

Giovanni Di Franco

**EDS: Esplorare,  
descrivere e  
sintetizzare i dati**

Guida pratica  
all'analisi dei dati nella  
ricerca sociale



IL RICCIO E LA VOLPE

*Studi, ricerche e percorsi di sociologia*

**FrancoAngeli**

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

Giovanni Di Franco

**EDS: Esplorare,  
descrivere e  
sintetizzare i dati**

Guida pratica  
all'analisi dei dati nella  
ricerca sociale

**FrancoAngeli**

*Grafica della copertina: Maria Teresa Pizzetti*

Seconda edizione Copyright © 2001, 2005 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 specificate sul sito [www.francoangeli.it](http://www.francoangeli.it)*

## Indice

<b>Prefazione di Enzo Campelli</b>	pag.	7
<b>Introduzione</b>	»	9
<b>1. Esplorare, descrivere e sintetizzare i dati nella ricerca sociale</b>	»	13
1.1. Introduzione	»	13
1.2. Definizione e funzioni dell'analisi EDS	»	15
1.3. Cosa si deve sapere prima di progettare un'analisi EDS	»	21
1.4. Pianificare un'analisi EDS	»	29
<b>2. Dall'informazione ai dati: progettare e costruire una matrice dei dati</b>	»	31
2.1. Introduzione: la matrice dei dati e l'algebra matriciale	»	31
2.2. Progettare una matrice dei dati	»	36
2.3. Unità di analisi e relative proprietà	»	38
2.4. Costruire una matrice dei dati	»	39
2.5. Altre matrici utili per l'analisi dei dati	»	54
<b>3. Preparare i dati per l'analisi</b>	»	56
3.1. Funzioni della pre-analisi dei dati	»	56
3.2. Etichettare le variabili e le modalità	»	57
3.3. Il trattamento dei dati mancanti	»	58
3.4. Ricodificare i dati di una variabile	»	63
3.5. Costruire nuove variabili a partire dalle variabili presenti nella matrice dei dati	»	66
3.6. Ponderare i casi, normalizzare e standardizzare i dati	»	70
3.6.1. Tecniche di ponderazione	»	70
3.6.2. La normalizzazione la standardizzazione dei dati	»	74
3.7. La deflazione dei dati	»	82

<b>4. Principi di analisi monovariata</b>	pag.	87
4.1. Descrivere, sintetizzare e rappresentare la distribuzione di una variabile	»	87
4.2. L'analisi monovariata delle variabili categoriali non ordinate	»	91
4.3. L'analisi monovariata delle variabili con categorie ordinate	»	99
4.4. L'analisi monovariata delle variabili categoriali derivate da batterie di domande	»	105
4.5. L'analisi monovariata delle variabili cardinali e quasi-cardinali	»	110
<b>5. Principi di analisi bi- e tri-variata</b>	»	123
5.1. Esplorare la relazione tra due variabili: sintetizzarla mediante un coefficiente e rappresentarla graficamente	»	123
5.2. L'analisi bivariata delle variabili categoriali non ordinate	»	127
5.3. L'analisi bivariata delle variabili con categorie ordinate	»	140
5.4. L'analisi bivariata delle variabili quasi-cardinali e delle variabili cardinali	»	145
5.5. Esplorare la relazione fra tre variabili: sintetizzarla mediante un coefficiente e rappresentarla graficamente	»	167
5.6. L'analisi trivariata delle variabili categoriali	»	168
5.7. L'analisi trivariata delle variabili quasi-cardinali e cardinali	»	175
<b>6. Tecniche di analisi multivariata per analisi EDS</b>	»	181
6.1. Analisi in componenti principali	»	182
6.2. Lettura dei risultati e dei grafici ottenuti con l'analisi in componenti principali	»	193
6.3. Analisi delle corrispondenze multiple	»	209
6.4. Lettura dei risultati ottenuti dall'analisi delle corrispondenze multiple	»	215
6.5. Tecniche di analisi dei gruppi	»	223
6.6. Lettura dei risultati e dei grafici ottenuti dall'analisi dei gruppi	»	232
6.7. Per andare oltre questo libro	»	249
<b>Bibliografia</b>	»	251

## *Prefazione*

di Enzo Campelli

Come ricorda Isaiah Berlin, un verso isolato nei frammenti sopravvissuti di Archiloco afferma lapidariamente, e in verità piuttosto oscuramente, che “la volpe sa molte cose, ma il riccio ne sa una grande”: πολλ’ οιδ’ αλωπηξ, αλλ’ εχινος εν μεγα<sup>1</sup>.

Archiloco, il più antico poeta greco di cui si abbiano notizie storicamente fondate, è uomo appassionato, aggressivo, talvolta disperato, che irride con i suoi giambi sobri e spesso feroci i valori tradizionali della società, aristocratica e guerriera, a cui pure appartiene. Non è chiaro quale sia il senso, per lui, dell’opporre il sapere della *volpe*, che la metafora può suggerire come predatorio e strumentale, forse occasionale ma spregiudicato ed audace, al sapere del *riccio*, caparbio e difensivo, solido e approfondito, forse limitato ma sistematico, in sé soddisfatto e compiuto per la sua stessa asserita *grandezza*. Né è chiaro, in definitiva, quale sia per Archiloco dei due il sapere *migliore*, per lui che, se pure ironizza spavalidamente contro i miti guerreschi e la vanità della gloria degli eroi<sup>2</sup>, accetta fedelmente di combattere per la sua patria fino alla morte. E’ piuttosto Berlin a suggerire, di questo verso spezzato, un’interpretazione importante. Come è noto, in uno scritto degli anni ’50 egli lo assume come capace di illustrare una delle più profonde differenze che dividerebbero gli scrittori e i pensatori. “Esiste infatti – osserva Berlin – un grande divario tra coloro, da una parte, che riferiscono tutto a una visione centrale, a un sistema più o meno coerente e articolato, con regole che li guidano a capire, a pensare e a sentire – un principio ispiratore, unico e universale, il solo

1. Cfr. l’edizione critica dei frammenti di Archiloco, curata da G. Tarditi: *Archiloco*, Roma, Edizioni dell’Ateneo, 1968; si tratta metricamente di un epodo, frammento n.196, p.169.

2. Si può ricordare a questo proposito il Frammento n.8, in cui, confessando la propria infelice *performance* in una battaglia contro i Traci, in cui aveva dovuto abbandonare lo scudo, massima vergogna possibile per un guerriero greco, Archiloco scrive sgarbatamente che “del mio scudo si fa bello qualcuno dei Sai; l’ho abbandonato presso un cespuglio, arma senza difetto, a malincuore. Ma ho salvata la vita. Vada in malora lo scudo: ne comprerò uno migliore” [ασπιδα μεν Σαιων τις ανειλετο, την παρα θαμνοι εντος αμωμητων καλλιπον ουχ εθελων, αυτος δ’ εξεφυγον θανατου τελος. Ασπισ εκεινη ερρετω, εξ αυτις κτησομαι ου κακιω]; cfr. *ibid.*, p.65.

che può dare un significato a tutto ciò che essi sono e dicono –, e coloro, dall'altra parte, che perseguono molti fini, spesso disgiunti e contraddittori, magari collegati soltanto genericamente, *de facto*, per qualche ragione psicologica o fisiologica, non unificati da un principio morale ed estetico”<sup>3</sup>.

Si può pensare allora che la polarità acutamente intravista da Berlin abbia anche una validità epistemologica, e serva altrettanto bene a segnalare due diversi orientamenti di interesse e di analisi, due distinti *programmi di ricerca* che percepibilmente si contendono il campo da tempo, nelle scienze sociali ed in particolare in sociologia. Si tratta di una dialettica cognitiva e affettiva che oppone, nello studio dei fenomeni sociali, l'unitario al molteplice, il frammentario al sistemico, il disordinato al conforme. Non è prudente né utile, tanto meno in questa stagione di mutamento sociale denso e accelerato, prendere definitivamente partito in una simile alternativa, né assumere la *necessità* di una riduzione reciproca. Al contrario, è forse da ritenere che solo in questa irrisolta *tensione essenziale* possano maturare risultati di autentico interesse teorico ed empirico.

Per tali ragioni questa nuova collana di studi, ricerche e percorsi di sociologia si propone programmaticamente di praticare entrambi i versanti, e di ricercare la congiunzione possibile fra il sapere del riccio e quello della volpe. Il volume che dà avvio al progetto è un singolare manuale di analisi dei dati, che precisa fin dal titolo e dalle primissime pagine le particolarità di un'impostazione precisa. Pur impegnandosi nella presentazione ragionata di tecniche di analisi multivariata, l'opzione decisiva del testo è quella di spostare fin da subito la concentrazione del lettore dal piano meramente tecnico della conoscenza e dalla pratica degli algoritmi e dei *software* a quello metodologico delle scelte e delle strategie cognitive. Infatti, scrive Giovanni Di Franco, “obiettivo di questo libro è porre l'attenzione sulla necessaria consapevolezza e conoscenza metodologica, prima ancora che statistica, di chi analizza i dati *prima, durante e dopo* l'uso di un programma di analisi dei dati” (p. 12). Come dire, in altri termini, che in nessun caso l'analisi dei dati può essere pensata come una semplice e meccanica applicazione di procedure, che possa prescindere da una contestuale operazione di riflessione teorica sulle procedure stesse. D'altra parte, “ambizione di questo lavoro è sollecitare nei lettori l'autoapprendimento che deriva dalla messa in pratica dei contenuti presentati” (*ibid.*): da ciò deriva che il tono prevalente nel testo non sia tanto quello della *presentazione* delle diverse procedure – obiettivo peraltro perseguito con accuratezza – quanto piuttosto quello della *discussione* delle procedure stesse, di un loro bilancio critico rispetto agli spazi di semplificazione e di indeterminazione che inevitabilmente ciascuna di esse contiene, da un lato, e rispetto agli obiettivi cognitivi perseguiti, dall'altro. A questa stessa esigenza risponde

3. I. Berlin, “Lev Tolstoj's Historical Scepticism”, in *Oxford Slavonic Papers*, II, 1951. Cfr. la traduzione italiana di una più ampia versione successiva in *Il riccio e la volpe*, Milano, Adelphi, 1986, pp. 71-72.



la decisione di pensare il lavoro come costituito da *due* componenti strettamente interconnesse: quella metodologica e quella applicativa. E' importante sottolineare la reciproca dipendenza dei due strumenti, ed il fatto che il contributo dell'autore si strutturi su *entrambi* i percorsi.

Un possibile equivoco va in ogni caso immediatamente risolto. Il fatto che il volume presenti sé stesso come “di base e divulgativo” o che si affermi programmaticamente che “il testo su supporto cartaceo *non conterrà neanche una formula*” (p. 11), non significa affatto che esso rinvii a un progetto di semplificazione o di banalizzazione “operativa” degli argomenti trattati. Al contrario, Di Franco ha costantemente cura di restituire la trattazione al necessario livello di complessità. *EDS: esplorare, descrivere e sintetizzare i dati*, in altri termini, per quanto voglia essere una *guida pratica all'analisi dei dati nella ricerca sociale*, non è in alcun modo un veloce compendio per “non specialisti”, che consenta o addirittura teorizzi “scorciatoie” di alcun tipo. E non solo sul piano delle procedure tecniche: in realtà la consapevolezza dei grandi “nodi” e delle difficoltà proprie del lungo e multiforme dibattito circa il *metodo* delle scienze sociali, pur se opportunamente non specificata – dal momento che si tratta pur sempre di un manuale di analisi dei dati – è costantemente presente nello sfondo. Guida *pratica* quindi, non certo nel senso di tendere più o meno sotteraneamente ad evitare, sottovalutare o aggirare la *teoria*, ponendosi rispetto ad essa come presunta alternativa autosufficiente, ma piuttosto nel senso di proporsi di avviare i lettori all'intervento concreto e competente rispetto a problemi di analisi concettualmente specificati: obiettivo, questo, che alla teoria pertiene a pieno titolo.



## *Introduzione*

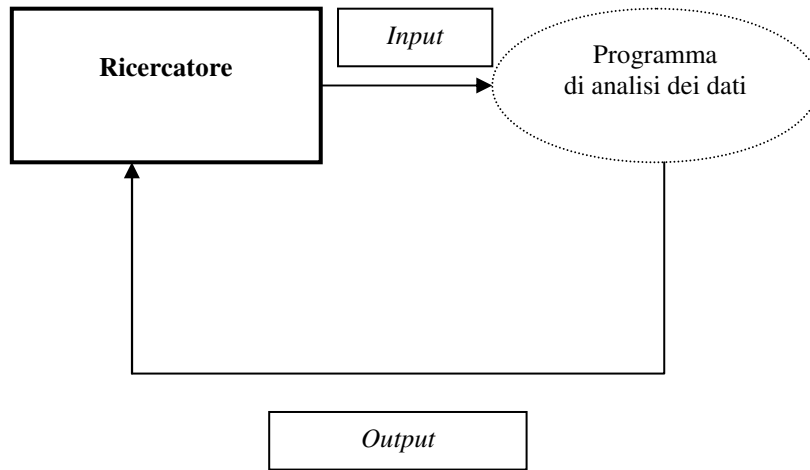
Questo libro, di base e divulgativo, affronta alcune procedure di analisi dei dati finalizzate all'esplorazione, descrizione e sintesi degli stessi illustrandone la logica e i passaggi operativi di tutte le fasi. Dalla progettazione e realizzazione di una matrice dei dati fino all'impiego di procedure di analisi multivariata di tipo esplorativo-descrittivo.

Il testo *non conterrà neanche una formula*. Assumiamo che il lettore conosca solo le quattro operazioni aritmetiche (somma, sottrazione, moltiplicazione e divisione) e le operazioni di elevazione al quadrato ed estrazione della radice quadrata di un numero. Si cercherà, quindi, di affrontare gli argomenti statistico-matematici in modo semantico, delegando ai programmi informatici l'onere della conoscenza degli algoritmi statistici e la loro corretta esecuzione. Per mantenere fede alla suddetta promessa, in alcuni casi, sarà necessario essere reticenti e/o evasivi su alcuni aspetti tecnici. Saranno comunque forniti ai lettori dei riferimenti bibliografici utili per approfondire gli aspetti tecnici dell'analisi dei dati. Tali riferimenti sono stati individuati, nella vasta letteratura disponibile in materia, preferendo i contributi di autori italiani rivolti ad un pubblico di medie competenze statistico-matematiche.

Obiettivo di questo libro è quindi porre l'attenzione sulla necessaria consapevolezza e conoscenza metodologica, prima ancora che statistica, di chi analizza i dati *prima, durante e dopo* l'uso di un programma di analisi dei dati.

Il modello proposto è costituito dalla coppia ricercatore / *software* di analisi dei dati e può essere così rappresentato: il ricercatore deve essere in grado di impostare i comandi sul programma di analisi in modo coerente in funzione dei suoi obiettivi (*input*) e deve poi essere in grado di valutare ed interpretare i risultati che il calcolatore gli restituisce (*output*).

Ambizione di questo lavoro è sollecitare nei lettori l'autoapprendimento che deriva dalla messa in pratica dei contenuti presentati. Si è soliti dire che solo l'applicazione pratica delle procedure di analisi dei dati ne consente un'assimilazione efficace; questo libro vuole proprio sollecitare l'esperienza diretta del lettore fornendogliene gli strumenti.



Per il loro contributo di osservazioni e di critiche esprimo il mio sentito grazie a Damiano Abbatini, Chiara Brunelli, Enzo Campelli, Alberto Marradi, Federica Pintaldi e Marta Simoni. Naturalmente, chi scrive è il solo responsabile di ogni eventuale errore presente nel libro.

## *1. Esplorare, Descrivere e Sintetizzare i dati nella ricerca sociale*

### **1.1. Introduzione**

Quando si illustra l'*iter* di una ricerca empirica, si è portati a presentarlo come un percorso lineare, nel quale le fasi si succedono ordinatamente l'una dopo l'altra; ma nella effettiva pratica di una qualsiasi ricerca, le diverse fasi non si susseguono in questo modo. Altri modi di rappresentazione potrebbero essere quello in circolo nel quale l'inizio e la fine si congiungono (Corbetta, 1999, p. 83), o un diagramma a blocchi con numerose frecce che collegano i diversi blocchi (Ricolfi, 1997, p. 26); ma anche in questi casi si tratterebbe di astrazioni. Forse la metafora migliore per rappresentare l'intero *iter* di una ricerca è una rete di reti, nella quale la macro-rete è disarticolabile in reti locali, che interagiscono sia tra loro sia con la macro-rete. Un tale modello però, se da un lato riesce a rendere bene l'idea della complessità, proprio per questo non è facilmente rappresentabile.

In genere, la letteratura è concorde sul fatto che la ricerca empirica nelle scienze sociali è finalizzata alla produzione di una conoscenza controllabile (pubblica, intersoggettiva e potenzialmente replicabile) che produce asserti e talvolta propone spiegazioni giustificandole su base empirica.

L'*iter* di una ricerca empirica può essere articolato su cinque diversi livelli, ciascuno dei quali può essere ulteriormente disarticolato; essi sono:

- 1) il disegno della ricerca;
- 2) la costruzione della base empirica;
- 3) l'organizzazione dei dati;
- 4) l'analisi dei dati;
- 5) l'esposizione dei risultati.

La successione tra questi cinque livelli non segue sempre e necessariamente un ordine lineare (Ricolfi, 1997).

Nel disegno della ricerca si definiscono gli obiettivi dell'indagine e si pianificano le strategie da attuare per il conseguimento degli stessi. In questa fase si cerca di ridurre le domande generali in domande più specifiche, articolando i concetti più generali in concetti più specifici e più facilmente definibili operativamente (Marradi, 1984).

La costruzione della base empirica consiste, innanzitutto, nel delimitare l'ambito spazio-temporale dell'indagine, definendo l'unità di analisi e le sue proprietà interessanti per gli obiettivi prefissati, e le procedure attraverso le quali le relative informazioni saranno rilevate.

Nella fase dell'organizzazione dei dati avviene la trasformazione delle informazioni acquisite in dati. Si può propriamente parlare di dati, infatti, solo quando essi sono organizzati in una struttura che si chiama matrice dei dati (vedi cap. 2). Considerando i tipici procedimenti di produzione della base empirica nelle scienze sociali, possiamo definire i dati come "informazioni interpretate"; essi sono tali perché sono inseriti in un contesto che li organizza e assegna loro un senso. A rigore, anche le informazioni sono il risultato di un precedente lavoro di interpretazione, per cui si dovrebbe più correttamente dire che i dati sono informazioni re-interpretate.

La fase dell'organizzazione dei dati, con la trasformazione delle informazioni in dati, che spesso conduce alla costruzione di una matrice dei dati, costituisce il principale punto di distinzione tra il modo cosiddetto "standard" e i modi cosiddetti "non standard" della ricerca sociale (Marradi, 1996a; Ricolfi, 1997). Questo nonostante il fatto che negli ultimi anni anche nelle ricerche "non standard" si dedichi molta attenzione a questa fase (con un sempre più intensivo uso di tecnologie informatiche e di altro tipo che consentono un trattamento meno personale, rispetto al precedente modo di procedere, delle informazioni), per cui diviene sempre più difficile individuare delle distinzioni "forti" sul piano delle diverse procedure (Cipriani e Bolasco, 1995; Ricolfi, 1997).

L'analisi dei dati è costituita da quell'insieme di operazioni, con un uso più o meno intenso di strumenti statistici, attraverso le quali si producono asserti e spiegazioni che riguardano i dati stessi. Il modo tipico della ricerca "standard", ossia di quella ricerca che costruisce e analizza una matrice dei dati con le tecniche statistiche, nella produzione di asserti e di nessi tra asserti consiste nell'adozione di procedure impersonali e automatiche. Ciò comporta il fatto che due ricercatori diversi che analizzino la stessa matrice dei dati con le stesse tecniche statistiche che coinvolgono le stesse variabili e gli stessi casi, dovrebbero produrre gli stessi asserti. A dire il vero, l'impersonalità e l'automatismo nella produzione degli asserti costituiscono un'idealizzazione, un *wishful thinking* (il credere vero qualcosa perché lo si desidera intensamente) e un modo di presentare le procedure standard di analisi dei dati, più che un resoconto fedele delle effettive procedure di tale analisi.

Infine, l'ultima fase consiste nell'esposizione dei risultati; essa dovrebbe avere la funzione di rendicontare nel modo più esauriente possibile l'intero *iter* della ricerca, di comunicare i principali risultati ottenuti con l'analisi dei dati anche in riferimento alla letteratura precedente ed, eventualmente, di suggerire nuove ipotesi di lavoro per future ricerche sullo stesso oggetto.

Nella fig. 1.1 si presenta l'articolazione dei cinque livelli, usando uno schema tipico dei diagrammi a blocchi.

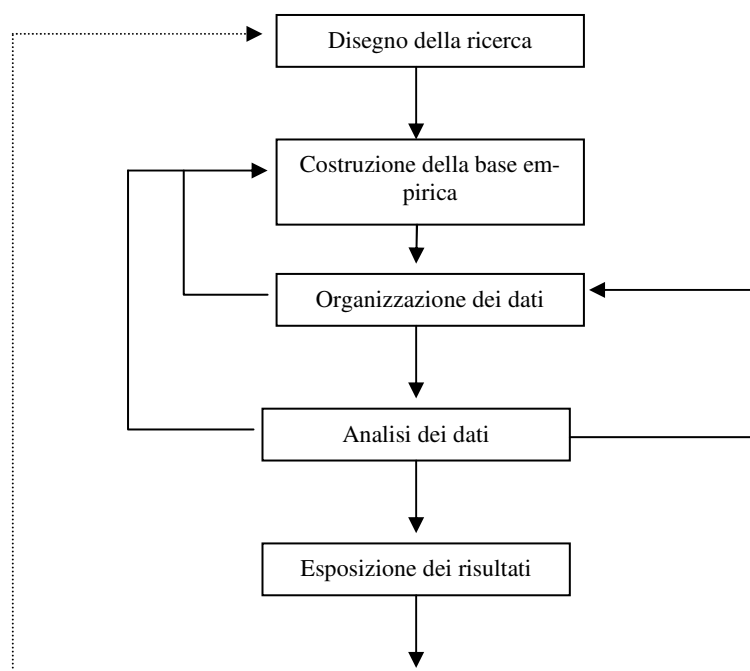


Figura 1.1 – L'iter di una ricerca empirica nelle scienze sociali (nostro adattamento da Ricolfi, 1997, p. 26)

## 1.2. Definizione e funzioni dell'analisi EDS (Esplorativa, Descrittiva, Sintetica)

Si è soliti distinguere, in materia di analisi dei dati nelle ricerche sociologiche che si caratterizzano per l'uso della matrice dei dati, tra due approcci. Il primo muove da un sistema di ipotesi circa le relazioni tra le variabili — nel quale il ricercatore è in grado di individuare, selezionare e connettere le variabili in base al suo patrimonio di conoscenze — che si vogliono sottoporre a controllo empirico. Quindi l'obiettivo consiste nel valutare se le ipotesi speci-

ficcate a monte della ricerca sono compatibili con i dati a disposizione. I procedimenti di elaborazione statistica in questo caso vengono detti modelli di analisi dei dati (Capecchi, 1967; Ricolfi, 1992; Di Franco, 1997) e adottano le tecniche dell'inferenza statistica, in quanto il ricercatore è interessato alla generalizzazione dei risultati della ricerca al di là del campione esaminato.

Il secondo approccio, che costituisce l'oggetto di questo lavoro, è invece finalizzato all'esplorazione, alla descrizione e alla sintesi dei dati (d'ora in poi EDS). In questo caso, l'analista non deve formulare a monte delle ipotesi circa la/le relazione/i tra le variabili oggetto di studio; queste, caso mai, dovrebbero costituire l'esito dell'analisi dei dati. In questo ambito l'atteggiamento prevalente del ricercatore è costituito da un'umile disponibilità ad accogliere risultanze empiriche non prevedibili a priori. Gli strumenti di analisi dei dati idonei per gli obiettivi di questo approccio sono usati in funzione tecnica<sup>1</sup>.

Abbiamo presentato i due approcci seguendo l'ordine della loro comparsa nella ricerca sociale. Solo con la diffusione dei *personal computers*, a partire dalla fine degli anni '60, è stato infatti possibile trattare simultaneamente un numero elevato di variabili. L'enorme incremento delle possibilità di calcolo elettronico ha sollecitato lo sviluppo di molti procedimenti di esplorazione dei dati. Tra i promotori di questo approccio possiamo ricordare il francese Ben-zécri (1969; 1972; 1973), fondatore della scuola detta *analyse des données*, e l'americano Tukey (1977), padre della cosiddetta *data analysis*.

È evidente la complementarità tra i due approcci e i vantaggi che derivano dall'adozione di strategie integrate. Inoltre, è bene chiarire che la distinzione tra modelli e tecniche di analisi dei dati non riguarda tanto i procedimenti (ad esempio, tra i primi di solito si classificano la regressione multipla, l'analisi fattoriale classica, i modelli di equazioni strutturali e quelli loglineari; tra i secondi l'analisi in componenti principali, l'analisi delle corrispondenze multiple e le tecniche di classificazione automatica) quanto l'uso che ne viene fatto. Per cui, più correttamente si dovrebbe distinguere tra un approccio esplorativo e un approccio confermativo-modellistico all'analisi dei dati, indipendentemente da quali strumenti in questi contesti vengono adoperati.

Come detto, in questo testo presenteremo solo alcune tecniche di analisi dei dati particolarmente utili per l'approccio EDS. Questa scelta si giustifica, coerentemente con l'obiettivo didattico del libro, citando un brano di Duncan:

1. In letteratura si usano spesso altre coppie concettuali come 'approccio inferenziale' e 'approccio esplorativo-descrittivo'; 'analisi statistica di scuola anglosassone' e 'analisi statistica di scuola francese'; 'contesto della scoperta' e 'contesto della giustificazione'; 'descrizione' e 'spiegazione'. In tutti questi casi, anche se su differenti livelli di generalità, si ripropone con termini diversi la differenza che noi abbiamo etichettato 'approccio modellistico' e 'approccio tecnico' all'analisi dei dati. Recentemente alcuni autori (Fraire, 1994; Bolasco, 1999) ripropongono la distinzione tra i due approcci differenziando l'analisi 'multivariata' e l'analisi 'multidimensionale'; con la prima si intende un approccio modellistico all'analisi dei dati; con la seconda si definisce un approccio EDS.



Uno può anche compiere un passabile lavoro nel primo settore [cioè quello delle tecniche EDS], pur senza sapere molto della materia, e ne fanno fede i numerosi specialisti in analisi statistica dei dati che non hanno interessi concreti in alcun settore; ma uno non può nemmeno cominciare nel secondo settore [quello dei modelli di analisi dei dati] senza una solida conoscenza della teoria scientifica rilevante in quel campo, poiché il punto di partenza è, per l'appunto, il modello e non il metodo statistico (Duncan, 1975, p. 6).

La citazione appena riportata potrebbe far ritenere l'approccio EDS alla portata di ricercatori quasi neofiti dell'analisi dei dati; al contrario l'approccio confermativo all'analisi dei dati sarebbe appannaggio dei soli ricercatori esperti. In un certo senso ciò può essere condivisibile, a patto di non sottovalutare il ruolo del ricercatore nella fase esplorativa dell'analisi dei dati. A questo proposito sono illuminanti le parole di Tukey:

L'analisi esplorativa dei dati è un lavoro da detective, un detective dei numeri, del conteggio e dei grafici. Per fare il suo lavoro il detective ha bisogno sia di strumenti che di 'comprensione'. Se non ha polvere per le impronte digitali, non ne potrà raccogliere, e, contemporaneamente, se non ha pensato a cosa il criminale può avere toccato, non guarderà mai nel posto giusto. Allo stesso modo l'analisi dei dati necessita sia di strumenti che di 'comprensione'. [...] I processi sono nettamente divisi tra la ricerca delle prove (di cui si occupa la polizia, gli investigatori, etc.), e la valutazione delle prove stesse (il lavoro di giudici e giurie). L'analisi esplorativa è un lavoro da detective, l'analisi confermativa è un lavoro da giudici. Finché il detective non trova degli indizi, giudice e giuria non hanno nulla su cui lavorare. Allo stesso modo finché un'indagine esplorativa non trova 'indicazioni', generalmente a carattere quantitativo, non c'è nulla su cui lavorare per un'analisi confermativa. Gli esperimenti e alcuni progetti d'indagine costituiscono alcune eccezioni, o meglio, alcune parziali eccezioni di questa regola. Ma ciò è possibile perché una certa parte dell'analisi dei dati è stata pianificata come parte dell'esperimento o dell'inchiesta. Ma facendo questo imponiamo all'analisi una restrizione – cosa che contraddice gli intenti – che rischia di non farci vedere dei risultati che potrebbero essere molto interessanti. Come tutte le storie di detectives ci insegnano, molte circostanze che circondano il crimine sono accidentali ed ingannevoli. Allo stesso modo, molte delle indicazioni che si svelano dalla base dei dati sono accidentali ed ingannevoli. Accettare tutte le apparenze come prove definitive sarebbe folle, sia per un detective sia per un ricercatore. Ma, al contempo, non raccogliere tutte le apparenze perché alcune – ma potrebbero essere persino tutte – sono accidentali, è un abuso di autorità che dovrebbe essere (e spesso lo è) punito appropriatamente. *L'analisi esplorativa non sarà mai l'intera storia, ma niente altro può essere così importante come fondamento, come primo passo* (Tukey, 1977, pp. 1-3, corsivi nostri).

Questo bel brano rende giustizia all'analisi EDS sottolineandone il ruolo e la funzione nell'ambito di una visione matura per le esigenze della ricerca so-

ciale. Tukey e Wilk (1970, p. 372) sintetizzano così gli obiettivi dell'approccio esplorativo:

1. acquisire una descrizione più analitica di ciò che si conosce superficialmente o che si intuisce;
2. trovare risultati inattesi, e suggerire modelli inattesi per la sintesi e l'esposizione dei dati;
3. utilizzare i risultati dell'analisi per stabilire l'adattamento di un certo sistema di ipotesi formulate a priori;
4. fornire stimoli e indicazioni per future ricerche;
5. stimolare la creatività del ricercatore mentre analizza i dati e valuta i passi successivi per l'analisi degli stessi.

A questo proposito, è doveroso tenere presente le specifiche caratteristiche delle ricerche empiriche nelle scienze sociali. In esse le tecniche di costruzione dei dati presentano modalità per le quali il ricercatore è fortemente coinvolto. Anche se molti sociologi continuano ad usare un linguaggio che risente dell'impostazione comportamentista<sup>2</sup>, parlando di dati "osservati"<sup>3</sup> o peggio di dati "oggettivi"<sup>4</sup>, è doveroso dichiarare e rendicontare nei rapporti di ricerca come si è pervenuti, ossia attraverso quali scelte di natura sia teorico-concettuale sia tecnica, alla costruzione dei dati. Al contrario, quando si illustrano i risultati delle analisi dei dati, si tende spesso a trascurare, o dare per scontato, tutto quello che viene fatto a monte della costruzione della matrice

2. Usando una metafora, la prospettiva comportamentista nelle indagini sociologiche considera l'essere umano oggetto di studio come una specie di automa (banca dati) che a richiesta fornisce tutte le informazioni in modo attendibile. Questo assunto potrà sembrare eccessivo e soprattutto irrealistico, ma è proprio quello che serve per poter conferire la massima fiducia nella qualità (validità e attendibilità) dei dati. Comunque, dopo un secolo di ricerca sociale, la sostenibilità di questo assunto è del tutto tramontata anche se non è scomparsa l'ideologia comportamentista. Infatti, di recente, nell'ambito degli studi di marketing, l'atteggiamento comportamentista ha risolto in modo coerente il problema della comune natura umana tra il ricercatore e il suo oggetto di studio. Da qualche tempo, infatti, si sta sviluppando la cosiddetta "etologia del consumatore", che consiste nell'analisi del comportamento dei consumatori nei grandi magazzini solo attraverso le registrazioni audiovisive (sfruttando le telecamere installate per i sistemi di sicurezza). In questo modo è stato risolto il fastidio di dover interagire con i poco *attendibili* e troppo 'umani' intervistati.

3. A nostro avviso si può parlare di osservazione solo quando un osservatore rileva i dati direttamente o attraverso degli appositi strumenti. Nella ricerca sociale quasi sempre, ovvero tutte le volte che si raccolgono le informazioni attraverso interviste con o senza questionario, i dati sono prodotti attraverso una collaborazione, una relazione, una mediazione tra chi formula delle domande e chi fornisce delle risposte. Su questo punto particolare si rimanda ai contributi di Pitrone (1986), Palumbo (1995), Guala (1996) e Gobo (1997); in generale sulle procedure di costruzione dei dati nella ricerca sociale alle raccolte curate da Marradi (1988) e Marradi e Gasperoni (1995).

4. L'ansia di apparire il più possibile oggettivi, ossia di aderire il più possibile alla realtà, non è una prerogativa esclusiva del comportamentismo. Essa si ritrova tale e quale anche in certi approcci cosiddetti "qualitativi" o "non standard" quando si pretende di fornire oggettivamente, ovvero senza alcuna interferenza da parte del ricercatore, i dati empirici.

dei dati. Questo atteggiamento spesso induce a ritenere che l'analisi dei dati sia il luogo della "precisione statistica" e della totale impersonalità, dove con virtuosismi tecnici si riesce a risolvere ogni problema o a trovare una risposta ad ogni domanda. In realtà, come giustamente afferma Marradi:

Ogni risultato scientifico è frutto di una serie di scelte e decisioni tecniche, che il ricercatore compie in un senso o nell'altro sulla base della sua preparazione, ma anche del suo intuito. Presentare tali risultati come il naturale sbocco di un procedimento codificato e impersonale è dare un'immagine totalmente mistificata del procedimento scientifico: è ideologia della scienza. Le decisioni tecniche sono soggettive, e da esse dipendono in buona misura i risultati. Ogni volta che la mia soggettività interviene in modo rilevante, userò la prima persona singolare per ricordarlo al lettore e assumerne la responsabilità (Marradi, 1976, p. 292).

Giunti a questo punto, riteniamo di aver esposto le ragioni per cui in questo libro, rivolto in primo luogo a lettori principianti in analisi dei dati, tratteremo esclusivamente dell'approccio EDS; le riassumiamo brevemente:

- costituisce in ogni caso il primo passo di ogni elaborazione dei dati;
- richiede al ricercatore che intende avvalersene un carico minore, anche se non certo nullo, di conoscenze e di esperienza;
- data la natura degli obiettivi, questo approccio si presta facilmente ad una presentazione non formalizzata.

Passiamo ora a definire con la maggiore precisione possibile cosa intendiamo con i termini 'esplorazione', 'descrizione' e 'sintesi' quando essi sono applicati alle procedure di analisi dei dati.

Per 'esplorare i dati' si intende operare con l'intento di scoprire qualcosa che non è evidente o non è noto. Usando una metafora (Marradi, 1996b), è come quando un esploratore entra in una foresta vergine e in questa apre un primo sentiero. Non è detto che questo primo sentiero sia quello più breve o quello migliore, tra tutti i possibili, per attraversare la foresta; in ogni caso, un secondo esploratore, o lo stesso primo esploratore in un secondo viaggio, potrà avvalersi del primo sentiero aperto. Fuori di metafora, un ricercatore, ad esempio, usando uno strumento come il microscopio non descrive semplicemente quello che osserva, ma esplora il suo oggetto di studio per individuarne le caratteristiche che al momento gli sono sconosciute.

La descrizione dei dati consiste nel trovare alcuni valori caratteristici che rappresentino la/le distribuzione/i di una o più variabili. Quindi, se la variabile in questione è ad esempio l'età degli intervistati in un campione di mille individui, si descriverà la distribuzione di questa variabile calcolandone alcuni valori caratteristici come le misure della tendenza centrale (media, moda e mediana), le misure di dispersione (il minimo e il massimo, il campo di variazione, lo scarto-tipo e la varianza), oppure rappresentando graficamente con un istogramma la forma della distribuzione. Queste operazioni verranno dettagliatamente descritte nel cap. 4.

La sintesi dei dati è necessaria quando le variabili e i casi sono molti, e diviene indispensabile ridurne la quantità in modo ottimale avvalendosi di strumenti, per lo più multivariati come l'analisi delle componenti principali, l'analisi delle corrispondenze multiple o l'analisi dei gruppi (vedi cap. 6), che producono benefici ma hanno anche i costi inevitabili di qualsiasi operazione di sintesi. Questo è quasi sempre il caso di ricerche svolte con la tecnica dell'inchiesta (con unità di analisi individuale e ambito di solito piuttosto ampio come la nazione o la regione) in cui si costruiscono matrici dei dati con mille-duemila casi e cento-duecento variabili.

Abbiamo definito separatamente le funzioni dell'esplorazione, della descrizione e della sintesi solo per comodità espositiva; ma esse non sono così nettamente separabili. Una descrizione è in qualche modo già una sintesi ed entrambe derivano da un momento esplorativo.

Riassumendo, l'approccio EDS consente un'esplorazione-descrizione-sintesi della matrice dei dati (casi per variabili; vedi cap. 2) che spesso si avvale di rappresentazioni grafiche. I dati possono essere analizzati avendo diversi obiettivi: la ricerca delle somiglianze o delle differenze tra i casi (gli individui si somigliano quando i profili<sup>5</sup> delle modalità assunte sulle variabili sono simili), la classificazione dei casi in un numero ristretto di gruppi omogenei, la descrizione sintetica delle relazioni tra le variabili, etc.

In un'indagine progettata con finalità prevalentemente descrittivo-esplorative, si avranno ipotesi di ricerca non eccessivamente formalizzate e specificate oppure addirittura implicite<sup>6</sup>; il raffinamento delle ipotesi di partenza e la formulazione di nuove ipotesi più specifiche circa l'oggetto di studio costituirà l'esito di una tale ricerca.

In molti settori disciplinari sociologici nei quali si effettuano ricerche empiriche, non si dispone di conoscenze teoriche sufficienti per l'impostazione di un disegno della ricerca esclusivamente confermativo; per questa ragione le ricerche o alcuni suoi aspetti che seguono l'approccio EDS sono le più diffuse. In ogni caso ci sono ragioni pratiche, soprattutto di natura economica, che fanno sì che anche nei rari casi in cui si possa svolgere una ricerca seguendo l'approccio confermativo, ci si cauti ampliando la raccolta dei dati anche ad altre variabili che non rientrano nelle ipotesi che si intendono sottoporre a controllo empirico. In questo modo sarà possibile, in fase di analisi dei dati, valutare ipotesi alternative rispetto a quelle formulate a monte della raccolta dei dati.

5. Il profilo di un caso consiste nella sequenza dei suoi stati sulle proprietà rilevate. Ad esempio, il profilo del caso X potrebbe essere: maschio sulla variabile 'genere', cattolico sulla variabile 'confessione religiosa', e così via per tutte le altre variabili rilevate nell'ordine in cui esse appaiono nella matrice.

6. Peraltro, uno studio descrittivo-esplorativo di un fenomeno sociale richiede quanto meno la selezione di quali aspetti e caratteristiche del fenomeno oggetto di studio debbano essere prese in considerazione: la "mappa dei concetti" (Marradi, 1994).