

La cassetta degli attrezzi. Strumenti per le scienze umane

Direttore

Giovanni Di Franco, Università di Salerno

Comitato editoriale

Elena Battaglini, Ires-Cgil

Sara Bentivegna, Università di Roma

Alberto Marradi, Università di Firenze

Federica Pintaldi, Istat

Luciana Quattrociocchi, Istat

Marta Simoni, Iref-Acli

La collana, rivolta a ricercatori accademici e professionisti, studiosi, studenti, e operatori del variegato mondo della ricerca empirica nelle scienze umane, si colloca sul versante dell'alta divulgazione e intende offrire strumenti di riflessione e di intervento per la ricerca.

Obiettivo è consolidare le discipline umane presentando gli strumenti di ricerca empirica, sia di raccolta sia di analisi dei dati, in modo intellegibile e metodologicamente critico così da consentirne l'applicazione proficua rispetto a definiti obiettivi cognitivi.

I testi sono scritti da professionisti della ricerca che, attingendo alla personale esperienza maturata in anni di attività, offrono ai lettori strumenti concettuali e tecnici immediatamente applicabili nella propria attività di ricerca.

Tutti i volumi pubblicati sono sottoposti a referaggio.

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “informazioni” per ricevere via e-mail le segnalazioni delle novità o scrivere, inviando il loro indirizzo, a: “FrancoAngeli, viale Monza 106, 20127 Milano”.

Federica Pintaldi

COME SI ANALIZZANO I DATI TERRITORIALI

La cassetta degli attrezzi
Strumenti per le scienze umane/3

FrancoAngeli

Progetto grafico della copertina: Elena Pellegrini

Copyright © 2009 by FrancoAngeli s.r.l., Milano, Italy.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

119. *La cassetta degli attrezzi. Strumenti per le scienze umane*

1. Giovanni Di Franco, *L'analisi dei dati con Spss. Guida alla programmazione e alla sintassi dei comandi*
2. Silvia Cataldi, *Come si analizzano i focus group*
3. Federica Pintaldi, *Come si analizzano i dati territoriali*
4. Giovanni Di Franco, *Il campionamento nelle scienze umane: come ridurre la distanza fra la teoria e la pratica*
5. Alberto Marradi, *Come evitare gli errori tipici in un questionario*

Indice

1. Introduzione	pag.	11
2. Potenzialità e limiti dell'analisi territoriale	»	13
2.1 L'oggettività dei dati territoriali	»	13
2.2 L'analisi secondaria dei dati	»	19
2.3 Le fonti dei dati	»	27
2.4 Cosa leggere per saperne di più	»	32
3. Tipi di unità di analisi territoriale	»	37
3.1 Definizione di unità di analisi	»	37
3.2 Classificazione delle unità di analisi territoriali	»	41
3.3 Unità amministrative con funzioni generali	»	45
3.4 Unità amministrative con funzioni specifiche	»	61
3.5 Unità non amministrative	»	64
3.6 Cosa leggere per saperne di più	»	68
4. Caratteristiche dell'unità di analisi territoriale	»	71
4.1 L'influenza dell'unità di analisi territoriale	»	71
4.2 Problemi di omogeneità degli esemplari	»	75
4.3 Le dimensioni spaziale e temporale	»	77
4.4 Cosa leggere per saperne di più	»	81
5. Tipi di variabili	»	85
5.1 Proprietà, definizione operativa e variabile	»	85
5.2 Relazione tra proprietà e referente empirico	»	88
5.3 Costruzione di variabili aggregate	»	92
5.4 Definizione degli indicatori	»	97
5.5 Cosa leggere per saperne di più	»	101

6. Relazioni tra variabili	pag.	103
6.1 La fallacia ecologica: cosa è	»	103
6.2 La fallacia ecologica: esempi empirici	»	112
6.3 Relazione tra dati individuali e territoriali	»	116
6.4 Dal sondaggio atomistico al sondaggio contestuale	»	124
6.5 Cosa leggere per saperne di più	»	126
7. L'analisi dei dati territoriali: esempi empirici	»	137
7.1 Costruzione della matrice dei dati	»	137
7.2 Lo squilibrio dimensionale uno	»	140
7.3 Lo squilibrio dimensionale due	»	146
7.4 La scelta dell'unità di analisi	»	148
7.5 L'eterogeneità interna ai casi	»	158
7.6 Popolazione e territorio nelle unità territoriali	»	159
7.7 Relazioni tra variabili territoriali	»	164
7.8 Cosa leggere per saperne di più	»	169

Per Robinson (1950, 352) i dati ecologici non si utilizzano per caratterizzare il territorio ma per scoprire il comportamento individuale: “la correlazione ecologica è utilizzata solo perché non sono disponibili le correlazioni tra le proprietà degli individui”.

Menzel (1950, 674) rivendica l'autonomia per i dati territoriali e per le inferenze che da questi si possono trarre. Secondo l'autore una correlazione ecologica può essere interessante per il fatto che tale relazione indica una causa comune al manifestarsi dei due fenomeni. Inoltre le correlazioni a livello ecologico possono indicare anche una relazione di causa ed effetto tra le variabili: “la correlazione ecologica può avere un grande valore anche senza riflettere la correlazione individuale”.

Per Goodman (1953) lo studio delle relazioni a livello ecologico può essere effettuato per due motivi: a) interesse per le relazioni a livello ecologico; b) interesse per le relazioni a livello individuale.

Blau (1960) ha sottolineato che può considerarsi limitativo analizzare i dati aggregati di per sé; lo stesso vale per lo studio del comportamento individuale senza tener conto del contesto sociale e dell'interazione tra individuo ed ambiente.

Per Micheli (1978, 617) i dati aggregati presentano due vantaggi: sono meno onerosi (in termini di tempo e denaro) e vi è “un maggiore controllo delle perturbazioni dovute all'equazione personale del rilevatore”.

1. Introduzione

Il volume è rivolto a chi – studenti, ricercatori sociali, politici, amministratori e operatori di enti locali, etc. – necessita conoscere in modo approfondito i contesti territoriali sia per soddisfare particolari obiettivi cognitivi sia per intervenire sugli stessi, attraverso la pianificazione di interventi, programmi, e politiche sociali.

Conoscere le risorse e i vincoli connessi all'uso dei dati territoriali (anche detti ecologici) costituisce la pre-condizione indispensabile per utilizzarli efficacemente. Fin dalle origini, questi dati caratterizzano gran parte della ricerca nelle scienze umane (basti pensare alla classica ricerca di Durkheim sul suicidio pubblicata nel 1901). Eppure le loro peculiarità sono spesso trascurate nei manuali di metodologia della ricerca.

La proliferazione delle fonti e degli enti che producono i dati su unità di analisi territoriale, e la sempre maggiore diffusione degli stessi attraverso le reti informatiche e altri supporti tecnologici, esige un'appropriata preparazione metodologica che consenta di selezionare adeguatamente l'enorme massa di dati a disposizione dei ricercatori.

L'analisi dei dati territoriali consente di caratterizzare i contesti all'interno dei quali avviene l'interazione sociale. Il suo valore conoscitivo consiste nella possibilità di poter studiare la dimensione societaria. L'individuazione dell'unità territoriale in cui collocare lo studio dei fenomeni sociali oggetto di studio è la questione centrale: l'unità di analisi influenza tutto il processo di elaborazione e interpretazione dei dati. Pertanto, la portata euristica dell'analisi territoriale dipende innanzitutto da quanto l'unità scelta rappresenta una

area territoriale significativa per il fenomeno oggetto di studio. Le unità territoriali racchiudono un territorio con una propria storia. Se è vero che con il processo di globalizzazione si azzerano le distanze, spingendo l'esperienza sociale in una dimensione globale nello spazio e nel tempo, la radicalizzazione di questa posizione rischia di incorrere in quella che è stata chiamata *globalization fallacy*. Diversamente i contesti locali continuano a mantenere una loro importanza, quali espressione di una particolare struttura socio-economica e di una determinata tradizione culturale. La dimensione territoriale dei fenomeni sociali è dunque un aspetto imprescindibile. Questo livello di analisi permette di cogliere le tendenze di fondo di una società, la statica e la dinamica dei fenomeni sociali.

Spesso i dati desunti con la tecnica dell'inchiesta con un questionario somministrato ad un campione di una popolazione sono connotati negativamente in quanto prodotti in una sorta di 'vuoto sociale'. Con tale espressione si afferma la necessità per la ricerca sociale di tenere conto dei contesti sociali entro i quali sono raccolte le opinioni, gli atteggiamenti, i valori, i comportamenti degli attori individuali, nella consapevolezza che i contesti sociali non sono in alcun modo desumibili a partire dalle sole caratteristiche degli individui che ne fanno parte. Pertanto, è necessario disporre di informazioni che siano rilevate su un livello di analisi che superi il piano degli individui e che non sia totalmente riducibile alle sole caratteristiche individuali.

Nel testo si avanza la proposta di passare dal cosiddetto sondaggio atomista (ossia il sondaggio condotto tenendo conto solo delle caratteristiche degli individui) al sondaggio contestuale (tenendo conto delle caratteristiche degli individui in relazione ai diversi contesti sociali individuabili mediante una previa analisi territoriale). La fusione intelligente di dati individuali e di dati territoriali non potrà che elevare la qualità delle ricerche e delle analisi sociali, migliorando altresì la comprensione dei fenomeni oggetto di studio.

Ringrazio Giovanni Di Franco per la revisione del lavoro.

2. Potenzialità e limiti dell'analisi territoriale

2.1 L'oggettività dei dati territoriali

Di solito, illustrando le caratteristiche dell'analisi territoriale, si evidenzia il vantaggio di disporre di dati spesso definiti 'oggettivi' o *hard* in opposizione ai cosiddetti dati 'soggettivi' (anche detti *soft*, 'sensibili', o 'percettivi') rilevati tramite indagini condotte su un campione di individui.

Ovviamente, per le unità territoriali non è possibile effettuare una rilevazione diretta mediante intervista, ossia la tecnica di analisi più diffusa nelle indagini individuali. Per cui i dati riferiti ad una unità territoriale, perdendo quel carattere di soggettività e intenzionalità legati all'individuo e all'interazione intervistato-intervistatore, apparentemente divengono immuni dai problemi legati all'affidabilità e alla fedeltà. Tale connotazione rischia di sopravvalutarne la validità e l'attendibilità, dimenticando che tutti i dati sono costruiti: essendo il risultato di un insieme di scelte, a volte consapevoli e a volte inconsapevoli, effettuate dal ricercatore e/o dall'ente preposto alla loro rilevazione.

Inoltre, in relazione ai dati territoriali, la frequente e immediata disponibilità della matrice dei dati può far dimenticare tutto quello che c'è a monte della stessa. Come ricorda Di Franco (2005,38): "Si deve evitare di cadere in un atteggiamento scienziato di santificazione della matrice dei dati dovuto al convincimento che essa sia una rappresentazione «oggettiva» della realtà".

Ad esempio, si dimentica che alcune informazioni disponibili a livello aggregato derivano da informazioni rilevate tramite questionario su unità di analisi individuale e successivamente aggregate.

I dati del censimento della popolazione, punto di riferimento per le indagini con unità territoriale, derivano da informazioni raccolte con un questionario auto-compilato presso le abitazioni delle famiglie italiane. Non si capisce per quale ragione l'informazione fornita dal soggetto sul proprio livello di istruzione debba ritenersi un dato 'soggettivo', mentre il tasso di analfabetismo a livello comunale (derivato da una opportuna aggregazione dei dati di censimento) divenga quasi per magia un dato 'oggettivo'.

Quando i dati sono presentati in forma aggregata assumono una forma cristallizzata, divenendo più affidabili rispetto a quando la stessa informazione si trovava in forma disaggregata a livello individuale. Il fascino del numero può incidere su questo incremento di credibilità assegnata all'informazione aggregata. Come vedremo nel paragrafo 5.3, nel passaggio dall'unità di analisi individuale ad una aggregata il cambiamento di natura della variabile da categoriale a cardinale può indurre questo malinteso.

Un secondo aspetto che induce a ritenere 'oggettivi' i dati territoriali deriva dal fatto che, di solito, essi sono prodotti da una fonte ufficiale e riguardano prevalentemente registrazioni di caratteristiche di status piuttosto che di opinioni e di atteggiamenti. Ogni paese rileva un insieme di indicatori in modo sistematico al fine di valutare la situazione socio-economica e programmare l'attività politica.

L'inflazione, il prodotto interno lordo, il tasso di occupazione, l'indice di vecchiaia, il tasso di analfabetismo, etc. sono alcune di queste statistiche ufficiali. Esse assumono necessariamente un carattere di 'oggettività' per chi deve utilizzarle per programmare le politiche di un paese. Quando i dati sono la base per le decisioni – la legge finanziaria, il documento di programmazione economica, etc. – l'assunzione dell'oggettività è necessaria: la classe politica non può metterli in discussione altrimenti non potrebbe agire.

Peraltro, nei casi di critica delle statistiche ufficiali, il giudizio sull'affidabilità delle stesse spesso è espresso più in funzione di interessi politici che dall'attenta valutazione del processo di produzione dei dati.

Il ricercatore, però, non ha bisogno di assumere l'oggettività dei dati per poter lavorare. Al contrario, l'uso critico delle informazioni e la consapevolezza che qualsiasi dato è costruito aumenta la qualità del proprio lavoro.

Anche quando le informazioni aggregate non provengono da inchieste con questionario il loro grado di affidabilità deve essere accertato. Più autori hanno evidenziato i problemi di attendibilità e di validità dei dati provenienti da fonte amministrativa per mancato aggiornamento delle liste, confusione nei criteri classificatori, duplicazioni e/o incompletezza delle informazioni, etc. (Zajczyk 1996).

Ad esempio, l'Istat ha rielaborato i dati sui permessi di soggiorno al 1° gennaio 2000 forniti dal Ministero dell'Interno. Nella pubblicazione del volume *La presenza straniera in Italia: caratteristiche demografiche* (Istat 2001) si precisa che la stima fornita dall'Istituto è di circa 90.000 permessi di soggiorno in più rispetto ai dati diffusi dal Ministero a suo tempo; la differenza deriva sia dal recupero di documenti già rilasciati ma non registrati in tempo, sia dall'aggiunta dei permessi di soggiorno in scadenza e prorogati nei mesi successivi al 1° gennaio 2000. La Caritas, da anni attenta all'analisi della presenza degli stranieri in Italia, nella pubblicazione dell'XI rapporto sull'immigrazione denuncia una differenza tra i dati sui nuovi permessi di soggiorno concessi nel 2000 provenienti dall'archivio CED (Centro Elaborazioni Dati presso il Ministero dell'Interno) e quelli desumibili dall'archivio del Servizio Immigrazione e polizia di frontiera dello stesso Ministero dell'Interno.

In linea generale, quanto più i dati a disposizione sono frutto di elaborazione, tanto più diviene difficile ricostruire il processo che ha portato alla loro costruzione, e, quindi, valutare la loro utilità ai fini della ricerca. Informazioni disponibili in valore assoluto sono preferibili a tassi o percentuali, numeri indice già calcolati; soprattutto quando non sono espliciti i criteri utilizzati per normalizzare i dati.

Ciò vale in particolar modo per gli indici costruiti a partire da una combinazione di informazioni elementari. In alcuni casi questi indici vogliono rappresentare concetti molto generali, ma poco si conosce del rapporto di indicazione tra indicatori e concetto, della definizione operativa dei singoli indicatori, dell'algoritmo impiegato per sintetizzare le diverse variabili in un unico indice.

Numerose organizzazioni intergovernative, istituzioni private e fondazioni nazionali e internazionali pubblicano regolarmente classifiche su temi economici, sociali e di *governance* che spesso ottengono grande attenzione sulla stampa e contribuiscono a formare la percezione nazionale e internazionale della condizione dei singoli paesi.

Un buon esempio è costituito dall'indice di competitività Global Competitiveness Index (Gci) realizzato dal World Economic Forum (Wef), che intende fornire: "un quadro completo del panorama competitivo nei paesi di tutto il mondo a qualsiasi stadio di sviluppo". L'indice, che realizza la graduatoria di 131 paesi, trova grande attenzione sui media contribuendo a formare la percezione nazionale e internazionale della condizione dei singoli paesi, incidendo sulle scelte di localizzazione degli investimenti esteri e su altre attività economiche. Nel 2007 l'Italia si trova in 46^a posizione, manifestando un basso grado di competitività. Ma, come spesso accade, nei mezzi di comunicazione poco spazio è dedicato a spiegare il significato di questo indice. Oltre al fatto che negli anni le variabili selezionate per ciascuna dimensione cambiano, e ciò rende difficile il confronto con gli anni passati, il concetto di 'competitività' è talmente polisemantico che diviene difficile capire in che senso questo indice sintetico lo rappresenti. Lo schiacciamento del concetto sull'indice favorisce quel processo di reificazione dei dati: una graduatoria diventa *tout court* la 'competitività' di un paese.

Il Gci si basa su 12 dimensioni (pilastri) della competitività: istituzioni, infrastruttura, stabilità macroeconomica, sanità e istruzione primaria, istruzione superiore e formazione, efficienza dei mercati dei beni, efficienza del mercato del lavoro, livello di sofisticazione del mercato finanziario, capacità tecnologica, ampiezza del mercato, sofisticazione e innovazione. Ciascuna dimensione è definita da un insieme di indicatori, da un minimo di due ad un massimo di 28. La graduatoria finale dei paesi deriva dalla media dei punteggi ottenuti da ciascun paese nelle dodici graduatorie parziali, alle quali viene applicato un peso differente in relazione allo stadio di sviluppo del paese, calcolato tenendo conto del reddito pro-capite.

Complessivamente l'indice è composto da 113 indicatori: 33 provengono da fonti esterne (*hard data*) e 80 dall'Executive Opinion Survey (*soft data*), un sondaggio annuale realizzato dal World Economic Forum in collaborazione con circa 130 istituti nei diversi paesi analizzati. Il sondaggio si basa sulle interviste effettuate ad un campione di manager di imprese a

vocazione internazionale; nel 2007 in Italia sono state realizzate 110 interviste a manager di impresa. Non si trovano informazioni metodologiche né sui criteri di campionamento per la scelta dei soggetti intervistati (lista da cui sono selezionati, età, sesso, regione, professione, etc.), né sulla tecnica di intervista (*face to face*, telefonica, via e-mail). Il questionario utilizzato nel sondaggio è molto ampio, comprendendo 12 aree tematiche: azienda; andamento generale dell'economia; governo e istituzioni pubbliche; infrastrutture; innovazione e tecnologia; finanza; impresa; istruzione e capitale umano; corruzione, etica e responsabilità sociale; viaggi e turismo; ambiente; salute. Tutte le sezioni tematiche presentano la stessa struttura, con una batteria di domande con modalità di risposta articolate lungo una scala che varia da 1 (minimo) a 7 (massimo).

Oltre al fatto che l'utilizzo ripetuto della stessa scala può portare l'effetto *response set* (risposta meccanica), occorrerebbe interrogarsi su quanto la risposta ad una scala da 1 a 7 corrisponda al posizionamento dell'Italia in una graduatoria di 131 paesi. Ad esempio, al quesito sulla qualità dell'istruzione l'Italia ottiene il punteggio di 3,4 collocandosi in 77^a posizione, dopo Corea, Zimbabwe, Zambia, Kenya, Nigeria, e altri. Con tutte le critiche che si possono muovere al sistema scolastico in Italia, rimane dubbio che la risposta ad un quesito di questionario da parte di 110 manager possa essere indicativo della qualità dell'istruzione nel nostro Paese.

Analizzando i risultati dell'Italia più nel dettaglio, il maggior numero di posizioni basse in graduatoria derivano da cosiddetti indicatori *soft*: nel 71% dei casi l'Italia si colloca oltre il 50° posto. Ciò denota una percezione particolarmente negativa dell'Italia da parte degli intervistati sulla maggior parte delle tematiche in esame, almeno in confronto ai giudizi espressi negli altri paesi. La media delle graduatorie degli indicatori *soft* è pari a 69. In relazione agli indicatori *hard*, diversamente, in 4 casi su 10 l'Italia si trova tra i primi venti paesi.

Si ribadisce l'importanza di studiare l'andamento dei singoli indicatori e il loro peso relativo nel determinare il risultato finale. Infine, quando possibile, è sempre preferibile il confronto in termini di punteggio piuttosto che di graduatorie. Infatti, posizioni lontane in graduatoria possono corrispondere a differenze non significative in termini di valori nei singoli indicatori e viceversa.

Quando i dati provenienti da diverse fonti sono discordanti, questo schiacciamento del concetto sull'indice può comportare una strumentalizzazione dei risultati: si valuta attendibile il dato che corrobora la propria tesi senza considerarne le modalità di costruzione. Un caso interessante è la polemica scaturita a proposito dell'aumento dell'in-

flazione a seguito dell'introduzione dell'Euro. Su base annua, a dicembre 2002 l'Istat ha stimato l'inflazione del 2,8%, mentre l'Eurispes (un istituto di ricerca privato) ha diffuso un dato sull'aumento dei prezzi dei prodotti alimentari pari al 29%. Questa vistosa differenza fra il dato dell'inflazione rilevato dall'Istat e quello dell'Eurispes ha suscitato un notevole dibattito sui mezzi di comunicazione di massa. Nei telegiornali e nei quotidiani si è dedicato molto spazio alle interviste di persone per strada, nei mercati, nei bar, etc. Tutte le opinioni raccolte manifestavano una percezione dell'aumento dei prezzi molto superiore a quella registrata dall'Istat. Non è da escludersi una strumentalizzazione a fini politici di questa disputa fra un istituto privato, che ha buone ragioni per accrescere la propria visibilità sull'opinione pubblica, e l'Istituto Nazionale di Statistica.

Ricordiamo che il paniere dei beni dell'Istat è composto da 960 prodotti. La costruzione del paniere, e dei pesi assegnati a ciascun prodotto, derivano dai dati relativi ai consumi delle famiglie stimati nell'ambito della Contabilità Nazionale, che a sua volta si basano sulle informazioni provenienti dall'indagine campionaria sui consumi, svolta annualmente dall'Istat su un campione di circa 28.000 famiglie. I prezzi dei prodotti componenti il paniere vengono rilevati in 29.000 unità di rilevazione, alle quali si aggiungono circa 10.000 abitazioni per la rilevazione degli affitti. Ogni mese, in media, vengono rilevate 300.000 quotazioni. Il paniere dell'Eurispes è basato su 148 voci di prodotto. Il paniere e i relativi pesi di incidenza dei prodotti sono stati definiti in base ai consumi di 20 famiglie romane. La rilevazione si è svolta tra l'ultima settimana di novembre e la prima di dicembre, con la registrazione di 304 schede. Per di più non sempre era disponibile il prezzo del prodotto per l'anno precedente. A ciò si è rimediato controllando vecchi scontrini, con interviste dirette, manifesti promozionali, etc. Da questa breve nota si evince come la stima dell'Eurispes sia molto più vicina alle impressioni della massaia che non a una rilevazione condotta con criteri seri e trasparenti. Piuttosto che cavalcare la vulgata populista sull'aumento incontrollato dei prezzi, i mezzi di comunicazione di massa avrebbero reso un migliore servizio se si fossero occupati di spiegare ai loro utenti i criteri usati dai due istituti per rilevare l'inflazione.

Per loro natura i concetti, compresi quelli economici, non sono direttamente osservabili e tanto meno oggettivi. L'inflazione è un concetto di proprietà complesso che viene operazionalizzato rilevando le

variazioni di prezzo di un campione di prodotti, denominato paniere, e successivamente calcolando una media ponderata delle variazioni di ciascun prodotto. È ovvio, dunque, che la stima dipende dal paniere selezionato, dalla rappresentatività delle unità di campionamento, dall'affidabilità dei dati registrati e dai pesi assegnati a ciascuna variazione. Come più volte sottolineato, senza conoscere le procedure di costruzione dei dati non si può esprimere giudizio sulla loro attendibilità. A tale proposito, è bene diffidare delle fonti che non esplicitano i criteri che portano alla costruzione delle informazioni diffuse.

L'unica forma di oggettività che può essere pretesa dai dati territoriali (come da qualsiasi altro tipo di dato) è la trasparenza nelle procedure di rilevazione e registrazione delle informazioni, in modo che chiunque possa valutarne l'affidabilità. Oggettività, quindi, diviene sinonimo di intersoggettività, e di pubblicità delle procedure.

2.2 L'analisi secondaria dei dati

La differenza tra analisi primaria e secondaria dei dati, per quanto presenti molti punti in comune, non coincide con la differenza tra dati territoriali e dati individuali. Inizialmente il termine analisi secondaria identificava lo studio a posteriori dei dati di sondaggio. Successivamente, la definizione di analisi secondaria venne ampliata includendo qualsiasi tipo di analisi che si avvale di dati raccolti precedentemente e di solito per scopi diversi da quelli che muovono le analisi secondarie.

L'equiparazione fra l'analisi secondaria e l'analisi di dati aggregati nasce dal fatto che per le unità territoriali sono più facilmente disponibili informazioni rispetto ai dati di sondaggio. Per cui spesso le due cose, nei fatti, coincidono. A tale proposito, un indubbio vantaggio dell'analisi territoriale consiste nella vasta disponibilità di informazioni raccolte originariamente per altri fini. Soprattutto quando l'aggregato territoriale identifica un'unità amministrativa, il numero di indicatori disponibili è molto vasto e negli ultimi anni in continua espansione (vedi par. 2.3). L'analisi secondaria dei dati, dunque, offre al ricercatore l'ovvio vantaggio di ridurre notevolmente l'onere, in termini di costi e di tempo, della fase di raccolta dei dati.