS-International, risiede innanzitutto nella possibilità di accedere alle variabili armonizzate per effettuare confronti internazionali.

Inoltre IPUMS-International permette, anche per analisi svolte al solo livello nazionale, di sfruttare tutte le variabili puntatore, evitando la necessità di elaborare, tramite l'utilizzo di programmi statistici, istruzioni che creino ad esempio i collegamenti con i genitori, il coniuge, i figli, ecc..

Queste potenzialità, unitamente alla creazione dei comandi per leggere i dati e assegnare le etichette alle variabili con uno tra i pacchetti statistici SAS, SPSS, STATA, permettono di risparmiare molto tempo per effettuare le analisi.

Si è pensato allora di valutare la possibile applicazione dei dati offerti da IPUMS-International in lavori relativi alle migrazioni interne, alla fecondità differenziale e alle famiglie, analizzando i dati disponibili sia per l'Italia che per altri paesi.

Si sono analizzate alcune informazioni sui dati di alcuni paesi che si possono ritenere avanzati dal punto di vista statistico, come Canada, Stati Uniti, Francia, e Regno Unito per valutare le differenze con i dati italiani, e capire se permettano alcune analisi di nostro interesse.

#### **4.1. STUDI SULLE MIGRAZIONI INTERNE**

Relativamente all'Italia, il Professore Fiorenzo Rossi si era occupato, con Silvia Meggiolaro, nel volume "Da Nord Est a Nord Ovest. Gli emigrati veneti in Italia nel XX secolo." della popolazione veneta emigrata principalmente negli anni tra il 1951 e il 1971, ma anche nei periodi precedenti e seguenti, verso le regioni del nord ovest d'Italia. Di questa erano state esaminate, sfruttando i dati campionari dei censimenti 1981 e 1991, alcune caratteristiche morfologiche individuali (sesso, età, stato coniugale, istruzione, lavoro) e familiari (tipologia e dimensione delle famiglie, caratteristiche delle abitazioni), e caratteristiche dinamiche (nuzialità e fecondità), confrontandole con le stesse caratteristiche della popolazione di partenza e di quella di arrivo.

Su questo solco si sarebbe voluto continuare ad esaminare i superstiti, attraverso il censimento del 2001, degli emigrati veneti nelle regioni nordoccidentali italiane.

L'edizione oscurata dei files campionari non ha consentito a questo scopo alcun esame, in quanto la classificazione del luogo di nascita dei residenti indica solamente se la persona è nata nella stessa regione di residenza oppure no, e non consente di identificare il nome della regione di migrazione.

Non è quindi possibile individuare i nati in una determinata regione e viventi in un'altra regione al momento del censimento.

È possibile altresì utilizzare i dati forniti da IPUMS-International, che altro non sono che i file campionari dei censimenti nazionali, per effettuare le analisi relativamente alle migrazioni, interne e non, allo stesso modo che si sarebbe potuto fare richiedendo direttamente i dati agli enti nazionali di competenza.

In questo modo è possibile però evitare tutto il lavoro di creazione delle istruzioni per codificare le variabili ed eventualmente creare collegamenti con i familiari per valutarne l'influsso sugli avvenimenti.

##### **4.1.1. DETTAGLIO TERRITORIALE: CANADA, FRANCIA, REGNO UNITO E STATI UNITI**

- Canada: il Canada offre un dettaglio territoriale che arriva ad identificare le Province (le Province canadesi corrispondono in termini territoriali alle

Regioni italiane). Il luogo di nascita viene identificato attraverso la Provincia di nascita, e le migrazioni rispetto a 1 o 5 anni precedenti il censimento, specificano il nome della Provincia di provenienza.

- Francia: la Francia offre un dettaglio territoriale che arriva ad identificare le Regioni (le Regioni francesi corrispondono in termini territoriali a qualcosa di intermedio tra le province e le Regioni italiane). Nel luogo di nascita viene indicato il nome della Regione, e per le migrazioni sono indicati la Regione o il paese da cui si è emigrati.
- Regno Unito: il Regno Unito offre un dettaglio territoriale che arriva ad identificare le Regioni (che corrispondono indicativamente alle Regioni italiane). Nel luogo di nascita viene indicato il nome della Regione, e per le migrazioni sono indicati oltre ad un dettaglio regionale, anche un dettaglio sub-regionale (est o ovest di una Regione).
- Stati Uniti: gli Stati Uniti offrono un dettaglio territoriale nel luogo di nascita che arriva ad identificare il nome dello Stato. Per le migrazioni sono indicate anche le città di residenza ad un anno dalla migrazione.

Si può notare da questo breve excursus sui dettagli territoriali, come vi siano molti paesi che, diversamente dall'Italia, offrono senza problemi maggiori caratteristiche che permettono di effettuare analisi sulle migrazioni interne.

#### **4.2. STUDI SULLA FECONDITÀ**

L'uso del metodo dei figli propri (own-children), ideato per paesi con buone statistiche censuarie, può essere applicato con ottimi risultati a dati di censimento di paesi sviluppati (Rossi, Maffenini, 1984; Rossi, Calovi, 1987). Esso consente di individuare fecondità differenziale tra gruppi di popolazioni non individuabili dalle statistiche correnti. Oltre al caso degli emigrati (vedi 4.1) sono stati studiati in passato la popolazione dell'Alto Adige secondo i gruppi linguistici (Colombo, Maffenini, Rossi, 1980 e 1983) e quella della provincia di Milano con determinate esperienze migratorie (Clerici, 1988 e 1989). Oggi sarebbe di estremo interesse esaminare la fecondità differenziale secondo alcune caratteristiche rilevate dal censimento: ad esempio, stato occupazionale, titolo di studio, situazione familiare, distanza dal nato precedente, ecc.

#### **4.2.1. IL METODO DEI FIGLI PROPRI**

Il metodo dei figli propri consiste nello stimare quozienti di fecondità per specifiche caratteristiche. Attraverso il collegamento di ciascun bambino con la “madre” presunta, si perviene ad una stima dei figli nati dalle donne di ciascuna età. Oltre che per età della madre questi nati potranno venir classificati secondo tutti quei caratteri della donna che sono stati rilevati dal censimento. È possibile quindi ottenere indicazioni sulla fecondità e sul calendario riproduttivo delle donne appartenenti ai diversi gruppi di classificazione, e quindi poter effettuare confronti. (Clerici 1988)

Per effettuare questo tipo di analisi sui dati di censimento, le caratteristiche cui i dati devono rispondere sono:

- possibilità di ricavare l'età del minore e della madre;
- possibilità di appaiare il figlio alla presunta madre.

L'età nel dataset non deve quindi essere espressa in classi, e si devono avere persone raggruppate in famiglie, e quindi caratterizzate da un identificativo familiare che permetta di unire figlio e madre.

#### **4.2.2. APPAIARE MADRE E FIGLIO CON IPUMS-INTERNATIONAL**

È possibile con un programma statistico adeguato, tramite una serie di istruzioni, identificare le possibili madri presenti in un nucleo familiare e collegarle con i rispettivi presunti figli.

IPUMS-International ha creato però una serie di variabili (family interrelationship pointer variables, vedi cap.3.4) che permettono di identificare co-residenti madri, padri, e coniugi per ogni persona. Esse identificano madri e padri sociali, e non necessariamente i genitori biologici, e permettono di costruire variabili a livello individuale che rappresentano, ad esempio, l'occupazione del coniuge, l'età della madre, o il livello di istruzione del padre.

Per effettuare tali collegamenti, è stato sviluppato un complesso algoritmo di base personalizzato per tenere conto delle peculiarità dei campioni specifici.

Ciascuno dei campioni organizzati per famiglie, contiene una variabile che indica il rapporto di ciascun membro della famiglia con il capo famiglia, come elencato nel modulo censimento.

Utilizzare le variabili SPLOC, MOMLOC, e POPLOC, che IPUMS-International mette a disposizione dei suoi utenti, permette di poter ricavare, senza effettuare una lunga programmazione su un pacchetto statistico, le caratteristiche del coniuge o dei genitori a cui si è interessati collegando le caratteristiche di un membro della famiglia ad un altro. Nella maggior parte dei casi, gli utenti dovrebbero essere in grado di manipolare queste variabili per costruire proprie misure all'interno di un pacchetto statistico.

MOMLOC è la variabile che meglio si adatta per qualsiasi studio sulla fecondità differenziale si voglia effettuare, e risulta fondamentale per l'applicazione del metodo dei figli propri e di estrema importanza per l'analisi differenziale sulla base delle caratteristiche della madre.

#### **4.2.3. ANALISI DELLA POSSIBILITÀ DI APPLICARE IL METODO DEI FIGLI PROPRI: CANADA, FRANCIA, REGNO UNITO E STATI UNITI**

Il metodo dei figli propri non è utilizzabile con i dati di censimento italiani, a meno che non ci si accontenti di identificare le età in classi, cosa che renderebbe l'analisi decisamente poco precisa e poco informativa.

Il Canada e il Regno Unito non permettono l'utilizzo del metodo dei figli propri: le persone non sono organizzate in famiglie all'interno del censimento, e quindi non è possibile collegare il figlio con la presunta madre.

Con i dati provenienti da Francia e Stati Uniti è invece possibile effettuare l'analisi con l'uso del metodo "own-children": sono infatti rispettati i due principali requisiti riguardanti la possibilità di ricavare l'età del minore e della madre, e la possibilità di appaiare il figlio alla presunta madre poiché le persone sono identificate all'interno della famiglia.

#### **4.3. STUDI SULLA TIPOLOGIA E LA DIMENSIONE DELLE FAMIGLIE**

La tipologia e la dimensione familiare, note al momento del censimento, possono essere studiate in modo più approfondito, in particolare per quanto riguarda la

presenza di specifiche figure accanto alla tradizionale famiglia di un solo nucleo, o di famiglie multiple (presenza di più nuclei).

Gli studiosi della famiglia, inoltre, si stanno muovendo sempre più da un'analisi svolta a livello familiare, verso un'analisi svolta a livello individuale relativamente a caratteristiche familiari. Ad esempio, invece di misurare la percentuale di famiglie con a capo con un monogenitore madre, si potrebbe valutare la proporzione di donne che sono monogenitori, o la percentuale di bambini che risiedono con monogenitori madri. Le variabili puntatore (family interrelationship pointer variables) sono decisamente molto adatte a questo tipo di valutazioni.

In aggiunta a POPOLOC, MOMLOC e SPLOC di cui si è già parlato in questo lavoro, IPUMS-International offre una serie di altre variabili costruite per lo studio delle famiglie.

1 NFAMS	Numero di nuclei <sup>(1)</sup> nella famiglia <sup>(2)</sup>
2 NCOUPLES	Numero di coppie di sposi presenti nella famiglia
3 NMOTHERS	Numero di donne con figli propri presenti nella famiglia
4 NFATHERS	Numero di uomini con figli propri presenti nella famiglia
5 PERNUM	Numero d'ordine della persona all'interno della famiglia
6 RELATE	Relazione con il capofamiglia
7 FAMSIZE	Numero di membri dello stesso nucleo residenti nella stessa famiglia
8 FAMUNIT	Numero del nucleo
9 SPLOC	Identificativo del coniuge
10 MOMLOC	Identificativo della madre
11 POPLOC	Identificativo del padre
12 NCHILD	Numero di figli nella famiglia
13 NCHLT5	Numero di figli con meno di 5 anni nella famiglia
14 ELDCH	Età del figlio più anziano nella famiglia
15 YNGCH	Età del figlio più giovane nella famiglia
16 NSIBS	Numero di fratelli e sorelle propri presenti nella famiglia

NOTE:

<sup>(1)</sup>Insieme di persone coabitanti che sono legate dal vincolo di coppia (coniugate o non coniugate) e/o dal vincolo genitore-figlio. Più in particolare un figlio continua ad essere considerato facente parte del nucleo familiare dei genitori (o del genitore) solo fino a che non costituisce una nuova coppia o fino a che non diventa genitore egli stesso, ossia fino a quando non forma un altro nucleo familiare. Nell'ambito di una famiglia possono esistere uno o più nuclei, o può non esservene nessuno (caso di un unico abitante nella casa).

<sup>(2)</sup>Insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o vincoli affettivi, coabitanti e aventi dimora abituale nello stesso comune. Una famiglia può essere costituita anche da una sola persona.

### 4.3.1. DETTAGLIO FAMILIARE: CANADA, FRANCIA,ITALIA, REGNO UNITO E STATI UNITI

Il Canada e il Regno Unito non permettono analisi familiari poiché le persone non sono organizzate in famiglie all'interno del censimento. Si possono ricavare alcune frammentarie informazioni sulla dimensione e la tipologia della famiglia, ma non si possono collegare tra i loro i componenti coabitanti per effettuare analisi più specifiche.

Con i dati di Francia, Italia e Stati Uniti è possibile invece:

- descrivere la composizione della household, e in particolare classificarla secondo le seguenti modalità:

	FR	IT	US
Una sola persona nella famiglia	X	X	X
Coppia sposata o convivente, senza figli	X	X	X
Coppia sposata o convivente, con figli	X	X	X
Famiglia monogenitore	X	X	X
Nucleo allargato, solo parenti	X	X	X
Nucleo composito, nucleo e estranei	X	X	X
Famiglia senza nuclei	X	X	X
Coabitazioni di gruppo	X	X	. <sup>(1)</sup>
Non classificabile	. <sup>(1)</sup>	X	X

NOTE: <sup>(1)</sup> la modalità non è disponibile per questo paese.

- effettuare tutta una serie di analisi basate sui costrutti familiari, che permettano di valutare la composizione delle famiglie all'interno di una stessa abitazione, collegando tra loro i componenti e confrontando i valori delle variabili.

### 4.4 ANALISI DEI DATI

Per dare un esempio pratico dell'utilizzo dei dati offerti da IPUMS-International, ho eseguito alcune elaborazioni relative all'applicazione del metodo dei figli propri, e all'analisi della tipologia familiare.

Ho deciso di analizzare i dati francesi che si prestano bene per entrambi le elaborazioni.

Ho richiesto a IPUMS-International i dati relativi al censimento Francese del 1999, contenenti le seguenti variabili:

TIPOLOGIA	VARIABILE	ETICHETTA
H	<b>SAMPLE</b>	Identificativo del campione censimento da cui si è fatta l'estrazione
H	<b>SERIAL</b>	Identificativo della household
H	<b>HHTYPE</b>	Classificazione della household
H	<b>NFAMS</b>	Numero di famiglie nella household
P	<b>PERNUM</b>	Numero della persona all'interno della household
P	<b>WTPER</b>	Peso della persona
P	<b>AGE</b>	Età
P	<b>SEX</b>	Sesso
P	<b>EMPSTAT</b>	Stato occupazionale
P	<b>HRSFULL</b>	Lavoro Full-time o part-time
H	<b>FR99A_PERNNF</b>	Numero di individui non parte della family
P	<b>WTPER_MOM</b>	Peso della madre
P	<b>AGE_MOM</b>	Età dalla madre
P	<b>EMPSTAT_MOM</b>	Stato occupazionale della madre
P	<b>HRSFULL_MOM</b>	Lavoro Full-time o part-time della madre

Utilizzando il programma statistico SAS ho quindi elaborato i dati: tramite i comandi messi a disposizione da IPUMS-International ho letto il file di dati in formato ASCII e assegnato le etichette a tutte le variabili.

A questo punto non ho dovuto far altro che estrarre dai dati i record utili per le mie elaborazioni, e infine creare le tabelle di frequenza necessarie.

#### 4.4.1 ANALISI DELLA FECONDITÀ

La mia attenzione si è inizialmente concentrata sull'analisi della fecondità delle donne considerando bambini che avessero fino a 10 anni al momento del censimento.

In particolare ho creato un SAS-dataset contenente i minori di 11 anni, e uno contenente tutte le donne con età compresa tra i 15 e i 60 anni. In questo modo ho avuto a disposizione tutte le informazioni utili per classificare i minori per età della madre al parto dall'elaborazione dei dati del primo dataset, e dal secondo ho potuto ricavare il numero totale di donne in età fertile classificate per età. Per non appesantire le elaborazioni non ho tenuto conto in questa sede delle

probabilità di sopravvivenza per ricalcolare la stima delle donne presenti x anni prima del censimento.

Ho quindi calcolato:

- i tassi specifici di fecondità per età della madre

$$f_x^{1999-i} = \frac{\text{nati da donne di età } x \text{ con figli di età } y}{\frac{1}{2} * [\text{donne di età } x \text{ nell'anno}(1999-y) - \text{donne di età } x \text{ nell'anno}(1999-y-1)]}$$

- i TFT (tassi di fecondità totale per età del figlio al momento del censimento)

$$TFT = \sum_x f_x^y$$

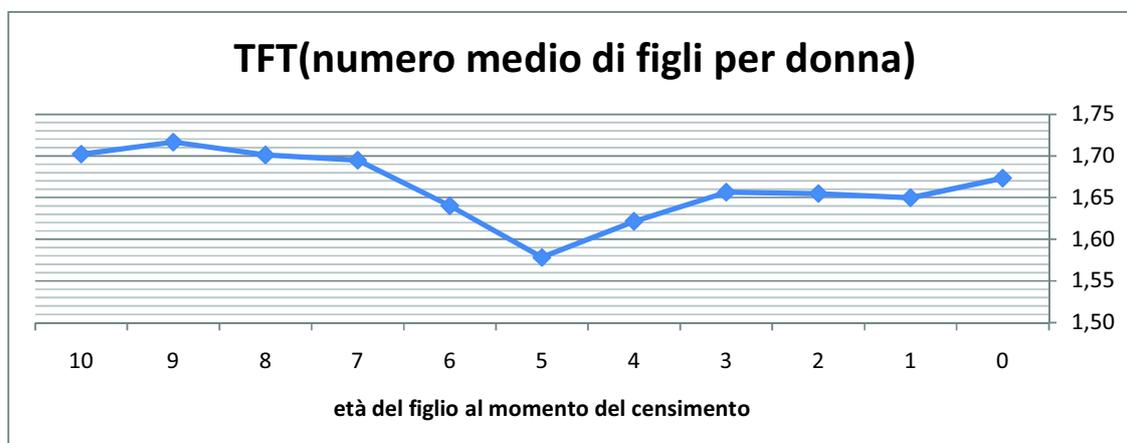
**TAB.1 : NUMERO DI FIGLI PER ETÀ DELLA MADRE AL PARTO E PER ETÀ DEL FIGLIO AL MOMENTO DEL CENSIMENTO.**

ETÀ DELLA MADRE AL PARTO	ETÀ DEL FIGLIO AL MOMENTO DEL CENSIMENTO										
	0	1	2	3	4	5	6	7	8	9	10
15	200	300	220	260	520	460	420	440	660	420	500
16	480	400	560	740	760	900	720	960	820	1200	1140
17	1180	1680	1460	1220	1440	1720	1960	2000	2280	2000	2440
18	2620	2540	3040	3440	3380	4020	4060	4680	4460	4820	5200
19	6500	6060	6800	5900	6360	7800	9100	9860	9760	9640	10860
20	8340	8980	9140	10100	11100	12660	12600	15040	15740	15560	16580
21	12700	12800	13260	14560	17180	16400	18200	20420	21000	20960	24160
22	16120	17440	19120	20360	21520	23180	26780	27540	27880	29780	31620
23	21080	23720	25960	27800	28840	28800	31920	33400	38100	39760	39540
24	30560	32480	33820	35520	37680	36620	39580	42600	43640	48220	49700
25	40600	41580	42500	44760	43980	43640	47220	52000	55560	54980	56100
26	48740	51020	49960	50680	49420	49800	54900	56780	59560	60980	59700
27	54500	54460	54360	55740	54060	54820	56880	61840	60540	59360	59240
28	57700	55440	56360	57760	56840	57480	56600	60800	57080	59280	56680
29	57260	53500	55100	57480	56940	52860	55220	52900	54520	55820	53840
30	52460	54600	55100	55300	54060	50020	50600	48800	47840	48820	47420
31	50400	48240	51100	49760	47800	43260	43420	45720	44240	42920	39860
32	43600	45260	44380	43500	39260	36700	38700	37460	39060	36280	36400
33	40260	38180	38680	36580	33360	32040	32980	32780	31640	30720	30020
34	35720	32840	31160	29620	29000	27840	27260	25560	25660	26220	24880
35	30140	27220	25920	26820	23300	22100	22560	22260	22280	21560	20540
36	22580	23040	21760	20320	20140	18580	18380	19100	18740	17460	17600
37	19040	17680	16980	15740	14900	14760	14600	15960	13840	14380	12960
38	14680	13840	13360	12660	11980	11700	11860	11780	11140	10980	10160
39	11760	10220	10420	9480	9340	9040	8680	8920	8960	8220	8040
40	8220	7820	7960	7520	8180	7060	6820	7020	6540	6680	5720
41	5740	5540	5180	4980	4380	4200	4840	4680	4100	4520	4600
42	3860	4060	3500	3600	3400	3700	3780	3220	2980	3120	2800
43	2200	2580	2040	1760	2100	2480	2120	2240	1760	1880	1280
44	1540	1580	1640	1300	1300	1040	1080	1220	1360	1080	1160
45	840	820	860	780	840	780	980	840	840	680	500
46	580	740	580	460	600	500	560	700	320	460	340
47	420	320	400	340	260	540	320	160	240	240	220
48	340	260	280	140	180	220	240	260	80	220	360
49	180	380	240	220	320	240	160	280	80	320	220
50	280	60	200	180	200	160	140	180	180	140	80
<b>TOTALE</b>	<b>703420</b>	<b>697680</b>	<b>703400</b>	<b>707380</b>	<b>694920</b>	<b>678120</b>	<b>706240</b>	<b>730400</b>	<b>733480</b>	<b>739680</b>	<b>732460</b>

**TAB.2 : TASSI DI FECONDITÀ SPECIFICI PER ETÀ DELLA MADRE (%), TFT E ETÀ MEDIA AL PARTO PER ETÀ DEL FIGLIO AL MOMENTO DEL CENSIMENTO.**

ETÀ DELLA MADRE AL PARTO	ETÀ DEL FIGLIO AL MOMENTO DEL CENSIMENTO										
	0	1	2	3	4	5	6	7	8	9	10
15	0,05	0,08	0,06	0,07	0,14	0,13	0,12	0,12	0,17	0,1	0,12
16	0,12	0,1	0,14	0,2	0,21	0,25	0,2	0,25	0,2	0,28	0,27
17	0,3	0,43	0,39	0,33	0,39	0,48	0,52	0,49	0,54	0,47	0,58
18	0,67	0,68	0,83	0,94	0,94	1,06	1	1,1	1,05	1,14	1,24
19	1,75	1,65	1,86	1,63	1,68	1,91	2,14	2,31	2,3	2,31	2,62
20	2,27	2,46	2,53	2,67	2,72	2,98	2,95	3,55	3,77	3,76	3,96
21	3,48	3,54	3,5	3,57	4,05	3,84	4,29	4,89	5,07	5,01	5,65
22	4,46	4,61	4,69	4,8	5,04	5,47	6,41	6,66	6,66	6,96	7,26
23	5,57	5,82	6,12	6,52	6,8	6,89	7,71	7,98	8,91	9,13	8,94
24	7,5	7,65	7,93	8,38	9,02	8,85	9,46	9,96	10,02	10,91	11,14
25	9,56	9,75	10,03	10,71	10,63	10,43	11,04	11,94	12,57	12,32	12,72
26	11,42	12,04	11,95	12,25	11,81	11,64	12,6	12,84	13,35	13,83	13,77
27	12,86	13,03	13,14	13,32	12,64	12,58	12,86	13,86	13,73	13,69	13,61
28	13,81	13,4	13,46	13,5	13,05	13	12,68	13,79	13,17	13,62	12,98
29	13,84	12,78	12,88	13,19	12,88	11,84	12,52	12,2	12,53	12,78	12,43
30	12,53	12,77	12,65	12,51	12,11	11,34	11,67	11,21	10,95	11,27	11,11
31	11,78	11,07	11,56	11,15	10,84	9,98	9,98	10,47	10,21	10,05	9,35
32	10,01	10,24	9,94	9,87	9,06	8,43	8,86	8,65	9,15	8,51	8,51
33	9,11	8,56	8,77	8,44	7,67	7,34	7,61	7,68	7,42	7,18	7,05
34	8	7,45	7,19	6,81	6,64	6,43	6,38	6	6	6,16	5,92
35	6,84	6,28	5,96	6,14	5,38	5,18	5,29	5,21	5,23	5,13	4,87
36	5,21	5,29	4,98	4,69	4,72	4,36	4,3	4,48	4,46	4,14	4,15
37	4,38	4,05	3,92	3,69	3,5	3,45	3,43	3,8	3,28	3,39	3,06
38	3,36	3,19	3,13	2,97	2,8	2,75	2,82	2,8	2,63	2,6	2,38
39	2,71	2,39	2,45	2,22	2,19	2,15	2,06	2,1	2,12	1,92	1,87
40	1,93	1,84	1,86	1,77	1,95	1,68	1,61	1,66	1,53	1,55	1,33
41	1,35	1,3	1,22	1,18	1,04	0,99	1,14	1,09	0,95	1,05	1,09
42	0,9	0,95	0,83	0,85	0,8	0,87	0,88	0,75	0,69	0,74	0,77
43	0,52	0,61	0,48	0,41	0,5	0,58	0,49	0,52	0,42	0,52	0,42
44	0,37	0,37	0,39	0,31	0,3	0,24	0,25	0,29	0,38	0,35	0,38
45	0,2	0,19	0,2	0,18	0,2	0,18	0,23	0,23	0,27	0,23	0,17
46	0,14	0,17	0,14	0,11	0,14	0,12	0,15	0,23	0,11	0,16	0,13
47	0,1	0,07	0,09	0,08	0,06	0,15	0,1	0,05	0,08	0,09	0,09
48	0,08	0,06	0,07	0,03	0,05	0,07	0,08	0,09	0,03	0,09	0,13
49	0,04	0,09	0,06	0,06	0,1	0,08	0,06	0,1	0,03	0,12	0,08
50	0,13	0,03	0,1	0,11	0,13	0,11	0,1	0,14	0,14	0,1	0,06
<b>TFT(numero medio di figli per donna)</b>	<b>1,67</b>	<b>1,65</b>	<b>1,65</b>	<b>1,66</b>	<b>1,62</b>	<b>1,58</b>	<b>1,64</b>	<b>1,69</b>	<b>1,7</b>	<b>1,72</b>	<b>1,7</b>
<b>ETÀ MEDIA AL PARTO</b>	<b>30,06</b>	<b>29,87</b>	<b>29,76</b>	<b>29,61</b>	<b>29,47</b>	<b>29,33</b>	<b>29,2</b>	<b>29,08</b>	<b>28,92</b>	<b>28,88</b>	<b>28,69</b>

**GRAFICO1 : TASSI DI FECONDITÀ SPECIFICI PER ETÀ DELLA MADRE (%) RELATIVI ALL'ETÀ DEI FIGLI AL MOMENTO DEL CENSIMENTO.**



NOTE: ho invertito le età dei figli per evidenziare la temporalità, da dieci anni prima del censimento all'anno di censimento.

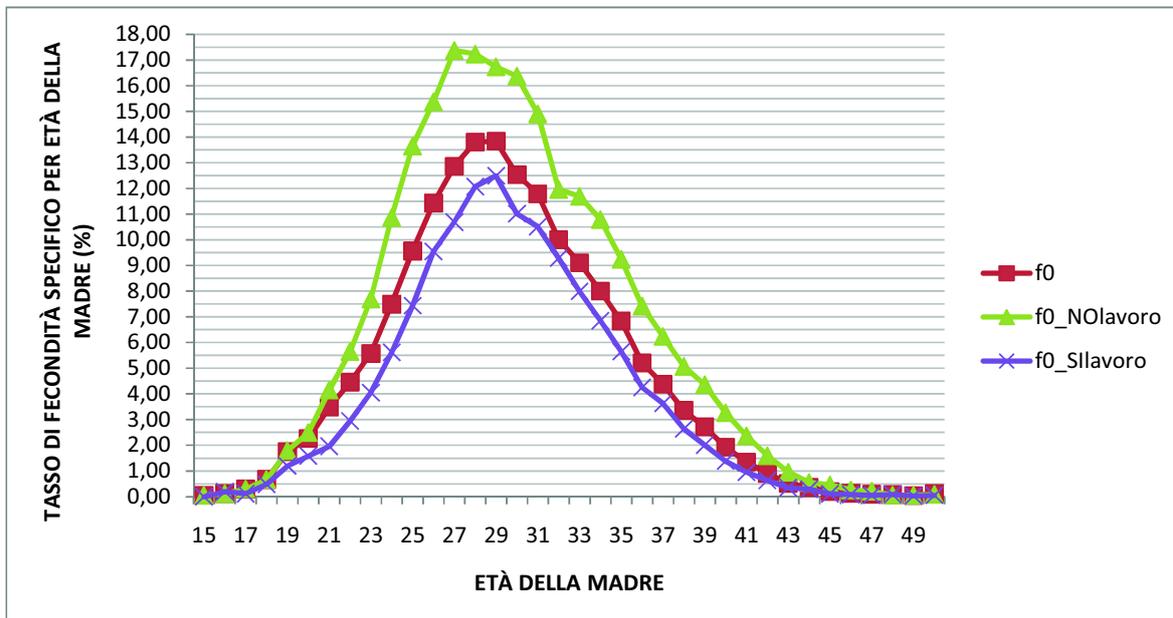
Dal grafico si nota come il tasso di fecondità totale delle donne francesi sia diminuito da 1,67 figli per donna nell'anno 1989 (età dei figli 10 anni), fino ad un picco negativo di 1,58 figli per donna nel 1994 (età dei figli 5 anni). Poi sembra esservi stata una piccola ripresa negli anni successivi, spinta verso il raggiungimento del livello iniziale.

Per usufruire delle caratteristiche della madre ho provato a calcolare i tassi specifici per età della madre nell'anno di censimento condizionatamente alla posizione lavorativa della madre. Dopo aver valutato se la madre lavorasse oppure no, ho confrontato i tassi di fecondità specifici per età delle lavoratrici part-time con quelli delle lavoratrici full-time.

**TAB.3 : TASSI DI FECONDITÀ SPECIFICI PER ETÀ DELLA MADRE (%) RELATIVI AI NATI CON 0 ANNI AL MOMENTO DEL CENSIMENTO, TFT CONDIZIONATI ALLO STATO OCCUPAZIONALE DELLA MADRE.**

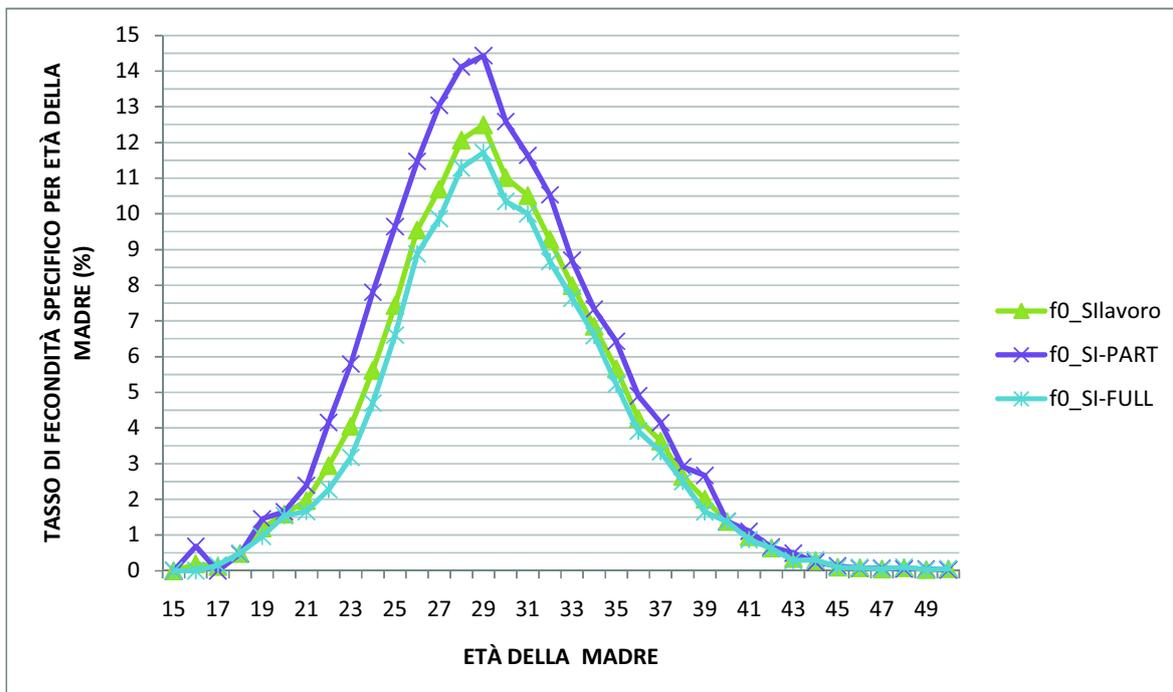
ETÀ MADRE	f <sup>0</sup>	f <sup>0</sup> _NOlavoro	f <sup>0</sup> _SIlavoro	f <sup>0</sup> _SI-PART	f <sup>0</sup> _SI-FULL
15	0,05	0,05	0	0	0
16	0,12	0,12	0,19	0,68	0,00
17	0,30	0,31	0,12	0,00	0,16
18	0,67	0,67	0,48	0,47	0,49
19	1,75	1,80	1,20	1,45	0,96
20	2,27	2,50	1,59	1,66	1,55
21	3,48	4,15	1,96	2,39	1,66
22	4,46	5,65	2,95	4,15	2,28
23	5,57	7,69	4,05	5,79	3,18
24	7,50	10,87	5,61	7,80	4,70
25	9,56	13,65	7,43	9,64	6,60
26	11,42	15,37	9,54	11,48	8,88
27	12,86	17,37	10,69	13,04	9,86
28	13,81	17,24	12,07	14,12	11,29
29	13,84	16,74	12,49	14,43	11,72
30	12,53	16,38	11,02	12,58	10,35
31	11,78	14,90	10,51	11,64	10,00
32	10,01	11,96	9,28	10,53	8,67
33	9,11	11,70	7,99	8,70	7,64
34	8,00	10,79	6,85	7,33	6,58
35	6,84	9,24	5,65	6,43	5,22
36	5,21	7,42	4,26	4,90	3,91
37	4,38	6,24	3,63	4,15	3,34
38	3,36	5,07	2,64	2,92	2,50
39	2,71	4,36	2,01	2,67	1,65
40	1,93	3,27	1,38	1,38	1,38
41	1,35	2,37	0,95	1,11	0,87
42	0,90	1,60	0,64	0,67	0,62
43	0,52	0,96	0,35	0,50	0,28
44	0,37	0,55	0,29	0,24	0,31
45	0,20	0,46	0,11	0,13	0,09
46	0,14	0,26	0,09	0,07	0,09
47	0,10	0,22	0,05	0,07	0,05
48	0,08	0,06	0,09	0,05	0,10
49	0,04	0,06	0,03	0,05	0,03
50	0,13	0,09	0,05	0,03	0,06
<b>TFT</b>	<b>1,67</b>	<b>2,22</b>	<b>1,38</b>	<b>1,63</b>	<b>1,27</b>

**GRAFICO2 : TASSI DI FECONDITÀ SPECIFICI PER ETÀ DELLA MADRE (%) RELATIVI AI NATI CON 0 ANNI AL MOMENTO DEL CENSIMENTO E ALLO STATO OCCUPAZIONALE (LAVORATRICE SI/NO) DELLE MADRI.**



L'essere lavoratrici o meno influisce sulla fecondità. Il grafico risulta sempre maggiore per le donne che non lavorano rispetto alle lavoratrici.

**GRAFICO3 : TASSI DI FECONDITÀ SPECIFICI PER ETÀ DELLA MADRE (%) RELATIVI AI NATI CON 0 ANNI AL MOMENTO DEL CENSIMENTO E ALLO STATO OCCUPAZIONALE (LAVORATRICE PART-TIME/FULL-TIME) DELLE MADRI.**



Questo grafico infine mostra come l'essere lavoratrici a tempo pieno influisca sulla minore fecondità rispetto all'essere lavoratrici part-time.

#### 4.4.2 ANALISI SULLE FAMIGLIE

Per non appesantire l'analisi, visto che non c'è un obiettivo specifico per queste elaborazioni, ma solo l'interesse di valutare la facilità di analizzare i dati tramite l'uso delle variabili di IPUMS-International, ho elaborato delle semplici tabelle che riportassero la composizione delle famiglie dei bambini con età compresa tra 0 e 10 anni.

Ho quindi valutato la percentuale dei bambini secondo la tipologia familiare di appartenenza.

**TAB.4 : NUMERO DI BAMBINI PER ETÀ DEL BAMBINO AL MOMENTO DEL CENSIMENTO E PER TIPOLOGIA FAMILIARE**

	ETÀ DEL BAMBINO										
	0	1	2	3	4	5	6	7	8	9	10
<b>Coppia sposata o convivente, con figli</b>	619020	609480	610140	610300	592660	575140	594380	612400	613740	616060	607860
<b>Famiglia monogenitore</b>	45340	51880	60300	66160	72160	77340	85140	91060	95920	100660	101760
<b>Nucleo allargato, solo parenti</b>	43020	40580	37120	35300	35080	32280	34960	35520	35200	34360	35120
<b>Nucleo composito, nucleo e estranei</b>	3840	3460	3760	4440	3720	4180	4160	4360	4280	3860	4360
<b>Famiglia senza nuclei</b>	180	100	120	140	220	100	140	160	220	260	160
<b>Coabitazioni di gruppo</b>	6120	6040	5320	5360	5440	4720	5500	6840	7060	6880	8040
<b>TOTALE</b>	717520	711540	716760	721700	709280	693760	724280	750340	756420	762080	757300

**TAB.5 : % DI BAMBINI PER TIPOLOGIA FAMILIARE E PER ETÀ DEL BAMBINO AL MOMENTO DEL CENSIMENTO**

	ETÀ DEL BAMBINO										
	0	1	2	3	4	5	6	7	8	9	10
<b>Coppia sposata o convivente, con figli</b>	86,27	85,66	85,12	84,56	83,56	82,90	82,06	81,62	81,14	80,84	80,27
<b>Famiglia monogenitore</b>	6,32	7,29	8,41	9,17	10,17	11,15	11,76	12,14	12,68	13,21	13,44
<b>Nucleo allargato, solo parenti</b>	6,00	5,70	5,18	4,89	4,95	4,65	4,83	4,73	4,65	4,51	4,64
<b>Nucleo composito, nucleo e estranei</b>	0,54	0,49	0,52	0,62	0,52	0,60	0,57	0,58	0,57	0,51	0,58
<b>Famiglia senza nuclei</b>	0,03	0,01	0,02	0,02	0,03	0,01	0,02	0,02	0,03	0,03	0,02
<b>Coabitazioni di gruppo</b>	0,85	0,85	0,74	0,74	0,77	0,68	0,76	0,91	0,93	0,90	1,06
<b>TOTALE</b>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Si possono quindi valutare le composizioni familiari, con attenzione magari alle modalità meno tradizionali (famiglie monogenitoriali, famiglie coabitanti con parenti) che potrebbero assumere una certa rilevanza sociale. Questa analisi sarebbe molto interessante se confrontata con i censimenti precedenti dello stesso paese, al fine di valutare l'evoluzione dei fenomeni, e se confrontata con i censimenti di altri paesi per poter magari valutare l'applicazione e l'elaborazione di specifiche politiche assistenziali.

**TAB.6 : NUMERO DI NUCLEI NELLA FAMIGLIA PER ETÀ DEL BAMBINO AL MOMENTO DEL CENSIMENTO, E % SUL TOTALE PER ETÀ**

NUMERO NUCLEI NELLA FAMIGLIA	ETÀ DEL BAMBINO										
	0	1	2	3	4	5	6	7	8	9	10
<b>1</b>	709240	704280	709540	713740	701980	686320	716640	741260	747540	753640	748000
%	98,846	98,98	98,993	98,897	98,971	98,928	98,945	98,79	98,826	98,893	98,772
<b>2</b>	5640	4620	4860	5320	4860	5320	4840	5560	5100	4600	5420
%	0,786	0,6493	0,6781	0,7371	0,6852	0,7668	0,6682	0,741	0,6742	0,6036	0,7157
<b>3</b>	1160	1200	1060	1160	940	760	1060	1140	1560	1280	1380
%	0,1617	0,1686	0,1479	0,1607	0,1325	0,1095	0,1464	0,1519	0,2062	0,168	0,1822
<b>4 e più</b>	1480	1440	1300	1480	1500	1360	1740	2380	2220	2560	2500
%	0,2063	0,2024	0,1814	0,2051	0,2115	0,196	0,2402	0,3172	0,2935	0,3359	0,3301

**TAB.7 : NUMERO DI PERSONE NON FACENTI PARTE DEL NUCLEO, MA BITANTI, PER ETÀ DEL BAMBINO AL MOMENTO DEL CENSIMENTO, % SUL TOTALE PER ETÀ DEL BAMBINO**

NUMERO PERSONE NON FACENTI PARTE DEL NUCLEO	ETÀ DEL BAMBINO										
	0	1	2	3	4	5	6	7	8	9	10
<b>1</b>	680280	675100	681540	685480	672960	658800	686120	709640	714640	722100	714540
%	95,626	95,691	95,797	95,692	95,613	95,611	95,456	95,446	95,367	95,617	95,366
<b>2</b>	23940	24240	23740	24440	24760	24280	25980	27140	27680	26240	28080
%	3,3652	3,4359	3,3369	3,4118	3,5178	3,5237	3,6145	3,6503	3,6938	3,4746	3,7477
<b>3</b>	4120	3480	3300	3840	3600	3880	3880	4280	4280	4040	3520
%	0,5791	0,4933	0,4638	0,5361	0,5115	0,5631	0,5398	0,5757	0,5712	0,535	0,4698
<b>4 e più</b>	3060	2680	2860	2580	2520	2080	2800	2440	2760	2820	3120
%	0,4301	0,3799	0,402	0,3602	0,358	0,3019	0,3895	0,3282	0,3683	0,3734	0,4164

Nonostante non vi sia un obiettivo specifico nell'elaborazione, è facile mostrare come, con semplicità, si possano creare tabelle sui costrutti familiari che potrebbero essere di estremo interesse per decisioni di tipo assistenziale e di valutazione del mutamento delle dinamiche familiari.

#### 4.5 CONCLUSIONI

A prescindere dai risultati di queste semplici e banali analisi, quello che mi premeva evidenziare era:

- utilizzando i dati offerti da IPUMS-International serve solo qualche basilare nozione sull'utilizzo di un pacchetto statistico per elaborarli;
- si sarebbero potuti chiedere, nello stesso file, anche dati relativi ad altri censimenti francesi disponibili, oppure da censimenti di altri paesi. In questo modo si sarebbero potuti confrontare i risultati ottenuti, in maniera trasversale rispetto al tempo o alla geografia, aggiungendo un semplice comando che stratificasse le analisi secondo il censimento di provenienza;
- l'applicazione del metodo dei figli propri è applicabile con estrema facilità ai dati disponibili, e il tempo di programmazione e elaborazione a computer viene ridotto drasticamente, con il vantaggio che non si rischia di incappare in errori di distrazione.

Io spero che questa breve guida di valutazione delle caratteristiche offerte da IPUMS-International possa essere utile per farne comprendere le potenzialità in termini di confronto fra censimenti e la possibilità di applicazione per analisi di tipo valutativo dei fenomeni.

## APPENDICE A

### INFORMAZIONI SUI CAMPIONI IPUMS

#### CENSIMENTI PER I QUALI SONO PRESENTI I CAMPIONI

1. Argentina	1970-1980-1991-2001	16. Ghana	2000	31. Palestine	1997
2. Armenia	2001	17. Greece	1971-1981-1991-2001	32. Panama	1960-1970-1980-1990-2000
3. Austria	1971-1981-1991-2001	18. Guinea	1983-1996	33. Philippines	1990-1995-2000
4. Belarus	1999	19. Hungary	1970-1980-1990-2001	34. Portugal	1981-1991-2001
5. Bolivia	1976-1992-2001	20. India	1983-1987-1993-1999	35. Romania	1977-1992-2002
6. Brazil	1960-1970-1980-1991-2000	21. Iraq	1997	36. Rwanda	1991-2002
7. Cambodia	1998	22. Israel	1972-1983-1995	37. Slovenia	2002
8. Canada	1971-1981-1991-2001	23. Italy	2001	38. South Africa	1996-2001-2007
9. Chile	1960-1970-1982-1992-2002	24. Jordan	2004	39. Spain	1981-1991-2001
10. China	1982-1990	25. Kenya	1989-1999	40. Uganda	1991-2002
11. Colombia	1964-1973-1985-1993-2005	26. Kyrgyz Republic	1999	41. United Kingdom	1991-2001
12. Costa Rica	1963-1973-1984-2000	27. Malaysia	1970-1980-1991-2000	42. United States	1960-1970-1980-1990-2000-2005
13. Ecuador	1962-1974-1982-1990-2001	28. Mexico	1960-1970-1990-1995-2000-2005	43. Venezuela	1971-1981-1990-2001
14. Egypt	1996	29. Mongolia	1989-2000	44. Vietnam	1989-1999
15. France	1962-1968-1975-1982-1990-1999	30. Netherlands	1960-1971-2001		

Sample design	Sample fraction (%)	Households	Persons	Weighted	De jure - De facto	Census date (d-m-yr)	Smallest geography	Collective dwellings
Argentina 1970	2	129,728	466,892	–	de facto	30/09/1970	department	yes
Argentina 1980	10	672,062	2,667,714	yes	de facto	22/10/1980	department	yes
Argentina 1991 <sup>(1)</sup>	10	1,148,351	4,143,727	yes*	de facto	19/05/1991	department	yes
Argentina 2001	10	1,040,852	3,626,103	–	de facto	17/18-11-01*	department	yes
Armenia 2001	10	81,929	326,56	–	de jure	09/11/2001	province	no
Austria 1971	10	264,655	749,894	–	de jure	12/05/1971	NUTS3 region	yes
Austria 1981	10	283,693	756,556	–	de jure	12/05/1981	NUTS3 region	yes
Austria 1991	10	310,099	780,512	–	de jure	15/05/1991	NUTS3 region	yes
Austria 2001	10	341,035	803,471	–	de jure	15/05/2001	NUTS3 region	yes
Belarus 1999	10	385,508	990,706	–	de facto	16/02/1999	region	no
Bolivia 1976	10	121,378	461,699	–	de facto	10/07/1976	province	yes
Bolivia 1992	10	177,926	642,368	–	de facto	03/06/1992	province	yes
Bolivia 2001	10	239,475	827,692	–	de facto	01/09/2001	province	yes
Brazil 1960 <sup>(2)</sup>	5	613,273	3,001,439	–	both	01/09/1960	state	yes
Brazil 1970	5	1,022,207	4,953,759	yes	both	01/09/1970	state	yes
Brazil 1980	5	1,343,377	5,870,467	yes	both	01/09/1980	municipality	yes
Brazil 1991	5.08	2,012,276	8,522,740	yes	de jure	01/09/1991	municipality	yes
Brazil 2000	6	2,652,356	10,136,022	yes	de jure	01/08/2000	municipality	yes
Cambodia 1998	10	223,513	1,141,254	–	de facto	03/05/1998	district	yes
Canada 1971 <sup>(3)</sup>	1	n.a.	214,019	–	de jure	01/06/1971	province	yes
Canada 1981 <sup>(3)</sup>	2	n.a.	486,875	–	de jure	03/06/1981	province*	yes
Canada 1991 <sup>(3)</sup>	3	n.a.	809,654	–	de jure	04/06/1991	province*	no
Canada 2001 <sup>(3)</sup>	2.05	n.a.	801,055	yes	de jure	15/05/2001	province*	no
Chile 1960 <sup>(3)</sup>	1	n.a.	88,184	–	de facto	29/11/1960	municipality	yes
Chile 1970	10	199,041	890,481	–	de facto	22/04/1970	municipality	yes
Chile 1982	10	282,356	1,133,062	–	de facto	21/05/1982	municipality	yes
Chile 1992	10	373,964	1,335,055	–	de facto	22/04/1992	municipality	yes
Chile 2002	10	486,115	1,513,914	–	de facto	24/04/2002	municipality	yes

Sample design	Sample fraction (%)	Households	Persons	Weighted	De jure - De facto	Census date (d-m-yr)	Smallest geography	Collective dwellings
China 1982	1	2,428,658	10,039,191	-	de jure	01/07/1982	city/prefecture	yes
China 1990	1	3,152,818	11,835,947	-	de jure	01/07/1990	city/prefecture	yes
Colombia 1964 <sup>(3)</sup>	2	n.a.	349,652	-	de facto	15/07/1964	municipality	no
Colombia 1973	10	349,853	1,988,831	-	de facto	24/10/1973	municipality	no
Colombia 1985	10	571,046	2,643,125	yes	de facto	15/10/1985	municipality	no
Colombia 1993	10	774,321	3,213,657	-	de facto	24/10/1993	municipality	no
Colombia 2005	10	1,054,901	4,117,607	yes	de facto	05-2005 to 02-2006*	municipality	no
Costa Rica 1963 <sup>(3)</sup>	6	n.a.	82,345	-	de jure	31/03/1963	canton	yes
Costa Rica 1973	10	36,323	186,762	-	de jure	14/05/1973	canton	yes
Costa Rica 1984	10	56,186	241,22	-	de jure	11/06/1984	canton	yes
Costa Rica 2000	10	106,973	381,5	-	de jure	28/06/2000	canton	yes
Ecuador 1962 <sup>(3)</sup>	3	n.a.	136,443	-	de facto	25/11/1962	province	yes
Ecuador 1974	10	145,902	648,678	yes	de facto	08/06/1974	canton	yes
Ecuador 1982	10	195,401	806,834	-	de facto	28/11/1982	canton	yes
Ecuador 1990	10	243,898	966,234	-	de facto	25/11/1990	canton	yes
Ecuador 2001	10	354,222	1,213,725	-	de facto	25/11/2001	canton	yes
Egypt 1996	10	1,270,787	5,902,243	yes	de facto	18/19-11-96	district	no
France 1962	5	748,917	2,320,901	-	de jure	07/03/1962	region	yes*
France 1968	5	815,699	2,487,778	-	de jure	01/03/1968	region	yes*
France 1975	5	915,624	2,629,456	-	de jure	20/02/1975	region	yes*
France 1982	5	969,632	2,631,713	-	de jure	04/03/1982	region	yes*
France 1990	4.02	949,893	2,360,854	-	de jure	05/15-03-90	region	yes*
France 1999	5	1,219,323	2,934,758	-	de jure	08/03/1999	region	yes
Ghana 2000	10	397,097	1,894,133	-	de facto	26/03/2000	district	yes
Greece 1971	10	249,35	845,483	-	de facto	14/03/1971	municipality	no
Greece 1981	10	294,323	923,108	-	de facto	05/04/1981	municipality	no
Greece 1991	10	320,387	951,875	-	de facto	17/03/1991	municipality	no
Greece 2001	10	367,438	1,028,884	-	de facto	18/03/2001	municipality	no

Sample design	Sample fraction (%)	Households	Persons	Weighted	De jure - De facto	Census date (d-m-yr)	Smallest geography	Collective dwellings
Guinea 1983	10	110,777	457,837	–	de facto	01/02/1983	prefecture	yes
Guinea 1996	10	108,793	729,071	–	de facto	01/12/1996	prefecture	yes
Hungary 1970	5	172,831	515,119	–	de jure	01/01/1970	none	yes
Hungary 1980	5	211,355	536,007	–	de facto	01/01/1980	none	yes
Hungary 1990	5	219,389	518,24	–	de facto	01/01/1990	none	yes
Hungary 2001	5	227,252	510,502	–	de facto	01/02/2001	none	yes
India 1983 <sup>(4)</sup>	0.09	120,847	623,494	yes	de jure	01-83 to 12-83	region	no
India 1987 <sup>(4)</sup>	0.09	129,06	667,848	yes	de jure	07-87 to 06-88	region	no
India 1993 <sup>(4)</sup>	0.07	115,409	564,74	yes	de jure	07-93 to 06-94	region	no
India 1999 <sup>(4)</sup>	0.07	120,578	596,688	yes	de jure	07-99 to 06-00	region	no
Iraq 1997 <sup>(5)</sup>	10	265,402	1,944,278	–	de jure	17/10/1997	district	yes
Israel 1972 <sup>(6)</sup>	10	89,19	315,608	–	de jure	19/20-05-72	none	yes*
Israel 1983 <sup>(6)</sup>	10	124,61	403,474	–	de jure	04/06/1983	subdistrict*	yes*
Israel 1995 <sup>(6)</sup>	10	177,412	556,365	–	de jure	04/11/1995	subdistrict*	yes*
Italy 2001 <sup>(6)</sup>	5	1,168,044	2,990,739	–	de jure	21/10/2001	region	yes
Jordan 2004	10	97,343	510,646	–	de facto	01/10/2004	district	yes
Kenya 1989	5	224,861	1,074,098	–	de facto	25/10/1989	district	yes*
Kenya 1999	5	317,106	1,407,547	–	de facto	25/08/1999	district	yes*
Kyrgyz Republic 1999	10	110,285	476,886	–	both	24/03/1999	district	yes
Malaysia 1970 <sup>(7)</sup>	2	38,006	175,997	–	de facto	25/08/1970	district	yes
Malaysia 1980 <sup>(7)</sup>	2	38,049	182,601	–	de facto	10/06/1980	district	yes
Malaysia 1991	2	78,192	347,892	–	de facto	14/08/1991	district	yes
Malaysia 2000	2	95,564	435,3	–	de jure	05/07/2000	district	no
Mexico 1960 <sup>(3)</sup>	1.05	n.a.	502,8	–	de jure	08/06/1960	state	no
Mexico 1970	1	82,856	483,405	–	de jure	28/01/1970	state	no
Mexico 1990	10	1,648,280	8,118,242	–	de jure	12/03/1990	municipality	no
Mexico 1995	0.04	72,277	332,061	yes	de jure	05/11/1995	municipality	no
Mexico 2000	10.06	2,312,035	10,099,182	yes	de jure	14/02/2000	municipality	no

Sample design	Sample fraction (%)	Households	Persons	Weighted	De jure - De facto	Census date (d-m-yr)	Smallest geography	Collective dwellings
Mexico 2005	10	2,546,985	10,284,550	–	de jure	17/10/2005	municipality	no
Mongolia 1989 <sup>(8)</sup>	10	42,783	190,631	yes*	de jure	05/01/1989	province	yes
Mongolia 2000	10	55,795	243,725	–	de jure	05/01/2000	province	yes
Netherlands 1960 <sup>(3)(6)</sup>	1.02	n.a.	143,251	yes	de jure	31/05/1960	none	no
Netherlands 1971 <sup>(3)(6)</sup>	1.02	n.a.	159,203	yes	de jure*	28/02/1971	none	no
Netherlands 2001 <sup>(3)(6)(9)</sup>	1.02	n.a.	189,725	yes	de jure	01/01/2001	none	no
Palestine 1997 <sup>(6)</sup>	10	40,753	259,191	yes	de facto	10/12/1997	governorate	yes
Panama 1960	5	11,869	53,553	–	de facto	11/12/1960	district	yes
Panama 1970	10	31,755	150,473	–	de facto	10/05/1970	district	yes
Panama 1980	10	47,726	195,577	yes	de facto	11/05/1980	district	yes
Panama 1990	10	61,458	232,737	–	de facto	13/05/1990	district	yes
Panama 2000	10	84,346	284,081	–	de facto	14/05/2000	district	yes
Philippines 1990	10	1,156,126	6,013,913	yes	de jure	01/05/1990	municipality	yes
Philippines 1995	10	1,362,190	6,864,758	–	de jure	01/09/1995	municipality	yes
Philippines 2000	10	1,511,890	7,417,810	yes	de jure	01/05/2000	municipality	yes
Portugal 1981	5	179,409	492,289	–	de jure	16/03/1981	NUTS3 region*	yes
Portugal 1991	5	214,155	491,755	–	de jure	15/04/1991	NUTS3 region*	yes
Portugal 2001	5	258,843	517,026	–	de jure	12/03/2001	NUTS3 region*	yes
Romania 1977 <sup>(10)</sup>	10	619,904	1,937,021	–	de jure	05/01/1977	county	no
Romania 1992	10	728,846	2,238,578	–	de jure	07/01/1992	county	no
Romania 2002	10	732,016	2,137,967	–	de jure	18/03/2002	county	no
Rwanda 1991	10	153,041	742,918	–	de facto	01/08/1991	province	yes
Rwanda 2002	10	191,719	843,392	–	de facto	15/16-08-02	province	yes
Slovenia 2002 <sup>(3)</sup>	10	63,637	179,632	–	de facto	31/03/2002	region	no
South Africa 1996 <sup>(11)</sup>	10	993,801	3,621,164	yes	de facto	10/10/1996	district	yes
South Africa 2001	10	991,543	3,725,655	yes	de facto	10/10/2001	municipality	yes
South Africa 2007	2	345,17	1,047,657	yes	de facto	07/02/2007	municipality	yes
Spain 1981 <sup>(3)</sup>	5	n.a.	2,084,221	yes	de facto	01/03/1981	municipality	yes

Sample design	Sample fraction (%)	Households	Persons	Weighted	De jure - De facto	Census date (d-m-yr)	Smallest geography	Collective dwellings
Spain 1991	5	592,276	1,931,458	yes	de facto	01/03/1991	municipality	yes
Spain 2001	5	714,473	2,039,274	-	de jure	01/11/2001	municipality	yes
Uganda 1991	10	339,166	1,548,460	yes	de facto	11/01/1991	county	yes
Uganda 2002	10	529,271	2,497,449	-	de facto	12/09/2002	county	yes
United Kingdom 1991 <sup>(12)</sup>	1	215,761	541,894	-	de jure	21/04/1991	SARs region	yes
United Kingdom 2001 <sup>(3)(6)</sup>	3	n.a.	1,843,525	-	de jure	29/04/2001	SARs region	yes
United States 1960	1	579,212	1,799,888	-	de jure	01/04/1960	state	yes
United States 1970	1	744,475	2,029,666	-	de jure	01/04/1970	250,000+ population	yes
United States 1980	5	4,711,341	11,343,120	-	de jure	01/04/1980	100,000+ population	yes
United States 1990	5	5,527,406	12,501,046	yes	de jure	01/04/1990	100,000+ population	yes
United States 2000	5	6,184,438	14,081,466	yes	de jure	01/04/2000	100,000+ population	yes
United States 2005	1	1,245,246	2,878,380	yes	de jure	01/07/2005	100,000+ population	no
Venezuela 1971	10	284,336	1,158,527	yes	de jure	11/02/1971	municipality	no
Venezuela 1981	10	323,321	1,441,266	-	de jure	21/10/1981	municipality	yes
Venezuela 1990	10	468,808	1,803,953	yes	de jure	21/10/1990	municipality	yes
Venezuela 2001	10	646,08	2,306,489	-	de jure	01/10/2001	municipality	yes
Vietnam 1989	5	534,223	2,626,985	yes	de jure	01/04/1989	province	yes
Vietnam 1999	3	534,139	2,368,167	yes	de jure	01/04/1999	province	no
TOTAL		77,469,676	279,464,944					

NOTES:

- (1) Missing data for several key variables requires use of alternative weight variable
- (2) Excludes 11 states, including entire north region -- one sixth of national population
- (3) Persons not organized into households
- (4) Employment survey
- (5) Excludes three governorates in Kurdish region
- (6) Age grouped into categories
- (7) Excludes the two Borneo states: Sabah and Sarawak
- (8) Highly clustered sample design
- (9) 'virtual census' compiled from registration data
- (10) Excludes two counties and parts of three others -- 7% of national population
- (11) 19 districts in Eastern Cape not organized into households -- 1.3% of sample
- (12) Excludes Northern Ireland
- \* See sample design for more information.

## BIBLIOGRAFIA

- **“Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2009.”**
- LIVI BACCI M., “Introduzione alla demografia”, LOESCHER EDITORE, Torino, 1981
- COLOMBO B., MAFFENINI W., ROSSI F., Dinamica demografica differenziale in Alto Adige, “Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali”, vol.IV, n.1-2, 1981, pp.59-80.
- ROSSI F., MEGGIOLARO S., “Da Nord Est a Nord Ovest. Gli emigrati veneti in Italia nel XX secolo”, Materiali di Demografia Storica, CLEUP Editore, Padova, 2006.
- ROSSI F., CALOVI C., Applicazione del metodo own-children a piccoli aggregati: problemi e risultati, Università di Padova, Dipartimento di Scienze Statistiche, “Rapporti Tecnici”, n.1-87, Padova, 1987, pp.123.
- CLERICI R., Fecondità e migrazioni: un’analisi delle donne censite nel 1981 in Provincia di Milano, pg.258-280 del volume “Studi Emigrazione”, anno XXV, giugno 1988, n.90.
- <https://international.ipums.org/international/>
- <http://www.istat.it/>
- <http://usa.ipums.org/usa/chapter5/chapter5.shtml>