

Strategia di campionamento per l'indagine sulla valutazione dei corsi di formazione

di Michele Lalla* e Giuseppe Fiorani**

1. Introduzione¹

La valutazione dell'impatto della formazione professionale finanziata dal Fondo Sociale Europeo è un tema su cui da anni —anche in Italia— sono impegnati numerosi studiosi, operatori del settore, e dirigenti pubblici. Dati gli obiettivi della programmazione europea per l'occupazione (occupabilità, adattabilità, equità, pari opportunità), la gran parte della discussione si è incentrata sulle diverse dimensioni della valutazione, al fine di stabilire come giungere alla formulazione di un giudizio, capace di rispondere alle domande poste dai programmatori. Tale discussione ha consentito di raggiungere un buon livello di condivisione circa la definizione degli ambiti metodologici, secondo lo scopo della valutazione —sia di rendicontazione o di valutazione di impatto (Martini, Cais, 1999) — e secondo il profilo temporale delle fasi di programmazione (opportunità dell'intervento, confronto tra risultati e obiettivi, monitoraggio del processo, impatto rispetto al contesto, molteplicità degli attori coinvolti nella programmazione).

Qualora ci si muova in ambito di valutazione di impatto, emergono le peculiarità del «prodotto» formazione professionale, il cui risultato non è esprimibile con misure fisiche, ma solo attraverso la rilevazione di cambiamenti di stato occupazionale o professionale delle persone che hanno usufruito del servizio (Gori, Vittadini, 1999). Tali cambiamenti di stato sono influenzati congiuntamente da numerosi fattori individuali (anagrafici, motivazionali, abilità soggettive), dalla qualità del processo formativo, nonché da concomitanti movimenti della domanda e dell'offerta nei mercati del lavoro locali. Al fine di approssimare un giudizio ragionevolmente corretto sull'impatto del programma, occorre isolare l'effetto «formazione professionale» dagli altri, in modo da tentare di comprendere cosa sarebbe successo rispetto a una situazione di mancanza di intervento. Ciò è approssimabile solo attraverso l'adozione di disegni di ricerca «quasi-sperimentali» e di modelli di analisi multivariata (Brucchi Luchino, 2001).

Quando si passa dal piano metodologico a quello empirico, il numero di contributi italiani si restringe assai (Berliri, Pappalardo, 2001). L'Isfol, tra l'altro, effettua questo tipo di analisi a livello nazionale, e coordina le esperienze regionali in appositi gruppi di lavoro, per uniformare le modalità di raccolta dei dati. Nell'ambito di uno di questi gruppi di lavoro sono maturate le considerazioni che qui vengono presentate.

Il lavoro si sofferma su un aspetto particolare di questa discussione e, se si vuole, abbastanza laterale nel processo di analisi valutativa: l'importanza della strategia di campionamento per la valutazione di efficacia dei programmi di formazione professionale. In sintesi, si sostiene che l'impatto della formazione, la cui misurazione è già difficile di per sé, può diventare inafferrabile se non si riduce a monte l'eterogeneità

* Professore straordinario di Statistica sociale presso il Dipartimento di Economia Politica dell'Università di Modena e Reggio Emilia.

** Professore a contratto presso la Facoltà di Economia dell'Università di Modena e Reggio Emilia e ricercatore Poleis.

dell'unità di analisi. Tale eterogeneità può dipendere, infatti, oltre che dai fattori oggettivi e soggettivi accennati, anche da differenze nei criteri amministrativi di gestione delle liste dei formati. Per esempio, corsi appartenenti a una stessa misura del FSE possono riguardare gruppi di utenti molto differenti in termini di occupabilità, e tecnologie corsuali completamente differenti. Non vi è modo di risalire alle spiegazioni dell'eterogeneità dei risultati, se la strategia di campionamento non si sforza di raggruppare per classi omogenee gli utenti. In sostanza, si sottolinea che l'analisi di esito occupazionale si colloca all'interno dell'organizzazione di un *Sistema Informativo* della formazione, nel quale coesistono dati provenienti da fonti diverse e con diversi criteri: programmatori, soggetti gestori, agenzie formative, società di ricerca. Tali dati, necessari sia alle funzioni di monitoraggio gestionale sia a quelle di analisi di efficacia, non sono integrabili in modo meccanico. Tale problema, già sollevato da altri contributi (Bulgarelli, Gori, 2001), è destinato a ampliarsi in prospettiva, a seguito del maggior ruolo delle agenzie formative nella valutazione degli esiti, previsto dalle disposizioni ministeriali e regionali sull'accreditamento.

Il piano di campionamento descrive il processo di selezione di un sottoinsieme, detto *campione*, dell'universo o popolazione obiettivo o popolazione di riferimento (\mathcal{P}) e il complesso di decisioni prese nella sua realizzazione. In genere, comprende la definizione delle unità statistiche costituenti \mathcal{P} , la determinazione dell'ammontare di unità statistiche da includere nel campione, la sua struttura, la stratificazione e gli stadi o i livelli in cui si è operata la selezione, le regole adottate per eseguirla. Il campione è costituito da un numero fissato di unità statistiche (campionarie), alle quali somministrare l'intervista o il questionario. Il processo di realizzazione di un campione si collega intrinsecamente anche al processo di raccolta dei dati e di elaborazione. Nell'espletamento dei primi due processi si incontrano numerose difficoltà, specie quando le unità statistiche sono entità complesse, come le imprese che non costituiscono, però, l'oggetto della presente indagine.

Definite le unità statistiche, e quindi \mathcal{P} , il passo successivo consiste nella costruzione della lista o base campionaria, che non è facile da realizzare e presenta già gli ostacoli più difficili. Le unità selezionate dalla lista, poi, nella fase della rilevazione possono risultare: *non appartenenti* alla popolazione obiettivo; *uscite* (o morte) perché hanno perso la condizione di appartenenza o sono emigrate; *non rintracciabili* perché potrebbero avere traslocate in altra sede, avere registrato erroneamente le loro generalità negli archivi. I dati amministrativi possono contenere numerosi errori di imputazione (indirizzi, comune, data di nascita), che comportano un elevato tasso di mancate risposte (Martini, 1990). Le modifiche che intervengono nel tempo possono alterare le caratteristiche della popolazione di riferimento e, pertanto, anche i mutamenti rilevati tramite il campione danno una indicazione dei cambiamenti avvenuti in \mathcal{P} , ma influenzano alcuni aspetti del campionamento (come le probabilità di selezione o il calcolo dei pesi) e possono distorcere le stime (Duncan e Kalton, 1987; Kalton e Citro, 1993). In aggiunta a questi problemi strutturali, si hanno anche le *mancate collaborazioni* all'indagine che possono essere: *parziali*, quando i soggetti intervistati rispondono solo a alcune parti o domande del questionario; o *totali*, quando i soggetti si rifiutano di rispondere o di partecipare all'indagine. L'evoluzione della popolazione \mathcal{P} e le risposte mancanti incidono nella determinazione delle stime dei suoi parametri e, in particolare, nella precisione delle stime e nel calcolo dei pesi per riportare i valori campionari a quelli relativi a \mathcal{P} .

L'obiettivo del lavoro è presentare alcuni aspetti teorici e pratici inerenti a questi problemi nell'indagine sulla «*Valutazione dei corsi di formazione al/sul lavoro in Emilia-Romagna*». L'universo di riferimento per la procedura di campionamento era, quindi, limitato: nello *spazio*, alla regione Emilia-Romagna (E-R); nel *contenuto*, agli allievi dei corsi di formazione promossi, direttamente o indirettamente, dalla regione; nel *tempo*, a un triennio di attività programmate. Le delimitazioni dell'indagine derivano, in definitiva, dagli obiettivi del Programma Operativo Regionale (POR); in particolare, dall'Obiettivo 3 del Fondo Sociale Europeo (FSE) per gli anni 2000-2003. Nel seguito si espongono le caratteristiche del piano di campionamento per conseguire tali obiettivi; in particolare: nel paragrafo 2 si descrive sinteticamente \varnothing , l'entità degli errori campionari, e i caratteri che devono definire l'appartenenza delle unità a \varnothing ; nel paragrafo 3 si illustra la strategia di campionamento da adottare, e come determinare le precisioni effettive risultanti dopo la rilevazione; nel paragrafo 4 si riportano alcuni aspetti della determinazione dei pesi in presenza delle difficoltà di rilevazione empiriche; nel paragrafo 5 si accenna a una possibile normalizzazione all'unità dei pesi, per «incorporare» il piano di campionamento nel processo di stima senza alterare la numerosità campionaria; nel paragrafo 6 si delineano alcuni aspetti essenziali degli errori non campionari; nel paragrafo 7 si riportano i primi risultati relativi al campione per l'indagine condotta sui corsi terminati nell'anno 2001; infine, nel paragrafo 8 segue una breve conclusione.

2. Definizione della popolazione di riferimento

La Regione E-R sta predisponendo le linee guida della valutazione di efficacia delle attività formative finanziate dal FSE, Obiettivo 3. La comparabilità dell'attività valutativa è ritenuta un vincolo da rispettare e per conseguirla occorre aderire alle linee guida (coordinamento) dell'ISFOL. Attualmente, l'indicazione primaria di armonizzazione delle indagini consiste nella proposta di adozione di un questionario per la rilevazione del tasso di occupazione dei partecipanti ai corsi di formazione al lavoro a 12 mesi dalla fine del corso. Nulla è detto, sinora, sulla strategia di campionamento da adottare per selezionare i partecipanti ai corsi di formazione. L'omogeneità dello strumento di rilevazione consente di ottenere un risultato ottimale per la comparabilità dei dati, ma non è sufficiente per ottenere anche stime comparabili: occorre uniformare anche il disegno campionario, ai fini della comparabilità delle stime del tasso di occupazione dei partecipanti alla formazione professionale. La strategia di campionamento o la struttura del campione è, quindi, un altro elemento essenziale per avere un periodo di riferimento comune che implica automaticamente la stessa popolazione di riferimento, \varnothing ; diversamente, si avranno comparazioni che, non riferendosi allo stesso periodo, comparano oggetti diversi! Pertanto, occorre definire, per tutti, in modo inequivocabile l'universo o \varnothing da cui campionare. Altri aspetti, di non minore importanza, riguardano: la sua stratificazione, la precisione delle stime o errore di campionamento, l'errore di rilevazione accettato. Non occuparsi di questi aspetti rischia di vanificare il tempo e le risorse investite sui questionari.

2.1. Unità statistica e base campionaria

In base agli obiettivi dell'indagine, \mathcal{S} è costituita dai soggetti che hanno frequentato i corsi in esame e completato il ciclo di lezioni. Le informazioni (personali) sui soggetti, che consentono di rintracciare i soggetti inclusi nel campione, e sul tipo di corso seguito sono contenute nel Sistema Informativo della Formazione Professionale (SIFP). L'archivio costituisce, quindi, la base campionaria dell'indagine o lista (*frame or list*), ossia l'elenco delle unità di \mathcal{S} dal quale eseguire la selezione delle unità campionarie.

Per ottenere un buon campionamento bisogna procurarsi una buona lista, cioè una lista esente da difetti perché un campione non può dare risultati più precisi della lista da cui è stato estratto. In realtà, le liste disponibili presentano diversi limiti; in particolare, gli elenchi che si desumono dagli archivi amministrativi non sono sempre esaustivi e i dati contenuti sono controllati accuratamente solo se interessano l'attività dell'ente che li produce (Martini, 1990). I difetti più comuni presenti nelle liste, quindi, anche in quella fornita dal SIFP, sono i seguenti: (a)*incompletezza*, la lista non contiene alcune unità o alcuni sottoinsiemi di \mathcal{S} ; (b)*sopracompletezza*, la lista contiene alcune unità non pertinenti o non appartenenti a \mathcal{S} ; (c)*ridondanza* o registrazione multipla di unità, la lista contiene la stessa unità statistica registrata più volte con piccole variazioni nei codici di identificazione; (d)*inesistenza* o registrazione di unità bianche, la lista contiene i codici che identificano unità fittizie perché in realtà non esistono; (e)*molteplicità* o presenza di grappoli, la lista contiene dei nominativi ai quali, per ognuno di essi, corrispondono più unità statistiche; (f)*inattualità*, la lista contiene dati relativi a unità che si riferiscono a un periodo precedente a quello corrente, in cui si utilizzano le informazioni in essa contenute e si svolge l'indagine; (g)*imprecisioni*, la lista contiene alcune unità non individuate correttamente perché i dati relativi a esse non sono esatti sia perché il dichiarante può essersi sbagliato o aver dichiarato il falso, sia perché possono esserci stati errori di codifica, di trascrizione, di immissione. Si può enfatizzare così l'irreperibilità delle unità statistiche. *Il campione rappresenta*, quindi, *la popolazione della lista*: se la lista non contiene informazioni accurate su \mathcal{S} , anche il campione non fornirà informazioni accurate su \mathcal{S} .

La lista da cui effettuare l'estrazione del campione deve anzitutto includere le *stesse tipologie di attività formative*, al fine di evitare che la stima del tasso di occupazione si riferisca a popolazioni differenti. Ad esempio, la regione E-R esclude l'apprendistato, gli incentivi individuali, l'orientamento, l'obbligo scolastico; e include l'obbligo formativo, la formazione iniziale e superiore. Qualora altre regioni compiano scelte diverse, è decisivo che la stratificazione consenta di risalire sempre alle stesse popolazioni di riferimento. Ciò vale naturalmente anche per le altre fondamentali caratteristiche: il territorio (la provincia), il genere dei partecipanti, e l'intensità della formazione (durata dei corsi, in classi di ore di insegnamento). Se il campionamento fosse effettuato per asse o misura di finanziamento, invece che per tipologia formativa, la comparabilità delle stime non sarebbe possibile.

La lista contiene anche informazioni sugli attori che promuovono o realizzano i corsi. L'obiettivo della valutazione potrebbe (e dovrebbe) riguardare anche chi conduce e organizza in pratica le attività formative. Gli enti e le società che "producono" formazione sono tante e il loro controllo è eseguibile solo per campione. Allora, la strategia di campionamento dovrebbe considerare un metodo che tenga anche conto della presenza di realizzatori; tuttavia, allo stadio attuale si è escluso tale possibilità perché si sarebbe dovuto ricorrere a una forma di campionamento a grappolo che è meno preciso del tradizionale campionamento stratificato (Cochran, 1977); inoltre, con

il metodo del campionamento stratificato su altri caratteri si possono ottenere ugualmente informazioni su un certo numero di società formatrici perché è presumibile che siano estratti un numero di casi sufficienti per accertare alcuni aspetti della loro attività. In ogni caso, si dovrebbe almeno standardizzare il tipo di informazioni da trasmettere al SIFP e controllarne la qualità, al fine di avere una base di dati efficiente anche per il realizzare il campione.

La strategia di campionamento «dipende» dagli obiettivi dell'indagine e dalle tecniche di valutazione che si intendono adottare. Se si vuole costruire una graduatoria (di performance) degli enti formatori o delle tipologie dei corsi per confrontare la loro efficienza nell'inserimento dei formati sul mercato del lavoro, allora occorre rilevare il maggior numero di soggetti utilizzando un questionario molto agile (Martini, 1999). Il limite intrinseco di questa procedura consiste nelle mancate collaborazioni all'indagine, che sono dell'ordine del 30%. In tal caso, i valori dell'indice utilizzato per costruire la graduatoria potrebbero avere un errore che rende difficoltosa la decisione eventuale da prendere; non meno semplice è il caso di un ente che mostra scarsa efficienza (per esempio, bassi tassi di occupazione) perché forma su segmenti del mercato del lavoro in condizioni più disagiate. Se si vuole eseguire una valutazione d'impatto della formazione, cioè valutare il contributo «netto» che la formazione dà alla riuscita occupazionale (Martini, Garibaldi, 1993), occorre programmare le indagini in modo completamente diverso. A tal proposito vi è una vasta letteratura; tra gli altri si vedano: Rettore, Trivellato, e Martini (2002); Shadish, Cook, e Campbell (2002); Heckman (1999). L'impostazione attuale dell'indagine e del questionario si configura come un processo che confluisce in una «analisi descrittiva dell'esperienza dei formati» (Martini, 1999) e, in questo ambito, la strategia di campionamento proposta è coerente con gli obiettivi tipici di un tale studio.

2.2. Precisione delle stime e errore di rilevazione

L'errore campionario è funzione della numerosità campionaria e della precisione attesa della variabile da stimare. Per la formazione al lavoro, la variabile da stimare è essenzialmente il tasso di occupazione; allora, è rilevante il confronto con le stime dell'Istat ottenute con la Rilevazione Trimestrale delle Forze di Lavoro (RTFL) per persone della stessa fascia di età e livello territoriale. La scelta della regione E-R è di ottenere la precisione campionaria a livello di ogni provincia, per evitare che differenze strutturali dei mercati del lavoro locali si sommino o si compensino.

L'errore di rilevazione dipende anche dalla modalità di somministrazione del questionario. Trattandosi di questionario telefonico di tipo CATI, diventano importanti la serietà e affidabilità della società di rilevazione, le procedure di appuntamento, i tentativi di chiamata prima della sostituzione, la percentuale massima di sostituzione ammessa per i rifiuti alla partecipazione all'indagine.

La diffusione della telefonia mobile che spesso sostituisce il telefono fisso e presenta un tasso di «cambiamento» elevato, può creare alcune difficoltà di copertura nelle indagini telefoniche, che in futuro dovrebbe aumentare. Il segmento dei giovani costituisce l'insieme dove la telefonia mobile è più diffusa; quindi, i giovani sono meno raggiungibili e sono proprio i soggetti più interessati all'azione della formazione.

2.3. *Struttura del campione*

La struttura del campione dipende fundamentalmente dagli obiettivi della ricerca e dalle risorse a disposizione. Nel nostro caso, questi aspetti dovrebbero essere grosso modo gli stessi per tutte le regioni che regolano l'Obiettivo 3 del FSE. L'obiettivo fondamentale della ricerca è determinare il grado di successo della formazione al lavoro, attraverso la rilevazione della condizione occupazionale dei partecipanti *a 12 mesi dalla fine del corso*. Tale valore è stato fissato considerando soprattutto la natura del mercato del lavoro, che presenta una media di attesa poco più di anno, ma anche l'efficienza logistica dell'indagine: un periodo inferiore comporterebbe una *acquisizione incompleta* delle informazioni utili agli obiettivi da conseguire, ma anche un'attività continua nell'intervistare i soggetti, un affaticamento superfluo di intervistato e intervistatori perché i mutamenti attesi non si sono ancora verificati. I costi aumentano tanto e per mantenere l'equilibrio finanziario si finirà per aumentare l'errore di campionamento. Naturalmente, presenta anche una controindicazione derivante dagli errori di memoria perché i cambiamenti che si rilevano sono avvenuti nel passato e possono essere distorti o rimossi. Si tratta, in ogni caso, di effettuare campionamenti ripetuti annui, al fine di poter confrontare, in modo lordo, l'esito occupazionale dei formati con l'andamento del mercato del lavoro contestuale, per identici segmenti di popolazione.

Rispetto al tempo o periodo di riferimento dell'indagine, *il problema cruciale è la scelta discriminante tra anno di programmazione o anno di realizzazione dei corsi*.

La scelta dell'*anno di programmazione* del corso, (a), comporta la selezione di individui che hanno frequentato i corsi finanziati nell'anno di riferimento, ad esempio il 2000, indipendentemente dall'anno in cui il corso termina. In questo caso si privilegia l'omogeneità amministrativa.

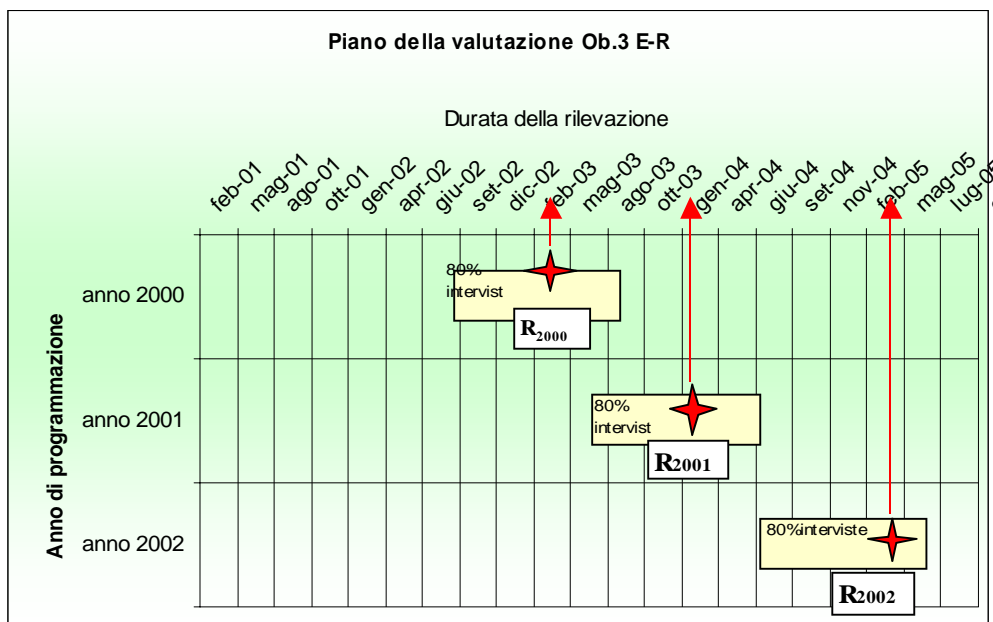
La scelta dell'*anno di realizzazione* del corso, (b), comporta la selezione di individui che hanno frequentato i corsi terminati nell'anno di riferimento, indipendentemente dall'anno in cui sono stati finanziati. In questo caso si privilegia l'omogeneità temporale nel confronto tra attività formativa e rispettivo andamento del mercato del lavoro.

La scelta influenza non solo il piano di campionamento, ma pregiudica anche il conseguimento degli obiettivi dell'indagine perché cambia totalmente la leggibilità delle stime. Nel primo caso (a), il campione esprimerà un tasso di occupazione medio di persone che hanno trovato lavoro in anni differenti; inoltre, i tempi di rilevazione si allungano, in quanto l'attività programmata nel 2000 termina in modo scaglionato, e presumibilmente entro i primi mesi del 2003. La stima finale non è quindi ottenibile prima del 2004, dato che deve passare un anno dalla fine del corso; prima sono possibili stime solo parziali. Nel secondo caso (b), il tasso di occupazione dei formati è riferito allo stesso intervallo temporale di confronto dell'andamento del mercato del lavoro, e questo riduce o elimina l'ambiguità di interpretazione dei risultati. Inoltre, la stima finale —sempre riferibile all'anno di programmazione— è disponibile nei primi mesi del 2002, in tempo utile quindi per fornire al decisore elementi utili sulle eventuali correzioni da adottare. È appena il caso di notare che la situazione si rovescia per le stime riferite alla programmazione finanziaria, perché i corsi terminati in un dato anno di calendario si riferiscono a programmazioni diverse nel tempo.

Con il criterio di campionare in base all'anno di programmazione finanziaria, si allungano i tempi in cui si possono effettuare le interviste, dovendo attendere almeno 12 mesi dalla conclusione del corso. I corsi dell'anno 2000 finiscono per il 12,5% nel 2000 stesso, per il 68% nel 2001, e il restante 20% nel 2002. Il rapporto finale deve essere

consegnato, però, in tempi utili per il programmatore; ci si deve attendere, quindi, che esso sia consegnato a aprile 2003 con stime però più imprecise, in quanto basate sull'80% circa delle interviste. Il punto debole di questo assunto sta soprattutto nella stima "incompleta" del tasso di occupazione dei formati (Figura 1).

Figura 1. Piano di valutazione con campionamento per anno di programmazione finanziaria



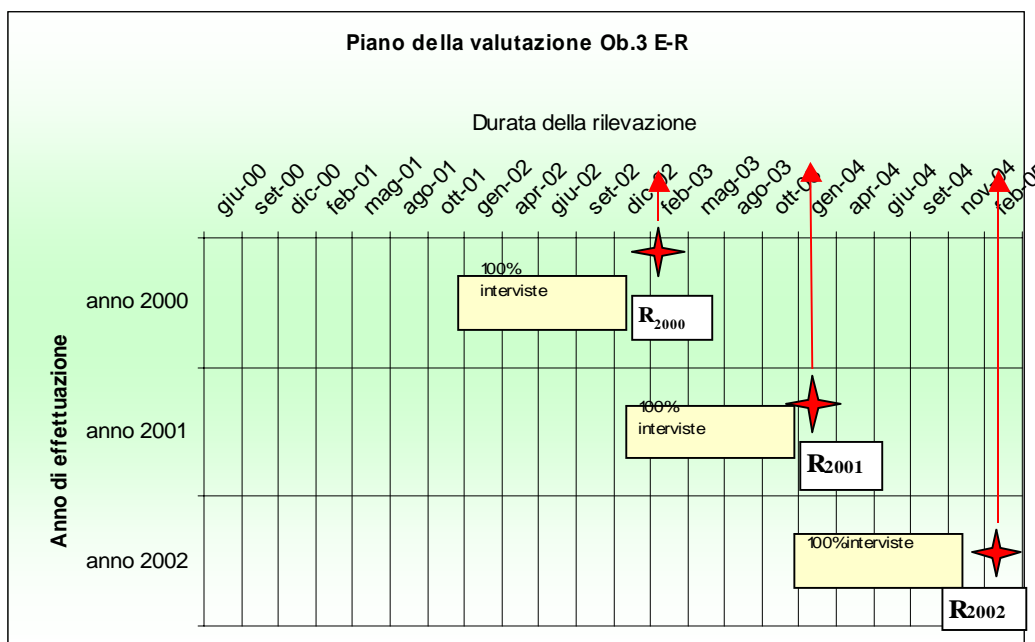
L'attività del 2000 è partita con ritardo, essendo il primo anno della nuova programmazione. Anche prevedendo dal 2001 un andamento a regime più ravvicinato tra programmazione e realizzazione, il ciclo di vita della programmazione finanziaria non cambierà molto per gli anni successivi, a meno di eventi inattesi. Con questo criterio passeranno, quindi, sempre più di due anni tra l'anno di programmazione e quello di consegna del rapporto a esso riferito. Tale rapporto sarà basato, inoltre, su stime parziali: quelle definitive saranno disponibili dopo tre anni. Tali stime rappresentano fedelmente l'attività (programmata in termini finanziari) effettuata per tutti i tipi di durata dei corsi; ma nell'interpretazione delle probabilità di successo, rispetto al mercato del lavoro, queste stime sono la media delle probabilità di successo di tre anni (2000, 2001, 2002). Si pone quindi il problema dell'anno da scegliere per la comparazione.

Con il criterio di campionare in base all'anno di realizzazione dei corsi, non si incorrere nel fenomeno dei mancanti (coloro di cui non si conosce la data di termine del corso). I risultati delle interviste, e quindi le stime delle probabilità di occupazione dei formati, sarebbero disponibili dopo 14 mesi (e non dopo 26). Tali risultati, inoltre, sarebbero completi e non parziali. Infine, le stime si riferirebbero a un anno solo, e sarebbero confrontabili all'andamento del mercato del lavoro territoriale per le stesse categorie di persone (Figura 2).

Naturalmente, il programmatore sarà sempre in grado di distinguere l'anno di programmazione finanziaria dei corsi conclusi nell'anno solare, al fine di utilizzare la valutazione di efficacia esterna (di esito), anche per integrare la valutazione di efficacia interna (di gestione). Le elaborazioni possono essere anche eseguite per anno di programmazione finanziaria; naturalmente, ora, il problema si rovescia: gli esiti della

programmazione complessiva saranno noti dopo diversi anni e le stime relative all'anno, non essendo stato isolato nel contesto del campione, potrebbero presentare forse qualche lieve distorsione.

Figura 2. Piano di valutazione con campionamento per anno di realizzazione dei corsi



Queste osservazioni sulla base dei primi problemi rilevati nella predisposizione dell'attività di valutazione di impatto non intendono essere prescrittive. L'obiettivo è piuttosto quello di sollecitare e raccogliere l'opinione di coloro che, a vario titolo, sono coinvolti nel processo di valutazione (autorità decisionale, valutatori, ricercatori), per renderlo sempre più trasparente, comprensibile, e utile per gli interessati. La scelta non inficia completamente l'impianto complessivo della procedura di campionamento.

3. Campionamento: stratificazione e numerosità campionaria

Il campione si può costruire con il metodo del campionamento stratificato perché è il più idoneo e usuale a indagare la realtà che, a volte, si presenta già suddivisa in domini o aree territoriali predefinite e nel presente caso sono le province (indice *i*): Bologna (BO), Ferrara (FE), Forlì (FO), Modena (MO), Piacenza (PC), Parma (PR), Ravenna (RA), Reggio Emilia (RE), Rimini (RN). Le province promuovono i corsi, secondo gli obiettivi del POR; tuttavia vi sono corsi che non hanno specificamente un referente provinciale, per questi corsi si prevede uno strato (non geografico) denominato «Regione Emilia-Romagna (RER)». Per semplicità, il carattere che individua tali raggruppamenti si indicherà con il termine «provincia», anche se non ha, in verità, un senso geografico; infatti, per la regione E-R tali modalità costituiscono il «piano formativo», ma qui, per semplicità, le sue modalità si denomineranno province.

Le caratteristiche dei corsi attivati sono un aspetto rilevante dell'attività formativa, e gli obiettivi che si propongono di conseguire possono costituire la base per realizzare una loro tipologia. Ai fini del campionamento ci si dovrebbe limitare a un

numero ridotto di modalità (indice j) per contenere la dimensione campionaria che influenza assai l'ammontare dei costi; tuttavia, per aumentare l'accuratezza e la precisione della valutazione dell'attività di formazione è opportuno disporre di modalità molto omogenee; pertanto, si sono individuate nove modalità: Esperienze di Lavoro (EdL o *work experience*), formazione Post Obbligo Formativo e Post Diploma (FPOPD), Istruzione Formazione Tecnica Superiore (IFTS), Formazione Finalizzata Al Reinserimento Lavorativo (FFARL), Formazione per Creazione di Impresa (FCI), Formazione per Occupati (FO), Formazione all'Interno dell'Obbligo Formativo (FIOF), Formazione Post Universitaria (FPU), Formazione Permanente (FP). Il carattere che denota tale raggruppamento si denominerà «tipologia formativa».

Gli strati individuati dall'incrocio della provincia con l'azione formativa sono i domini nei quali si fissa il livello di precisione delle stime che, in base alle informazioni disponibili negli archivi amministrativi, si può riferire a una proporzione generica con un margine di errore del 5% e un livello di "fiducia" del 95%. Si propone la scelta della precisione in termini assoluti perché nel calcolo della dimensione campionaria si considererà l'accuratezza relativa a una percentuale intorno al 50%. Nel caso in oggetto, l'errore relativo sarebbe più indicato per valutare percentuali piccole, ma i risultati della dimensione campionaria sarebbero troppo elevati. La stratificazione per tipologia formativa (e per provincia) si può ulteriormente raffinare secondo due caratteri distintivi delle unità statistiche oggetto di indagine: il genere (indice k), e la durata del corso (indice l). Nella determinazione della dimensione del campione per sub-strato dei domini si è adottata l'allocatione proporzionale che è autoponderante e possiede alcuni vantaggi, consistenti in una varianza ridotta (rispetto al campione casuale semplice), semplicità, e robustezza dei risultati (Kish, 1990, 1992).

3.1. Determinazione dell'ampiezza del campione

L'ampiezza totale del campione si può determinare considerando sia la precisione desiderata delle stime, sia i costi da sostenere per condurre l'indagine, ma nel presente caso si è ignorata la funzione di costo. In base alle informazioni disponibili, la precisione desiderata delle stime si è calcolata considerando una generica proporzione di un determinato carattere. L'ampiezza del campione è data da (Cochran, 1977):

$$n_{ij} = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{d^2} \cdot \frac{1}{1 + \frac{1}{N_{ij}} \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{d^2} - 1 \right)} \quad (1)$$

Il termine n_{ij} indica l'ampiezza risultante dalla precisione desiderata nello strato determinato per l' i -esima provincia e la j -esima tipologia formativa (strato ij , per $i=1, \dots, I=10$; $j=1, \dots, J=9$), π è la proporzione non nota della popolazione obiettivo \mathcal{P} , N_{ij} è il numero totale di unità statistiche nella popolazione, d è l'errore (assoluto) desiderato nella stima della proporzione π , e $z_{\alpha/2}$ è il valore dell'ascissa della distribuzione normale corrispondente al livello di significatività α . La precedente si può scrivere diversamente se si considera l'errore relativo, r , rispetto a π che dà $d=r\pi$ e si sostituisce, ma qui per brevità si tralascia.

L'applicazione della (1) per strato ij , potrebbe comportare una dimensione campionaria finale eccessivamente elevata, per i costi connessi alle interviste. Per gli obiettivi dell'indagine sulla «Valutazione dei corsi di formazione al/sul lavoro in

Emilia-Romagna» è sufficiente anche applicare una procedura “iterativa”: si fissa la precisione delle stime sulla tipologia formativa, che è rilevante, e si controlla che la precisione delle stime per provincia sia dell’ordine del 5%, che si può ritenere accettabile. Nell’indagine condotta nel 2001 si sono fissati gli errori per provincia e per tipologia formativa; per alcune modalità di quest’ultima si è proceduto al censimento: IFTS, FCI, FIOF, FPU.

3.2. Stratificazione per genere e classi di durata dei corsi

Per ottenere un campione che rappresenti, con adeguata numerosità, la popolazione secondo alcuni caratteri considerati fondamentali nell’analisi degli effetti dei corsi frequentati sull’occupazione e sulla posizione professionale, la stratificazione per tipologia formativa (e per provincia) si raffinerà secondo altri due caratteri distintivi delle unità statistiche oggetto di indagine: il genere (indice $k=1, \dots, K=2$) e la durata del corso, data in ore e suddivisa in classi (indice $l=1, \dots, L_j$; dove l’indice j posto alla L denota che le classi di durata dei corsi dipendono dalla tipologia formativa). La stratificazione secondo la durata dei corsi è opportuna perché si presume che sia anche in relazione con gli effetti che si possono manifestare sulle possibilità occupazionali e di avanzamento di carriera o miglioramento della posizione professionale. Un’ipotesi plausibile di suddivisione in classi di ore è: 0–149, 150–499, 500 e oltre, «non terminato»; per le tipologie FO e FP si sono fissate classi di ore pari a: 0–49, 50–149, 150 e oltre, «non terminato». L’ultima classe, «non terminato», si riferisce a quei corsi che nell’anno di riferimento si stanno ancora svolgendo; la definizione di \mathcal{S} è, dunque, essenziale: se \mathcal{S} è costituita dai soggetti che hanno frequentato i corsi programmati in un determinato anno, allora la classe di durata «non terminato» non sarà vuota perché vi possono essere ritardi nell’attivazione o corsi che hanno una durata superiore all’anno; se \mathcal{S} è costituita dai soggetti che hanno frequentato corsi che sono terminati in un determinato anno, allora la classe «non terminato» è vuota. Nella determinazione della dimensione del campione per questi sub-dominî si adotterà l’allocazione proporzionale.

Il numero di soggetti, n_{ijkl} , da intervistare negli strati $ijkl$ (determinati dalla modalità i -esima della provincia, j -esima della tipologia formativa, k -esima del genere, e l -esima della durata del corso), è stato determinato in due fasi: prima, con l’applicazione dell’espressione (1) in ogni strato ij (provincia per tipologia formativa) si è determinata la dimensione campionaria obiettivo, n_{ij} , dello strato ij ; poi, la numerosità ottenuta si è «espansa» negli strati determinati dagli incroci delle altre variabili secondo il metodo dell’allocazione proporzionale che dà una dimensione per strato $ijkl$ pari a

$$n_{ijkl} = n_{ij} \frac{N_{ijkl}}{N_{ji}} \quad (2)$$

dove n_{ij} , N_{ij} , e N_{ijkl} sono le dimensioni del campione a livello provinciale per tipologia formativa (strato ij), della popolazione sempre nello strato ij , e della popolazione nello strato $ijkl$, rispettivamente.

3.3. Errore «effettivo»

Le indagini empiriche presentano numerose difficoltà che comportano, alla fine della raccolta dati, una riduzione della dimensione campionaria desiderata (programmata) per strato; di conseguenza aumenta l’entità dell’errore relativo delle stime. Per valutare l’ordine di grandezza dell’errore campionario, a posteriori, si può ricavare d dalla (1).

L'errore «a posteriori», $d_{p:ij}$, considerando i soggetti partecipanti, $n_{p:ij}$, nello strato ij è pari a

$$d_{p:ij}^2 = \frac{z_{\alpha/2}^2 \pi (1 - \pi)}{n_{p:ij}} \left(\frac{N_{ij} - n_{p:ij}}{N_{ij} - 1} \right), \quad (3)$$

dove N_{ij} indica sempre l'entità della popolazione obiettivo dell' i -esima provincia e j -esima tipologia formativa, π è la proporzione non nota della popolazione obiettivo \mathcal{P} , e $z_{\alpha/2}$ è il valore dell'ascissa della distribuzione normale corrispondente al livello di significatività α .

4. Fattori di espansione

Sia data una popolazione \mathcal{P} di N unità. Sia Y la variabile in oggetto con distribuzione statistica incognita e valori (Y_1, Y_2, \dots, Y_N) . Si voglia stimare il totale della variabile Y , dato da $Y = \sum_{i=1}^N Y_i$ in base al campione osservato (y_1, y_2, \dots, y_n) , con l'eventuale uso di variabili ausiliarie. Nello schema adottato, y_1 indica il valore osservato di Y nell'unità ottenuta dalla prima estrazione, y_2 indica il valore osservato di Y nell'unità ottenuta dalla seconda estrazione, e così via fino all' n -esima estrazione. Gli stimatori che si considerano, in genere, sono lineari del tipo

$$\hat{Y} = \sum_{i=1}^n w_i y_i \quad (4)$$

dove le quantità w_i sono dette *pesi* che non dipendono dal numero d'ordine delle osservazioni, ma possono dipendere dal tipo di campionamento adottato e dall'etichetta che individua l'unità statistica selezionata (Cicchitelli, Herzel, Montanari, 1992). Più in particolare, se si considera un piano di campionamento probabilistico che genera un campione di n unità estratte senza ripetizione (reimmissione), e sono note le probabilità di inclusione del primo ordine delle unità di \mathcal{P} dalla quale si sono estratte le osservazioni; allora, lo stimatore corretto del totale, \hat{Y} , diventa lo stimatore di Horvitz-Thompson (Horvitz, Thompson, 1952). Si tratta, quindi, di uno stimatore ottenuto dalla combinazione lineare delle osservazioni campionarie con pesi pari all'inverso delle probabilità di inclusione, dipendenti dalle etichette delle unità cui si riferiscono le osservazioni, ossia dal piano di campionamento adottato. L'allocatione, che si adotterà nell'indagine sulla «Valutazione dei corsi di formazione al/sul lavoro in Emilia-Romagna», richiederà, presumibilmente, l'uso di pesi diversi per ciascun dominio di studio (o strato), anche se si tratta di un campione autoponderante perché la correzione per le mancate collaborazioni conduce all'uso di pesi diversi per strato.

I pesi riportano le stime dei parametri della popolazione alla data di riferimento della lista. Il loro valore è influenzato da almeno tre fenomeni che coinvolgono le unità di \mathcal{P} e modificano la realtà indagata: la non appartenenza a \mathcal{P} ; l'emigrazione o l'uscita da \mathcal{P} ; la non rintracciabilità, che può includere sia l'emigrazione, sia l'uscita, sia gli errori di registrazione negli archivi. Il trattamento delle unità che rientrano in tali categorie può seguire diverse strategie alternative nel calcolo dei pesi: si ignorano e si trattano come non rispondenti, ma ne consegue una possibile sovrastima dei parametri caratterizzanti \mathcal{P} ; si assume che nella lista vi siano unità estranee a \mathcal{P} . L'entità degli errori può essere trascurabile e la dimensione campionaria per strato piccola; allora la

stima risente della rarità degli eventi e della scarsa numerosità per strato del campione, sicché si rischia di enfatizzare l'effetto nelle stime e ottenere una considerevole sottostima dei valori della popolazione.

La statistica che si ottiene applicando i pesi è detta anche *stimatore per espansione* perché nel caso di un campionamento casuale semplice o autoponderante diventa semplicemente il prodotto della corrispondente grandezza campionaria moltiplicata per l'inverso della frazione di campionamento: $\hat{Y} = (N/n) \cdot y$, dove si sono omessi gli indici ij , per semplicità. La frazione di campionamento è, in totale, n/N ; quindi, si trattano i dati come se ogni unità del campione rappresenti N/n unità della popolazione. Tale fattore, N/n , è detto anche *coefficiente di espansione*. Nel caso in oggetto, all'interno di ogni strato (o dominio di studio) si ha un peso che corrisponde proprio a questa rappresentazione; ma per il calcolo dei pesi finali occorre considerare la probabilità di rintracciare una unità e la probabilità di ottenere la sua partecipazione. Si può mostrare che all'interno di ogni strato il fattore di espansione o peso è uguale al rapporto tra il numero di unità della popolazione nello strato $ijkl$ e il numero di unità partecipanti all'indagine:

$$w_{p;ijkl} = \frac{N_{ijkl}}{n_{p;ijkl}} = \frac{1}{\pi_{p;ijkl}^*} \quad (5)$$

dove, come si è detto, l'indice i denota la provincia ($i=1, \dots, I=10$), l'indice j la tipologia formativa ($j=1, \dots, J=9$), l'indice k denota il genere ($k=1, \dots, K=2$), e l'indice l denota la durata del corso in classi di ore di lezione ($l=1, \dots, L_j$); inoltre, $n_{p,ijkl}$ denota il numero di unità che partecipano all'indagine e rispondono alle domande del questionario; $w_{p,ijkl}$ indica il peso da attribuire all'unità statistica appartenente allo strato $ijkl$; e $1/\pi_{p;ij}^*$ può interpretarsi come una «pseudo-probabilità» di selezione o probabilità di rilevare effettivamente i dati dell'unità statistica perché deriva dalla probabilità di inclusione modificata o corretta per le difficoltà incontrate. Naturalmente, questa è la soluzione più semplice per compensare le stime dalle difficoltà delle indagini e dalle non risposte; altre strategie, più sofisticate e complesse, che non si possono spesso applicare alle indagini su larga scala, si trovano in Little e Rubin (1987) e Rubin (1988). Gli stimatori diventano, così, non lineari e le varianze possono aumentare (Kish, 1990). In genere, poi, le correzioni apportate non correlano con le variabilità negli strati e incrementano generalmente la varianza (Bethlehem, Keller, 1987; Potter, 1990).

5. Normalizzazione dei pesi all'unità

Per eseguire test statistici e/o stimare i parametri di modelli rappresentativi della realtà indagata non si può pesare con $w_{p;ijkl}$ perché altera la numerosità campionaria e, quindi, le probabilità di significatività relative alle ipotesi da sottoporre a verifica. In pratica, quindi, per rimediare a tali inconvenienti è utile «scalare» i pesi in modo che la loro somma sia uguale all'unità, anche se i totali non sono così riportati alla popolazione di riferimento (Verma, 1995). Per incorporare la struttura del campione nella determinazione degli stimatori e non alterare la numerosità campionaria, si può utilizzare un insieme di pesi che, partendo da $w_{p;ijkl}$, mantengano inalterate le

caratteristiche del campione, ossia soddisfacciano due vincoli: (a) la loro somma sia uguale al numero totale degli strati ($IJKL$), (b) il loro prodotto per la numerosità degli strati sia uguale al numero totale di interviste n . Si può utilizzare un peso dato dal rapporto tra i pesi «originari», $1/\pi_{p;ijkl}^*$, e un peso medio, $1/\bar{\pi}_{p;ijkl}^*$, in modo da soddisfare entrambe le condizioni (a) e (b). Le grandezze figurano al denominatore, sicché si può calcolare la media, usando la funzione somma delle quantità inverse perché tutte positive. Il peso medio è, allora, la media armonica delle probabilità di selezione per i vari strati $ijkl$. Il peso normalizzato a uno per ogni strato $ijkl$ sarà dato dal rapporto tra i pesi effettivi finali $w_{p;ijkl}$ e il peso medio dato dall'inverso della media armonica, $1/\bar{\pi}^*$. Si può mostrare che i pesi $w_{p;ijkl}^*$ sono dati dal rapporto tra i pesi degli strati rispetto alla popolazione totale di riferimento e i pesi degli strati nel campione rispetto alla dimensione totale del campione, che conduce all'espressione

$$w_{p;ijkl}^* = \frac{N_{ijkl}}{n_{p;ijkl}} \frac{n}{N}. \quad (6)$$

I pesi $w_{p;ijkl}^*$ mantengono inalterata la dimensione complessiva del campione, ma possono alterare completamente la struttura delle dimensioni campionarie per strato.

6. Errori non campionari

A questo stadio del processo di indagine si può solo sostenere che il procedimento proposto potrebbe essere suscettibile di miglioramento; purtroppo, i vincoli logistici e di risorse umane e finanziarie pongono limiti insormontabili ai possibili miglioramenti del piano di campionamento; e in ciò emergono i tratti fisiologici di ogni piano concreto. Per esempio, si sarebbe potuto stratificare ulteriormente sia sul titolo di studio (o scolarità), sia sulla classe di età del frequentante; ma il numero dei dominî di studio sarebbe diventato molto elevato: già la proposta attuale conta di $10 \times 5 \times 2 \times 3$ (province \times tipologia formativa \times genere \times classe di durata) = 300 dominî. Il miglioramento del piano, in tal senso, avrebbe comportato un aggravio di costi, che non sarebbe stato sostenibile e, forse, non avrebbe prodotto un sostanziale guadagno nelle stime.

Nelle indagini campionarie occorre prestare attenzione anche all'insieme complessivo delle operazioni che si devono eseguire nella raccolta dei dati, che si articola in varie fasi e coinvolgono molteplici persone, come gli intervistatori e gli intervistati. Le cause di errore sono, pertanto, molteplici e non sempre controllabili; i loro effetti sono denominati *errori non campionari*. Occorre indirizzare gli sforzi nel ridurli perché possono diventare preponderanti rispetto agli errori campionari. Le caratteristiche ideali del processo di indagine sono: (a) assenza di errori nella lista di \mathcal{P} , ossia a ogni nominativo della lista corrisponde una e una sola unità di \mathcal{P} e viceversa, senza altri tipi di errori descritti in breve nel paragrafo 2; (b) la selezione delle unità è coerente con il piano di campionamento, ossia sono definite le probabilità di inclusione del primo e del secondo ordine; (c) le variabili sono rilevate senza errore per tutte le unità campionarie; (d) la codifica e la trascrizione su supporto magnetico è esente da errore (Cicchitelli, Herzel, Montanari, 1992). Non esiste ancora una teoria completa degli errori non campionari; pertanto, ogni indagine è un caso a sé e presenta un proprio *profilo dell'errore*. L'individuazione di tali errori richiedono una analisi dettagliata sul campo in cui si opera che descriva in modo completo e circoscritto tutte le operazioni

necessarie e le relative (potenziali) fonti di errore e, possibilmente, anche il loro effetto sull'errore complessivo (Bailar, 1983).

Gli errori non campionari sono distinti generalmente in tre tipologie (Lessler, Kalsbeek, 1992): (i)errori nella lista o errori di copertura, causati dalle già citate (§2) imprecisioni contenute nella lista; (ii)errori da mancata risposta, derivanti sia dall'impossibilità di procedere alla rilevazione per non reperibilità o assenza di alcune unità statistiche incluse nel campione (Kish, 1965), sia dalla non partecipazione all'indagine delle unità statistiche selezionate e rintracciate —rifiuto totale—, sia dall'assenza di cooperazione su una particolare domanda del questionario —rifiuto parziale—; (iii)errori di misurazione, generati da numerosi fattori che alterano il valore da osservare introducendo una differenza con il valore reale.

Gli errori della lista sono i più perniciosi perché si è impotenti nei loro confronti. In genere, si deve usare una lista che sia ragionevolmente aggiornata (attuale), che offra un accettabile grado di copertura di ϕ (completezza), sia relativamente priva di duplicazioni di unità (ridondanza), non contenga grappoli di unità corrispondenti a uno stesso nominativo (molteplicità), non includa unità senza un reale corrispondente empirico o estranee a ϕ (inesistenza, sopracompletezza), sia esente da errori di imputazione: nei nomi, negli indirizzi, nei numeri di telefono. La scelta della lista, come la selezione delle unità dalla stessa, deve tenere conto delle esigenze del committente, le quali definiscono gli obiettivi dell'indagine. Il processo non è, però, isolato dalle risorse disponibili per l'indagine e è in conflitto con la dimensione del campione perché aumentandola diminuisce, sí, l'errore campionario, ma cresce il rischio di un incremento dell'errore non campionario.

Ridurre gli errori da mancata risposta è il primo compito da perseguire in una indagine perché migliorare la *qualità dei dati* raccolti è lo sforzo da compiere per ottenere risultati più affidabili e fedeli alla realtà (Liepins, Uppuluri, 1990). Proprio gli strumenti che si usano nell'indagine costituiscono una fonte primaria di errore. L'esperienza può aiutare a progettare strategie efficienti, ma nelle realtà complesse le difficoltà sono ineliminabili. Una vasta letteratura fornisce utili raccomandazioni sui procedimenti da seguire, ma spesso in pratica ci si trova alla corda. Ecco un elenco di principî che si possono rintracciare in un qualunque manuale di metodologia per la ricerca sociale (Bailey, 1994): (a)cominciare l'intervista presentandosi e descrivendo concisamente obiettivi e oggetto dell'indagine; (b)sottolineare l'importanza della collaborazione degli intervistati perché consentirà, nel caso specifico, di migliorare il servizio di formazione professionale promosso dalla regione E-R; (c)rassicurare che i dati saranno assolutamente segreti, non ceduti a altri enti, e non usati per scopi diversi da quelli dell'indagine; (d)le risposte devono essere completamente libere e senza vincoli di sorta; (e)disporre alla fine le domande delicate, potenzialmente imbarazzanti o compromettenti —per esempio, le domande sul salario percepito e sul voto di diploma o di laurea—; (f)prestare attenzione alle batterie di domande, specie quelle con la stessa scala o con le stesse opzioni di scelta, per evitare risposte seriali —*response set*— (g)controllare il flusso informativo generato dalla serie di domande affinché sia coerente e efficace, evitando eccessivi salti logici e strutturali; (h)usare un linguaggio chiaro e semplice, evitando espressioni gergali o dialettali o tecniche; (i)predisporre un questionario il più breve possibile, evitando l'uso di domande lunghe e relative combinazioni di esse; (j)aiutare il rispondente nel ricordare eventi inerenti al passato; (k)minimizzare l'introduzione di aspetti sensibili; (l)verificare all'inizio, e revisionare successivamente, gli strumenti adottati (pre-test, test).

Gli errori di misurazione si sovrappongono, in parte, a quelli da mancata risposta perché l'assenza di una risposta potrebbe dipendere proprio da una formulazione ambigua o inadeguata. La misurazione comporta, in generale, che il processo applicato goda di alcune proprietà fondamentali (Zeller, Carmines, 1980): la *validità*, quando rileva effettivamente l'intensità o la proprietà del concetto in esame, ossia, consegue gli obiettivi fissati; l'*attendibilità*, quando applicato più volte agli stessi fenomeni, nelle stesse condizioni, riproduce (entro certi limiti) gli stessi risultati; la *precisione*, quando c'è la possibilità di valutare i sottomultipli dell'unità di misura. La terminologia non è ancora consolidata sicché, in alcuni contesti si usano termini più suggestivi: sinonimi di attendibilità sono i sintagmi *stabilità* (della misura) o *fedeltà* (dello strumento); sinonimo di precisione è il termine *accuratezza*. Una distinzione tipica degli errori di misurazione è basata sulla causa che li ha prodotti: (1)errori di *strumenti*, in genere, riconducibili al questionario per domande formulate in modo ambiguo, ordinate in modo inadeguato, o batterie dei test non tarati bene, e così via; (2)errori di *tecniche*, in genere, legati al tipo di tecnica utilizzata, come il questionario postale, l'intervista auto-somministrata, l'intervista telefonica, il CATI —*Computer Assisted Telephone Interviewing*— oppure il CAPI —*Computer Assisted Personal Interviewing*—, la batteria di test; (3)errori dell'*intervistatore*, derivanti dalla influenza che esercita sull'intervistato sia nell'incentivare o disincentivare la sua partecipazione, sia nel fornire o non fornire una data risposta; (4)errori dell'*intervistato*, connessi alla capacità di comprensione dell'intervistato o di ricordare gli eventi accaduti, alla sua idoneità e volontà di fornire risposte veritiere. L'indagine deve rilevare, in particolare, i periodi di occupazione e disoccupazione o i miglioramenti avvenuti nella carriera lavorativa; pertanto, si va incontro a diversi errori (Neter e Waksberg, 1964): l'errore *telescopico*, il rispondente ricorda l'evento, ma lo colloca in un momento errato del tempo tendendo a avvicinarlo al presente o a allontanarlo, rispetto alla data vera; l'errore di *condizionamento*, si ha in una intervista ripetuta nel tempo, quando si ha un decremento del numero di eventi riportati rispetto a quello reale; l'errore di *richiamo* (*recall loss effect*), quando si verifica una perdita di informazione dovuta o all'incapacità del rispondente di ricordare, o alla numerosità eccessiva di eventi da riportare (*report loading effect*).

La qualità dei dati raccolti si migliora soprattutto con l'accortezza nel reclutamento e addestramento dei rilevatori. In questa indagine ci si rivolgerà a una ditta esterna e specializzata. In tal caso, è importante controllare l'operato degli intervistatori perché la ditta può avere anche una lunga esperienza sul campo, ma i suoi dipendenti sono spesso reclutati da poco e non hanno ancora acquisito l'abilità necessaria. In questo settore di attività si osserva, in genere, un alto tasso di rotazione del personale (*turnover*). A tale scopo occorre uno o più supervisori. Non si sfuggirà, purtroppo, alla riluttanza delle unità selezionate a collaborare all'indagine che, nel caso specifico, si può stimare preventivamente intorno al 50%, in base alla supposta scarsa affidabilità della lista, alle esperienze condotte in precedenza, e anche alla letteratura esistente (Goyder, 1987; Groves, 1989; Groves *et al.*, 2002).

Altri tipi di errori non campionari possono emergere nelle fasi successive del processo di indagine: durante la codifica, la revisione, la registrazione, e l'elaborazione dei dati. Questi sono non meno rilevanti dei precedenti, ma non coinvolgono rispondenti e intervistatori, bensì altro personale addetto (Cicchitelli, Herzel, Montanari, 1992).

Il trattamento degli errori non campionari richiede assunti sulle caratteristiche di φ , sulla natura, e sulla distribuzione degli errori. Tali assunti non hanno sempre un

corrispondente empirico e, pertanto, occorre sempre operare con la maggiore coerenza possibile rispetto alle condizioni ideali di svolgimento dell'indagine. Solo così si ottengono dati validi, attendibili, e precisi; ma l'ideale non corrisponde al reale, sicché occorre anche accettare l'imprecisione, fissando eventualmente un limite massimo oltre il quale ricorrere a interventi migliorativi, seppur costosi. D'altronde, anche gli istituti specializzati, come l'ISTAT, o con ampie risorse umane e finanziarie, come la Banca d'Italia, che sono più accreditati presso la popolazione e supportati dalla legge —gli intervistati sono «obbligati» a partecipare all'indagine—, non riescono a ottenere il successo prescritto dalle condizioni ideali.

6.1. Misure relative alle mancate risposte

Si possono definire alcune percentuali (o indicatori) che esprimono la qualità del processo di indagine. Il Tasso di Efficienza dell'Intervistatore (TEI) esprime la percentuale di volte che un intervistatore ottiene le interviste agli indirizzi campionari contattati:

$$\text{TEI} = 100 \frac{\text{Numero di interviste}}{\text{Numero di contatti}}$$

ma la specificazione del numeratore e del denominatore possono evidenziare aspetti diversi del processo di intervista. Tali tipi di rapporti sono denominati anche Tassi di Completamento (TC, *completion rate*); riguardano il successo delle interviste e possono definirsi variando relativamente numeratore e denominatore per ottenere indicazioni diverse sul processo di indagine. L'indice TEI, così definito, esprime la potenzialità di partecipazione o di successo (rispondenti eleggibili) di quella indagine; oppure il grado medio di successo degli intervistatori nell'ottenere la cooperazione degli elementi di \mathcal{P} . Un indicatore diverso, e un po' più preciso di TEI, può definirsi come Tasso di Interviste Completate (TIC)

$$\text{TIC} = 100 \frac{\text{Numero di interviste completate}}{\text{Numero di unità campionarie eleggibili}}$$

dove le unità campionarie eleggibili si riferiscono, in alcuni testi, a quelle unità che potenzialmente possono essere intervistate; ossia, l'insieme delle unità intervistate completamente, più quelle intervistate parzialmente, più quelle che rifiutano di partecipare, più quelle che presentano uno stato di appartenenza non determinato, più quelle mai rintracciate. L'aggiunta o l'eliminazione, di queste ultime, consentono di ottenere tassi di risposta diversi che consentono di cogliere aspetti diversi del processo. Le combinazioni sono diverse e tante, qui ci si limiterà soltanto a alcune variazioni per definire alcuni tassi principali di indicatori di processo delle interviste:

$$\text{TEIC} = 100 \frac{\text{Numero di Interviste Completate}}{\text{Numero di Contatti}}$$

esprime il Tasso di Efficienza degli Intervistatori nei Contatti (TEIC) avuti con gli intervistati;

$$\text{PPPI} = 100 \frac{\text{Numero di Rispondenti a tutte le domande}}{\text{Numero di Rispondenti Cominciato Intervista}}$$

esprime la Propensione degli elementi della Popolazione a Partecipare all'Indagine (PPPI), diversamente è interpretabile come l'efficienza o abilità degli intervistatori a ottenere la collaborazione degli intervistatori;

$$TUR = 100 \frac{\text{Numero di Interviste Completate}}{\text{Numero di Unità nel Campione (Eleggibili + Ineleggibili)}}$$

dove «eleggibile», qui, deve essere inteso come il numero di unità appartenenti alla popolazione, viceversa per «ineleggibile», e esprime il Tasso di Unità statistiche Rilevate (TUR);

$$TUU = 100 \frac{\text{Numero di Interviste Completate Eleggibili}}{\text{Numero di Unità nel Campione (Eleggibili + Ineleggibili)}}$$

esprime il Tasso di Unità statistiche Utili ai fini della stima dei parametri di θ . Nelle interviste telefoniche è interessante considerare il rapporto tra il numero di contatti avuti e il numero complessivo di tentativi eseguiti per accertare il peso di lavoro compiuto dagli intervistatori, definibile Tasso di Successo nei Contatti Telefonici (TSCT)

$$TSCT = 100 \frac{\text{Numero di Successi nei Contatti Telefonici}}{\text{Numero Totale di Tentativi di Contatti Telefonici}};$$

con piccole variazioni di numeratore e denominatore, si possono ottenere indicatori un po' diversi e interessanti a seconda degli obiettivi che si vogliono conseguire.

In termini complementari, si possono calcolare i tassi di rifiuto; per esempio, il Tasso di Non Risposta, TNR, può essere così definito

$$TNR = 100 \frac{\text{Numero di Rifiuti}}{\text{Numero di Interviste + Numero di Rifiuti}}$$

Si possono determinare diversi tassi variando il denominatore, ossia, distinguendo tra tutti i potenziali rispondenti o tra tutti rispondenti con numero di telefono:

$$TNR_1 = 100 \frac{\text{Numero di Rifiuti}}{\text{Numero di Potenziali Rispondenti}}$$

$$TNR_2 = 100 \frac{\text{Numero di Rifiuti}}{\text{Numero di Potenziali Rispondenti con Numero di Telefono}}$$

7. Alcuni dati sulla rilevazione del 2001

Per fornire un'idea dell'onerosità del processo si riportano alcuni dati dell'indagine svolta nella regione E-R, sui corsi terminati nell'anno 2001. I valori della popolazione per provincia e per tipologia formativa, N_{ij} , sono riportati nella Tabella 1.

Tabella 1. Dimensioni della popolazione, N_{ij} , per provincia e per tipologia formativa

PROV.	POPOLAZIONE									
	1 EdL	2 FPOPD	3 IFTS	4 FFRL	5 FCI	6 FO	7 FIOF	8 FPU	9 FP	Totale
MO		582		693	17	4547	90	98	15	6042

RE	25	306		281	34	3714		115	58	4533
BO		723		968	108	5825		107	476	8207
FE	4	676		580	81	3261	43	31	490	5166
FC	115	606		86		1396	84	25	458	2770
PC	122	216		270	55	2267	39	39	149	3157
PR	35	259		449	65	2525		14	67	3414
RA	53	571		143	59	3842	72	106	317	5163
RN	147	358		323		1578	29		647	3082
RER	59	756	80	267	14	4981	109	81	90	6437
Totale	560	5053	80	4060	433	33936	466	616	2767	47971

Le corrispondenti dimensioni campionarie, n_{ij} , che indicano l'ampiezza risultante dalla precisione desiderata nello strato determinato dall' i -esima modalità del piano formativo e dalla j -esima tipologia formativa (strato ij , per $i=1, \dots, I=10$; $j=1, \dots, J=9$), sono riportati nella Tabella 2. Si può osservare che nelle modalità 3 (IFTS), 5 (FCI), 7 (FIOF), e 8 (FPU) si sono selezionate tutte le unità di φ (censimento). Per le altre, l'errore assoluto relativo alla stima di una percentuale è pari al 5%. La frazione di campionamento complessiva, come obiettivo, risulta pari al 22,8%: un valore assai elevato!

Tabella 2. Dimensioni campionarie, n_{ij} , con la precisione desiderata (r in percentuale) delle stime di una generica proporzione, per provincia e per tipologia formativa

PROV.	CAMPIONE									Totale
	1 EdL	2 FPOPD	3 IFTS	4 FFRL	5 FCI	6 FO	7 FIOF	8 FPU	9 FP	
r in %	5%	5%	100%	5%	100%	5%	100%	100%	5%	
MO	0	238	0	254	17	368	90	98	15	1080
RE	24	174	0	166	34	362	0	115	51	926
BO	0	258	0	284	108	375	0	107	218	1350
FE	4	252	0	237	81	357	43	31	221	1226
FC	90	242	0	71	0	312	84	25	214	1038
PC	94	141	0	162	55	341	39	39	109	980
PR	33	158	0	212	65	346	0	14	58	886
RA	47	236	0	106	59	363	72	106	178	1167
RN	108	190	0	179	0	320	29	0	248	1074
RER	52	262	80	161	14	371	109	81	74	1204
Totale	452	2151	80	1832	433	3515	466	616	1386	10931

La stratificazione aggiuntiva su altri due caratteri ha condotto al calcolo di dimensioni campionarie obiettivo, n_{ijkl} , per strato $ijkl$ (determinato dalla modalità i -esima del piano formativo, j -esima della tipologia formativa, k -esima del genere, e l -esima della durata del corso). Il numero di soggetti da intervistare negli strati è stato calcolato con il metodo dell'allocazione proporzionale. Il numero di individui da intervistare è aumentato perché nel calcolo di n_{ijkl} si è usato l'arrotondamento per eccesso sicché quasi in ogni strato si è intervistata una unità in più. La dimensione campionaria obiettivo, in totale, è diventata pari a 11119 unità, corrispondente al 23,2% della popolazione di riferimento: 188 unità statistiche in più della dimensione obiettivo.²

Rispetto ai dominî di studio occorre «rivalutare» l'errore finale che ci si può aspettare da una stima, a causa delle mancate interviste. Nella Tabella 3 si riportano,

perciò, le dimensioni campionarie effettivamente rilevate per provincia e per tipologia formativa. Si assumono le ipotesi che i dati mancanti si distribuiscano in modo casuale e siano incorrelati con le variabili oggetto di stima. Si può «rivalutare», allora, la precisione che fornisce il campione effettivo ottenuto dalla rilevazione. Per fornire un ordine di grandezza dell'errore ottenuto, rispetto a quello programmato, che a causa delle mancate collaborazioni sarà un po' più alto, si riportano, nella Tabella 4, gli errori assoluti rispetto a una proporzione pari a 0,5. In essa si può notare che le tipologie interessate alla rilevazione totale (censimento) erano IFTS, FCI, FIOF, FPU: per FCI non si è rilevato ancora alcun caso; per le rimanenti tre (IFTS, FIOF, FPU) si hanno errori assoluti elevati perché la numerosità della popolazione di riferimento nelle celle è bassa e, pertanto, anche la perdita di poche unità comporta una elevata perdita di informazione. L'errore programmato del 5% rimane pressoché invariato per le tipologie FPOPD e FO; per FFRL si ha un incremento di un paio di punti percentuali per quasi tutte le modalità del piano formativo, eccetto Forlì-Cesena e Ravenna che presentano un errore più del doppio perché si è avuto un basso tasso di copertura (Tabella 5); un esito analogo si ha per la FO, dove le modalità del piano formativo con errori molto elevati sono, ora, Reggio nell'Emilia e Piacenza; anche per EdL si ha un aumento dell'errore assoluto e, in particolare, le modalità del piano formativo con errori alti sono Reggio nell'Emilia e Parma. Tali aumenti derivano, come si è detto, essenzialmente dalla mancata partecipazione all'indagine dei soggetti che sono stati contattati e dalla presenza di errori nella lista che non hanno consentito di rintracciare sempre gli individui selezionati. L'entità dell'errore, allora, non solo aumenta, ma aumentano anche i rischi di possibili distorsioni perché coloro che sono rintracciabili e disponibili a partecipare possono essere anche soggetti con caratteristiche ben definite e ben distinte dai non rispondenti.

Tabella 3. Dimensioni campionarie effettivamente rilevate, $n_{p,ij}$, per provincia e per tipologia formativa

PROV.	CAMPIONE									Totale
	1 EdL	2 FPOPD	3 IFTS	4 FFRL	5 FCI	6 FO	7 FIOF	8 FPU	9 FP	
<i>r in %</i>	5%	5%	100%	5%	100%	5%	100%	100%	5%	
MO		231		135		231	25	66		688
RE	13	162		101		278		74	15	643
BO		236		180		312		45	231	1004
FE		230		201		280	9	20	227	967
FC	76	248		43		270	41	21	202	901
PC	72	99		88		251	15	25	6	556
PR	18	164		109		263		8	44	606
RA	35	216		44		278	27	62	175	837
RN	81	182		141		286	12		259	961
RER	33	216	31	114		189	48	40	50	721
Totale	328	1984	31	1156	0	2638	177	361	1209	7884

Nella Tabella 5 sono riportati, infine, le «percentuali di copertura», ossia la percentuale di intervista per dominio di campionamento delle interviste effettuate rispetto all'obiettivo programmato. Vi è una stretta analogia tra i dati della Tabella 4 e quelle della Tabella 5: più ci si avvicina all'obiettivo programmato del numero di interviste e la percentuale si avvicina al 100%, più la precisione «risultante» delle stime coincide con la precisione desiderata; più si hanno mancate partecipazioni e la

percentuale si allontana dal 100%, più la precisione «risultante» delle stime diminuisce, ossia aumenta l'errore assoluto.

Tabella 4. Errori assoluti, d_{ij} , relativi alle stime di una generica proporzione $P=0,5$ per piano formativo e per tipologia formativa (per IFTS, FCI, FIOF, FPU si era proceduto al censimento, ma per le difficoltà dell'indagine si è avuta una parziale partecipazione)

PROV.	1 EdL	2 FPOPD	3 IFTS	4 FFRL	5 FCI	6 FO	7 FIOF	8 FPU	9 FP
MO		5,1		7,7		6,4	17,1	7,1	
RE	19,6	5,4		8,0		5,8		7,0	22,4
BO		5,3		6,7		5,5		11,4	4,7
FE		5,4		5,7		5,7	30,0	13,5	4,9
FC	6,7	4,9		10,8		5,5	11,2	8,9	5,3
PC	7,6	7,4		8,8		6,0	20,5	12,1	40,1
PR	16,7	4,7		8,3		5,8		24,0	8,9
RA	9,9	5,4		12,6		5,8	15,3	8,2	5,1
RN	7,5	5,2		6,3		5,4	22,5	0,0	4,8
RER	11,7	5,8	14,1	7,1		7,1	10,8	11,3	9,5

Tabella 5. Percentuali di copertura del campione rispetto all'obiettivo fissato, calcolate rispetto alla Tabella 2 per provincia e per tipologia formativa

PROV.	1 EdL	2 FPOPD	3 IFTS	4 FFRL	5 FCI	6 FO	7 FIOF	8 FPU	9 FP	Totale
MO	0,0	97,1	0,0	53,1	0,0	62,8	27,8	67,3	0,0	63,7
RE	54,2	93,1	0,0	60,8	0,0	76,8	0,0	64,3	29,4	69,4
BO	0,0	91,5	0,0	63,4	0,0	83,2	0,0	42,1	106,0	74,4
FE	0,0	91,3	0,0	84,8	0,0	78,4	20,9	64,5	102,7	78,9
FC	84,4	102,5	0,0	60,6	0,0	86,5	48,8	84,0	94,4	86,8
PC	76,6	70,2	0,0	54,3	0,0	73,6	38,5	64,1	5,5	56,7
PR	54,5	103,8	0,0	51,4	0,0	76,0	0,0	57,1	75,9	68,4
RA	74,5	91,5	0,0	41,5	0,0	76,6	37,5	58,5	98,3	71,7
RN	75,0	95,8	0,0	78,8	0,0	89,4	41,4	0,0	104,4	89,5
RER	63,5	82,4	38,8	70,8	0,0	50,9	44,0	49,4	67,6	59,9
Totale	72,6	92,2	38,8	63,1	0,0	75,0	38,0	58,6	87,2	72,1

8. Conclusioni

Gli aspetti più critici derivano dalle difficoltà di realizzazione che sono già note e riscontrate da ogni esperienza sul campo: non si riusciranno a rilevare tutte le unità statistiche programmate nei diversi strati; ma occorre accontentarsi perché sarebbe arduo migliorare l'entità dei dati rilevati senza aumentare i costi oltre le quote accettabili. Per quanto concerne la sostituzione delle unità che si rifiutano di collaborare, si dovrà operare con l'accortezza di non ricorrere a tale pratica perché se da un lato si migliora la precisione delle stime, dall'altro si consegue un aumento della distorsione, perché le unità più disponibili a collaborare potrebbero avere caratteristiche distintive che inficiano o distorcono le stime dei parametri della popolazione. L'esperienza insegna che, anche in questo caso, non si riesce a rinunciare alla

sostituzione programmata. L'impegno da ottemperare concerne un controllo accurato sugli intervistatori per verificare se il lavoro svolto è puntuale, accorto, e onesto.

Nelle elaborazioni dei dati, che non coinvolgono verifiche di ipotesi, si possono usare i pesi $w_{p;ijkl}$ che riportano le stime alla popolazione obiettivo perché sono più immediati e leggibili. Nella verifica di ipotesi si dovrebbero usare i pesi $w_{p;ijkl}^*$ anche se ora non sono ancora disponibili; occorre tenere presente che alterano assai la struttura del campione effettivo per strato, ma è un modo per considerare le differenze introdotte dal piano di campionamento nella diversa numerosità per strato e dalle mancate collaborazioni che non sono uniformi per strato.

NOTE

¹ Il lavoro è il frutto della collaborazione dei due autori; tuttavia, ai fini delle attribuzioni si precisa che i paragrafi 1, 2, 7, e 8 sono attribuibili a Giuseppe Fiorani e i rimanenti paragrafi a Michele Lalla. Si ringraziano Giovanni Solinas e Paolo Silvestri, tra gli altri, per le fruttuose discussioni svolte sul tema della strategia campionamento. Una speciale gratitudine si esprime a Gabriele Marzano, responsabile delle attività di valutazione dell'Agenzia Emilia-Romagna Lavoro, per i suggerimenti che ci ha fornito di volta in volta e gli incoraggiamenti che ci ha dispensato nella progettazione dell'indagine, nella sua conduzione, e durante l'analisi dei dati. Il presente lavoro è una sintesi di alcune parti delle *Linee guida* predisposte da Poleis srl (2002) per conto della Regione Emilia-Romagna, citate dalle *Linee guida* prodotte Isfol (2003).

² Il numero di strati in cui si è eseguito il campionamento è circa 50 perché interessa solo cinque tipologie: EdL, FPOPD, FFRL, FO, FP; in realtà sono 48 perché due strati della Tabella 2 sono vuoti. Vi sono perciò $384=48 \times 2(\text{genere}) \times 4(\text{durata del corso})$ strati potenziali e, quindi, 384 unità in più del previsto; ma alcuni strati sono vuoti e altri sono poco numerosi, sicché i calcoli conducono alla selezione di tutte le unità presenti nello strato e non si ha più l'effetto dell'arrotondamento per eccesso; perciò, il numero di unità aggiunte si riduce da 384 a 188.

Riferimenti bibliografici

- Bailar B.A., Error profiles: uses and abuses, in Wright T., *Statistical Methodology Improvement Data Quality*, Academic Press, New York, 1983, pp. 117-130.
- Bailey K. D., *Methods of Social Research*, 4th edition, The Free Press, New York, 1994. Tr. it., *Metodi della ricerca sociale*, il Mulino, Bologna, 1995.
- Berliri C., Pappalardo C., La misurazione dell'impatto occupazionale dei programmi formativi, Relazione al Convegno Isfol: «*Formazione e occupazione: valutazione degli esiti occupazionali nelle regioni italiane*», Roma, 10 ottobre 2001.
- Bethlehem J.G., Keller W.J., Linear weighting of sample survey data, *Journal of Official Statistics*, **3**, 1987, pp. 141-153.
- Brucchi Luchino, *Manuale di economia del lavoro*, il Mulino, Bologna, 2001.
- Bulgarelli A., Gori E., Sistemi informativi per la valutazione dell'efficienza e dell'efficacia dei progetti di formazione professionale, in Convegno Intermedio SIS 2001, *Processi e Metodi Statistici di Valutazione*, Sessioni plenarie, specializzate, satellite, Facoltà di Economia, Università di "Tor Vergata", Roma, pp. 223-240.
- Cicchitelli G., Herzel A., Montanari G.E., *Il campionamento statistico*, il Mulino, Bologna, 1992.
- Cochran W. G., *Sampling Techniques*, John Wiley & Sons, 3rd ed., New York, 1977.
- Duncan G.J., Kalton G., Issue of design and analysis of surveys across time, *International Statistic Review*, **55**, 1987, pp. 97-117.

- Gori E., Vittadini G., *Qualità e valutazione nei servizi di pubblica utilità*, Etas, Milano, 1999.
- Goyder J., *The Silent Minority*, Basil Blackwell, Oxford, 1987.
- Groves R.M., *Survey Errors and Survey Costs*, Wiley & Sons, New York, 1989.
- Groves R.M., Dillman D.A., Eltinge J.L., Little R.J.A. *Survey Nonresponse*, Wiley & Sons, New York, 2002.
- Heckman J. J., The Economics and Econometrics of Active Labor Market Programs, in Ashenfelter O., Card D., *Handbook of Labor Economics*, Elsevier, Amsterdam, 1999, pp. 1865-2097.
- Horvitz D.G., Thompson D.J., A Generalization of Sampling Without Replacement from a finite Universe, *Journal of the American Statistical Association*, **47**, 1952, pp. 663-685.
- Isofol – Struttura nazionale di valutazione, *Linee Guida per la valutazione degli effetti occupazionali del FSE 2000-2006. Metodologie per la valutazione di programma*, Vol.3, Isofol, Roma, 2003.
- Kalton G., Citro C.F. Panel Surveys: Adding the Fourth Dimension, *Survey Methodology*, 19, 1993, pp. 205-215.
- Kish L., *Survey sampling*, Wiley & Sons, New York, 1965.
- Kish L., Weighting: why, when, and how, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1990, pp. 121-130.
- Kish L., Weighting for unequal P_j , *Journal of Official Statistics*, **8**, 2, 1992, pp. 121-130.
- Lessler J.T., Kalsbeek W.D., *Nonsampling Errors in Surveys*, Wiley & Sons, New York, 1992.
- Liepins G.E., Uppuluri V.R.R., *Data Quality Control. Theory and Pragmatics*, Marcel Dekker, New York, 1990.
- Little R.J., Rubin D.B., *Statistical Analysis with Missing Data*, John Wiley & Sons, New York, 1987.
- Martini A., Graduatoria di performance, impatto della formazione, analisi descrittive dell'esperienza dei formati: tre direzioni distinte per la raccolta e l'utilizzo dei dati sugli esiti occupazionali, *Incontro di lavoro: «Quali modalità di raccolta e quali utilizzi per i dati sugli esiti occupazionali dei formati»*, Torino, 2 e 3 dicembre 1999.
- Martini A., Garibaldi P., L'informazione statistica per il monitoraggio e la valutazione degli interventi di politica del lavoro, *Economia & Lavoro*; **27**(1), 1993, pp. 3-22.
- Martini A., Cais G., Controllo (di gestione) e valutazione (delle politiche): un (ennesimo, ma non ultimo) tentativo di sistemazione concettuale, *Progetto Valutazione*, Torino, 1999, <http://www.prova.org>.
- Martini, M., I dati amministrativi come fonte di informazione statistica sulle imprese, *Economia & Lavoro*, Anno XXIV, 1990, pp. 45-58.
- Neter J., Waksberg J., A Study of Response Errors in Expenditures Data from Household Survey, *Journal of the American Statistical Association*, **59**, 1964, pp. 18-55.
- Potter F.J., A study of procedures to identify and trim extreme sampling weights, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1990, pp. 121-130.
- Poleis srl, *Linee guida per la valutazione delle azioni rivolte alle persone dell'obiettivo 3 – F.S.E. 2000-2006 in Emilia-Romagna*, Agenzia Emilia-Romagna Lavoro, Bologna, 2002.
- Rettore E., Trivellato U., Martini A., La valutazione delle politiche del lavoro in presenza di selezione: migliorare la teoria, i metodi o i dati?, Working Paper n. 46, Dipartimento di Scienze Statistiche, Università di Padova, 2002, pp. 3-35.
- Rubin D.B., *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York, 1988.
- Shadish W. R., Cook T. D., Campbell D. T., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin Co., New York, 2002.
- Verma V., *Weighting for Wave 1*, Working Group “European Community Household Panel”, Doc. PAN 36/95, Statistical Office of the European Communities, Luxembourg, 1995.

Zeller R.A., Carmines E.G. (1980). *Measurement in the Social Sciences: the Link between Theory and Data*, Cambridge University Press, Cambridge.