

L'EFFETTO TECNICA NELLE INDAGINI MIXED-MODE





L'EFFETTO TECNICA NELLE INDAGINI MIXED-MODE

ISBN 978-88-458-1942-1

© 2017

Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0. <https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire, trasmettere e adattare liberamente dati e analisi dell'Istituto nazionale di statistica, anche a scopi commerciali, a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat), marchi registrati e altri contenuti di proprietà di terzi appartengono ai rispettivi proprietari e non possono essere riprodotti senza il loro consenso.



INDICE

	Pag.
Premessa	7
1. L'uso della multi-tecnica con componente web nelle rilevazioni demografiche e sociali	9
1.1 Un quadro d'insieme	9
1.2 Struttura e contenuti del volume	11
2. Le indagini multi-tecnica e la qualità dei dati	13
2.1 La qualità nella fase di raccolta dati	13
2.2 L'effetto tecnica	14
2.3 I disegni d'indagine multi-tecnica	15
2.4 I tipi di disegno di questionario nelle indagini multi-tecnica	17
2.5 I paradati	18
3. Il disegno delle indagini multi-tecnica	19
3.1 La progettazione della rilevazione	19
3.1.1 Scelta delle tecniche e disegno del questionario	22
3.1.2 Modalità di gestione della multi-tecnica	22
3.2 Il disegno del campione	24
3.3 Il disegno del questionario	25
3.4 Le strategie per la massimizzazione dei tassi di risposta	29
3.4.1 Le "leve" per aumentare i tassi di risposta	31
3.4.2 Strategie flessibili: responsive e adaptive design	33
3.5 La prevenzione dell'effetto tecnica tramite pre-test e indagini pilota	34
4. Metodi diagnostici per la valutazione e la misurazione dell'effetto tecnica	39
4.1 Introduzione	39
4.2 Metodi esplorativi e inferenziali	40
4.2.1 Analisi esplorative per unità	40
4.2.2 Analisi esplorative per variabile	49
4.2.3 Analisi mediante indicatori statistici	54
4.2.4 Analisi mediante test statistici	56
4.3 Metodi basati su modello	61
4.3.1 Valutazione dell'effetto selezione	61
4.3.2 Valutazione dell'effetto misurazione	62



4.3.3 <i>Metodi per distinguere l'effetto selezione dall'effetto misurazione</i>	66
5. Metodi per la valutazione e il trattamento dell'effetto tecnica in fase di stima	69
5.1 <i>Introduzione</i>	69
5.2 <i>L'effetto tecnica nell'approccio controfattuale</i>	71
5.2.1 <i>Premessa</i>	71
5.2.2 <i>L'equivalenza della misurazione delle tecniche</i>	71
5.2.3 <i>L'effetto tecnica in un disegno concorrente</i>	73
5.2.4 <i>L'effetto tecnica in un disegno sequenziale</i>	74
5.3 <i>Metodi di inferenza causale per la stima dell'effetto tecnica</i>	74
5.3.1 <i>Modelli di analisi</i>	74
5.3.2 <i>Stima degli effetti selezione e misurazione</i>	76
5.4 <i>Metodi per il trattamento degli effetti selezione e misurazione</i>	77
5.4.1 <i>Premessa</i>	77
5.4.2 <i>Metodi basati sul propensity score</i>	79
5.4.3 <i>Metodi basati su modelli di regressione</i>	81
5.4.4 <i>Metodi standard di imputazione multipla</i>	82
5.4.5 <i>Metodo di imputazione con pesi frazionati</i>	82
5.4.6 <i>Metodo di imputazione con modelli di risposta e selezione</i>	84
5.5 <i>Il trattamento della mancata risposta totale nelle indagini multi-tecnica</i>	85
5.5.1 <i>Premessa</i>	85
5.5.2 <i>Indicatori di rappresentatività</i>	85
5.5.3 <i>Il trattamento della mancata risposta totale nei disegni con-correnti e sequenziali</i>	87
6. Analisi dell'effetto tecnica in due casi di studio	89
6.1 <i>Introduzione</i>	89
6.2 <i>La sperimentazione sull'indagine Aspetti della vita quotidiana 2014</i>	89
6.2.1 <i>Obiettivi della sperimentazione</i>	89
6.2.2 <i>Cenni sulla rilevazione</i>	90
6.2.3 <i>Analisi esplorative preliminari</i>	90
6.2.4 <i>Caratteristiche dell'applicazione metodologica</i>	93
6.2.5 <i>L'effetto selezione</i>	95
6.2.6 <i>Analisi fattoriale confermativa multigruppo</i>	97
6.2.7 <i>Conclusioni e ulteriori sviluppi</i>	100
6.3 <i>L'esperienza sulla gestione del censimento della popolazione</i>	101
6.3.1 <i>Obiettivi della sperimentazione</i>	101
6.3.2 <i>Cenni sulla rilevazione</i>	102
6.3.3 <i>Analisi esplorative preliminari</i>	103
6.3.4 <i>Analisi della rappresentatività dei campioni</i>	105
6.3.5 <i>L'acquisizione di informazione ausiliaria</i>	106
6.3.6 <i>Analisi degli effetti selezione e misurazione: il propensity score matching</i>	107
6.3.7 <i>L'effetto misurazione: il metodo basato sull'imputazione multipla</i>	112
6.3.8 <i>Conclusioni e ulteriori sviluppi</i>	116
Capitolo 7 – Sintesi e raccomandazioni generali	117
7.1 <i>Sintesi del volume</i>	117

7.2 Alcune raccomandazioni generali	118
Conclusioni e prospettive future	121
Ringraziamenti	123
Riferimenti bibliografici	125
Appendice: Procedure di calcolo di alcuni indicatori statistici per l'analisi esplorativa dei dati	133

PREMESSA¹

Coerentemente con il processo di modernizzazione del sistema di produzione delle statistiche demo-sociali dell'Istat, che ha determinato l'accentramento della funzione di raccolta dati dell'Istituto, la Direzione centrale per la raccolta dati (Dcird) ha avviato un progetto di analisi, ricerca e sperimentazione nell'ambito delle strategie di raccolta dati a tecnica mista che prevedono l'uso della tecnica *Computer assisted web interview* (Cawi) in questa area della produzione statistica.

Il sempre maggiore ricorso a strategie *miste* di raccolta dati, con sempre più spazio riservato all'uso della tecnica *web*, ha infatti imposto all'attenzione dell'Istituto il problema della progettazione ottimale di questo tipo di strategie, ai fini della prevenzione e del trattamento dei potenziali effetti distorsivi connessi al loro uso, a garanzia della diffusione di informazioni di sempre maggiore qualità contenendo costi e fastidio statistico sui rispondenti.

La complessità della fase di progettazione di strategie multi-tecnica è dovuta, in termini generali, ai differenti aspetti da tenere sotto controllo in questo tipo di rilevazioni (metodologici, operativi, organizzativi, tecnologici): al fine di garantire una trattazione e un'analisi del problema il più esaustiva possibile, il progetto di analisi che ha condotto alla stesura del presente volume ha visto pertanto il coinvolgimento di figure professionali e competenze di vari settori dell'Istat, trasversali e di produzione. Questo volume è pertanto il risultato delle attività di studio e sperimentazione condotte dal gruppo di progetto nel corso del 2017.

Il volume delinea un *framework* concettuale di riferimento nell'ambito delle rilevazioni a tecnica mista, e fornisce un quadro d'insieme delle problematiche metodologiche, operative e organizzative connesse alla progettazione e alla realizzazione di strategie *miste* che prevedono l'uso del Cawi in area demo-sociale, evidenziando i rischi connessi a una progettazione inaccurata o inappropriata. In questo senso, il volume (che segue il precedente volume "L'utilizzo della tecnica Cawi nelle indagini su individui e famiglie", pubblicato nel 2017 nella collana Letture statistiche-Metodi) rappresenta un ulteriore contributo al processo di progressiva armonizzazione e standardizzazione in Istat in questo ambito, sia dal punto di vista della terminologia, delle definizioni e dei concetti di base, sia relativamente agli aspetti metodologici, operativi e organizzativi, con particolare attenzione agli effetti delle tecniche sulle stime prodotte.

Il principale risultato atteso è una maggiore sensibilizzazione rispetto alla necessità di investire risorse in fase di progettazione di rilevazioni a tecnica mista, a garanzia di più elevati livelli di qualità e minori costi.

¹ La premessa è a cura di Saverio Gazzelloni.

1. L'USO DELLA MULTI-TECNICA CON COMPONENTE WEB NELLE RILEVAZIONI DEMO-SOCIALI¹

1.1 Un quadro d'insieme

Nell'ambito delle rilevazioni dirette condotte dagli Istituti nazionali di statistica, l'utilizzo di strategie di raccolta dati (o di *data collection*) efficaci ed efficienti riveste un ruolo chiave per l'acquisizione (e quindi per la diffusione) di informazioni statistiche di elevata qualità sia sul versante delle famiglie e degli individui, sia su quello delle imprese e delle Istituzioni. Tuttavia, il problema del controllo dei costi connessi alla strategia di rilevazione sul campo caratterizza soprattutto l'area delle statistiche demo-sociali: è in questo ambito, infatti, che sono tradizionalmente utilizzate tecniche di raccolta dati (in particolare, quelle di tipo *faccia a faccia*) che, se da un lato garantiscono tassi di risposta più elevati e una maggiore accuratezza delle misurazioni, dall'altro implicano costi elevati dal punto di vista sia organizzativo, sia operativo, e possono determinare eccessivo fastidio statistico su alcune tipologie di rispondenti.

L'esigenza di individuare un *trade-off* ottimale fra vincoli di costo e esigenze di qualità con specifico riferimento all'area delle statistiche demo-sociali, ha indotto molti Istituti nazionali di statistica a investire su strategie di tipo *misto*: in questo tipo di strategie, per uno stesso processo di indagine vengono utilizzate congiuntamente più tecniche di raccolta dati, scelte tenendo conto degli specifici obiettivi conoscitivi, delle caratteristiche della popolazione obiettivo, del contesto informativo, delle risorse disponibili. L'idea di base è che, sfruttando in modo ottimale le potenzialità di ciascuna tecnica, sia possibile garantire adeguati livelli di qualità dei dati (in termini di copertura, tassi di risposta, accuratezza dei dati osservati), minimizzando il fastidio statistico sui rispondenti e tenendo sotto controllo i costi. Nell'ambito delle strategie di tipo misto, sono spesso privilegiate quelle che prevedono l'uso del *web* (*Computer assisted web interview*, Cawi), sia in quanto meno costose, sia perché esse, in molti contesti applicativi e laddove adeguatamente progettate, offrono garanzie di una elevata qualità dei dati osservati (Groves e Lyberg, 2010; Couper, 2011; Kreuter, 2013).

Anche l'Istat, tenuto conto delle sempre più pressanti esigenze di contenimento dei costi nello specifico ambito delle statistiche demo-sociali, e della sempre maggiore penetrazione di internet nella popolazione italiana, sta progressivamente estendendo in diversi contesti produttivi l'uso congiunto del Cawi e di altre tecniche di raccolta dati. Esperienze significative in questo senso sono state condotte nel caso di importanti processi di produzione, primo fra tutti il *Censimento della Popolazione 2011* (Zindato *et al.*, 2017), nonché rilevazioni di tipo campionario, come ad esempio le rilevazioni sulle famiglie *Aspetti della vita quotidiana e I cittadini e il tempo libero* (Brait *et al.*, 2017), e l'*Indagine sui percorsi di studio e di lavoro dei diplomati* e l'*Indagine sull'inserimento professionale dei laureati*

¹ Il capitolo è stato redatto da Orietta Luzi, che ha anche curato l'intero volume.

(Gallo, 2017). Recentemente è notevolmente aumentata in Istat l'attenzione verso il problema della progettazione metodologica e operativa di strategie a tecnica mista (in tutte le loro componenti, quali la strategia di stima, il questionario, etc.), in modo da ottimizzare il rapporto qualità-costi attraverso la prevenzione e il controllo dei possibili effetti distorsivi derivanti dall'uso congiunto di più tecniche, e di acquisire elementi utili a garantire la possibilità di trattamento ex-post di tali effetti (come nel caso della rilevazione *Aspetti della vita quotidiana* per l'anno 2017).

Ricerche e studi riportati in letteratura (per una panoramica, cfr. ad esempio De Leeuw, 2005, e Dillman, 2009) dimostrano infatti che l'uso di strategie a tecnica mista può essere all'origine di specifiche componenti non campionarie dell'errore totale, comunemente denominate *effetto tecnica*, che si manifestano come differenze sistematiche nelle stime finali. In termini generali, questo tipo di errori è il risultato di due componenti: *effetto selezione* (dovuto sostanzialmente al fatto che le diverse tecniche sono caratterizzate da copertura e tassi di non risposta differenti), e *effetto misurazione* (dovuto in pratica all'uso di strumenti differenti per rilevare uno stesso fenomeno) (Biemer, 2017). Per di più, tali effetti sono spesso sovrapposti (*confounded*), per cui è necessario adottare specifici accorgimenti, in fase sia di disegno della rilevazione sia di trattamento dei dati, per eliminare i loro effetti distorsivi sugli output finali.

La letteratura sull'effetto tecnica e sulle problematiche connesse alla valutazione e al trattamento delle sue componenti è molto vasta. Le strategie proposte, tuttavia, sono sostanzialmente di due tipi, a seconda delle fasi del processo in cui si interviene/degli obiettivi da perseguire:

- 1) riduzione dell'effetto tecnica mediante ottimizzazione del disegno dell'indagine (tecniche utilizzate, questionario/i, campione/i, etc.²);
- 2) trattamento (*adjustment*) dell'effetto tecnica a valle della fase di raccolta dei dati.

La varietà degli approcci metodologici proposti nel caso di indagini mixed-mode (tra i più recenti citiamo, ad esempio, Shouten *et al.*, 2013, Klausch *et al.*, 2015, Buelens e Van den Brakel, 2017, Roberts e Vandeplass, 2017), e la crescente rilevanza assunta negli Istituti nazionali di statistica dal problema della misurazione e del contenimento dei relativi effetti distorsivi, ha richiamato anche l'attenzione di Eurostat sulla necessità di avviare un processo di condivisione di esperienze e di scambio di idee su questi temi in un'ottica di progressiva standardizzazione metodologica e operativa in quest'area.

La rilevanza di questi temi e la necessità di ulteriori sviluppi metodologici e operativi in questo ambito è stata sottolineata sempre a livello europeo, dalla Conferenza dei Direttori Generali degli Istituti nazionali di statistica europei (Dgins),³ nel cui ambito sono state formulate raccomandazioni relative alla necessità di rivedere in un'ottica di maggiore efficienza e qualità la fase di raccolta dei dati presso famiglie e individui e di sviluppare un'architettura comune per le statistiche sociali europee.

Rilevanti risultati sono stati conseguiti, inoltre, nell'ambito del progetto Europeo *Essnet on data collection for social surveys using multi-modes (Dcss)* (Eurostat, 2014), ed è in stato approvato presso la Commissione Europea un ulteriore progetto di collaborazione su queste tematiche.

2 Il tema del disegno di indagini a tecnica mista, denominato Mixed Mode Designs for social Surveys - Mimod, che garantiscano rapporti ottimali fra qualità e costo/burden sui rispondenti è oggetto di un recente volume del Journal of Official Statistics (cfr. AA.VV, 2017).

3 "Wiesbaden Memorandum", Settembre 2011 ([https://www.destatis.de/EN/AboutUs/Events/DGINS/ Document_Memorandum.pdf;jsessionid=250B3BD806299841B8D924749F4908E2.cae4?__blob=publicationFile](https://www.destatis.de/EN/AboutUs/Events/DGINS/Document_Memorandum.pdf;jsessionid=250B3BD806299841B8D924749F4908E2.cae4?__blob=publicationFile)).

E' in questo contesto nazionale ed europeo che in Istat è stato realizzato il progetto di analisi, ricerca e sperimentazione nell'ambito delle indagini a tecnica mista i cui risultati sono descritti in questo volume.

1.2 Struttura e contenuti del volume

Il volume ha come primo obiettivo la definizione di un *framework* concettuale comune di riferimento nell'ambito delle rilevazioni *mixed-mode*.

Il volume offre poi una panoramica delle soluzioni metodologiche e operative per l'ottimizzazione del disegno complessivo di strategie *miste* di raccolta dati, ai fini della prevenzione, del monitoraggio, della valutazione e del trattamento dell'effetto tecnica, illustrando per alcune di esse i risultati di applicazioni sperimentali a dati di indagini a tecnica mista condotte correntemente dall'Istat nell'area sociale.

Il volume è strutturato come segue.

Il Capitolo 2 contiene un'analisi dei vari aspetti connessi alla qualità nella fase di raccolta dati, e introduce i concetti e le definizioni di base nel contesto del disegno di indagini a tecnica mista, a partire dalla definizione dell'effetto tecnica e delle sue componenti.

Nel Capitolo 3 sono illustrati i vari aspetti connessi al disegno di rilevazioni *mixed-mode*, a partire dalla scelta delle tecniche e alla loro gestione congiunta, fino ad arrivare al disegno del campione e del/dei questionario/i. Nel Capitolo sono anche discusse le strategie per la massimizzazione dei tassi di risposta in rilevazioni multi-tecnica, e i metodi per la prevenzione dell'effetto tecnica tramite pre-test e indagini pilota.

Nel Capitolo 4 sono illustrati alcuni strumenti di analisi di tipo sia esplorativo, sia inferenziale, sia basati su modelli statistici per la diagnosi degli effetti di misurazione dovuti all'uso della tecnica mista. Tali analisi sono utili sia per offrire indicazioni su presenza ed entità di tali effetti, sia per orientare le attività di trattamento, sia per supportare l'eventuale revisione di elementi del disegno della rilevazione.

Il Capitolo 5 è dedicato alla presentazione di metodi inferenziali per la valutazione e il trattamento dell'effetto tecnica in fase di stima. Il contesto teorico di riferimento è costituito dall'inferenza causale che è utilizzata secondo una prospettiva controfattuale, ovvero ipotizzando l'esistenza di un risultato potenziale non realmente osservato. Questo perché, a meno che non siano stati predisposti disegni sperimentali o che le variabili d'indagine siano presenti in archivi amministrativi, la valutazione dei dati raccolti con più tecniche richiede la stima di dati controfattuali. L'obiettivo del Capitolo è quello di fornire, pur attraverso una breve, e non esaustiva, descrizione dei metodi reperibili nella vastissima letteratura sul tema, una visione d'insieme dei metodi utilizzabili, nella fase finale dell'indagine, per correggere le stime di parametri di interesse da effetti distorsivi di diversa natura. Un altro aspetto brevemente analizzato è il trattamento della mancata risposta totale nelle indagini *mixed-mode*.

Il Capitolo 6 contiene la descrizione di alcuni casi di studio realizzati applicando alcune delle metodologie illustrate nei Capitoli 4 e 5 a rilevazioni sulle famiglie condotte dall'Istat in area sociale. In particolare, vengono illustrati i risultati ottenuti nel contesto della rilevazione *Aspetti della vita quotidiana* (anno 2014), e *I Cittadini e il tempo libero* (anno 2015).

Il Capitolo 7 contiene una sintesi dei contenuti del volume, corredata da raccomandazioni generali su aspetti chiave della progettazione e realizzazione di strategie *mixed-mode* per rilevazioni demo-sociali su cui focalizzare l'attenzione.

Considerazioni conclusive e prospettive di ricerca future in Istat chiudono il volume. L'Appendice contiene informazioni e dettagli tecnici sulle procedure informatiche sviluppate per il calcolo di alcuni indicatori statistici, fra quelli illustrati nel Capitolo 4, utili in fase di analisi preliminare dei dati.

2. LE INDAGINI MULTI-TECNICA E LA QUALITÀ DEI DATI¹

2.1 La qualità nella fase di raccolta dati

Come è noto la fase della raccolta dati è una fase cruciale per la qualità: le scelte che vengono fatte possono condizionare i risultati che si otterranno e, d'altro canto, attraverso un'accurata progettazione e un attento monitoraggio di tale fase si può contribuire in misura significativa a prevenire o ridurre diverse tipologie di errore, in particolare le mancate risposte totali e gli errori di misurazione.

Nelle Linee Guida per la qualità dei processi statistici (Istat, 2012), in corrispondenza della fase di "Acquisizione diretta dei dati", il primo principio enunciato, relativo alla scelta della tecnica di raccolta dei dati, afferma che: *"Le tecniche e gli strumenti di raccolta dei dati devono essere scelti in funzione degli obiettivi dell'indagine e delle caratteristiche della popolazione oggetto di studio e in modo tale da massimizzare la qualità dei dati e, al tempo stesso, limitare il carico statistico sui rispondenti e i costi"*.

Si tratta quindi di scegliere la tecnica o la combinazione di tecniche più favorevoli a soddisfare l'obiettivo di massimizzare la qualità in presenza di vincoli che possono essere rappresentati dai costi, dai tempi di esecuzione e dal contenimento del carico statistico sui rispondenti.

Massimizzare la qualità vuol dire minimizzare l'errore totale (*total survey error*) associato all'indagine, dato dalla componente campionaria (nel caso di indagini campionarie) e dalla componente non campionaria (dovuta ad errori o a carenze nelle varie fasi del processo produttivo) (Biemer e Lyberg, 2003). A parità di errore campionario (che viene generalmente prefissato), minimizzare l'errore totale vuol dire minimizzare gli errori non campionari cioè errori di lista o di copertura, di mancata risposta (totale e parziale) e di misurazione. In fase di raccolta dati, questi errori possono essere originati dalla tecnica, o dalle tecniche utilizzate, così come da altri fattori che, a loro volta, interagiscono con la tecnica di raccolta dati, e che comprendono il tipo di argomenti trattati (che possono risultare più o meno sensibili o presentare difficoltà di comprensione), la presenza o meno dell'intervistatore e il questionario (formulazione dei quesiti, modalità di risposta, etc.).

Ne consegue che l'influenza della tecnica di raccolta dati sulla qualità dei dati è non solo significativa, ma coinvolge numerosi aspetti che interagiscono tra di loro. Come già detto, il presente volume ha l'obiettivo di analizzare i possibili effetti sulla qualità dei dati derivanti dall'uso congiunto di più tecniche di raccolta dati nelle indagini sociali (indagini multi-tecnica o *mixed-mode surveys*) che prevedano in particolare l'uso della tecnica web, ovvero l'auto-compilazione da parte dei rispondenti di un questionario on-line.

Nei prossimi paragrafi si forniscono alcune definizioni utili alla comprensione degli argomenti trattati nel resto del volume.

¹ Il capitolo è stato redatto da Marina Signore.

2.2 L'effetto tecnica

In termini generali, qualunque tecnica di raccolta dati può contribuire a generare errori di copertura, di mancata risposta e di misurazione causati in modo specifico dalla tecnica scelta. Ad esempio le indagini telefoniche (o Cati - *Computer assisted telephone interview*) presentano problemi specifici di copertura poiché è in costante aumento il numero di famiglie (o individui) che non sono presenti in alcuna lista telefonica o per i quali è difficile associare un nominativo ad un recapito telefonico. Tuttavia con l'espressione "effetto tecnica" non ci si riferisce ad una generica influenza della tecnica sulla qualità dei dati bensì ad "effetti distorsivi"² nei dati osservati e nelle stime finali. Chiaramente gli errori sistematici sono molto più preoccupanti e, in generale, più difficili da trattare rispetto agli errori casuali che tendono a compensarsi in media. Con riferimento all'effetto tecnica, gli effetti distorsivi si esplicitano in due modi distinti. Una prima conseguenza si ha quando una tecnica di raccolta dati comporta un'autoselezione dei rispondenti, quindi una sistematica diminuzione dei potenziali rispondenti all'indagine e un conseguente aumento degli errori di non osservazione, ovvero errori di copertura o di lista e mancate risposte totali.³

La seconda conseguenza si ha quando una tecnica di raccolta dati comporta un aumento degli errori di misurazione "sistematici" sui dati osservati, e quindi introduce una distorsione ad essi associata.⁴

Da quanto detto si evince che, in linea teorica, è possibile parlare di effetto tecnica anche nel caso in cui si adotti una sola tecnica di raccolta dati. Tuttavia è nel contesto dell'uso congiunto di più tecniche di raccolta che l'effetto tecnica acquista spessore e interesse, in quanto diviene più rilevante studiare (e limitare) gli errori derivanti dall'uso di più tecniche (anche in riferimento alla scelta di usare una sola tecnica o tecniche diverse). Diamo pertanto la seguente definizione di "effetto (multi)-tecnica" denominato, nel seguito, semplicemente "effetto tecnica" (in inglese *mode effect*):

"Effetto tecnica" (*mode effect*): si definisce effetto tecnica l'insorgere di differenze sistematiche nei dati raccolti dovute all'uso di tecniche differenti (Hox *et al.*, 2015). All'interno dell'effetto tecnica si distinguono due componenti:

- "Effetto selezione dovuto alla tecnica" (*mode selection effect*, nel seguito denominato semplicemente "effetto selezione"): errori di copertura e di mancata risposta totale (non risposta) associati alle diverse tecniche (errori di non osservazione). Il risultato è una differenza nella composizione dei campioni osservati;
- "Effetto misurazione dovuto alla tecnica" (*mode measurement effect*, nel seguito

2 Ricordiamo che "errori casuali" (di segno diverso) provocano prevalentemente un aumento di variabilità nelle stime, mentre "errori sistematici" (che hanno lo stesso segno) provocano in prevalenza distorsioni nei dati (sovrastime o sottostime del parametro oggetto di indagine) (Biemer e Lyberg, 2003).

3 Gli errori di non osservazione si distinguono in "errori di copertura o di lista" (errori dovuti alla mancata inclusione, duplicazione, aggiornamento della lista utilizzata per contattare le unità oggetto di indagine) e in "mancate risposte totali" o "non risposte" (errori dovuti al fatto che, per varie cause, viene a mancare tutta l'informazione relativa ad una unità appartenente alla popolazione oggetto di indagine (Biemer e Lyberg, 2003). Anche le "mancate risposte parziali" (errori dovuti alla mancata misurazione del valore di una variabile per una determinata unità) quando manca l'informazione relativa a una o più variabili in un determinato questionario rientrano negli errori di non osservazione (Biemer e Lyberg, 2003).

4 Gli "errori di misurazione" sono errori di osservazione e sono tali per cui il valore rilevato per una data variabile su una data unità è diverso dal valore vero. Gli errori di misurazione possono essere sia di tipo casuale che sistematico. Gli errori di misurazione sono tra i più difficili da stimare poiché, in generale, richiedono l'effettuazione di studi o indagini appositamente disegnate (Biemer e Lyberg, 2003).

denominato semplicemente “effetto misurazione”): errori di misurazione dovuti alle tecniche (risultato degli errori di osservazione) che introducono una distorsione (differenze sistematiche nei dati o *measurement bias*).

L'effetto misurazione si verifica se un rispondente tende a fornire risposte diverse ad uno stesso quesito al variare della tecnica di raccolta utilizzata.

Ad esempio rispondenti ad una intervista Cawi possono interpretare una domanda in modo diverso da rispondenti a cui la stessa domanda viene posta di persona da un intervistatore per il solo effetto della diversa modalità di somministrazione. In letteratura, si sottolinea come l'uso congiunto di tecniche con e senza intervistatore (ad esempio Cati/Cawi o Capi⁵ /Cawi) rappresenti un fattore di criticità in quanto la presenza o meno dell'intervistatore corrisponde a stimoli diversi e quindi contribuisce a creare un diverso meccanismo di risposta (Vandenplas et al., 2016). Inoltre, gruppi di rispondenti diversi possono avere una propensione⁶ più o meno alta a partecipare ad una indagine a seconda della tecnica utilizzata⁷ (effetto selezione). Ad esempio, per le persone anziane o con una conoscenza limitata del computer e di internet, la tecnica Cawi costituisce un ostacolo di per sé che induce un effetto di auto-selezione.

Pertanto, nelle indagini *mixed-mode*, le differenze osservate tra gruppi di rispondenti possono essere determinate sia da differenti caratteristiche dei gruppi sia da errori di misurazione. Generalmente accade che i due effetti si confondano tra di loro (*confounding effect*) rendendo più complesso il problema inferenziale della stima non distorta dei parametri oggetto di indagine e compromettendo i vantaggi connessi all'uso congiunto di più tecniche (Vannieuwenhuyze et al., 2014). In ogni caso è opportuno sottolineare che non è possibile studiare l'impatto degli errori di misurazione se non si è prima provveduto a scorporare (e quindi a eliminare) l'effetto selezione (Vandenplas et al., 2016). A tal fine è necessario disporre di informazione ausiliaria che non dipenda dalla tecnica (*mode insensitive*), disponibile in registri statistici o in fonti amministrative (si veda il Capitolo 6). Il Capitolo 5 presenta alcuni modelli che possono essere utilizzati allo scopo di stimare l'effetto tecnica e le sue componenti.

2.3 I disegni d'indagine mixed-mode

Com'è noto, le indagini *mixed-mode* vengono adottate per ovviare ai limiti intrinseci di una singola tecnica e per cercare di massimizzare uno o più obiettivi di qualità, di contenimento del carico statistico o di riduzione dei costi (cfr. Capitolo 3).

Sono stati quindi studiati diversi tipi di disegni che si possono adottare combinando più tecniche di raccolta dati: in questi casi si cerca di utilizzare le caratteristiche di una tecnica per compensare le limitazioni dell'altra.

In generale, la scelta delle tecniche non è limitata alla fase di raccolta dati sul campo. In realtà, il problema è più ampio e riguarda la scelta delle modalità con cui contattare i rispondenti anche nelle fasi preliminari (contatto iniziale, appuntamenti per l'intervista, lettera di presentazione dell'indagine, screening di sottopopolazioni, etc.) e successive alla raccolta

5 Computer assisted personal interview.

6 Con il termine propensione alla risposta si indica la probabilità di un rispondente di partecipare all'indagine e che può variare in funzione di diversi fattori tra cui anche la tecnica utilizzata per raccogliere i dati.

7 Per la formalizzazione si rimanda al Capitolo 5.

dati sul campo (solleciti alle unità non rispondenti, re-interviste su un sotto-campione, verifica dell'effettuazione dell'intervista da parte del rilevatore, indagini longitudinali, etc.). In questi casi si parla di "sistemi di indagine multi-tecnica" (*mixed-mode survey systems*) (de Leeuw, 2005).

Per le finalità del presente volume, l'attenzione si concentra sulla fase della raccolta dati sul campo. In questo ambito vi sono due principali tipi di disegni di indagini multi-tecnica:

- disegno di indagine con "tecniche concorrenti" (*Concurrent mixed mode*) in cui una tecnica di rilevazione è utilizzata per raggiungere una parte del campione e una diversa tecnica è utilizzata per raggiungere un'altra parte del campione;
- disegno di indagine con "tecniche sequenziali" (*Sequential mixed mode*) in cui si utilizza la stessa tecnica di rilevazione per tutto il campione nella prima fase dell'indagine, seguita poi da una tecnica diversa su coloro che non hanno risposto alla prima fase.

La scelta delle singole tecniche da usare congiuntamente, in modo concorrente o sequenziale, si basa su di un insieme di fattori che vengono esplicitati nel Capitolo 3 congiuntamente ai possibili disegni campionari. Inoltre si possono adottare due o più tecniche di raccolta dati in entrambi i disegni descritti.

In termini generali, la scelta può essere guidata dal tipo di errore non campionario sulla cui riduzione si intende agire in maniera prioritaria. Ad esempio si può cercare di contenere l'effetto selezione cercando di ridurre i potenziali errori di copertura oppure di innalzare i tassi di risposta. Di seguito si riportano alcuni esempi di disegni e delle rispettive finalità.

Disegni di indagine con tecniche concorrenti vengono generalmente utilizzati per ovviare a problemi di copertura legati ad una singola tecnica (ad es. per fronteggiare i problemi di copertura legati al Cati si può affiancargli una tecnica faccia a faccia). In base alla letteratura, i disegni concorrenti non sembrano avere effetti significativi sui tassi di risposta (de Leeuw, 2005) al di là di un generico effetto positivo di usare le indagini multi-tecnica rispetto alla mono-tecnica (Vandenplas, 2016). Pertanto, le tecniche concorrenti possono essere utilizzate per aumentare la copertura (cioè coprire strati di popolazione che hanno caratteristiche diverse) sebbene questo non comporti necessariamente un incremento dei tassi di risposta.

Per le indagini che utilizzano tecniche differenti in modo concorrente con assegnazione della tecnica secondo le caratteristiche di contatto delle unità della popolazione, è necessario valutare insieme all'errore di misura anche l'effetto selezione causato dalla distribuzione dissimile delle sottopopolazioni individuate.

I disegni di indagine con tecniche sequenziali invece sono generalmente utilizzati per ovviare a problemi di mancata risposta totale. In questi casi, si utilizza una tecnica (in genere poco costosa) come tecnica principale per tutto il campione seguita poi da una tecnica (in genere più costosa) per la quale ci si attende un più alto tasso di risposta. Si utilizza, ad esempio, come prima tecnica una tecnica web seguita da un'intervista con rilevatore (Capi/Papi⁸) sui non rispondenti ad una certa data. Con disegni di questo tipo, oltre ad aumentare i tassi di risposta, si può anche cercare di contenere i costi di rilevazione.

Nella scelta del tipo di disegno (concorrente o sequenziale) influiscono anche gli aspetti gestionali della rilevazione, che possono risultare più o meno complessi a seconda dei casi.

Entrambi i disegni, concorrenti e sequenziali, possono indurre un effetto tecnica, in particolare per quanto riguarda l'effetto misurazione (errori di misurazione dovuti alla tecnica utilizzata). L'uso di tecniche sequenziali può indurre un effetto selezione derivante dalla

8 Paper assisted personal interview.

differente rappresentatività dei due sotto-campioni. Inoltre, come già detto, l'effetto misurazione può confondersi con l'effetto selezione (de Leeuw, 2005).

2.4 I tipi di disegno di questionario nelle indagini mixed-mode

Un aspetto cruciale delle indagini *mixed-mode* riguarda il disegno del questionario, ovvero se il questionario debba essere adattato al tipo di tecnica oppure se sia preferibile utilizzare il medesimo questionario a prescindere dalla tecnica. Entrambe le opzioni presentano vantaggi e svantaggi rispetto ai problemi di qualità e organizzativi di cui tener conto in fase di progettazione dell'indagine (si veda al riguardo il Capitolo 3).

In generale, si distinguono due approcci principali al problema:

- “disegno del questionario unificato” (*Unified mode design or uni-mode design*) in cui il questionario somministrato ai rispondenti è lo stesso indipendentemente dalla tecnica. In questo approccio si fornisce lo stesso “stimolo”, ovvero stesso quesito e modalità di risposta, per tutte le tecniche. A tutti i rispondenti viene pertanto offerto uno “stesso stimolo” (*same offered stimulus*);

- “disegno del questionario tecnica-specifico” (*mode-specific o generalised mode design*) in cui i quesiti vengono costruiti in modo differente al variare della tecnica con l'obiettivo di conseguire una equivalenza semantico-cognitiva. In questo approccio si cerca l'uguaglianza dello “stimolo percepito” dai rispondenti (*same perceived stimulus*), adattando specifici aspetti del questionario a ciascuna tecnica, per ottenere risposte equivalenti.

Come descritto in dettaglio nel Capitolo 3, il questionario unificato viene spesso scelto per ridurre l'influenza del questionario sulle risposte ottenute con le varie tecniche, e perché più semplice da gestire da un punto di vista organizzativo (stesse istruzioni, stessi formati dei dati, etc.).

Adottare questionari specifici per tecnica significa invece ricercare una equivalenza delle risposte ottenute utilizzando domande e modalità di risposta disegnate in funzione della tecnica con cui vengono somministrate. È questa una ipotesi di ricerca che deve essere dimostrata prima di poter essere applicata ad una indagine.

A tal fine è necessario conoscere in che modo ciascuna tecnica influenzi le modalità di risposta da parte dei rispondenti e quindi conoscere il processo cognitivo domanda-risposta, ovvero come la domanda viene percepita, come viene elaborata l'informazione e come viene fornita la risposta, al variare della tecnica.

Il Capitolo 3, nel paragrafo dedicato ai pre-test del questionario, descrive come sia possibile sincerarsi dell'equivalenza cognitiva dei questionari costruiti con questo approccio prima dell'avvio dell'indagine estensiva.

In generale si può affermare che il mezzo può modificare il significato, ovvero che alcuni quesiti possono essere percepiti in modo diverso a seconda della tecnica e che alcune formulazioni dei quesiti sono più adatte ad una determinata tecnica piuttosto che ad altre. Può essere pertanto utile verificare sempre questi aspetti con i rispondenti, per esempio in fase di test o di indagine pilota per avere una quantificazione dell'errore di misura al variare del disegno del questionario.

2.5 I paradata

Scegliere la migliore combinazione di tecniche per massimizzare la qualità dei dati raccolti non è un problema di facile soluzione e, nonostante la ricca letteratura disponibile sull'argomento, non sono disponibili regole di carattere generale né *best practices* da poter mutuare direttamente.

Per orientarsi in questa scelta è importante in primo luogo avere un quadro chiaro dei vantaggi e degli svantaggi legati a ciascuna tecnica e a determinate combinazioni multi-tecnica (ad es. la combinazione Cati/Capi in alternativa alla combinazione Capi/Cawi) rispetto agli obiettivi dell'indagine e alla popolazione obiettivo.

La conoscenza teorica deve poi essere suffragata da evidenze empiriche ottenute dalla fase di raccolta dati (in precedenti edizioni dell'indagine o in indagini similari) oppure da test o indagini pilota (si veda il Paragrafo 3.5).

A tal fine, i paradata possono rivestire un ruolo importante. Ricordiamo che i paradata sono informazioni sui dati e sono caratterizzati dal fatto di essere ottenuti nel corso del processo stesso di raccolta spesso in modo automatico (Couper, 1998). Esempi noti di paradata sono tutto quell'insieme di informazioni sull'intervista ottenute con le tecniche assistite da computer (numero di tentativi di chiamata, durata dell'intervista, errori/correzioni di singoli quesiti, interruzioni nella compilazione di un questionario on-line, etc.). Rientrano tra i paradata anche informazioni di tipo qualitativo come i feedback dei rilevatori o l'osservazione dell'intervista da parte di supervisori.

I paradata possono essere analizzati con diverse finalità come ad esempio monitorare la performance degli intervistatori. Qui si vuole richiamare l'attenzione sul fatto che i paradata possono essere utilizzati anche per guidare nella scelta del disegno di indagini multi-tecnica, ad esempio per conoscere la preferenza dei rispondenti (o gruppi di) nei confronti della tecnica (Vandenplas *et al.*, 2016) o per disegnare strategie *miste* flessibili (quali *responsive* e *adaptive design* come descritto nel Paragrafo 3.4.2). Un altro esempio sull'uso dei paradata viene fornito nel Capitolo 4 quando vengono descritte le analisi preliminari ed esplorative.

3. IL DISEGNO DELLE INDAGINI A TECNICA MISTA¹

3.1 La progettazione della rilevazione

È opportuno premettere che in questo volume si affronta il tema della progettazione e dell'utilizzo della *tecnica mista* limitatamente all'attività di rilevazione dei contenuti propri dell'indagine, non soffermandosi, invece, sull'adozione di più tecniche in funzione delle diverse finalità del contatto con il rispondente (cfr. Capitolo 2), quali ad esempio i solleciti oppure i follow-up, siano essi finalizzati a sanare errori di coerenza oppure ad effettuare un'indagine di qualità.

Come introdotto nel Capitolo 2, si può affermare, in estrema sintesi, che la scelta della tecnica di acquisizione per la rilevazione dei contenuti di indagine è necessariamente funzionale a tre aspetti: obiettivi dell'indagine, caratteristiche della popolazione, massimizzazione della qualità. Tale scelta è inoltre condizionata da due vincoli: limitazione del carico statistico sui rispondenti e contenimento dei costi (Istat, 2012).

Alla luce di quanto detto, un insieme di fattori devono essere analizzati in parallelo per individuare la tecnica di acquisizione più adatta, quali:

- 1) il tema trattato;
- 2) la complessità dei concetti da rilevare;
- 3) la complessità del questionario (salti, controlli di coerenza);
- 4) il pressing sul rispondente per contenere le mancate risposte;
- 5) il tasso di risposta atteso;
- 6) la durata dell'intervista;
- 7) l'ambiente circostante al rispondente;
- 8) il supporto/influenza del rilevatore sul rispondente;
- 9) le caratteristiche/dotazione tecnica del rispondente;
- 10) la tempestività della raccolta dati;
- 11) il costo.

Un aspetto determinante che influisce su molti dei fattori citati è la presenza o meno dell'intervistatore: se da un lato, infatti, l'intervistatore 'facilita' la rilevazione fornendo supporto all'intervistato, motivandolo a rispondere e garantendo più elevati tassi di risposta (contenimento della mancata risposta totale/parziale –cfr. Capitolo 2), d'altro canto incide notevolmente sul costo per intervista, necessita di un'accurata formazione e implica un maggior rischio di influenza sulle risposte.

Per tale motivo, nel sintetizzare i pro e i contro di ogni tecnica per ciascuno degli aspetti elencati sopra, si è ritenuto opportuno enfatizzare nella Tavola 3.1 la distinzione tra tecniche autocompilate e tecniche con rilevatore.

Come si evince dalla Tavola 3.1, ciascuna tecnica implica vantaggi e svantaggi, per cui la scelta di quella più adatta all'indagine dipende dal diverso peso che, a seconda dei casi,

¹ Hanno collaborato alla stesura del capitolo: Stefania Macchia (paragrafo 3.1), Claudia De Vitiis (paragrafo 3.2), Manuela Murgia (paragrafo 3.3), Gabriella Fazzi (paragrafo 3.4), Martina Lo Conte (paragrafo 3.5).

si attribuisce a ciascuno degli aspetti elencati; lo sforzo è quello di individuare la tecnica che riesca a bilanciare al meglio le esigenze connesse alla qualità dei dati rilevati ed i vincoli in termini di costo e di tempo (Dillman *et al.*, 2009).

Tavola 3.1 - Pro-Contro delle tecniche su fattori da analizzare in fase di progettazione

Aspetti/Tecnica	Tecniche autocompilate		Tecniche con intervistatore	
	Cartacea Postale - Consegnato/ripreso da rilevatore	Cawi	Telefonica Cati	Faccia a faccia Capi/Papi
Tema trattato	Possibili tutti i temi, anche quelli delicati.	Possibili tutti i temi, anche quelli delicati, se si attua una strategia per far sì che il rispondente si senta 'protetto'. ¹	Possibili tutti i temi, anche quelli delicati, previa specifica formazione degli intervistatori.	Possibili tutti i temi, anche quelli delicati, se si attua una strategia per 'proteggere' il rispondente. ²
Complessità dei concetti da rilevare	Il trattamento di concetti complessi è subordinato al fatto di poter includere istruzioni chiare e sintetiche.	Il trattamento di concetti complessi è subordinato al fatto di poter includere istruzioni chiare, sintetiche e di facile visualizzazione o accesso.	Possono essere trattati anche concetti più complessi.	Possono essere trattati anche concetti più complessi.
Complessità del questionario (salti e controlli di coerenza)	Struttura dei salti molto semplice, controlli non gestibili.	Salti anche complessi; controlli semplici, è necessario garantire semplicità della navigazione.	Può essere anche molto complesso sia in termini di salti che di controlli.	Può essere molto complesso con il Capi; con il Papi l'intervistatore può supportare nella gestione dei salti (che però non possono essere molto complessi); controlli non gestibili.
Contenimento mancata risposta parziale	E' autocompilato, quindi il rispondente ha tutto il tempo che vuole per rispondere. Tuttavia con il postale è impossibile esercitare pressing per ottenere singole risposte; se consegnato/ripreso dal rilevatore, possibile un pressing limitato.	Possibile pressing su un numero contenuto di quesiti tramite controlli soft in caso di mancata risposta.	Possibile pressing da parte dell'intervistatore per ottenere le risposte, ma contenuto perché c'è il rischio che l'insistenza si traduca in risposte veloci e vaghe o interruzioni dell'intervista.	Possibile pressing da parte dell'intervistatore per ottenere le risposte, ma contenuto perché c'è il rischio che l'insistenza si traduca in risposte veloci e vaghe.
Tasso di risposta totale atteso	Abbastanza basso, soprattutto nel caso del postale.	Basso	Abbastanza alto	Alto
Lunghezza del questionario/ Durata dell'intervista	Lunghezza limitata (1-7 pagine).	Durata limitata Non più di 15/20 minuti.	Durata limitata. Non più di 15/20 minuti.	Possibile una durata superiore rispetto alle altre tecniche.
Controllo su chi risponde / Ambiente del rispondente	Non è possibile avere il controllo su chi risponde. Tuttavia il rispondente può scegliere il momento/luogo opportuno per rispondere.	Non è possibile avere il controllo su chi risponde. Tuttavia il rispondente può scegliere il momento/luogo opportuno per rispondere.	C'è il controllo su chi risponde e semi-garantita possibilità di effettuare l'intervista in un ambiente tranquillo tramite la gestione degli appuntamenti. L'intervistatore fornisce supporto; può esercitare influenza sul rispondente, ma contenuta in quanto il mezzo (telefono) consente un contatto indiretto.	C'è il controllo su chi risponde, ma serve strategia e una buona formazione degli intervistatori per organizzare l'intervista in un ambiente tranquillo.
Supporto /Influenza sul rispondente Architettura	Né supporto, né influenza da parte dell'intervistatore.	Né supporto, né influenza da parte dell'intervistatore.	L'intervistatore fornisce supporto; può esercitare influenza sul rispondente, maggiore rispetto alla tecnica Cati.	L'intervistatore fornisce supporto; può esercitare influenza sul rispondente, maggiore rispetto alla tecnica Cati.
Dispositivi tecnici del rispondente	Nessuno.	Richiesta connessione WEB.	Richiesto il telefono e numero accessibile.	Nessuno.
Tempestività	Bassa (Oltre a consegna e eventuale ritiro ci sono i tempi di registrazione).	Media, con apposita strategia di solleciti.	Alta.	Alta se Capi. Media se Papi.
Costo Può includere: stampa, registrazioni, costo rilevatori (compresi spostamenti sul territorio) e dispositivi hw	Basso.	Molto Basso.	Medio.	Alto.

¹ Assicurare l'anonimato del rispondente o la sicurezza dei dati via web; senza questa strategia, non è possibile affrontare temi delicati

² Attraverso contatti preliminari con il rispondente per concordare luogo/momento in cui organizzare l'incontro con il rilevatore e, in casi specifici, attraverso l'utilizzo di accorgimenti che impediscano al rilevatore di conoscere le risposte fornite (risposte fornite in buste chiuse, parziale autocompilazione del rispondente sui supporti CAPI, etc).

Se tuttavia, con l'avvento delle tecniche assistite da computer era considerato di per sé innovativo abbandonare la tecnica Papi per adottare il Cati piuttosto che il Capi, appare attualmente sempre meno percorribile la soluzione di affidarsi ad una singola tecnica, soprattutto per le indagini su famiglie e individui.

La necessità infatti da un lato di contenere i costi, che porta a preferire la tecnica Cawi, e dall'altro di tenere conto dei problemi di copertura tipici di ogni singola tecnica, nonché delle caratteristiche dei rispondenti, rende sempre più opportuna la scelta di soluzioni multi-tecnica.

A ciò si aggiunge che, nell'attuale momento storico in cui la comunicazione attraverso la 'rete' ha assunto una rilevanza mai verificatasi prima, non mettere a disposizione per le indagini su famiglie e individui anche la tecnica Cawi potrebbe da un lato dare un'immagine non 'moderna' dell'ente promotore dell'indagine e dall'altro 'allontanare' alcuni profili di rispondenti che maggiormente si avvalgono del Web.

Dunque l'adozione del *mixed mode* consente di ovviare a problemi dovuti al fatto che alcuni segmenti della popolazione possano non essere coperti con una determinata tecnica in funzione di diverse caratteristiche che li contraddistinguono, ad esempio:

- non hanno il telefono fisso e non possono quindi essere raggiunti con la tecnica Cati, a meno di non reperirne il numero di cellulare;
- non dispongono di connessione web e non possono quindi essere raggiunti con la tecnica Cawi;
- si tratta di particolari tipologie di popolazione, quali ad esempio gli stranieri con problemi connessi alla lingua, per cui tecniche senza intervistatore possono non essere efficaci.

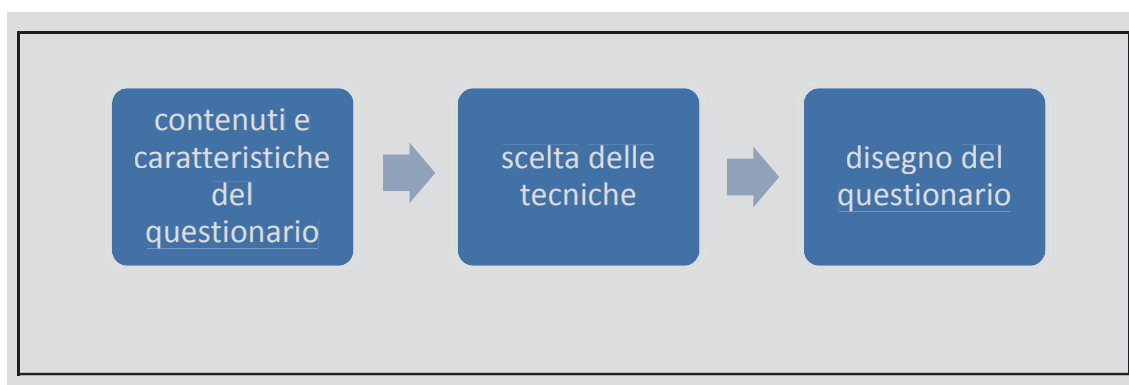
A ciò si aggiunge che affidarsi a più tecniche può andare incontro alle esigenze di particolari profili di rispondenti che abbiano dimostrato una specifica riluttanza per alcune tecniche piuttosto che per altre (per il Capi perché non accettano la visita dell'intervistatore, per il Cati perché stanchi di ricevere telefonate per promozioni, telemarketing, etc. oppure per il Cawi a causa di dubbi sulla sicurezza della trasmissione dei dati).

Un ulteriore fattore che potrebbe giustificare soluzioni multi-tecnica può derivare dalle caratteristiche del fenomeno da studiare che potrebbero essere meglio rilevate con diverse tecniche (Dillman, 2000); qualora ad esempio il questionario affronti tematiche che richiedano il supporto dell'intervistatore, ma ne includa anche qualcuna che affronta temi più delicati oppure richieda di reperire preventivamente le informazioni da fornire, potrebbe essere preferibile estrapolare una parte del questionario/sezione dedicata a questi argomenti e rilevarla con una tecnica autocompilata. Indubbiamente questa soluzione presenta una serie di complessità gestionali, tuttavia ne sono documentate alcune esperienze di utilizzo nelle indagini sociali (Martin, 2011), ed è citata in letteratura soprattutto in riferimento alla necessità di gestire set di quesiti particolarmente sensibili (Recanatini et. al., 2000, p.34). In Istat, ad esempio, una soluzione di questo tipo è stata adottata per l'indagine "Discriminazioni di genere, orientamento sessuale ed appartenenza etnica": tale indagine adottava l'intervista faccia a faccia come tecnica principale, mentre i quesiti più sensibili venivano rilevati con un questionario cartaceo che il rilevatore consegnava al rispondente perché lo autocompilasse e lo restituisse in busta chiusa. Diverso il caso dell'"Indagine sull'inserimento professionale dei dottori di ricerca 2009", in cui si è adottata la tecnica di rilevazione Cati, ma si chiedeva agli intervistati di compilare una sezione di questionario con tecnica Cawi per entrare maggiormente in dettaglio sull'attività di ricerca e sulle pubblicazioni, al fine di non appesantire eccessivamente l'intervista telefonica (Istat, 2013).

3.1.1 Scelta delle tecniche e disegno del questionario

Come esplicitato, la scelta delle tecniche da includere nel set di quelle da adottare è conseguente alla valutazione di una molteplicità di fattori (Figura 3.1) che includono i contenuti e le caratteristiche del questionario; parallelamente, però, la scelta delle tecniche impatta a sua volta sul disegno del questionario (cfr. Paragrafo 3.3).

Figura 3.1 - Scelta delle tecniche da utilizzare in una rilevazione multi-mode



La situazione più semplice è indubbiamente quella di adottare un questionario identico per tutte le tecniche (cfr. Capitolo 2), ma questo non comporta contro indicazioni soltanto a certe condizioni, ossia se:

- 1) il/i temi da trattare sono tutti 'neutrali' rispetto alla tecnica,
- 2) la lunghezza del questionario è contenuta,
- 3) la complessità del questionario è contenuta,
- 4) i quesiti sono semplici da capire,
- 5) le informazioni richieste sono immediatamente disponibili.

Se queste condizioni sono soddisfatte ed è quindi possibile adottare lo stesso questionario per tutte le tecniche, è intuitivo che la tecnica che presenta più vincoli ha un maggiore impatto sul disegno del questionario (cfr. Paragrafo 3.3).

In caso contrario il disegno del questionario può seguire i tre approcci diversi descritti nel paragrafo 3.3.

Un'altra casistica riguarda le indagini 'panel' per le quali la strategia è spesso quella di utilizzare per la prima intervista una tecnica che consenta di gestire un questionario più lungo, mentre per le successive interviste, nelle quali il questionario è impostato con molti quesiti a conferma, si può optare per tecniche adatte a interviste più brevi. Possibili soluzioni per le indagini panel sono, ad esempio, il Capi per la prima intervista e il Cati o il Cawi per le successive.

3.1.2 Modalità di gestione della multi-tecnica

Un'altra scelta progettuale per le rilevazioni multi-tecnica riguarda la modalità con cui gestire le diverse tecniche: sequenziale o concorrente (cfr. Capitolo 2).

Anche in questo caso, è opportuno distinguere la situazione in cui le caratteristiche di tutti i segmenti della popolazione in termini di propensione alla risposta o di disponibilità delle informazioni necessarie per il contatto con le diverse tecniche siano conosciute ex ante oppure no.

Qualora queste caratteristiche siano note ex ante, l'approccio ottimale è quello concorrente: si identificano i segmenti disgiunti del campione raggiungibili con ciascuna tecnica e si associa quindi la tecnica a ciascun individuo appartenente al segmento.

Se invece queste caratteristiche non sono note ex ante, in generale si adotta la strategia di tipo sequenziale anche se è possibile teoricamente utilizzare la concorrente lasciando al rispondente la scelta.

Con l'approccio *sequenziale*, la soluzione più funzionale è quella di adottare come prima tecnica quella meno costosa, in modo tale da ottenere quante più risposte possibile con questa tecnica e quindi contenere i costi. Generalmente, la tecnica meno costosa è il Cawi, per cui la strategia è quella di aprire la rilevazione con questa tecnica ed individuare una data limite, oltre la quale si identificano i non rispondenti che vengono contattati con un'altra tecnica; in ciascuno dei periodi, è inoltre opportuno sia definita una strategia di solleciti nell'ottica di massimizzare i tassi di risposta.

L'approccio *concorrente* consiste invece nel mettere a disposizione di tutto il campione tutte le tecniche contemporaneamente e lasciare che sia il rispondente a scegliere quale tecnica utilizzare.

E' opportuno precisare tuttavia che in entrambe le situazioni (propensione alla risposta con le diverse tecniche conosciuta a priori oppure no) l'associazione rispondente-tecnica non è mai rigida e irrevocabile; motivazioni legate alla strategia di andare incontro alle esigenze del rispondente oppure a garantire la copertura di tutti gli strati del campione possono portare a contattare il rispondente con una tecnica diversa rispetto a quella inizialmente utilizzata.

Inoltre, si sottolinea che le diverse scelte progettuali (sequenziale/concorrente) comportano alcune implicazioni di cui tener conto.

La somministrazione *concorrente* con più tecniche comporta indubbiamente un vantaggio in termini di tempo (la durata della rilevazione può essere più breve rispetto alla soluzione sequenziale); d'altro canto, diverse sono le problematiche a seconda che siano identificati ex ante gli strati raggiungibili con ciascuna tecnica oppure no. Il primo caso è sicuramente il più semplice in quanto a ciascuno strato viene associata una tecnica, tutti i costi e gli aspetti organizzativi sono noti in anticipo; in pratica ogni singola tecnica è gestita autonomamente rispetto alle altre.

Viceversa aprire contemporaneamente l'insieme delle tecniche su tutto il campione necessita imprescindibilmente di un sistema di gestione degli esiti di contatto integrato, che consenta di monitorare la rilevazione in tempo reale e gestire i contatti, dirottandoli, se necessario, da una tecnica ad un'altra. E' indispensabile, per esempio, che anche per il Cawi sia noto se ciascuna unità abbia effettuato o meno l'intervista, se l'abbia iniziata e non completata o se si sia limitata a registrarsi al sito senza compilare; a seconda infatti della situazione, sono diverse le iniziative da attivare (ad esempio assegnare il nominativo ad un intervistatore per l'effettuazione/completamento dell'intervista, oppure effettuare un sollecito via e-mail o telefonico, etc.).

La modalità *sequenziale* è indubbiamente più semplice da gestire, in quanto la rilevazione effettuata con la tecnica che si intende massimizzare (solitamente il Cawi) si conclude ad una certa data, per proseguire con le altre tecniche sulle unità che non hanno risposto con la prima. Non è necessario quindi un sistema integrato di gestione degli esiti per tecnica, in quanto gli esiti vengono analizzati alla fine del primo periodo di rilevazione. Tuttavia questa strategia presenta l'inconveniente che il numero di unità intervistabili per tecnica non è quantificabile a priori in modo puntuale e questo può costituire un fattore problema-

tico, soprattutto per la stima dei costi, in quanto, abitualmente, le tecniche adottate nella seconda fase della rilevazione sono quelle con il rilevatore (Cati e/o Capi) che più pesano sul budget. Il Capi, inoltre, ha un'ulteriore implicazione organizzativa legata agli spostamenti dei rilevatori sul territorio. A questa problematica si può ovviare stimando la propensione a rispondere per le diverse tecniche sulla base di precedenti rilevazioni su popolazioni analoghe. Va ricordato, infine, che anche nel caso della modalità sequenziale il più delle volte la seconda fase della rilevazione (quella generalmente svolta con modalità Capi o Papi) prevede la possibilità di continuare ad utilizzare il canale Cawi, situazione che, come ricordato, necessita di un adeguato sistema di gestione dei contatti e degli esiti.

3.2 Il disegno del campione

Nella progettazione di un'indagine multi-tecnica la definizione del disegno campionario è strettamente connessa con la scelta delle tecniche di rilevazione e deve tenere conto principalmente dei due seguenti aspetti:

- la tipologia di lista di selezione, che può essere unica o distinta per tecnica e può contenere unità elementari o cluster di unità;
- la modalità del mix di tecniche di rilevazione, in particolare se è concorrente o sequenziale.

Relativamente alla lista di selezione, generalmente è disponibile un unico archivio contenente le unità della popolazione, comune alle possibilità di contatto mediante le differenti tecniche: ad esempio l'archivio delle liste anagrafiche dei comuni italiani (Lac) per le indagini ISTAT su famiglie e individui, che contiene tutte le informazioni necessarie per il contatto diretto delle famiglie e degli individui e, per una parte delle unità, un contatto di telefonia fissa.

In alcuni casi invece può essere disponibile un archivio per ogni tecnica, ad esempio la lista telefonica per le interviste Cati oppure liste di indirizzi e-mail per indagini web. In questa ultima situazione possono essere definiti disegni campionari separati che possono essere riferiti a popolazioni non sovrapposte o con parziale sovrapposizione, circostanza che deve essere gestita seguendo opportuni metodologie di campionamento e di stima (approcci *multi-frame*).

Anche qualora sia disponibile un archivio unico, possono essere definiti campioni indipendenti con riferimento a ciascuna delle sottopopolazioni che sono raggiungibili con le diverse tecniche di rilevazione. In particolare, nei casi in cui è stato stabilito di utilizzare le differenti tecniche in modo concorrente su sottopopolazioni differenti, ad esempio le famiglie con telefono fisso con il Cati e con altra tecnica le famiglie senza telefono, è possibile definire campioni indipendenti basati su schemi campionari differenti.

Se viceversa le tecniche sono somministrate in modo sequenziale, necessariamente deve essere definito un unico campione selezionato dalla popolazione generale e pertanto i due sottocampioni che rispondono alle differenti tecniche risultano non indipendenti.

Nel caso meno frequente in cui le tecniche siano proposte simultaneamente ai rispondenti lasciando loro la scelta di quale utilizzare, il campione sarà necessariamente unico e la conseguente suddivisione nei sottocampioni relativi alle differenti tecniche sarà il risultato di un meccanismo dipendente dai rispondenti stessi.

Di fatto, in conclusione, la relazione tra i sottocampioni relativi alle diverse tecniche è determinata dalla modalità del mix di tecniche.

Per quanto riguarda la definizione del o dei disegni campionari, per ciascuno di essi vengono seguiti gli stessi criteri utilizzati per le indagini mono tecnica.

Il disegno campionario può essere, quindi, a uno o più stadi, con stratificazione delle unità (di primo e/o di secondo stadio) sulla base di informazioni contenute nell'archivio.

Generalmente si ricorre a disegni a più stadi, in cui i comuni sono le unità di primo stadio, quando è prevista l'intervista faccia a faccia (tecnica Papi o Capi) che richiede la concentrazione del campione sul territorio; nel caso dei disegni sequenziali questo implica che se alla fine della sequenza di tecniche è prevista tale modalità di contatto, sarà necessario definire un disegno comuni-famiglie (o individui).

Qualora invece non sia prevista l'intervista diretta (ossia nei casi in cui le tecniche sul campo siano solo il Cati e il web), si può definire un disegno a un solo stadio di selezione, generalmente con stratificazione, che presenta noti vantaggi relativi all'efficienza delle stime.

Per quanto riguarda la definizione della numerosità campionaria, quando per un'indagine multi-tecnica concorrente si definiscono disegni campionari differenti sulle sottopopolazioni individuate per le diverse tecniche, è necessario tenere in considerazione il fatto che i due sotto-campioni dovranno essere utilizzati congiuntamente per la produzione delle stime dell'indagine. Di conseguenza, la definizione della numerosità campionaria complessiva e delle numerosità campionarie da assegnare al sottocampione di ciascuna tecnica, dovrà seguire gli usuali criteri di efficienza delle stime come se fosse un unico campione. Ciò significa in concreto che è opportuno che la dimensione del campione assegnato a ogni tecnica sia proporzionale al peso della corrispondente sottopopolazione nella popolazione complessiva e che l'allocazione del campione negli strati sia coerente con tale principio. Inoltre, le strategie da mettere in atto per ottenere le numerosità campionarie attese (sovra-campionamento o sostituzioni) possono essere definite e gestite indipendentemente per i due campioni.

Per le indagini multi-tecnica sequenziali invece, la numerosità campionaria viene definita seguendo gli usuali criteri basati sull'efficienza delle stime nel rispetto dei vincoli di costo. Una particolare attenzione deve essere tuttavia dedicata alla definizione del sovra-campionamento necessario per ottenere la numerosità campionaria attesa, che viene definito sulla base dei tassi attesi di risposta delle diverse tecniche utilizzate in sequenza.

3.3 Il disegno del questionario

Come definito nel Capitolo 2 per *mode effect* si intende l'insorgere di differenze sistematiche nei dati raccolti dovute all'uso di tecniche differenti.

Gli accorgimenti da prendere nel disegno del questionario quando si adottano soluzioni multi-tecnica sono funzionali a contenere la componente del *mode effect* relativa al *mode measurement effect*, ossia agli errori di misurazione, dovuti alla tecnica, che introducono una distorsione sui dati raccolti.

Il *mode measurement effect* è sempre presente nei dati raccolti anche quando viene utilizzata una singola tecnica di raccolta dati, in quanto ogni tecnica ha caratteristiche e peculiarità proprie che non sempre la rendono perfettamente aderente alle esigenze dell'indagine.

L'uso della multi-tecnica, quindi, potrebbe offrire l'opportunità di compensare i punti di debolezza propri di ciascuna tecnica con i punti di forza propri delle altre (De Leeuw, 2005), purché si presti attenzione a non incrementare l'errore di misura aggiungendo 'un mix' di

questi. Un modo per evitare questo rischio è costruire strumenti di raccolta dati che siano insensibili alla tecnica usata, ossia siano *mode insensitive* (Roberts, 2007).

La costruzione/progettazione dei questionari d'indagine è, come già detto, influenzata dalla tecnica di raccolta dati e di conseguenza anche dalle strategie multi-tecnica che si intendono mettere in campo. Ogni strategia influenzerà in modo diverso il disegno del questionario e, per descrivere come, è conveniente dapprima suddividere le diverse strategie in due grandi gruppi (De Leeuw, 2009):

- da una parte le strategie costituite da una tecnica principale e da altre utilizzate come modalità ausiliarie o secondarie. Questo è il caso, ad esempio, di quando si utilizza la tecnica Cawi perché adatta alla popolazione di indagine e il Cati per il recupero delle non risposte residuali (colmare alcuni strati);
- dall'altra le strategie in cui tutte le tecniche di rilevazione sono ugualmente importanti.

La progettazione del questionario segue regole e principi diversi in funzione di questi due grandi gruppi. Se si adottano le strategie del primo gruppo, allora il questionario può essere progettato seguendo l'approccio del *mode-enhancement construction* (Dillman, 2000), secondo il quale lo strumento di rilevazione è ottimizzato in base alle caratteristiche della tecnica principale per sfruttarne al meglio le potenzialità, massimizzando così la qualità dei dati. In generale, è possibile affermare che un tale approccio può essere adottato quando è più importante massimizzare la qualità dei dati piuttosto che l'equivalenza dei dati raccolti con varie tecniche (Betts *et al.*, 2010).

Se si sceglie di adottare le strategie del secondo gruppo (tutte le tecniche ugualmente importanti) il disegno del questionario può seguire tre approcci diversi:

- lo *unimode or unified approach*,
- il *mode specific approach*,
- il *generalized approach*.

Secondo l'*unimode or unified approach* (Dillman, 2000 e 2009)² i questionari vengono progettati in modo tale che siano il più possibile simili (formulazione dei quesiti e gestione del flusso) così da offrire gli stessi stimoli in ciascuna tecnica. L'obiettivo è quello di contrarre la variabilità tra le risposte degli intervistati con le varie tecniche contenendo, di conseguenza, il *mode effect*. I principali criteri su cui si basa questo approccio sono:

- utilizzare lo stesso (o quasi lo stesso) testo del quesito in tutte le tecniche;
- utilizzare le stesse opzioni di risposta per tutte le tecniche e inserirle nel testo del quesito;
- sviluppare le stesse istruzioni per le condizioni di salto/flusso;
- per ciascuna tecnica, invertire l'ordine degli item di risposta in metà dei questionari;
- utilizzare le stesse etichette nelle scale;
- offrire o non offrire l'opzione 'non so/non ricordo' in modo analogo fra le tecniche;
- evitare l'uso di cartellini nelle interviste faccia-a-faccia se usate in combinazione con le interviste telefoniche.

In questo caso, sulle scelte di disegno da effettuare influirà la tecnica che pone maggiori limitazioni. Ad esempio nel caso di un mix Cawi-Cati la tipologia e quantità di controlli saranno stabiliti in funzione della tecnica web.

Il vantaggio dell'*unimode approach* (se il presupposto della sua costruzione è vero) è una semplificazione delle procedure di costruzione dei questionari elettronici, della fase di

² Nell'edizione del 2000 del libro di Dillman, 'Mail and Internet Surveys - The Tailored Design Method' è possibile trovare un elenco di nove linee guida all'approccio unimode che sono poi state riviste nell'edizione del 2009.

data entry, della fase di controllo e correzione. Lo svantaggio risiede nel fatto che, in quanto tale, non permette di sfruttare al meglio le peculiarità di ciascuna tecnica soprattutto quando il mix è composto da tecniche basate su questionari cartacei ed elettronici, oppure su comunicazioni visive (Capi-Cawi) e orali (Cati). Quindi non sempre il mix delle tecniche è tale da consentire il rispetto di tutti i principi dell'*unimode approach* che, inoltre, potrebbe non rappresentare una soluzione adeguata ai problemi di *social desirability*³ dovuti alla presenza di un intervistatore (Dillman *et al.*, 2005), o di *satisficing*⁴ dovuti al contrario all'assenza di un intervistatore. Inoltre potrebbe non essere efficace per contenere il *primacy/recency effect*⁵ nel caso di indagini panel dove l'ordine delle modalità non può essere modificato onde evitare effetti confondenti sui reali cambiamenti del fenomeno. In queste situazioni potrebbe essere auspicabile adottare uno dei due approcci descritti di seguito.

Il primo è il *mode specific approach*, noto anche come *maximisation method*, secondo il quale il questionario è ottimizzato per ogni tecnica in modo da ridurre la componente *measurement effect* dell'errore di misura dovuta a ciascuna di esse. L'ipotesi sottostante l'approccio è che, riuscendo a ridurre, fino ad annullarla, la componente *measurement effect* di ciascuna tecnica sui dati rilevati, se ne aumenti l'accuratezza (i dati saranno il più possibile vicini al valore vero (Bethlehem *et al.*, 2012) e si rendano direttamente confrontabili perché differiranno solo per effetto del caso.

Qualora il mix sia rappresentato, ad esempio, da tecniche assistite da computer (*Computer Aided/Assisted Interviews*, Cai) con o senza il rilevatore, il disegno dei questionari dovrebbe seguire i seguenti principi:

- 1) per le tecniche Cai con rilevatore:
 - massimizzare la quantità di controlli;
 - limitare l'uso dei 'non so/non ricordo';
 - ruotare le alternative di risposta che saranno lette al rispondente;
 - usare testi di rilancio/*probing* per le domande a risposta multipla;
 - usare testi di rilancio/*probing* per le domande a testo libero;
 - usare sistemi di codifica assistita anche avanzati;
- 2) per le tecniche Cai senza rilevatore:
 - minimizzare la quantità di controlli;
 - limitare l'uso dei 'non so/non ricordo' ai quesiti non fondamentali per l'indagine;
 - ruotare le alternative di risposta;
 - usare domande chiuse;
 - usare sistemi di codifica assistita solo se *user-friendly*;
 - usare istruzioni molto sintetiche;
 - ottimizzare il *layout* del questionario sapendo che di fatto rappresenta una "guida" alla compilazione (Brancato *et al.*, 2006).

3 Si riferisce alla tendenza dei rispondenti a dare risposte socialmente desiderabili invece di scegliere quelle che riflettano le loro vere opinioni o stati, al fine di fornire una rappresentazione di loro stessi più favorevole agli occhi di chi li ascolta (Wiley online library).

4 Con il termine 'satisficing' si intende una strategia decisionale mirata a un risultato soddisfacente o adeguato, piuttosto che alla soluzione ottimale. Questo perché mirare alla soluzione ottimale può richiedere "un'inutile" perdita di tempo, energia e risorse. La teoria del 'satisficing' trova applicazione in diversi settori, tra cui l'economia, l'intelligenza artificiale e la sociologia. Adottare un atteggiamento satisficing implica che un consumatore, nell'affrontare una gamma di scelte per una specifica esigenza, seleziona un prodotto o un servizio "abbastanza buono", piuttosto che sforzarsi per trovare la scelta migliore o ottimale.

5 Il primacy e il recency effect rappresentano la tendenza del rispondente a scegliere più spesso, rispettivamente, il primo o l'ultimo item di una lunga lista di modalità di risposte.

Lo svantaggio del *mode specific approach* risiede nella difficoltà di raggiungere l'obiettivo che si prefigge: l'assunto su cui si basa richiede infatti una ben strutturata fase di pre-test e di indagini pilota (con notevole innalzamento dei costi) che ne dimostri la veridicità. Inoltre, e specialmente nelle indagini continue, qualsiasi cambiamento deve essere implementato e soprattutto testato nei diversi questionari che il metodo genera (Bethlehem e Biffignandi, 2012).

Il secondo approccio, alternativo all'*unimode*, è il *generalized approach* (de Leeuw, 2005), che può essere considerato un caso particolare del *mode specific approach*, secondo il quale il questionario è progettato appositamente per essere diverso per ogni tecnica di rilevazione al fine di raggiungere l'equivalenza cognitiva dello stimolo percepito. Questo perché, usando le parole della de Leeuw, "... lo stesso stimolo offerto non è necessariamente lo stesso stimolo percepito" (de Leeuw, 2015).

L'approccio generalizzato non è un approccio facile, poiché la sua implementazione richiede che i ricercatori identifichino chiaramente le differenze tra le tecniche che influenzano il processo cognitivo della risposta e dimostrino, attraverso test cognitivi, che i diversi formati di domanda generino risposte equivalenti (de Leeuw, 2005).

Un esempio di *generalized approach* si può trovare nel lavoro di Statistics Finland relativo alla sperimentazione web dell'indagine Forze di Lavoro, per la quale è stata utilizzata la tecnica mista Cati-Cawi (Järvensivu, 2015). La sperimentazione si è focalizzata sul quesito sulle ore lavorate (Figura 3.2): sono state somministrate in Cawi due versioni di questionario, una con il quesito identico a quello utilizzato per il Cati e una con una formulazione appositamente pensata per la tecnica Cawi e totalmente diversa da quella dell'indagine telefonica. Il test cognitivo ha dimostrato come l'equivalenza di stimolo percepito sia stata raggiunta con la seconda formulazione, come indicato nelle figure sottostanti.

Figura 3.2 - Approccio *unimode*: formulazione Cawi identica alla Cati

The image shows a survey interface for 'Työvoimatutkimus 2013' (Labor Force Survey 2013) from Tilastokeskus/Statistikcentralen. The question asks for the number of hours worked in the main job during the week of 7.10.-13.10.2013. The input fields show '37' for hours and '30' for minutes. A legend at the bottom indicates 'Tuntia = Hours' and 'Minuuttia = Minutes'. The interface includes navigation buttons: '< Takaisin', 'Jatka >', and 'Jatka myöhemmin'.

3. Il disegno delle indagini multi-tecnica

I risultati del test cognitivo sulla formulazione identica a quella del Cati hanno mostrato che il quesito è difficile da compilare sia per le difficoltà del calcolo delle ore totali sia per problemi di ricordo e che inoltre sia frequente sbagliare considerando non le ore lavorate, ma quelle stabilite dal contratto. Tali difficoltà possono, invece, essere attenuate nel Cati dalla presenza di un rilevatore (Figura 3.3).

Figura 3.3 - Approccio *generalizzato*: formulazione Cawi totalmente diversa dal Cati

Tilastokeskus
Statistikcentralen

Työvoimatutkimus 2013

Montako tuntia työskentelit viikolla 7.10.-13.10.2013 (7.10.-13.10.2013)?

- Kirjoita taulukkoon tekemiesi työtuntien määrä. Jos et muista tarkkaa aikaa, arvio riittää.
- Laski mukaan mahdolliset ylityöt, kotona tehty ansiotyö ja päivystys työpaikalla.
- Älä merkitse ruokataukoja, tilapäisiä tunninpoissaoloja sekä sairaus- ja lomapäivien tunteja.
- Älä merkitse kotona tapahtuvaa päivystystä tai kotitaloustyötä.

	Päivämäärä	Tunnit	Minuutt
maanantai	7.10.2013	5	5
tiistai	8.10.2013	6	33
keskiviikko	9.10.2013	5	
torstai	10.10.2013	3	56
perjantai	11.10.2013	4	
lauantai	12.10.2013		
sunnuntai	13.10.2013		
Yhteensä		24	34

Monday, Tuesday...
Yhteensä = In total

How many hours did you work in your main job during the week 7.10.-13.10.2013?

< Takaisin Jatka > Jatka myöhemmin

In questo caso il test cognitivo ha dimostrato come la domanda organizzata per giorni della settimana abbia semplificato il calcolo delle ore totali, delegando questa funzione allo strumento, e abbia inoltre facilitato il ricordo esatto delle ore lavorate facilitando il ricordo delle ore di assenza dal lavoro e di straordinario.

3.4 Le strategie per la massimizzazione dei tassi di risposta

Ormai da oltre 30 anni ricercatori e studiosi assistono, e provano ad analizzare e porre rimedio, al lento ed apparentemente inesorabile declino dei tassi di risposta nelle indagini su individui e famiglie. Nel 2002 de Leeuw e de Heer hanno pubblicato uno studio in cui stimano quanto forte e vario sia tale trend nelle indagini statistiche ufficiali, comparando i tassi di risposta, di rifiuto e di mancato contatto nelle serie storiche di 10 indagini in 16 Paesi diversi, inclusa l'Italia. L'analisi mostra un aumento nel tempo sia dei mancati contatti sia dei rifiuti (i primi aumenterebbero di uno 0,2 per cento ogni anno, i secondi di uno 0,3 per cento). Pur partendo da livelli differenti, i tassi di mancato contatto mostrano nei diversi Paesi un andamento simile; la situazione è invece differente per i tassi di rifiuto, per i quali si registrano tendenze diverse da Paese a Paese e per le diverse indagini.⁶

6 Dillmann e Charley-Baxter (2000) in un'analisi su dati di un'indagine postale condotta nei precedenti 12 anni, mettono in discussione la "retorica del declino dei tassi di risposta", attribuendo ad altri fattori il numero crescente di rifiuti: l'uso di questionari sempre più lunghi e la selezione di campioni sempre meno interessati al tema.

I tentativi e le strategie messe in atto per cercare di arrestare, se non di invertire, tale dinamica, devono spesso fare i conti con la necessità di contenere i costi delle indagini senza per questo intaccare la qualità dei dati. È anche alla luce di questo contesto che va quindi interpretato il sempre maggiore ricorso ai disegni multi-tecnica, e all'adozione della raccolta dei dati tramite web.

Come già accennato in precedenza, infatti, ogni modalità di raccolta dati consente di raggiungere sottoinsiemi diversi di popolazione (copertura) e di ottenere tassi di risposta più o meno alti.

Sono in genere le tecniche più costose a garantire performance migliori sotto questo profilo. Le interviste faccia a faccia, notoriamente più costose, in genere hanno rifiuti più bassi di quelle telefoniche, che a loro volta ottengono risultati migliori delle interviste web.

È noto (Rubin *et al.*, 2002) che il fenomeno della mancata risposta può generare distorsioni importanti sulle stime finali dei fenomeni oggetto di indagine, in particolare quando il meccanismo generatore della mancata risposta è di tipo non casuale (*Missing not at Random, MNAR*).⁷ Con specifico riferimento alle tecniche di raccolta dei dati, Marradi sostiene che “*la mera appartenenza di un soggetto a certe categorie riduce oppure accresce a priori, e in modo significativo, la probabilità che egli sia reperibile al telefono o al suo indirizzo nelle ore in cui può cercarlo un intervistatore, che conceda l'intervista, che rinvii il questionario postale, etc.*” (1999, p.14).

Il ricorso a disegni multi-tecnica è finalizzato, in presenza di budget sempre più ridotti, a trovare la combinazione ideale per contenere la diminuzione dei tassi di risposta e dei diversi errori non campionari. La maggior parte degli studi sull'applicazione della multi-tecnica in effetti riporta non solo un aumento del tasso di risposta, ma anche una riduzione degli errori dovuti alla mancata risposta (Gallagher *et al.*, 2005 per un esempio sulle famiglie; Sala e Lynn, 2009, per una indagine sulle imprese).

In genere si adottano disegni con multi-tecniche sequenziali, dove la prima modalità di raccolta usata è quella che ha il minor costo per contatto e, di seguito, si cerca di recuperare la mancata risposta con una seconda modalità di raccolta (cfr. anche Capitolo 2, e Paragrafo 3.1.2). In letteratura, una delle prime applicazioni della multi-tecnica prevedeva un primo invio postale del questionario e un successivo follow-up con interviste faccia a faccia ai non rispondenti (Alexander e Wetrogan, 2000). I primi studi sul tema risalgono alla fine degli anni '60, quando Hochstim (1967) analizzò i risultati di tre indagini che prevedevano in sequenza diversa l'intervista faccia a faccia, telefonica e postale: i risultati furono sostanzialmente comparabili in quanto a tassi di risposta e risultati tematici; ma, a parità di risultati, i costi della strategia che partiva con il questionario postale erano considerevolmente più bassi. Studi successivi hanno messo in discussione l'ottimismo dei primi risultati evidenziando l'introduzione di possibili errori di misurazione legati all'uso di tecniche diverse.

Disegni con multi-tecniche concorrenti, come evidenziato già nel Capitolo 2 e nel Paragrafo 3.1.2, sono invece spesso adottati per ovviare a problemi di copertura. Contrariamente a quanto ci si potrebbe aspettare, lasciare al rispondente la scelta della modalità di risposta (ad esempio via web o tramite intervista telefonica) non sembra portare benefici sul tasso di risposta, a prescindere da quali siano le opzioni messe a disposizione (de Leeuw,

⁷ Va posta attenzione alla relazione esistente fra il tasso di risposta e l'errore derivante dalla mancata risposta (*response bias*). Se è vero che esiste una relazione fra i due fenomeni, è anche vero che l'aumento del tasso di risposta non solo non è una garanzia di una diminuzione dell'errore ad esso correlato, ma potrebbe in alcuni casi anche aumentarlo laddove siano solo alcune fasce di popolazione ad essere raggiunte (Groves, 2006). Sul tema generale delle mancate risposte si veda anche Groves e Lyberg, 1988; Groves, 1989.

2005). Un beneficio tuttavia persiste se, a parità di tasso di risposta, l'introduzione di una seconda tecnica riesce a contenere i costi, velocizzare i tempi di rilevazione sul campo, e aumentare la soddisfazione del rispondente.

Il ricorso a più tecniche di rilevazione è spesso importante anche per ridurre la mancata risposta su singoli item: i questionari auto-amministrati sono da subito apparsi più adatti a rilevare domande sensibili rispetto alle modalità che prevedono l'intervento di un intervistatore. Anche nel passaggio dal questionario cartaceo a quello elettronico tale vantaggio sembra confermato, a condizione di offrire al rispondente tutte le garanzie possibili sull'anonimato e la sicurezza nella trasmissione delle sue risposte (Tourangeau e Smith, 1996). Per questo motivo è prassi consolidata ricorrere alla multi-tecnica nelle indagini che vogliono rilevare informazioni su temi delicati o socialmente desiderabili in cui l'intervistatore, dopo aver condotto parte dell'intervista, lascia al rispondente un questionario cartaceo oppure un dispositivo per la compilazione elettronica dello stesso.

3.4.1 Le "leve" per aumentare i tassi di risposta

Molteplici sono gli elementi che intervengono nel condizionare i tassi di risposta finali. Di seguito vengono elencati i principali, con particolare attenzione al caso in cui la prima tecnica cui si ricorre in un disegno di indagine multi-tecnica è quella web (come ben descritto in Callegaro *et al.*, 2015). Solo alcuni elementi rappresentano delle "leve" su cui il ricercatore può intervenire; altri sono infatti elementi "dati", che difficilmente possono essere modificati. Rientrano in quest'ultima classe:

- 1) l'importanza dell'istituto che conduce l'indagine o che l'ha commissionata: tanto maggiore è la reputazione del committente, in termini di affidabilità, notorietà, capacità riconosciuta di proteggere i dati e le informazioni fornite, tanto maggiore sarà la propensione dei soggetti a rispondere. È perciò importante che sia sempre ben chiaro e visibile il logo del committente, almeno sulla lettera di invito e sulla prima pagina del questionario. Non sembra invece avere effetti sul tasso di risposta la riproduzione del logo su tutte le pagine (Heerwegh e Loosveldt, 2006). Più utile risulta invece un'efficace campagna di comunicazione, magari mirata ad alcuni target particolari, che sensibilizzi sul tema oggetto di indagine e metta in rilievo la sua rilevanza per la comunità;
- 2) la salienza, ovvero l'importanza percepita del tema per i soggetti intervistati, è l'altro elemento cruciale nella partecipazione ad una rilevazione. I soggetti più interessati al tema non sono solo quelli che rispondono più facilmente, ma anche che compilano con più attenzione e interrompono meno frequentemente. Enfatizzare troppo l'importanza del tema per aumentare i tassi di risposta tuttavia potrebbe aumentare, come rovescio della medaglia, l'effetto selezione.

Fra gli aspetti su cui il ricercatore può intervenire rientrano le decisioni generali sul disegno del questionario, gli incentivi, le modalità di contatto:

- 1) sono diverse le decisioni generali sulla costruzione del questionario che impattano sul tasso finale di risposta: la lunghezza, la difficoltà/semplificata di navigazione, il layout grafico. Queste non incidono direttamente sulla decisione di partecipare, ma hanno un forte impatto sul numero di interruzioni e di abbandoni (Callegaro *et al.*, 2005; Vicente e Reis, 2010);
- 2) gli incentivi: poco usati nelle indagini condotte in Italia, il ricorso agli incentivi appare invece una pratica abbastanza diffusa all'estero e, in generale, nei panel. Alcuni au-

tori tuttavia invitano a riflettere sui rischi che tale strategia comporta, dal momento che l'incentivo attira soprattutto particolari fasce meno abbienti di popolazione, e rischia di aumentare, assieme ai tassi di risposta, anche l'effetto selezione (Göriz, 2008; Parsons e Manierre, 2013) e l'effetto misurazione, poiché i rispondenti tendono a dare risposte acquiescenti (Comley, 2000);

- 3) le modalità di contatto con il rispondente. In genere il rispondente a una rilevazione web viene contattato in tre momenti diversi: riceve un avviso (una lettera informativa) prima dell'avvio dell'indagine; riceve un avviso di avvio dell'indagine via e-mail, e poi uno o più solleciti nel corso dell'indagine.

In generale, anche in sistemi di raccolta *unimode*, un contatto iniziale proposto con una modalità differente riesce a incrementare i tassi di risposta. Ne sono esempi l'invio postale della lettera informativa alle famiglie prima di una intervista telefonica (che in alcuni casi aumenta la partecipazione anche del 50 per cento, cfr. Edwards *et al.*, 2009), o una telefonata alle imprese prima dell'invio del questionario cartaceo o elettronico.

Non sempre è possibile scegliere quale modalità di contatto adottare: per la popolazione generale spesso si hanno a disposizione solo gli indirizzi postali, a volte i numeri di telefono, raramente le e-mail. Su altri sottoinsiemi di popolazione invece si hanno a disposizione soltanto le e-mail. Se è disponibile più di una possibilità è bene combinarle in modo da sfruttare i vantaggi di ciascuna: una lettera cartacea infatti accresce la percezione dell'importanza dell'indagine e comunica meglio il prestigio del committente. Rispetto ad una e-mail, non rischia di essere persa nello *spam*, e non genera timori legati a virus o tentativi di *phishing*.

Se è prevista la presenza di un intervistatore, il momento più favorevole al contatto sembra essere il pomeriggio infrasettimanale per le interviste faccia a faccia e la sera o i week-end per quelle telefoniche. Gli inviti via e-mail dovrebbero essere mandati quando il rispondente non è troppo impegnato: per questo si suggerisce un giorno intermedio della settimana, ad esempio il mercoledì; tuttavia l'analisi dell'andamento dei tassi di risposta in numerose indagini web mostra picchi ricorrenti il lunedì, giorno in cui si suggerisce di far partire la rilevazione sul campo.

Nella prima lettera (o e-mail) che riceve il rispondente deve essere spiegato in modo chiaro cosa è richiesto di fare, qual è l'oggetto dell'indagine, quali le procedure seguite per la selezione del campione, come si garantirà la privacy delle risposte, come contattare (telefonicamente, via e-mail etc.) il committente, le modalità di risposta previste. Gli inviti possono essere personalizzati: la personalizzazione tende a favorire la partecipazione, ma al contempo sembra indurre risposte socialmente più desiderabili poiché riduce la percezione di anonimato.

Le lettere devono essere corte, contenere solo gli elementi essenziali posti in modo ben visibile e scritti in modo chiaro. Si suggerisce di non superare una pagina per gli inviti cartacei e ridurre la lunghezza per quelli via mail (Dillman *et al.*, 2009).

La prima indagine tramite questionario postale di cui si ha documentazione risale al 1788, quando Sir John Sinclair per raggiungere il 100 per cento delle risposte dai parroci scozzesi inviò 23 solleciti. Oggi, in molti casi, due contatti complessivi con i rispondenti sono ritenuti sufficienti; il primo sollecito aggiunge dal 10 al 20 per cento delle risposte, mentre dal secondo in poi sembra perdersi efficacia con incrementi attorno all'1 per cento; secondo altri studi, invece, un secondo sollecito riesce a incrementare anche dell'8 per cento (Kaplowitz *et al.*, 2012) i tassi di risposta. Studi recenti tuttavia mostrano che tali risposte, giunte tardivamente e dopo diversi solleciti, sono meno complete o accurate.

Inoltre, se la strategia di contatto resta invariata nel corso della rilevazione, allora anche l'aggiunta di ulteriori risposte non sembra comportare differenze nelle stime finali prodotte.

Se con i solleciti inviati via posta in genere si suggerisce un intervallo di una settimana fra due invii, via mail appare più efficace – alla luce dell'analisi dell'andamento dei picchi di risposta in corrispondenza dell'invito e del loro veloce declino nei due giorni successivi – un sollecito recapitato a distanza di quattro giorni dal precedente. Si ricorda inoltre che un numero eccessivo di solleciti potrebbe infastidire il rispondente.

3.4.2 Strategie flessibili: responsive e adaptive design

Il sempre più difficile coinvolgimento dei rispondenti e la crescente disponibilità di paradati (cfr. Paragrafo 2.5) derivanti dall'uso di questionari elettronici ha portato diversi studiosi a immaginare modalità di raccolta dei dati meno rigide, non definite e bloccate a priori o identiche per tutti i casi. Si parla in questi casi di strategie flessibili di indagine, con riferimento a due differenti modalità: il *responsive* e l'*adaptive design*.⁸ Nel primo caso il disegno di indagine prevede una organizzazione per fasi, e le modifiche sono possibili e differenziate nel passaggio da una fase all'altra. Come spiega Groves: “*a design phase is a time period of a data collection during which the same set of sampling frame, mode of data collection, sample design, recruitment protocols and measurement conditions are extant*” (2006, pp. 440). In un disegno d'indagine responsive, quindi, è necessario:

- a. individuare a priori un insieme di caratteristiche della strategia di indagine che potenzialmente potrebbero influenzare costi e livello d'errore;
- b. indicare degli indicatori del costo sostenuto e delle caratteristiche dell'errore e monitorarli già dalla fase iniziale di raccolta;
- c. modificare le strategie di indagine nelle fasi successive sulla base delle regole decisionali relative al bilanciamento di costi ed errori;
- d. integrare i dati ottenuti nelle diverse fasi.⁹

Gli indicatori utilizzati per definire la strategia da adottare derivano da paradati: “*Paradata can be proxy indicators of costs or errors. They can include data collection administrative data, such as records of contact attempts on sample cases, travel distance and time to reach sample cases and hours spent by interviewers on different tasks during interviewing*” (Groves e Heeringa, 2006, pp. 448).

Ciascuna delle leve indicate precedentemente, quindi, fa riferimento a una specifica fase del disegno di indagine. A conclusione di una singola fase, i risultati dell'analisi degli indicatori previsti è utilizzata per decidere la strategia della fase successiva, che può differire dalla precedente per ampiezza del campione (ad esempio si può decidere che solo un sottoinsieme dei non rispondenti passi alla fase successiva), tecnica di rilevazione, ampiez-

⁸ Per un'interessante rassegna della letteratura su responsive e adaptive design si veda Tourangeau et al., 2017.

⁹ Un classico esempio, riportato da Groves e Heeringa (2006), di disegno responsive è il Chicago Mind and Body Survey, un'indagine epidemiologica condotta con un campionamento probabilistico a due stadi sulle famiglie di Chicago. Nella prima fase dell'indagine sono state somministrate 2100 interviste face to face della durata di circa due ore. Era previsto un incentivo monetario di 60\$ e la gestione dei solleciti è stata lasciata alla discrezione dell'intervistatore. Considerando tempi e costi, i tassi di risposta e il numero di interviste sono stati inferiori alle aspettative. Nella seconda fase è stato estratto un sotto-campione dei non rispondenti stratificato in base alla propensione a rispondere. La propensione è stata valutata in base alle impressioni degli intervistatori. È stata quindi adottata una frazione di campionamento doppia nello strato con maggiore propensione alla risposta rispetto allo strato con propensione minore. Inoltre l'incentivo è stato aumentato da 60\$ a 100\$ ed è stato definito un numero fisso di solleciti. Nella terza fase è stato tentato un nuovo contatto con tutti i non rispondenti della seconda fase, aumentando l'incentivo a 150\$.

za degli incentivi, modalità di contatto, ma anche lunghezza del questionario (ad esempio se si decide di utilizzare una versione short per i rispondenti più difficili da coinvolgere).

In molti aspetti simile al *responsive design* è quello che Schouten *et al.*, (2013) chiamano *adaptive design*: in questo caso il punto centrale non è l'organizzazione in fasi ma il ricorso a informazioni sul campione note a priori (tratte da dati amministrativi, o da indagini precedenti) finalizzate a suddividere lo stesso in gruppi caratterizzati da una differente propensione alla risposta. Per ciascun gruppo sono quindi definite strategie differenti.

3.5 La prevenzione dell'effetto tecnica tramite pre-test e indagini pilota

Per contenere la presenza di errori di misura dovuti all'uso di diverse tecniche tra cui quella web, una strategia raccomandabile è quella di agire in maniera 'preventiva', raccogliendo informazioni attraverso pre-test o indagini pilota prima del disegno e della realizzazione della rilevazione vera e propria.

Dalla diffusione delle indagini web, iniziata circa 20 anni fa, ad oggi sono stati realizzati molti studi per trovare il disegno più efficace e il modo migliore per somministrare quesiti di vario genere attraverso questionari Cawi (Couper, 2008, Bethlehem e Biffignandi 2012, Tourangeau *et al.*, 2013, Callegaro *et al.*, 2015). L'ampia letteratura a disposizione risulta dunque fondamentale per pianificare indagini *mixed-mode* che prevedano l'uso del web. Spesso tuttavia i risultati sono diversi da Paese a Paese e da indagine a indagine, specie quando si vada ad inserire la nuova modalità di raccolta dati in un processo già esistente. Realizzare esperimenti ed analisi incentrate sulle indagini specifiche, pertanto, è sempre di notevole aiuto per valutare a priori gli effetti che le diverse tecniche avrebbero sulla raccolta dei dati complessivi e per attuare di conseguenza gli opportuni correttivi.

Come scrivevano Presser *et al.* (2004), un pre-test è "l'unico modo per valutare in anticipo se un questionario causa problemi per gli intervistatori o per i rispondenti". Nel caso di test Cai è necessario anche verificare che il questionario elettronico funzioni come disegnato e che tutta l'architettura informatica funzionerà e reagirà in maniera "efficace" ad eventi inaspettati.

Nel caso di indagini multi-tecnica, l'obiettivo principale di un pre-test è quello di verificare 'l'equivalenza' dei questionari che verranno somministrati con le diverse tecniche scelte per la rilevazione. A seconda dell'approccio scelto per il disegno dei questionari, come si è detto nel Paragrafo 3.3, questi potrebbero essere simili nella formulazione (*unimode approach*) o potrebbero utilizzare formulazioni/gestione del flusso disegnati specificatamente in base alla tecnica (*mode specific* o *generalized approach*). In entrambi i casi, il processo cognitivo indotto nel rispondente al momento della formulazione della risposta potrebbe essere differente in base ai diversi stimoli (orali, visivi, di definizioni etc.) forniti al momento della compilazione e portare così a errori di misurazione nelle stime. Con l'obiettivo quindi di prevenire e minimizzare tali errori risulta fondamentale progettare ed effettuare un pre-test dei questionari finalizzato a verificare dove, e se possibile per quali motivi, si verificano gli eventuali effetti della tecnica.

Il pre-test può essere organizzato a ridosso della rilevazione, in una fase precedente all'avvio di quella estensiva (*stand alone*), oppure, per limitare i costi, può essere realizzato all'interno della rilevazione stessa (*embedded*); in quest'ultimo caso però i risultati verranno sfruttati per l'edizione successiva dell'indagine.

Gli aspetti da testare sono molti e, in base alle tecniche utilizzate e alle caratteristiche dell'indagine, occorrerà prestare più attenzione ad alcuni elementi piuttosto che ad altri. Fondamentale, ad esempio, è la presenza, in una o più tecniche, dell'intervistatore: questo infatti, come si è detto nel Paragrafo 3.1, se da un lato può dare delucidazioni o spingere verso una risposta anche quando l'intervistato preferirebbe non rispondere, dall'altro potrebbe provocare delle distorsioni sistematiche nelle risposte (*effetto intervistatore*) (Pitrone, 2009, Roberts *et al.*, 2006).

Gli aspetti da valutare nel pre-test sono soprattutto quelli potenzialmente suscettibili di effetto tecnica, ossia che potrebbero fornire stimoli differenti al rispondente nel momento della compilazione. Due sono i punti di vista da considerare nella fase di pre-test: quello dei contenuti e quello della funzionalità/usabilità degli strumenti Cai.

Dal punto di vista dei contenuti, è utile testare:

- i quesiti sensibili e su temi delicati/riservati, particolarmente critici per le tecniche con intervistatore, in quanto suscettibili di "desiderabilità sociale" (cfr. Paragrafo 3.3);
- le variabili con lunghe liste di modalità, che potrebbero determinare un effetto *primacy recency*: la scelta del primo item avviene più frequentemente nel Cawi quando non vengono lette tutte le modalità proposte, mentre la tendenza a scegliere il più "recente" è più frequente nel Cati, quando è l'intervistatore che legge tutti gli item e per chi ascolta è più semplice ricordare l'ultimo;
- le domande che prevedono matrici, a rischio del cosiddetto effetto *satisficing*, per il quale alcuni intervistati (specialmente i più anziani, meno istruiti e meno motivati) danno la risposta 'meno faticosa' possibile, piuttosto che quella corretta che richiederebbe la massima attenzione. Tale effetto è più frequente nei questionari autocompilati, in assenza di comunicazione non verbale: nell'intervista faccia a faccia l'interesse che un bravo intervistatore è capace di comunicare con l'espressione del viso o gesti di incoraggiamento spingono l'intervistato a riflettere e cooperare, facendo il massimo di sforzo cognitivo necessario (Pitrone, 2009);
- le scale di misurazione, che, proposte con diversi layout grafici, possono non essere equivalenti nella percezione del rispondente (ad esempio quando sono visualizzate in verticale o in orizzontale);
- le classificazioni e i metodi di codifica: i quesiti che prevedono l'uso di classificazioni o codifica possono portare a risultati piuttosto diversi se auto-compilati o compilati da un rilevatore ben istruito;
- la presenza delle modalità "non so" e/o "non risponde": lo sforzo progettuale deve essere quello di valutare come renderli equivalenti nelle diverse tecniche utilizzate. Non è facile infatti trattarli nello stesso modo in quanto in un questionario cartaceo implicitamente viene data la possibilità di non rispondere, in un Cati o Capi la non risposta viene "filtrata" e minimizzata dall'intervistatore e nel Cawi invece può essere gestita in due modi: dando la possibilità di non rispondere al quesito e proseguire l'intervista oppure visualizzando le modalità "non so"/"non risponde" e prevedendo l'obbligatorietà della risposta. Questi diversi trattamenti possono indurre una risposta diversa a seconda della modalità di compilazione.¹⁰

¹⁰ Pitrone (2009) riporta nel suo libro i risultati di alcune ricerche che confrontano i dati ottenuti da interviste faccia a faccia e telefoniche (Holbrook et al., 2003) relativamente alla non risposta e ai comportamenti di *satisficing* nel caso di diverse tecniche di somministrazione del questionario: la percentuale media di "non so" passa dal 17% nelle domande somministrate faccia a faccia al 26% nei sondaggi telefonici mentre le risposte meccaniche (*satisficing effect*) a batterie di domande con la stessa modalità di risposta, passano dal 35% circa nelle interviste personali al 41% in quelle telefoniche.

Dal punto di vista dello 'strumento software', invece, è necessario effettuare dei test della funzionalità e dell'usabilità dei questionari.

Il test della funzionalità serve a valutare il corretto funzionamento tecnico degli strumenti per determinati aspetti che dipendono dalla tecnica. Ad esempio, se c'è l'ausilio del computer, si dovranno testare l'aspetto grafico (*visual design*), la presenza di istruzioni e definizioni, le regole e i controlli *hard* e *soft*,¹¹ i tempi di caricamento, il display corretto; se poi si tratta di una indagine web si dovranno valutare, ad esempio, anche l'accesso e la navigazione (con i pulsanti "avanti" e "indietro"), il salvataggio e l'invio, il tutto possibilmente con browser diversi e con dispositivi diversi (*smartphone, tablet, etc.*).

Una volta che gli strumenti sono considerati adeguati, occorre effettuare il test dell'usabilità dei questionari con dei potenziali rispondenti. L'obiettivo è quello di verificare che i questionari siano semplici da utilizzare e da compilare, siano auto-esplicativi (le istruzioni e le definizioni sufficienti e comprensibili in tutte le tecniche, per il rispondente o i rilevatori) e quindi risultino il più possibile "invitanti" per la compilazione.

Una strategia di pre-test molto utilizzata per la verifica dei contenuti e la formulazione delle domande è il *cognitive interviewing*, che si basa su una valutazione dei quesiti attraverso la ricostruzione del complesso processo cognitivo che determina la risposta (Willis, 1994). Il processo di risposta, infatti, inizia per l'intervistato con la comprensione della domanda, passa attraverso il recupero delle informazioni nella memoria e nelle proprie conoscenze, l'integrazione di informazioni diverse per arrivare ad un giudizio complessivo, e infine alla formulazione della risposta finale (Tourangeau, 2000). La comprensione di tale processo, secondo la prospettiva Casm (*Cognitive aspects of survey methodology*) alla base della *cognitive interview*, è fondamentale per disegnare i quesiti in modo da ridurre le fonti di errore nella risposta.

La *cognitive interview* viene definita come una "espansione" della usuale intervista in una indagine (Willis, 2004). A differenza di quest'ultima, però, l'intervista viene condotta da un intervistatore specializzato e istruito in maniera specifica che somministra i quesiti da testare in un laboratorio ad hoc. Inoltre, si applicano diverse tecniche, chiedendo al rispondente di riportare la risposta secondo specifiche modalità¹² (Willis, 1994).

Negli ultimi decenni sono state sviluppate anche altre tecniche di pre-test del questionario, come ad esempio il *verbal interaction coding* e l'analisi della conversazione, la *response latency*, le vignette, gli esperimenti su piccola scala e i modelli statistici per la valutazione dei questionari¹³ (Presser *et al.*, 2004; Tourangeau *et al.*, 2000; Palmieri, 2013; Biemer, 2004).

11 I controlli per gli errori soft sono segnalazioni di plausibilità di situazioni rare ma non necessariamente errate e possono essere annullati; quelli hard, invece, si riferiscono a situazioni impossibili e quindi non prevedono l'annullamento ma sono da correggere assolutamente se si vuole proseguire l'intervista.

12 Si tratta principalmente di due modalità: 1. "Think aloud", in cui l'intervistatore legge il quesito al soggetto che deve rispondere "pensando ad alta voce", ossia descrivendo il processo che lo guida alla propria risposta. L'intervistatore interviene pochissimo, in modo da non influenzare il suo comportamento; 2. "Verbal probing", dove l'intervistatore chiede informazioni specifiche sul quesito posto (che cosa significa per il rispondente un certo termine, oppure se può riformulare il quesito con parole proprie, o ancora quanto è stato facile o difficile arrivare alla risposta).

13 In particolare, con il verbal interaction coding e l'analisi della conversazione si studia l'interazione tra intervistatore e intervistato; la response latency si basa sull'analisi del tempo trascorso tra la lettura della domanda e la formulazione della risposta come indicatore della difficoltà di comprensione del quesito; le vignette, attraverso brevi storie e scenari con situazioni e personaggi ipotetici, sono utilizzate per lo studio di valori e atteggiamenti o per cogliere errate interpretazioni sul senso delle domande e il significato di singoli termini; gli esperimenti su piccola scala prevedono la somministrazione di questionari diversi a gruppi differenti considerati equivalenti secondo alcune caratteristiche; infine, i modelli statistici per la valutazione dei questionari mirano a identificare con tecniche statistiche i quesiti affetti da errori di misura.

Le tecniche citate portano in alcuni casi a risultati qualitativi (osservazione di comportamenti e analisi delle risposte testuali) mentre in altri forniscono dati quantitativi (frequenze e analisi statistiche). La strategia ottimale sarebbe affiancare tecniche di entrambi i tipi. Infatti, da un lato le analisi quantitative consentono di verificare in maniera più rigorosa l'eventuale presenza di errori di misurazione; dall'altro, i dati qualitativi risultano fondamentali per la comprensione del processo cognitivo che guida l'intervistato alla risposta e quindi molto utili per arrivare alla migliore formulazione dei quesiti.

Nell'effettuare un pre-test è fondamentale anche la selezione dei rispondenti scelti per testare il questionario. In genere si scelgono gruppi di numerosità contenuta che, anche se non rappresentativi della popolazione di riferimento, devono essere il più possibile eterogenei al loro interno e avere le stesse caratteristiche del campione da selezionare per la rilevazione vera e propria (per sesso, età, titolo di studio, etc.).

Diversa e più ampia rispetto ai test finora descritti è l'indagine pilota, il cui obiettivo è quello di testare tutte le fasi della rilevazione; è quindi particolarmente opportuno che sia inserita nel processo di indagine ogni qual volta si tratti della prima volta che si adotta una tecnica mista. Infatti, sia nel caso in cui si inserisca una nuova tecnica in aggiunta a quella adottata fino a quel momento, sia per una indagine *mixed-mode* da realizzare per la prima volta, ci si troverà ad affrontare una serie di problematiche nuove che dovranno essere gestite durante tutto il processo di indagine.

L'indagine pilota consente così di verificare non solo l'equivalenza e la funzionalità/usabilità dei questionari utilizzati, ma tutto l'impianto della rilevazione: dagli aspetti organizzativi alla selezione dei campioni da intervistare con le varie tecniche, ai tassi di risposta che si potranno ottenere con ciascuna tecnica, alle strategie di sollecito e di sensibilizzazione dei rispondenti per aumentare la copertura dell'indagine. Il tutto finalizzato a raccogliere elementi utili per la valutazione, misurazione e trattamento dell'effetto tecnica, nonché per l'individuazione delle relative cause, sia in un'ottica di prevenzione che di correzione (problemi affrontati nei Capitoli 4 e 5).

4. METODI DIAGNOSTICI PER LA VALUTAZIONE E LA MISURAZIONE DELL'EFFETTO TECNICA¹

4.1 Introduzione

Come illustrato nei capitoli precedenti, l'analisi dei dati raccolti mediante rilevazione campionaria multi-tecnica richiede l'uso di metodologie specifiche. La tecnica mista di rilevazione può infatti causare l'insorgere di differenze sistematiche nei dati raccolti. Tali differenze devono essere quindi individuate e misurate, al fine di trarre informazioni utili in fase di analisi dei risultati dell'indagine.

Come introdotto nel Capitolo 2, l'effetto tecnica si distingue in due componenti, effetto selezione (associato a errori di non osservazione, quali errori di copertura e di mancata risposta), e effetto misurazione (errori di osservazione dovuti alla tecnica), entrambe con un impatto sulla variabilità e la distorsione delle stime finali prodotte.

I metodi diagnostici per valutare e misurare l'effetto tecnica e le sue componenti sono molteplici, e possono essere grossolanamente raggruppati in due ampie categorie: metodi di analisi esplorativa e inferenziale, basati sull'utilizzo di tecniche grafiche, indicatori e test statistici, e metodi basati sull'utilizzo di modelli statistici. Tali strumenti forniscono informazioni sull'effetto tecnica a vari livelli, pertanto è auspicabile, in fase di analisi preliminare dei dati osservati, sfruttare il potere informativo di ciascuno di essi per pervenire ad un quadro quanto più completo possibile dei vantaggi e degli svantaggi connessi al mix di tecniche adottato.

I primi permettono al ricercatore di avere rapidamente una visione complessiva sulla qualità dei dati raccolti. Per quanto queste analisi siano spesso soltanto preliminari e condotte in vista di analisi più complesse, esse sono un passaggio "obbligato e irrinunciabile", poiché "la conoscenza approfondita dei propri dati – che si ottiene principalmente con un'analisi ben condotta – è una condizione per impostare ed eseguire efficacemente le successive analisi più complesse, estraendo dai dati la quota più alta possibile delle informazioni interessanti che essi contengono, e soprattutto evitando di estrarre informazioni che essi non contengono" (Marradi, 1995, p.25). Nel capitolo saranno illustrati, a titolo esemplificativo, alcuni dei principali indicatori utili a una prima analisi esplorativa dei dati e saranno riportati alcuni esempi riferiti a più modelli di indagine multi-tecnica, utilizzando sia risultati prodotti in letteratura, sia risultati provenienti dall'analisi di dati di indagini Istat. L'intento è quello di fornire spunti sulle possibili analisi e interpretazioni dei risultati, anche se le analisi dovranno poi essere adattate ai differenti contesti di indagine e agli specifici obiettivi conoscitivi.

I metodi di tipo inferenziale e quelli basati sull'utilizzo di modelli statistici hanno come obiettivo quello di approfondire i risultati dell'analisi di tipo esplorativo, per fornire informa-

¹ Hanno collaborato alla stesura del capitolo: Maria Clelia Romano (paragrafo 4.1); Silvia Montecolle e Emanuela Bologna (paragrafo 4.2.1); Gabriella Fazzi (paragrafo 4.2.2); Orietta Luzi (paragrafi 4.2.3 e 4.2.4); Roberta Varriale (paragrafi 4.1 e 4.3.2); Ilaria Vannini (paragrafi 4.3.1 e 4.3.3).

zioni sulla significatività statistica delle differenze fra le misurazioni ottenute con le differenti tecniche utilizzate, e/o di scorporare le diverse componenti dell'effetto tecnica. È naturale che, per la natura stessa di tali metodi, essi vengano solitamente utilizzati per analisi su un sottoinsieme di variabili rilevate, di maggiore interesse per i ricercatori e/o ritenute maggiormente sensibili all'effetto tecnica.

Le informazioni ottenute dalle tipologie di strumenti appena introdotte potranno essere utilizzate sia in fase di stima per la correzione delle stime prodotte (cfr. Capitolo 5), sia per un eventuale ridisegno della rilevazione analizzata o di altre analoghe ad essa (cfr. Capitolo 4). Obiettivo di questo Capitolo è quindi, da un lato, offrire una panoramica degli strumenti di analisi esplorativa più comunemente utilizzati per la valutazione dell'effetto tecnica (Paragrafi 4.2.1, 4.2.2 e 4.2.3), dall'altro, individuare e proporre ulteriori strumenti per una diagnosi più approfondita, in particolare analisi di tipo inferenziale (Paragrafo 4.2.4) e metodi basati sull'utilizzo di modelli statistici (Paragrafo 4.3).

4.2 Metodi esplorativi e inferenziali

In questo Capitolo sono illustrati alcuni strumenti di natura esplorativa ed inferenziale utili all'analisi preliminare dei dati raccolti in rilevazioni che hanno adottato un disegno multi-tecnica.

4.2.1 Analisi esplorative per unità

Come illustrato nel Capitolo 3, tra le motivazioni che inducono ad adottare un disegno multi-tecnica c'è sicuramente la possibilità di migliorare la copertura e la partecipazione alla rilevazione, soprattutto nel caso di indagini su larga scala, grazie al fatto che ciascuna tecnica dovrebbe riuscire a raggiungere meglio particolari segmenti di popolazione.

Tuttavia, anche tra la popolazione effettivamente raggiungibile attraverso le varie tecniche, la propensione a partecipare all'indagine può differire, e la mancata risposta concentrarsi su specifici sottogruppi di popolazione, in modo peculiare a seconda delle caratteristiche socio-demografiche (autoselezione). Questo effetto impatta sulle stime prodotte al punto che alcuni autori ritengono anche poco corretto da un punto di vista metodologico l'utilizzo di tecniche diverse per la stima di un fenomeno.² Ovviamente l'effetto selezione varia al variare del disegno multi-tecnica scelto ed è possibile adottare strategie in fase di raccolta dati in grado di contenerlo. Tuttavia non è possibile eliminarlo del tutto, pertanto nel momento in cui si conduce una rilevazione multi-tecnica è bene tenere sotto controllo una serie di indicatori che possono essere utili non solo a termine del *field* (per una valutazione complessiva dell'andamento della rilevazione stessa e per trarne indicazioni utili in fase di correzione della mancata risposta totale), ma anche in corso di rilevazione per valutare l'adozione di possibili strategie in grado di migliorare la partecipazione di fasce particolarmente sfuggenti della popolazione o semplicemente "stressare" al massimo la resa della tecnica meno costosa.

La scelta degli indicatori da tenere sotto controllo e la loro interpretazione dipendono molto dal tipo di disegno multi-tecnica adottato, se si tratta cioè di campioni indipendenti o meno, oppure se le diverse tecniche sono utilizzate sequenzialmente o simultaneamente. Di

² Cfr. Lugtig et al., 2011.

seguito saranno illustrati, a titolo esemplificativo, alcuni dei principali indicatori utili a una prima analisi esplorativa dei dati e saranno riportati alcuni esempi riferiti a più modelli di indagine multi-tecnica. L'intento è fornire spunti sulle possibili analisi e interpretazioni dei risultati, ma resta chiaro che a seconda dell'obiettivo che ci si pone (correzione delle mancate risposte, riprogettazione della rilevazione, etc.) si possono privilegiare alcuni indicatori piuttosto che altri, così come scegliere alcune variabili di classificazione piuttosto che altre e affinare le analisi per agevolarne l'interpretazione.

Nell'ottica di massimizzare la qualità dei dati, minimizzando l'errore totale associato all'indagine c'è anche la necessità di raccogliere, durante la rilevazione, una serie di informazioni che permettono di tenere sotto controllo gli errori non campionari come gli errori di copertura, oltre agli errori di mancata risposta. L'utilizzo di una tecnica piuttosto che di un'altra permette di raccogliere queste informazioni in modo più o meno agevole e dettagliato. Per le indagini con tecnica Cawi, mancando la figura dell'intervistatore, non è possibile avere un elevato dettaglio sulle ragioni della mancate risposte totali (rifiuti, mancato recapito dell'informativa, analfabetismo informatico, etc.), tranne nei casi in cui tale motivo viene comunicato al *Contact center* (nel corso di contatti *inbound* o *outbound*). In generale, più è articolata l'informazione registrata durante la fase di raccolta dati, maggiori sono le informazioni a disposizione per analizzare i non rispondenti, per intervenire nel trattamento degli effetti distorsivi per l'indagine corrente, per prevenire le eventuali distorsioni nelle future edizioni dell'indagine stessa. Ad esempio, nel caso di indagini campionarie condotte con intervista diretta, gli esiti associati ad interviste non avvenute permettono di analizzare le motivazioni della selezione dei rispondenti, e comprendere se esse sono riconducibili alle liste da cui il campione è stato estratto o sono una conseguenza delle caratteristiche delle unità oggetto di indagine.

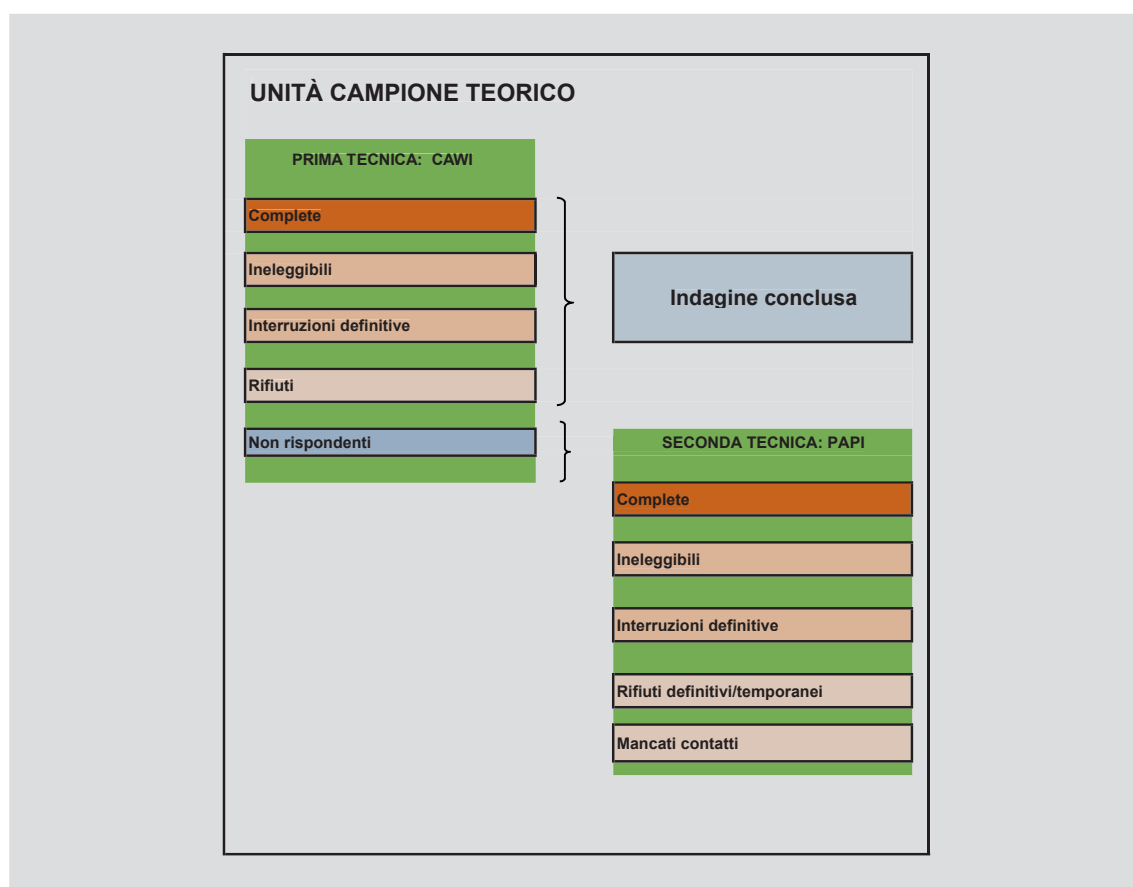
Queste informazioni assumono maggiore rilievo nel caso di una indagine multi-tecnica. Le analisi degli esiti portano infatti a riflessioni diverse a seconda che si tratti di una indagine concorrente o sequenziale.

Nel caso di indagini concorrenti, laddove è possibile ricondurre tutte le unità oggetto di indagine ad un esito, il confronto permette di valutare la diversa bontà delle liste utilizzate per l'estrazione delle unità (nel caso le unità provengano da liste diverse) o la mancata partecipazione dovuta ad altri motivi, isolando ad esempio quelli più strettamente legati alla tecnica utilizzata.

Nel caso di indagini sequenziali, avere informazioni sul motivo della mancata partecipazione all'indagine con la tecnica scelta per il primo tentativo di contatto può essere di aiuto per gestire la seconda fase di indagine. Le unità da contattare con la seconda tecnica possono essere selezionate non solo in base al criterio della mancata risposta alla prima tecnica, ma anche evitando di ritornare su unità non rispondenti che hanno indicato motivazioni che comporterebbero la loro esclusione anche dalla seconda tecnica, come l'ineleggibilità o rifiuto per grave malattia che possono essere considerati esiti non recuperabili (Figura 4.1).

Conoscere nel dettaglio gli esiti della mancata partecipazione è utile anche per studiare i profili dei non rispondenti, poiché consente di analizzarne non solo le caratteristiche strutturali, se note, ma anche le motivazioni per cui le unità contattate, non hanno partecipato alla rilevazione. A partire dagli esiti e a seconda dell'unità di rilevazione (individuo o famiglia) è possibile costruire diversi tassi e indicatori. I tassi di risposta o di mancata partecipazione possono essere declinati rispetto alle caratteristiche strutturali del campione (ad esempio territorio di residenza, numero componenti famiglia, sesso e classi di età

Figura 4.1 - Diagramma di flusso esemplificativo dei possibili esiti di una indagine a tecnica mista sequenziale Cawi/Papi



degli individui) e danno una prima indicazione della performance della rilevazione e della selezione degli intervistati.

Anche nel caso di una tecnica simultanea sulle stesse unità di rilevazione (che possono per esempio scegliere quale canale utilizzare per partecipare alla rilevazione), i tassi possono fornire un quadro di insieme molto utile anche in corso di rilevazione ed orientare eventuali interventi a supporto di una tecnica o dell'altra.

La stessa interpretazione dei tassi sarà diversa a seconda del tipo di disegno adottato nell'indagine analizzata. Per esempio, se si tratta di un'indagine con campioni indipendenti sottoposti a due diverse tecniche, l'analisi evidenzia effettivamente il contributo relativo di ciascuno dei canali utilizzati nel determinare il numero di rispondenti finali. Se si tratta invece di un medesimo campione al quale è stata proposta la doppia tecnica (sequenzialmente o in maniera concomitante), il tasso fornisce prime indicazioni anche sui fattori individuali, familiari e di contesto che favoriscono la propensione a partecipare a un'indagine con una tecnica piuttosto che con un'altra. Nel caso, inoltre, in cui il disegno di indagine abbia previsto due campioni indipendenti, di cui uno sottoposto a una doppia tecnica mista e un secondo campione, trattato con mono-tecnica, il confronto tra i tassi di risposta mostra i differenti risultati che una stessa tecnica consegue in base al disegno adottato.

Per esempio nelle indagini sui laureati e sui diplomati a un campione è stata proposta una tecnica mista sequenziale Cawi e Cati, e a un altro campione esclusivamente la tecnica Cati. Come evidente dalla Tavola 4.1 la risposta Cati degli individui sottoposti a tecnica mista è più bassa (meno della metà) della risposta Cawi ed è molto più contenuta (quasi un terzo) della risposta Cati, derivante da coloro ai quali non è stata proposta la tecnica web.

4. Metodi diagnostici per la valutazione e la misurazione dell'effetto tecnica

Tavola 4.1 - Tassi di risposta per tipologia di tecnica di indagine, gruppo di laurea e tipo di diploma, sesso e ripartizione geografica - Indagine sull'inserimento dei laureati e dei diplomati- Anno 2015

GRUPPI DI CORSO DI LAUREA SESSO RIPARTIZIONE GEOGRAFICA	LAUREATI DI PRIMO E SECONDO LIVELLO			
	Tasso di risposta Cawi (a)	Tasso di risposta Cati dopo fase Cawi (b)	Tasso di risposta Cati 'puro' (c)	Tasso di risposta totale
CORSO DI LAUREA				
Agrario	47,3	24,9	67,5	71,7
Architettura	49,1	20,8	56,9	68,2
Chimico-farmaceutico	49,7	20,5	57,8	68,7
Difesa e sicurezza	34,3	14,3	33,3	46,5
Economico-statistico	51,4	21,0	62,0	71,1
Educazione fisica	40,8	27,0	64,7	67,5
Geo-biologico	57,0	19,5	63,5	75,0
Giuridico	44,8	23,1	56,4	66,5
Ingegneria	57,6	18,0	62,6	73,9
Insegnamento	42,7	27,6	58,7	68,9
Letterario	49,1	22,0	62,8	70,1
Linguistico	48,1	20,5	57,1	67,3
Medico	42,9	27,0	63,2	69,0
Politico-sociale	47,9	21,9	59,2	68,6
Psicologico	45,9	22,9	67,1	68,7
Scientifico	65,6	14,3	67,5	78,7
SESSO				
Maschi	51,7	20,2	60,8	70,6
Femmine	47,5	23,6	61,7	69,9
RIPARTIZIONE GEOGRAFICA				
Nord-ovest	52,3	20,5	61,5	71,3
Nord-est	51,2	22,0	62,4	72,0
Centro	50,5	21,4	61,2	70,7
Sud	43,8	24,0	58,1	66,7
Isole	47,1	23,5	65,8	70,0
Totale	49,4	22,0	61,3	70,2
TIPI DI DIPLOMA				
DIPLOMATI				
SESSO RIPARTIZIONE GEOGRAFICA	Tasso di risposta Cawi (a)	Tasso di risposta Cati dopo fase Cawi (b)	Tasso di risposta Cati 'puro' (c)	Tasso di risposta totale
Istituto professionale	19,0	28,0	37,3	43,5
Istituto tecnico	25,5	26,9	53,6	47,4
Liceo	35,2	25,5	42,1	55,6
Istruzione magistrale	29,0	26,9	44,7	52,4
Istruzione artistica	24,7	24,7	37,4	45,8
SESSO				
Maschi	23,9	27,8	37,7	47,1
Femmine	27,0	26,1	40,2	48,8
RIPARTIZIONE GEOGRAFICA				
Nord-ovest	30,7	25,5	39,3	51,2
Nord-est	28,3	26,4	40,0	50,3
Centro	25,3	29,2	41,9	50,3
Sud	20,8	27,1	37,2	44,0
Isole	21,6	25,8	36,9	43,6
Totale	25,6	26,8	39,0	48,0

(a) Dato dal rapporto fra il numero di rispondenti Cawi e il numero di laureati sottoposti alla tecnica mista e sequenziale Cawi e Cati.

(b) Dato dal rapporto fra il numero di rispondenti Cati sottoposti alla tecnica mista e sequenziale Cawi e Cati e il numero di laureati sottoposti alla tecnica mista e sequenziale Cawi e Cati.

(c) Dato dal rapporto fra il numero di rispondenti sottoposti alla Cati e il numero di laureati sottoposti alla tecnica Cati.

In altre parole la resa della tecnica Cati diminuisce se usata dopo una tecnica Cawi, perché lavora su un bacino di rispondenti più difficile, visto che i più propensi a rispondere lo hanno già fatto utilizzando la prima tecnica proposta, che peraltro ha già prodotto con i solleciti inviati un certo “stress” sulle unità rispondenti.³

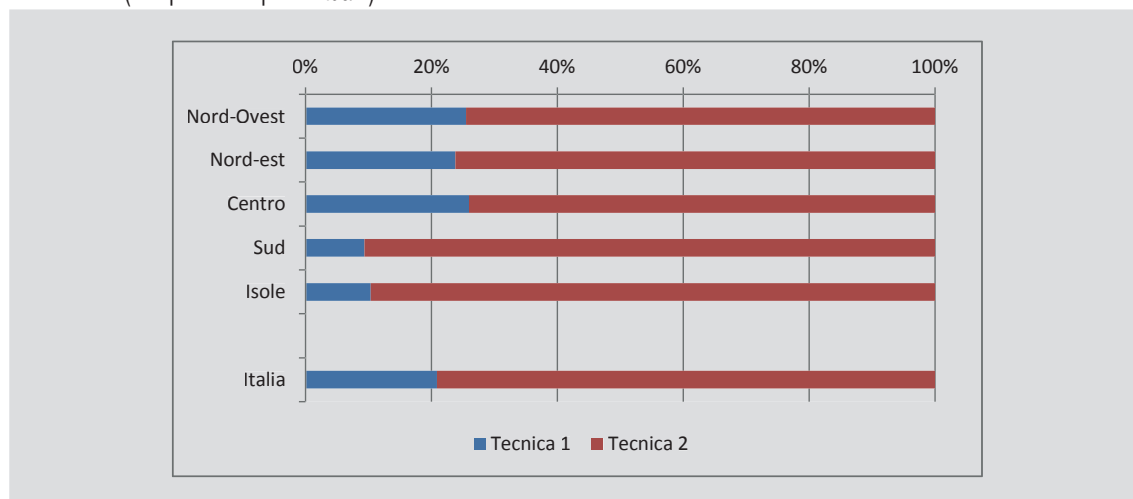
Il confronto tra target di popolazione differenti (diplomati da un lato, laureati dall'altro) aggiunge ulteriori informazioni sulla differente resa di ciascuna tecnica a seconda della specificità delle unità di rilevazione. I diplomati hanno manifestato una minore propensione alla partecipazione e una minore propensione al ricorso al web come canale di risposta. Va tuttavia precisato che questo differente tasso di risposta deriva anche dalla minore disponibilità, rispetto ai laureati, di recapiti telefonici o mail e quindi, dalla minore efficacia della strategia di solleciti adottata.

Infine, per entrambe queste rilevazioni è evidente il vantaggio che la doppia tecnica consente di ottenere in termini di copertura della popolazione di riferimento.

Oltre ai tassi di risposta, anche la semplice distribuzione dei questionari compilati per tecnica di rilevazione può dare indicazioni importanti a fine rilevazione sulla resa complessiva di ciascuna tecnica e sul contributo fornito alla buona riuscita della rilevazione.

A titolo esemplificativo, nella Figura 4.2 è riportato un esempio di rappresentazione grafica di questo indicatore. Nel caso riportato, la percentuale di questionari compilati via web presenta una certa eterogeneità a livello territoriale con valori più elevati nelle regioni settentrionali.

Figura 4.2 - Famiglie che hanno compilato il questionario per ripartizione e tecnica di rilevazione
(composizione percentuale)



Può essere utile inoltre analizzare la composizione dei rispondenti per tecnica, confrontando le distribuzioni delle principali variabili strutturali riferite all'intero campione di rispondenti e ai due sotto-campioni distinti per tecnica con la distribuzione delle medesime variabili nella popolazione di riferimento. Ovviamente tale analisi sarà tanto più accurata quanto più ricche sono le informazioni contenute nella lista originaria, poiché ci consente di valutare le distribuzioni rispetto a più variabili e quindi fornire informazioni per la correzione delle stime. E' opportuno elaborare questi indicatori tenendo conto di alcune variabili chiave, per esempio quelle connesse ai domini di stima che nel caso delle indagini su individui

³ Cfr. Gallo, 2017.

4. Metodi diagnostici per la valutazione e la misurazione dell'effetto tecnica

e famiglie sono generalmente sesso, classe di età, regione, classe di ampiezza del comune, etc. Sarà utile però tenere sotto controllo tutte le principali variabili classificatorie relative al fenomeno indagato (per esempio, il titolo di studio), sebbene in questo caso il confronto sia spesso possibile solo tra le distribuzioni ottenute con le due tecniche e non con la popolazione di riferimento

La Tavola 4.2 riporta un esempio di come rappresentare le principali variabili rispetto alle quali è possibile fare questo confronto per le indagini sulla popolazione, che utilizzano le Liste Anagrafiche Comunali (Lac) per l'estrazione dei campioni.

Tavola 4.2 - Questionari compilati per caratteristiche individuali e/o familiari e tecnica utilizzata (composizione percentuale)

SESSO RIPARTIZIONE GEOGRAFICA	Dottori 2008	Dottori 2010	Totale	Totale
SESSO				
Maschio	%	%	%	%
Femmina	%	%	%	%
Totale	100	100	100	100
CLASSI DI ETÀ'				
Fino a 17	%	%	%	%
.....	%	%	%	%
Oltre 50	%	%	%	%
Totale	100	100	100	100
CITTADINANZA				
Italiana	%	%	%	%
Straniera	%	%	%	%
Totale	100	100	100	100
RIPARTIZIONE GEOGRAFICA				
.....	%	%	%	%
REGIONE				
.....	%	%	%	%
AMPIEZZA DEMOGRAFICA DEL COMUNE				
.....	%	%	%	%
N.COMPONENTI				
.....	%	%	%	%
TIPOLOGIA FAMILIARE				
.....	%	%	%	%

Tuttavia, le informazioni disponibili sull'universo di riferimento possono non esaurire le esigenze informative. Per esempio, non è possibile valutare la rappresentatività della popolazione in base al titolo di studio o allo stato civile dei rispondenti, non essendo tali variabili presenti nelle Lac e dunque essendo ignota la distribuzione delle stesse nella popolazione di riferimento. In questi casi è possibile valutare l'utilizzo di altre fonti (per esempio gli archivi del Miur per il titolo di studio) per arricchire questo tipo di analisi con informazioni aggiuntive, qualora le si ritengano particolarmente correlate con i fenomeni oggetto di studio.

Per descrivere in maniera ancora più dettagliata il profilo dei non rispondenti per tecnica di rilevazione è utile un *linkage* della lista del campione con altri dati amministrativi o da altra fonte disponibile (Illemann Christensen *et al.*, 2015). In una ricerca sulla salute condotta dal National Institute of Public Health danese, i dati anagrafici sui membri del campione sono stati incrociati con informazioni desunte dai registri sulla salute (Tavola 4.3). A partire da questo *data set* integrato, gli autori sottolineano alcune importanti differenze fra i rispondenti e non rispondenti nelle due diverse tecniche previste (in presenza o assenza di un intervistatore): i non rispondenti, in genere (e in misura maggiore i non

Tavola 4.3 - Analisi dei non rispondenti con variabili ausiliarie da altra fonte (valori percentuali)

	INTERVISTA FACCIA A FACCIA			QUESTIONARIO AUTO-AMMINISTRATO		
	Rispondenti	Non rispondenti	Popolazione	Rispondenti	Non rispondenti	Popolazione
Medicinali prescritti (%)	62,4	57,5	61,2	60,9	53,1*	57,9
Contatto con il medico generico (%)	63,1	60,0	62,4	63,2	57,4*	61,0
Numero medio di contatti con un medico generico	3,69	3,76	3,71	3,29	3,14	3,23

Fonte: adattato da Illemann Christensen et al. (2013), p. 330.

rispondenti al questionario autosomministrato), tendono a fare ricorso ai servizi sanitari in misura minore.

Può essere utile anche rappresentare graficamente la distribuzione di alcune variabili. Il grafico riportato nella Figura 4.3, per esempio, consente di confrontare la distribuzione per età dei rispondenti a uno stesso questionario, somministrato con due differenti tecniche con la distribuzione per età della popolazione di riferimento. Nel caso specifico, la rappresentazione grafica aiuta a cogliere con un colpo d'occhio, le differenze tra le sottopopolazioni raggiunte con le due tecniche, e di entrambe rispetto alla popolazione di riferimento.

Figura 4.3 - Popolazione per età per tecnica

Laddove possibile è utile confrontare l'esito dell'indagine *mixed-mode* nel suo complesso (senza distinguere i campioni raggiunti con le diverse tecniche) con l'indagine *uni-mode*. Ad esempio (Tavola 4.4) Sala e Lynn (2010) controllano eventuali effetti tecnica nell'indagine multi-tecnica Cawi/Cati confrontando la distribuzione di alcune variabili demografiche con quelle dell'indagine

Tavola 4.4 - Composizione del campione per variabili demografiche

	Postale	Multi-tecnica	Popolazione di riferimento
SESSO			
Uomini	37,2	44,2	41,9
Donne	62,8	55,8	58,1
ETÀ			
18-35	30,2	32,0	32,0
36-45	35,7	35,4	36,0
46 e oltre	34,1	32,6	32,0

Fonte: adattato da Sala e Lynn (2010), p.128.

postale condotta contemporaneamente.⁴

Questo indicatore può fornire informazioni di grande utilità sulla diversa composizione della popolazione catturata da ciascuna tecnica e sulla capacità di ciascuna di esse, e di entrambe a fine rilevazione, di rappresentare adeguatamente la popolazione di riferimento. Per esempio, se le differenze nella composizione demografica dei rispondenti alle varie tecniche non sono significative, è meno probabile attendersi una distorsione dovuta alla diversa entità della mancata risposta tra le due tecniche. Di conseguenza le differenze nei dati, se presenti, non sarebbero attribuibili alle differenze di composizione del campione. Diverso il caso in cui si ravvisino differenze nella composizione dei due sotto-campioni.

Questo stesso indicatore (distribuzione dei rispondenti per alcune caratteristiche) può essere calcolato anche per variabili non disponibili per la popolazione di riferimento, e fornire informazioni altrettanto utili in merito al ruolo che le caratteristiche in esame hanno nell'accrescere o meno la propensione alla risposta con le tecniche utilizzate.

Osservando la distribuzione per titolo di studio nell'indagine Istat "Musica e video nelle abitudini dei cittadini"⁵ (Tavola 4.5) appare evidente come le due tecniche abbiano un diverso grado di successo nel catturare determinati target di popolazione. La tecnica Cawi, diversamente dalla Cati, ottiene migliori risultati con gli individui con elevato titolo di studio e con competenze informatiche. Infatti, tra i rispondenti al questionario web è più elevata la percentuale di quanti hanno conseguito almeno il diploma e usano il Pc.

Tavola 4.5 - Individui per titolo di studio più alto conseguito e uso del PC per tecnica di rilevazione - Indagine Musica e video nelle abitudini dei cittadini, Anno 2017 (composizione percentuale)

	Cawi	Cati
TITOLO DI STUDIO PIÙ ALTO IN FAMIGLIA		
Dottorato di ricerca o laurea	17,9	13,1
Diploma superiore	40,2	35,6
Licenza media	29,8	32,1
Licenza elementare, nessun titolo	11,8	19,2
USO DEL PC		
Sì	92,4	73,4
No	7,6	26,6
Totale	100,0	100,0

In un disegno multi-tecnica, la resa di ciascuna tecnica dipende anche dall'insieme delle metodologie adottate nel disegno dell'intero processo di indagine, comprese ad esempio le strategie di sollecito adottate nella componente Cawi, o quelle per la gestione delle eventuali risposte *proxy* (cioè non fornite dal diretto interessato). Per esempio, qualora nell'ambito di una rilevazione sulle famiglie, in cui vanno intervistati tutti i componenti, venga data agli intervistati la possibilità di compilare il questionario non solo per sé stessi ma anche per altri componenti della famiglia momentaneamente impossibilitati (modalità *proxy*), può essere interessante valutare le differenze tra le tecniche nell'adottare questa scelta. Ad esempio (Tavola 4.6), nell'indagine "Cittadini e tempo libero", nella fase Papi il 12,7 per cento delle persone intervistate ha risposto in modalità *proxy* ad un questionario destinato all'auto-compilazione, mentre nella fase Cawi si arriva al 21,2 per cento. Questa maggiore percentuale di risposte *proxy* nella fase Cawi probabilmente si spiega con la ten-

⁴ L'indagine utilizzata dagli autori è la Ismie (Improving survey measurement of income and employment, cfr. Jäckle *et al.*, 2004).

⁵ Per maggiori informazioni sulla rilevazione, cfr. il link: <http://www.istat.it/it/archivio/197097>.

denza a delegare la compilazione del questionario non solo in caso di assenza, ma anche per scarsa propensione all'utilizzo degli strumenti informatici. Le differenze che emergono tra le due tecniche, inducono anche a interrogarsi sull'effetto che l'eventuale scelta di non consentire tale modalità di risposta, avrebbe avuto sulla partecipazione: probabilmente avrebbe determinato un incremento delle mancate risposte differenziato tra le due tecniche e dunque più elevato per la componente Cawi. Nei prossimi capitoli si vedrà anche l'impatto che tale scelta può avere nelle stime.

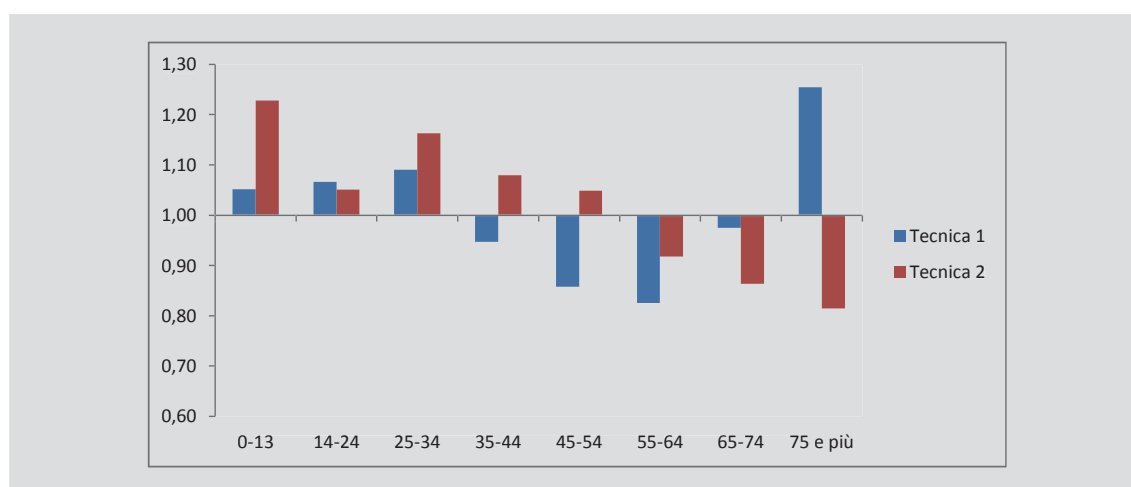
Tavola 4.6 - Persone di 14 anni e più che per modalità di risposta Proxy/NON Proxy al questionario autocompilato per tecnica di rilevazione. Indagine Cittadini e tempo libero. Anno 2015 (valori assoluti e percentuali)

	Cawi		Papi	
	Valori assoluti	%	Valori assoluti	%
QUESTIONARIO AUTOCOMPILATO				
Non Proxy	5.763	78,9	24.368	87,3
Proxy	1.546	21,2	3.539	12,7
Totale	7.309	100,0	27.907	100,0

Fonte: Indagine Istat "I Cittadini e il tempo libero", Anno 2015.

Nel caso di indagini campionarie su individui o famiglie, un altro modo per capire quanto una specifica sotto-popolazione sia sopra o sottostimata nel campione è quello di valutare un peso medio atteso, rapportando la popolazione nota a quella campionata. Ad esempio, nel caso di indagini campionarie su individui il peso medio che ogni individuo dovrebbe avere è dato dal rapporto tra la popolazione italiana (dato demografico) e gli individui rispondenti. Se si vuole una valutazione degli individui campionati per classi di età, si fa lo stesso calcolo distinto per le diverse classi di età: popolazione demografica per classe di età rapportata agli individui rispondenti per classe di età. Il rapporto tra i pesi per classi di età e il peso medio totale dà la distanza dal peso medio e una stima di quanto quella classe di età sia sotto o sopra rappresentata dalle unità rispondenti. Più il rapporto si allontana da 1, più quella popolazione è sotto- o sopra-stimata. Un esempio di rappresentazione per tecnica è riportato nella Figura 4.4.⁶

Figura 4.4 - Esempio di rappresentazione della distanza dal peso medio per classe di età. Confronto tra due tecniche



⁶ Per un approfondimento dei dati dell'indagine "I Cittadini e il tempo libero" 2015 si rimanda a Istat (2017).

Tali analisi ci possono dare un'idea sulle differenze nella copertura che si ottengono a seconda della tecnica di rilevazione utilizzata, e possono dare indicazioni sulle azioni che è necessario intraprendere e dove agire per migliorare le situazioni osservate. Ad esempio, se in una occasione di indagine si è adottato un approccio sequenziale, analizzando i risultati di copertura ottenuti, si potrebbe decidere nell'edizione successiva dell'indagine di passare ad un approccio concorrenziale in cui si seleziona a monte, in base ad una serie di caratteristiche demografiche di base, il campione a cui somministrare una tecnica piuttosto che l'altra.

4.2.2 Analisi esplorative per variabile

In questo paragrafo sono presentati esempi di analisi esplorative per dati rilevati con tecniche diverse che focalizzano l'attenzione sulle caratteristiche delle variabili obiettivo e delle relative distribuzioni, permettendo al ricercatore di avere velocemente una visione complessiva sulla qualità dei dati raccolti variabile per variabile.

Una prima analisi può essere il semplice confronto fra le distribuzioni di frequenza di alcune variabili selezionate. In letteratura (cfr. de Leeuw, 2005; Jackle *et al.*, 2010) si suggerisce di confrontare inizialmente le variabili che non dovrebbero risentire dell'effetto misurazione, ma solo di quello di selezione (ad esempio il genere, l'età, il titolo di studio). Le proposte di analisi dei tassi di risposta e delle composizioni percentuali mostrate nel paragrafo precedente esemplificano tali passaggi. L'obiettivo è capire se è possibile escludere che un effetto selezione ci sia stato, per poter poi confrontare le distribuzioni di altre variabili di interesse per l'indagine. Questa procedura si basa sull'assunto che le variabili di interesse abbiano una forte associazione con quelle controllate in precedenza, assunto spesso difficilmente sostenibile;⁷ ciò nonostante, si ritiene che una prima analisi esplorativa così condotta possa essere utile a orientare il lavoro di analisi successivo.

È utile, inoltre, controllare se la mancata risposta parziale sia in qualche modo legata a un effetto tecnica. Si parla di mancata risposta parziale se si rileva l'"assenza di alcune delle informazioni richieste per un sottoinsieme delle unità intervistate". Se i dati sono *missing at random* (MAR, cfr. Rubin e Little, 2002) ovvero se la probabilità di avere, per una certa variabile Y, una mancata risposta non dipende dal valore della mancata risposta stessa, l'informazione mancante è più facilmente compensabile tramite imputazione – assegnando cioè un valore sostitutivo del dato mancante (cfr. ad esempio Grande e Luzi, 2003). Nel caso in cui sia la tecnica stessa a condizionare la mancata risposta, allora l'assunto della casualità della distribuzione delle mancate risposte è più difficilmente sostenibile.

Casi del genere possono verificarsi per le domande sensibili, sulle quali non vi è mai l'obbligo di risposta, e in cui la presenza dell'intervistatore (ad esempio in un'intervista a domicilio) potrebbe spingere l'intervistato a rispondere più frequentemente che in un questionario autocompilato (web o cartaceo che sia).

Laddove non si riscontrassero forti differenze nelle distribuzioni controllate,⁸ è possibile procedere a confronti di distribuzioni di altre variabili centrali per l'indagine oppure sensibili a effetti di misurazione; nel caso invece si riscontrassero differenze nelle distribuzioni, è invece necessario selezionare, dai campioni generali di ciascuna tecnica, sotto-campioni simili per caratteristiche e, pertanto, confrontabili.

7 Per superare questo limite, in presenza di un'indagine con tecnica unica cui poter comparare i dati ottenuti con il disegno multi-tecnica, Vannieuwenhuyze et al. (2010) hanno proposto un metodo per stimare in maniera distinta effetto misurazione ed effetto selezione.

8 Per i test da utilizzare per valutare la significatività di tali differenze si rimanda al paragrafo successivo.

In ambo i casi, è utile procedere comunque ad analisi tri-variate che permettano di controllare la relazione fra le variabili considerate (quella oggetto di indagine e la tecnica utilizzata) rispetto ad una terza variabile (ad esempio l'area geografica di residenza).

Per variabili categoriche (nominali o ordinali), può essere utile sia un confronto su tabella, sia un confronto grafico. Ad esempio in letteratura si riportano le distribuzioni semplici ottenute con le due tecniche, presentate come distribuzioni percentuali oppure come proporzione (Tavola 4.7). Gli istogrammi a colonne raggruppate (in base alla tecnica) possono essere un'efficace rappresentazione grafica che facilita il confronto fra i dati.

Tavola 4.7 - Interesse per la politica per tecnica di rilevazione e popolazione di confronto (European Social Survey)
(proporzioni)

	Cawi	Capi	ESS
INTERESSE PER LA POLITICA			
P (per niente interessato)	0,084	0,033	0,067
P (poco interessato)	0,330	0,188	0,224
P (abbastanza interessato)	0,488	0,679	0,607
P (molto interessato)	0,098	0,100	0,101

Fonte: adattato da Vannieuwenhuyze et al. (2010), p. 1038.

Talvolta già il semplice confronto fra distribuzioni fa emergere risultati interessanti. Un esempio esplicativo riguarda le domande sensibili, per le quali l'effetto misurazione è ovviamente più rilevante. Nella domanda sull'autodefinizione dell'orientamento sessuale dei soggetti (Tavola 4.8),⁹ il confronto fra le risposte ottenute in presenza o in assenza dell'intervistatore in campioni omogenei per fascia d'età fa emergere elementi interessanti: la maggiore tendenza a non rispondere o a rispondere con la categoria "in altro modo" nel cartaceo autosomministrato; la maggiore percentuale di quanti si dichiarano "eterosessuali", categoria più socialmente desiderabile, nell'intervista con rilevatore.¹⁰

Tavola 4.8 - Autodefinizione dell'orientamento sessuale per tecnica di rilevazione (valori percentuali)

	Saq ¹⁰	Cati
Non risponde	18,1	2,92
Omosessuale/ bisessuale	1,8	0,3
Transessuale	0,1	0,0
Eterosessuale	74,3	96,5
In altro modo	5,7	0,3

Fonte: Istat: Indagine sulle discriminazioni di genere, orientamento sessuale e appartenenza etnica (indagine pilota e core), Fazzi et al., 2017.

Ad esempio, nell'indagine campionaria sull'inserimento professionale dei laureati,¹²

⁹ L'esempio è tratto dall'indagine sulle discriminazioni di genere, orientamento sessuale e appartenenza etnica, condotta nel 2011 in modalità Capi e questionario autocompilato cartaceo. Nel 2010 era stata condotta l'indagine pilota tramite intervista telefonica (Capi).

¹⁰ Analisi successive hanno mostrato che sono i più anziani, i meno istruiti e gli stranieri a non rispondere a questa domanda. Per la maggior parte coloro che non rispondono, ma compilano altre parti del questionario, si dichiarano attratti da persone del sesso opposto. Analogamente anche parte di coloro che si definiscono "in altro modo" ha caratteristiche simili, ed è plausibile che per questi sottoinsiemi, in assenza di un intervistatore, non sia chiaro il significato del termine "eterosessuale".

¹¹ *Self administered questionnaire*.

¹² L'indagine appare particolarmente utile come esempio per il particolare disegno d'indagine, che ha previsto un multi-tecnica sequenziale Cawi-Cati e un campione indipendente che ha potuto rispondere solo in Cati. Ciò, vedremo di seguito, permette di confrontare i campioni che hanno risposto in Cawi e in Cati, e in Cati con disegni di campionamento differenti.

condotta dall'Istat nel 2015, un confronto fra le distribuzioni¹³ del motivo principale per cui si è scelto un determinato corso di laurea mostra una maggiore tendenza degli intervistati a scegliere le modalità più socialmente desiderabili (“interesse per la disciplina” o “prospettive lavorative”) in presenza dell'intervistatore, e una maggiore propensione alle risposte “per ripiego” e “prosecuzione naturale degli studi”, forse percepite come meno apprezzabili, nell'autocompilazione sul web (Tavola 4.9).

Tavola 4.9 - Motivo principale d'iscrizione all'università per tecnica di rilevazione (valori percentuali)

MOTIVO PRINCIPALE PER L'ISCRIZIONE	Cawi	Cati post Cawi	Cati indipendente	Totale
interesse verso la disciplina specifica	63,8	67,4	69,1	65,4
prospettive lavorative	15,4	17,8	16,4	16,2
prosecuzione naturale degli studi dopo la laurea triennale	13,2	9,2	10,1	11,7
per ripiego (non ho superato i test di ingresso per il corso che mi interessava)	2,0	0,9	0,7	1,5
suggerito da familiari/amici	1,3	1,5	0,9	1,3
la laurea triennale non offriva opportunità di lavoro	1,3	0,9	0,9	1,1
rappresentava l'unica offerta didattica locale	1,0	0,8	0,5	0,9
il diploma non offriva opportunità di lavoro	0,7	0,5	0,5	0,6
suggerito dai professori di scuola secondaria	0,2	0,3	0,2	0,2
seguire gli amici	0,1	0,2	0,1	0,2
altro motivo	1,0	0,6	0,6	0,9
Totale	100,0	100,0	100,0	100,0

Fonte: Istat: Elaborazioni ad hoc su dati dell'Indagine sull'inserimento professionale dei laureati. Anno 2015.

Dopo aver confrontato i risultati ottenuti, a prescindere dal fatto che siano o meno evidenti delle differenze, è opportuno incrociare i dati con una terza variabile che si ritiene possa influenzare la relazione fra la tecnica di somministrazione e la variabile in esame. La relazione tra le due variabili, infatti, potrebbe cambiare in presenza di una terza variabile, che potrebbe spiegare la variazione, e quindi correggere una relazione spuria, o farla emergere, evidenziando una relazione soppressa.

L'analisi tri-variata permette quindi di controllare il persistere dell'effetto a parità di un'altra condizione, nell'esempio presentato il tipo di laurea (triennale o specialistica/da vecchio ordinamento o a ciclo unico). Nel caso specifico la presenza di modalità non disponibili per le lauree triennali giustifica molte delle differenze riscontrate. È interessante osservare comunque il persistere di un possibile effetto di desiderabilità sociale dovuto alla presenza dell'intervistatore anche nelle risposte dei laureati triennali o di quelli specialistici. In particolare la modalità di risposta “per ripiego” sembra vere un effetto di desiderabilità sociale solo per i laureati della triennale, per i quali probabilmente la selezione al momento dell'iscrizione ai corsi di laurea è più forte (Tavola 4.10).

In presenza di variabili numeriche, il confronto fra le medie nei diversi gruppi – suddivisi in base alla tecnica - permette di analizzare le differenze. Sempre dall'indagine sui laureati sono state considerate ad esempio due variabili numeriche: il voto di laurea e la retribuzione media percepita. Qui le differenze appaiono minori, e sembrano dovute a un effetto di selezione del campione più che di tecnica. Infatti (Tavola 4.11), appare persistentemente più basso il voto medio di quanti hanno risposto nella Cati sequenziale, ovvero di coloro che non avendo risposto in Cawi sono stati

¹³ Si ringrazia Francesca Gallo per le elaborazioni dei dati.

Tavola 4.10 - Motivo principale d'iscrizione all'università per tecnica di rilevazione e tipo di laurea (valori percentuali)

MOTIVO PRINCIPALE PER L'ISCRIZIONE	A CICLO UNICO/ VECCHIO ORDINAMENTO O SPECIALISTICA				TRIENNALE			
	Cawi	Cati post Cawi	Cati	Totale	Cawi	Cati post Cawi	Cati	Totale
Interesse verso la disciplina	57,2	62,7	62,1	59,2	71,8	72,8	77,0	72,6
Prosecuzione naturale degli studi dopo la laurea triennale	24,0	17,2	19,0	21,6
Prospettive lavorative	12,3	14,7	14,9	13,2	19,2	21,3	18,0	19,7
La laurea triennale non offriva opportunità di lavoro	2,3	1,7	1,7	2,1
Rappresentava l'unica offerta didattica locale	1,2	1,0	0,6	1,1	0,7	0,5	0,4	0,6
suggerito da familiari/amici	0,9	1,2	0,6	1,0	1,7	1,8	1,4	1,7
Per ripiego (non ho superato i test di ingresso)	0,6	0,4	0,3	0,5	3,6	1,5	1,1	2,7
Il diploma non offriva opportunità di lavoro	0,3	0,3	0,1	0,3	1,1	0,7	0,9	1,0
Per seguire gli amici	0,1	0,2	0,1	0,1	0,2	0,3	0,2	0,2
Suggerito dai professori di scuola secondaria	0,1	0,1	0,1	0,1	0,4	0,4	0,4	0,4
Altro motivo	0,7	0,5	0,5	0,7	1,4	0,6	0,7	1,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Fonte: Istat: Elaborazioni ad hoc su dati dell'Indagine sull'inserimento professionale dei laureati. Anno 2015.

Tavola 4.11 - Voto di laurea per tecnica di rilevazione e tipo di laurea (valori medi).

	Voto medio totale	Voto medio quinquennale	Voto medio triennale
Cawi	103,3	105,6	100,3
Cati post Cawi	102,7	105,2	99,8
Cati	103,0	105,6	100,1

Fonte: Istat: Elaborazioni ad hoc su dati dell'Indagine sull'inserimento professionale dei laureati. Anno 2015.

recuperati con un'intervista telefonica. Invece il campione Cati indipendente mostra valori molto simili a quello Cawi. Si potrebbe ipotizzare che i soggetti con una migliore riuscita scolastica siano stati i più motivati a partecipare.

Anche controllando il voto medio per tecnica all'interno di ciascun gruppo di laurea (variabile che come noto spiega molta della varianza del voto medio, e che ha effetti anche sulla selezione del campione) nella maggior parte dei casi tale tendenza persiste (Figure 4.5 e 4.6).

Stesso discorso potrebbe valere per la retribuzione media percepita. In questo caso però un'analisi per tipo di laurea mostra retribuzioni abbastanza omogenee per i laureati della triennale che hanno risposto con le diverse tecniche, e retribuzioni più basse per i laureati della specialistica che hanno risposto in Cawi. L'analisi per gruppo scientifico offre un quadro variegato, facendo propendere per l'esclusione di un effetto simile a quello sul voto analizzato in precedenza (cfr. Figura 4.7).

Fino ad ora abbiamo cercato di attribuire significato alle differenze fra medie o fra distribuzioni senza preoccuparci della significatività statistica delle stesse. Il prossi-

4. Metodi diagnostici per la valutazione e la misurazione dell'effetto tecnica

mo paragrafo è pertanto dedicato ai test statistici da applicare, in base ai differenti disegni d'indagine e delle variabili implicate nell'analisi.

Figura 4.5 - Voto di laurea per tecnica di rilevazione e gruppo di laurea specialistica, quinquennale o del vecchio ordinamento (valori medi)

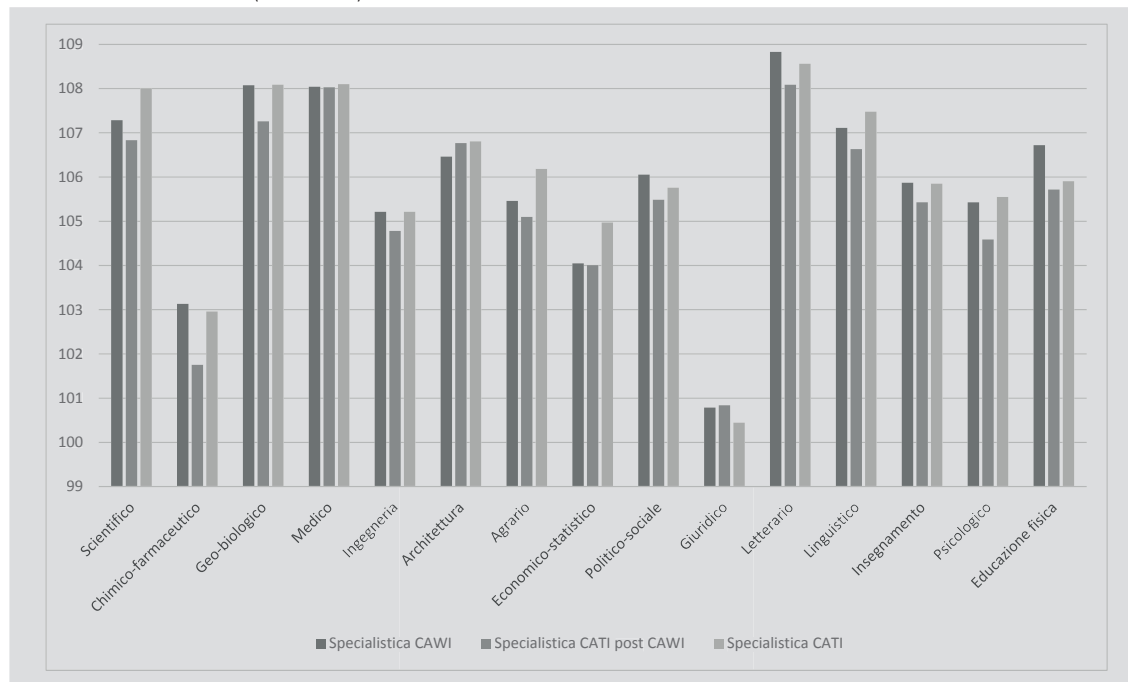


Figura 4.6 - Voto di laurea per tecnica di rilevazione e gruppo di laurea triennale (valori medi)

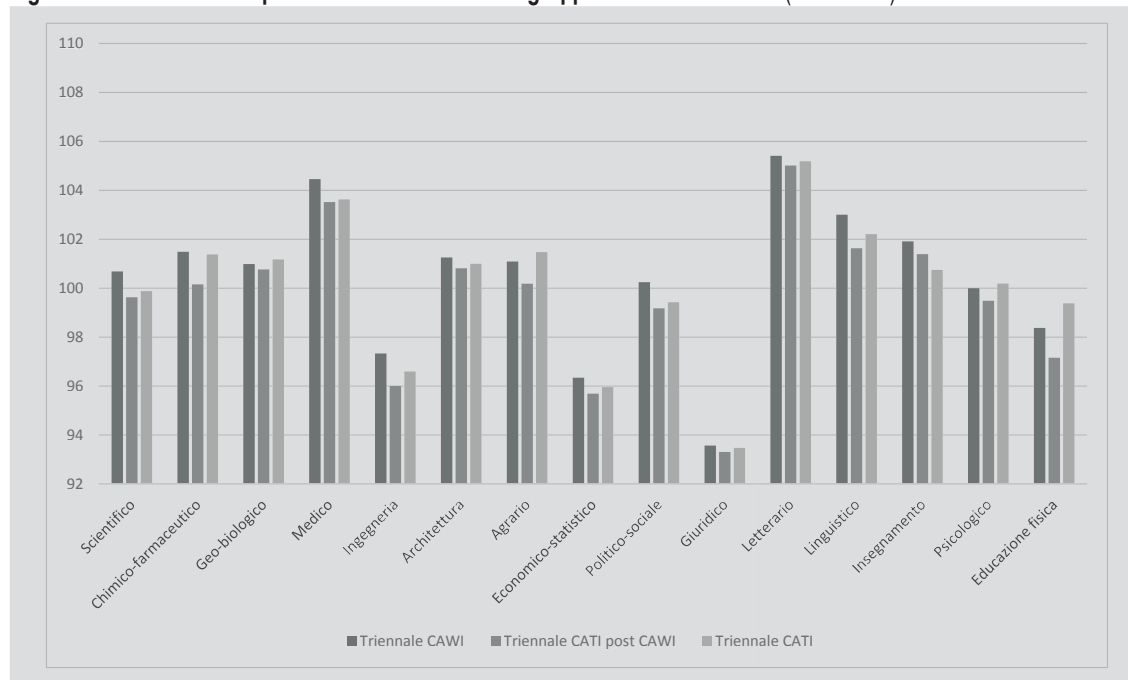
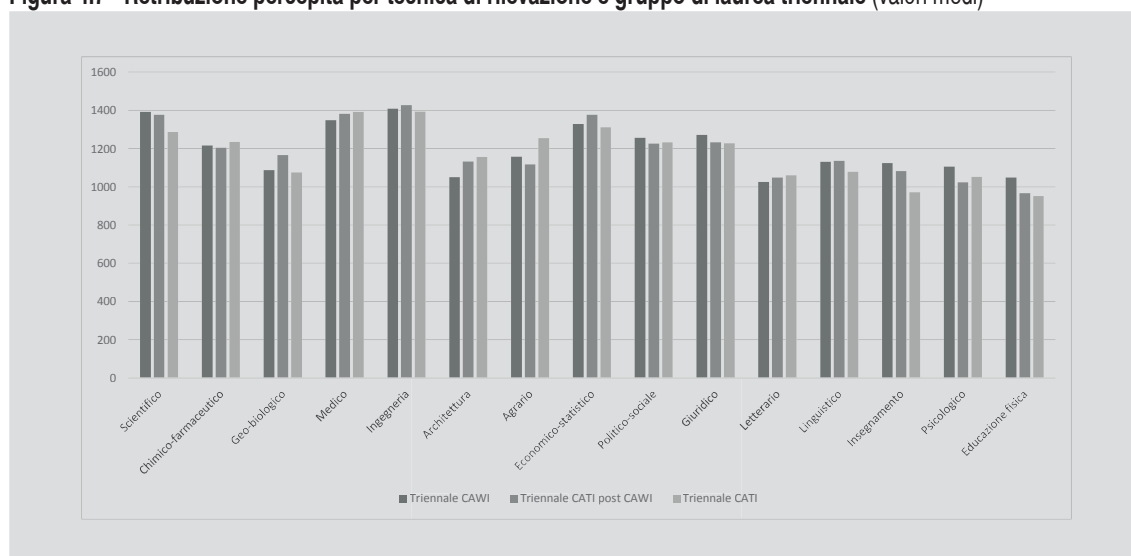


Figura 4.7 - Retribuzione percepita per tecnica di rilevazione e gruppo di laurea triennale (valori medi)



4.2.3 Analisi mediante indicatori statistici

In questa sezione vengono proposti alcuni indicatori utilizzabili per ottenere informazioni sintetiche sull'entità delle eventuali differenze fra distribuzioni marginali o congiunte e fra livelli di associazione tra variabili osservate utilizzando tecniche differenti. È opportuno che questi indicatori siano calcolati per "strati" (o domini o gruppi) omogenei, in modo da depurare i confronti da caratteristiche strutturali delle unità osservate, che possono influire sulla propensione delle unità stesse a rispondere con l'una o l'altra tecnica. I gruppi omogenei di analisi possono anche essere definiti sfruttando le informazioni ricavate dai paradati (cfr. Paragrafo 2.5).

Alcuni semplici indicatori (ripresi da Della Rocca *et al.*, 2003) sono elencati nella Tavola 4.8, separatamente per variabili di natura categorica (nominali o ordinali) o continua.¹⁴

Sia n la numerosità complessiva del campione teorico S di una indagine di tipo *mixed-mode*, e siano A e B due sottoinsiemi disgiunti di unità di S che abbiano risposto utilizzando due tecniche differenti. I sottoinsiemi A e B possono corrispondere a campioni indipendenti di numerosità n_1 + $n_2 = n$, oppure a sotto-campioni non indipendenti di S .

Siano Y ed X variabili osservate su S . Nella Figura 4.8 si riportano schematicamente statistiche univariate e indicatori sintetici che possono essere utilizzati, a seconda della natura di Y e X , per

Figura 4.8 - Indicatori per variabili osservate con tecniche differenti, per criterio di confronto e tipo di variabile

CRITERIO DI VALUTAZIONE	TIPO DI VARIABILI		
	CATEGORICHE		CONTINUE
	NOMINALI	ORDINALI	
Differenze fra distribuzioni marginali/univariate	Differenze relative fra indici di posizione (mediane, quartili, etc.)		Differenze relative fra indici statistici (medie, mediane, quartili, varianze, etc.)
	Indice di dissomiglianza di Leti (I_{m1}) Indice di dissomiglianza quadratico (I_{m2})		
Differenze fra distribuzioni congiunte/livelli di associazione	Indice di dissomiglianza bivariato (I_{j1}) Coefficiente di contingenza V di Cramer ¹⁵		Differenze relative fra: • Indice di correlazione di Pearson • Covarianza

¹⁴ Per il calcolo di questi indicatori è stata sviluppata una procedura generalizzata in linguaggio SAS, per la quale si rimanda all'Appendice.

¹⁵ Per il calcolo di questi indicatori è stata sviluppata una procedura generalizzata in linguaggio SAS, per la quale si rimanda all'Appendice.

avere indizi sulla presenza di un effetto tecnica sulle distribuzioni marginali/univariate, sulle distribuzioni congiunte e i livelli di associazione fra variabili.

Si analizzi il caso in cui Y ed X siano *variabili categoriche* rispettivamente con K e J modalità.

a) Differenze fra distribuzioni marginali

In questo caso, per variabili ordinali è ovviamente possibile utilizzare le usuali statistiche di posizione (mediane, quartili, distanze interquartiliche).

Per variabili sia ordinali, sia nominali, è possibile poi effettuare confronti mediante indici sintetici. Siano $f_{y_k^A}$ e $f_{y_k^B}$ le distribuzioni marginali della variabile Y nei (sotto)campioni A e B . Una misura dell'entità delle differenze fra tali distribuzioni può essere ottenuta mediante il seguente *Indice di dissomiglianza* (Leti, 1983):

$$I_{m1} = \frac{1}{2} \sum_{k=1}^K |f_{y_k^A} - f_{y_k^B}|.$$

Tale indice assume valore minimo (0) nel caso di uguaglianza fra le due distribuzioni, e valore massimo (1) nel caso di massima dissimilarità fra esse (caso in cui in ciascuna distribuzione tutte le unità presentano una stessa modalità che però è diversa per le due distribuzioni). È evidente che valori dell'indice prossimi allo zero indicano una potenziale assenza di effetto tecnica sulle distribuzioni marginali della variabile considerata.

Il corrispondente indice quadratico, di analoga interpretazione, è dato dall'espressione:

$$I_{m2} = \sqrt{\frac{1}{2} \sum_{k=1}^K (f_{y_k^A} - f_{y_k^B})^2}$$

b) Differenze fra distribuzioni congiunte

Siano $f_{y_k^A}$ e $f_{y_k^B}$ le distribuzioni congiunte delle variabili di Y e X (nominali o ordinali) nei (sotto)campioni A e B . Un indice sintetico della differenza fra tali distribuzioni è rappresentato dalla seguente espressione:

$$I_{j1} = \frac{1}{2} \sum_y \sum_x |f_{yx^A} - f_{yx^B}|.$$

Tale indice varia in $[0,1]$. Valori prossimi allo zero di I_{j1} indicano uguaglianza fra le distribuzioni (potenziale assenza di effetto tecnica).

c) Differenze fra livelli di associazione

Il livello di associazione statistica fra due variabili categoriche Y ed X può essere misurata utilizzando il Coefficiente di contingenza V di Cramer (Leti, 1983):

$$Cramer = \left\{ \frac{\chi^2}{n \times \min(k-1, j-1)} \right\}^{1/2},$$

dove χ^2 è l'indice quadratico di contingenza (*Chi-quadrato*) basato sugli scarti tra le frequenze della tabella a doppia entrata fra Y e X , e le corrispondenti frequenze teoriche di indipendenza. *Cramer* assume valori fra 0 (indipendenza) e 1 (associazione completa). Il confronto fra i valori del

coefficiente sui sottoinsiemi di valori osservati con due differenti tecniche A e B fornisce un'indicazione di massima sulla presenza di un effetto tecnica sui livelli di associazione fra Y e X nei gruppi a confronto.

Per quanto riguarda l'analisi mediante indicatori statistici di *variabili continue*:

- indizi sulla presenza di *effetto tecnica sulle distribuzioni univariate* possono essere direttamente ricavati dall'analisi delle differenze fra le usuali caratteristiche delle distribuzioni semplici poste a confronto (medie, mediane, quartili, deviazione standard, etc.);
- la presenza di un *effetto tecnica sulle relazioni fra due variabili continue* Y e X può essere indagata in fase di analisi preliminare dei dati analizzando comparativamente i coefficienti di correlazione e/o le covarianze fra Y e X calcolati nei (sotto)insiemi A e B .

4.2.4 Analisi mediante test statistici

In questo paragrafo sono richiamate alcune ben note *statistiche test* per evidenziarne il possibile utilizzo ai fini della verifica di ipotesi sulla significatività delle differenze fra distribuzioni marginali e/o congiunte o sui livelli di associazione fra variabili osservate utilizzando tecniche differenti. Attraverso queste analisi dunque si mira ad acquisire informazioni non solo sulla potenziale presenza, ma appunto anche sulla significatività statistica di un eventuale *mode effect* presente nei dati.

È opportuno che i test, analogamente a quanto indicato nel precedente Paragrafo 4.2.3, siano effettuati su sottopopolazioni omogenee di unità campione,¹⁶ al fine di tener conto di variabili strutturali note (connesse alla diversa propensione a rispondere con una tecnica piuttosto che con un'altra), e di isolare potenziali effetti sulle misurazioni non dovuti alla tecnica.

Nella trattazione dell'argomento, si farà riferimento separatamente ai casi in cui le diverse tecniche di raccolta siano applicate su *campioni indipendenti* o *non indipendenti*. Come discusso nel Paragrafo 3.2, infatti, nell'ambito delle rilevazioni multi-tecnica le varie tecniche possono essere utilizzate su campioni della popolazione obiettivo estratti o meno in maniera indipendente.

Per ogni tipo di test, vengono anche indicate le procedure del linguaggio di programmazione SAS (*Proc*) e/o i riferimenti a funzioni del linguaggio di programmazione R che possono essere utilizzati per la sua applicazione.

a) Verifica di ipotesi per tecniche utilizzate su campioni indipendenti

Nel caso di campioni indipendenti è possibile utilizzare gli strumenti inferenziali classici (parametrici o non parametrici) per la verifica di ipotesi sulla significatività delle differenze fra distribuzioni o fra livelli di associazione. In questo caso è in pratica possibile ipotizzare l'assenza di un effetto selezione e, di conseguenza, valutare la significatività delle differenze fra le misurazioni nei due campioni dovute alla sola tecnica di rilevazione. Questa situazione si verifica, come già accennato nel Paragrafo 3.2, in disegni *mixed-mode* in cui è stato stabilito di utilizzare le differenti tecniche in modo concorrente su sottopopolazioni differenti, oppure nei casi in cui siano disponibili i dati di una indagine pilota (cfr. Paragrafo 3.5).

Nella Figura 4.9 sono indicati alcuni test utilizzabili, per natura della variabile oggetto di studio (categorica - nominale o ordinale - o numerica continua) e per criterio di valutazione (effetto sulle distribuzioni marginali, oppure effetto sui livelli di associazione fra variabili).

¹⁶ Corrispondenti ad esempio ai domini di stima della rilevazione, o a particolari stratificazioni della popolazione.

4. Metodi diagnostici per la valutazione e la misurazione dell'effetto tecnica

Figura 4.9 - Alcuni test statistici per la verifica di ipotesi nel caso di tecniche differenti su campioni indipendenti

CRITERIO DI CONFRONTO	TIPO DI VARIABILI		
	NOMINALI	ORDINALI	CONTINUE
	<u>Test non parametrici</u> Test Chi-Quadrato (Proc FREQ)		
	<u>Test parametrici</u> Test su parametri di regressione logistica (Proc LOGISTIC)		
Differenze fra distribuzioni marginali/univariate	<u>Test non parametrici</u> Test esatto di Fisher (Proc FREQ)	<u>Test non parametrici</u> - Test di Mann-Whitney (Proc POWER; Proc NPAR1WAY) - Test di Kolmogorov-Smirnov (Proc NPAR1WAY)	
		<u>Test parametrici</u> - Test su parametri di regressione lineare (Proc REG) - Test T di Student (Proc TTEST; Proc POWER) - Analisi della varianza (Proc ANOVA) - Test F di Fisher (Proc ANOVA) - Altri test per varianze: Bartlett, Levene (https://cran.r-project.org/doc/contrib/Frascati-FormularioStatisticaR.pdf)	
Differenze fra livelli di associazione		<u>Test parametrici</u> Trasformata z di Fisher su parametri di correlazione (Proc CORR; https://cran.r-project.org/doc/contrib/Frascati-FormularioStatisticaR.pdf)	

Test sulle differenze fra distribuzioni marginali

Nel caso di una *variabile categorica* Y (nominale o ordinale), la presenza di *mode effect* sulle distribuzioni marginali osservate utilizzando due tecniche differenti T_1 e T_2 può essere verificata utilizzando un modello di regressione logistica del tipo:¹⁷

$$P(Y=1|X) = \text{logit}(\pi(X)) = \beta_0 + \beta_0 X_{\text{tecnica}} + \beta_1 X_{\text{Aux1}} + \beta_2 X_{\text{Aux2}} + \dots + \beta_k X_{\text{Auxk}} + \varepsilon$$

dove:

- X_{tecnica} è una variabile dicotomica che vale 1 (uso della tecnica T_1) oppure 0 (uso della tecnica T_2);
- $X_{\text{Aux1}}, X_{\text{Aux2}}, \dots, X_{\text{Auxk}}$ sono variabili strutturali (ad es. ripartizione territoriale, genere, età, etc.) utilizzate per “depurare” il confronto da differenze strutturali nelle unità campione;¹⁸
- i residui ε sono soggetti alle usuali ipotesi di normalità, indipendenza e omoschedasticità.

Sotto l'ipotesi nulla $H_0: \beta_0 = 0$ (assenza di influenza della tecnica sui valori di Y) la significatività del test (rifiuto dell'ipotesi nulla) può essere interpretata in termini di presenza di effetto tecnica su Y al prefissato livello di confidenza. Un aspetto cruciale nell'utilizzo di modelli di questo tipo è naturalmente quello della scelta delle variabili ausiliarie da utilizzare come covariate (per una discussione più articolata su questi aspetti, cfr. Paragrafo 5.4.2).

Uno strumento comunemente utilizzato nel contesto di rilevazioni di tipo mixed-mode per verificare la significatività dell'associazione fra i valori di una variabile Y osservati con due tecniche differenti è il ben noto *test Chi-Quadrato* (anche indicato con il simbolo χ^2).

Siano T_1 e T_2 le tecniche utilizzate per osservare Y su due campioni indipendenti A e B di numerosità rispettivamente n_1 e n_2 e siano Y_A e Y_B le relative distribuzioni marginali. Il test Chi-Quadrato

¹⁷ Per una descrizione più esaustiva del modello di regressione logistica, cfr. Paragrafo 5.4.2.

¹⁸ Più in generale, possono essere utilizzate le informazioni ausiliarie disponibili nei registri statistici e/o in fonti amministrative

Tavola 4.12 - Esempio di tabella di contingenza per il test χ^2 per variabili categoriche osservate con tecniche differenti su campioni indipendenti

TECNICA	MODALITÀ DELLA VARIABILE Y			Totale
	Y _a	Y _b	Y _c	
T1	n _{1a}	n _{1b}	n _{1c}	n _{1.}
T2	n _{2a}	n _{2b}	n _{2c}	n _{2.}
Totale	n _{.a}	n _{.b}	n _{.c}	n _{..}

può essere impostato sia su una tabella di contingenza a doppia entrata fra le Y_A e Y_B , sia su una tabella di contingenza del tipo di quella rappresentata nella Tavola 4.12. In entrambi i casi, l'ipotesi nulla è quella di indipendenza, che, se accettata, equivale ad ammettere la presenza di un effetto tecnica significativo sulla variabile analizzata. Al contrario, in caso di rifiuto dell'ipotesi nulla, non sarà possibile affermare, sulla base dei dati campionari posti a confronto, la presenza di un effetto tecnica statisticamente significativo (al prefissato livello di confidenza).

Il test di *Kolmogorov-Smirnov* (KS nel seguito) è un test non parametrico applicabile però a dati almeno ordinali. Nel caso della tecnica mista, esso può essere utilizzato per confrontare le distribuzioni di due campioni indipendenti al fine di verificare l'ipotesi che essi provengano dalla stessa popolazione. Il test non richiede alcuna ipotesi sulla distribuzione campionaria (può essere considerato in pratica l'alternativa non parametrica al *test T di Student* per variabili numeriche continue). È più potente del test χ^2 e dunque da preferire.

Il *test esatto di Fisher* è fra i test non parametrici esatti utilizzabili per piccoli campioni (quindi nel caso di distribuzioni sparse, asimmetriche oppure con code "pesanti"), per l'analisi di variabili nominali o dicotomiche (tabelle 2x2), oppure per generiche tabelle di contingenza rxc (cfr. Freeman e Halton, 1951). Questo test può essere dunque usato nel contesto della tecnica mista per verificare se i dati dei due campioni indipendenti siano compatibili con l'ipotesi nulla che le loro popolazioni di origine abbiano la stessa distribuzione campionaria, e che di conseguenza le differenze osservate siano puramente casuali. Se i campioni sono sufficientemente numerosi, si può usare in alternativa il test χ^2 , che però è esatto solo asintoticamente (cioè per dimensioni campionarie molto grandi), mentre il *test esatto di Fisher* è, come dice il nome, sempre esatto.

Il *test di Mann-Whitney* (o *test U di Mann-Whitney*) è uno dei più potenti test non parametrici per campioni statistici indipendenti e variabili ordinali o continue (cfr. Conover, 1999). L'ipotesi nulla del test è che i due campioni siano estratti da una stessa popolazione, e che dunque le loro distribuzioni di probabilità siano uguali. Per campioni di grandi dimensioni può essere preferito al *test KS*.

Nel caso di una *variabile continua*, il confronto fra le distribuzioni di tale variabile nei campioni osservati con le diverse tecniche può essere basato sui classici test parametrici *T di Student* e/o *ANOVA* (per la verifica della differenze fra medie), e *F di Fisher* (per la verifica dell'omogeneità delle varianze), sui quali non è ovviamente necessario soffermarsi.¹⁹ Oltre al *test F di Fisher*, nel contesto mixed-mode su campioni

¹⁹ È importante ricordare però che, come noto, le ipotesi alla base di tali test sono, oltre all'indipendenza dei campioni posti a confronto, la forma gaussiana delle distribuzioni, e, per i test sulle medie, l'omoschedasticità delle varianze. Per quanto riguarda l'ipotesi di normalità, in base al teorema del limite centrale, essendo i campioni delle rilevazioni Istat di numerosità elevata, tale ipotesi essa si considera sempre verificata. Per quanto riguarda l'ipotesi che i diversi campioni abbiano varianze uguali, è noto che allontanarsi sensibilmente da questa condizione influenza gravemente la significatività dei test sulle medie. Oltre alla verifica delle condizioni di validità per il confronto tra medie, questi test vengono utilizzati specificatamente per il confronto tra varianze di campioni indipendenti: ad esempio, una tecnica più

indipendenti è possibile ovviamente utilizzare altri test parametrici per la verifica dell'omogeneità delle varianze. Fra questi, quelli da preferire sono:

- il *test approssimato di Bartlett*, in cui i p gruppi a confronto ($p \geq 2$) possono contenere un numero differente di osservazioni.²⁰ Questo test si basa sull'ipotesi di normalità delle distribuzioni da cui sono estratti i campioni. Al fine di verificare la significatività delle differenze tra i p gruppi, il test utilizza la distribuzione $\chi^2_{(p-1)}$.
- il *test di Levene* è un test ritenuto più robusto, rispetto a scostamenti dall'ipotesi di normalità delle distribuzioni, degli altri test di confronto tra varianze basati sulle distribuzioni F e $\chi^2_{(p-1)}$. Esso può essere applicato anche a campioni non bilanciati.

La presenza di un effetto tecnica su distribuzioni marginali può infine essere verificata attraverso l'uso di un *modello di regressione lineare* del tipo:²²

$$Y = \beta + \beta_0 X_{Tecnica} + \beta_1 X_{Aux1} + \beta_2 X_{Aux2} + \dots + \beta_k X_{Auxk} + \varepsilon$$

Un modello di questo tipo consente di verificare, focalizzando l'attenzione sul parametro β_1 , la significatività dell'effetto dell'uso di tecniche differenti su Y , depurando tale valutazione dall'effetto di variabili ausiliarie ritenute esplicative della propensione dei rispondenti a utilizzare l'una o l'altra tecnica.

Test sulle differenze fra livelli di associazione

Nel caso di *variabili continue*, la verifica dell'influenza della tecnica sui livelli di correlazione può essere effettuata utilizzando il test basato sulla *trasformazione z_r di Fisher* (Fisher, 1970).²³ Date le correlazioni campionarie r_1 e r_2 ottenute da due campioni indipendenti rispettivamente di numerosità n_1 e n_2 , il test z_r consente di verificare l'ipotesi di uguaglianza delle correlazioni delle corrispondenti popolazioni ($H_0: \rho_1 = \rho_2$) al prefissato livello di confidenza.

Nel caso di *variabili categoriche*, per gli scopi di questo volume, il *Coefficiente di contingenza di Cramer* illustrato nel Paragrafo 4.2.3 può essere considerato sufficiente a valutare la presenza di effetto tecnica sui livelli di associazione fra coppie di variabili.

b) *Verifica di ipotesi per tecniche utilizzate su campioni non indipendenti*

Come visto nel Paragrafo 3.2, nel contesto della multi-tecnica è possibile che si ricorra a disegni campionari che prevedono l'uso delle diverse tecniche su campioni non indipendenti. Questo si verifica, ad esempio, quando le tecniche di raccolta dati sono applicate in modo sequenziale su un unico campione rappresentativo della popolazione obiettivo, oppure nel caso in cui si siano effettuate (ad esempio a fini di

efficace in termini di qualità del dato raccolto dovrebbe fornire varianze minori rispetto a tecniche di minor qualità, a parità di altre condizioni.

20 A differenza, ad esempio, dei test F_{\max} di Hartley (valido anche per campioni non indipendenti) e del test di Cochran, che richiedono invece che i gruppi a confronto siano bilanciati (cioè abbiano lo stesso numero di osservazioni).

21 Infatti, con p misure di varianze campionarie s^2 che abbiano gradi di libertà ν_i , eventualmente tra loro diversi, estratte casualmente da popolazioni distribuite in modo normale, tale test segue una distribuzione $\chi^2_{(p-1)}$.

22 Sotto le usuali assunzioni di validità.

23 Per una correlazione campionaria r che usa un campione da una distribuzione normale bivariata con coefficiente di correlazione di Pearson $\rho=0$, la statistica z_r è ottenuta mediante una trasformazione monotona di r .

Figura 4.10 - Esempi di test statistici per la verifica di ipotesi per campioni non indipendenti

CRITERIO DI CONFRONTO	TIPO DI VARIABILI		
	CATEGORICHE		CONTINUE
	NOMINALI	ORDINALI	
Differenze fra distribuzioni marginali/univariate	<u>Test non parametrici</u> Test McNemar (Proc FREQ; https://cran.r-project.org/doc/contrib/Frascati-FormularioStatisticaR.pdf)		<u>Test non parametrici</u> Test di Wilcoxon (Proc UNIVARIATE)
	<u>Test parametrici</u> Test su modelli di <i>regressione logistica</i> (Proc LOGISTIC)		<u>Test parametrici</u> Test T di Student "per campioni appaiati" (Proc TTEST) Test su modelli di regressione lineare (Proc REG)

valutazione dell'errore di misura) delle reinterviste su sottoinsiemi di unità campione opportunamente selezionate.

In questi contesti, i test proposti per campioni indipendenti sono utilizzabili solo nel caso in cui si possa agire sui sottoinsiemi di unità, rilevate con le differenti tecniche, in modo da "depurare" le misurazioni da effetti non dovuti alla tecnica stessa, approssimando per questa via l'ipotesi di indipendenza.

In alternativa, possono essere utilizzati test specifici al caso appunto di campioni *non indipendenti*. Nella Figura 4.10 sono proposte alcune possibili soluzioni al problema della verifica di ipotesi su differenze fra distribuzioni marginali di variabili osservate con due tecniche differenti sui campioni non indipendenti, per natura delle variabili.

Nel caso di *variabili almeno ordinali* (quindi ordinali e numeriche continue), il *test di Wilcoxon* (o *Wilcoxon Signed Rank Test*, cfr. Lehmann, 1998) è uno dei più potenti test non parametrici per la verifica dell'ipotesi che due campioni non indipendenti di numerosità n_1 ed n_2 (uguali o differenti) siano estratti dalla stessa popolazione, cioè che sia vera l'ipotesi nulla che la media (o la mediana) delle due distribuzioni siano uguali fra loro e dunque ad uno stesso valore a μ_0 . L'ipotesi di base è di simmetria delle distribuzioni poste a confronto. Questo test può essere considerato in pratica l'analogo del *test di Mann-Whitney* per campioni non indipendenti.

Nel caso di una *variabile dicotomica* Y con le tecniche T_1 e T_2 , la situazione è quella rappresentata schematicamente nella Tavola 4.13. In tale contesto, l'analisi delle distribuzioni marginali può essere effettuata mediante il test non parametrico di *McNemar* (McNemar, 1947). Se, per un certo livello di confidenza, il test conduce ad accettare l'ipotesi nulla di uguaglianza fra le distribuzioni, sarà lecito supporre l'assenza di un effetto tecnica significativo su Y . Nel contesto della tecnica mista, questo test può anche essere utilizzato in una situazione sperimentale in cui un (sotto)campione di unità venga intervistato due volte con le due tecniche poste a confronto.

Per una *variabile Y nominale o ordinale*, la presenza di un *mode effect* può essere verificata attraverso l'uso di un *modello di regressione logistica* analogo a quello illustrato con riferimento alla verifica delle differenze fra marginali di variabili categoriche osservate su campioni *indipendenti*.

Tavola 4.13 - Esempio di tabella di contingenza per il test McNemar per la verifica di ipotesi su distribuzioni di variabili dicotomiche osservate con tecniche differenti su campioni non indipendenti

TIPO DI TECNICA	MODALITÀ DELLA VARIABILE Y		Totale
	Y_a	Y_b	
T1	n_{1a}	n_{1b}	n_1
T2	n_{2a}	n_{2b}	n_2
Totale	n_a	n_b	$n..$

Analogamente, nel caso di *variabili continue*, la presenza di effetto tecnica sulle distribuzioni univariate può essere verificata utilizzando un *modello di regressione lineare* analogo a quello illustrato con riferimento al problema della verifica delle differenze fra distribuzioni univariate di variabili continue osservate su campioni *indipendenti*.

4.3 Metodi basati su modello

In questo paragrafo sono descritti alcuni metodi che utilizzano modelli statistici utili a diagnosticare la presenza dell'effetto tecnica nelle stime finali di una indagine, nel caso dell'utilizzo di indagini multi-tecnica. L'obiettivo è fornire informazioni sulla significatività statistica delle differenze fra le misurazioni ottenute con le differenti tecniche utilizzate, e di scorporare le diverse componenti dell'effetto tecnica.

4.3.1 Valutazione dell'effetto selezione

L'effetto selezione dipende dal fatto che tecniche di raccolta dei dati differenti presentano tassi di copertura e di risposta differenti. Le dimensioni della non risposta e della copertura dell'indagine sono strettamente correlate alla tecnica di raccolta dei dati. Infatti, metodi di indagine diversi possono attirare diversi tipi di rispondenti a seconda delle loro caratteristiche (ad esempio il possesso o meno di un telefono o l'accesso a Internet, la preferenza per una certa tecnica di raccolta etc.). Utilizzando indagini multi-tecnica si possono incrementare i tassi di risposta, dal momento che ciascun elemento del campione può scegliere il metodo preferito per rispondere, e ridurre l'errore di copertura poiché l'utilizzo di più tecniche di rilevazione consente di raggiungere soggetti altrimenti difficili da contattare.

Tuttavia, le caratteristiche di un campione ottenuto con indagini multi-tecnica devono essere tenute sotto controllo in quanto lo stesso può non rispecchiare il disegno di indagine iniziale. E' quindi necessario considerare le implicazioni dell'utilizzo di modalità diverse di indagine sul campione sia in fase di progettazione dell'indagine, ad esempio utilizzando esperimenti controllati per l'assegnazione delle unità campionarie alle diverse tecniche, sia a conclusione della fase di raccolta dei dati, nel momento del confronto tra le stime dell'indagine. Infatti, sia l'errore di copertura che l'errore di non risposta causati dalla tecnica di indagine sono fonte di distorsione: solo per fare un esempio, quando si utilizza il *web* o il telefono si genera facilmente sottocopertura della popolazione, poiché la disponibilità di questi mezzi di comunicazione non è universale, a differenza delle interviste faccia a faccia o tramite questionario cartaceo è possibile raggiungere più facilmente la popolazione di riferimento. Inoltre, le indagini *web* di solito hanno tassi di risposta più bassi, ma allo stesso tempo consentono di raggiungere soggetti con particolari caratteristiche e difficili da contattare altrimenti.

In presenza di un disegno multi-tecnica, l'analisi delle caratteristiche dei rispondenti è fondamentale per identificare e valutare la distorsione dovuta ad errori di copertura e non risposta. Inoltre, ai fini di valutare l'effetto tecnica sulla precisione delle stime, sarebbe necessario confrontare rispondenti e non rispondenti (cfr. Holbrook *et al.*, 2007) attraverso (1) interviste dei non rispondenti con una indagine

di *follow-up*; (2) il confronto tra: coloro che hanno risposto subito al questionario o che non si sono mai rifiutati di rispondere e i rispondenti tardivi o con i soggetti che inizialmente hanno rifiutato di partecipare all'indagine, sotto l'ipotesi che questi ultimi siano più simili ai non rispondenti; (3) il confronto tra: gli intervistati che hanno partecipato a tutte le fasi di *follow-up* di una indagine panel e coloro che invece hanno abbandonato l'indagine; (4) il confronto delle caratteristiche degli intervistati con quelle della popolazione complessiva o con quelle dei non rispondenti, nel caso si abbiano a disposizione informazioni esterne su entrambi i gruppi.

4.3.2 Valutazione dell'effetto misurazione

In base alla recente letteratura, si distingue tra tre principali approcci utili a diagnosticare la presenza dell'effetto misurazione nelle stime finali di una indagine, nel caso dell'utilizzo di indagini multi-tecnica:

1. l'approccio controfattuale, *potential outcome approach*, (Rubin, 1974; Van-nieuwenhuyze *et al.*, 2014), che presuppone l'esistenza di effetti di misurazione sistematici dovuti alla tecnica di raccolta dati utilizzata e si pone in una situazione ipotetica (controfattuale) in cui ciascun rispondente potrebbe, in linea di principio, rispondere utilizzando tutte le tecniche a disposizione. Questo approccio sarà illustrato in dettaglio nel Capitolo 5;
2. l'analisi fattoriale confermativa multigruppo, un approccio a variabili latenti utilizzato in presenza di risposte *multi-item* per verificare se il modello di misurazione, ovvero il modello che esamina la relazione tra le variabili latenti e le loro misure, possa essere considerato equivalente nei diversi metodi di raccolta dei dati. Tale approccio è quello applicato più spesso nel caso di confronti tra le risposte in indagini multi-nazionali o panel. I risultati empirici di un caso di studio che utilizza i dati provenienti dall'indagine Istat "Aspetti della vita quotidiana", condotta nel 2014 e che applica, a titolo esemplificativo, l'analisi fattoriale confermativa multigruppo, saranno descritti nel Capitolo 6;
3. l'approccio basato su una indagine di riferimento, che utilizza una indagine distinta per valutare l'effetto selezione e misurazione in una indagine multi-tecnica.

Analisi fattoriale confermativa multigruppo

Tra i metodi applicati per l'analisi dell'effetto misurazione sulle stime finali di un'indagine multi-tecnica, l'analisi fattoriale confermativa multigruppo è la più utilizzata in presenza di domande con più item (*multi-item*). L'obiettivo di tale metodologia è la valutazione dell'invarianza (equivalenza) del modello di misura del modello fattoriale fra più gruppi, rappresentati dai diversi metodi di raccolta dati, dopo aver tenuto in considerazione l'effetto selezione. Dopo una breve descrizione del modello fattoriale, nel paragrafo vengono illustrati diversi livelli (dal meno al più restrittivo) di invarianza di misura, corrispondenti a diversi vincoli imposti sui parametri del modello.

L'idea di base dell'analisi fattoriale standard è trovare un insieme di fattori (costrutti) latenti, contenenti essenzialmente le stesse informazioni di un determinato insieme di variabili risposta osservate (indicatori/*item*), dove il numero dei fattori

latenti è inferiore al numero delle variabili stesse. In particolare, i fattori latenti devono poter cogliere le dipendenze tra le variabili risposta osservate: se i fattori fossero tenuti fissi, le variabili osservate sarebbero indipendenti. Nell'analisi fattoriale, le deviazioni delle variabili osservate dal valore reale dei fattori latenti sono definite "errori di misurazione", tali errori possono essere casuali o sistematici (Groves *et al.*, 2004).

Uno strumento statistico per esaminare l'invarianza di misura di un modello fattoriale in diversi gruppi indipendenti è l'analisi fattoriale confermativa multigruppo (*Multigroup Confirmatory Factor Analysis*, MCFA). In questo contesto, i gruppi sono rappresentati dai diversi metodi di raccolta dati. L'invarianza di misura viene raggiunta quando i rapporti tra le variabili osservate e i costrutti latenti sono identici nei gruppi rilevati, ovvero vi è la stessa grandezza e direzione dell'errore di misura nei diversi gruppi. Quando vi è invarianza di misura, i dati provenienti da diverse modalità di raccolta possono essere analizzati congiuntamente.

Formalmente, nell'analisi fattoriale il modello di misura (modello di risposta) per le variabili osservate condizionate alle variabili latenti è un modello lineare generalizzato che è specificato tramite un predittore lineare, una funzione *link* e una distribuzione appartenente alla famiglia esponenziale.

Sia y_{ih} la risposta osservata sull'indicatore h ($h = 1, \dots, H$) dell'individuo i ($i = 1, \dots, n$), e sia v_{ih} il predittore lineare del modello di risposta. Il valore atteso della risposta y_{ih} condizionato alle variabili latenti è "legato" al predittore lineare v_{ih} tramite una funzione *link*:

$$g(E(y_{ih} | \boldsymbol{\eta}_i)) = v_{ih}$$

dove $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{iJ})$ è un vettore contenente le J variabili latenti dell'individuo i . Generalmente, J è molto più piccolo di H . Gli indicatori che possono essere utilizzati nel modello possono essere continui, politomici, ordinali, conteggi, durate in tempo continuo, *rankings* e confronti a coppie. La scelta tra le diverse funzioni *link* segue naturalmente i tipi di scala delle variabili osservate; sono inoltre consentite diverse forme di distribuzione per ciascun indicatore h . Gli errori *specifici* di ogni singolo indicatore h (dati dalla relazione della variabile risposta con il predittore lineare) sono impliciti e dipendono dalla distribuzione della risposta.

Per ogni individuo i , il modello fattoriale standard è espresso da:

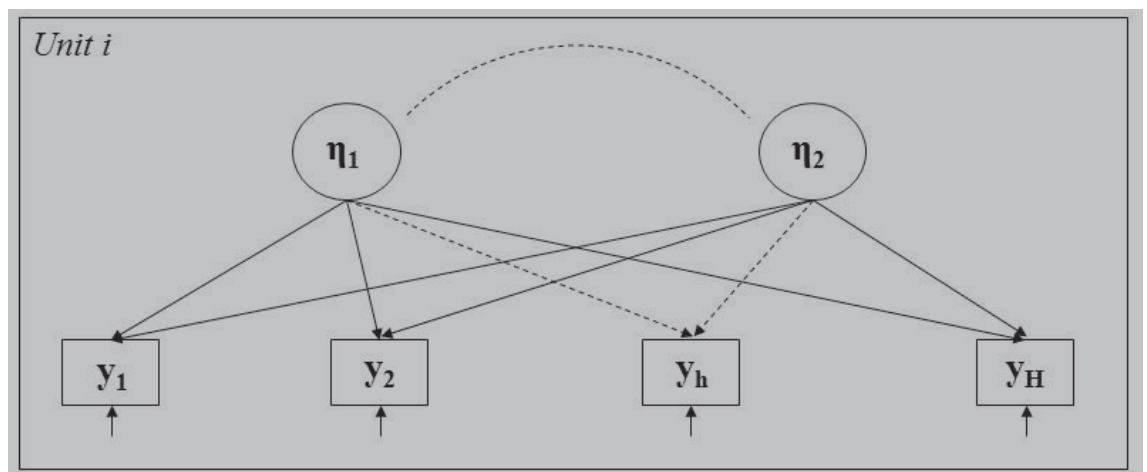
$$v_{ih} = \mu_h + \sum_j \lambda_{hj} \eta_{ij}$$

dove η_{ij} è il valore della variabile latente j -esima ($j = 1, 2, \dots, J$) per l' i -esimo individuo e λ_{hj} rappresenta il relativo peso fattoriale (*factor loading*). Nel modello fattoriale standard viene solitamente assunto che le variabili latenti $\boldsymbol{\eta}_i$ siano indipendenti e distribuite in modo identico con: $\boldsymbol{\eta}_i \sim MN(0, \Psi)$, dove Ψ è una matrice $J \times J$. Generalmente si ipotizza che le H variabili osservate siano indipendenti l'una dall'altra condizionatamente alle variabili latenti $\boldsymbol{\eta}_i$. Tale ipotesi viene comunemente definita ipotesi di indipendenza locale (Bartholomew e Knott, 1999).

I due approcci tradizionali per l'analisi fattoriale sono l'analisi fattoriale esplorativa e confermativa. Il primo non richiede ipotesi a priori né su come gli indicatori siano legati ai fattori sottostanti né sul numero di fattori: strutture di fattori latenti

differenti con differenti strutture di correlazione, generate da diverse rotazioni nello spazio dei fattori, possono produrre la stessa distribuzione congiunta delle variabili osservate e non vi sono motivi per preferire una rotazione - e quindi una soluzione - all'altra. Al contrario, l'analisi fattoriale confermativa analizza i modelli a priori, in cui sono esplicitamente specificati sia il numero di fattori latenti che la corrispondenza tra i fattori stessi con gli indicatori.

Figura 4.11 - Modello a due fattori latenti correlati



A titolo esemplificativo, la Figura 4.11 rappresenta un modello a due fattori latenti, con fattori correlati. I rettangoli rappresentano le variabili osservate che possono essere variabili risposta (continue, binarie, ordinali, nominali, di conteggio) o variabili ausiliarie. I cerchi rappresentano le variabili latenti. Le frecce rappresentano le relazioni di regressione tra variabili, non necessariamente lineari; diversi modelli di regressione vengono utilizzati a seconda della natura delle variabili coinvolte.

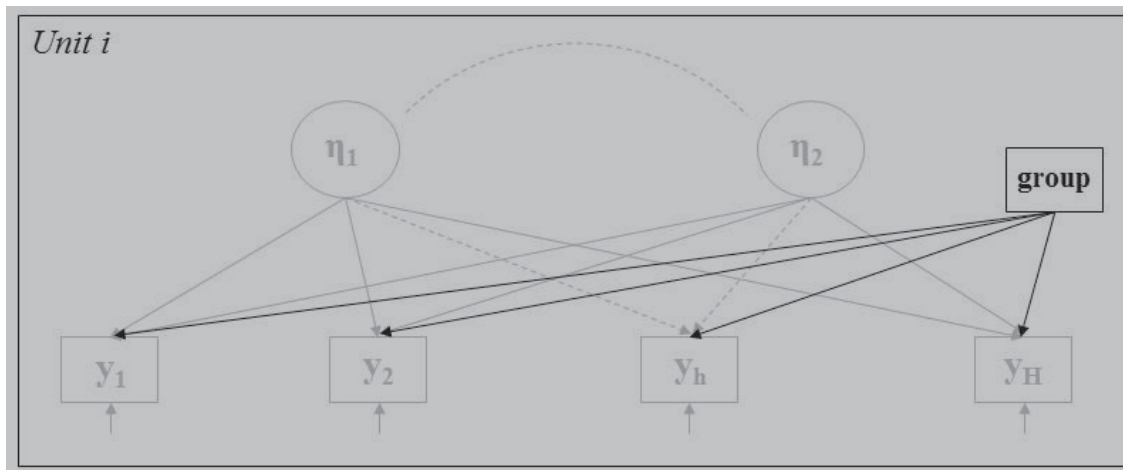
Come anticipato in precedenza, l'analisi MCFA è uno strumento utile per valutare l'invarianza di misura di un modello fattoriale nei vari gruppi (rappresentati, in questo contesto, da diverse modalità di raccolta dati). L'invarianza di misura afferma che, dato il fattore latente, la distribuzione di probabilità dei punteggi osservati non dipende dall'appartenenza ai gruppi, e che i gruppi possono differire solo per quanto riguarda le medie e le (co)varianze dei fattori latenti (Lubke e Muthén, 2004).

La Figura 4.12 mostra un modello fattoriale a due fattori, correlati tra loro, con una covariata osservata che rappresenta il gruppo (modalità di raccolta dati).

La valutazione dell'invarianza di misurazione avviene per passi, con un diverso grado di vincoli sui parametri, dalla struttura di invarianza più debole alla più forte. Se non si trova alcuna evidenza a favore dell'invarianza di misurazione, i dati raccolti con diverse modalità non possono essere analizzati congiuntamente (cfr. Hox et al., 2017). I ricercatori possono decidere quali elementi o matrici di parametri debbano essere testati e in quale ordine questi test dovrebbero essere eseguiti. In questo paragrafo, viene illustrata la gerarchia proposta da Hox *et al.* (2015) nel contesto di dati raccolti con tecniche miste.

Invarianza configurale (*Configural invariance*). Verifica se lo stesso modello fattoriale si applica nei diversi gruppi: questo implica che le stesse variabili osservate siano associate allo stesso fattore latente, ovvero che in ogni gruppo vi sia lo stesso numero di fattori con pesi fattoriali comparabili e i fattori latenti abbiano significati simili. In indagini svolte

Figura 4.12 - Modello multigruppo a due fattori latenti correlati



con tecniche miste di raccolta dati, l'invarianza configurale è generalmente verificata. In tal caso, i costrutti latenti sono gli stessi, ma è necessario valutare se sono misurati in modo diverso utilizzando metodi di raccolta dati differenti.

Invarianza metrica (*Metric invariance*). Verifica che i pesi fattoriali (λ_{η_j}) siano uguali tra i gruppi, ovvero se i rispondenti all'indagine con diverse tecniche di raccolta dati attribuiscono lo stesso significato al fattore latente sotto analisi. L'invarianza metrica consente un'interpretazione valida delle covarianze delle variabili osservate: nei casi in cui sia verificata solo l'invarianza metrica, le differenze tra le medie delle variabili osservate non possono essere interpretate, pur restando valide le interpretazioni su covarianze e correlazioni degli indicatori.

Invarianza scalare (*Scalar invariance*). Verifica se l'intercetta degli indicatori (μ_{η_j}) è uguale tra i gruppi, ovvero che vi sia lo stesso errore sistematico negli indicatori. Quando vi è invarianza metrica, il significato del costrutto latente (i pesi fattoriali) e i livelli degli indicatori (intercette) sono uguali nei diversi gruppi e un'eventuale differenza nelle medie e nelle varianze dei fattori latenti tra i gruppi indica la presenza di un effetto tecnica; più in particolare, qualora l'effetto selezione sia stato preso in considerazione, un'eventuale differenza nelle medie e nelle varianze dei fattori latenti indica la presenza di un effetto misurazione. Quando l'assegnazione ai differenti metodi di raccolta dati è casuale, non ci dovrebbero essere differenze tra le medie e le varianze dei fattori latenti nei gruppi. Quando l'invarianza scalare è verificata solo parzialmente, è ancora possibile confrontare le medie dei fattori latenti, a condizione che almeno due intercette e i relativi pesi fattoriali siano uguali (Byrne *et al.*, 1989). In questo caso, l'obiettivo dell'analisi è scoprire quali pesi fattoriali e intercette si differenziano tra i gruppi: una volta individuato l'elemento problematico, i questionari possono essere modificati di conseguenza (Lugtig *et al.*, 2011).

L'approccio basato su un'indagine di riferimento

Un ulteriore approccio utile a valutare gli effetti selezione e misurazione in un'indagine multi-tecnica utilizza i dati provenienti da un'indagine di riferimento distinta. Tale indagine è (1) un'indagine svolta con un'unica modalità di raccolta dati o (2) un sotto-campione incluso nella progettazione dell'indagine obiettivo, ottenuto assegnando alcune unità ad un sistema di raccolta dati con modalità singola. Nella letteratura più recente sono stati suggeriti

diversi approcci di analisi che utilizzano un'indagine di riferimento (cfr. Vannieuwenhuyze e Loosveldt, 2012): tra questi, l'approccio che adotta un modello di mistura ha l'obiettivo di valutare se esistono differenze tra i rispondenti che utilizzano diverse modalità di raccolta dati. In tale approccio, si ipotizza l'esistenza di una variabile categorica latente le cui classi sono le modalità di indagine. L'appartenenza delle unità alle classi latenti è nota nel campione ottenuto con tecnica mista di raccolta dati, mentre nei dati dell'indagine di riferimento l'appartenenza alle classi latenti è sconosciuta e può essere prevista utilizzando informazioni ausiliarie. L'ipotesi sottostante è che gli intervistati dell'indagine di riferimento avrebbero risposto con una tecnica di raccolta dati differente se gli fosse stata offerta l'opportunità di scegliere. Una volta ottenuta la classificazione degli intervistati nelle diverse classi latenti, è possibile procedere all'analisi delle differenze tra le classi.

4.3.3 Metodi per distinguere l'effetto selezione dall'effetto misurazione

Gli errori dovuti alla mancata risposta e alla copertura sono facilmente confondibili con gli errori di misurazione, tra i quali si possono elencare: quelli dovuti alla tendenza di una persona a ricordare meglio il primo e l'ultimo item di una serie (*serial position effect*), quelli dovuti alla posizione della domanda rispetto alle altre, quelli che dipendono dalla tendenza dell'intervistato a confermare le informazioni che gli vengono presentate dal rilevatore (*acquiescence*) e alla desiderabilità sociale (*social desirability*) (cfr. Heerwegh e Loosveldt, 2011).

Per ridurre la possibilità di osservare differenze significative nelle stime finali di indagine, è importante distinguere e separare l'effetto selezione dall'effetto misurazione. Pur essendo un compito piuttosto complicato, esistono diversi metodi statistici che hanno l'obiettivo di identificare l'effetto selezione dall'effetto misurazione. Di seguito, vengono descritti alcuni esempi.

Il disegno sperimentale. In un disegno sequenziale a tecnica mista, un gruppo casuale di interlocutori passa da una modalità di raccolta dati ad un'altra, generando un campione osservato con una diversa composizione rispetto a quella progettata. In un disegno sperimentale, invece, gli intervistati sono assegnati in modo casuale a una tecnica piuttosto che a un'altra, generando gruppi che sono abbastanza paragonabili rispetto alle variabili di confondimento note o non note. Anche in questo caso è però necessario controllare per la non risposta, in modo da garantire la comparabilità tra i risultati dell'indagine (cfr. Heerwegh, 2009). Solo in fase di progettazione di indagine: qualora questo non fosse possibile, è necessario controllare per le variabili di confondimento dopo la fase di raccolta dei dati.

Validazione con dati esterni. In presenza di dati esterni da utilizzare come *golden standard*, è possibile individuare la tecnica di raccolta che porta a risultati migliori. Tuttavia, questi dati non sono facilmente ottenibili, in particolare su domande attitudinali che usualmente rappresentano le variabili di maggiore interesse, quando vengono analizzati gli effetti dell'utilizzo di indagini *mixed-mode*. Inoltre, quando invece i dati esterni all'indagine sono disponibili, si riferiscono in genere a sotto-popolazioni con caratteristiche molto particolari (cfr. Kreuter *et al.*, 2008).

Modelli statistici. I campioni di indagini multi-tecnica possono essere resi comparabili utilizzando la modellazione statistica. Tra i diversi approcci utili a questo scopo, si possono elencare:

1. la *ponderazione* come aggiustamento da *propensity score*, che utilizza i pesi basati sul *propensity score*, per creare un campione nel quale la distribuzione delle cova-

- riate è indipendente dal metodo di raccolta dati assegnato (cfr. Lee, 2006);
2. la *stratificazione*, che fissa i livelli delle variabili che possono confondere l'effetto del metodo di raccolta dei dati (*confounders*) sulle risposte e produce gruppi entro i quali tali variabili non variano rispetto al metodo di raccolta;
 3. *modelli multivariati*, per correggere le differenze esistenti nella composizione dei campioni, come la regressione lineare o logistica e l'analisi della covarianza, che può utilizzare un numero elevato di covariate in contemporanea (Dillman *et al.*, 2009). Ad esempio, De Leeuw (1992) utilizza l'analisi della varianza ponendo la tecnica di raccolta dei dati e la partecipazione o meno all'indagine come fattori e le dimensioni risultanti da un'analisi di omogeneità svolta su una variabile rilevante di indagine misurata sia sui rispondenti che sui non rispondenti come variabili dipendenti;
 4. *modelli a variabile latente* che possono essere utilizzati in combinazione con la re-intervista (cfr. Biemer, 2001) o con la validazione da dati esterni (Voogt e Saris, 2005), per correggere la distorsione dovuta alla non risposta.

Lo svantaggio di questi modelli è l'assunzione che ogni tecnica di indagine possa potenzialmente coprire l'intera popolazione, cosa che non sempre è verificata (ad esempio come nel caso dei tassi di copertura nelle indagini telefoniche e web).

Propensity score matching. E' utilizzato per separare l'effetto dovuto alla multi-tecnica dalla distorsione dovuta alla non risposta, mettendo in corrispondenza rispondenti simili che hanno risposto con metodi diversi, in un disegno quasi-sperimentale. Dopo il *matching* tra questi, la differenza nella composizione dei campioni può dipendere dalle differenze nelle variabili risposta (Lugtig *et al.*, 2011). Per dettagli, cfr. il Paragrafo 5.4.2.

5. METODI PER LA VALUTAZIONE E IL TRATTAMENTO DELL'EFFETTO TECNICA IN FASE DI STIMA¹

5.1 Introduzione

L'obiettivo principale di un'indagine statistica è ottenere stime accurate dei parametri di interesse. Nelle indagini multi-tecnica le stime sono il risultato della combinazione di dati raccolti con differenti strumenti di rilevazione. In questi casi, l'accuratezza del processo inferenziale è assicurata, oltre che da altri numerosi fattori - copertura, mancata risposta totale, stimatore - soltanto se la misurazione effettuata con più tecniche è equivalente, ovvero se non è violata l'assunzione di invarianza della misurazione delle tecniche (Hox *et al.*, 2015).

Tradizionalmente, i metodi di inferenza statistica utilizzati per trattare i dati raccolti con tecniche miste assumono che l'effetto misurazione sia ignorabile, ovvero che sia determinato dal caso e non dalla tecnica. Partendo da una considerazione di ordine generale, è possibile affermare che tutte le tecniche di indagine generano, per ogni unità della popolazione, valori vicini ai valori veri. Tuttavia, quando più tecniche di rilevazione sono utilizzate per rilevare lo stesso fenomeno non è detto che, tutte, realizzino una misurazione equivalente (cfr. Capitolo 4).

D'altra parte, nelle indagini multi-tecnica, l'ipotesi di ignorabilità dell'effetto misurazione non può essere valutata attraverso test statistici standard delle differenze nei parametri stimati sui campioni di rispondenti con modalità diverse (di seguito indicati semplicemente come *campioni di rispondenti*), senza tener conto dell'effetto selezione dovuto alla non casualità della risposta al mix di tecniche somministrate.

L'effetto selezione rappresenta, di fatto, una conseguenza naturale di questo tipo di indagini, in quanto l'utilizzo di più tecniche di rilevazione consente di raggiungere differenti target di popolazione. Tale effetto non costituisce in sé un limite, ma della diversa composizione dei campioni di rispondenti si deve tener conto, sia per verificare la reale presenza di errori di misura, sia per neutralizzare differenze nelle stime determinate da variazioni della composizione dei campioni di rispondenti nelle successive edizioni dell'indagine.

Nei disegni di indagini di tipo concorrente, la diversa composizione dei campioni è determinata ex-ante ed è legata, generalmente, a ragioni di copertura. Infatti, per garantire la rappresentatività dell'intera popolazione, campioni indipendenti sono selezionati da popolazioni raggiungibili con modalità di rilevazione differenti (ad esempio la tecnica Cati per le famiglie con telefono fisso).

Nei disegni di tipo sequenziale, la diversa composizione dei campioni è determinata dall'autoselezione dei rispondenti che scelgono modalità di risposta differenti (*mode choice*). In questo caso i campioni di rispondenti non sono indipendenti e la loro dimensione diventa una variabile casuale.

¹ Hanno collaborato alla stesura del capitolo: Claudia De Vitiis (paragrafi 5.2.2., 5.3.2., 5.4.2., 5.5.1., 5.5.3.), Francesca Inglese (paragrafi 5.1, 5.2.1., 5.2.3., 5.2.4., 5.3.1., 5.4.1., 5.4.3., 5.4.4., 5.4.5., 5.4.6., 5.5.2.).

L'effetto selezione assume, pertanto, una duplice connotazione: da una parte può essere visto come un vantaggio, dall'altra, come una complicazione poiché, come anticipato nei capitoli precedenti, risulta completamente confuso con l'effetto misurazione determinato da differenze negli errori di misura associabili a diverse tecniche di rilevazione.

La confusione tra i due effetti può condurre a difformità nelle stime di parametri della popolazione calcolate sui campioni di rispondenti che sono causate o dalla diversa composizione dei campioni o da differenze negli errori di misura (de Leeuw, 1992; Weisberg, 2005). Entrambi gli effetti, pur agendo in modo dissimile, sono contenuti nella relazione tra variabile d'indagine e tecnica. L'effetto misurazione produce differenze nelle medie stimate sui diversi campioni di rispondenti, mentre l'effetto selezione, influenzando la struttura di correlazione dei dati, produce misure non equivalenti tra i campioni di rispondenti (Hox *et al.*, 2015).

Da un punto di vista inferenziale è necessario che gli effetti di selezione e misurazione siano indagati separatamente (Vannieuwenhuyze e Loosveldt, 2013), sia per una corretta formulazione dell'errore totale non campionario, sia per l'applicazione di metodi atti a correggere le stime dei parametri di interesse dell'indagine dagli effetti distorsivi determinati da una molteplicità di cause.

Il problema della confusione tra i due effetti costituisce il tema centrale della teoria dell'inferenza causale (Morgan e Winship, 2009; Pearl, 2009; Weisberg, 2010). Detta teoria può essere applicata al contesto delle indagini multi-tecnica in quanto l'errore di misura è concettualizzato come un effetto causale della tecnica sulla variabile d'indagine, mentre l'effetto selezione è visto come una correlazione spuria tra la variabile d'indagine e la tecnica.

Per la stima dei due effetti, l'inferenza causale è usata secondo una prospettiva controfattuale, ovvero ipotizzando l'esistenza di un risultato potenziale non realmente osservato. Tale approccio fornisce un insieme di metodi che mirano a superare il problema controfattuale con modelli di analisi basati su covariate, o che hanno l'obiettivo di costruire dati controfattuali sulla base di modelli di imputazione.

L'applicabilità e la validità di metodi inferenziali basati su modelli è, tuttavia, strettamente connessa alla disponibilità di variabili ausiliarie, provenienti da archivi o dall'indagine stessa, la cui specifica attitudine sia proprio quella di spiegare effetti distorsivi di diversa natura. Condizione essenziale è che tali variabili ausiliarie, se rilevate nell'indagine stessa, non siano influenzate dalla tecnica (*mode insensitive*).

Nelle indagini *mixed-mode* è necessario, poi, considerare l'effetto selezione dovuto alla mancata risposta totale. Per produrre stime meno distorte possibile, gli effetti introdotti dal mix di tecniche devono essere trattati congiuntamente all'effetto selezione determinato dalla mancata risposta totale. Il trattamento della mancata risposta totale, nelle indagini multi-tecnica soprattutto di tipo sequenziale, è un'operazione complessa in quanto il modello di risposta dovrebbe tener conto del processo di risposta, ovvero del fatto che la distribuzione del campione di rispondenti della seconda fase dipende dai risultati della risposta che si realizza in prima fase con una differente tecnica (Cobben *et al.*, 2006).

In generale, l'utilizzo di indagini a tecnica mista non garantisce di per sé una riduzione dell'errore totale, perché a fronte di una riduzione di effetti distorsivi sulle stime dell'indagine dovuti ad un aumento generale del tasso di risposta o alla soluzione di problemi di copertura, altre forme di distorsione, come visto, possono essere introdotte.

L'obiettivo del presente Capitolo è quello di fornire, attraverso una breve e non esaustiva descrizione dei metodi reperibili nella vastissima letteratura sul tema, una visione d'insieme dei metodi utilizzabili per correggere le stime di parametri di interesse in presenza di effetti distorsivi di diversa natura. Il quadro teorico di riferimento dei metodi presentati,

descritto nella prima parte del Capitolo, è l'approccio inferenziale di tipo controfattuale. Infatti, la correzione degli effetti distorsivi introdotti dal mix di tecniche richiede la stima di dati controfattuali, a meno che non siano stati predisposti disegni sperimentali o che per le variabili d'indagine siano ottenibili valori di riferimento da archivi amministrativi.

5.2 L'effetto tecnica nell'approccio controfattuale

5.2.1 Premessa

L'equivalenza della misurazione di tecniche differenti assume particolare rilevanza nelle indagini *mixed-mode*. Per valutare tale aspetto è necessario analizzare le differenze nell'errore di misura determinato dalle tecniche utilizzate (effetti misurazione):

Infatti, per un'indagine multi-tecnica gli effetti misurazione potrebbero condurre a interpretare erroneamente le variazioni delle stime nel tempo, o causare instabilità nella serie storica se la dimensione e composizione dei gruppi di rispondenti con tecniche specifiche, sia in disegni concorrenti che sequenziali, varia nel tempo.

La valutazione dell'equivalenza della misurazione delle tecniche costituisce una fase complessa del processo inferenziale. Gli effetti di misurazione possono essere confusi con gli effetti di selezione, sia se si confrontano singole tecniche sia se si confrontano tecniche concorrenti o sequenziali nelle indagini *mixed-mode*.

Per rendere più agevole la comprensione di tale problema e dei metodi di analisi e di trattamento dell'effetto tecnica descritti nei successivi capitoli, è utile formalizzare alcuni concetti. A tal fine, si presenta dapprima una formalizzazione del concetto di equivalenza della misurazione delle tecniche e, successivamente, si descrivono le componenti dello stimatore dell'effetto tecnica secondo due approcci: il primo basato sul confronto tra due tecniche singole in un disegno di tipo concorrente; il secondo che considera lo specifico contesto delle indagini *mixed-mode* sequenziali.

In entrambi i casi si fa riferimento ad un contesto semplice in cui si prendono in considerazione solo due tecniche. Inoltre, gli effetti selezione e misurazione sono formalizzati considerando una delle due tecniche come riferimento (benchmark). Questa impostazione, tratta dal lavoro di Klausch (2014), porta alla definizione e formalizzazione di una quantità "controfattuale" che è necessario stimare per separare i due effetti.

In questa dissertazione si analizzano soltanto i fattori distorsivi causati dalla componente sistematica dell'errore di misura e da errori di selezione. Non si tiene conto, pertanto, di altre componenti di errore, come gli errori di misura casuali e l'errore campionario.

5.2.2 L'equivalenza della misurazione delle tecniche

Da un punto di vista formale il concetto di equivalenza della misurazione può essere illustrato introducendo un semplice modello di misura. A tal fine si considerino, per la generica variabile casuale y , i valori y_{m_1} e y_{m_2} osservati rispettivamente con le tecniche m_1 e m_2 . Il modello di misura per la variabile y_{m_1} è esprimibile come (Klausch, 2014):

$$E(y_{m_1}) = E(y) + MB_{m_1}$$

dove $E(y)$ e $E(y_{m_1})$ sono rispettivamente il valore atteso della variabile y nella popolazione e il valore atteso della variabile y rilevata con la tecnica m_1 , $MB_{m_1} = E(y_{m_1}) - E(y)$ è l'errore sistematico di misura della tecnica m_1 .

L'equivalenza della misurazione con due tecniche si realizza quando gli errori di misura ad esse associati sono uguali:

$$MB_{m_1} = MB_{m_2} \Leftrightarrow E(y_{m_2}) - E(y_{m_1}) = 0.$$

Tale assunzione è fondamentale per la comparabilità dei dati raccolti con diverse tecniche di rilevazione.

La valutazione dell'equivalenza della misurazione delle tecniche è, come anticipato, resa complicata dal combinarsi di effetti determinati dal processo di selezione e dal processo di misurazione.

Per formalizzare i due effetti è necessario, a partire dalla scomposizione della distorsione totale, definire lo stimatore dell'effetto tecnica totale e le sue componenti.

Per tale scopo si introduce la media \bar{y}_{mm} ottenuta come combinazione di dati raccolti con le tecniche m_1 e m_2 :

$$\bar{y}_{mm} = \pi_{m_1} \bar{y}_{m_1} + \pi_{m_2} \bar{y}_{m_2},$$

in cui π_m e \bar{y}_m indicano rispettivamente la proporzione di rispondenti e la media campionaria per la tecnica m ($m=1,2$).

Si introduce poi il valore atteso dello stimatore \hat{y}_{m_1} della media campionaria \bar{y}_{m_1} relativa ai rispondenti con la tecnica m_1 :

$$E(\hat{y}_{m_1}) = E(y_{m_1} | R_{m_1} = 1).$$

La distorsione totale dello stimatore \hat{y}_{m_1} è definita dalla seguente espressione:

$$\begin{aligned} TB(\hat{y}_{m_1}) &= E(y_{m_1} | R_{m_1} = 1) - E(y) = \\ &= E(y_{m_1} | R_{m_1} = 1) - E(y | R_{m_1} = 1) + E(y | R_{m_1} = 1) - E(y) = \quad (5.1) \\ &= MB_{m_1} + SB_{m_1}, \end{aligned}$$

dove MB_{m_1} è l'errore di misura condizionato alla risposta (R_m è un indicatore di risposta ad una specifica tecnica) dei rispondenti con la tecnica m_1 , e SB_{m_1} è l'errore di selezione rispetto alla media della variabile y nella popolazione.

La distorsione totale dello stimatore \hat{y}_{m_2} è definibile in modo analogo.

Da tale impostazione è facile dedurre che l'impatto del mix di tecniche sull'errore totale di una stima dipende da come agiscono gli errori specifici associabili a ciascuna delle tecniche utilizzate.

Nei paragrafi successivi lo stimatore dell'effetto tecnica totale è riformulato secondo due impostazioni, la prima basata sul confronto tra tecniche singole in un disegno concorrente, la seconda che prende in considerazione un disegno di indagine *mixed-mode* sequenziale.

5.2.3 L'effetto tecnica in un disegno concorrente

Si consideri un disegno di tipo concorrente con due tecniche m_1 e m_2 . Un semplice stimatore dell'effetto tecnica, detto effetto tecnica totale, è:

$$\hat{T}_{SM} = \hat{y}_{m_2} - \hat{y}_{m_1}. \quad (5.2)$$

Il valore atteso dello stimatore (5.2) può essere espresso come:

$$\begin{aligned} E(\hat{T}_{SM}) &= (E(y_{m_2} | R_{m_2} = 1) - E(y)) - (E(y_{m_1} | R_{m_1} = 1) - E(y)) = \\ &= (TB_{m_2} - TB_{m_1}) = \\ &= (MB_{m_2} - MB_{m_1}) + (SB_{m_2} - SB_{m_1}), \end{aligned}$$

dove: TB_m è l'errore totale specifico della tecnica m dello stimatore media rispetto alla media della popolazione; MB_m e SB_m sono rispettivamente l'errore di misura e l'errore di selezione associati alla tecnica e definiti nella 5.1.

Le differenze $(MB_{m_2} - MB_{m_1})$ e $(SB_{m_2} - SB_{m_1})$ si definiscono, rispettivamente, effetto misurazione ed effetto selezione:

$$ME = (MB_{m_2} - MB_{m_1}), \quad (5.3)$$

$$SE = (SB_{m_2} - SB_{m_1}). \quad (5.4)$$

Una stima dei fattori distorsivi e degli effetti di selezione e misurazione è possibile, tuttavia, soltanto nella situazione in cui esiste una misura vera della variabile y per tutti gli individui della popolazione.

Per ovviare al problema della non conoscenza di una misurazione vera di y si può considerare come vera una misurazione effettuata con una tecnica di riferimento, detta di benchmark.

Assumendo, ad esempio, che i valori di benchmark siano y_{m_1} , ovvero quelli osservati con la tecnica m_1 , allora le formule 5.3 e 5.4 diventano:

$$\begin{aligned} ME &= E(y_{m_2} | R_{m_2} = 1) - E(y_{m_1} | R_{m_2} = 1) = ME_{m_2}, \\ SE &= E(y_{m_1} | R_{m_2} = 1) - E(y_{m_1} | R_{m_1} = 1) = SE(y_{m_1}), \end{aligned}$$

dove: ME_{m_2} è l'effetto misurazione condizionato dei rispondenti con la tecnica m_2 , che può essere visto come una variazione della distorsione dovuta all'errore di misura in m_2 ; $SE(y_{m_1})$ è l'effetto selezione rispetto alla variabile misurata con m_1 , che, invece, è una variazione della distorsione dovuta all'errore di selezione generato dall'utilizzo della tecnica m_2 , invece che della tecnica m_1 , per misurare y_{m_1} ; $E(y_{m_1} | R_{m_2} = 1)$ è una quantità "controfattuale", detta anche "risultato potenziale", che in realtà non è osservata, ma che sotto determinate condizioni può essere stimata (Klausch, 2014).

5.2.4 L'effetto tecnica in un disegno sequenziale

Si prenda in considerazione un disegno di indagine multi-tecnica sequenziale, in cui m_1 è la tecnica proposta a tutte le unità del campione nella prima fase della rilevazione, e m_2 è la tecnica somministrata nella seconda fase alle unità non rispondenti con la tecnica m_1 . La tecnica m_2 è assunta come benchmark.

In questo caso lo stimatore dell'effetto tecnica totale, assume la forma

$$\hat{T}_{MM} = \hat{y}_{m_1} - \hat{y}_{(m_1)m_2},$$

dove \hat{y}_{m_1} è la stima della media della variabile y osservata con la tecnica m_1 , e $\hat{y}_{(m_1)m_2}$ è la stima della media della variabile y osservata con la tecnica m_2 per il campione dei non rispondenti con la tecnica m_1 .

Gli effetti condizionati di misurazione e di selezione, analogamente al caso del disegno concorrente, possono essere espressi nella forma:

$$ME_{m_1} = E(y_{m_1} | R_{m_1} = 1) - E(y_{m_2} | R_{m_1} = 1),$$

$$SE(y_{m_2}) = E(y_{m_2} | R_{m_1} = 1) - E(y_{m_2} | R_{m_1} = 0, R_{m_2} = 1).$$

In questo contesto, ME_{m_1} esprime una variazione della distorsione causata dall'errore di misura rispetto alla tecnica di riferimento m_2 , mentre $SE(y_{m_2})$ rappresenta una variazione della distorsione dovuta all'errore di selezione e alla sequenzialità della tecnica m_2 (Klausch, 2014).

Come visto nel paragrafo precedente, gli effetti selezione e misurazione sono confusi nell'effetto tecnica totale, pertanto è necessario stimare un risultato potenziale, $E(y_{m_2} | R_{m_1} = 1)$, al fine di separarli.

La stima degli effetti di selezione e di misurazione può essere effettuata con metodi propri della teoria dell'inferenza causale, come sarà descritto nei prossimi paragrafi.

5.3 Metodi di inferenza causale per la stima dell'effetto tecnica

5.3.1 Modelli di analisi

In generale, per risolvere il problema della confusione degli effetti di misurazione e selezione, la teoria dell'inferenza causale fornisce diversi metodi. Si tratta di metodi che cercano di aggirare il problema controfattuale attraverso l'analisi di modelli che fanno uso di variabili ausiliarie capaci di catturare singoli effetti.

Vannieuwenhuyze e altri autori presentano, in diversi lavori, due principali modelli di analisi: il *back-door model* che utilizza variabili ausiliarie *confondenti* e il *front-door model* che, invece, utilizza variabili ausiliarie *mediatrici*.

Le variabili confondenti del primo modello spiegano l'effetto selezione come causa comune della variabile d'indagine e della tecnica. Le variabili confondenti, in generale, spiegano sia

la variabile dipendente sia la variabile indipendente di un modello, sono variabili di controllo. Il metodo di analisi con variabili confondenti richiede che sia soddisfatta l'ipotesi d'ignorabilità dell'effetto selezione, ovvero che la scelta della tecnica sia determinata soltanto da caratteristiche individuali e che le variabili ausiliarie non siano influenzate dalla tecnica (*mode-insensitive*), ovvero derivino da fonte esterna all'indagine o non siano affette da errori di misura. Rientrano in tale tipologia di variabili, ad esempio, le variabili di tipo socio-demografico, come sesso, età, etc. Tuttavia, esse non sempre sono sufficienti a spiegare l'effetto selezione. Le variabili confondenti dovrebbero cercare, infatti, di misurare la capacità delle unità statistiche ad essere contattate e a rispondere con ogni modalità, oppure di misurare la preferenza degli individui a rispondere all'indagine con una specifica tecnica (Vannieuwenhuyze e Lynn, 2014).

Le variabili mediatrici del secondo modello spiegano l'errore di misura come variabile intermedia tra la variabile d'indagine e la tecnica. Le variabili mediatrici trasmettono l'effetto di una variabile indipendente su una variabile dipendente, hanno la funzione di chiarire la natura della relazione tra la variabile indipendente e la variabile dipendente di un modello.

Il metodo di analisi con variabili mediatrici assume l'ignorabilità dell'effetto misurazione (detta anche assunzione di *esaustività*) e l'assenza di effetti di selezione sulle variabili (detta anche assunzione di *isolamento*). Le variabili mediatrici non sono di facile identificazione, esse dovrebbero cercare di misurare, ad esempio, l'onere della risposta, la soddisfazione, l'acquiescenza o la desiderabilità sociale delle unità che partecipano all'indagine (Vannieuwenhuyze e Lynn, 2014).

I due metodi, sopra descritti, sono sottoposti a limiti dovuti alle assunzioni sottostanti i modelli. Tuttavia, nei modelli di analisi è possibile combinare variabili ausiliarie di diversa tipologia per superare alcune problematiche. Le variabili mediatrici possono essere utilizzate, ad esempio, per garantire l'assunzione di non influenzabilità dalla tecnica delle variabili confondenti, mentre le variabili confondenti possono essere utilizzate per superare le assunzioni del *front-door model*, ad esempio, costruendo prima gruppi omogenei attraverso modelli di analisi con variabili confondenti. Inoltre, le variabili confondenti possono essere utilizzate per garantire la rappresentatività dei campioni di rispondenti con tecniche diverse, mentre le variabili mediatrici possono essere utilizzate per garantire l'equivalenza della misura nei dati, se utilizzate all'interno del modello di analisi con variabili strumentali (Vannieuwenhuyze e Lynn, 2014).

Quest'ultimo tipo di modello è applicabile in un'ottica comparativa tra un campione di rispondenti con una singola tecnica e il campione di rispondenti con tecnica mista.² Il modello con variabili strumentali utilizza una variabile binaria che suddivide il campione dei rispondenti all'indagine in due gruppi a seconda della tecnica, singola o mista, con cui hanno risposto. L'uso di una variabile strumentale richiede che siano soddisfatte alcune assunzioni: l'equivalenza della misurazione e la rappresentatività dei campioni.

I modelli di analisi sopra descritti³ sono definiti modelli di aggiustamento basati su covariate (Vannieuwenhuyze et al., 2014).

2 In Vannieuwenhuyze e Loosveldt (2012) sono descritti i metodi *extended mixed-mode calibration* ed *extended MM comparison* per trattare una situazione analoga.

3 In Vannieuwenhuyze e Lynn (2014) è riportata una ricca bibliografia dei modelli descritti (Angrist *et al.*, 1996; Bowden e Turkington, 1990; Heckman, 1996, 1997; Pearl, 1995, 2009; Rubin, 1974, 1978; Morgan & Winship, 2009).

5.3.2 Stima degli effetti selezione e misurazione

Partendo da una semplificazione dell'interessante e utile formalizzazione dell'approccio inferenziale causale in un'ottica controfattuale, illustrato in Vannieuwenhuyze, Loosveldt e Molenberghs (2014), si riportano di seguito le espressioni dell'effetto selezione e dell'effetto misurazione che è possibile stimare tramite i modelli *back-door* e *front-door*.

Secondo gli autori, una stima non distorta di un parametro di interesse può essere ottenuta attraverso la costruzione della piena informazione, che per ogni unità del campione di rispondenti prevede due potenziali risultati a seconda della tecnica D ($D=m_1, m_2$): un valore osservato e un valore controfattuale. Assumendo, per semplicità, che una delle due tecniche utilizzate, m_1 , non sia affetta da errore di misura, il valor medio di una generica variabile y può essere espresso come

$$\bar{y}_{m_1} = \bar{y}_{m_1 m_1} \tau_{m_1} + \bar{y}_{m_1 m_2} \tau_{m_2}, \quad (5.6)$$

in cui \bar{y}_{dg} è il valor medio condizionato $E(y|D=d, G=g)$, essendo G la tecnica effettivamente utilizzata e D quella potenziale; τ_g rappresenta la probabilità non condizionata della tecnica effettiva, $P(G=g)$. La stima della media condizionata $\bar{y}_{m_1 m_1}$ è direttamente ottenibile dai dati osservati, mentre la stima di $\bar{y}_{m_1 m_2}$ richiede dati controfattuali. Tale formalizzazione consente di ottenere le seguenti espressioni, rispettivamente per l'effetto selezione condizionato e per l'effetto misurazione condizionato:

$$SE_{m_1}(\bar{y}) = \bar{y}_{m_1 m_1} - \bar{y}_{m_1 m_2}, \quad (5.7)$$

$$ME_{m_1}(\bar{y}) = \bar{y}_{m_2 m_2} - \bar{y}_{m_1 m_2}. \quad (5.8)$$

Da queste espressioni è immediato desumere che se l'effetto selezione fosse nullo la media di popolazione potrebbe essere stimata mediante una rilevazione mono-tecnica m_1 , mentre solo se l'effetto misurazione fosse nullo si potrebbero ottenere stime non distorte utilizzando la tecnica mista. Anche per la stima delle espressioni (5.7) e (5.8) è necessario disporre di dati controfattuali che sono ottenibili solamente sulla base delle assunzioni sopra descritte.

In questo quadro concettuale, Vannieuwenhuyze *et al.* (2014) introducono gli insiemi di variabili ausiliarie confondenti, B , e mediatrici, F , per la formalizzazione del *back-door model* e del *front-door model* sulla base delle assunzioni menzionate nel paragrafo precedente. Per il primo modello le assunzioni di ignorabilità della selezione della tecnica e di non influenzabilità dalla tecnica delle variabili consentono di ottenere la media controfattuale $\mu_{m_1 m_2}$ mediante i dati osservati. Assumendo per semplicità che le variabili B siano sintetizzate in una variabile discreta ($b=1, 2, \dots$), si ottiene

$$\bar{y}_{m_1 m_2} = \sum_b (\bar{y}_{m_1 m_2 b} \pi_{b|m_1 m_2}) = \sum_b (\bar{y}_{m_1 m_2 b} \pi_{b|m_2 m_2}), \quad (5.9)$$

5. Metodi per la valutazione e il trattamento dell'effetto tecnica in fase di stima

in cui \bar{y}_{dgb} è il valor medio condizionato $E(y|D=d, G=g, B=b)$ e $\pi_{b|dg}$ è la probabilità condizionata $P(B=b|D=d, G=g)$. Il secondo passaggio della (5.9) discende dalle due ipotesi alla base del modello: $\bar{y}_{m_1 m_2 b} = \bar{y}_{m_1 m_1 b}$ per l'ipotesi d'ignorabilità che implica $y \perp G | (D, B)$; $\pi_{b|m_1 m_2} = \pi_{b|m_1 m_1}$ per l'ipotesi di non influenzabilità dalla tecnica che implica $B \perp D | G$. Introducendo la (5.9) nelle (5.6)-(5.8) si ottengono le stime della media di popolazione, dell'effetto selezione e dell'effetto misurazione:

$$\begin{aligned}\bar{y}_{m_1} &= \sum_b \bar{y}_{m_1 m_1 b} (\pi_{b|m_1 m_1} \tau_{m_1} + \pi_{b|m_2 m_2} \tau_{m_2}) \\ SE_{m_1}(\bar{y}) &= \sum_b \bar{y}_{m_1 m_1 b} (\pi_{b|m_1 m_1} - \pi_{b|m_2 m_2}) \\ ME_{m_1}(\bar{y}) &= \sum_b \pi_{b|m_2 m_2} (\bar{y}_{m_2 m_2 b} - \bar{y}_{m_1 m_1 b}).\end{aligned}\quad (5.10)$$

La suddivisione del campione di rispondenti nelle modalità della variabile discreta B può essere effettuata, ad esempio, mediante il *propensity score matching* basato sul modello logistico, come sarà illustrato nel successivo paragrafo.

Analogamente, gli autori illustrano come sulla base del *front door model*, assumendo le ipotesi di *esaustività* e *isolamento* delle variabili mediatrici F , che spiegano le differenze di risposta alle due tecniche, sintetizzate per semplicità nella variabile discreta f , si ottengono le stime della media di popolazione, dell'effetto selezione e dell'effetto misurazione:

$$\begin{aligned}\bar{y}_{m_1} &= \sum_f \pi_{f|m_1 m_1} (\bar{y}_{m_1 m_1 f} \tau_{m_1} + \bar{y}_{m_2 m_2 f} \tau_{m_2}) \\ SE_{m_1}(\bar{y}) &= \sum_f \pi_{f|m_1 m_1} (\bar{y}_{m_1 m_1 f} - \bar{y}_{m_2 m_2 f}) \\ ME_{m_1}(\bar{y}) &= \sum_f \bar{y}_{m_2 m_2 f} (\pi_{f|m_2 m_2} - \pi_{f|m_1 m_1}).\end{aligned}\quad (5.11)$$

In conclusione, il *back-door model* e il *front-door model* utilizzano variabili di aggiustamento per separare l'effetto selezione e l'effetto misurazione e ottenerne una stima. Inoltre, sfruttando le ipotesi su cui sono basati, pur assumendo un'ottica di tipo controfattuale, superano la necessità di disporre o stimare dati controfattuali.

5.4 Metodi per il trattamento degli effetti selezione e misurazione

5.4.1 Premessa

In fase di stima, la scelta dei metodi da adottare per correggere gli effetti distorsivi della multi-tecnica sulle stime dell'indagine, determinati dal combinarsi degli effetti selezione e misurazione, è un'operazione alquanto complessa.

La confusione dei due effetti, infatti, rende complicato effettuare aggiustamenti degli effetti misurazione se prima non si stima in che misura le differenze tra i campioni di ri-

spondenti con tecniche differenti sono il risultato dell'effetto selezione. Ciò comporta la necessità di modellizzare il processo di selezione e di stimare l'errore di misura tenendo sotto controllo l'effetto selezione.

Il problema della confusione dei due effetti è affrontato, come già anticipato, con metodi derivanti dalla letteratura sull'inferenza causale e sulla mancata risposta totale. L'obiettivo comune di tali metodi è di rendere comparabili le unità del campione che hanno risposto con modalità differenti; essi assumono che le differenze che rimangono tra gruppi comparabili sono causate soltanto da errori di misura. I rispondenti all'indagine possono essere resi comparabili operando un matching tra le unità rispondenti con modalità differenti, ponderando i campioni di rispondenti o includendo variabili ausiliarie capaci di catturare l'effetto selezione nel modello di analisi dell'errore di misura (Vannieuwenhuyze e Loosveldt, 2012).

Su questa logica di comparabilità delle unità dei campioni di rispondenti con tecniche diverse sono basati i metodi di aggiustamento dell'effetto selezione, definiti come *weighting methods*. Tra questi metodi sono annoverabili i metodi noti come *post-stratificazione*, *calibrazione*, *propensity score*. Tali metodi assumono l'ignorabilità dell'effetto selezione e hanno l'obiettivo di bilanciare, rispetto ad un set di variabili ausiliarie esplicative dell'effetto selezione e *mode-insensitive* (variabili confondenti), le distribuzioni dei campioni di rispondenti. La principale funzione dei suddetti metodi di aggiustamento è di neutralizzare l'effetto selezione. Essi assumono che l'errore di misura determinato dal mix di tecniche rimane costante nel tempo e correggono la distorsione determinata dalla dipendenza della selezione dei campioni di rispondenti dalla tecnica.

Nelle indagini ripetute nel tempo di tipo *mixed-mode* sequenziale, l'assunzione di invarianza dell'errore di misura è, tuttavia, poco sostenibile. La composizione dei rispondenti per tecnica può, infatti, cambiare nelle successive edizioni dell'indagine, comportando variazioni nell'errore totale di misura.

Buelens e Van den Brakel (2011) propongono un interessante metodo che mira ad ottenere stime non distorte mantenendo costante l'errore di misura nelle diverse occasioni d'indagine. Gli autori puntano a definire una procedura di stima robusta rispetto a variazioni nelle distribuzioni di rispondenti a tecniche diverse. Per evitare la confusione di cambiamenti veri nel tempo delle stime con variazioni nella composizione dei campioni di rispondenti, gli autori propongono uno stimatore di regressione generalizzata (GREG) calibrato per tecnica. In pratica, attraverso il meccanismo dello stimatore GREG, lo stimatore proposto realizza simultaneamente una calibrazione rispetto a variabili ausiliarie che correggono l'effetto selezione e una calibrazione dei campioni di rispondenti a livelli fissati per stabilizzare l'errore di misura.

I metodi inferenziali sopra descritti - *post-stratificazione*, *calibrazione*, *propensity score* - sono utilizzati, di fatto, anche per stimare l'effetto misurazione determinato dal mix di tecniche. Altri metodi sono utilizzati a tal scopo, come i metodi basati su modelli di regressione generalizzati (Kolenikov e Kennedy, 2014) e i metodi basati su modelli di imputazione multipla (Rubin, 1976; Rubin, 1987).

Tali metodi assumono anch'essi l'ignorabilità dell'effetto selezione ma a differenza dei precedenti metodi consentono di correggere le stime di specifiche variabili dell'indagine da effetti misurazione. Generalmente, tali metodi prendono una tecnica come riferimento (benchmark), ad esempio la tecnica ritenuta più affidabile (*gold-standard-mode*) o quella storicamente utilizzata dall'indagine.

Nell'approccio metodologico basato su modelli di regressione generalizzati gli aggiustamenti di effetti misurazione sono ottenuti mettendo in relazione la variabile d'indagine con la tecnica e un insieme di variabili ausiliarie. Si tratta di un metodo che può essere utilizzato solo per produrre stime aggregate.

Nell'approccio metodologico basato su modelli di imputazione, essendo l'errore di misura concettualizzato come un problema di dati mancanti, l'aggiustamento di effetti misurazione è effettuato convertendo le risposte fornite con una specifica tecnica, assunte come mancanti, in risposte controfattuali. Tale impostazione conduce alla costruzione di un data set completo di rispondenti con la stessa tecnica, ad esempio quella assunta come riferimento, composto da risposte effettivamente osservate e da risposte potenziali, o controfattuali, stimate per l'altra tecnica affetta da errori di misura. L'obiettivo ulteriore perseguito attraverso l'uso di imputazioni multiple è di sostituire i dati missing con un insieme di valori plausibili (dati controfattuali) in modo da riflettere l'incertezza nella predizione dei dati mancanti e stimare la varianza degli stimatori. I metodi di imputazione multipla generano un risultato sintetico che ha proprietà simili a quelle dei dati assunti come riferimento per la realizzazione del campione completo.

I metodi di aggiustamento di effetti misurazione, generalmente, consentono la correzione delle stime di singole variabili di indagine e non sono quindi facilmente applicabili a indagini il cui obiettivo sia produrre una notevole molteplicità di stime.

Nei successivi paragrafi si espongono alcuni dei metodi introdotti, in particolare: il metodo di aggiustamento dell'effetto selezione basato sul *propensity score*; il metodo di correzione dell'effetto misurazione basato su un modello di regressione (per variabili continue); il metodo di correzione dell'effetto misurazione basato su modelli di imputazione multipla. Relativamente a quest'ultimo approccio metodologico si fa un accenno ai metodi di imputazione esistenti per il trattamento delle mancate risposte parziali, che possono essere adattati al contesto di studio e si riportano alcuni metodi di imputazione multipla applicati alle indagini *mixed-mode*. Quest'ultimi prendono in considerazione la situazione in cui il meccanismo di selezione della tecnica non è ignorabile, ovvero del fatto che sussiste una relazione di dipendenza tra la distribuzione della variabile d'indagine (e le caratteristiche degli individui) e la scelta della tecnica da parte dei rispondenti. Questa può essere la situazione che si presenta quando si rilevano quesiti sensibili come, ad esempio, l'identità sessuale di un individuo, la violenza subita dalle donne, il reddito, etc. I metodi che di seguito saranno brevemente descritti sono: il metodo di imputazione multipla con pesi frazionati (*Fractional multiple imputation*, Fmi) proposto da Park *et al.* (2016); il metodo di imputazione multipla con modello di risposta (e modello di selezione) proposto da Suzer-Gurtekin *et al.* (2012).⁴

5.4.2 Metodi basati sul *propensity score*

L'approccio del *propensity score* (Rosenbaum e Rubin, 1983; Rubin, 2006) è adottato, in generale, negli studi osservazionali al fine di costruire dei gruppi omogenei di unità nei quali vi sia un equilibrio rispetto alle variabili ausiliarie. Possono essere utilizzati diversi metodi di aggiustamento basati sul *propensity score*: matching, stratificazione, aggiustamenti basati su covariate (D'Agostino, 1998; Austin, 2011; Lee e Valliant, 2008). Un ambito

⁴ In Kolenikov e Kennedy (2014) è descritto un metodo di imputazione multipla basato su un modello econometrico di utilità implicita che assume l'ipotesi NMAR, ovvero che la causa generatrice dei dati missing non è completamente identificata dai dati osservati.

classico di applicazione di tale approccio è il trattamento della mancata risposta totale (De Vitiis *et al.*, 2002; De Vitiis *et al.*, 2012).

Nel contesto delle indagini a tecnica mista, l'approccio del *propensity score* consente di affrontare la questione dell'effetto selezione con l'obiettivo di isolare gli effetti confondenti del meccanismo di selezione. Sotto l'ipotesi d'ignorabilità dell'effetto selezione della tecnica, il *propensity score* può essere interpretato come la probabilità di assegnazione (o scelta) di una tecnica condizionatamente alle covariate osservate. La propensione a rispondere a una delle tecniche viene modellizzata mediante un modello logistico e il *propensity score* viene ottenuto come probabilità predetta. Obiettivo è la costruzione di gruppi omogenei di rispondenti rispetto alla propensione alla tecnica, all'interno dei quali non vi siano differenze di composizione dei campioni tra le due tecniche, in modo da poter valutare la presenza di errori di misura senza il "disturbo" dell'effetto selezione.

Il metodo del *propensity score* può essere inquadrato nell'ambito dell'inferenza causale in un'ottica controfattuale come un'applicazione del *back-door model*; condizione essenziale è che si disponga di un insieme di variabili ausiliarie *confondenti* che siano *mode insensitive*. Alcuni autori (Vandenplas *et al.*, 2016) propongono una correzione per le stime dell'indagine basata sul *propensity score* attraverso la costruzione di pesi che correggono l'effetto selezione aggiustando le distribuzioni dei campioni di rispondenti alle diverse tecniche.

Entrando nello specifico del metodo per il caso di un'indagine a doppia tecnica di tipo sequenziale, seguendo Vandenplas *et al.* (2016), una delle due tecniche, m_1 , viene considerata come riferimento mentre le risposte fornite con la seconda, m_2 , devono essere rese equivalenti a quelle che sarebbero state fornite mediante la prima (dati controfattuali). Questa equivalenza viene ottenuta attraverso il *propensity score weighting* che realizza il *matching* tra i rispondenti alla tecnica m_1 e i rispondenti alla tecnica m_2 .

Il propensity score di ciascun rispondente è definito come la probabilità di partecipare alla tecnica m_1 stimata mediante un modello binomiale logistico su un insieme di variabili ausiliarie: $\text{logit}[p(x)] = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$.

I rispondenti sono quindi suddivisi in classi sulla base delle probabilità predette (quintili o decili della distribuzione); condizione essenziale per l'applicabilità del metodo è che tali classi risultino bilanciate, ossia in ciascuna di esse deve essere verificata l'indipendenza tra ciascuna variabile ausiliaria e la tecnica di rilevazione, tramite un opportuno test statistico (ad esempio il test del chi-quadrato per variabili ausiliarie categoriali).

Per ciascun gruppo bilanciato k , si può quindi calcolare il peso che rende equivalenti la proporzione pesata dei rispondenti alla tecnica m_1 alla proporzione dei rispondenti alla tecnica m_2 nella stessa classe

$$w_k = \frac{n_{k,m_2}/n_{m_2}}{n_{k,m_1}/n_{m_1}} \quad (5.12)$$

Utilizzando tali correttori, nelle classi bilanciate è possibile ottenere una valutazione complessiva dell'effetto tecnica nelle sue due componenti di effetto selezione e effetto misurazione (Vandenplas *et al.*, 2016): l'effetto selezione, come differenza tra le stime pesate e non pesate dei rispondenti alla tecnica m_1 .

$$SE_{m_1}(\bar{y}) = \frac{\sum_{i=1}^{n_{m_1}} y_{i,m_1}}{n_{m_1}} - \frac{\sum_{i=1}^{n_{m_1}} w_{k,i} y_{i,m_1}}{n_{m_1}}; \quad (5.13)$$

5. Metodi per la valutazione e il trattamento dell'effetto tecnica in fase di stima

e l'effetto misurazione, come differenza tra la stima pesata dei rispondenti alla tecnica m_1 e la stima non pesata dei rispondenti a m_2

$$ME_{m_1}(\bar{y}) = \frac{\sum_{i=1}^{n_{m_1}} w_{k,i} y_{i,m_1}}{n_{m_1}} - \frac{\sum_{i=1}^{n_{m_2}} y_{i,m_2}}{n_{m_2}}. \quad (5.14)$$

Oltre a queste misure complessive, all'interno dei gruppi bilanciati è possibile ottenere una valutazione dell'errore di misura relativo a una generica variabile y attraverso un test di indipendenza tra la tecnica e y . Per ciascuna modalità di una variabile categoriale, ad esempio, possono essere definite le differenze tra le frequenze osservate per ciascuna tecnica e le frequenze attese in caso di indipendenza tra la tecnica e la variabile e si può definire un test del χ^2 per verificare la significatività di tali differenze

$$d_{k,m_1} = (F^*_{k,m_1} - F_{k,m_1})/n_{k,m_1}, \quad d_{k,m_2} = (F^*_{k,m_2} - F_{k,m_2})/n_{k,m_2},$$

in cui $F^*_{k,m}$ e $F_{k,m}$ rappresentano rispettivamente le frequenze assolute attese ed effettive e $n_{k,m}$ il numero di rispondenti alla generica tecnica m nella classe k .

Il peso (5.12) può essere utilizzato per correggere le stime dell'indagine per l'effetto selezione, anche se in letteratura non si trovano molti esempi di tale trattamento delle stime e alcuni autori esprimono delle riserve a riguardo (Kolenikov e Kennedy, 2014).

Come già evidenziato, i limiti di questo approccio sono principalmente legati all'assunzione d'ignorabilità della selezione, al rischio di trovare gruppi non bilanciati e la violazione dell'assunzione di sovrapposizione (Vandenplas *et al.*, 2016).

E' utile sottolineare come la validità del processo inferenziale, sopra descritto, può, tuttavia, essere rafforzata da azioni esogene alla fase di stima, ma incorporate nel disegno complessivo dell'indagine multi-tecnica. Si tratta di azioni che devono essere volte ad acquisire, con la stessa indagine, informazioni utili a modellizzare la propensione a rispondere con tecniche differenti (*paradata, mode-preference*) e superare i limiti posti dall'assunzione d'ignorabilità della selezione che è alla base dei *back-door models*, oppure di azioni di prevenzione dell'errore di misura che puntano ad azzerarne la presenza o comunque a ridurlo in modo consistente l'entità (cfr. Capitolo 3).

5.4.3 Metodi basati su modelli di regressione

Il metodo basato sui modelli di regressione generalizzati è descritto nel lavoro di Kolenikov e Kennedy (2014).

Nel modello di regressione, assume particolare rilevanza la variabile *tecnica di risposta*, che identifica ciascun rispondente in base alla tecnica di risposta utilizzata. Il metodo mira all'adattamento di un modello in cui la variabile *tecnica di risposta* diventa, con altre variabili ausiliarie, un predittore del modello e in cui le risposte fornite con tecniche diverse sono armonizzate sottraendo i coefficienti stimati di una determinata modalità dalla risposta aggregata. Le variabili ausiliarie rilevanti del modello sono quelle che hanno una relazione sia con la tecnica che con la variabile di interesse dell'indagine. L'introduzione nel modello di un effetto interazione tra tecnica e variabili ausiliarie (ad esempio variabili socio-demografiche) può spiegare come la risposta con una tecnica è influenzata dalle caratteristiche osservabili dei rispondenti (Kolenikov e Kennedy, 2014).

Assumendo che la generica variabile di interesse y sia continua, il modello di regressione può essere formalizzato come:

$$y_i^m = \boldsymbol{\beta}' \mathbf{x}_i + \varepsilon_i^m, \quad (5.15),$$

dove:

y_i^m è il valore della variabile osservata sulla i -esima unità del campione di rispondenti con la tecnica m ; m ha valore 0 per la tecnica assunta come riferimento e valore 1 per l'altra tecnica,

$\boldsymbol{\beta}$ è il vettore dei coefficienti di regressione,

\mathbf{x}_i è il vettore delle variabili ausiliarie.

Supponendo che il residuo della regressione, ε , contiene errori non specifici della tecnica, si può scrivere $\varepsilon_i^1 = \varepsilon_i^0 + \gamma m$, dove γ è l'effetto tecnica. Il modello (5.15) può essere espresso nella forma:

$$y_i^m = \boldsymbol{\beta}' \mathbf{x}_i + \gamma m_i + \varepsilon_i. \quad (5.16)$$

L'aggiustamento dell'errore di misura è ottenuto dal valore predetto del modello (5.16) con l'esclusione del contributo della tecnica $\hat{\gamma}m$:

$$\tilde{y}_i^0 = E[y_i^0 | \mathbf{x}_i, y_i^m] = \mathbf{x}_i' \hat{\boldsymbol{\beta}} + \hat{\varepsilon}_i^m = y_i^m - \hat{\gamma}m_i,$$

dove $\hat{\varepsilon}$ è il residuo della regressione o l'errore predetto.

L'estensione di un aggiustamento dell'errore di misura a variabili categoriali tramite modelli di regressione logistica è possibile soltanto per stime di proporzioni (Kolenikov e Kennedy, 2014).

5.4.4 Metodi standard di imputazione multipla

Tra i metodi esistenti per il trattamento di dati mancanti (o *missing*), e sviluppati in pacchetti SAS ed R, una tecnica di imputazione multipla molto diffusa è basata sull'utilizzo di equazioni concatenate (White *et al.*, 2011; van Buuren, 2012) in cui ogni distribuzione marginale è stimata tramite modelli lineari generalizzati o metodi basati su donatore. Il metodo presenta delle limitazioni quando applicato ad indagini complesse, esso inoltre invoca l'assunzione MAR (*missing at random assumption*, cfr. Rubin e Little, 2002), ovvero che la causa generatrice dei dati mancanti è identificata dai dati osservati.

5.4.5 Metodo di imputazione con pesi frazionati

L'imputazione multipla frazionata è un metodo introdotto da Kalton e Kish (1984) che consiste nel sostituire ogni valore mancante con un insieme di valori plausibili a probabilità differenziate ottenute tramite pesi frazionati.

L'applicazione di un'imputazione multipla frazionata al contesto delle indagini *mixed-mode* permette di trattare il problema dell'errore di misura quando la scelta della tecnica è autoselezionata (Park *et al.*, 2016). Inoltre, l'introduzione di pesi frazionati comporta una

riduzione del carico computazionale iterativo per valutare il valore atteso condizionato associato con i dati mancanti.

Il metodo proposto dagli autori è applicato ad un contesto in cui le tecniche di rilevazione utilizzate sono soltanto due e i campioni di rispondenti realizzati sono dipendenti (disegno di indagine *mixed-mode* sequenziale).

Supponendo che il campione di rispondenti con la tecnica m_2 sia assunto come benchmark (tecnica utilizzata nella fase di follow-up per i non rispondenti con la tecnica m_1 sottoposta a tutto il campione nella prima fase di rilevazione) e il campione di rispondenti con la tecnica m_1 sia affetto da errori di misura, per quest'ultimo campione si definisce, relativamente ad una generica variabile dell'indagine, y , il modello dell'errore di misura:

La variabile y_{m_2} è la variabile controfattuale non osservata di y_{m_1} nel campione di rispondenti con la tecnica m_1 .

$$y_{m_i} = \alpha_0 + \alpha_1 y_{m_2i} + u_i, \quad (5.17)$$

dove:

y_{m_i} è il valore della variabile osservata sulla i -esima unità del campione di rispondenti con la tecnica m_1 ,

y_{m_2i} è il valore della variabile osservata sulla i -esima unità del campione di rispondenti con la tecnica m_2 ,

$$u_i \approx (0, \sigma_u^2).$$

Ponendo $y_{m_2i} = y_i$ e introducendo il vettore di variabili ausiliarie \mathbf{x}_i , si definisce il modello strutturale:

$$y_i | \mathbf{x}_i \approx f(y_i | \mathbf{x}_i; \theta), \quad (5.18)$$

dove θ è il parametro ignoto che caratterizza la distribuzione condizionata di y_{m_1} su y_{m_2} .

La variabile y_{m_2} è la variabile controfattuale non osservata di y_{m_1} nel campione di rispondenti con la tecnica m_1 .

I modelli (5.17) e (5.18) formano un set di modelli dell'errore di misura: il primo, chiamato modello di misura, descrive la relazione tra i campioni di rispondenti con le due tecniche; il secondo, chiamato modello di errore strutturale, incorpora informazione aggiuntiva da x_i .

Il modello predittivo, o di imputazione, è ottenuto tramite la formula di Bayes

$$f(y_i | y_{m_i}, \mathbf{x}_i) = \frac{f(y_i | \mathbf{x}_i; \theta) g(y_{m_i} | y_i; \alpha)}{\int f(y_i | \mathbf{x}_i; \theta) g(y_{m_i} | y_i; \alpha) dy_i}, \quad (5.19)$$

dove $g(y_{m_i} | y_i; \alpha)$ è il modello dell'errore di misura (5.17) relativo al campione di rispondenti alla tecnica m_1 .

Il modello di imputazione ha lo scopo di generare $y_{m_i} = y_i$ per $i \in m_1$; per il campione di rispondenti con la tecnica m_1 è possibile, pertanto, stimare il valore atteso condizionato di y_i tramite il modello (5.19).

Nel modello di imputazione, oltre al modello dell'errore di misura, può essere inserito il modello di selezione della tecnica $P(m_i = m_1 | \mathbf{x}_i, y_i)$, utile soprattutto quando la scelta della tecnica da parte dei rispondenti non è casuale (assunzione di non ignorabilità dell'effetto selezione). Questo comporta una modifica della distribuzione condizionata

$$f(y_i | y_{m_i}, \mathbf{x}_i, m_i = m_2) = \frac{f(y_i | \mathbf{x}_i; \theta) g(y_{m_i} | y_i; \alpha) P(m_i = m_1 | \mathbf{x}_i, y_i)}{\int f(y_i | \mathbf{x}_i; \theta) g(y_{m_i} | y_i; \alpha) P(m_i = m_1 | \mathbf{x}_i, y_i) dy_i}. \quad (5.20)$$

Se il modello di selezione soddisfa la seguente uguaglianza,

$$P(m = m_1 | \mathbf{x}, y) = P(m = m_1 | \mathbf{x}),$$

allora l'effetto selezione della scelta della tecnica è ignorabile, nel senso che dipende soltanto da caratteristiche dell'individuo (Rubin, 1976), in questo caso il modello di imputazione è riconducibile al modello (5.19).

Per la variabile d'indagine y , lo stimatore corretto della media della popolazione è esprimibile in funzione di due componenti: la prima calcolata sulla variabile osservata nel campione di rispondenti con la tecnica m_2 ; la seconda calcolata sul valore atteso condizionato predetto dal modello nel campione di rispondenti con la tecnica m_1 . La stima ottenuta è, dunque, relativa ad una specifica tecnica, quella associata al campione di rispondenti assunto come benchmark:

$$\hat{Y} = \frac{1}{N} \left\{ \sum_{i \in m_2} w_i y_i + \sum_{i \in m_1} w_i E(y_i | y_{m_i}, \mathbf{x}_i) \right\}.$$

5.4.6 Metodo di imputazione con modelli di risposta e selezione

L'interessante metodo proposto da Suzer-Gurtekin et al. (2012) utilizza un modello di risposta, di regressione lineare o logistico a seconda del tipo di variabile di interesse, per imputare valori controfattuali per ogni campione di rispondenti caratterizzato dalla tecnica di risposta utilizzata (m_1 e m_2). Il modello di imputazione è in questo caso basato sull'assunzione d'ignorabilità dell'effetto selezione.

L'assunzione di non ignorabilità dell'effetto selezione implica che nel modello di imputazione siano incluse le stime di massima verosimiglianza di parametri specifici di due modelli: il modello per la variabile risposta dell'indagine (equazione di *outcome*) e il modello per il meccanismo generatore dei dati mancanti (equazione di selezione). Le due equazioni definiscono, di fatto, un modello di selezione del campione (*sample selection model*) i cui parametri sono stimati simultaneamente con il metodo di stima basato sulla massima verosimiglianza completa (Cobben, 2009; Heckman, 1979). Da tale impostazione discende che il modello di imputazione stima, per ogni unità, il valore atteso di una variabile di interesse condizionato ad un set di variabili ausiliarie e al processo di selezione.

Il metodo di imputazione proposto dagli autori, a differenza di altri metodi che assumono una tecnica come benchmark, porta alla costruzione di data set completi di rispondenti per ogni specifica tecnica. In pratica, per i rispondenti con la tecnica m_1 si imputano valori

come se avessero risposto con la tecnica m_2 ; la combinazione dei valori osservati con la tecnica m_2 e dei valori ottenuti dall'imputazione porta alla costruzione di un campione completo di rispondenti con la sola tecnica m_2 . Procedendo allo stesso modo si arriva a costruire un campione completo di rispondenti con la sola tecnica m_1 .

L'obiettivo è di calcolare, per ogni campione completo di rispondenti, stime specifiche per tecnica (m_1 e m_2). Tali stime sono poi combinate insieme per ottenere un'unica stima. Gli stimatori che gli autori propongono sono ottenuti come: (i) semplice media aritmetica delle stime specifiche per tecnica (m_1 e m_2); (ii) combinazione lineare delle stime specifiche delle tecniche m_1 e m_2 , pesate con l'inverso della varianza delle rispettive stime; (iii) combinazione lineare delle stime specifiche delle tecniche m_1 e m_2 , pesate con l'inverso dell'errore quadratico medio delle rispettive stime.

Questo approccio è stato applicato dagli autori ad un disegno di indagine *mixed-mode* sequenziale in cui sono utilizzate due tecniche di rilevazione. Il metodo ha il vantaggio di poter essere esteso a indagini che utilizzano più di due tecniche di rilevazione e di tener conto anche del problema della mancata risposta totale.

5.5 Il trattamento della mancata risposta totale nelle indagini multi-tecnica

5.5.1 Premessa

Per le indagini a tecnica mista il trattamento della mancata risposta totale è una questione resa complessa dal processo di risposta che genera un ulteriore effetto di selezione a causa della mancata risposta totale. Per valutare l'errore totale non campionario, gli effetti introdotti dalla tecnica mista devono essere trattati insieme all'effetto di selezione sulla mancata risposta totale. I modelli per il trattamento della tecnica mista e il modello di risposta sono gli elementi fondamentali per combinare dati provenienti da tale indagini con disegni diversi.

Nei paragrafi che seguono si introduce un importante strumento di analisi della risposta basato su indicatori di rappresentatività che può essere esteso anche alle indagini *mixed-mode* e si presentano alcune soluzioni al problema del trattamento della mancata risposta.

Gli indicatori di rappresentatività, a differenze di alcune analisi presentate nel Capitolo 4, sono costruiti a partire dalla propensione alla risposta stimata con modelli multivariati di tipo binomiale. Essi, pur non consentendo di effettuare aggiustamenti della mancata risposta totale, sono utilizzabili per valutazioni della distorsione che la mancata risposta totale può introdurre nelle stime dei parametri di interesse di un'indagine.

5.5.2 Indicatori di rappresentatività

Generalmente, il tasso di risposta è assunto come indicatore di qualità dell'indagine. Poiché tale indicatore non fornisce informazioni relative alla distorsione che la mancata risposta totale può introdurre nelle stime dei parametri di interesse di un'indagine, è possibile utilizzare altre tipologie di indicatori. Si tratta di indicatori di rappresentatività, noti come *R-indicators* e *partial R-indicators*, che sono stati introdotti da Bethlehem *et al.* (2009), e Schouten *et al.* (2011).

Gli indicatori di rappresentatività sono basati su una misura di variabilità della propensione alla risposta e descrivono come il campione dei rispondenti ad un'indagine riflette la popolazione di interesse rispetto ad un certo numero di variabili caratterizzanti la popolazione stessa. In sostanza, essi misurano di quanto il campione di rispondenti di un'indagine si discosta dalla risposta rappresentativa.

Il concetto di risposta rappresentativa è legato al meccanismo che genera la mancata risposta totale, o i dati missing: se tale meccanismo è di tipo MCAR (*Missing Completely at Random*) allora, per una specifica variabile dell'indagine, i rispondenti e i non rispondenti in media sono gli stessi; se è di tipo MAR (*Missing At Random*) allora vuol dire che all'interno di sottopopolazioni i rispondenti e i non rispondenti in media sono gli stessi; diversa è la situazione in cui il meccanismo è di tipo NMAR (*Not Missing At Random*) dove la probabilità di risposta dipende dai dati missing.

La risposta rappresentativa è associabile, quindi, ad un campione casuale di rispondenti all'indagine; la risposta rappresentativa è condizionata quando è associabile ad un campione casuale stratificato. Tali indicatori possono essere calcolati a differenti livelli di dettaglio ma non sono utilizzabili per effettuare aggiustamenti della mancata risposta totale.

La risposta è detta rappresentativa rispetto a un insieme di variabili ausiliarie quando la propensione alla risposta per le sottopopolazioni identificate dalle variabili stesse è costante. La risposta è detta condizionatamente rappresentativa per una certa variabile ausiliaria, sottratta all'insieme di variabili ausiliarie considerate nel modello, quando la propensione alla risposta per la sottopopolazione formata dalla variabile presa in considerazione è costante nella sottopopolazione formata dalle altre variabili ausiliarie.

L'indicatore *R-indicator* è una semplice misura basata sulla *standard deviation* della propensione alla risposta:

$$R(\rho_x) = 1 - 2S(\rho_x)$$

dove ρ_x è la propensione alla risposta stimata rispetto alle variabili ausiliarie \mathbf{x} e $S(\rho_x)$ è la standard deviation di ρ_x .

Tale misura varia nell'intervallo dei valori (0,1); la rappresentatività della risposta si ha quando la misura assume valore 1.

L'indicatore *partial R-indicator* scompone la variabilità della propensione alla risposta rispetto a differenti valori delle variabili ausiliarie, la standard deviation $S(\rho_x)$ diventa.

$$S^2(\rho_x) = S_w^2(\rho_x|H) + S_b^2(\rho_x|H)$$

dove H è la stratificazione della popolazione in sottogruppi, $S_w^2(\rho_x|H)$ è la varianza della propensione alla risposta nei sottogruppi e $S_b^2(\rho_x|H)$ è la varianza della propensione alla risposta tra i sottogruppi.

L'utilizzo degli indicatori di rappresentatività può essere esteso anche all'analisi dei campioni di rispondenti, e del (dei) campione di non rispondenti, nelle indagini *mixed-mode*. In questo caso gli indicatori sono derivati da modelli di risposta che possono essere diversamente formulati. La rappresentatività può essere valutata prendendo i rispondenti con l'una o l'altra tecnica, o i rispondenti al mix di tecniche, come riferimento rispetto ai non rispondenti, ancora i rispondenti con una tecnica rispetto ai rispondenti con l'altra tecnica oppure ai rispondenti al mix di tecniche.

5.5.3 Il trattamento della mancata risposta totale nei disegni concorrenti e sequenziali

Nei disegni di indagini multi-tecnica di tipo concorrente, il modello di risposta può essere inserito nei modelli per l'effetto tecnica più facilmente, ma per le indagini di tipo sequenziale questa operazione è molto più complessa. Inoltre, in quest'ultimo schema, la distribuzione delle unità campionarie tra le modalità successive di rilevazione dipende dai risultati del processo di risposta nelle modalità precedenti. In questo caso, la sequenza di tecniche (*mode response path*) dovrebbe essere considerata nel modello di risposta (Cobben *et al.*, 2006).

Di fatto, per le indagini con schema di tipo sequenziale non è possibile definire un metodo per il trattamento della mancata risposta che produca una correzione dei pesi di riporto all'universo come nelle indagini a tecnica unica.

I metodi proposti in letteratura, ad esempio il *sample selection model* (Cobben, 2009; Heckman, 1979; De Vitiis *et al.*, 2016), consentono di correggere contemporaneamente gli effetti selezione determinati dal mix di tecniche e dalla mancata risposta totale, relativamente alla stima di parametri di singole variabili d'indagine. Tali metodi presuppongono, però, l'ipotesi di assenza di errori di misura.

Il metodo descritto al Paragrafo 5.4.6 consente invece, di effettuare un aggiustamento della mancata risposta totale una volta costruiti, mediante imputazione multipla, i data set completi di rispondenti per ogni specifica tecnica. Tuttavia anche tale approccio è limitato dal fatto che stime meno distorte di parametri di interesse sono relative a singole variabili di indagine.

I metodi qui delineati, in conclusione, come anche i metodi illustrati nei paragrafi precedenti per il trattamento dell'effetto misurazione, pur soffrendo del limite di essere utilizzabili solo per specifiche variabili d'indagine, sono molto interessanti perché in fase di analisi dell'effetto tecnica possono essere usati per il confronto con i risultati di metodi di stima che possono essere utilizzati in modo complessivo per tutte le variabili dell'indagine, agendo, ad esempio, sui pesi di riporto all'universo.

6. ANALISI DELL'EFFETTO TECNICA IN DUE CASI DI STUDIO¹

6.1 Introduzione

Questo Capitolo ha l'obiettivo di presentare l'utilizzo di alcune delle tecniche descritte nei Capitoli 4 e 5 volte a valutare, ed eventualmente misurare, o correggere l'effetto tecnica in presenza di dati rilevati con tecnica mista.

I casi studio presentati nei paragrafi 6.2 e 6.3 sono relativi rispettivamente alle indagini "Aspetti della vita quotidiana" (Avq) e "I Cittadini e il tempo libero" (Ctl). Le due indagini sono state scelte come casi di studio in quanto hanno utilizzato le tecniche di rilevazione Cawi/Papi, la prima su campioni indipendenti, la seconda in sequenza. Per ogni indagine, il Capitolo contiene informazioni sugli obiettivi della sperimentazione, sulle caratteristiche della rilevazione e sui risultati delle analisi esplorative, nonché la descrizione delle applicazioni sperimentali delle metodologie per la valutazione e/o misurazione dell'effetto tecnica utilizzate e dei relativi risultati.

6.2 La sperimentazione sull'indagine Aspetti della vita quotidiana 2014

6.2.1 Obiettivi della sperimentazione

Il presente paragrafo ha l'obiettivo di presentare alcune analisi (analisi preliminari, test, modello logistico, analisi fattoriale multigruppo) che possono essere adottate in presenza di indagini indipendenti condotte con tecnica diversa, ma con stesso questionario, e illustrare i passaggi logici delle procedure, in modo da fornire al lettore un quadro di riferimento di analisi. Lo studio è stato condotto sui dati dell'edizione 2014 dell'indagine Avq, realizzata con tecnica Papi, e sui dati di un'indagine sperimentale Cawi condotta utilizzando gli stessi quesiti e la stessa struttura del questionario dell'indagine Avq. Il confronto ha lo scopo di rilevare un possibile effetto misurazione per alcuni *item* di variabili presenti nel questionario. Le elaborazioni e gli approfondimenti qui presentati hanno lo scopo di evidenziare la sequenza delle analisi scelte, piuttosto che i risultati ottenuti.

Di seguito si procede alla descrizione dell'indagine Avq e di alcune analisi esplorative volte a identificare differenze nei tassi di risposta con le due tecniche per particolari distribuzioni di rispondenti, e ad analizzare le differenti distribuzioni di frequenza tra rispondenti Cawi e Papi per un sottoinsieme ristretto di variabili. Successivamente si passa alla descrizione dei metodi applicati e alla presentazione di alcuni risultati.

¹ Hanno collaborato alla stesura del capitolo : Silvia Montecolle (paragrafi 6.2.1, 6.2.2, 6.2.3); Roberta Varriale (paragrafi 6.1, 6.2.6, 6.2.7); Iaria Vannini (paragrafi 6.2.4, 6.2.5); Claudia De Vitiis (paragrafi 6.3.1, 6.3.8), Francesca Inglese (paragrafi 6.3.4), Emanuela Bologna (paragrafi 6.3.2, 6.3.3), Marco Dionisio Terribili (paragrafi 6.3.5, 6.3.6), Alessio Guandalini (paragrafi 6.3.7).

6.2.2 Cenni sulla rilevazione

Avq è un'indagine sulle famiglie condotta annualmente dal 1993 con tecnica Papi. Essa fa parte del sistema integrato di Indagini Multiscopo e raccoglie informazioni su vari fenomeni sociali sia a livello familiare sia individuale. La popolazione di interesse è costituita dalle famiglie residenti in Italia e dagli individui che le compongono.

L'edizione del 2014 è stata condotta nel mese di marzo su un campione di oltre 24.000 famiglie.

L'intervista alla famiglia prevede l'utilizzo di due questionari cartacei. Il primo è il questionario base della rilevazione che è somministrato da un rilevatore tramite intervista. Questo modello è composto da: (i) una "Scheda Generale", in cui si rilevano le relazioni di parentela e altre informazioni di natura socio-demografica e socio-economica relative ai componenti della famiglia; (ii) le "Schede Individuali", una per ciascun componente della famiglia e (iii) un "Questionario familiare" che contiene quesiti familiari ai quali risponde un solo componente adulto. L'altro è un modello somministrato per auto-compilazione. Il modello è consegnato dal rilevatore a ciascun componente della famiglia e contiene quesiti che possono essere agevolmente compilati in autonomia dal rispondente anche senza l'intervento diretto del rilevatore.

Nei mesi di novembre-dicembre 2014 è stata condotta una indagine pilota con tecnica esclusivamente Cawi utilizzando il questionario dell'indagine Avq. La rilevazione era propedeutica alla realizzazione con tecnica mista Cawi-Papi dell'indagine multiscopo sulle famiglie Ctl, e aveva l'obiettivo di testare l'utilizzo della tecnica Cawi per lo svolgimento di un'indagine complessa rivolta alla famiglia e a tutti i suoi componenti. La numerosità del campione è stata di 11.898 famiglie, tenendo conto che i tassi di partecipazione attesi in questo tipo di tecnica non sono particolarmente alti, tale numerosità, piuttosto elevata per una pilota, ha permesso di disporre di un numero di interviste utili adeguato all'analisi dei risultati. A queste famiglie è stata mandata una lettera informativa con l'invito a partecipare alla rilevazione in Cawi. La tecnica Papi ha rappresentato necessariamente la guida per l'impostazione del questionario somministrato con tecnica Cawi: sia per quanto riguarda l'impianto generale e la trasposizione su web dei diversi modelli, sia con riferimento a istruzioni e accorgimenti tecnici nello svolgersi dell'intervista tra sezioni e quesiti.

6.2.3 Analisi esplorative preliminari

Alla fine del periodo di rilevazione, le famiglie intervistate in Cawi sono risultate 1.576, pari al 13,2 per cento delle famiglie appartenenti al campione teorico.

Nella Tavola 6.1 sono messe a confronto alcune caratteristiche delle famiglie intervistate con la tecnica Cawi con quelle intervistate nell'ambito dell'indagine Avq condotta con tecnica Papi nel 2014. Gli effetti del diverso comportamento sul territorio rispetto alla compilazione web si notano anche nella composizione percentuale dei rispondenti. Tra i rispondenti Cawi rispetto a quelli Papi è maggiore il peso delle famiglie che vivono nel Nord, nei comuni centro dell'area metropolitana e nel loro *hinterland*.

La distribuzione delle famiglie per caratteristiche delle persone che la compongono (Tavola 6.2) mostra come il Cawi sia approcciato con maggiore facilità dalla famiglia in cui è presente almeno un componente al di sotto dei 30 anni. Al contrario, le famiglie composte da tutti anziani, notoriamente meno avvezzi alle nuove tecnologie, sono una quota più bassa di rispondenti.

Inoltre, sono principalmente le famiglie con componenti che hanno titolo di studio me-

6. Analisi dell'effetto tecnica in due casi di Studio

91

Tavola 6.1 - Famiglie intervistate per ripartizione e tipologia comunale. Confronto tra Avq indagine Papi edizione 2014 e Avq pilota 2014. Dati in migliaia e valori percentuali

RIPARTIZIONE TIPOLOGIA COMUNALE	Avq Papi		Avq Cawi	
	V.A.	%	V.A.	%
Nord-ovest	4.241	23,1	597	37,9
Nord-est	3.534	19,3	365	23,2
Centro	3.365	18,4	310	19,7
Sud	5.244	28,6	190	12,1
Isole	1.953	10,7	114	7,2
Comune centro dell'area metropolitana	2.329	12,7	284	18,0
Periferia dell'area metropolitana	1.550	8,5	175	11,1
Fino a 2.000 abitanti	1.613	8,8	86	5,5
Da 2.001 a 10.000 abitanti	4.697	25,6	373	23,7
Da 10.001 a 50.000 abitanti	4.980	27,2	381	24,2
Oltre 50.000 abitanti	3.168	17,3	277	17,6
Totale	18.337	100,0	1.576	100,0

Fonte: Indagine Papi Aspetti della vita quotidiana 2014 e Indagine pilota Cawi Aspetti della vita quotidiana 2014.

Tavola 6.2 - Famiglie che hanno completato i questionari per numero componenti, presenza di giovani, anziani e titolo di studio più alto conseguito e tecnica utilizzata. Composizione percentuale

NUMERO DI COMPONENTI PRESENZA DI ALMENO UNA PERSONA FINO A 30 ANNI IN FAMIGLIA FAMIGLIE PER PRESENZA DI ANZIANI TITOLO DI STUDIO PIÙ ALTO IN FAMIGLIA (*)	Avq Papi		Avq Cawi		Popolazione di riferimento	
	V.A.	%	V.A.	%	V.A.	%
1 componente	5.603	30,6	404	25,6	7.748	31,0
2 componenti	5.216	28,4	433	27,5	6.805	27,2
3 componenti	3.627	19,8	392	24,9	5.017	20,1
4 componenti	2.951	16,1	278	17,6	4.103	16,4
5 o più componenti	940	5,1	69	4,4	1.347	5,4
Nessuna persona fino a 30 anni	10.816	59,0	828	52,5	14.368	57,4
Almeno una persona fino a 30 anni	7.521	41,0	748	47,5	10.654	42,6
Monocomponente >=65 anni	2.782	15,2	147	9,3	3.656	14,6
Monocomponente <65 anni	2.821	15,4	257	16,3	4.092	16,4
Pluricomponente tutti >=65 anni	2.022	11,0	114	7,2	2.478	9,9
Pluricomponente almeno uno <65 anni	10.712	58,4	1058	67,1	14.795	59,1
Dottorato di ricerca o laurea	3.642	19,9	543	37,0	(**)	(**)
Diploma superiore	7.364	40,2	664	45,2	(**)	(**)
Licenza mediai	4.046	22,1	162	11,0	(**)	(**)
Licenza elementare, nessun titolo	3.285	17,9	100	6,8	(**)	(**)
Totale	18.337	100,0	1.576	100,0		

** Non è possibile confrontare la distribuzione del titolo di studio osservata nell'indagine con quella dell'universo di riferimento poiché questa informazione era assente nella lista anagrafica originaria da cui è stato estratto il campione.

Fonte: Indagine Papi "Aspetti della vita quotidiana" 2014 e Indagine pilota Cawi "Aspetti della vita quotidiana" 2014.

dio o alto a prediligere la modalità di compilazione Cawi. Se si considerano le famiglie con almeno un componente laureato, si osserva uno scarto di 17 punti percentuali rispetto ad Avq Papi. Infine, se si considera l'utilizzo del PC e di Internet, appare evidente come con la tecnica Cawi si siano selezionate per lo più famiglie con competenze informatiche (Tavola 6.3).



Le caratteristiche familiari sono correlate con quelle degli individui che le compongono.

Tavola 6.3 - Famiglie per uso settimanale del pc, skills informatici e uso settimanale di Internet. Confronto tra Avq indagine Papi edizione 2014 e Avq pilota Cawi 2014. Dati in migliaia e valori percentuali

	Avq Papi		Avq Cawi	
	V.A.	%	V.A.	%
Frequenza di uso del pc				
Uso del pc meno di una volta a settimana	6.384	34,8	273	17,3
Almeno una persona in famiglia usa il pc almeno 1 volta a settimana	11.953	65,2	1.303	82,7
Skills pc				
Nessuno in famiglia ha alti livelli di competenza informatica	12.035	65,6	739	46,9
Almeno una persona in famiglia ha alti livelli di competenza informatica	6.302	34,4	837	53,1
Frequenza di uso di internet				
Uso di internet meno di 1 volta a settimana	6.246	34,1	260	16,5
Almeno una persona in famiglia usa Internet almeno 1 volta a settimana	12.091	65,9	1.316	83,5
Totale	18.337	100,0	1.576	100,0

Fonte: Indagine Papi Aspetti della vita quotidiana 2014 e Indagine pilota Cawi Aspetti della vita quotidiana 2014.

La tecnica Cawi porta, come atteso (Tavola 6.4), ad una presenza maggiore di giovani tra i rispondenti, non si evidenziano invece differenze di genere. Le classi dai 35 ai 64 anni sono

Tavola 6.4 - Persone per classe di età. Confronto tra Avq indagine Papi edizione 2014 e Avq pilota Cawi 2014. Dati in migliaia e valori percentuali

CARATTERISTICHE INDIVIDUALI	Papi		Cawi		Popolazione di riferimento	
	Valori assoluti	%	Valori assoluti	%	Valori assoluti	%
SESSO						
MASCHI	21186	48,5	1.920	49,0	29.123	48,6
FEMMINE	22518	51,5	1.995	51,0	30.847	51,4
Totale	43.704	100,0	3.915	100,0	59.970	100,0
ETA'						
0-5	2203	5,0	195	5,0	3.276	5,5
6-13	3273	7,5	306	7,8	4.527	7,5
14-24	4749	10,9	453	11,6	6.485	10,8
25-34	4375	10,0	396	10,1	6.863	11,4
35-44	6.343	14,5	627	16,0	9.163	15,3
45-54	6.914	15,8	705	18,0	9.394	15,7
55-64	5.794	13,3	589	15,0	7.521	12,5
65 e più	10.053	23,0	644	16,4	12.741	21,2
Totale	43.704	100,0	3.915	100,0	59.970	100,0
Titolo di studio*						
Dottorato di ricerca o laurea	4.613	13,4	676	24,6		(**)
Diploma superiore	11.743	34,1	1.210	44,0		(**)
Licenza media	10.015	29,1	560	20,4		(**)
Licenza elementare, nessun titolo	8.035	23,4	303	11,0		(**)
Totale	34.406	100,0	2.749	100,0		

(*) Persone di 25 anni e più.

(**) Non è possibile confrontare la distribuzione del titolo di studio osservata nell'indagine con quella dell'universo di riferimento poiché questa informazione era assente nella lista anagrafica originaria da cui è stato estratto il campione.

Fonte: Indagine Papi "Aspetti della vita quotidiana" 2014 e Indagine pilota Cawi "Aspetti della vita quotidiana" 2014

più rappresentate di quanto non avvenga nella Papi, in cui al contrario c'è una maggiore presenza di individui appartenenti alle età più anziane (65 anni e più). Osservando la distribuzione per il titolo di studio, chi ha un titolo di studio più alto ha una maggiore propensione ad utilizzare la tecnica Cawi: la quota di chi possiede un titolo universitario è doppia in Cawi rispetto alla Papi.

Le differenze socio-demografiche dei rispondenti nelle due indagini campionarie implicano una differenza tra le distribuzioni di frequenza delle risposte fornite ai quesiti proposti.

Obiettivo dei paragrafi che seguono è studiare l'effetto dell'utilizzo di tecniche diverse di raccolta dati, confrontando le risposte raccolte attraverso l'indagine pilota condotta con metodo Cawi con quelle della Papi condotta nel marzo dello stesso anno, cercando di isolare attraverso modelli specifici l'effetto selezione e misurazione dovuti alla differenti tecniche di rilevazione.

6.2.4 Caratteristiche dell'applicazione metodologica

Oltre alla selezione dei rispondenti, l'assunzione di base è anche che le tecniche di raccolta dati possono comportare una certa distorsione nelle risposte fornite ai questionari proposti. Naturalmente, molteplici sono i fattori e gli stimoli che nel presente contesto di indagine possono influenzare i rispondenti in maniera differente dipendentemente dalla tecnica di rilevazione utilizzata. Si pensi, solo per fare qualche esempio, all'obbligatorietà di risposta per la tecnica Cawi, oppure alla con-sapevolezza del rispondente del numero di quesiti rimanenti nel questionario carta-ceo (informazione non disponibile per il questionario web).

In generale, l'errore di misura causato dall'utilizzo di una tecnica dipende indirettamente: (1) dal processo cognitivo ottimale necessario per rispondere in modo veritiero (*cognitive burden*) e (2) dalla motivazione, quindi dallo sforzo necessario all'esecuzione del processo di risposta. Da questi si generano sia errori casuali sia sistematici nelle risposte e quindi nelle stime (Krosnick, 1999). In particolare, in base alla propria motivazione, il rispondente adotta diverse strategie di risposta, tra le quali: (1) la tendenza a selezionare sempre la stessa opzione di risposta alle scale di misura, in presenza di domande simili (*nondifferentiation response set*); (2) la tendenza a rispondere sempre in modo affermativo (*acquiescence response set*); (3) la tendenza a scegliere le opzioni estreme nelle scale di misura (*extremeness response set*); (4) la tendenza a rispondere in base a ciò che è socialmente più accettabile (*social desirability response set*) (de Leeuw, 2005). Quest'ultima strategia di risposta, in presenza di domande sensibili, comporta generalmente la generazione di errori sistematici ("*social desirability bias*"). Infatti, il rispondente potrebbe non rivelare il proprio pensiero ritenendo che questo possa essere considerato socialmente non appropriato da parte di un intervistatore o nel caso fosse talmente personale da temere che una terza persona possa rivelarlo a terze parti (Tourangeau e Yan, 2007).

Tra le domande del questionario utilizzate per condurre le due indagini Cawi e Papi, sono state scelte alcune variabili *multi-item* utili alla misurazione delle abitudini alimentari (costrutto latente), assumendo che le risposte possano essere influenzate dalla tecnica utilizzata a causa di comportamenti guidati dalla desiderabilità sociale (Tavola 6.5).

Tavola 6.5 - Descrizione delle variabili utilizzate per l'analisi confermativa fattoriale multigruppo

Variabili ^(a)	Modalità
Frequenza di consumo di carni rosse	1='almeno una volta al giorno'
Frequenza di consumo di latticini	2='talvolta nella settimana'
Frequenza di consumo di snacks e dolci	3='mai o occasionalmente'
Frequenza di consumo di frutta e verdura	1='mai o occasionalmente'
Frequenza di consumo di pesce	2='talvolta nella settimana'
	3='almeno una volta al giorno'
Attenzione al consumo di sale ^(b)	1='utilizzo illimitato, sale non iodato'
	2='utilizzo illimitato, sale iodato' + 'utilizzo limitato, sale non iodato'
	3='utilizzo limitato, sale iodato'

(a) Le variabili originali hanno cinque modalità ('più di una volta al giorno'; 'una volta al giorno'; 'qualche volta nella settimana'; 'meno di una volta al giorno'; 'mai'): sono state ricodificate in variabili con tre categorie per evitare di ottenere frequenze molto basse. È stata attribuita l'accezione maggiormente positiva alla modalità con punteggio più elevato.

(b) Questa variabile è stata creata ricodificando due domande: una sull'attenzione nel consumo di sale e l'altra sull'utilizzo di sale iodato.

La Tavola 6.6 mostra le distribuzioni di frequenza delle variabili selezionate, distinte per tecnica di raccolta dei dati.

Tavola 6.6 - Distribuzioni di frequenza delle variabili selezionate, distinte per tecnica di raccolta dei dati

Frequenza di consumo	Almeno una volta al giorno	Talvolta nella settimana	Valori assoluti	Mai o occasionalmente
Tecnica di raccolta dati				
Carni rosse				
Papi	13,35	69,74	16,91	100,0
Cawi	11,82	70,75	17,43	100,0
Latticini				
Papi	23,19	63,26	13,55	100,0
Cawi	19,72	66,45	13,83	100,0
Snacks e dolci				
Papi	15,05	37,72	47,24	100,0
Cawi	14,05	42,65	43,3	100,0
Frutta e verdura				
Papi	82,19	15,32	2,49	100,0
Cawi	77,43	19,58	2,99	100,0
Pesce				
Papi	3,27	53,68	43,05	100,0
Cawi	2,43	53,34	44,23	100,0
Attenzione al consumo di sale	Nessuna attenzione al consumo e non iodato	Nessuna attenzione al consumo e iodato/ Attenzione al consumo e non iodato	Attenzione al consumo e iodato	Totale
Sale				
Papi	19,26	42,33	38,41	100,0
Cawi	12,67	42,79	44,54	100,0

Nel caso di dati raccolti con diverse tecniche di indagine, non si può assumere che le risposte dei rispondenti ad un certo item o scala possano essere comparate direttamente. Infatti, la comparabilità delle risposte ottenute dipende dal loro grado di equivalenza. L'equivalenza delle misurazioni nelle indagini con tecnica mista comporta la stessa grandezza e la stessa direzione dell'errore di misura.

Per analizzare la natura del fenomeno relativo alle abitudini alimentari, è stato selezionato il metodo diagnostico MCFA (cfr. Paragrafo 4.3.2. per i dettagli teorici). L'obiettivo è quello di

valutare se il modello di misura utilizzato per l'analisi delle abitudini alimentari con le risposte raccolte con l'indagine pilota abbia gli stessi valori dei parametri rispetto a quello ottenuto con le risposte all'indagine Papi, e se i livelli della media dei fattori latenti utili alla misurazione del fenomeno siano uguali tra le due tecniche. La MCFA può essere svolta solo dopo aver controllato per l'effetto selezione, pertanto, prima di descriverne i risultati (Paragrafo 6.2.6), viene presentata un'analisi della copertura e della non risposta nelle due indagini (Paragrafo 6.2.5) volta all'individuazione delle variabili socio-demografiche utili a controllare per l'effetto selezione. Questa analisi è stata seguita da un'analisi esplorativa per l'individuazione della struttura latente del fenomeno oggetto di studio.

6.2.5 L'effetto selezione

L'effetto selezione si presenta quando diversi tipi di rispondenti decidono di rispondere ad un'indagine scegliendo, tra quelle disponibili, la tecnica da loro preferita. Sono forme di errori di non risposta: diversi tipi di rispondenti non utilizzano un certo metodo di risposta auto-selezionandosi invece per un altro metodo. Inoltre, l'effetto selezione dipende dalla copertura della popolazione dal momento che la tecnica di raccolta dei dati è determinante per la predisposizione della lista di estrazione utilizzabile. Nell'analisi dell'effetto selezione, quindi, la non risposta e la copertura devono essere considerati congiuntamente.

Nonostante l'indagine pilota (con tecnica Cawi) e l'indagine con tecnica Papi siano state realizzate indipendentemente l'una dall'altra, senza dare la possibilità di scegliere il metodo di risposta preferito ai membri del campione, non è comunque possibile testare l'invarianza nelle relative misurazioni senza tener conto dell'effetto della non risposta e della copertura, cioè l'effetto selezione, sulle stime finali d'indagine (Voogt e Saris, 2005). Nell'analisi che segue, sono state individuate le variabili socio-demografiche determinanti per l'analisi della non risposta. Queste sono state, quindi, utilizzate per controllare per l'effetto selezione nel successivo svolgimento dell'analisi fattoriale confermativa multigruppo.

Come illustrato in precedenza, i tassi di risposta alle due indagini sono molto diversi (Tavola 6.7): con la pilota è stato ottenuto circa il 13 per cento delle risposte mentre con la Papi si è raggiunto circa il 75 per cento.

Tavola 6.7 - Composizione dei campioni delle indagini Cawi e Papi (valori assoluti e percentuali)

Tecnica di raccolta dei dati	Numero di rispondenti	Numero di non rispondenti	% di rispondenti	% di non rispondenti	Campione totale
Cawi	3.916	24.440	13,8	86,2	28.356
Papi	44.992	14.820	75,2	24,8	59.812

La non risposta, specialmente quella piuttosto elevata dell'indagine pilota, è una fonte di errore potenziale. Come descritto nella letteratura più recente, le analisi delle differenze tra tassi di risposta in indagini con tecnica di rilevazione mista evidenziano che il metodo di risposta preferito dipende dalle caratteristiche socio-demografiche dei rispondenti, come il sesso, il reddito, la nazionalità e l'istruzione. Fortunatamente, questo tipo di informazioni è disponibile sia per i rispondenti che per i non rispondenti ed è stato possibile utilizzarle per ulteriori analisi della non risposta nelle due indagini. Le informazioni ausiliarie sono disponibili nelle liste anagrafiche dei comuni italiani (Lac) – per quanto riguarda l'età, il sesso e la cittadinanza – e nell'ARCHivio

Integrato di Microdati Economici e DEMografici (Archimede,² Garofalo, 2014) – per quanto riguarda variabili come la tipologia della famiglia, il numero di componenti, il numero di stranieri nella famiglia, il reddito della famiglia, il livello di istruzione del capofamiglia e la condizione occupazionale degli individui.

La disponibilità di variabili sia al livello individuale sia di famiglia ha permesso di utilizzare dei modelli logistici multilivello per la stima della probabilità di risposta delle due indagini Cawi e Papi, condizionatamente a tali variabili socio-demografiche. La struttura multilivello si ha in quanto gli individui sono naturalmente raggruppati in gruppi, ovvero le famiglie.

Per l'indagine svolta con metodo di raccolta dati Cawi (Tavola 6.8), la probabilità di non risposta è spiegata, con un livello di significatività dello 0,05, dalle seguenti variabili:

1. cittadinanza (Italiana o meno);
2. età, in particolare per individui dai 55 ai 64 anni rispetto ai giovani di età inferiore ai 15 anni;
3. istruzione elevata e di medio livello del capofamiglia, rispetto alla bassa istruzione;
4. numero di stranieri in famiglia;
5. numero di componenti della famiglia;
6. reddito complessivo della famiglia.

Per l'indagine svolta con metodo di raccolta dati Papi (Tavola 6.9), la probabilità di non risposta è spiegata, con un livello di significatività dello 0,05, da:

1. sesso;
2. cittadinanza (Italiana o meno);
3. età, per tutte le classi di età, a partire dai 25 ai 34 anni fino ai 65-74 anni e oltre, rispetto ai giovani di età inferiore ai 15 anni;
4. istruzione elevata del capofamiglia, rispetto a un basso livello di istruzione;
5. occupazione (sì/no);
6. famiglie con bambini rispetto a famiglie senza bambini;
7. numero di stranieri in famiglia;
8. numero di componenti della famiglia.

In entrambi i modelli utilizzati, è presente una componente di variabilità tra le famiglie (come mostrato dalla significatività del parametro “teta02”) che si aggiunge a quella tra individui. Si tratta della varianza dell'effetto casuale dovuto alla presenza di una struttura gerarchica dei dati - gli individui raggruppati nelle famiglie - che ha reso possibile la stima di modelli multilivello.

Concludendo, l'adozione di modelli logistici multilivello ha consentito di individuare le variabili socio-demografiche di cui tenere conto nello svolgimento della MCFA, in modo tale da rendere simili i campioni delle indagini Cawi e Papi e così che i dati raccolti con esse possano essere considerati epurati da qualsiasi effetto selezione.

² Archimede è un progetto Istat che ha l'obiettivo di ampliare l'offerta informativa dell'Istat mediante la produzione di collezioni di dati elementari di tipo longitudinale (ad es. percorsi sociali ed economici) e cross section da rendere disponibili all'utenza, utili alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a livello nazionale, regionale e locale.

6. Analisi dell'effetto tecnica in due casi di Studio

Tavola 6.8 - Modello selezionato per la probabilità di risposta nell'indagine Cawi

Parametri	Stime	Errori	t-Value	Pr > t
Intercetta	-13,7499	0,7637	-18,00	<,0001
teta02	49,1007	3,3462	14,67	<,0001
Donna	0,09763	0,08265	1,18	0,2375
Cittadinanza italiana	2,4897	0,5866	4,24	<,0001
15-24 anni	0,1038	0,1940	0,54	0,5926
25-34 anni	-0,4750	0,2146	-2,21	0,0269
35-44 anni	-0,02597	0,1924	-0,13	0,8926
45-54 anni	0,04961	0,1940	0,26	0,7982
55-64 anni	0,4671	0,1997	2,34	0,0194
65-74 anni	0,1070	0,2095	0,51	0,6096
>75 anni	-0,4002	0,2244	-1,78	0,0746
Livello medio di istruzione del capofamiglia	3,9901	0,2397	16,65	<,0001
Livello elevato di istruzione del capofamiglia	4,6563	0,2664	17,48	<,0001
Occupato	0,2092	0,1168	1,79	0,0734
Famiglia con figli	0,2478	0,1987	1,25	0,2124
Altri tipi di famiglia	-0,4605	0,2407	-1,91	0,0558
Numero di stranieri in famiglia	-0,9284	0,2571	-3,61	0,0003
Numero di component nella famiglia	-0,1747	0,07418	-2,35	0,0185
Reddito	0,000013	1,613E-6	8,05	<,0001

Tavola 6.9 - Modello selezionato per la probabilità di risposta nell'indagine Papi

Parametri	Stime	Errori	t-Value	Pr > t
Intercetta	2.3637	0.4179	5.66	<.0001
teta02	445.2	21.3889	20.81	<.0001
Donna	0.3285	0.06616	4.97	<.0001
Cittadinanza italiana	2.871	0.3305	8.69	<.0001
15-24 anni	0.02767	0.1622	0.17	0.8646
25-34 anni	-1.3171	0.1524	-8.64	<.0001
35-44 anni	-0.372	0.1498	-2.48	0.013
45-54 anni	1.1168	0.1622	6.88	<.0001
55-64 anni	2.5923	0.1795	14.44	<.0001
65-74 anni	3.4365	0.1949	17.63	<.0001
>75 anni	2.1159	0.1823	11.61	<.0001
Livello medio di istruzione del capofamiglia	0.5457	0.1787	3.05	0.0023
Livello elevato di istruzione del capofamiglia	0.3919	0.2471	1.59	0.1127
Occupato	0.05904	0.09063	0.65	0.5148
Famiglia con figli	1.457	0.2732	5.33	<.0001
Altri tipi di famiglia	0.03361	0.3468	0.1	0.9228
Numero di stranieri in famiglia	0.1783	0.1563	1.14	0.2538
Numero di component nella famiglia	0.6906	0.1042	6.63	<.0001
Reddito	2.3637	0.4179	5.66	<.0001

6.2.6 Analisi fattoriale confermativa multigruppo

Nel presente paragrafo viene applicata un'analisi MCFA, così come illustrata da un punto di vista teorico nel Paragrafo 4.3.2, con l'obiettivo principale di capire quali sono i costrutti latenti relativi ai comportamenti alimentari dei rispondenti e alla loro "dieta". Suc-

cessivamente, si vuole analizzare sia se esiste invarianza di misura (configurale, metrica, scalare) tra i gruppi (Cawi / Papi) del modello fattoriale, sia se il livello medio dei fattori latenti è uguale tra i diversi metodi di raccolta dati.

Le variabili selezionate per le analisi sono descritte nella Tavola 6.5.

Tra le diverse parametrizzazioni possibili del modello fattoriale in presenza di risposte di tipo ordinale, le più utilizzate sono quelle del modello *cumulative link* e del modello *adjacent-category logit* (Agesti, 2002). In dettaglio, data una variabile risposta ordinale y_{ih} con categorie $c = 1, 2, \dots, C-1$, per l'individuo i e l'indicatore h , il modello cumulative link standard esprime la probabilità della variabile risposta di essere maggiore o uguale alla categoria c come:

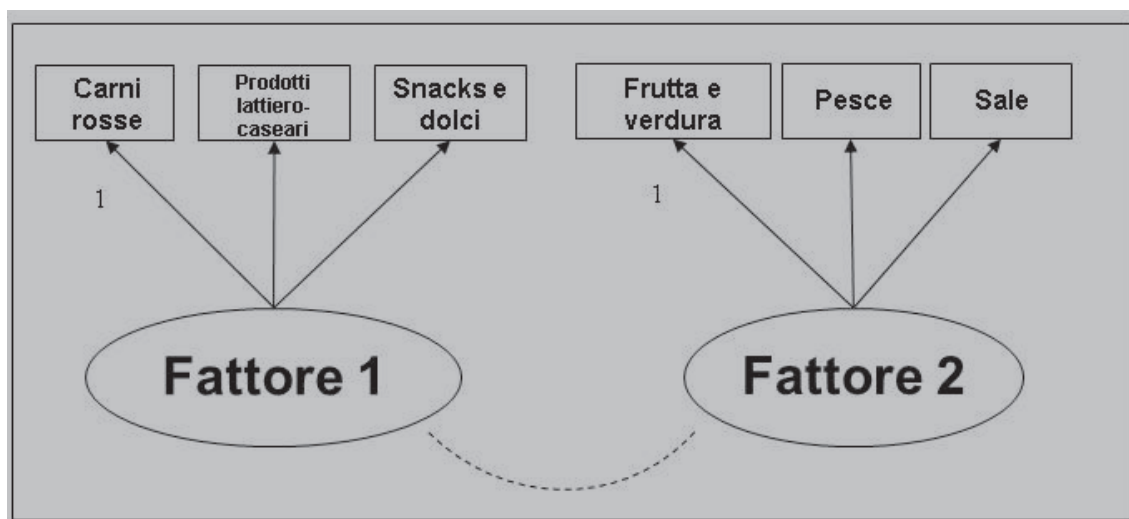
$$F^{-1}[P(y_{ih} \geq c | \eta_i)] = \alpha_{hc} + \sum_j \lambda_{hj} \eta_{ij}$$

dove F è comunemente scelta come una funzione di ripartizione normale o logistica (Muthén e Asparouhov, 2002), η_i è un vettore contenente le variabili latenti J per l'individuo i , η_{ij} è il valore della j -ima ($j = 1, 2, \dots, J$) variabile latente per l'individuo i e λ_{hj} sono i pesi fattoriali che mettono in relazione il fattore latente η_j con la variabile osservata y_{ih} ; α_{hc} sono le intercette del modello, dette soglie. La caratteristica principale di questo modello è che l'effetto di η_{ij} è lo stesso per ogni probabilità cumulata indipendentemente dalla categoria di y . Inoltre, ogni *cumulative link* ha una propria intercetta.

In questa applicazione esistono 3 categorie per ogni indicatore e sono stimate due soglie α_c ; F è stata scelta come una funzione di ripartizione normale standard. Il software utilizzato per eseguire le analisi è *LatentGOLD*, versione sintassi 5.0 (Vermunt e Magidson, 2013).

L'analisi fattoriale confermativa indica la presenza di due fattori latenti correlati (Figura 6.1) relativi alle abitudini alimentari e alla dieta dei rispondenti: un fattore rappresenta il comportamento legato al "Consumo limitato di alimenti non sani (proteine e dolci)" (misurato dagli indicatori "Frequenza di consumo di carni rosse", "Frequenza di consumo di latticini" e "Frequenza di consumo di snacks e dolci"), l'altra rappresenta "l'attenzione a una dieta sana" (misurata dagli indicatori "Frequenza di consumo di frutta e verdura", "Frequenza di consumo di pesce" e "Attenzione al consumo di sale").

Figura 6.1 - Modello fattoriale, due fattori latenti correlati



6. Analisi dell'effetto tecnica in due casi di Studio

Come mostrato nella Tavola 6.10, tutti i pesi fattoriali sono positivi: maggiore è il punteggio fattoriale maggiore è la probabilità $P(y_{ih} \geq c | \mathbf{n}_i)$, ovvero maggiore è il livello del fattore latente maggiore è l'attenzione alla salute alimentare. L'indicatore più importante relativo al primo costrutto latente è quello relativo al consumo di latticini, mentre l'indicatore con il coefficiente più alto che misura il secondo costrutto è quello relativo al consumo di pesce.

Tavola 6.10 - Modello fattoriale, parametri

Indicatore	Pesi fattoriali		Soglie	
	Fattore latente1	Fattore latente2	α_1	α_2
Carni rosse	1		1,34	-1,11
Prodotti lattiero-caseari	1,26		0,81	-1,47
Snacks e dolci	0,74		1,17	-0,12
Frutta e verdura		1	2,31	1,09
Pesce		1,15	0,26	-2,09
Sale		0,37	0,77	-0,46
Varianza fattore latente	0,34	0,26		
Covarianza tra fattori latenti		-0,10		

Dopo aver analizzato la struttura latente sottostante il fenomeno del comportamento alimentare, è stata applicata una MCFA per valutare l'eventuale presenza di invarianza di misura tra le diverse modalità di raccolta dati. Le covariate ausiliarie utilizzate per controllare l'effetto selezione (errore di copertura e errore di non risposta considerati congiuntamente) sono: sesso, cittadinanza, età, istruzione del capofamiglia, condizione occupazionale, caratteristiche familiari (presenza di bambini, stranieri, numero di componenti). Inoltre, al fine di soddisfare le condizioni di ignorabilità del disegno, è stata utilizzata una covariata legata alla tipologia comunale di appartenenza (ampiezza demografica del comune) di appartenenza dei rispondenti.

Dopo aver stimato due modelli con la stessa struttura (stessi costrutti latenti, stessi schemi di parametri fissi, liberi e vincolati) per i rispondenti Cawi e i rispondenti Papi che hanno presentato una buona adattabilità ai dati, è stata utilizzata la gerarchia di ipotesi descritta nel Paragrafo 4.3.2 per testare la presenza di invarianza di misura. In particolare, è stata verificata la presenza di invarianza configurale e scalare. Infatti, nel contesto di modelli fattoriali con variabili di tipo ordinale i parametri che risultano rilevanti per lo studio dell'invarianza di misura sono le soglie e i pesi fattoriali (Muthén e Asparouhov, 2002): le variazioni dei pesi fattoriali, infatti, implicano anche cambiamenti nelle intercette. Se l'invarianza scalare è raggiunta, viene verificata la presenza di differenze nelle medie e nelle varianze dei fattori latenti. Per valutare la significatività delle restrizioni sui parametri è stato utilizzato il test del rapporto di verosimiglianza per i modelli nidificati (cioè per i modelli in cui il modello da sottoporre a test è ottenuto vincolando alcuni parametri del modello precedente) e il confronto dell'indice *Bayesian Information Criteria* (BIC) per i modelli non nidificati.

Come illustrato nel Paragrafo 4.3.2, l'invarianza configurale assume che le stesse variabili osservate siano associate con gli stessi fattori latenti. L'invarianza scalare presuppone che i pesi fattoriali e le soglie degli indicatori siano uguali tra i gruppi formati dalle modalità di raccolta dati Papi e Cawi. Il test del rapporto di verosimiglianza tra i primi due modelli suggerisce che è verificata solo l'invarianza configurale: i due gruppi misurano i due costrutti latenti in modo differente (Tavola 6.11).

Tavola 6.11 - Modelli con invarianza di misura per i dati raccolti con tecnica Cawi e Papi

MODELLO	LL	BIC	Npar
Invarianza configurale: modello senza vincoli	-228,798	459,173	146
Invarianza scalare	-228,906	459,217	130
Invarianza scalare con medie dei fattori latenti diverse	-228,899	459,223	132

La Figura 6.2 mostra le probabilità a posteriori degli indicatori osservati condizionatamente ad alcuni valori (0,1,2) dei due fattori latenti e all'appartenenza al gruppo (Cawi o Papi), per un "rispondente tipico", ossia un rispondente con le caratteristiche più frequenti: femmina, italiana, di circa 44 anni, occupata, con livello basso di istruzione del capofamiglia, in una famiglia di 4 componenti con figli e nessuna persona straniera. Mentre gli indicatori "Frequenza del consumo di carne rossa" e "Frequenza di consumo di frutta e verdura" mostrano lo stesso comportamento di risposta alla variazione del fattore latente, le probabilità a posteriori condizionate degli indicatori "Frequenza di consumo di *snacks* e dolci" e "Attenzione al consumo di sale" mostrano uno spostamento nelle intercette, e quelle degli indicatori "Frequenza del consumo di latticini" e, soprattutto, "Frequenza del consumo di pesce" mostrano un comportamento completamente diverso nei due gruppi.

6.2.7 Conclusioni e ulteriori sviluppi

Concludendo, nell'indagine Avq non esiste invarianza di misura delle variabili relative ai comportamenti alimentari e alla dieta nei due gruppi formati dai dati raccolti con le tecniche Cawi e Papi, probabilmente a causa di comportamenti influenzati da *social desirability* o da un diverso effetto del *response set* indotto dal questionario cartaceo o elettronico. Come prima conseguenza, non è possibile confrontare le stime finali ottenute dai dati raccolti con tecniche differenti. Per il futuro, possono essere utilizzati due approcci: (1) raccogliere i dati attraverso una singola tecnica o (2) valutare la possibilità di utilizzare questionari differenti per le due tecniche.

Naturalmente i risultati ottenuti dipendono da alcune scelte effettuate che sono state condizionate dalla qualità dei dati, e ancor più dalla disponibilità delle informazioni utilizzate, nonché dal set di variabili considerate che si riteneva potesse avere delle differenze dovute alla tecnica utilizzata. Come già detto, scopo del presente Capitolo è stato quello di illustrare i passaggi logici delle procedure condotte piuttosto che un'analisi dei risultati empirici ottenuti.

6. Analisi dell'effetto tecnica in due casi di Studio

Figura 6.2 - Probabilità a posteriori degli indicatori osservati condizionati ai fattori latenti e all'appartenenza al gruppo, per un "rispondente tipo"



6.3 La sperimentazione sull'indagine I Cittadini e il tempo libero 2015

6.3.1 Obiettivi della sperimentazione

I dati utilizzati per la sperimentazione sono relativi all'indagine Ctl, che per la prima volta nell'edizione del 2015 è stata realizzata con tecnica mista sequenziale Cawi/Papi. L'obiettivo principale

della sperimentazione è stato di analizzare e trattare gli effetti selezione e misurazione introdotti dal mix di tecniche attraverso alcuni metodi inferenziali descritti nel Capitolo 5.

A tal fine è stato applicato il metodo del *propensity score* (cfr. Paragrafo 5.4.2) con il duplice obiettivo di separare i due effetti e valutare l'introduzione di un correttore dell'effetto selezione. Tale metodo offre la possibilità di analizzare all'interno di gruppi omogenei di rispondenti l'errore di misura per singole variabili target dell'indagine depurato dell'effetto selezione. Successivamente, per alcune variabili per le quali sono emerse alcune evidenze sulla presenza di effetti misurazione, è stata sperimentata una metodologia di aggiustamento di tale effetto basata su modelli di imputazione multipla (cfr. Paragrafo 5.4.4).

Di seguito si descrive l'indagine Ctl e si riportano alcune analisi esplorative volte a identificare differenze nei tassi di risposta con le due tecniche per particolari distribuzioni di rispondenti e ad analizzare la rappresentatività dei campioni di rispondenti con le due tecniche Cawi e Papi rispetto al campione dei non rispondenti. Successivamente si passa alla descrizione dei metodi applicati e alla presentazione di alcuni risultati. E' da sottolineare che per le analisi esplorative e l'analisi dei modelli sottostanti il metodo del *propensity score* e il metodo di imputazione multipla sono state utilizzate variabili ausiliarie mode-insensitive come sesso, età, cittadinanza, localizzazione geografica desunte da Lac e altre, come reddito, titolo di studio e stato occupazionale acquisite dal database Archimede che è frutto dell'integrazione di più archivi amministrativi. Tale fase di abbinamento del file contenente i nominativi del campione teorico con i nominativi presenti nel database Archimede è descritta in una sezione che precede la descrizione dei metodi.

Nello studio qui descritto non sono state condotte analisi relative al trattamento della mancata risposta totale che, per un'indagine a tecnica mista di tipo sequenziale, come illustrato nel Capitolo 5 (cfr. Paragrafo 5.5.2), non è possibile trattare in modo complessivo in un'ottica di correzione dei pesi campionari, ma richiede l'utilizzo di modelli complessi da stimare per le singole variabili di indagine.

6.3.2 Cenni sulla rilevazione

L'indagine Ctl fa parte del sistema integrato di indagini sociali denominato "Indagini Multiscopo sulle Famiglie". L'indagine è stata svolta per la prima volta nel 1995³ e le rilevazioni successive sono state condotte con cadenza periodica pluriennale. L'indagine ha permesso nel tempo di descrivere in modo sistematico un campo complesso e differenziato qual è quello del tempo libero e dei rapporti che sussistono tra questo e la partecipazione culturale.

La rilevazione si concentra a rilevare gli atteggiamenti e i comportamenti della popolazione riconducibili alla sfera del tempo libero, con particolare riguardo all'insieme di comportamenti e di attività legate alla partecipazione culturale, alla pratica sportiva e alle attività più direttamente legate alla sfera del sé e dell'autorealizzazione. I temi oggetto dell'indagine vengono analizzati sia sotto gli aspetti più tradizionali sia sotto quelli emergenti, dando ampio spazio ad approfondimenti nelle varie sezioni dei questionari. Si va dalla fruizione della televisione e della radio alle letture, dall'uso di internet e delle nuove tecnologie agli aspetti legati alla fruizione di spettacoli dal vivo (concerti, spettacoli teatrali, manifestazioni sportive).

L'indagine permette di capire i profondi cambiamenti che stanno avvenendo nella nostra società relativamente al tempo libero, contestualizzato rispetto al più ampio tessuto della vita quotidiana.

³ L'indagine, nella sua prima edizione era denominata "Tempo libero e cultura". Le edizioni successive sono state effettuate nel 2000, nel 2006 e nel 2015.

6. Analisi dell'effetto tecnica in due casi di Studio

L'indagine Ctl rientra tra rilevazioni comprese nel Programma statistico nazionale che raccoglie l'insieme delle indagini statistiche necessarie al Paese e viene condotta su tutti i componenti di un campione di circa 24.000 famiglie, estratte dai registri anagrafici comunali e distribuite in circa 850 comuni italiani di diversa ampiezza demografica, attraverso un disegno di campionamento a due stadi.

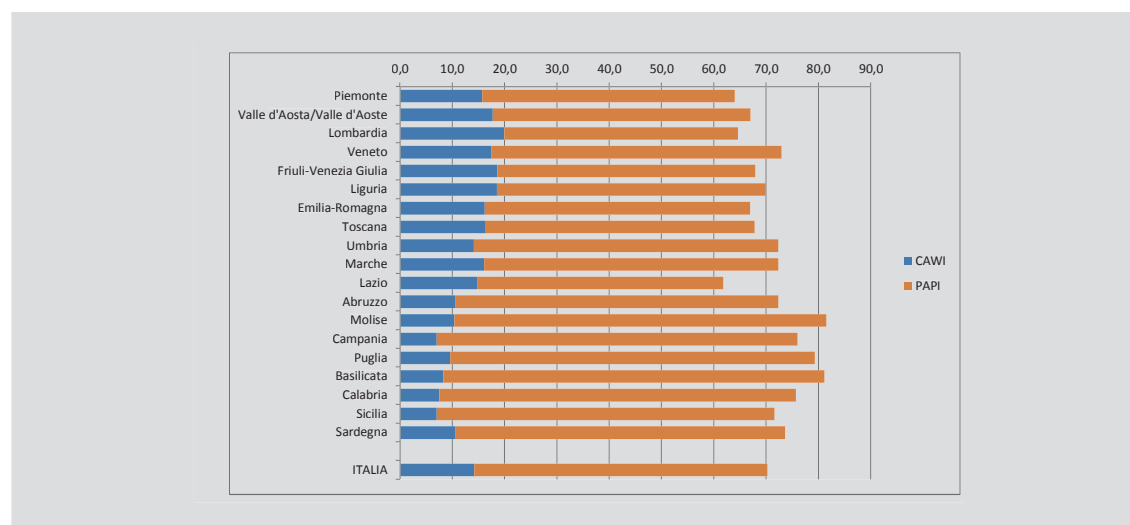
Fino alla rilevazione condotta nel 2006 l'indagine è stata svolta con tecnica Papi, tramite l'uso di più modelli di rilevazione: un modello di rilevazione somministrato tramite intervista diretta con l'ausilio di un intervistatore; uno o più modelli⁴ somministrati per auto-compilazione. Nel 2015, per la realizzazione della quarta wave dell'indagine, è stata introdotta un'importante innovazione che ha riguardato la tecnica di raccolta dati, consistita nell'applicazione della tecnica mista sequenziale Cawi/Papi.

Le famiglie sono state in prima battuta invitate a partecipare alla rilevazione rispondendo alle domande presenti nei modelli di rilevazione tramite web (Cawi). Successivamente, alle famiglie che non avevano partecipato all'indagine via web, è stata data la possibilità di essere intervistate tramite tecnica Papi, con l'ausilio di un rilevatore comunale che ha provveduto a somministrare uno dei due modelli di rilevazione per intervista diretta, mentre l'altro modello è stato consegnato a ciascun componente delle famiglie che ha provveduto a compilarlo personalmente.

6.3.3 Analisi esplorative preliminari

Alla fine del periodo di rilevazione il tasso di risposta complessivo dell'indagine Ctl 2015 è risultato pari al 72,6 per cento. La tecnica Cawi è stata scelta dal 20,9 per cento delle famiglie rispondenti, mentre la tecnica Papi dal 79,1 per cento. Le famiglie intervistate sono risultate in tutto 16.970, per un totale di 39.848 individui.

Figura 6.3 - Tasso di compilazione delle famiglie per regione e tecnica di rilevazione



⁴ Di numero variabile a seconda dell'edizione di indagine. Ad esempio nel 2006 i modelli per autocompilazione erano due: un modello destinato alle persone di 3 anni e più che approfondiva i diversi aspetti legati alle abitudini nel tempo libero; un modello destinato alle persone di 18 anni e più che approfondiva invece alcuni aspetti riguardanti la formazione scolastica e professionale; nel 2015 è stato utilizzato un solo modello per autocompilazione.

Osservando la distribuzione regionale del tasso di compilazione⁵ delle famiglie rispondenti rispetto alla tecnica di rilevazione (Figura 6.3), si osserva una maggiore propensione a rispondere in Cawi per le famiglie residenti al Nord.

La provincia di Trento supera il 20 per cento e la Lombardia, il Friuli Venezia Giulia e la Liguria si attestano tra il 18,5 per cento e il 20 per cento. Anche le altre regioni del Nord e del Centro sono tutte sopra la media nazionale, solo l'Umbria ha un valore lievemente un po' più basso, ma comunque in linea con la media nazionale. Viceversa, al Sud si registrano tassi di compilazione in Cawi più bassi e viceversa tassi di compilazione in Papi decisamente più elevati. I tassi di compilazione in Cawi più bassi si osservano soprattutto in Sicilia (7,0 per cento), Campania (7,0 per cento) e Calabria (7,5 per cento).

Analizzando la distribuzione di alcune caratteristiche familiari per tecnica di rilevazione (Tavola 6.12), si osserva come si siano dimostrate più propense a partecipare in Cawi le famiglie di 2 o più componenti rispetto alle famiglie monocomponenti. Molto probabilmente ciò avviene perché le famiglie di 1 componente sono più spesso composte da anziani, che presentano *skills* informatici inferiori rispetto alla popolazione più giovane. A conferma di ciò si osserva, invece, come hanno risposto maggiormente in Cawi le famiglie in cui è presente almeno un giovane in famiglia (43,4 per cento contro il 38,6 per cento di quanto osservato nella Papi) e viceversa meno le famiglie monocomponenti anziane o pluri-componenti con almeno un anziano.

Tavola 6.12 - Caratteristiche delle famiglie intervistate per tecnica utilizzata (valori assoluti e composizione percentuale)

CARATTERISTICHE FAMILIARI	Cawi		Papi		Popolazione di riferimento	
	V.A.	%	V.A.	%	V.A.	%
NUMERO DI COMPONENTI						
<i>1 componente</i>	941	27,4	4.467	33,0	8.212	32,3
<i>2 componenti</i>	1.023	29,8	3.795	28,0	6.780	26,7
<i>3 componenti</i>	727	21,1	2.538	18,8	4.990	19,6
<i>4 componenti</i>	605	17,6	2.048	15,1	4.071	16,0
<i>5 o più componenti</i>	142	4,1	684	5,1	1.379	5,4
PRESENZA DI ALMENO UNA PERSONA FINO A 30 ANNI IN FAMIGLIA						
<i>Nessuna persona fino a 30 anni in famiglia</i>	1.947	56,6	8.313	61,4	14.733	57,9
<i>Almeno una persona fino a 30 anni in famiglia</i>	1.491	43,4	5.219	38,6	10.700	42,1
PRESENZA DI ANZIANI IN FAMIGLIA						
<i>Monocomponente >=65 anni</i>	358	10,4	2.323	17,2	4.063	16,0
<i>Monocomponente <65 anni</i>	583	17,0	2.144	15,8	4.149	16,3
<i>Pluricomponente tutti >=65 anni</i>	288	8,4	1.546	11,4	2.445	9,6
<i>Pluricomponente almeno uno <65 anni</i>	2.209	64,3	7.519	55,6	14.776	58,1
Totale	3.438	100,0	13.532	100,0	25.433	100,0

Gli effetti del diverso comportamento rispetto alla compilazione Cawi versus Papi si notano anche se si analizza la distribuzione degli individui rispondenti rispetto ad alcune caratteristiche. Dall'analisi della distribuzione degli individui distinti per sesso, età e titolo di studio (Tavola 6.13) si osserva innanzitutto una presenza maggiore di giovani adulti tra i rispondenti in Cawi: le classi dai 35 ai 64 anni sono più rappresen-

⁵ Calcolato come il rapporto tra le famiglie che hanno completato il questionario e il totale delle famiglie campione.

6. Analisi dell'effetto tecnica in due casi di Studio

Tavola 6.13 - Caratteristiche dei rispondenti per tecnica di rilevazione (valori assoluti e composizione percentuale)

CARATTERISTICHE INDIVIDUALI	Cawi		Papi		Universo di riferimento	
	Valori assoluti	%	Valori assoluti	%	Valori assoluti	%
SESSO						
MASCHI	4.024	48,3	15.274	48,5	29.361	48,6
FEMMINE	4.300	51,7	16.250	51,5	31.052	51,4
ETA						
<i>Fino a 5</i>	431	5,2	1.400	4,4	3.183	5,3
<i>6-13</i>	584	7,0	2.217	7,0	4.578	7,6
<i>14-24</i>	906	10,9	3.346	10,6	6.503	10,8
<i>25-34</i>	847	10,2	3.179	10,1	6.829	11,3
<i>35-44</i>	1.280	15,4	4.239	13,4	8.967	14,8
<i>45-54</i>	1.540	18,5	4.973	15,8	9.621	15,9
<i>55-64</i>	1.228	14,8	4.308	13,7	7.662	12,7
<i>65 e più</i>	1.508	18,1	7.862	24,9	13.070	21,6
Totale	8.324	100,0	31.524	100,0	60.413	100,0
Titolo di studio*	1.626	20,6	2.966	9,8	(**)	(**)
<i>Dottorato di ricerca o laurea</i>	3.177	40,3	9.550	31,7	(**)	(**)
<i>Diploma superiore</i>	1.756	22,2	9.405	31,2	(**)	(**)
<i>Licenza media</i>	1.334	16,9	8.203	27,2	(**)	(**)
<i>Licenza elementare, nessun titolo</i>	3.438	100,0	13.532	100,0	25.433	100,0

*Per le persone di 6 anni e più.

**Non è possibile confrontare la distribuzione del titolo di studio osservata nell'indagine con quella dell'universo di riferimento poiché questa informazione era assente nella lista anagrafica originaria da cui è stato estratto il campione.

tate di quanto non avvenga nella Papi, in cui al contrario c'è una maggiore presenza di individui appartenenti alle età più anziane di 65 anni e più. Osservando la distribuzione per il titolo di studio, si osserva, inoltre, come chi ha un titolo di studio più alto ha avuto una maggiore propensione ad utilizzare la tecnica Cawi: la quota di chi possiede un titolo di studio universitario è più del doppio nella Cawi rispetto alla Papi.

6.3.4 Analisi della rappresentatività dei campioni

Gli indicatori di rappresentatività sono basati, come detto nel Capitolo 5 (cfr. Paragrafo 5.5.1), su una misura di variabilità della propensione alla risposta e misurano di quanto il campione di rispondenti di un'indagine si discosta dalla risposta rappresentativa.

Nella Tavola 6.14 sono riportati gli indici di rappresentatività, *R-indicators*, desunti dall'analisi di diversi modelli di risposta. I modelli utilizzati sono modelli probabilistici binomiali di tipo *logit* che stimano la propensione delle unità statistiche (famiglie) a rispondere con una

Tavola 6.14 - Caratteristiche dei rispondenti per tecnica di rilevazione (valori assoluti e composizione percentuale)

Modello di analisi	Indice di rappresentatività
Rispondenti Cawi vs. non rispondenti prima fase	0,816
Rispondenti Cawi vs. non rispondenti seconda fase	0,704
Rispondenti Cawi+Papi vs. non rispondenti seconda fase	0,823
Rispondenti Papi vs. non rispondenti seconda fase	0,792
Rispondenti Cawi vs. rispondenti Papi	0,741

tecnica o al mix di tecniche, dato un insieme di caratteristiche familiari. Le variabili ausiliarie considerate nei modelli, definite a livello familiare, sono riportate nel Paragrafo 6.3.6.

La variabile indicatrice del modello di risposta è stata diversamente definita a seconda dei campioni di rispondenti e non rispondenti che si volevano analizzare.

Poiché la rappresentatività della risposta si ha quando l'indice assume valore 1, dalla Tavola 6.14 si evidenziano alcune differenze nei confronti tra campioni. In particolare l'indice assume un valore più elevato, 0,823, quando si analizza il campione di rispondenti Cawi-Papi con il campione dei non rispondenti finali. Questo potrebbe indicare che all'interno di sottopopolazioni, definite dalle variabili ausiliarie utilizzate, i rispondenti all'indagine Cawi e Papi messi insieme sono più simili, in media, ai non rispondenti. Una possibile interpretazione di questo risultato potrebbe essere il basso tasso di rispondenti Cawi (circa il 20 per cento).

6.3.5 L'acquisizione di informazione ausiliaria

L'abbinamento tra il file relativo al campione teorico dell'indagine Ctl e la base dati Archimede è stato realizzato tramite il codice identificativo degli individui. Tale operazione ha prodotto un aggancio non completo tra i due file. Questo è stato determinato da alcuni problemi derivanti principalmente da ragioni di copertura degli archivi integrati nella base dati utilizzata.

Per gli individui abbinati, si è posto un problema ulteriore dovuto alla mancanza, per una parte di essi, di alcune variabili ausiliarie.

Nella Figura 6.4 si riporta il risultato del linkage effettuato; con il simbolo (•) si indicano le informazioni disponibili, mentre con il simbolo (o) si indicano le informazioni non disponibili.

L'abbinamento tra il campione teorico dell'indagine e l'archivio è avvenuto corretta-

Figura 6.4 - Scenario informativo dell'indagine Ctl dopo la fase di linkage tra il campione teorico e la base-dati Archimede

Linkage	Risposta	Modalità di rilevazione	Y	X ₁	...	X _p
LINKED	Rispondente	Cawi	•	o	...	o
		Papi	•	•		•
	Non rispondente		o	o		o
			o	•		•
NOT LINKED	Rispondente	Cawi	•	O	O	O

6. Analisi dell'effetto tecnica in due casi di Studio

mente nel 95,4 per cento dei casi; tra gli abbinati e i non abbinati i rispondenti sono distribuiti omogeneamente tra il Cawi e il Papi, rispettivamente 21 e 79 per cento negli abbinati, 19,7 e 80,3 nei non abbinati (Tavola 6.15). La stessa omogeneità non si riscontra invece tra i rispondenti e i non rispondenti: infatti la proporzione di rispondenti passa dal 32,7 per cento per i linkati all'11,8 per cento nei non linkati. La principale motivazione di questo squilibrio è data dal fatto che tra i rispondenti non linkati figurano i componenti aggiuntivi delle famiglie di fatto (che rappresentano il 3 per cento circa dei rispondenti totali) per i quali non era disponibile il codice identificativo dalla lista di selezione, ma sono stati ugualmente inclusi nel campione perché trovati nella famiglia al momento dell'intervista.

Tavola 6.15 - Tasso di risposta e percentuale d'abbinamento tra il campione teorico dell'indagine Ctl e la base-dati Archimede

Risposta e tecnica di rilevazione	LINKAGE				Totale	
	Linked		Not Linked		V.A.	%
	n	%	n	%		
Non rispondente	18.209	32,7%	316	11,8%	18.525	31,7%
Rispondente	37.495	67,3%	2.359	88,2%	39.854	68,3%
Cawi	7.862	21,0%	464	19,7%	8.326	20,9%
Papi	29.633	79,0%	1.895	80,3%	31.528	79,1%
Totale	55.704	100%	2.675	100%	58.379	

Le variabili ausiliarie acquisite dal data base Archimede sono: il tipo di lavoro (dipendente, autonomo, non in età lavorativa), il titolo di studio e il reddito e altre variabili ausiliarie che non sono state considerate nelle analisi poiché presentavano un elevato numero di casi in cui l'informazione era mancante.

Il data set ottenuto alla fine del processo di abbinamento e di analisi della completezza dell'informazione acquisita presenta, dunque, una numerosità ridotta rispetto al campione teorico dell'indagine in oggetto.

In tale fase di studio si è deciso di effettuare la sperimentazione sul data set ridotto. Per trattare il problema di dati mancanti sulle variabili dovrebbero essere applicati metodi che incorporano i dati incompleti nell'analisi tenendo conto del meccanismo che ne determina la mancanza.

La scelta è stata determinata da alcune considerazioni, la prima è che la perdita di individui del campione teorico dell'indagine Ctl è non consistente, la seconda è che tale perdita, come visto, si distribuisce tra i rispondenti Cawi e i rispondenti Papi in modo abbastanza uniforme.

6.3.6 Analisi degli effetti selezione e misurazione: il propensity score matching

Il propensity score matching è stato applicato al campione di rispondenti dell'indagine Ctl. La propensione a rispondere alla tecnica Cawi (rispetto alla tecnica Papi) condizionatamente ad un insieme di covariate osservate (X_1, \dots, X_p) è stata stimata con un modello logistico.

E' da precisare che, poiché la risposta è una variabile associabile alla famiglia e non al singolo individuo, il modello è stato definito a livello di cluster, ovvero di famiglia. Le variabili ausiliarie note per gli individui sono state trasformate a livello familiare tenendo conto della modalità prevalente all'interno della famiglia: ad esempio, la tipologia lavorativa prevalente in una famiglia è stata valorizzata come "dipendente" se la maggior parte dei membri della famiglia stessa sono lavoratori dipendenti; analogamente la nazionalità della famiglia è stata definita "italiana" se tutti i membri della famiglia godono della cittadinanza italiana ("mista" o "straniera" negli altri casi).

Stimate le probabilità predette tramite il modello *logit*, che sarà descritto nel successivo paragrafo, si è proceduto alla costruzione di gruppi omogenei di rispondenti rispetto alla propensione di risposta alla tecnica. I gruppi, o classi, sono definiti dai decili della distribuzione delle probabilità predette. In questo caso per la costruzione di gruppi omogenei rispetto alla propensione alla risposta con tecnica Cawi si è fatto riferimento al criterio di *matching* noto come sotto-classificazione che è stato sviluppato in linguaggio di programmazione SAS. E' possibile utilizzare altri criteri, come il *nearest-neighbor* che realizza un *matching* 1 a 1 tra i rispondenti appartenenti al gruppo di controllo e i rispondenti appartenenti al gruppo sottoposto al trattamento (disponibile nel pacchetto R - MatchIt). Nel pacchetto *R - Matching* sono implementati criteri basati sul *propensity score* e altri criteri di *matching* come, ad esempio, la distanza di Mahalanobis.

All'interno dei gruppi individuati, dopo aver verificato l'ipotesi di bilanciamento, ossia l'indipendenza tra la tecnica e ciascuna variabile ausiliaria, si è proceduto ad analizzare l'effetto tecnica per alcune variabili target dell'indagine.

Prima, per ciascun gruppo risultato bilanciato, è stato calcolato un peso secondo la formula 5.12, che rende equivalente la proporzione pesata dei rispondenti alla tecnica Cawi alla proporzione dei rispondenti alla tecnica Papi. In tal modo è stato poi possibile effettuare una valutazione complessiva dell'effetto selezione e dell'effetto misurazione per alcune variabili target dell'indagine utilizzando le espressioni riportate nel Capitolo 5.

Infine, per ciascuna modalità delle variabili categoriali considerate sono state calcolate le differenze tra le frequenze osservate per ciascuna tecnica e le frequenze attese in caso di indipendenza tra la tecnica e la modalità della variabile. Per verificare la significatività di tali differenze è stato utilizzato il test del χ^2 .

Il modello di analisi

Nel modello binomiale di tipo *logit*, la variabile indicatrice di risposta alla tecnica assume valore 1 se la famiglia ha risposto con tecnica Cawi e valore 0 se ha risposto con tecnica Papi.

A partire dai 37.495 individui rispondenti abbinati con la base dati Archimede si è giunti, raggruppando gli individui appartenenti ad una stessa famiglia, ad un data set composto da 15.821 record familiari (di cui 3.261 Cawi e 12.560 Papi).

Le variabili ausiliarie utilizzate nel modello, e definite a livello di famiglia, sono:

- Ripartizione geografica, 5 modalità: Nord-Ovest, Nord-Est, Centro, Sud, Isole;
- Tipologia comunale, 6 modalità: Città metropolitana, Area metropolitana, re-stanti comuni: sotto i 2000 abitanti, dai 2000 ai 10000 abitanti, dai 10000 ai 50000 abitanti, oltre i 50000 abitanti;
- Classe di reddito familiare, 5 modalità (basate sui quintili della distribuzione): <11955€, 11955-20982€, 20982-30028€, 30028-46119€, >46119€;
- Titolo di studio più alto in famiglia, 3 modalità: inferiore al diploma, diploma, oltre il diploma (di scuole media superiore);
- Tipologia occupazionale prevalente, 4 modalità: prevalenza di lavoratori dipendenti, prevalenza di lavoratori indipendenti, prevalenza di individui fuori dall'età lavorativa, famiglia mista;
- Cittadinanza prevalente, 3 modalità: Italiani, stranieri, famiglia mista;
- Tipologia familiare, 8 modalità: monocomponente fino a 54 anni, monocomponente con più di 54, coppia con figli con almeno un componente fino a 24 anni, coppia con figli con nemmeno un componente sotto i 25 anni, coppia senza figli, monoge-

6. Analisi dell'effetto tecnica in due casi di Studio

nitore con almeno un componente sotto i 25 anni, monogenitore con nemmeno un componente sotto i 25 anni, altra tipologia.⁶

Stima dei parametri e bontà di adattamento del modello

La definizione del modello di analisi ha richiesto una fase di studio al fine di ottenere un modello che massimizzasse la bontà di adattamento (*goodness of fit*) e, di conseguenza, la capacità predittiva dello stesso. Tale studio è stato svolto con una tecnica *backward-stepwise* che rimuove progressivamente le variabili inserite nel modello se non apportano un miglioramento allo stesso

Tavola 6.16 - Stima dei parametri del modello di regressione logistica

Variabile ausiliaria	Modalità	Stima	Std. Error	Pr>ChiQuadro	
(Intercetta)		-2.0584	0.0978	<.0001	***
	Nord Ovest	0.4518	0.0411	<.0001	***
Ripartizione territoriale	Nord-Est	0.4156	0.042	<.0001	***
	Centro	0.1914	0.0442	<.0001	***
	Sud	-0.5468	0.0467	<.0001	***
	Città metropolitana	0.2753	0.0559	<.0001	***
	Area metropolitana	0.1694	0.0621	0.0064	**
Tipologia comunale	Altri comuni sotto i 2000 ab.	-0.1858	0.0658	0.0047	**
	Altri comuni 2000-10000 ab.	-0.2053	0.0432	<.0001	***
	Altri comuni 10000-50000 ab.	-0.0318	0.0417	0.4461	
	coppia con figli con almeno un componente under 25	0.2231	0.0687	0.0012	**
	coppia con figli con nemmeno un componente under 25	0.045	0.0478	0.3464	
	coppia senza figli	-0.0209	0.0637	0.7426	
Tipologia familiare	monogenitore con almeno un componente under 25	0.2401	0.0511	<.0001	***
	monocomponente over 54	-0.2051	0.097	0.0345	*
	Altre tipologie	0.0861	0.0654	0.1885	
	monocomponente	-0.3726	0.1263	0.0032	**
Titolo di studio più alto in famiglia	Al di sotto del diploma	-0.5525	0.036	<.0001	***
	Diploma	0.0503	0.0306	0.1004	
	< 11.955 €	-0.3711	0.0612	<.0001	***
Classe di reddito familiare	11.955 - 20.892 €	-0.1745	0.0507	0.0006	***
	20.892 - 30.028 €	0.0459	0.0439	0.2954	
	30.028 - 46.119 €	0.1972	0.0418	<.0001	***
Cittadinanza prevalente	Famiglia mista	-0.0484	0.1573	0.7585	
	italiani	0.6625	0.0941	<.0001	***
	Dipendenti	0.1348	0.0365	0.0002	***
Tipologia occupazionale prevalente	Indipendenti	-0.0747	0.06	0.2133	
	Fuori dall'età lavorativa	-0.1784	0.0443	<.0001	***

Modalità di riferimento: isole, altri comuni con più di 50000 abitanti, monogenitore con nemmeno un componente under 25, oltre il diploma, >46.119 €, famiglia straniera, tipologia occupazionale mista

⁶ Si osserva che la variabile tipologia familiare non proviene da archivio bensì è una variabile ricostruita attraverso informazioni rilevate dall'indagine. Tale variabile è stata inserita nel modello poiché, da alcune evidenze empiriche, può essere ritenuta significativa sulla scelta della modalità di rilevazione da parte della famiglia.

in termini di bontà di adattamento del modello ai dati. Nessuna delle variabili ausiliarie sopra descritte è stata omessa dalla procedura poiché ogni covariata, di fatto, migliorava la capacità predittiva del modello.

La stima dei parametri di regressione, ottenuta col metodo di massima verosimiglianza, ha dato luogo ai coefficienti riportati nella Tavola 6.16.

Analizzando i coefficienti di regressione si nota come le modalità che influenzano maggiormente la propensione della famiglia a rispondere al Cawi sono: l'essere composta solo da italiani, avere un reddito alto, vivere in una città metropolitana (o nella area metropolitana) e vivere nel Nord Italia.

Il modello ha una buona capacità predittiva, infatti definendo le risposte come Cawi se la probabilità predetta è maggiore dello 0,5 si ottiene la seguente matrice di confusione (Tavola 6.17). Da tale matrice emerge che l'accuratezza del modello, ossia la proporzione di veri positivi e veri negativi (in grassetto) è alquanto elevata (80.13 per cento).

Tavola 6.17 - Matrice di confusione del modello per tecnica di rilevazione

Prob.osservate\ Prob. predette	Cawi	Papi	Totale
Cawi	1.51	18.56	20.08
Papi	1.30	78.62	79.92
Totale	2.81	97.18	100

Costruzione di gruppi omogenei e verifica dell'ipotesi di bilanciamento

A partire dalla distribuzione delle probabilità predette dal modello di risposta Cawi sono stati definiti i decili, al fine di creare classi o gruppi omogenei di unità (famiglie) rispetto alle variabili ausiliarie del modello.

Il metodo del *propensity score*, come visto, richiede che in ciascun gruppo generato sia verificata l'ipotesi di bilanciamento, ossia che ognuna delle variabili ausiliarie utilizzate nel modello sia indipendente dalla modalità di rilevazione Cawi/Papi. Questa ipotesi è stata valutata con il test d'ipotesi del Chi-quadrato. Nella Tavola 6.18 vengono riportati i valori del *p-value* associati all'ipotesi nulla d'indipendenza: *p-value* inferiori ad alfa (fissato a 0.01) indicano che l'ipotesi nulla d'indipendenza tra covariata e tecnica è stata rifiutata.

Tavola 6.18 - P-value relativi all'ipotesi nulla d'indipendenza tra covariata e tecnica

Gruppo	Tipologia comunale	Tipologia occupazionale prevalente	Tipologia familiare	Titolo di studio più alto in famiglia	Classe di Reddito familiare	Cittadinanza prevalente	Ripartizione territoriale	N
1	0.520	0.964	0.675	0.480	0.911	0.522	0.916	1461
2	0.132	0.042	0.626	0.624	0.291	0.910	0.077	1462
3	0.458	0.069	0.650	0.807	0.822	0.224	0.612	1469
4	0.728	0.039	0.463	0.101	0.437	0.715	0.639	1447
5	0.810	0.335	0.011	0.857	0.862	0.673	0.248	1453
6	0.475	0.107	0.348	0.690	0.358	0.767	0.113	1458
7	0.319	0.270	0.730	0.211	0.378	0.636	0.172	1458
8	0.193	0.059	0.195	0.671	0.815	0.595	0.299	1458
9	0.700	0.624	0.019	0.183	0.446	0.297	0.547	1460
10	0.029	0.156	0.260	0.887	0.523	0.289	0.061	1456
Totale	2.74E-25	1.15E-51	7.64E-28	2.60E-149	1.01E-136	1.19E-16	3.25E-96	14582

L'esito della verifica del bilanciamento è stato positivo in quanto in nessuno dei dieci gruppi individuati dai decili della distribuzione delle probabilità predette, l'ipotesi di indi-

6. Analisi dell'effetto tecnica in due casi di Studio

pendenza viene rifiutata. A titolo esplicativo è utile osservare come, al contrario, la stessa ipotesi viene rifiutata osservando il campione nella sua interezza (nell'ultima riga): questo indica che i gruppi sono stati creati correttamente, garantendo che le famiglie che si trovano all'interno dello stesso decile hanno una propensione alla risposta Cawi simile.

Analisi degli effetti selezione e misurazione

Come già anticipato, per ciascun gruppo bilanciato è stato calcolato un peso (formula 5.12) che di fatto costituisce un correttore dell'effetto selezione. Tale correttore consente di effettuare, nelle classi bilanciate, una valutazione complessiva dell'effetto tecnica. L'effetto selezione è ottenuto, seguendo le formule del Paragrafo 5.4.2, come differenza tra le stime pesate e non pesate dei rispondenti alla tecnica Cawi (formula 5.13), mentre l'effetto misurazione è ottenuto come differenza tra la stima pesata dei rispondenti alla tecnica Cawi e la stima non pesata dei rispondenti alla tecnica Papi (formula 5.14).

Nella Tavola 6.19 si riportano, per le modalità di alcune variabili target dell'indagine Ctl, le distribuzioni di frequenza delle variabili "Abitudine alla TV", "Uso del PC", "Accesso ad internet", ed i rispettivi effetti selezione e misurazione.

Tavola 6.19 - Effetto selezione e misurazione per alcune variabili d'indagine

Variabile	Frequenze relative			Effetto	
	Cawi	Cawi corretta	Papi	Selezione	Misurazione
Abitudine alla TV					
No	0,136	0,113	0,090	0,023	0,024
Tutti i giorni	0,693	0,666	0,725	0,028	-0,059
Talvolta	0,163	0,144	0,165	0,019	-0,021
Uso del PC					
Negli ultimi 3 mesi	0,698	0,537	0,497	0,161	0,040
Tra 3 mesi e un anno fa	0,028	0,030	0,029	-0,003	0,001
Più d'un anno fa	0,054	0,056	0,045	-0,002	0,011
Mai	0,213	0,299	0,401	-0,087	-0,101
Accesso a internet					
No	0,152	0,226	0,368	-0,074	-0,142
Si	0,848	0,703	0,632	0,145	0,071
SPORT					
Nell'ultimo mese	0,413	0,306	0,252	0,107	0,054
Da 2 a 6 mesi fa	0,072	0,061	0,057	0,012	0,004
Da 6 mesi a un anno fa	0,028	0,023	0,022	0,005	0,001
Più d'un anno fa	0,219	0,189	0,211	0,030	-0,023
Mai	0,268	0,352	0,458	-0,084	-0,106

La variabile d'indagine su cui sembra esserci un maggior impatto di entrambi gli effetti selezione (0.145) e misurazione (-0.071) è la variabile "Accesso ad internet". Notevoli sono anche gli effetti selezione relativi all'"Uso del PC" negli ultimi tre mesi (0.161) e all'attività sportiva nell'ultimo mese (0.107).

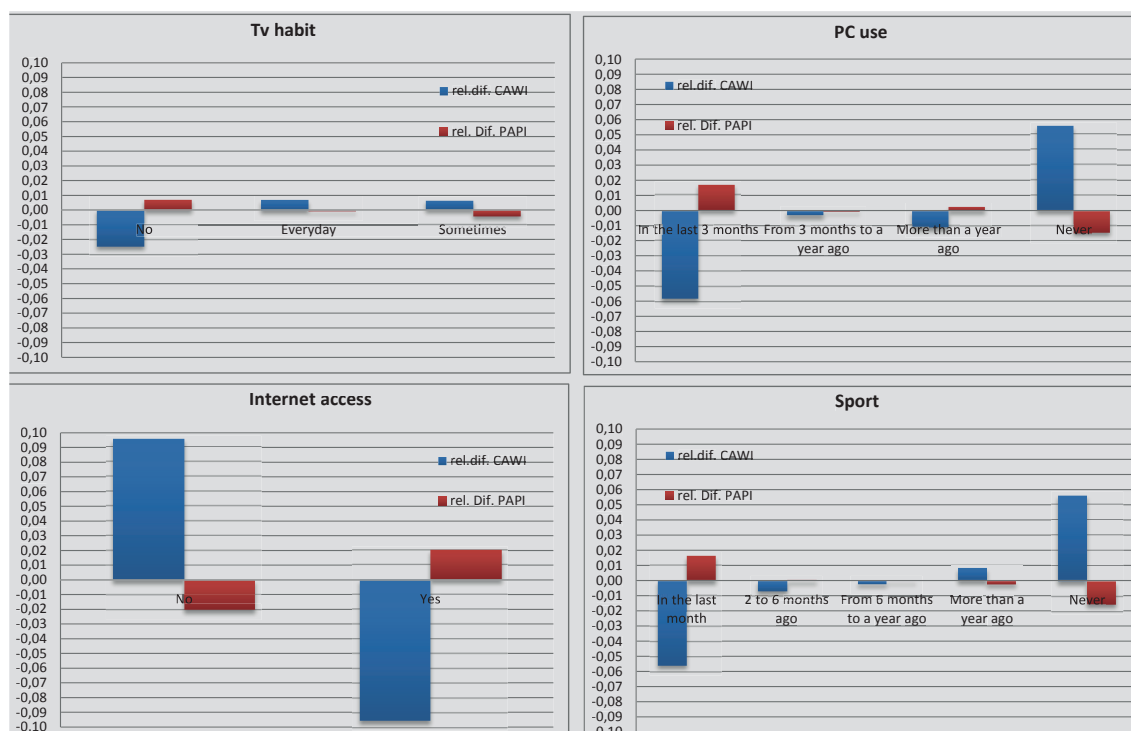
L'analisi della dipendenza della risposta dalla tecnica è stata effettuata confrontando le frequenze osservate e quelle attese, in caso di indipendenza in distribuzione, delle modalità di alcune variabili target dell'indagine. A tal fine sono state quindi calcolate, in ciascun gruppo bilanciato, le differenze percentuali tra le frequenze relative osservate e quelle attese nel

caso di indipendenza⁷ per le medesime variabili “Abitudine alla TV”, “Uso del PC”, “Accesso ad internet”, “Ultima volta che si è fatto sport”. Per sintetizzare i risultati ottenuti queste differenze sono state poi mediate sui dieci gruppi. Si riportano nei grafici della Figura 6.5. alcuni interessanti risultati.

Come si evince dai grafici, le maggiori differenze tra frequenze osservate e frequenze attese si ottengono per la variabile “Accesso ad internet” dei rispondenti Cawi (grigio scuro). Una situazione simile si ha quando si osservano le modalità estreme (“mai” e “negli ultimi 3 mesi”) della variabile “Uso del PC” e della variabile “sport”.

Tali risultati ci inducono a pensare che siano presenti per questa indagine degli errori di misura dovuti prevalentemente alla tecnica Cawi.

Figura 6.5 - Differenze percentuali tra le frequenze relative osservate e quelle attese di alcune variabili target per tecnica



6.3.7 L'effetto misurazione: il metodo basato sull'imputazione multipla

Come già detto nel Capitolo 5, nell'approccio basato sui modelli di imputazione multipla, l'errore di misura è trattato come un problema di mancata risposta. Le risposte fornite con una specifica tecnica sono considerate mancanti e convertite in risposte controfattuali. In genere, vengono prese come riferimento le risposte fornite con la tecnica prevalente o con la tecnica utilizzata nelle occasioni precedenti dell'indagine, mentre le risposte fornite con l'altra tecnica vengono imputate sulla base di un modello e, dunque, convertite in risposte controfattuali. Il punto di forza dell'imputazione multipla è che non viene generato un unico set di valori controfattuali, ma molteplici set. Questo consente di tenere in considerazione la variabilità aggiuntiva dovuta alla presenza di valori imputati e quindi originaria-

⁷ Data la generica frequenza assoluta osservata n_{ij} e quella attesa n_{ij}^* , la relativa frequenza relativa osservata è $f_{ij} = n_{ij} / \sum n_{ij}$ mentre la frequenza relativa attesa è $f_{ij}^* = n_{ij}^* / \sum n_{ij}^*$. La differenza relativa percentuale sarà dunque $(f_{ij} - f_{ij}^*) \cdot 100$.

6. Analisi dell'effetto tecnica in due casi di Studio

mente mancanti o considerati come tali.

Tutti i principali software statistici implementano metodi di imputazione multipla che, con alcuni semplici accorgimenti, possono essere applicati al caso *mixed-mode*. Nel linguaggio di programmazione R esistono diversi pacchetti con i quali è possibile svolgere l'imputazione multipla. Ad esempio, Mice (*Multivariate imputation via chained equations*) con il quale è possibile implementare il *Predictive mean matching* (Pmm) per variabili numeriche, la regressione logistica (logreg) per variabili dicotomiche, la regressione polinomica bayesiana (*polyreg*) per variabili categoriali non ordinate con più di due modalità e i modelli a *odds* proporzionali per variabili categoriali ordinate. Un altro pacchetto è Hmisc (*Harrell miscellaneous*), con il quale, attraverso il comando `aregImpute`, è possibile gestire l'*additive regression*, il bootstrap ed il Pmm. Un'ulteriore alternativa è rappresentata dal pacchetto MI (*Multiple Imputation con diagnostica*), che implementa modelli di imputazione multipla attraverso il metodo Pmm. In Stata (*Data Analysis and Statistical Software*), invece, usando il comando "mi" è possibile utilizzare diversi metodi di imputazione multipla, come normale multivariata, equazioni concatenate, regressione lineare, *predictive mean matching*, regressione troncata, modello logistico, *ordered logit* e *logit multinomiale*. Invece, in SAS – software utilizzato nell'applicazione successiva – vi è la *proc MI*. Attraverso la *proc MI* è possibile effettuare l'imputazione multipla utilizzando i metodi di imputazione più adatti alla natura della variabili che si stanno trattando.

Analisi preliminari per la definizione del modello di imputazione

Nel caso studio relativo ai dati dell'indagine Ctl, si considera come tecnica di riferimento la Papi. Le risposte Cawi, quindi, sono considerate come mancanti, imputate e convertite in risposte controfattuali per far emergere l'errore di misura.

I quesiti presi in considerazione sono "Ha praticato sport nell'ultimo mese?", "Guarda la tv abitualmente?", "Se avesse un po' più di tempo libero, a che cosa lo dedicherebbe?" e "Ha un accesso internet da casa?". Per tali quesiti, nella Tavola 6.20 sono riportate le distribuzioni delle risposte tra tecnica Papi e Cawi.

Tavola 6.20 - Distribuzione percentuale delle risposte tra tecnica Papi e Cawi per alcuni quesiti

Quesito	Papi	Cawi	TOTALE
Ha praticato sport nell'ultimo mese? (sport ultimo mese)	78,51	21,49	100,00
Guarda la tv abitualmente? (abitudine tv)	78,28	21,72	100,00
Se avesse un po' più di tempo libero, a che cosa lo dedicherebbe? (vorrei)	68,97	31,03	100,00
Ha un accesso internet da casa? (accesso ad internet)	67,17	32,83	100,00

Per migliorare la qualità dei valori controfattuali generati dall'imputazione multipla in genere si fa ricorso ad alcune variabili ausiliarie che hanno una relazione con le variabili di interesse o che si pensa possano essere associate alla tecnica. Nel nostro caso sono state considerate regione, classe d'età, stato civile, fonte principale di reddito, titolo di studio, cittadinanza, tipologia di lavoro. L'esperienza basata sulla conoscenza dei dati ed il ricorso alla letteratura possono aiutare ad identificare le variabili più adatte da inserire nel modello di imputazione. Ovviamente, queste variabili ausiliarie non devono essere affette da errori di misura. Per questo motivo, è molto importante effettuare una ricerca sulle variabili ausiliarie già in fase di progettazione dell'indagine in modo da poterne disporre in fase di stima.

Ad ogni modo, un'ulteriore verifica sulla bontà delle variabili ausiliare può essere svolta attraverso delle analisi secondarie, come ad esempio lo studio della correlazione – nel caso di variabili quantitative – o della connessione – in caso di variabili qualitative. Le variabili significative – generalmente con una correlazione >0.4 o con una connessione significativa – se inserite nel modello di imputazione aiutano, come già detto, a supportare l'assunzione MAR alla base dell'applicazione del metodo.

Nel nostro caso studio per semplicità le osservazioni che presentavano risposte mancanti sulle variabili ausiliarie non sono state incluse nell'analisi. Avendo tutte variabili qualitative abbiamo svolto un'analisi della connessione tra le variabili ausiliarie e le variabili di interesse (Tavola 6.21). Questa analisi mette in evidenza una relazione significativa della classe d'età con tutte le variabili di interesse considerate. La tipologia di lavoro, invece, non è connessa con la pratica dello sport nell'ultimo mese, mentre lo è con tutte le altre. L'accesso ad internet e la pratica dello sport nell'ultimo mese sono le sole connesse anche con la regione, mentre il sesso influenza solo l'attività a cui le persone si dedicherebbero nel caso in cui avessero più tempo libero. L'analisi fornisce un'indicazione sulle variabili più importanti da inserire nel modello di imputazione.

Un'altra analisi che ci può fornire indicazione importati sulle variabili da inserire nel modello di imputazione è relativa all'associazione tra variabili ausiliarie e tecnica (Papi/Cawi). L'applicazione di un modello logistico ha evidenziato che regione, classe d'età, stato civile e titolo di studio sono variabili significative per la tecnica. L'inserimento di queste variabili nel modello di imputazione, come detto sopra, può aiutare a supportare l'assunzione MAR alla base dell'applicazione del metodo.

Tavola 6.21 - Connessione tra le variabili ausiliarie e le variabili di interesse, V di Cramer

	Variabili ausiliarie									Variabili di interesse			
	regione	classe d'età	stato civile	fonte principale di reddito	sesto	titolo di studio	cittadinanza	tipologia di lavoro	sport ultimo mes	abitudine tv	vorrei	accesso ad internet	
Regioni	1,000	0,047	0,071	0,067	0,019	0,047	0,102	0,089	0,184	0,064	0,054	0,108	
Classe d'età	0,047	1,000	0,475	0,540	0,048	0,514	0,119	0,445	0,317	0,150	0,174	0,406	
Stato civile	0,071	0,475	1,000	0,269	0,183	0,186	0,294	0,177	0,282	0,092	0,106	0,250	
Fonte principale di reddito	0,067	0,540	0,269	1,000	0,142	0,255	0,125	0,900	0,236	0,132	0,114	0,367	
Sesso	0,019	0,048	0,183	0,142	1,000	0,082	0,000	0,127	0,000	0,038	0,324	0,000	
Titolo di studio	0,047	0,514	0,186	0,255	0,082	1,000	0,085	0,368	0,181	0,122	0,122	0,337	
Cittadinanza	0,102	0,119	0,294	0,125	0,000	0,085	1,000	0,061	0,023	0,061	0,075	0,011	
Tipologia di lavoro	0,089	0,445	0,177	0,900	0,127	0,368	0,061	1,000	0,069	0,117	0,149	0,217	
Sport ultimo mese	0,184	0,317	0,282	0,236	0,000	0,181	0,023	0,069	1,000	0,067	0,225	0,225	
Abitudine tv	0,064	0,150	0,092	0,132	0,038	0,122	0,061	0,117	0,067	1,000	0,087	0,073	
Vorrei	0,054	0,174	0,106	0,114	0,324	0,122	0,075	0,149	0,225	0,087	1,000	0,202	
Accesso ad internet	0,108	0,406	0,250	0,367	0,000	0,337	0,011	0,217	0,225	0,073	0,202	1,000	

Confronto tra stime

Una volta individuate le variabili da inserire nel modello di imputazione, quindi, si è proceduto all'imputazione multipla. Nella Tavola 6.22 sono riportate le stime ottenute con il modello ridotto – ovvero il modello con le sole variabili ausiliarie significative –, il modello

Tavola 6.22 - Confronto tra stime corrette con il modello di imputazione completo e ridotto, con il *propensity score* e stime non corrette

	Modalità	Modello Completo	Modello Ridotto	Stime non corrette	Propensity score
Ha praticato sport nell'ultimo mese?	Si	26,49	26,32	28,28	26,40
	No	73,51	73,68	71,72	73,60
	Totale	100,00	100,00	100,00	100,00
Guarda la tv abitualmente?	No	7,24	7,24	10,35	9,68
	Ogni giorno	79,70	79,70	73,01	73,48
	A volte	13,06	13,06	16,64	16,54
	Totale	100,00	100,00	100,00	100,00
Se avesse un po' di tempo libero, a cosa lo dedicherebbe?	Guardare la tv	3,89	3,89	4,66	5,05
	Andare al cinema	1,31	1,31	1,70	1,65
	Ascoltare musica	1,06	1,06	1,34	1,30
	Suonare o cantare	0,71	0,71	1,19	1,11
	Ballare	1,57	1,57	1,97	2,11
	Fare shopping	2,07	2,07	2,50	2,64
	Andare a teatro	0,52	0,52	0,79	0,72
	Dipingere, modellare, scolpire	0,65	0,65	1,12	1,02
	Leggere	3,25	3,25	5,32	4,79
	Scrivere	0,40	0,40	0,65	0,56
	Praticare uno sport	7,04	7,04	10,07	9,44
	Cucinare	2,52	2,52	3,24	3,42
	Praticare un hobby	7,13	7,13	9,43	9,26
	Navigare su internet	0,93	0,93	1,27	1,32
	Svolgere attività di volontariato	3,03	3,03	3,82	3,78
	Incontrare amici	27,08	27,08	15,41	15,54
	Conoscere nuove	0,069	0,650	0,807	0,822
	Persone	4,03	4,03	4,60	4,72
	Riposare	19,57	19,57	14,78	15,23
	Prendersi cura di sé	8,49	8,49	9,88	10,12
Altro	4,77	4,77	6,27	6,22	
	Totale	100,00	100,00	100,00	100,00
Ha un collegamento internet da casa?	Si	31,99	35,07	32,16	33,39
	No	68,01	64,93	67,84	63,69
	Totale	100,00	100,00	100,00	100,00

completo – con tutte le variabili ausiliarie disponibili –, le stime ottenute senza fare ricorso all'imputazione multipla e le stime ottenute utilizzando il correttore dell'effetto selezione basato sul *propensity score*. Si precisa che tutte le stime considerate sono calcolate senza utilizzare i pesi campionari.

Le stime ottenute con il modello completo ed il modello ridotto sono molto vicine tra loro, segno che le variabili più significative sono sufficienti a spiegare il modello strutturale. Fanno eccezione le stime per il quesito "Ha un accesso internet da casa?".

Inoltre, le stime ottenute con l'imputazione multipla – che ricordiamo assume che il meccanismo di selezione sia di tipo MAR – sono molto vicine a quelle ottenute con il *propensity score*. Questo è il caso del quesito "Ha praticato sport nell'ultimo mese?".

Nel caso di quesiti con più modalità di risposta le differenze tra i due metodi sono maggiori. In particolare, tali differenze sono più evidenti sul quesito "Guarda la tv abitualmente?" e "Ha accesso ad internet da casa?", ciò sta ad indicare che è possibile che vi sia un effetto misurazione delle due tecniche.

6.3.8 Conclusioni e ulteriori sviluppi

I risultati riportati per il caso studio dell'indagine Ctl sono relativi alle prime sperimentazioni condotte sul problema dell'effetto tecnica. Di conseguenza tali analisi necessitano di essere supportate da ulteriori approfondimenti sia per tenere conto della variabilità campionaria sia per sottoporre a verifica le assunzioni alla base dei metodi adottati.

In particolare gli effetti selezione e misurazione dovrebbero essere analizzati in un contesto simulativo per ottenere una valutazione della significatività delle stime ottenute.

In generale, i metodi in grado di tener conto anche della non ignorabilità dell'effetto selezione – come ad esempio il metodo di imputazione multipla frazionata (cfr. Capitolo 5) sarebbero da preferire a quelli che assumono l'assenza di questa. In molti casi reali, l'assunzione d'ignorabilità dell'effetto selezione è molto forte e può portare a non correggere la distorsione, o peggio a correggerla in maniera errata.

7. SINTESI E RACCOMANDAZIONI GENERALI¹

7.1 Sintesi del volume

Il presente volume fornisce un quadro concettuale e definitorio di riferimento nel contesto dell'uso della multi-tecnica nelle rilevazioni statistiche su famiglie e individui, e introduce una terminologia comune in questo ambito (Capitolo 2). Sulla base di questo *framework*, i vari Capitoli analizzano le problematiche connesse ai vari aspetti del disegno di strategie *mixed-mode*, evidenziando sia i vantaggi e i rischi connessi al loro utilizzo, sia l'importanza della fase di progettazione come strumento di prevenzione dei potenziali effetti distorsivi sui risultati finali di questo tipo di strategie.

Nel Capitolo 3, in particolare, sono discussi i vari aspetti connessi al disegno del questionario e, più in generale, di strategie a tecnica mista, con riferimenti anche alle strategie di massimizzazione dei tassi di risposta e ai metodi di prevenzione dell'effetto tecnica tramite pre-test e indagini pilota.

Con riferimento al problema della valutazione e del trattamento dell'effetto tecnica, le possibili soluzioni metodologiche sono state illustrate seguendo un criterio di progressiva complessità (Capitoli 4 e 5): a partire da analisi di tipo esplorativo, sono stati proposti possibili strumenti inferenziali per ottenere evidenze sulla presenza di *mode effects*, illustrando poi alcune soluzioni metodologiche più complesse basate sull'uso di modelli e metodi di stima per la diagnosi e il trattamento dell'effetto tecnica. I risultati dell'applicazione sperimentale di tali metodologie a dati di indagini sociali dell'Istat (Capitolo 6) hanno consentito di approfondire la conoscenza da un lato delle metodologie utilizzate, dall'altro della complessità del loro utilizzo in situazioni reali in cui sia opportuno trattare le distorsioni derivanti dal ricorso alla tecnica mista di rilevazione.

Tuttavia, quelle presentate in questo volume sono le prime sperimentazioni condotte sul problema dell'effetto tecnica in indagini sociali dell'Istat. In termini generali, l'obiettivo delle applicazioni realizzate è stato non tanto quello di poter trarre conclusioni sulla base dei risultati ottenuti, quanto quello di evidenziare, da un lato, la necessità di progettare le indagini tenendo conto delle problematiche connesse all'uso di tecniche miste, dall'altro la complessità delle analisi sull'effetto tecnica. Tali analisi, infatti, molto complesse dal punto di vista metodologico e molto onerose in termini di tempo, richiedono il reperimento e l'utilizzo di consistenti informazioni ausiliarie. Nelle specifiche esperienze condotte, elementi vincolanti sono stati infatti sia le caratteristiche dei dati utilizzati e quindi le stesse variabili sottoposte ad analisi, sia la disponibilità e la qualità delle informazioni ausiliarie utilizzate. Inoltre, trattandosi di applicazioni realizzate in contesti non sperimentali, i risultati ottenuti dipendono da fattori non completamente controllabili dagli statistici, oltre che dalle scelte e dalle assunzioni fatte, la cui validità necessita di essere supportata da ulteriori studi e approfondimenti.

¹ Il capitolo è stato redatto da Orietta Luzi.

7.2 Alcune raccomandazioni generali

Alla luce di quanto esposto nei vari Capitoli di questo volume, sia sul piano teorico sia su quello applicativo, è dunque possibile estrarre alcune raccomandazioni di carattere generale su cui porre particolare attenzione in fase di pianificazione e realizzazione di processi di indagine di tipo *mixed-mode*. Di seguito sono riportate le principali.

- Progettazione

Come sottolineato nel Capitolo 2: «*La fase di raccolta dei dati è cruciale per la qualità. Attraverso un'accurata progettazione e un attento monitoraggio di tale fase si può contribuire in misura significativa a prevenire o ridurre diverse tipologie di errore, in particolare le mancate risposte totali e gli errori di misurazione*».

- La scelta delle singole tecniche, da usare in modo concorrente o sequenziale, si basa sulla necessità di massimizzare la qualità in presenza di vincoli di costo e di tempo e di contenere il carico statistico sui rispondenti (cfr. Capitolo 3). La scelta può essere guidata dal tipo di errore non campionario sulla cui riduzione si intende agire in maniera prioritaria (ad es. si può decidere di contenere l'effetto selezione riducendo i potenziali errori di copertura o di aumentare i tassi di risposta).
- Il problema del disegno riguarda non solo la scelta delle tecniche, ma anche altri elementi fondamentali quali: 1) preliminarmente alla raccolta dei dati sul campo, la scelta delle modalità con cui contattare i rispondenti (contatto iniziale, appuntamenti per l'intervista, lettera di presentazione dell'indagine, etc.) e, 2) successivamente alla raccolta dati, le modalità di sollecito per i non rispondenti, di re-intervista su un sotto-campione, di verifica dell'effettuazione dell'intervista da parte del rilevatore, etc. (cfr. Capitolo 3).
- Al fine di ridurre errori di misura dovuti alla tecnica mista, qualora i costi lo permettano, una strategia raccomandabile è quella di agire in maniera 'preventiva', raccogliendo informazioni attraverso pre-test o indagini pilota (cfr. Paragrafo 3.5). L'indagine pilota consente di verificare tutto l'impianto della rilevazione (aspetti organizzativi, scelta delle tecniche, stima dei tassi di risposta per tecnica, strategie di sollecito e di sensibilizzazione dei rispondenti etc.).
- In particolare, strumenti come pre-test o indagini pilota rappresentano strumenti fondamentali per la progettazione ottimale del questionario. Nelle indagini *mixed-mode*, infatti, è cruciale prevedere le strutture, il wording, etc. più appropriati per le differenti tecniche di raccolta/tipologie di rispondenti. Le diverse opzioni (questionario adattato al tipo di tecnica oppure unico questionario a prescindere dalla tecnica) presentano vantaggi e svantaggi in termini di qualità e aspetti organizzativi di cui è necessario tener conto in fase di progettazione (cfr. Paragrafo 3.3).
- Nella progettazione di un'indagine a tecnica mista la definizione del disegno campionario è strettamente connessa con la scelta delle tecniche di rilevazione. Elementi chiave in questo caso sono (cfr. Paragrafo 3.2): la tipologia di lista di selezione (ad es. lista unica o distinta per tecnica); la modalità di utilizzo delle tecniche di rilevazione, concorrente o sequenziale, che influenza il tipo di relazione tra i campioni (dipendenti o indipendenti) relativi alle differenti tecniche.
- Metodologie per la diagnostica e il trattamento del mode effect

Come sottolineato nel Capitolo 5: «L'obiettivo principale di un'indagine statistica è ottenere stime accurate dei parametri di interesse. Nelle indagini multi-tecnica le stime sono il risultato della combinazione di dati raccolti con differenti strumenti di rilevazione. In questo caso, l'accuratezza del processo inferenziale è assicurata, oltre che da altri numerosi fattori - copertura, mancata risposta totale, stimatore - soltanto se la misurazione effettuata con più tecniche è equivalente...».

- La realizzazione di un'indagine *mixed-mode* deve tener conto già in fase di disegno degli esiti dell'applicazione di metodologie esplorative e inferenziali su indagini analoghe (per popolazione target, mix di tecniche, etc.), in modo da adottare misure preventive che consentano di contenere gli errori di selezione e misura.
- A conclusione della raccolta dati, deve essere condotta una prima analisi esplorativa (cfr. Paragrafi 4.3.1 e 4.3.2.) al fine di avere una visione complessiva sulla qualità dei dati raccolti e trarre indicazioni per l'applicazione di metodologie di analisi più complesse, basate sull'utilizzo di modelli statistici.
- La scelta del modello statistico per la valutazione dell'effetto selezione e misurazione (cfr. Paragrafo 4.3.3) è legato principalmente al tipo di disegno utilizzato (sequenziale o meno), alla presenza di quesiti con più item, alla disponibilità di un'indagine di riferimento che possa essere considerata priva di errori.
- La selezione delle variabili da trattare deve essere guidata da analisi preliminari su un sottoinsieme di item considerati potenzialmente più sensibili all'effetto tecnica. L'obiettivo di tali analisi è avere informazioni sulla significatività statistica delle differenze fra le misurazioni ottenute con le differenti tecniche utilizzate, scorporate da possibili effetti selezione (cfr. Capitolo 4).
- La scelta dei metodi per il trattamento dell'effetto tecnica (cfr. Capitolo 5) dipende dai risultati delle analisi condotte sulla presenza degli effetti selezione e misurazione. Infatti, in assenza di effetto misurazione è possibile intervenire sui pesi campionari per correggere l'effetto selezione. In presenza, invece di effetto misurazione, sono possibili correzioni per singole stime, anche se in generale tale soluzione non è auspicabile per le indagini multi-obiettivo. Infine, la scelta dei metodi di correzione degli effetti selezione e misurazione è condizionata dalla validità delle assunzioni che ne sono alla base.
- Un ruolo chiave nell'ambito delle metodologie per la valutazione e il trattamento dell'effetto tecnica è rivestito dalla scelta delle variabili ausiliarie da utilizzare nei modelli. Tali variabili devono essere mode-insensitive. E' importante individuare le variabili ausiliarie già in fase di progettazione dell'indagine in modo da poterne disporre in fase di analisi e/o di stima.
- Nelle indagini a tecnica mista il trattamento della mancata risposta totale è un'operazione complessa soprattutto nei disegni di tipo sequenziale (cfr. Paragrafo 5.5). In questo caso, infatti, gli effetti introdotti dalla tecnica mista devono essere trattati insieme all'effetto selezione determinato dalla mancata risposta totale.

CONCLUSIONI E PROSPETTIVE FUTURE¹

Questo volume è il risultato di una vasta e complessa attività di ricerca e sperimentazione condotta dall'Istat, nell'ambito della Direzione centrale per la raccolta dati dell'Istat (Dcird), realizzata grazie alle competenze, conoscenze ed esperienze maturate in differenti ambiti dei settori di produzione e metodologici.

Le attività realizzate, illustrate in questo volume, sono maturate nel nuovo contesto organizzativo e funzionale dell'Istituto nazionale di statistica, determinato dal processo di modernizzazione realizzato nel 2016, in cui tutte le attività di raccolta delle informazioni statistiche (provenienti sia da rilevazioni dirette, sia da fonti secondarie quali fonti amministrative e *big data*) risultano "accentrate" in un singolo polo, rappresentato appunto dalla Dcird. L'obiettivo principale è, sul versante delle indagini, garantire una maggiore efficienza della fase di acquisizione dei dati presso le unità statistiche, eliminando le ridondanze e riducendo i costi di progettazione e implementazione delle procedure di *data collection*. Sul versante delle fonti amministrative, la centralizzazione degli archivi di base presso la Dcird garantisce un maggior controllo, livello di standardizzazione dei processi, e grado di efficienza nella gestione delle informazioni in essi contenute e destinate ai settori di produzione.

Con specifico riferimento all'area delle statistiche demo-sociali, questo nuovo contesto informativo non solo risulta funzionale al (ri)disegno integrato delle principali indagini condotte dall'Istat in questo ambito della Statistica ufficiale, ma anche alla progettazione di processi di raccolta dati con un migliore *trade-off* fra costi e qualità dell'informazione statistica. In relazione a questo ultimo aspetto, è relativamente recente l'attenzione posta dai ricercatori dell'Istat alla progettazione di strategie di raccolta dati di tipo tecnica mista, che implicano lo sviluppo di nuove strategie di campionamento, di disegno dei questionari, di gestione operativa e tecnica del processo di raccolta dei dati per la prevenzione e il trattamento degli effetti distorsivi dovuti all'uso della tecnica mista. E' in questa nuova ottica che sono stati progettati, ad esempio, il *master sample* alla base del prossimo Censimento della popolazione italiana, nonché le principali rilevazioni *mixed-mode* ad esso collegate (Forze di lavoro, Spese delle famiglie, Eu-Silc, Multiscopo sulle famiglie).

La realizzazione di tali progetti ha evidenziato ulteriormente la necessità di avviare, internamente all'Istituto, un processo di condivisione delle problematiche generali connesse all'uso del *mixed-mode*, e lo studio di possibili soluzioni al problema della prevenzione e del trattamento del cosiddetto effetto tecnica. In questo volume si fa riferimento, in particolare, alle strategie *mixed-mode* che prevedono l'uso congiunto di tecniche di rilevazione quali Cati o Papi e del web, una tecnica efficiente in termini di costi che, grazie alla progressiva diffusione di Internet nelle famiglie italiane, in particolari condizioni può contribuire alla riduzione dei tassi di non risposta ed all'incremento del tasso di copertura delle indagini.

Il principale risultato atteso dalla realizzazione di questo volume non è soltanto la condivisione di definizioni, concetti e problematiche nell'ambito della tecnica mista per le indagini demo-sociali, ma anche e soprattutto l'avvio di una riflessione trasversale ai vari settori dell'Istituto sulle tematiche metodologiche e progettuali trattate.

La prospettiva è quella di ulteriori approfondimenti di tali tematiche, volti alla verifi-

¹ Il contributo è stato redatto da Orietta Luzi.

ca dell'efficacia di queste e di nuove soluzioni in contesti applicativi progettati con il fine specifico di poter prevenire e/o trattare adeguatamente l'effetto tecnica, come appunto nel caso del disegno del *master sample* e in quello della rilevazione Aspetti della vita quotidiana (anno 2017).

Inoltre, grazie all'avvio di un nuovo progetto europeo sul tema del *mixed-mode data collection* (denominato "Mixed MOde Designs for Social Surveys" – Mi-mod), coordinato dall'Istat, sarà portato avanti un ampio e approfondito confronto a livello Europeo, già avviato nell'ambito dell'*Essnet on data collection for social surveys using multi-modes* (Dcss) (Eurostat, 2014).

Ringraziamenti

Si ringrazia il Dott. Saverio Gazzelloni per aver promosso e supportato il progetto che ha portato alla realizzazione di questo volume, e per i preziosi suggerimenti forniti su struttura e contenuti della pubblicazione.

Si ringraziano il Dott. Piero Demetrio Falorsi e il Dott. Marcello D'Orazio per la revisione, i suggerimenti e i commenti forniti sia sugli aspetti metodologici sia sulle applicazioni sperimentali presentati nel volume.

RIFERIMENTI BIBLIOGRAFICI

- AA.VV. 2017. *Journal of Official Statistics*. Volume 33, Issue 3 (Sep 2017). <https://www.degruyter.com/view/j/jos.2017.33.issue-3/issue-files/jos.2017.33.issue-3.xml>.
- Alexander, C.H., Jr e S. Wetrogan. 2000. "Integrating the American Community Survey and the Intercensal Demographic Estimates Program." Proceedings of the American Statistical Association. www.amstat.org/sections/srms/proceedings/papers/2000_046.pdf.
- Angrist, J. D., Imbens, G. W., e D. B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91, 444–455. <http://www.doi.org/10.1080/01621459.1996.10476902>.
- Austin, P. C. 2011. "An introduction to propensity score methods for reducing the effects of confounding in observational studies." *Multivariate Behavioral Research* 46, 399-424. Taylor & Francis Group, LLC.
- Bethlehem, J., e S. Biffignandi 2012. *Handbook of web Surveys*. Wiley Handbooks in Survey Methodology: Wiley.
- Bethlehem, J., Cobben, F., e B. Schouten. 2009. "Indicators for the Representativeness of Survey Response." Proceedings of Statistics Canada Symposium 2008 Data Collection: Challenges, Achievements and New Directions. Ottawa.
- Betts, P., e C. Lound. 2010. "ONS-The application of alternative modes of data collection on UK Government social surveys." *Literature review and consultation with National Statistical Institutes*.
- Biemer, P. B. 2001. "Non-response bias and measurement bias in a comparison of face to face and telephone interviewing." *Journal of Official Statistics* 17 (2), 295-320.
- Biemer, P. 2004. *Modeling Measurement error to Identify Flawed Questions*. In "Methods for testing and evaluating survey questionnaires." Presser S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin and E. Singer (eds.), Wiley Series in Survey Methodology: Wiley.
- Biemer, P., De Leeuw, E., Eckman, S., Edwards B., e F. Kreuter. 2017. *Total survey error in practice*. Wiley & Sons Ltd. (Estados Unidos).
- Biemer, P. P., e L.E. Lyberg. 2003. *Introduction to Survey Quality*. New York: John Wiley.
- Bowden, R.J., e D. A. Turkington. 1990. *Instrumental Variables*. Cambridge: Cambridge University Press.
- Brait, F., Orsini, S., Quattrociochi, L., Strozza, M., Bologna, E., Montecolle, S., Ioppolo, L., Joffre, V., Del Bufalo, E., e C. Facioni. 2017. *L'introduzione della tecnica Cawi nelle indagini multiscopo sulle famiglie*. In "L'utilizzo della tecnica Cawi nelle indagini su individui e famiglie." *Collana: Letture statistiche–Metodi*. <http://www.istat.it/it/archivio/203729>.
- Brancato, G., Macchia, S., Murgia, M., Signore, M., Simeoni, G., Blanke, K., Körner, T., Nimmergut, Lima, P., Paulino, R., e J. H. P. Hoffmeyer-Zlotnik. 2006. *Handbook of Recommended Practices for questionnaire development and testing in the European statistical system*. <http://ec.europa.eu/eurostat/web/quality/ess-practices/selected-ess-practices>.
- Buelens, B., e J. A. Van den Brakel. 2017. "Comparing Two Inferential Approaches to Handling Measurement Error in Mixed-Mode Surveys." *Journal of Official Statistics* 33(2), 513-531.
- Buelens, B., e J. Van den Brakel. 2013. "Measurement error calibration in mixed-mode sample surveys." *Discussion paper. Statistics Netherlands. The Hague/Heerlen, 2013*.
- Callegaro, M., Lozar Manfred, K., e V. Vehohar. 2015. *Web survey methodology*. Los Angeles: Sage.
- Cobben, F., Schouten, B., e J. Bethlehem. 2006. "A model for statistical inference based on mixed mode interviewing." Proceedings of Q2006 – European conference on Quality in Survey Statistics, Cardiff, Regno Unito, 24-26 Aprile.

- Cobben, F. 2009. "Nonresponse Adjustment in Mixed Mode Surveys." In: *Nonresponse in Sample Surveys Methods for Analysis and Adjustment*. Statistics Netherlands.
- Comley, P. 2000. "Pop-up surveys: What works, what doesn't work and what will work in the future." In *Proceedings of the ESOMAR Worldwide Internet Conference: Net Effects 3*. ESOMAR (Ed.), Amsterdam: ESOMAR.
- Conover, W. J. 1999. *Practical Nonparametric Statistics*. Third Edition, New York: John Wiley & Sons.
- Couper, M. P. 2008. *Designing effective web surveys*. New York: Cambridge University Press.
- Couper, M. P. 2011. "The Future of Modes of Data Collection." *Public Opinion Quarterly* 75, 889-908. <http://dx.doi.org/10.1093/poq/nfr046>.
- Couper, M. P. 1998. "Measuring Survey Quality in a CASIC Environment." *Proceedings of the Survey Research Methods Section*, 41-49. <http://www.amstat.org/sections/srms/Proceedings/>
- Couper, M. P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J. M., e E. Singer. 2004. (eds.) *Methods for testing and evaluating survey questionnaires*. Wiley Series in Survey Methodologies.
- D'Agostino, R. B. 1998. "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group." *Statistics Medicine* 17(19), 2265-2281.
- de Leeuw, E. D. 1992. *Data Quality in Mail, Telephone, and Face to Face surveys*. TT-Publications Amsterdam.
- de Leeuw, E. 2005. To mix or not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* 21(2), 233-55.
- de Leeuw, E. D., e W. de Heer. 2002. *Trends in household survey nonresponse: A longitudinal and international comparison*. In "Survey nonresponse." Groves, R. M., Dillman, D. A., Eltinge, J. L., e R. J. A. Little (eds.) New York: Wiley.
- Della Rocca, G., Luzi, O., Scavalli, E., Signore, M., e G. Simeoni. 2003. "Evaluating, monitoring and documenting the effects of editing and imputation in ISTAT surveys." *UN/ECE Work Session on Statistical Data Editing*, Madrid, Spagna, 20-22 Ottobre.
- De Vitiis, C., Falorsi, P.D., Falorsi, S., e A. Russo. 2002. "Un'analisi comparativa di alcuni metodi di trattamento della mancata risposta totale nella stima delle variazioni lorde nel campionamento ruotato." *Collana Convegni, del Dipartimento di Statistica, Probabilità e Statistiche Applicate*, Ed. Università degli Studi di Roma La Sapienza.
- De Vitiis, C., Cocchi, D., Inglese, F., e M. D. Terribili. 2012. "Treatment of total nonresponse via sequential weight adjustment in the Italian disability survey." *Italian Journal of Applied Statistics, Special Issue* 24(1), 75-101.
- De Vitiis, C., Guandalini, A., Inglese, F., e M. D. Terribili. 2016. "Metodi per il trattamento delle diverse componenti della mancata risposta totale applicati all'indagine Istat sulla Disabilità." In corso di pubblicazione sulla *Rivista di Statistica Ufficiale Istat*.
- Dillman, D.A. 2000. *Mail and Internet Surveys - The Tailored Design Method* (2nd Edition). John Wiley and Sons: New York.
- Dillman, D.A., e L. Carley-Baxter. 2000. "Structural determinates of mail survey response rates over a 12 year period, 1988-1999." *Proceedings of American Statistical Association survey methods section*. Alexandria. Va: 344-349.
- Dillman D. A., e L. M. Christian. 2005. "Survey Mode as a Source of Instability in Responses across Surveys." *Fields method* 17(1), 30-52.
- Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., e B.L. Messer. 2009. "Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the internet." *Social Science Research* 38 (1), 1-18.
- Dillman, D.A., Smyth, J., e L. Christian. 2009. *Internet, Mail and Mixed-Mode Surveys - The Tailored Design Method* (3rd Edition), John Wiley and Sons: New Jersey.
- Edwards, P. J., Roberts, I., Clarke, M. J., Diguiseppi, C., Wentz, R., Kwan, I., e S. Pratap. 2009. "Methods to increase response to postal and electronic questionnaires." *Cochrane Database of Systematic Reviews* (8)3.

- Eurostat. 2014. *Essnet on data collection for social surveys using multi-modes (Dcss)*. https://ec.europa.eu/eurostat/cros/content/data-collection_en.
- Fazzi, G., Rosati, S., e M.C. Romano. 2017. "Studying a sensitive issue with a structured questionnaire: wording and mode effect in sexual orientation questions." Workshop *QUESTIONNAIRE EVALUATION STANDARD* (QUEST). Roma, 4-6 Aprile.
- Fisher, R. A. 1970. *Statistical Methods for Research Workers*. Fourteenth Edition, Davien, CT: Hafner Publishing Company.
- Freeman, G. H., e J. H. Halton. 1951. "Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance." *Biometrika* 38, 141–149.
- Gallo, F. 2017. *La tecnica mista Cawi/Cati: l'esperienza delle indagini sui laureati e sui diplomati*. In "L'utilizzo della tecnica Cawi nelle indagini su individui e famiglie." Istat, Collana: Letture statistiche – Metodi. <http://www.istat.it/it/archivio/203729>.
- Gallagher, P. M., Fowler, F. J., e L.V. Stringfellow. 2005. "The nature of nonresponse in a Medicaid survey: causes and consequences." *Journal of Official Statistics* 21(1), 73–87.
- Garofalo, G. 2014. "Il Progetto ARCHIMEDE obiettivi e risultati sperimentali." *Istat Working Papers* 9/2014. <https://www.istat.it/it/files/2014/11/IWP-n.-9-2014.pdf>.
- Görizt, A. S. 2008. "The long-term effect of material incentives on participation in online panels." *Field Methods* 20(3), 211 – 225 .
- Grande, E., e O. Luzi. 2003. "Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat." *Contributi Istat* 11/2004. http://www3.istat.it/dati/pubbsci/contributi/Contributi/contr_2003/2003_6.doc
- Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley.
- Groves, R. M. 2006. "Nonresponse rates and nonresponse bias in household surveys." *Public Opinion Quarterly* 70(5), 646-675.
- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., e R. Tourangeau. 2004. *Survey methodology*. Hoboken, NJ: Wiley.
- Groves, R. M., e S.G. Heeringa. 2006. "Responsive design for household surveys: tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society Series A* 169, 439 – 457.
- Groves, R. M., e L. Lyberg. 1988. *An Overview of Nonresponse Issues in Telephone Surveys*. In "Telephone Survey Methodology." Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls II, W.L., e J.Waksberg (eds.), New York: John Wiley.
- Groves, R., e L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74(5), 849–879. <http://dx.doi.org/10.1093/poq/nfq065>.
- Heerwegh, D., e G. Loosveldt. 2006. "An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys." *Journal of Official Statistics* 22 (2), 191 – 210 .
- Heckman, J. J. 1979. "Sample selection bias as specification error." *Econometrica* 47, 153-161.
- Heckman, J. J. 1997. "Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations." *The Journal of Human Resources* 32(3), 441–462.
- Heerwegh, D. 2009. "Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects." *International Journal of Public Opinion Research* 21(1), 111-121.
- Heerwegh, D., e G. Loosveldt. 2011. "Assessing mode effects in a national crime victimization survey using structural equation models: social desirability bias and acquiescence." *Journal of official statistics* 27(1), 49-63.
- Hochstim, J. R. 1967. "A Critical Comparison of Three Strategies of Collecting Data from Households." *Journal of the American Statistical Association* 62, 976–989.
- Holbrook, A. L., Green, M., e J. Krosnick. 2003. "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires." *Public Opinion Quarterly* LXVII: 79-125.

- Holbrook, A. L., Krosnick, J. A., e A. Pfent. 2007. *The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms*. In "Advances in Telephone Survey Methodology." Lepkowski, J. M., Tucker, C., Brick, J. M., de Leeuw, E. D., Japac, L., Lavrakas, P. J., Link, M. W., e R. L. Sangster (eds.), John Wiley & Sons, Inc., Hoboken, NJ, USA. Ch. 23, 499-528.
- Hox, J., de Leeuw, E. D., e T. Klausch. 2015. "Mixed Mode Research: Issues in Design and Analysis." Invited paper presented at the International Conference on Total survey error: improving quality in the era of big data. Baltimore, 19-22 September. https://www.researchgate.net/publication/312054656_Mixed_Mode_Research_Issues_in_Design_and_Analysis_HOX-TSE15.
- Hox, J. J., de Leeuw, E. D., e T. Klausch. 2017. *Mixed-Mode Research*. In "Total Survey Error in Practice", Biemer, P.B., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., and B.T. West (eds.), ISBN: 978-1-119-04167-2.
- Hox, J. J., de Leeuw E. D., e E. A. O. Zijlmans. 2015. "Measurement equivalence in mixed mode surveys." *Frontiers in psychology* 6, 1-11.
- Kaplowitz, M. D., Lupi, F., Couper, M. P., e L. Thorp. 2012. "The effect of invitation design on web survey response rates." *Social Science Computer Review* 30 (3), 339 – 349.
- Kim, J. K. 2011. "Parametric fractional imputation for missing data analysis." *Biometrika* 98, 119-132.
- Klausch, T. 2014. *Informed Design of Mixed-Mode Surveys - Evaluating mode effects on measurement and selection error*. Ipskamp Drukkers, Enschede.
- Klausch, T., Hox, J., e B. Shouten. 2015. "Selection error in single- and mixed mode surveys of the Dutch general population." *Journal of the Royal Statistical Society, Statistics in Society, Series A* 178(4), 945–961.
- Kolenicov, S., e C. Kennedy. 2014. "Evaluating three approaches to statistically adjust for mode effects." *Journal of Survey Statistics and Methodology* 2, 126–158.
- Kreuter, F. 2013. "Facing the Nonresponse Challenge." *The Annals of the American Academy of Political and Social Science* 645, 23–35. <https://doi.org/10.1177/0002716212456815>.
- Kreuter, F., Presser, S., e R. Tourangeau. 2008. "Social desirability bias in Cati, IVR and web surveys." *Public Opinion Quarterly* 72(5), 847-865.
- Krosnick, J. A. 1999. "Survey research." *Annual Review of Psychology* 50, 537-367. Annual Reviews.
- Illemann Christensen A., Ekholm, O., Glümer, C., e K. Juel. 2014. "Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys." *European Journal of Public Health* 24(2), 327-332.
- Istat. 2012. *Linee guida per la qualità dei processi statistici*. Roma. https://www.istat.it/it/files/2013/03/Linee_Guida_Qualita_1_1_IT1.pdf.
- Istat. 2013. *Nota metodologica Indagine sull'inserimento professionale dei dottori di ricerca 2009*. <https://www.istat.it/microdata/download.php?ref=https://www.istat.it/it/archivio/87536&id=/wwwarmida/206/2014/01/Nota.pdf>.
- Istat. 2017. "L'utilizzo della tecnica Cawi nelle indagini su individui e famiglie." *Lecture statistiche – Metodi*. Roma. <https://www.istat.it/it/files/2017/09/Lutilizzo-della-tecnica-Cawi.pdf>.
- Jackle, A., Roberts, C., e P. Lynn. 2010. "Assessing the Effect of Data Collection Mode on Measurement." *International Statistical Review* 78, 3–20.
- Jäckle, A., Sala, E., Jenkins, S.P., e P. Lynn. 2004. "Validation of survey data on income and employment: the ISMIE experience." *ISER Working Paper 2004-14*. University of Essex, Colchester. <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-14.pdf>.
- Järvensivu, M. 2015. "On questionnaire design and pre-testing in mixed mode." Workshop QUestionnaire Evaluation Standard (QUEST 2015). Helsinki, 28-30 Aprile.
- Lee, S. 2006. "Propensity score adjustment as a weighting scheme for volunteer panel web surveys." *Journal of Official Statistics* 22 (2), 329-349.

- Lee, S., e R. Valliant. 2008. *Post-Survey Weighting Methods Using Propensity Scores: A Review*. In "Advances in Telephone Survey Methodology", John Wiley.
- Lehmann, E. L. 1998. *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
- Link, M. W., e A. H. Mokdad. 2005. "Effects of Survey Mode on Self-Reports of Adult Alcohol Consumption: A Comparison of Mail, Web and Telephone Approaches." *Journal of Studies on Alcohol* 66(2), 239–45.
- Lutig, P., Lensvelt-Mulders, G. J. L. M., Frerichs, R., e A. Greven. 2011. "Estimating Nonresponse Bias and Mode Effects in a Mixed-mode Survey." *International Journal of Market Research* 53(5), 669-686.
- Marradi, A. 1997. *Casuale e rappresentativo: ma cosa vuole dire?*. In "La politica e i sondaggi." P. Ceri (eds.), Torino: Rosenberg & Sellier, 9-52.
- Marradi, A. 1995. *L'analisi monovariata*. Franco Angeli: Bologna.
- Martin, P. 2011. *What makes a good mix? Chances and challenges of mixed mode data collection in the ESS*. CCSS, City University. https://www.city.ac.uk/__data/assets/pdf_file/0015/125133/CCSS-Working-Paper-No-02.pdf.
- McNemar, Q. 1947. "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages." *Psychometrika* 12, 153–157.
- Morgan, S. L., e C. Winship. 2009. *Counterfactuals and Causal Inference: Methods and Principles for Social Research. Analytical Methods for Social Research*. New York: Cambridge University Press.
- Murgia, M., e A. Nunnari. 2012. "Improve The Quality On Data collection: minimum requirements for a generalised software independently from the mode." Relazione presentata all'UNECE Seminar on New Frontiers for Statistical Data Collection, Ginevra, 31 Ottobre - 2 Novembre.
- Palmieri, M. 2013. *Il pretesting del questionario strutturato nella ricerca sociale L'analisi dell'interazione tra intervistatore e intervistato* - Tesi di dottorato. http://padis.uniroma1.it/bitstream/10805/2195/1/Tesi_MarcoPalmieri.pdf.
- Park, S., Kim, J. K., e S. Park. 2016. "An imputation approach for handling mixed-mode surveys." *The annals of Applied Statistics* 10(2), 1063-1085.
- Parsons, N. L., e M. J. Manierre. 2013. "Investigating the relationship among pre-paid token incentives, response rates, and nonresponse bias in a web survey." *Field Methods* 26(2), 191 – 204.
- Pitrone, M. C. 2009. *Sondaggi e interviste. Lo studio dell'opinione pubblica nella ricerca sociale*. Milano: Franco Angeli.
- Pearl, J. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82, 669–688. <http://www.dx.doi.org/10.1093/biomet/82.4.669>.
- Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. (2nd edition). New York: Cambridge University Press.
- Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J.T., Martin, E., Martin, J., e E. Singer. 2004. *Methods for testing and evaluating survey questions*. In "Methods for testing and evaluating survey questionnaires." Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., e E. Singer (eds.), Wiley Series in Survey Methodology: Wiley.
- Recanatini F., Wallsten, S. J., e L. Colin Xu. 2000. "Surveying Surveys and Questioning Questions: Learning From The World Bank Experience On Firm Survey." *Social Science Research Network Electronic Paper Collection*. <http://ssrn.com/abstract=432061>.
- Roberts, C., Jackle, A., e P. Lynn. 2006. "Causes of mode effects: separating out interviewer and stimulus effects in comparisons of face-to-face and telephone surveys." *AAPOR-ASA Section on Survey Research Methods*. <https://core.ac.uk/download/pdf/16703693.pdf>.
- Roberts, C., e Vandeplas, C. 2017. "Estimating Components of Mean Squared Error to Evaluate the benefits of Mixing Data Collection Modes." *Journal of Official Statistics* 33(2), 303-334.

- Rosenbaum, P. R., e D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1), 41-55.
- Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66, 688-701.
- Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63, 581-92.
- Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.
- Rubin, D. B. 1991. "Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism." *Biometrics* 47, 1213-34.
- Rubin, D. B. 2006. *Matching sampling for causal effects*. New York, NY: Cambridge University Press.
- Rubin, D. B., e R. J. A. Little. 2002. *Statistical analysis with missing data* (2nd ed.). New York: Wiley.
- Sala E., e P. Lynn. 2009. "The potential of a multi-mode data collection design to reduce non response bias. The case of a survey of employers." *Quality and Quantity* 43, 123-136.
- Schouten, B., Shlomo, N. , e C. Skinner. 2011. "Indicators for Monitoring and Improving Representativity of Response." *Journal of Official Statistics* 27, 231-253.
- Schouten, B., Calinescu, M., e A. Luiten. 2013. "Optimizing quality of response through adaptive survey design." *Survey Methodology* 39, 29-58.
- Shouten, B., Van den Brakel, J. A., Buelens, B., van der Laan, J., e T. Klausch. 2013. "Disentangling mode-specific selection and measurement bias in social surveys." *Social Science Research* 42, 1555-1570.
- Suzer-Gurtekin, Z. T., Heeringa, S. G., e R. Valliant. 2012. "Investigating the bias of alternative statistical inference methods in sequential mixed-mode surveys." *Section on Survey Research Methods – JSM 2012*.
- Tourangeau, R., e T. Yan. 2007. "Sensitive questions in surveys." *Psychological Bulletin* 133, 859-883.
- Tourangeau, R., e T. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection, Question Format, and Question Context." *Public Opinion Quarterly* 60, 275-304.
- Tourangeau, R., Conrad, F., e M. Couper. 2013. *The science of web surveys*. New York: Oxford University Press.
- Tourangeau, R., Rips, L.J., e K. Rasinski. 2000. *The psychology of survey response*. Cambridge University Press.
- Tourangeau R., Brick, J. M., Lohr, S, e J. Li. 2017. "Adaptive and responsive survey designs: a review and assessment." *Statistics in Society* 180(1), 203-223.
- van Buuren, S. 2012. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Vannieuwenhuyze, J. T. A., e P. Lynn. 2014. Measurement effects between Capi and Web questionnaires in the UK Household Longitudinal Study. *Working Paper Series* No. 2014 – 01. March 2014.
- Vannieuwenhuyze, J. T. A., e G. Loosveldt. 2012. "Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects." *Sociological Methods & Research* 42(1), 82-104.
- Vandenplas, C., Loosveldt, G., e J. T. A. Vannieuwenhuyze. 2016. "Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes. A sequential mixed mode experiment." *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda)* 10(2), 119.
- Vannieuwenhuyze J., Loosveldt, G., e G. Molenberghs. 2010. "A Method for Evaluating Mode Effects in Mixed-mode Surveys." *Public Opinion Quarterly* 74(5), 1027-1045.
- Vannieuwenhuyze, J. T. A., Loosveldt, G., e G. Molenberghs. 2014. "Evaluating Mode Effects in

- Mixed-Mode Survey Data Using Covariate Adjustment Mod-els.” *Journal of Official Statistics* 30(1), 1-21. <http://dx.doi.org/10.2478/jos-2014-0001>.
- Vicente, P., e E. Reis. 2010. “Using questionnaire design to fight nonresponse bias in web surveys.” *Social Science Computer Review* 28(2), 251 – 267.
- White, I. R., Royston, P., e A. M. Wood. 2011. “Multiple Imputation Using Chained Equations: Issues and Guidance for Practice.” *Statistics in Medicine* 30(5), 377–399.
- Willis, G. 1994. “Cognitive interviewing and Questionnaire design: A training Manual.” *Working Paper no. 7*. Cognitive Methods Staff, National Center for Health Statistics. Washington DC.
- Willis, G. 2004. *Cognitive Interviewing Revisited: A useful techniques, in Theory?*. In “Methods for testing and evaluating survey questionnaires.” Presser S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., e E. Singer (eds.), Wiley Series in Survey Methodology: Wiley.
- Zindato, D., Cassata, L., Cecconi, F., Cecconi, N., Benassi, F., Farano, S., Maiozzi, Picci, M., e A. Sasso. 2017. *Il Censimento della popolazione: dalla rilevazione “porta a porta” alla rilevazione paperless*. In “L'utilizzo della tecnica Cawi nelle indagini su individui e famiglie.” Collana: Letture statistiche – Metodi. <http://www.istat.it/it/archivio/203729>.

APPENDICE¹

In questa Appendice sono fornite informazioni di dettaglio sulle procedure di calcolo di alcuni indicatori statistici introdotti nel Paragrafo 4.2.3 nell'ambito delle analisi preliminari dei dati rilevati in indagini multi-tecnica.

Ci si riferisce in particolare agli indicatori riportati nello schema della Figura 4.8 per variabili di natura categorica (nominali o ordinali):

- l'indice di dissomiglianza di Leti ($Im1$) e il corrispondente indice quadratico ($Im2$) per il confronto fra distribuzioni univariate;
- l'indice di dissomiglianza bivariato ($Ij1$) per il confronto fra distribuzioni congiunte;
- il coefficiente di contingenza V di Cramer (*Cramer*) per l'analisi dei livelli di associazione fra variabili.

Per il calcolo di questi indicatori è necessario disporre di due data set separati (corrispondenti agli insiemi di unità campione rilevati con due tecniche differenti) contenenti ciascuno i dati da mettere a confronto (sia nel caso di tecniche utilizzate su campioni indipendenti, sia nel caso di tecniche utilizzate su sotto-campioni osservati in sequenza o in modo concorrente, dunque non indipendenti).

Cinque sono i programmi previsti dalla procedura, sviluppati in linguaggio SAS, e che devono risiedere tutti nella stessa libreria:

- *Lancio_Indicatori nostrato.sas*: programma da utilizzare per il lancio delle routine di calcolo degli indicatori;
- *Lancio_Indicatori con strato.sas*: programma da utilizzare per il lancio delle routine di calcolo degli indicatori per strato;
- *Indicatore_Im.sas*: il programma calcola gli indicatori $Im1$ e $Im2$;
- *Indicatore_Ij.sas*: il programma calcola l'indicatore $Ij1$;
- *Indicatore_Cr.sas*: il programma calcola il coefficiente *Cramer*.

I programmi di calcolo degli indici sono generalizzati e dunque non richiedono nessuna modifica da parte dell'utente. Ai fini del calcolo, l'utente deve infatti agire esclusivamente sui programmi *Lancio_Indicatori nostrato.sas* o *Lancio_Indicatori con strato.sas* nei quali vanno definiti i seguenti parametri:

- nome e percorso fisico della libreria dove si trovano memorizzati i data set oggetto di analisi;
- nome e percorso fisico della tabella Excel contenente alcuni parametri specifici utilizzati dalle procedure di calcolo degli indicatori;
- nomi dei data set oggetto di analisi;
- (solo per il programma *Lancio_Indicatori con strato.sas*) nome della variabile di strato.

Di seguito si forniscono alcune avvertenze di tipo tecnico per il corretto utilizzo delle procedure:

- i data set oggetto di analisi devono risiedere nella stessa cartella e quindi nella stessa libreria, ovviamente con nomi diversi;
- i data set a confronto devono avere uguale struttura interna;
- i nomi delle variabili analizzate devono essere uguali in ambedue i file;

¹ Il contributo è stato redatto da Giorgio Della Rocca.

- non è necessario che le variabili abbiano lo stesso formato, il programma provvederà automaticamente ad uniformare i formati;
- nel caso si vogliano utilizzare più variabili di strato (ad es. *Ripartizione Territoriale, classe di ampiezza dei comuni, etc.*), deve essere creata nei due data set analizzati, esternamente alle procedure di lancio, una stessa variabile ottenuta mediante concatenazione delle variabili scelte.

Nella Tavola A.1 è riportata la struttura della tabella Excel “*tabella_input.xlsx*” in cui vanno riportati i nomi delle variabili da sottoporre ad analisi, e alcuni parametri utilizzati dalle procedure di calcolo:

1. colonna *CTR*: il simbolo “*” posto in corrispondenza di una variabile indica che la variabile è selezionata per l’analisi;
2. colonna *Nome_var*: contiene il nome di ogni variabile candidata all’analisi (identico nei due file);
3. colonna *Missing*: contiene il simbolo “Y” tutte le volte che possono essere presenti valori mancanti nella distribuzione della variabile da analizzare (ad esempio nel caso in cui si confrontino distribuzioni di dati *grezzi*);
4. colonna *Min*: contiene il valore minimo delle modalità della variabile oggetto di analisi;
5. colonna *Max*: contiene il valore massimo delle modalità della variabile oggetto di analisi.

Tavola A1 - Tabella Excel per la selezione delle variabili e dei parametri per il calcolo degli indicatori statistici

CTR	Nome_var	Missing	Min	Max
*	sezzo		1	2
*	istr		1	12
	panpas	Y	1	5
	salumi		1	5
	pollo	Y	1	5

Al termine delle elaborazioni, i valori degli indicatori saranno disponibili su tre data set SAS distinti, nella stessa libreria dei dati, con i seguenti nomi:

- *Indiceresidenti_Im.sas7bdat*: contiene i valori degli indici I_{m1} e I_{m2} per le variabili oggetto di analisi;
- *Indice_Jm.sas7bdat*: contiene i valori dell’indice I_{j1} per le *t-uple* di variabili oggetto di analisi ($t=1,2,3$);
- *Indice_cr.sas7bdat*: contiene i valori del coefficiente *Cramer* per le variabili oggetto di analisi prese due a due.

Le tabelle di output contengono, oltre ai nomi delle variabili interessate dall’analisi e ai valori dei relativi indici, anche informazioni sulla data di effettuazione dell’elaborazione, il nome della variabile di strato e il valore della variabile di strato assunto nei cicli di elaborazione (che assumeranno rispettivamente i valori “TOT_INDAGINE” e “0” nel caso di lancio del programma senza stratificazione).

In relazione a questo aspetto, è importante ricordare che, ogni volta che viene eseguito il programma, è prevista per *default* la cancellazione dei valori degli indici precedentemente calcolati. Tale possibilità può essere soppressa escludendo il relativo passo del programma - opportunamente evidenziato nel programma SAS - ad esempio nel caso in cui si voglia

ripetere l'elaborazione su (sotto)insiemi di dati diversi e verificare l'impatto delle varie classificazioni sui valori degli indici calcolati in precedenza.

Di seguito si riportano alcune considerazioni per il corretto uso statistico delle procedure:

- i data set a confronto devono contenere dati relativi ad uno stesso stadio di lavorazione statistica, dunque corrispondenti entrambi o a dati *grezzi* (cioè a monte del processo di controllo e correzione), oppure a dati *puliti* (cioè a valle del processo di controllo e correzione). In particolare, nel caso si confrontino dati grezzi, è necessario porre attenzione all'attivazione o meno del parametro *missing* nella tabella Excel di cui alla Tavola A.1;
- nel caso si effettuino confronti su specifiche sotto-popolazioni (ad es. strati o domini), è necessario che le variabili di stratificazione prescelte siano statisticamente connesse all'effetto selezione;
- il calcolo degli indicatori, per quanto automatico e quindi poco costoso, va limitato alle variabili che si ritiene maggiormente soggette all'effetto tecnica;
- per quanto riguarda l'indicatore I_{jt} , ai fini della sua interpretabilità è suggerito di limitarne l'uso all'analisi delle relazioni bi- o al più tri-variate.

Si riportano nel seguito le istruzioni del programma di lancio della procedura (*Lancio_Indicatori.sas*). In grassetto sono evidenziati i passaggi modificabili a cura dell'utente per ogni specifica applicazione. Il programma richiama a sua volta gli specifici programmi per il calcolo dei vari indicatori.

```

/*****
PROGRAMMA DI LANCIO DELLE PROCEDURE DI CALCOLO DEGLI INDICATORI STA-
TISTICI SENZA STRATIFICAZIONE
*****/
options nomprint;
/*****Al momento non modificare *****/
%let str_peso1 = 1;
%let str_peso2 = %STR(;);
%let str_where = %STR(;);
/***** Fine del passo *****/

/* NOME DELLA LIBRERIA RELATIVO INDIRIZZO E DEFINIZIONE DEL NOME DEI DA-
TASET INTERESSATI AD ANALISI *** */
%let lib = indici;
Libname &lib '\\nas-balbo\effetto-stat\CALCOLO INDICATORI\dati';

%let DS1 = AVQ2014_papi;
%let DS2 = AVQ2014_cawi;

/* FINE NOME DELLA LIBRERIA RELATIVO INDIRIZZO E DEFINIZIONE DEL NOME DEI
DATASET INTERESSATI AD ANALISI **/** */

LETTURA TABELLA EXCEL CONTENENTE I NOMI DELLE VARIABILI SOGGETTE A D
ANALISI E IL LORO CAMPO DI OSSERVAZIONE

```



DEFINIZIONE VARIABILI MACRO DA UTILIZZARE NEI PROGRAMMI DEGLI INDICATORI **/

```

PROC IMPORT OUT= limiti
DATAFILE= “\\nas-balbo\effetto-stat\CALCOLO INDICATORI\tabella_input.xlsx”
    DBMS=EXCEL2000 REPLACE;
    GETNAMES=YES;
run;
data limiti;
    set limiti;
where ctr = ‘*’;
Run;
/**** DEFINIZIONE DELLE VARIABILI MACRO *****/
DATA _NULL_;SET limiti END=EOF;
    CNT + 1;
    CALL SYMPUT(‘VAR’||LEFT(CNT),nome_var);
    CALL SYMPUT(‘miss’||LEFT(CNT),missing);
    CALL SYMPUT(‘min_s’||LEFT(CNT),min);
    CALL SYMPUT(‘max_s’||LEFT(CNT),max);
    IF EOF THEN
    CALL SYMPUT(‘nobs’,CNT);
    RUN;

/***** FINE LETTURA TABELLA EXCEL *****/
/** CREAZIONE STRINGA VARIABILI INTERESSATE ANALISI E CREAZIONE DATASET
CON VARIABILE SELEZIONATE */
%let str_var = ;
%macro strv;
%do k = 1 %to &nobs;
    %let str_var = &str_var &&var&k ;
%end;
%mend;
%strv;
%put ***** &str_var;

data &DS1; set &lib..&DS1; keep &str_var;run;
data &DS2; set &lib..&DS2; keep &str_var;run;
/**** FINE CREAZIONE STRINGA VARIABILI INTERESSATE ANALISI E CREAZIONE DA-
TASET CON VARIABILE SELEZIONATE*/
/***** RINOMINA E UNIFORMA I FORMATI *****/
%macro rinomina;
%do k = 1 %to &nobs;

data &DS1; set &DS1;
N&&var&k = &&var&k * 1;
drop &&var&k;
run;

```



```

data &DS1; set &DS1;
rename N&&var&k =&&var&k;
Run;
/*****/
data &DS2; set &DS2;
N&&var&k = &&var&k * 1;
Drop &&var&k;
run;

data &DS2; set &DS2;
rename N&&var&k =&&var&k;
Run;
/*****/
%end;
%mend;
%rinomina
/***** FINE RINOMINA E UNIFORMO *****/
/***** CANCELLAZIONE DATASET DI USCITA CON VECCHI INDICATORI *****/
**** ATTENZIONE : ISTRUZIONI DA USARE OGNI VOLTA CHE SI ESEGUO I PGM
ALTRIMENTI ACCODA I RISULTATI NEI DATASET DI USCITA */
proc datasets library=&LIB noprint/* memtype=data details*/;
  delete Indici_Im Indice_Jm Indice_Cr;
run;
quit;
/***** FINE CANCELLAZIONE DATASET DI USCITA CON VECCHI INDICATORI
*****/

/****LANCIO SINGOLI PROGRAMMI – NB: AGGIORNARE PERCORSO *****/

%INCLUDE “\\nas-balbo\effetto-stat\CALCOLO INDICATORI\pgm\indicatore_Im.sas”;
%INCLUDE “\\nas-balbo\effetto-stat\CALCOLO INDICATORI\pgm\indicatore_Jm.sas”;
%INCLUDE “\\nas-balbo\effetto-stat\CALCOLO INDICATORI\pgm\indicatore_Cr.sas”;

/****FINE LANCIO SINGOLI PROGRAMMI *****/

Di seguito è riportato il listato del programma di lancio del calcolo degli indicatori
statistici nel caso in cui si vogliono utilizzare delle stratificazioni. I parametri di questo pro-
gramma sono gli stessi del programma precedente, con in più la necessità di indicare la
variabile di strato scelta nel punto indicato in grassetto nel programma.
/*****
PROGRAMMA DI LANCIO PER GLI INDICATORI STATISTICI CON STRATIFICAZIONE
*****/
options nomprint;
/*****Al momento non modificare *****/
%let str_peso1 = 1;
%let str_peso2 = %STR(;;);

```



```
%let str_where = %STR(;;)
/***** Fine del passo Al momento non modificare *****/
/* NOME DELLA LIBRERIA RELATIVO INDIRIZZO E DEFINIZIONE DEL NOME DEI DA-
TASET INTERESSATI AD ANALISI ***/
%let lib = indici;
Libname &lib "\nas-balbo\effetto-stat\CALCOLO INDICATORI\dati";

%let DS1 = AVQ2014_papi;
%let DS2 = AVQ2014_cawi;

/* DEFINIZIONE DELLA VARIABILE DI STRATO */
%let var_strato = cod_regione;
/*****
/* FINE NOME DELLA LIBRERIA RELATIVO INDIRIZZO E DEFINIZIONE DEL NOME DEI
DATASET INTERESSATI AD ANALISI E DELLA VARIABILE DI STRATO*****/
```