

**Energie Sisma Emilia.
Guida all'utilizzo dei database
dell'indagine 'Energie_I400'**

*Riccardo Righi, Francesco Pagliacci e
Margherita Russo*

CAPPaper n. 148
dicembre 2016



Università di Modena e Reggio
Emilia Facoltà di Economia
Marco Biagi



Università di Bologna
Dipartimento di Scienze
Economiche

CAPP - Centro di Analisi delle Politiche Pubbliche
Dipartimento di Economia Politica - Università di Modena e Reggio Emilia
Ufficio 54 - Ala Ovest
Viale Berengario, 51 41100 Modena - ITALY
phone: +39 059 2056854 fax: +39 059 2056947
email capp@unimo.it

**Energie Sisma Emilia.
Guida all'utilizzo dei database dell'indagine 'Energie_I400'**

Rev. 23/12/2016

Riccardo Righi*, Francesco Pagliacci^ e Margherita Russo°

Dipartimento di Economia Marco Biagi, Università di Modena e Reggio Emilia
e Centro di Analisi delle Politiche Pubbliche – CAPP Unimore

Progetto di ricerca applicata *Energie Sisma Emilia* | www.energie.unimore.it

- * riccardo.righi@unimore.it
- ^ francesco.pagliacci@unimore.it
- ° margherita.russo@unimore.it

Abstract

After the series of earthquakes that occurred in Emilia-Romagna (Italy) in May 2012, the research team based in the Department of Economics of the University of Modena and Reggio Emilia started research activities to investigate the socio-economic impact due to the shock. When analysing the effects of the earthquake on the living and working conditions of the population, a critical issue is the availability of relevant data. Thus, within the applied research ‘Energie Sisma Emilia’, in 2015 the research team realized a survey on a sample of 400 families living in the area affected by the earthquake. The survey was based on the same questionnaire adopted in the previous research ICESmo3 (carried on in Summer 2012) investigating the socio-economic conditions of families living in the same area. The 2015 questionnaire includes additional sections on the effects of the earthquake on social capital and consumption behaviour. Data are under analysis and will be presented in public events and publications. The dataset will be made available on request at the end of the research project.

The goal of this paper is to return the main methodological aspects, and to present the principles that guided management of data and that drive the use of databases. Section I describes the main issues addressed by the ‘Energie Sisma Emilia’ project. Section II introduces the general survey settings; the sample characteristics; and the elaboration of sample weights. Section III describes the details of the informational basis (organization of specific databases, codebook and command files developed to generate databases in format Stata13). Section IV presents the criteria that align the databases and the questionnaire: variables management was originally handled to allow researchers to easily retrieve information from the syntax of the variable labels, and to automatize future elaborations. Section V concludes with operational indications in using ‘Stata Survey’ to produce inferential analysis.

Keywords: earthquake, survey, post-stratification, Stata Survey

JEL Codes: C83

Acknowledgments

This paper has been realized in the research project Energie Sisma Emilia (<http://www.energie.unimore.it>) of the University of Modena and Reggio Emilia (Italy). The project is co-funded by Fondazione Cassa di Risparmio di Modena, Azienda Unità Sanitaria Locale di Modena e Università di Modena e Reggio Emilia.

Sommario

I - Introduzione: Energie Sisma Emilia

II - Aspetti metodologici

Piano di campionamento e attribuzione pesi base

Procedura di estrazione del campione (R&I)

Ricalcolo pesi con doppio processo di post-stratificazione

III - La base informativa: database, codebook e file di gestione dei dati

Database in formato Stata 13 e cartelle di riferimento

Database in altri formati

Codebook

File .do per la gestione dei dati

IV - La gestione delle variabili

Label delle variabili

Nomi delle variabili

Formati delle variabili

V - L'analisi dei dati con Stata Survey

Indicazioni operative

Bibliografia

Appendice - Lista delle variabili

I - Introduzione: Energie Sisma Emilia

È noto che shock esogeni come i terremoti producono criticità, ma al tempo stesso aprono importanti opportunità che dipendono da un complesso di fattori: decisioni strategiche, risorse pubbliche disponibili, fase del ciclo che accompagna la ricostruzione, capacità di coordinamento dei principali attori, capacità di fornire risposte efficaci ai rischi di infiltrazioni della criminalità organizzata, qualità del tessuto economico e sociale su cui l'evento si innesta (Russo e Silvestri 2016). Questo insieme di questioni sono oggetto di analisi del progetto Energie Sisma Emilia, avviato nel settembre 2014 da un vasto gruppo di ricerca dell'Università di Modena e Reggio Emilia. Il progetto di ricerca – cofinanziato dalla Fondazione Cassa di Risparmio di Modena, dalla ASL di Modena e dall'Università di Modena e Reggio Emilia - trae origine dal “Laboratorio sul sisma”¹ che era attivo sin dal giugno 2012 e ha coinvolto un gruppo di docenti del Dipartimento di Economia di Modena, oltre che ricercatori e professionisti esterni alla ricerca accademica e a studenti del dipartimento.

L'idea di fondo del progetto «Energie Sisma Emilia» è che il terremoto abbia prodotto una serie di fratture: non solo nelle strutture fisiche, ma anche nel sistema economico e sociale, di cui non è facile prevedere la ricomposizione (Russo e Silvestri 2016). Organizzazioni economiche, amministrazioni pubbliche, famiglie vanno in cerca di soluzioni e di risposte a obiettivi che non necessariamente sono gli stessi di prima del terremoto. Quegli obiettivi sono enormemente differenziati a seconda della natura dei soggetti (che si collocano a diversi livelli decisionali e di potere e hanno differenti tipi di interazioni interne ed esterne al sistema locale) e della prospettiva temporale con cui si possono affrontare. Il progetto Energie Sisma Emilia ha una duplice finalità. La prima finalità è contribuire all'analisi degli effetti che il terremoto ha prodotto sull'assetto economico e sociale della regione: tale analisi, che potrà favorire la maturazione di decisioni strategiche informate, sarà la base per costruire un modello di monitoraggio della ricostruzione destinato all'amministrazione pubblica regionale e locale. La seconda finalità è condividere quelle conoscenze con gli stakeholder, con la comunità scientifica e con tutti coloro i quali hanno contribuito alla realizzazione del progetto di ricerca.

La fase iniziale del progetto è stata dedicata: (a) alla rassegna della letteratura su studi di caso e modelli di analisi degli effetti di eventi sismici; (b) all'esplorazione e analisi sistematica dei dati disponibili su scala comunale (come ad esempio le statistiche ufficiali in campo sociodemografico e produttivo rilevate da ISTAT); (c) all'analisi della normativa generata per far fronte all'emergenza e alla ricostruzione (oltre 350 ordinanze, decreti e leggi) (Pavone e Russo 2015, Pavone *et al.* 2016); (d) all'analisi dei dati sui contributi richiesti per abitazioni, imprese, opere pubbliche e beni culturali (Ranuzzini *et al.* 2015, Reverberi e Russo 2015); (e) all'analisi del processo di ricostruzione del patrimonio edilizio (urbano e rurale) danneggiato (Giovannetti *et al.* 2015) e possibili effetti di gentrificazione (Cardullo e Russo 2015); (f) risposta del sistema agro-alimentare (Bertolini e Pagliacci 2015); (g) all'analisi delle caratteristiche socio-economiche dei comuni del cratere del sisma per individuare sotto-insiemi di comuni colpiti dal sisma con caratteristiche simili (Piazzi *et al.* 2015); (h) alla costruzione di un modello per identificare un campione di controllo nell'analisi degli effetti del sisma (Pagliacci e Russo 2015); (i) all'analisi dell'azione dell'amministrazione regionale e della struttura commissariale; (l) all'analisi dell'impatto sulle attività commerciali al dettaglio della bassa emiliana (Martinelli *et al.* 2015); (m) all'analisi delle modalità di finanziamento della ricostruzione e gli operatori finanziari coinvolti nel sisma dell'Emilia Romagna (Gualandri *et al.* 2015). La mole di dati e informazioni qualitative raccolte è davvero notevole.

In questo quadro, è stata inoltre predisposta la rilevazione di dati su di un campione di 400 famiglie residenti in quattro Comuni della Provincia di Modena: Cavezzo, Mirandola, Novi di Modena e San Felice sul Panaro (Energie_I400). Questa rilevazione², dedicata ad approfondire le condizioni socio

¹ Il blog ‘ricostruiremeglio’ (<http://www.ricostruiremeglio.wordpress.com>) raccoglie la documentazione prodotta nei due anni di attività, realizzata anche grazie anche al contributo di laureandi (Germana Ducato, Michela Passerini, Alfredo Pucci) e dello stage realizzato su quel progetto da Riccardo Rimondi.

² Parallelamente alla rilevazione è stato anche sviluppato un *field experiment* volto a comprendere gli effetti del sisma sulle scelte economiche individuali e, in particolare, sulla propensione individuale e sociale alla

economiche delle famiglie che vivono nel cratere del sisma, è stata sviluppata adattando il questionario già utilizzato nella indagine ICESmo³. La rilevazione, avviata con la collaborazione dei sindaci dei Comuni coinvolti, si è conclusa a Maggio 2016.

All'interno del progetto Energie Sisma Emilia, le 400 interviste condotte con altrettante famiglie estratte casualmente dalle anagrafi dei comuni di Cavezzo, Mirandola, Novi di Modena, San Felice sul Panaro sono state realizzate in base allo schema di rilevazione della indagine campionaria ICESmo sulle condizioni di vita e i redditi nella provincia di Modena, effettuata nel 2001, 2005 e 2012 (in parte prima, e in parte dopo il sisma).

Nel questionario Energie Sisma Emilia sono perciò incluse domande mirate a ricostruire - a livello personale e familiare - l'impatto del sisma sull'occupazione e il reddito, sullo stile di vita, sulle condizioni di salute e benessere e sulle relazioni sociali e personali.

Il questionario di Energie Sisma Emilia si articola in 17 sezioni:

- A. Informazioni generali (compilate dall'intervistatore)
- B. Notizie relative ai componenti della famiglia
- C. Esperienza scolastica
- D. Condizione professionale
- E. Redditi e lavoro
- E. Agevolazioni e indennizzi sisma
- F. Abitazione principale
- FF. Abitazione principale al 19 maggio 2012
- G. Altri beni immobili
- H. Ricchezza familiare
- N. Disagi legati ai servizi, giudizi sul comune, fiducia
- N. Aspettative centri storici
- N. Bisogni legati al sisma
- O. Commercio
- Q. Salute
- Y. Modello SF12 (autocompilato)
- K. Sisma e relazioni familiari e sociali (autocompilato)
- Z. Autocompilato Soggettivo (autocompilato)

La versione definitiva del questionario "ICESmo - Energie Sisma Emilia – 2015" è disponibile on line nel sito web del progetto Energie <http://www.energie.unimore.it/il-gruppo-di-ricerca/strumenti3>.

I dati sono in fase di elaborazione e saranno presentati in futuri eventi pubblici e pubblicazioni. Il dataset sarà reso disponibile su richiesta al termine del progetto di ricerca.

L'obiettivo di questo lavoro è quello di presentare gli strumenti di base per l'utilizzo dei dati raccolti attraverso l'analisi empirica. Il paper presenta la metodologia d'indagine e gli specifici strumenti statistici usati per la gestione dei dati raccolti attraverso le interviste. La sezione II presenta le impostazioni generali dell'indagine; le caratteristiche del campione intervistato e i pesi utilizzati nell'elaborazione. La sezione III descrive con un maggiore grado di dettaglio l'intera base informativa (organizzazione dei singoli database, codebook e file di comando per la gestione automatica dei database, in formato Stata13). La sezione IV presenta i criteri che consentono l'allineamento tra il questionario e i database: la gestione delle variabili è stata sviluppata per consentire una semplice interrogazione dei database attraverso la creazione di un'apposita sintassi delle etichette delle variabili, e per poter automatizzare future elaborazioni. La sezione V si conclude con alcune indicazioni, di carattere operativo, circa l'uso di 'Stata Survey' per realizzare analisi di statistica inferenziale.

cooperazione. Nel corso della rilevazione è stata richiesta la disponibilità degli intervistati a partecipare a tale esperimento, che poi è stato condotto in via indipendente dalla rilevazione del questionario.

³ Il gruppo di ricerca ha accesso ai singoli documenti anche nelle cartelle di lavoro su Dropbox.

II - Aspetti metodologici

In questa sezione sono descritti gli aspetti metodologici riguardanti la predisposizione del piano di campionamento, le fasi e le procedure utilizzate per la rilevazione, e infine le procedure utilizzate per la costruzione del vettore di pesi da utilizzare nelle analisi dei dati.

Piano di campionamento e attribuzione pesi base⁴

La **numerosità campionaria** complessiva è di **400 famiglie**, così distribuite nei quattro comuni oggetto di indagine: Cavezzo 53 famiglie, Mirandola 187 famiglie, Novi 77 famiglie e San Felice 83 famiglie. L'allocazione proporzionale prevede un peso circa uguale per tutte le unità, salvo arrotondamenti, nel caso di raggiungimento della dimensione obiettivo del campione (variabile 'pesobase').

In questo caso la dimensione obiettivo è stata conseguita e, quindi, ogni famiglia del campione (nei quattro comuni) rappresenta rispettivamente: Cavezzo: 54,1887 famiglie; Mirandola: 54,4545 famiglie; Novi di Modena: 54,4675 famiglie; San Felice sul Panaro: 54,1687 famiglie. Le 400 famiglie intervistate sono rappresentative delle 21.745 famiglie residenti nei quattro comuni.

Sulla base di queste numerosità obiettivo è stato chiesto agli uffici anagrafe dei quattro Comuni selezionati di effettuare l'estrazione di un campione casuale con tre famiglie di riserva per ogni famiglia prima estratta.

Procedura di estrazione del campione (R&I)

Le **anagrafi comunali** hanno estratto le **liste delle famiglie** secondo i seguenti criteri: ordinamento delle famiglie in base all'ordine alfabetico della via di residenza (e numero civico e numero interno); individuazione della prima famiglia campione sulla base di un numero casuale estratto; individuazione delle successive famiglie campione sulla base di un determinato passo di estrazione. Le riserve sono state individuate scegliendo tra le famiglie che nell'ordinamento alfabetico di via occupavano le posizioni successive a quella della famiglia campione.

Gli obiettivi campionari individuati per ogni comune erano obiettivi complessivi, non prevedevano al loro interno una pre-determinata distribuzione fra famiglie italiane e famiglie straniere (ovvero con capofamiglia con "cittadinanza straniera"). La composizione dei campioni estratti fra famiglie italiane e famiglie straniere è stata frutto dell'estrazione casuale.

Comuni	Totale famiglie da intervistare	Di cui: famiglie straniere	
		N.	%
Cavezzo	53	3	5,7
Mirandola	187	21	11,2
Novi	77	5	6,5
San Felice	83	9	10,8
Totale	400	38	9,5

È utile ricordare che le **riserve di una famiglia campione**, italiana o straniera, potevano essere **indifferentemente rappresentate da famiglie italiane o straniere**. Ciò significa che a fronte di una famiglia campione straniera, le riserve uno, due o tre potevano essere sia italiane sia straniere, in quanto l'ordinamento alfabetico dipendeva dal nome della via di residenza (e numero civico e numero interno).

Il campione intervistato si compone complessivamente di 21 famiglie con capofamiglia con cittadinanza straniera, pari al 5,3% del totale. Le famiglie straniere intervistate sono inferiori a quelle

⁴ Questa sezione si basa sulla nota ad uso interno, realizzata da Michele Lalla, che ha disegnato il piano di campionamento delle Indagini ICESmo e dell'indagine I-40.

casualmente estratte fra le famiglie campione (pari al 9,5%), in quanto **il tasso di irreperibilità delle famiglie straniere** è stato **molto elevato**. Il 55,3% delle famiglie straniere cercate, il doppio rispetto alla media del campione, non è stato trovato presso la residenza indicata dall'anagrafe.

Numerose famiglie straniere a seguito del sisma si sono trasferite altrove, spesso in altri comuni, non registrando l'avvenuto cambiamento. Il problema della irreperibilità delle famiglie straniere ha, quindi, condizionato il risultato, mentre il tasso di rifiuto è stato inferiore alla media.

TAB. 1 - FAMIGLIE INTERVISTATE, INDISPONIBILI E IRREPERIBILI PER COMUNE – 2016

Comuni	Famiglie campione	Interviste realizzate	Copertura del campione	Rifiuti	Irreperibili	Numerosità liste	Tasso di rifiuto (rifiuti /famiglie sentite)	Tasso di irreperibilità (irreperibili/ famiglie cercate)
	(A)	(B)	(B/A)	(C)	(D)	(E)	(C)/(B+C)	(D)/(B+C+D)
			%				%	%
Cavezzo	53	53	100,0	55	35	212	50,9	24,5
Mirandola	187	187	100,0	143	118	748	43,3	26,3
Novi *	77	77	100,0	45	36	241	36,9	22,8
San Felice	83	83	100,0	72	41	332	46,5	20,9
Totale	400	400	100,0	315	230	1.533	44,1	24,3

*a causa di duplicazioni presenti negli elenchi estratti dall'anagrafe di Novi, i nominativi a disposizione sono scesi da 308 a 241.

Di cui: Famiglie straniere

Comuni	Famiglie campione	Interviste realizzate	Copertura del campione	Rifiuti	Irreperibili	Numerosità liste	Tasso di rifiuto (rifiuti /famiglie sentite)	Tasso di irreperibilità (irreperibili/ famiglie cercate)
	(A)	(B)	(B/A)	(C)	(D)	(E)	(C)/(B+C)	(D)/(B+C+D)
			%				%	%
Cavezzo	3	0	0,0	2	9	19	100,0	81,8
Mirandola	21	13	61,9	2	27	84	13,3	64,3
Novi	5	2	40,0	1	3	21	33,3	50,0
San Felice	9	6	66,7	8	3	45	57,1	17,6
Totale	38	21	55,3	13	42	169	38,2	55,3

Ricalcolo pesi con doppio processo di post-stratificazione

Il peso base (come calcolato da piano di campionamento e rappresentato dalla variabile 'pesobase') è calcolato in modo tale che, se applicato al numero di famiglie intervistate, restituisce il reale numero di famiglie presenti nella popolazione oggetto di indagine (21.745 famiglie) presenti nei 4 Comuni. La stessa variabile 'pesobase' però, se applicata agli individui intervistati, restituisce un numero sensibilmente superiore a quello della popolazione (+4.000 individui, circa). Il fenomeno si spiega con la maggiore reperibilità e/o disponibilità a rispondere al questionario da parte di famiglie mediamente più numerose della norma. Per ovviare al problema della sovrastima della popolazione di individui, si è scelto di calcolare un secondo vettore di pesi, rappresentato dalla variabile 'pesoI400', utilizzando una **post-stratificazione a due passaggi**. Per eseguire tale procedura è stato

utilizzato il comando ‘Stata reweight’, utilizzando la funzione di distanza A (Pacífico 2014). La procedura si è sviluppata nei seguenti due passaggi.

- a) Il primo passaggio ha visto una post-stratificazione della variabile ‘pesobase’, sulla base dei **totali noti** (della popolazione) del **numero di famiglie per numerosità dei componenti**. I totali della popolazione utilizzati sono stati: (i) 6.620 famiglie con 1 solo componente (ii) 6.405 famiglie con 2 componenti (iii) 4.286 famiglie con 3 componenti (iv) 4.423 famiglie con 4 componenti o più⁵ (Fonte: Anagrafe comunale - Servizio Attività Produttive e Statistica della Provincia di Modena). Tale operazione di post-stratificazione ha permesso di ottenere un vettore di pesi familiari in cui il peso delle famiglie più numerose è ridimensionato e il peso delle famiglie meno numerose è accresciuto (rispetto al vettore di pesi iniziale). Il vettore di pesi creato con questo passaggio consente di pesare in modo corretto il numero di individui presenti nella popolazione (52.356 individui). La discrepanza di 4.000 individui che erano sovrastimati con il peso base iniziale, è stata perciò annullata. Allo stesso tempo, il riporto alla popolazione del numero di famiglie rimane corretto: il vettore di pesi elaborato continua a stimare, in totale, le 21.745 famiglie della popolazione di riferimento⁶.
- b) Nel secondo passaggio, i pesi ottenuti dopo la prima post-stratificazione sono stati ulteriormente post-stratificati. Grazie alla conoscenza dei corrispondenti totali noti nella popolazione è stato possibile elaborare 16 variabili che riportano i totali degli individui che, per famiglia, presentano queste **combinazioni di cittadinanza, genere ed età**:
- italiani, femmine, 0-10 anni (totale noto nella popolazione: 1.964 individui)
 - italiani, femmine, 11-18 anni (totale noto nella popolazione: 1.599 individui)
 - italiani, femmine, 19-34 anni (totale noto nella popolazione: 3.221 individui)
 - italiani, femmine, 34-49 anni (totale noto nella popolazione: 5.112 individui)
 - italiani, femmine, 50-64 anni (totale noto nella popolazione: 4.761 individui)
 - italiani, femmine, 65-74 anni (totale noto nella popolazione: 2.645 individui)
 - italiani, femmine, 75 anni e più (totale noto nella popolazione: 3.395 individui)
 - italiani, maschi, 0-10 anni (totale noto nella popolazione: 1.859 individui)
 - italiani, maschi, 11-18 anni (totale noto nella popolazione: 1.460 individui)
 - italiani, maschi, 19-34 anni (totale noto nella popolazione: 2.962 individui)
 - italiani, maschi, 34-49 anni (totale noto nella popolazione: 4.937 individui)
 - italiani, maschi, 50-64 anni (totale noto nella popolazione: 4.900 individui)
 - italiani, maschi, 65-74 anni (totale noto nella popolazione: 2.924 individui)
 - italiani, maschi, 75 anni e più (totale noto nella popolazione: 3.818 individui)
 - stranieri, maschi, tutte le età (totale noto nella popolazione: 3.821 individui)
 - stranieri, femmine, tutte le età (totale noto nella popolazione: 3.978 individui)
- (Fonte dei totali noti: <http://demo.istat.it/> - Popolazione residente - anno 2015)

Grazie alla definizione di questi conteggi famigliari e grazie alla conoscenza dei corrispondenti totali noti della popolazione è stato possibile procedere con una seconda procedura di post-stratificazione. Tale **secondo passaggio** ha consentito di **ritarare i pesi famigliari** campionari sulla base di **caratteristiche degli individuali della popolazione**.

Il nuovo e **definitivo peso familiare (variabile ‘pesoI400’)** è un vettore di pesi che consente di riportare il campione alla popolazione in modo maggiormente preciso rispetto alle caratteristiche sopra elencate. Per gli individui stranieri, non è stato possibile procedere con disaggregazioni ulteriori sulla base dell’età (come invece fatto per gli individui italiani), in quanto il numero di stranieri intervistati non consente all’algoritmo utilizzato di arrivare ad una convergenza finale.

⁵ Il totale delle famiglie, stando ai dati della fonte indicata, è di 21.734. Tale valore differisce di 11 unità (in meno) rispetto allo stesso totale di famiglie secondo quanto riportato da Istat. Il dato Istat è utilizzato nel passaggio descritto al punto (b). La discrepanza è stata considerata priva di rilievo ai fini della post-stratificazione dei pesi campionari.

⁶ Fonte: <http://demo.istat.it/>- Popolazione residente - anno 2015

Tutta la procedura di post-stratificazione dei pesi dell'indagine Energie_I400 è contenuta e descritta nel file:

/Interviste_400/Database/Riccardo/do/a00_pesi_I400_2016.10.12.do

In definitiva, la variabile peso da utilizzare è la variabile 'pesoI400'. Importante ricordare, che il peso finale, come il peso base iniziale, è un peso che è uguale per tutti i componenti della stessa famiglia⁷.

III - La base informativa: database, codebook e file di gestione dei dati

In questa sezione sono descritte tutte le informazioni riguardanti la base informativa: dove sono salvati i database, come sono organizzate le cartelle, quali file consentono di visualizzare le statistiche descrittive relative alle variabili contenute nei database, quali sono e come sono strutturati i file di comando che generano i database.

Database in formato Stata 13 e cartelle di riferimento

I database si dividono in **due categorie**: (1) codificati, (2) elaborati.

(1) La prima categoria è quella dei **database 'codificati'**. Questi sono il risultato della prima **codifica/gestione** dei dati inviati rilevati da R&I s.r.l.⁸. Sono **divisi per aree tematiche e dialogano tra di loro grazie alle due variabili 'codfam'** (codice familiare) e **'nip'** (codice dell'individuo all'interno della famiglia). In essi le variabili sono state codificate, ovvero sono state corrette da eventuali incongruenze ed è stato fatto un lavoro di allineamento delle caratteristiche (es: le variabili dummy, che nel questionario non erano sempre registrate allo stesso modo, sono state tutte ricodificate con valori 0/1). Questi database si trovano tutti nella seguente cartella:

/Interviste_400/Database/db_variabili_codificate

Questi database sono:

- *autocompilato_I400.dta*, relativo all'autocompilato (database individuale);
- *capsociale_I400.dta*, relativo alla sezione sul capitale sociale (database familiare);
- *carcom_I400.dta*, relativo alle caratteristiche dei componenti delle famiglie (database individuale);
- *carfam_I400.dta*, relativo alle caratteristiche delle famiglie (database familiare);
- *commercio_I400.dta*, relativo alla sezione sul commercio (database familiare);
- *lavoro_I400.dta*, relativo alla sezione del lavoro (database individuale);
- *rfam_I400.dta*, relativo ai redditi familiari (database familiare);
- *rper_I400.dta*, relativo ai redditi individuali (database individuale).

Nella stessa cartella sono inoltre presenti altri due database che contengono informazioni integrative. Questi database sono:

- *pesi_I400.dta*, che contiene i vettori di pesi realizzati per le analisi;
- *survey_reference_subsets_I400.dta*, che descrive l'identificazione dei insiemi di unità statistiche tenuti a rispondere alle diverse parti del questionario.

(2) Della seconda categoria fanno invece parte i **database 'elaborati'**. Questi sono delle evoluzioni dei precedenti. L'idea sottostante è che **ogni sviluppo delle analisi sia supportato da un database dedicato**, nel quale possono essere inserite variabili provenienti dai diversi database descritti e nel quale possono essere elaborate nuove variabili a seconda delle necessità e delle domande di ricerca.

⁷ Tale scelta è in linea con la Rilevazione sulle Forze di Lavoro dell'ISTAT (ISTAT 2014).

⁸ R&I s.r.l. (Ricerche e Interventi di politica industriale e del lavoro) è il centro di ricerca che ha gestito la fase di rilevazione dei dati.

Questi database al momento sono:

- *capsociale_elab_I400.dta*
- *carcom_elab_I400.dta*
- *rfam_elab_I400.dta*

e si trovano nella seguente cartella */Interviste_400/Database/elaborazioni*.

Database in altri formati

Nelle due cartelle in cui sono stati salvati i database, si trovano anche specifiche sottocartelle con database in altri formati. Per favorire l'elaborazione è stata inserita una sottocartella dal nome “*_db_altri_formati*” che contiene gli stessi database nei seguenti altri formati: excel, spss, stata (versioni precedenti al 13).

Codebook

In ognuna delle cartelle contenenti i database, è stata inserita una sottocartella dal nome “*_CODEBOOK*”, che contiene una descrizione delle variabili. Lo scopo dei codebook è quello di fornire **elaborazioni descrittive** di ogni singola variabile contenuta nel corrispondente database di interesse. Sono stati creati tre strumenti: (1) la lista delle variabili in Excel; (2) il codebook in Word, completo di tutte le variabili; (3) i codebook dei singoli database in formato .smcl, .ps, .pdf.

- (1) Il **file excel riassuntivo** che contiene informazioni base su tutte la variabili elaborate nei diversi database è

/Interviste_400/Database/Variabili_db_I400.xlsx

Le **informazioni sono relative a**: nome della variabile, formato, label, *reference set*⁹, dimensione del *reference set*, numero di osservazioni con valori non missing, numero di osservazioni con missing values, indicazione di variabile registrata ex-novo per l'indagine Energie_I400, (se applicabile) indicazione della variabile corrispondente nella precedente indagine ICESmo3, indicazione della presenza della variabile nei singoli database.

- (2) Nel **codebook complessivo (in formato Word)**, le variabili sono elencate sulla base dell'ordine alfabetico delle *variable labels*¹⁰. Per ogni variabile sono riportate le informazioni descritte al punto successivo. Il file contenente il codebook complessivo è:

/Interviste_400/Database/Codebook_complessivo_I400.docx

In questo codebook, per ogni variabile le informazioni sono ripetute per ciascuno dei db in cui la variabile è inserita (eliminare le ripetizione richiederebbe un lungo intervento manuale). In questo codebook, per ogni elaborazione relativa a variabili categoriali sono omesse (per motivi di spazio) le seguenti righe di output:

```
Key: proportions = cell proportions
     se           = linearized standard errors of cell proportions
     lb           = lower 95% confidence bounds for cell proportions
     ub           = upper 95% confidence bounds for cell proportions
     obs          = number of observations
```

- (3) Nelle cartelle che contengono i database sono anche presenti i **codebook dei singoli database**, creati nel seguente modo:

- a) Tutti i codebook iniziano con una **descrizione dell'intero contenuto del database**. A seguire, dopo l'impostazione per l'analisi inferenziale con utilizzo dei pesi calcolati, le singole variabili sono analizzate nel seguente modo:
- i. calcolo della media e degli intervalli di confidenza, per le variabili con formato 'float';

⁹ Vedere il Punto (3) del paragrafo 'Label delle variabili' della Sezione IV, per la spiegazione del concetto e della definizione dei *reference sets*.

¹⁰ Siccome le *variable labels* iniziano con il codice della domanda di riferimento (per come indicato nel questionario), questo tipo di ordinamento consente di avere una forte corrispondenza con l'ordine delle domande nel questionario.

- ii. calcolo delle proporzioni delle singole categorie, per le variabili con formato 'integer';
- iii. calcolo delle proporzioni delle singole categorie, per le variabili con formato 'string';
- b) Per le variabili categoriali ('integer'), la descrizione delle categorie è sempre riportata per esteso prima delle statistiche descrittive.
- c) Le variabili che caratterizzano l'indagine campionaria ('codfam', 'nip', 'codfam_nip', 'pesobase', 'pesoI400', 'N_famiglie', 'N_individui' e tutti i reference sets creati per ovviare al problema della distinzione tra missing values e osservazioni che non erano tenute a rispondere alla corrispondenti domande nel questionario) non sono state analizzate singolarmente all'interno dei codebook, in quanto, sebbene fondamentali nell'impostazione dell'analisi, è stato ritenuto inutile avere statistiche descrittive delle medesime in quel contesto. Sono perciò riportate all'interno della descrizione iniziale del contenuto dell'intero database, ma di esse non sono state prodotte le relative elaborazioni descrittive.

File .do per la gestione dei dati

I **comandi di gestione** dei database sono contenuti nei **file .do** indicati nella Tavola 1, la quale riporta lo schema gerarchico di esecuzione e la funzione associata ad ogni file.

Questi file .do sono **organizzati** in modo da consentire una gestione automatica e agevole di tutto l'insieme di comandi.

Tutti i file .do sono contenuti nella cartella:

/Interviste_400/Database/Riccardo/do

NOTA BENE: modifiche a singoli file .do possono compromettere il funzionamento dell'intera sequenza.

Tavola 1 Struttura dei file .do

Nome file e struttura di esecuzione	Descrizione del risultato generato dall'esecuzione
00_execute.do	<i>esecuzione dei file .do di livello inferiore</i>
01_programmi_I400.do	<i>esecuzione dei file .do di livello inferiore</i>
a00_pesi_I400_2016.10.12.do	<i>creazione pesi</i>
a01_carcom_I400_2016.06.20.do	<i>creazione db carcom</i>
a01bis_tipofam_I400_2016.07.18.do	<i>altre elaborazioni per carcom</i>
a02_lavoro_I400_2016.06.20.do	<i>creazione db lavoro</i>
a03_redditi_I400_2016.06.20.do	<i>creazione db rper e rfam</i>
a04_autocompilato_I400_2016.07.12.do	<i>creazione db autocompilato</i>
a05_capsociale_I400_2016.07.11.do	<i>creazione db capsociale</i>
a06_commercio_I400_2016.07.12.do	<i>creazione db commercio</i>
a07_carfam_I400_2016.11.03.do	<i>creazione db carfam</i>
02_programmi_elaborated.do	<i>esecuzione dei file .do di livello inferiore</i>
b01_carcom_elab.do	<i>creazione db carcom_elab</i>
b02_rfam_elab.do	<i>creazione db rfam_elab</i>
b03_capsociale_elab.do	<i>creazione db capsociale_elab</i>
c01_referenceSets_creation_I400.do	<i>creazione reference sets</i>
c02_referenceSets_application_I400.do	<i>applicazione dei reference sets ai db</i>
c03_def&app_FiltriDomandeInLabels_I400.do	<i>creazione e applicazione riferimenti per domande filtro</i>
03_codebook_I400.do	<i>creazione dei codebook di tutti i db</i>
04_save_other_I400.do	<i>salvataggio di tutti i db in formati diversi da Stata13</i>

IV - La gestione delle variabili

In questa sezione sono descritti gli elementi riguardanti le variabili: le *variable label* (etichette), i nomi delle variabili e i formati delle variabili. In particolare, la sezione illustra come la **costruzione del sistema di etichette e formati della variabili** risponde a due funzioni: (i) facilitare l'analisi e (ii) automatizzare il più possibile le elaborazioni.

- (i) Per favorire l'analisi dei dati, le *variable label* evidenziano tre elementi essenziali per un corretto utilizzo delle variabili:
- i codici della domanda del questionario a cui corrisponde (all'inizio della label);
 - la variabile che riguarda le unità di rilevazione tenute a rispondere a quella domanda (*reference set*, indicato tra parentesi graffe);
 - la presenza di eventuali domande filtro relative a quella domanda (tra barre verticali).
- Gli altri elementi inseriti nelle *label* delle variabili agevolano la comprensione del contenuto delle variabili stesse in modo da facilitare lo sviluppo delle elaborazioni.
- (ii) Per automatizzare le elaborazioni, le etichette e i formati sono stati pensati e costruiti per categorizzare le variabili in modo da generare elaborazioni attraverso il riferimento ai codici inseriti negli spazi indicati nella *label*, con le modalità descritte nella successiva sezione.

NOTA BENE: nel questionario, il carattere “.” (punto) è utilizzato nella nomenclatura dei codici delle domande. All'interno di tutti i database e di tutti i file di descrizione e/o elaborazione tale carattere è sempre stato sostituito con un “_” (underscore). Inoltre, sempre con riferimento alla nomenclatura dei codici delle variabili, tutte le lettere maiuscole sono state trasformate in minuscole. Es: la variabile ‘E1.2’ diventa ‘e1_2’.

In appendice sono riportate tutte le variabili codificate ed elaborate al 21.12.2016, divise per sezione del questionario e ordinate per *variable label*¹¹, e con informazioni relative a: sezione del questionario, etichetta della variabile (*variable label*), nome della variabile, formato, *reference set*¹² corrispondente, dimensione del *reference set*, numero di osservazioni con valori non missing, numero di osservazioni con missing values, indicazione di variabile registrata *ex novo* per l'indagine Energie_I400, (se applicabile) indicazione della variabile corrispondente nella precedente indagine ICESmo3¹³, indicazione della presenza della variabile nei singoli database.

Nel seguito di questa sezione tutti gli elementi riguardanti la gestione delle variabili sono descritti in modo dettagliato.

Label delle variabili

Le *label* (o etichette) delle variabili sono delle stringhe di testo (con lunghezza massima di 80 caratteri) associate alle corrispondenti variabili. Mentre solitamente l'unica funzione delle *variable labels* è quella di fornire una breve descrizione del contenuto della variabile, in questo lavoro le etichette sono state sviluppate in modo da includere informazioni rilevanti nella fase di elaborazione. È stato predisposto un vero e proprio codice che consente l'individuazione di variabili che necessariamente devono essere richiamate nelle fasi di analisi dei dati, facendo così diventare le *variable labels* degli strumenti di lavoro che vanno al di là dell'accenno al contenuto della variabile.. Di seguito sono illustrati i punti che spiegano l'impostazione di tale strumento.

- (1) All'inizio di ogni *variable label* è indicato il **codice della variabile** (o delle variabili) corrispondente nel questionario.
- (2) Nelle *label* possono comparire, indicate all'interno di **parentesi tonde**, le sigle ‘elab’, ‘dummy’ o ‘ricl’ che rispettivamente indicano:

¹¹ Come descritto nel Punto (1) del paragrafo ‘Label delle variabili’ di questa Sezione, ogni etichetta di variabile inizia con l'indicazione del codice della domanda (del questionario) corrispondente. Ordinare le variabili sulla base dunque dell'etichetta porta dunque ad un ordinamento che corrisponde a quello delle domande nel questionario.

¹² Vedere il Punto (3) del paragrafo ‘Label delle variabili’ di questa Sezione, per la spiegazione del concetto e della definizione dei *reference sets*.

¹³ L'indicazione relativa all'inserimento *ex novo* della variabile all'interno dell'indagine I400 (rispetto all'indagine ICESmo3) e, se applicabile, l'indicazione della variabile corrispondente nella precedente indagine ICESmo3 rispondono alla necessità di porre in evidenza gli elementi di continuità tra le due indagini, in modo da facilitare confronti.

- i) una variabile ELABORATA da altre esistenti;
- ii) una variabile DUMMY, con indicate le condizioni sotto le quali essa assume valore 1;
- iii) una variabile RICLASSIFICATA a partire da un'altra già esistente.

(3) In ogni *variable label* è indicato, all'interno delle **parentesi graffe**, l'insieme di riferimento di ogni variabile, ovvero l'insieme di unità statistiche che avrebbero dovuto rispondere alla domanda.

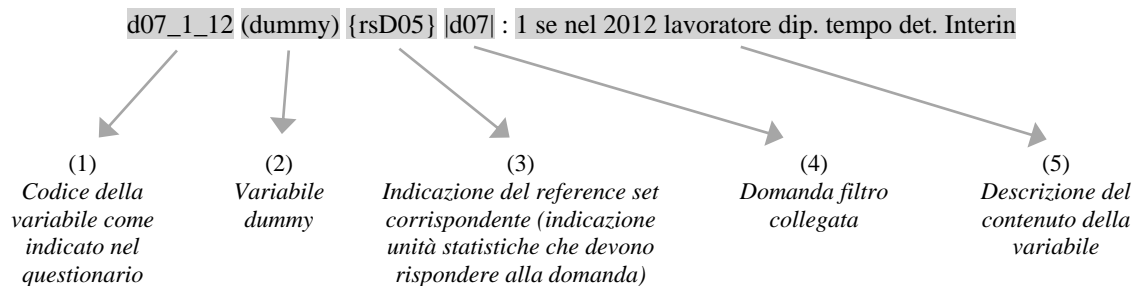
Gli **insiemi di riferimento** sono indicati da variabili col prefisso 'rs' (reference set) e ciascuno di essi è rappresentato da una variabile binaria che assume valore 1 quando le osservazioni appartengono ad esso. Le variabili che rappresentano gli insiemi di riferimento sono presenti in ogni database e, dunque, sono anche descritte all'inizio di ogni codebook. La creazione di queste variabili è stata necessaria per ovviare al problema della distinzione, per ogni variabile, tra missing values e unità statistiche che non erano tenute a rispondere alla relativa domanda del questionario.

NOTA BENE: solo per alcune variabili legate ai redditi familiari, sono indicati 2 *reference sets*. Tutte queste variabili hanno come insieme di riferimento l'insieme delle famiglie intervistate ('rsFAM'). La seconda indicazione ricorda che la variabile potrà avere valore maggiore di zero solo se la famiglia ha individui che rientrano nel secondo insieme di riferimento indicato (vedere Esempio #2).

In alcune *variable label* è indicata, compresa **tra due barre verticali**, la variabile (o le variabili) che nel questionario hanno la funzione di **filtro per la variabile esaminata**.

(5) In ogni variabile, dopo il simbolo di punteggiatura ":", è riportata la descrizione del contenuto della variabile. Solo nel caso delle variabili "Y_AGG" (vedere successivo punto 8), il simbolo di punteggiatura ":" compare due volte all'interno della *label*. Per queste variabili, la descrizione del contenuto inizia dopo il secondo ":".

Esempio #1: Label della variabile 'tipolavdiptdet12_interin'



(6) In alcune *variable label* sono indicate, all'interno di **parentesi quadre o a seguito della variabile principale**, altre variabili (oltre a quella principale) che concorrono a determinare la variabile in oggetto, o indicazioni riguardanti l'elaborazione della variabile stessa.

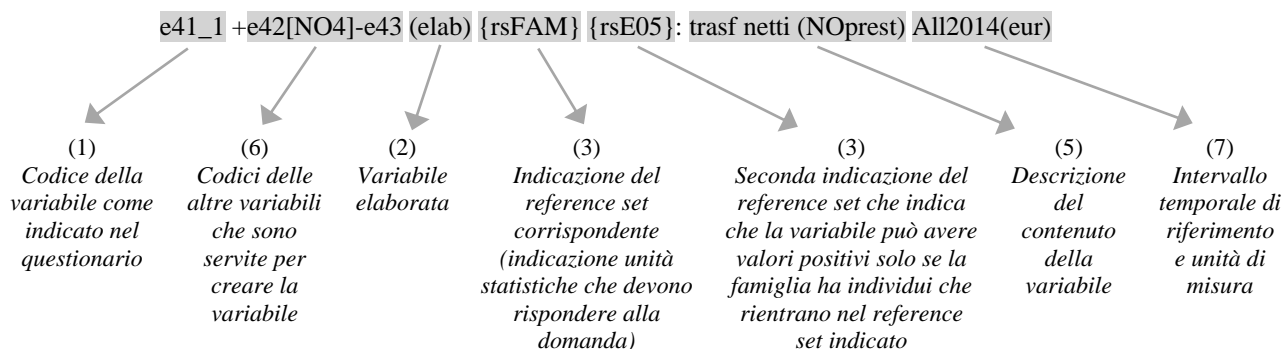
(7) Alla fine delle *variable label*, dove necessario, è stata indicata **l'unità di misura della variabile**. Per variabili che indicano frequenze, non è stata indicata alcuna unità di misura.

Le unità di misura che compaiono sono state abbreviate nel seguente modo:

- eur = euro
- Mgl.eur = migliaia di euro
- mq = metri quadri

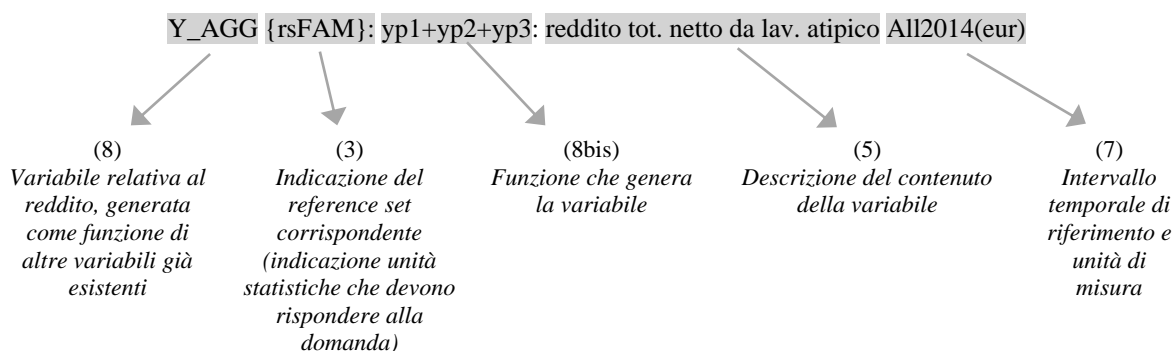
Oltre alle unità di misura, dove necessario, alla fine della *variable label* è descritto **l'intervallo temporale, o il momento nel tempo**, al quale la variabile fa riferimento (es: la sigla 'All2014' indica che la variabile fa riferimento a tutto il 2014. Vedere Esempio #2).

Esempio #2: Label della variabile 'ytrasf_fam'



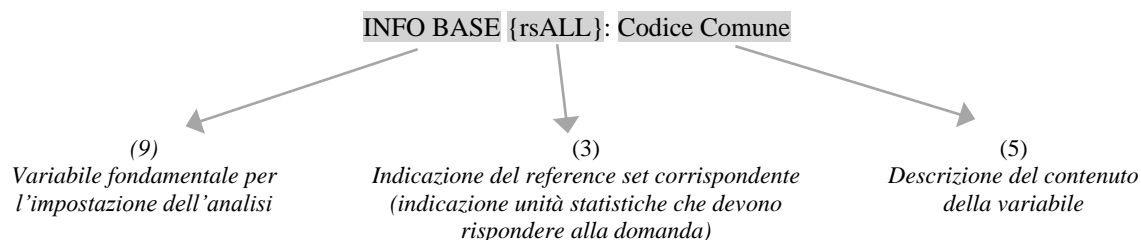
- (8) Ci sono variabili nella cui *label* è presente l'indicazione 'Y_AGG'. Tali variabili corrispondono a degli aggregati ed elaborazioni di variabili relative al reddito. Dove possibile la funzione che determina tale aggregato è stata descritta nella *label* a seguito dei due punti (8bis). In questi casi poi sono stati posti ulteriori due punti che descrivono a parole il contenuto della variabile.

Esempio #3: Label della variabile 'yp_fam'



- (9) Ci sono variabili nella cui *label* è presente l'indicazione 'INFO BASE'. Le variabili a cui è stato assegnata tale indicazione sono quelle che caratterizzano la struttura del questionario: codfam, nip, codfam_nip, comune, pesobase, pesoI400, N_famiglie, N_individui. Queste variabili, fondamentali nell'impostazione dell'analisi, non sono però state descritte nei codebook (con la sola eccezione della variabile 'comune') in quanto è stato ritenuto inutile avere statistiche descrittive delle medesime in quel contesto.

Esempio #4: Label della variabile 'comune'



Nomi delle variabili

- 1) I nomi delle variabili sono stati assegnati tenendo conto di come erano state elaborate le variabili nelle **precedenti indagini ICESmo**, che tenevano conto della corrispondente nomenclatura usata nella indagine della Banca d'Italia sui bilanci delle famiglie italiane. Dove non sono state trovate indicazioni, sono stati assegnati nomi che tengono conto del contenuto/significato della variabile.

- 2) Alcune nomi di variabili possono avere uno dei seguenti tre **suffissi**:
- ‘**_per**’ fa riferimento alla parola ‘PERsonale’: è utilizzato nelle variabili relative al percepimento di determinati redditi personale (per distinguerle da quelle relative ai redditi famigliari);
 - ‘**_fam**’ fa riferimento alla parola ‘FAMigliare’: è utilizzato nelle variabili relative al percepimento di determinati redditi famigliari (per distinguerle da quelle relative ai redditi personali);
 - ‘**_ricl**’: fa riferimento alla parola ‘RICLassificazione’: è utilizzato in caso di variabili che sono successive riclassificazioni di variabili già esistenti;
 - ‘**#**’: in alcuni casi invece è stato preferito inserire come suffisso un numero che ha un qualche significato rispetto alla riclassificazione operata (es: ‘cleta6’ corrisponde ad una riclassificazione in 6 classi dell’età degli individui, ‘cleta9’ corrisponde ad una riclassificazione in 9 classi dell’età degli individui).

Formati delle variabili

Le variabili sono state rielaborate assegnando ad ognuna di esse uno specifico formato sulla base della natura della variabile stessa. Questo permette di agevolare la creazione di comandi generalizzati attraverso lo sviluppo di routine adeguate per le specifiche tipologie di variabili (per mezzo del formato delle stesse). I formati utilizzati sono i seguenti:

DOUBLE è utilizzato per le variabili fondamentali che caratterizzano l'indagine campionaria (vedere paragrafo *Label delle variabili* al punto 9, relativo alle variabili identificate dalla sigla “INFO BASE” nella etichetta): ‘codfam’, ‘nip’, ‘codfam_nip’, ‘pesobase’, ‘pesoI400’, ‘N_famiglie’, ‘N_individui’ e tutti i reference sets creati per ovviare al problema della distinzione tra missing values e osservazioni che non erano tenute a rispondere alla corrispondenti domande nel questionario. Le variabili alle quali è stato assegnato il formato “double”, sebbene fondamentali nell’impostazione dell’analisi, non sono state descritte nei codebook in quanto è stato ritenuto inutile avere statistiche descrittive delle medesime in quel contesto.

LONG è utilizzato unicamente per la variabile ‘comune’, che è la variabile che identifica i Comuni relativi al campo di osservazione dell’indagine campionaria. Per motivi legati a limiti di Stata relativi ad alcuni formati, si è dovuto utilizzare tale formato per tale variabile. A differenza delle variabili con formato “double”, la variabile ‘comune’ è stata riportata e descritta nei codebook.

FLOAT è utilizzato per le variabili numeriche discrete/continue (es: reddito equivalente familiare, superficie dell’abitazione di residenza, numero di mensilità percepite nel 2014, etc.).

INTEGER è utilizzato per le variabili categoriali (es: condizione professionale, esito della verifica di agibilità dell’immobile di residenza, soggetto a cui la famiglia si è rivolta per prima per un aiuto legato a bisogni di prima necessità nella fase di emergenza dovuta al sisma, etc.). Caratteristiche di queste variabili:

- i) ai valori di queste variabili è sempre stata assegnata (salvo poche eccezioni) una etichetta che descrive la categoria corrispondente (es: al valore 1 della variabile ‘condpro’ è assegnata la categoria ‘1. Operaio’, al valore 4 della variabile ‘studio’ è assegnata la categoria ‘4. Licenza Media’);
- ii) le variabili dummy sono state tutte elaborate in formato INTEGER. In alcune di esse sono state descritte le categorie assegnate ai valori corrispondenti. Nella maggior parte di esse non è stata inserita alcuna descrizione dei valori in quanto già nella label è descritto a cosa è associato il valore 1 (es: “d04_12 (dummy){rsD01}: 1 se nel 2012 lavorava part time”).
- iii) per quanto riguarda i valori delle variabili dummy: nella maggior parte dei casi il questionario indicava il valore 1 per il ‘SI’ e il valore 2 per il ‘NO’. In fase di gestione dei database queste sono state tutte modificate e ricodificate con valori 0/1 (rispettivamente ‘NO’ e ‘SI’);
- iv) qualora le domande del questionario fossero simili a dummy, ma con anche una sola ulteriore categoria (es: SI, NO, NON SA), le corrispondenti variabili sono state trattate come le altre variabili categoriali e si è intervenuto il meno possibile nella ricodifica dei valori.

- v) nelle etichette che descrivono le categorie assegnate ai corrispondenti valori delle variabili (value labels), è sempre indicato all'inizio (prima di un punto) il numero di tale categoria (es: al valore 2 della variabile 'tipocasa' è assegnata l'etichetta "2. casa plurifamigliare")

STRING è utilizzato per le variabili categoriali che sono state registrate in formato testo. Tali variabili sono per la maggior parte variabili nelle quali sono state registrate ulteriori specificazioni a domande che nel questionario le precedevano. Tali variabili, per essere utilizzate devono essere riclassificate, rielaborando le informazioni tratte dalla loro forma testuale.

V - L'analisi dei dati con Stata Survey

In questa sezione sono fornite le indicazioni di base per produrre analisi inferenziali tramite l'utilizzo di Stata Survey. Al fine di produrre **analisi di tipo inferenziale**, Stata consente di utilizzare una serie di comandi che sono tutti compresi all'interno del pacchetto '**Stata Survey**', già presente all'interno di Stata.

L'utilizzo di tale pacchetto è semplice: richiede solo di indicare al software che i dati caricati sono dati campionari. Da quel momento in poi, se si producono elaborazioni tramite i comandi del pacchetto 'Survey', **Stata in automatico produce analisi inferenziali** tenendo sempre le medesime impostazioni (fino a nuove modifiche imposte dall'utente) relative alle caratteristiche del campione.

I comandi del pacchetto 'Survey' hanno una loro sintassi specifica. Il **manuale può essere scaricato online alla pagina web** <http://www.stata.com/manuals13/svy.pdf>. In generale: mentre la prima parte dell'analisi dei dati, ovvero quella relativa all'impostazione del campione, è da effettuare tramite il comando `svyset`, le successive elaborazioni sono introdotte dal prefisso `svy`.

Di seguito, sono indicati più nel dettaglio i passaggi fondamentali per una corretta elaborazione delle analisi dei database dell'indagine Energie_I400.

Indicazioni operative

- 1) **Impostare la lettura del database** sulla base del tipo di analisi da effettuare.
 - a) Se l'analisi riguarda le famiglie:
 - i) assicurarsi che le osservazioni siano relative alle 400 famiglie usando il comando:

```
capture keep if nip==1
```
 - ii) scrivere poi la seguente riga di comando:

```
svyset [pw=pesoI400], psu(codfam) strata(comune) fpc(N_famiglie)
```
 - b) Se l'analisi riguarda gli individui:
 - i) assicurarsi che le osservazioni siano relative ai 1031 individui dopo aver lanciato il comando:

```
sum nip
```
 - ii) scrivere poi la seguente riga di comando:

```
svyset [pw=pesoI400], psu(codfam_nip) strata(comune) fpc(N_individui)
```
- 2) Il comando `svyset` (sopra indicato) tiene conto di:
 - a) **pw** (Probability Weight): sono i pesi assegnati alle unità statistiche di riferimento. La nota metodologica relativa alla creazione del vettore di pesi è rappresentata dal file
/Interviste_400/Database/Riccardo/do/a00_pesi_I400_2016.10.12.do.
Qui basta ricordare quanto segue: il vettore di pesi da utilizzare è sempre lo stesso (dato dalla variabile 'pesoI400'), sia in caso di analisi delle famiglie che di analisi degli individui. Ad ogni individuo è stato infatti assegnato lo stesso peso assegnato alla sua famiglia.
 - b) **strata** (gli strati di campionamento): i 4 Comuni in cui si è svolta l'indagine (variabile 'comune');

- c) **psu** (Primary Sample Units): sono le unità statistiche di riferimento nell'analisi, vale a dire le famiglie o gli individui.
A seconda del tipo di fenomeno da indagare, deve essere indicata quale è la variabile che identifica univocamente le unità statistiche appropriate. La variabile 'codfam' identifica in modo univoco le famiglie, mentre la variabile 'codfam_nip' (data dalla combinazione delle variabili 'codfam' e 'nip') identifica in modo univoco gli individui intervistati;
- d) **fpc** (Finite Population Correction): il totale di famiglie, o di individui, presenti (totali noti) nella popolazione. Questi totali sono sempre presenti nei database sotto forma di variabili che si chiamano 'N_famiglie' (che nella popolazione sono 21.745) e 'N_individui' (che nella popolazione sono 52.356).
- 3) Per ogni variabile da analizzare prestare attenzione a:
- a) il **reference set** corrispondente (insieme di individui che erano tenuti a rispondere alla domanda), indicato tra parentesi graffe nella *variable label*. Vedere punto 3 del paragrafo 'Label delle variabili';
 - b) la **presenza di domande filtro** che riducono ulteriormente l'insieme di coloro tenuti a rispondere alla domanda, è indicato tra barre verticali nella *variable label*. Vedere punto 4 del paragrafo 'Label delle variabili'.
- 4) Prima di porre condizioni relative al restringimento delle osservazioni da analizzare mediante apposizione di condizioni, **verificare la corrispondenza tra le categorie di interesse e i valori delle variabili categoriali**. Comando:
- ```
groups variabile_name, show(none) clean
```

**NOTA BENE:** vista l'operazione di ricodifica, i valori assegnati nel questionario alle categorie delle variabili possono differire dai valori assegnati nei database alle medesime categorie.

## Bibliografia

- Bertolini P., Pagliacci P., 2015, *Le specificità del sistema agro-alimentare nella ricostruzione post-sisma*, Working Paper DEMB Series, n. 68
- Cardullo P., Russo M., 2015, *Earthquake in the city: the people yet to come*, Working Paper DEMB Series, n. 63
- Giovannetti E., Pagliacci F., Pergetti S., 2015, *Il processo della ricostruzione dell'abitare in Emilia*, Working Paper DEMB Series, n. 67
- Gualandri E., Pedrazzoli A., Vezzani P., 2015, *Finanziare la ricostruzione: attività e attori coinvolti nel sisma dell'Emilia Romagna*, Working Paper DEMB Series, n. 70
- ISTAT, 2014, *Rilevazione sulle Forze di Lavoro: Aspetti metodologici dell'indagine*
- Martinelli E., Tagliazucchi G., Marchi G., 2015, *Disastri Naturali e Dynamic Capabilities nel Commercio*, Working Paper DEMB Series, n. 66
- Pacifico D., 2014, *Reweight: a Stata Module to reweight Survey Data to External Totals*, Working Paper del Dipartimento del Tesoro del Ministero dell'Economia e della Finanza della Repubblica Italiana, n. 5
- Pagliacci F. e Russo M., 2015, *Socio-economic effects of an earthquake: does sub-regional counterfactual sampling matter in estimates? An empirical test on the 2012 Emilia-Romagna earthquake*, Working Paper DEMB Series, n. 82
- Pavone P. e Russo M., 2015, *Analisi lessico testuale delle ordinanze del Commissario Delegato alla ricostruzione in Emilia-Romagna: un contributo alla legge nazionale su emergenza e ricostruzione*, Working Paper DEMB Series, n. 72
- Pavone P., Righi R., Righi S., Russo M., 2016, *Text mining and network analysis to support improvements in legislative action. The case of the earthquake in Emilia-Romagna*, Lexicometrica, Proceeding of JADT2016 Conference, pp. 237-247
- Piazzini P., Pagliacci P., Russo M., 2015, *Analisi cluster delle caratteristiche socio-economiche dei comuni dell'Emilia-Romagna: un confronto tra comuni dentro e fuori dal cratere del sisma*, Working Paper DEMB Series, n. 61
- Ranuzzini M., Pagliacci P., Russo M., 2015, *L'informatizzazione delle procedure per la ricostruzione: prime evidenze dai contributi concessi per le abitazioni*, Working Paper DEMB Series, n. 71
- Reverberi M., Russo M., 2015, *I contributi alle imprese colpite dal sisma del 2012 in Emilia-Romagna: una base informativa per l'analisi e il monitoraggio della ricostruzione*, Working Paper DEMB Series, n. 69
- Russo M. e Silvestri P., 2016, *Innovation and development after the earthquake in Emilia*, Working Paper DEMB Series, n. 81

## **Appendice**

### **Lista delle variabili**























| Sezione | Questionario | Variable Label                                                             | Variable Name | Format   | Reference Set | Reference Set Size | Observations | Missing Values                        | Match con ICESm3       |
|---------|--------------|----------------------------------------------------------------------------|---------------|----------|---------------|--------------------|--------------|---------------------------------------|------------------------|
|         |              |                                                                            |               |          |               |                    |              | solo Energie I400<br>(X se SOLO I400) |                        |
|         |              | rsALL: TUTTI INDIVIDUI                                                     | rsALL         | Double - |               | 0                  | 400          |                                       | autocompliat_1400      |
|         |              | rsD01: INDIVIDUI OCCUPATI (magg15anni)                                     | rsD01         | Double - |               | 0                  | 1031         | 0                                     | x capsociale_elab_1400 |
|         |              | rsD02: INDIVIDUI DISOCCUPATI (magg15anni)                                  | rsD02         | Double - |               | 0                  | 1031         | 0                                     | x capsociale_1400      |
|         |              | rsD03: INDIVIDUI DISOCCUPATI o IN CERCA PRIMA OCC. (magg15anni)            | rsD03         | Double - |               | 0                  | 1031         | 0                                     | x carcom_elab_1400     |
|         |              | rsD04: INDIVIDUI TUTTI ESCLUSO PENS.& BENEST.& INABILI (magg15anni)        | rsD04         | Double - |               | 0                  | 1031         | 0                                     | x carcom_1400          |
|         |              | rsD05: INDIVIDUI LAVORATORI DIPENDENTI (magg15anni)                        | rsD05         | Double - |               | 0                  | 1031         | 0                                     | x cartram_1400         |
|         |              | rsE01: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. DIPENDENTE (magg15anni)   | rsE01         | Double - |               | 0                  | 1031         | 0                                     | x commercio_1400       |
|         |              | rsE01: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. DIPENDENTE (magg15anni)   | rsE01         | Double - |               | 0                  | 400          | 0                                     | x lavoro_1400          |
|         |              | rsE02: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. ATIPICO (magg15anni)      | rsE02         | Double - |               | 0                  | 1031         | 0                                     | x rlam_elab_1400       |
|         |              | rsE02: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. ATIPICO (magg15anni)      | rsE02         | Double - |               | 0                  | 400          | 0                                     | x rper_1400            |
|         |              | rsE03: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. INDIPENDENTE (magg15anni) | rsE03         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsE03: INDIVIDUI CHE HANNO AVUTO REDDITI DA LAV. INDIPENDENTE (magg15anni) | rsE03         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsE04: INDIVIDUI CHE HANNO AVUTO REDDITI DA PENSIONE (magg15anni)          | rsE04         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsE04: INDIVIDUI CHE HANNO AVUTO REDDITI DA PENSIONE (magg15anni)          | rsE04         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsE05: INDIVIDUI CHE HANNO AVUTO REDDITI DA TRASF. E ALTRO (magg15anni)    | rsE05         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsE05: INDIVIDUI CHE HANNO AVUTO REDDITI DA TRASF. E ALTRO (magg15anni)    | rsE05         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsE07: INDIVIDUI SOCI/GEST. SOC. (magg15anni)                              | rsE07         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsE07: INDIVIDUI SOCI/GEST. SOC. (magg15anni)                              | rsE07         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF01: FAM. CON ABIT. DI RES. IN AFFITTO                                   | rsF01         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF02: FAM. CON ABIT. DI RES. IN PROPR./USO_GRAT./USUFR./A_RISC/ETC        | rsF02         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsF02: FAM. CON ABIT. DI RES. IN PROPR./USO_GRAT./USUFR./A_RISC/ETC        | rsF02         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF03: FAM. CON ABIT. DI RES. DI PROPRIETA'                                | rsF03         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF04: FAM. CON ABIT. DI RES. PRE-SISMA DIVERSA DA ATTUALE                 | rsF04         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF05: FAM. CON ABIT. DI RES. PRE-SISMA DIVERSA DA ATTUALE E IN AFFITTO    | rsF05         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsF06: FAM. CON ABIT. DI RES. PRE-SISMA DIVERSA DA ATTUALE E NO-AFFITTO    | rsF06         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsFAM: TUTTE LE FAMIGLIE                                                   | rsFAM         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsFAM: TUTTE LE FAMIGLIE                                                   | rsFAM         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsG01: FAM. CON BENI IMM DIVERSI DA ABIT. PRINC AL 19mag2012               | rsG01         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsG01: FAM. CON BENI IMM DIVERSI DA ABIT. PRINC AL 19mag2012               | rsG01         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsG02: FAM. CON BENI IMM DIVERSI DA ABIT. PRINC ACQ. POST 19mag2012        | rsG02         | Double - |               | 0                  | 400          | 0                                     |                        |
|         |              | rsM15: INDIVIDUI MAGGIORI 15 ANNI                                          | rsM15         | Double - |               | 0                  | 1031         | 0                                     |                        |
|         |              | rsM15: INDIVIDUI MAGGIORI 15 ANNI                                          | rsM15         | Double - |               | 0                  | 400          | 0                                     |                        |