



## Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con



La versione completa dell'e-book (a pagamento) è leggibile con Adobe Digital Editions. Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.



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

Direttore

*Giovanni Di Franco*, Università di Roma “La Sapienza”

Comitato editoriale

*Elena Battaglini*, Ires-Cgil

*Sara Bentivegna*, Università di Roma “La Sapienza”

*Claudio Bezzi*, valutatore professionista

*Alberto Marradi*, Università di Firenze

*Federica Pintaldi*, Istat

*Luciana Quattrocioni*, 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](http://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”.

Giovanni Di Franco

# **TECNICHE E MODELLI DI ANALISI MULTIVARIATA**

**Nuova edizione ampliata e aggiornata**

**La cassetta degli attrezzi  
Strumenti per le scienze umane/8**

**FrancoAngeli**

Progetto grafico di copertina di Maria Teresa Pizzetti

2<sup>a</sup> edizione. Copyright © 2011, 2017 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](http://www.francoangeli.it).*

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

### Volumi pubblicati:

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. Teoria e pratica.*
5. Lucia Coppola, *NVivo: un programma per l'analisi qualitativa.*
6. Simone Gabbriellini, *Simulare meccanismi sociali con NetLogo. Una introduzione.*
7. Giovanni Di Franco, *Dalla matrice dei dati all'analisi trivariata. Introduzione all'analisi dei dati.*
8. Giovanni Di Franco, *Tecniche e modelli di analisi multivariata.*
9. Federica Pintaldi, *Come si interpretano gli indici internazionali. Istruzioni per l'uso a favore di ricercatori, giornalisti e politici.*
10. Maria Paola Faggiano, *Gli usi della tipologia nella ricerca empirica.*
11. Danilo Catania, *Dati e rappresentazioni territoriali con ArcGis.*
12. Claudio Bezzi, *Fare ricerca con i gruppi. Guida all'utilizzo di focus group, brainstorming, Delphi e altre tecniche.*
13. Giovanni Di Franco (a cura di), *Il poliedro coesione sociale. Analisi teorica ed empirica di un concetto.*
14. Elena Battaglini, *Sviluppo territoriale. Dal disegno della ricerca alla valutazione dei risultati.*
15. Claudio Bezzi, *Domanda e ti sarà risposto. Costruire e gestire il questionario nella ricerca sociale.*
16. Elena Pavan, *La Network Analysis con Nodexl*
17. Maria Paola Faggiano, *L'analisi del contenuto di oggi e di ieri. Testi e contesti on e offline.*

### OPEN ACCESS

- 10119.1. Giovanni Di Franco, *Giovani a tempo indeterminato. Valori e atteggiamenti dei giovani romani.*





# Indice

<b>1. Introduzione</b>	pag.	9
1.1 La nuova edizione di <i>Tecniche e modelli di analisi multivariata</i>	»	9
1.2 Cosa leggere per saperne di più	»	18
<b>2. Progettare l'analisi multivariata dei dati</b>	»	22
2.1 Definizione e funzioni dell'analisi multivariata	»	22
2.2 Le fasi dell'analisi multivariata	»	29
2.3 Il linguaggio dell'analisi multivariata	»	32
2.4 Natura delle variabili, tecniche e modelli compatibili	»	37
2.5 Cosa leggere per saperne di più	»	43
<b>3. Le tecniche per variabili cardinali</b>	»	45
3.1 Analisi in componenti principali	»	46
3.2 Analisi dei gruppi	»	87
3.3 Analisi discriminante	»	117
3.4 Tecniche di scaling multidimensionale	»	135
3.5 Cosa leggere per saperne di più	»	139
<b>4. Le tecniche per variabili categoriali</b>	»	142
4.1 Analisi delle corrispondenze binarie	»	142
4.2 Analisi delle corrispondenze multiple	»	158
4.3 Analisi delle corrispondenze con scelta delle modalità attive	»	187
4.4 Analisi delle corrispondenze condizionate	»	193
4.5 Analisi delle corrispondenze lessicali	»	197
4.6 Cosa leggere per saperne di più	»	207

<b>5. I modelli per variabili cardinali</b>	pag.	209
5.1 Regressione multipla lineare	»	211
5.2 Analisi fattoriale	»	231
5.3 Modelli di equazioni strutturali	»	239
5.4 Cosa leggere per saperne di più	»	271
<b>6. I modelli per variabili categoriali</b>	»	275
6.1 Modelli log-lineari	»	275
6.2 Regressione logistica con variabile dipendente dicotomica	»	312
6.3 Regressione logistica con variabile dipendente politomica	»	320
6.4 Regressione logistica con variabile dipendente ordinale	»	324
6.5 Analisi della varianza a più vie	»	328
6.6 Cosa leggere per saperne di più	»	340

# 1. Introduzione

## 1.1 La nuova edizione di *Tecniche e modelli di analisi multivariata*

Dal 2011, anno di pubblicazione della prima edizione di *Tecniche e modelli di analisi multivariata* (Di Franco 2011b), in Italia e nel resto del mondo, è notevolmente aumentato l'uso di procedimenti di analisi multivariata nella produzione scientifica dei ricercatori di molte discipline delle scienze umane. Le ragioni di tale aumento sono dovute a diversi fattori. In primo luogo segnaliamo la forte influenza della cosiddetta scientometria, ossia quella disciplina che pretende di certificare la qualità di una pubblicazione ricorrendo a indicatori di natura bibliometrica. Le recenti riforme in ambito di concorsi universitari, come l'abilitazione scientifica nazionale o la valutazione della qualità della ricerca, hanno imposto dei criteri – chiamati a volte indicatori, a volte soglie – che si ispirano ad astratti standard di natura quantitativa di discutibile validità e attendibilità.

Il principale effetto perverso di questa tendenza è quello di scambiare il contenuto per il contenitore. Così per valutare la qualità di un articolo scientifico si valuta la quantità di citazioni che ottiene la rivista che lo ha pubblicato (in termini di *impact factor* che a sua volta determina la posizione di rango della rivista nella cosiddetta classifica dell'eccellenza scientifica di un determinato settore: riviste di fascia A; di fascia B e di fascia C). A loro volta le riviste cosiddette scientifiche, per consolidare la posizione nella graduatoria, impongono agli autori di rispettare alcuni standard nelle loro pubblicazioni in modo da mantenere e possibilmente migliorare la posizione nella classifica. Sempre più

spesso fra questi standard è previsto l'uso di tecniche quantitative sempre più sofisticate e alla moda. Addirittura in certi casi la rivista indica agli autori quali coefficienti di significatività statistica o quali coefficienti di bontà di adattamento dei modelli adottati debbono essere usati. Se il ricercatore opta per altri coefficienti, altrettanto legittimi sul piano scientifico rispetto a quelli indicati dalla rivista, rischia la bocciatura del suo articolo. Così la valutazione della qualità di una pubblicazione è prevalentemente, se non totalmente, centrata sul controllo della forma e del rispetto di criteri che non hanno altra ragione d'essere se non quella di essere indicati come gli standard di qualità scientifica dalla rivista stessa.

Il secondo effetto perverso, connesso al precedente, consiste nella forte pressione al conformismo: i ricercatori per fare carriera hanno interesse a pubblicare molti articoli su riviste di fascia A del loro settore disciplinare (preferendo possibilmente quelle in lingua inglese visto l'annoso provincialismo italiano), e per questo sono indotti a seguire pedissequamente le richieste che massimizzano la probabilità di pubblicazione. Per raggiungere l'obiettivo basta seguire le mode del momento e conformarsi alle richieste di rispetto dogmatico dei criteri di presunta scientificità delle pubblicazioni. Appena si mette un piede fuori dal solco tracciato dal *main stream* si rischia di essere esclusi dal circuito dell'eccellenza, dovendo ripiegare su riviste meno accreditate per sperare di vedere pubblicato il proprio lavoro. Nella situazione attuale ricercatori vincitori di premi Nobel come, ad esempio, Peter Higgs e Albert Einstein rischierebbero la bocciatura dei loro fondamentali articoli sul bosone e sulla teoria della relatività, perché troppo difformi dagli standard attualmente in voga.

Una seconda ragione della diffusione dei procedimenti di analisi multivariata è dovuta al numero crescente di programmi informatici oggi disponibili sul mercato. Fra questi alcuni sono d'uso estremamente semplice poiché consentono all'utente di impostare un'analisi complessa scegliendo acriticamente la *routine* di un programma di analisi dei dati accettando così i comandi di *default* che sono stati impostati dai programmatori. Ricevuto questo *input*, il programma analizza in automatico i dati e fornisce in pochi secondi i risultati dell'analisi.

La notevole semplificazione prodotta dai programmi di analisi dei dati, da un lato, ha agevolato lo sviluppo delle applicazioni dei proce-

dimenti di analisi multivariata, dall'altro, ha determinato un abbassamento qualitativo di una grande parte di esse (vedi oltre).

In realtà il principale vantaggio del calcolo elettronico e dei programmi di analisi automatica dei dati è quello di aver reso possibile l'applicazione dei modelli e delle tecniche di analisi multivariata dei dati a matrici di grandi dimensioni per cui, anche se una buona parte di questi modelli e tecniche sono disponibili da molti anni (in diversi casi anche da più di un secolo), il loro concreto uso nella ricerca è reso possibile solo dagli strumenti informatici che sono in grado di effettuare in pochi secondi miliardi di operazioni matematiche che, se dovessero essere fatte a mano, richiederebbero molti mesi o anni di lavoro a valenti matematici.

In modo analogo a quanto è accaduto in passato per altri procedimenti di analisi dei dati – ad esempio, l'analisi fattoriale e l'analisi in componenti principali negli anni '60 e '70 del secolo scorso (vedi parr. 3.1 e 5.2; Di Franco e Marradi 2003; 2013); i modelli log-lineari a partire dagli anni '70 (vedi par. 6.1; Di Franco 2003); la regressione logistica a partire dagli anni '80 (vedi parr. 6.2, 6.3, 6.4); etc. –, il crescente interesse/entusiasmo per i procedimenti di analisi multivariata è tuttavia connesso a un loro uso sempre più acritico e disinvolto, tanto da inficiare seriamente la qualità scientifica di queste applicazioni.

In questo senso, si deve ad Alberto Marradi (1985) un primo grido d'allarme. Nella presentazione di un numero monografico della *Rassegna Italiana di Sociologia* dal titolo *Strumenti concettuali per la raccolta e l'analisi dei dati*, egli affermava:

“Già la precedente rivoluzione dei calcolatori (che produsse contemporaneamente lo *hardware* capace di gestire i grandi *files* delle scienze sociali, e un *software* più abbordabile da utenti sociologi anziché informatici) aveva provocato un abbassamento del livello qualitativo medio delle ricerche – a giudizio non degli esponenti di un umanesimo di *élite*, ma di protagonisti di quella stessa rivoluzione, come Norman Nie, il creatore del linguaggio Spss. Migliaia di aspiranti ricercatori, infarinati alla bell'è meglio da scuole estive di tecniche statistiche, hanno correlato e fattor'analizzato tutto ciò che capitava loro a tiro, producendo tonnellate di *nonsense*, di cui non pochi quintali hanno trovato la via della pubblicazione” (Marradi 1985, p. 136).

A nostro avviso, oggi un ricercatore che intenda avvalersi di questi strumenti di analisi dei dati deve porsi due ordini di problemi tra loro interconnessi: il primo è relativo alla comprensione degli strumenti concettuali che stanno a monte delle suddette tecniche e modelli; il secondo – ovviamente dipendente dal primo – è il riuscire a discriminare tra la ormai vasta offerta nei diversi programmi informatici di analisi statistica dei dati che, solo apparentemente, offrono le stesse procedure, quello o quelli che offrono i maggiori vantaggi in riferimento agli obiettivi e alle necessità di una particolare elaborazione.

Naturalmente, non è necessario comprendere tutte le fasi o tutti i passaggi degli algoritmi matematici dei singoli strumenti; ma non si possono ignorare gli assunti e i vincoli che questi presuppongono nei dati.

Parallelamente allo sviluppo dell'informatica, che prevede l'utilizzo del *computer* da parte di utenti totalmente privi di alfabetizzazione informatica, anche i programmi informatici di analisi dei dati oggi disponibili permettono al più inesperto utente l'esecuzione di procedimenti di analisi dei dati molto complessi senza assumere da parte dello stesso un benché minimo controllo delle procedure e, nel peggiore dei casi, senza richiederli l'onere di effettuare alcuna scelta tra le molte possibili. In ambedue i casi non si devono confondere questi vantaggi o venderli come "oggettività" o come "verità scientifica", dimenticando i rischi e le distorsioni che questi mezzi statistico-informatici comportano, soprattutto se nel loro utilizzo non vi è un controllo concettuale delle procedure e una giusta dose di buon senso. In particolare l'utilizzo di tali procedimenti di analisi prevede che l'utente sappia, quanto meno, valutare i risultati forniti dal calcolatore. Questo, spesso, significa saper navigare in un mare di cifre e di coefficienti statistici nel quale si rischia di rimanere sommersi. Non a caso, la migliore prova di quanto stiamo affermando è la sensazione di sgomento che gli utenti inesperti provano quando il computer fornisce i risultati delle procedure di analisi che rimangono del tutto incomprensibili. Allora si tocca con mano la distanza che separa l'uso di un programma di analisi dei dati e la competenza metodologica che l'utente deve possedere per sfruttare al meglio le possibilità della semplificazione procedurale offerta dai programmi informatici.

Recentemente, alcuni autori, esaminando diverse pubblicazioni su riviste scientifiche, hanno riscontrato che nella maggioranza dei casi è presente almeno un grave difetto che ne compromette il valore scientifico. MacCallum e Austin (2000), recensendo circa 500 applicazioni dei modelli di equazioni strutturali (vedi par. 5.3; Di Franco 2016) in sedici riviste di psicologia, hanno trovato che in circa la metà degli articoli era incompleto il riferimento alle stime dei parametri (ad esempio, erano state omesse le stime non standardizzate dei parametri); in un articolo su quattro non era stato indicato il tipo di matrice analizzata (se si trattava, nello specifico, di una matrice di correlazioni o una di varianze e covarianze); in circa il 10% degli articoli non veniva chiaramente descritto il modello specificato o gli indicatori delle dimensioni latenti (fattori).

Shah e Goldstein (2006) hanno recensito 93 articoli pubblicati su quattro riviste di economia trovando che nella maggioranza degli articoli era difficile determinare il modello realmente controllato e/o l'elenco completo delle variabili che lo costituivano. Inoltre, in 31 di 143 analisi il modello descritto nel testo non corrispondeva ai risultati statistici riportati nei grafici o nelle tabelle. Infine, in circa la metà degli articoli non era esplicitato il metodo usato per la stima dei parametri.

Entrambe le coppie di autori delle rassegne appena descritte hanno riscontrato problemi analoghi nelle applicazioni degli psicologi e degli economisti. In circa il 20% degli articoli esaminati da MacCallum e Austin (2000) le analisi utilizzavano campioni inferiori a 100 casi; Shah e Goldstein (2006) hanno riscontrato lo stesso problema nel 30% circa degli articoli. Gli autori degli articoli esaminati non hanno preso in considerazione modelli alternativi a quello oggetto della pubblicazione che avrebbero potuto essere ugualmente compatibili con i dati empirici (questi modelli alternativi sono definiti modelli equivalenti o quasi equivalenti). Ignorare i modelli equivalenti è una forma di distorsione confermativa tale per cui i ricercatori controllano un solo modello, dandone una valutazione eccessivamente positiva, e non riescono a prendere in considerazione altre possibili spiegazioni degli stessi dati (Shah e Goldstein 2006). Il potenziale della distorsione confermativa è ulteriormente amplificato dalla relativa mancanza di repliche. A questo proposito, la maggior parte delle pubblicazioni sono costituite da anali-

si di un singolo modello, senza quindi che vi siano procedure di convalida incrociata su campioni indipendenti della stessa popolazione o un approccio *split-sample* che divide il campione in due sotto-campioni equivalenti. La necessità nelle ricerche di disporre di campioni di numerosità elevata, pur costituendo un ostacolo per la replicabilità delle analisi condotte dai ricercatori, andrebbe comunque sostenuta con più vigore di quanto si faccia attualmente.

L'uso disinvolto e acritico dei procedimenti di analisi multivariata non è una caratteristica esclusiva degli economisti o degli psicometrici, come hanno dimostrato gli studi di MacCallum e Austin (2000) e di Shah e Goldstein (2006). Attribuire capacità taumaturgiche, “oggettive” e “scientifiche” ai risultati ottenuti attraverso l'uso di strumenti matematici è inversamente proporzionale al grado di consapevolezza metodologica dei ricercatori che ne fanno uso.

Paradossalmente, per le scienze umane, si può affermare che tanto meglio si conoscono gli assunti e i vincoli che uno specifico modello di analisi dei dati impone all'analisi, tanto meno si è portati ad usarlo (Di Franco 1997; 2003; 2006; 2011b; 2016).

Le basi di una buona analisi dei dati si gettano nel momento in cui si progetta una ricerca su un determinato oggetto, fenomeno o problema sociale, si formulano le ipotesi, si definisce il significato dei concetti rilevanti per l'indagine, si delinea la rappresentazione, in termini di procedimenti di operativizzazione, del o dei fenomeni che si intendono studiare, sia in termini spaziali e temporali sia in termini di indicatori e variabili che dovranno essere rilevate su una determinata unità di analisi, e così via.

Considerare l'analisi dei dati come la fase terminale del percorso di ricerca, eventualmente, da appaltare a qualche specialista o mago del *computer* al quale si chiede soltanto di applicare, che sia possibile o no non importa, qualche sofisticata tecnica multivariata non può produrre risultati rilevanti, e, spesso, produce solo *nonsense*.

A volte, per illustrare l'uso delle tecniche e dei modelli di analisi multivariata dei dati, si ricorre alla seguente analogia: *usare i procedimenti di analisi multivariata equivale a guidare una macchina; non occorre sapere come funziona il motore per essere degli ottimi automobilisti*. L'analogia è sbagliata in quanto il vero problema è quello di sapere effettuare la scelta più appropriata rispetto agli obiettivi prefissa-



ti: usare senza consapevolezza i procedimenti di analisi multivariata è come salire su un'automobile senza sapere quale strada si vuole percorrere e dove si vuole andare. In altri termini, il primo principio da non evadere mai è quello dell'indissolubilità fra teoria e ricerca, fra metodo e tecniche di ricerca.

I diversi problemi che si manifestano in tante applicazioni dell'analisi multivariata palesano la necessità di un libro che presenti in modo critico le diverse procedure soffermandosi in particolare sugli aspetti applicativi, evidenziando i tanti problemi che richiedono all'utente competenze diverse: sia di natura teorica e sostantiva di un dato ambito disciplinare, sia di natura epistemologica, metodologica, matematica e statistica. Obiettivo di questo libro non è proporre l'ennesimo manuale specialistico, bensì fornire ai lettori le conoscenze preliminari indispensabili per usare in modo critico – e perciò utile ed efficace – gli strumenti di analisi multivariata nelle proprie attività di ricerca.

Coerentemente con l'obiettivo sopra enunciato, il testo eviterà al massimo le formalizzazioni matematiche. I lettori se vorranno approfondire alcuni argomenti troveranno sia delle sezioni di testo in corpo arial e in carattere più piccolo sia alcuni riferimenti bibliografici alla fine di ciascun capitolo. Beninteso, evitare il formalismo matematico può suscitare nei lettori una percezione errata: l'uso consapevole degli strumenti di analisi multivariata è complesso anche prescindendo dagli aspetti matematici.

L'enorme semplificazione delle procedure che ormai caratterizza quasi tutti i programmi informatici illude molti utenti di essere in grado di usare l'analisi multivariata dei dati senza avere le conoscenze necessarie. Il suo uso richiede agli utenti delle qualità e capacità che non possono essere delegate ai programmi informatici, con tutto il loro armamentario di *app* e *gadget*. Oltre che di competenze teoriche e metodologiche, occorrono capacità di concentrazione, pazienza, nonché molto tempo a disposizione. Da questo punto di vista, la crescente semplicità dei programmi informatici si può considerare un difetto assai più che un pregio.

Usare un qualsiasi strumento di analisi dei dati non elimina gli ostacoli che si incontrano e i problemi inattesi che caratterizzano qualsiasi ricerca empirica. Il ricercatore deve trovare delle soluzioni che con-

sentano di condurre in porto il suo progetto di analisi che include la realizzazione di un modello capace di fornire delle informazioni utili sull'oggetto di studio.

Per ciascun procedimento di analisi multivariata si presentano i fondamenti concettuali; gli strumenti procedurali, anche in riferimento ai programmi informatici di analisi dei dati; e gli aspetti applicativi attraverso l'illustrazione di esempi tratti da ricerche empiriche condotte dall'autore o reperibili in letteratura.

Il nostro auspicio è che la nuova edizione di *Tecniche e modelli di analisi multivariata* possa diventare uno strumento utile per quei ricercatori che intendono applicare consapevolmente l'analisi multivariata nelle proprie ricerche evitando l'infatuazione per la matematica, tipica dei neofiti, che conduce all'inversione del rapporto fra i fini e i mezzi. Sempre più ricercatori cercano di definire la propria identità a partire dai loro presunti metodi scientifici. Tali metodi sono caratterizzati da un consistente uso di modelli e tecniche matematiche – per non parlare della nuova tendenza a lavorare con i cosiddetti *big data* con algoritmi informatici di estrazione ed analisi degli stessi finalizzati alla cosiddetta intelligenza artificiale. Nella maggior parte dei casi questo eccesso di sofisticazione matematica non è altro che una messa in scena che maschera un insieme di lavori scientifici del tutto poveri di significato sostanziale, o che non fanno altro che riprodurre conoscenze già acquisite se non di senso comune. Si spendono troppe energie su dettagli di poco conto (come gli aspetti tecnici e statistico-matematici delle analisi dei dati), mentre si dedica poco spazio – di solito con un lungo elenco di citazioni poste all'inizio dell'articolo – per definire gli aspetti concettuali e teorici del fenomeno oggetto della pubblicazione.

Se i ricercatori delle scienze umane vogliono davvero rendersi utili devono imparare a scegliere in modo critico e consapevole i loro procedimenti di analisi dei dati, facendo tabula rasa delle proprie o altrui certezze, come, ad esempio i criteri standard imposti dalle mode del momento o dalle riviste di fascia A.

Il volume è rivolto anche a quei ricercatori che, rifiutando a priori l'analisi quantitativa dei dati in base a preconcetti e ideologie ormai superate, appena vedono un numero vanno in fibrillazione e immediatamente denunciano l'impostura e il tentativo di offuscare la reale complessità dei fenomeni sociali.

Occorre essere consapevoli che ogni cifra è una costruzione sociale, che deriva dalla definizione concettuale di un fenomeno e da un insieme di regole operative definite in modo trasparente e intersoggettivo che ne consentono la rilevazione empirica: non si tratta della verità oggettiva, ma neanche di una mistificazione di una qualche forma di potere occulto che si vuole imporre sui comuni mortali. A questo proposito citiamo un brano di Weber (1922/1958, p. 254):

“[e viceversa] i dati statistici [...], ovunque concernano il corso oppure le conseguenze di un atteggiamento che racchiude in sé qualcosa di interpretabile in maniera comprensibile, risultano per noi “spiegati” solo se vengono anche realmente interpretati in modo dotato di senso nel caso concreto”.

La necessità di interpretare e comprendere semanticamente i risultati delle analisi dei dati, oltre a quella di spiegarli statisticamente, deve essere considerata una guida costante anche nel corso di tutte le operazioni tecniche di controllo empirico.

A nostro parere, questa conclusione dovrebbe essere condivisa da chi ha veramente esperienza diretta di ricerca empirica: non c'è operazione di ricerca, dalla fase di progettazione di un'indagine a quella dell'analisi e interpretazione dei risultati, che non richieda costanti sforzi di comprensione e interpretazione tesi all'acquisizione di una consapevolezza nelle scelte da compiere, tanto che possiamo definire tutta la metodologia della ricerca nei termini di *arte della scelta* (Maradi 1996).

Il presente volume è complementare ai testi *L'analisi dei dati con Spss. Guida alla programmazione e alla sintassi dei comandi* (Di Franco 2009), e *Dalla matrice dei dati all'analisi trivariata* (Di Franco 2011a) editi nella stessa collana. Diamo per scontato che gli argomenti dei due volumi siano conosciuti dai lettori di questo testo nel quale si illustreranno solo le più usate tecniche e modelli di analisi multivariata dei dati trattandole ad un livello da utente informato. Queste verranno introdotte direttamente nella loro cornice informatica, ossia saranno trattate in relazione alla loro effettiva disponibilità nei *packages* più diffusi presso i ricercatori delle scienze umane, non trascurando però di fare riferimento ai fondamenti statistico-matematici che implicano e cercando di evidenziarne pregi e difetti in relazione a specifici obiettivi cognitivi.

Un consapevole apprendimento delle possibilità di uso di questi strumenti è possibile solo attraverso un esercizio pratico. Per esercitarsi praticamente nell'analisi dei dati, oltre alle necessarie conoscenze di base di metodologia e di statistica, occorre disporre di un *file* di dati, di un *personal computer*, di un adeguato *software* informatico e una certa dose di buon senso che non faccia dimenticare che riuscire a risolvere i problemi di una esercitazione e/o di una simulazione di analisi dei dati è solo una piccola parte del bagaglio di esperienza che occorre per riuscire a percorrere un intero *iter* di ricerca empirica.

Se ripercorriamo a ritroso il consueto ordine secondo il quale si rende conto di un procedimento di ricerca, possiamo affermare che una buona analisi dei dati dipende dall'aver costruito una buona matrice dei dati, che significa aver definito precisamente l'unità di analisi dalla quale si sono individuati i casi su cui si sono rilevati i dati e aver progettato bene le variabili; ciò implica che la rilevazione dei dati sia stata effettuata in maniera pertinente ed efficace rispetto alla formulazioni delle ipotesi e degli obiettivi esplicitati e definiti nel disegno della ricerca rispetto al quale l'intero *iter* di ricerca comincia e, con andamento circolare e iterativo, finisce e riparte.

## 1.2 Cosa leggere per saperne di più

Il testo è stato scritto pensando ad un lettore privo di competenze specialistiche. Naturalmente, per contenere la dimensione del libro, alcuni argomenti sono dati per scontati. Diamo per scontato che il lettore abbia frequentato corsi di metodologia della ricerca e corsi di statistica sufficienti a trasmettergli idee abbastanza precise su cosa siano varianza e covarianza, correlazione e regressione, l'analisi trivariata nella tradizione lazarsfeldiana (spiegazione, interpretazione e specificazione), l'analisi fattoriale classica, i coefficienti di significatività statistica (come, ad esempio, il chi quadrato), e i procedimenti elementari dell'inferenza statistica.

Per colmare eventuali lacune su questi argomenti, si consiglia la lettura di uno o più manuali di metodologia, di analisi dei dati e di un testo sul campionamento e sull'inferenza statistica fra quelli di seguito elencati:

- Blalock, H. M., 1960, *Social Statistics*, New York, McGraw Hill Book; tr. it., *Statistica per la ricerca sociale*, Bologna, Il Mulino, 1969.
- Blalock, H. M., 1961, *Causal Inferences in Nonexperimental Research*, Chapel Hill, The University of North Carolina Press; tr. it., *L'analisi causale in sociologia*, Vicenza, Marsilio, 1967.
- Blalock, H. M., 1970, *An Introduction to Social Research*, New Jersey, Prentice-Hall; tr. it., *Introduzione alla ricerca sociale*, Milano, FrancoAngeli, 1993.
- Bohrnstedt, G. W., Knoke, D., 1994, *Statistics for social data analysis*, Itasca Ill., F. E. Peacock Publishers, tr. it., 1998, *Statistica per le scienze sociali*, Bologna, Il Mulino.
- Bolasco, S., 1999, *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Roma, Carocci.
- Di Franco, G., 1997, *Tecniche e modelli di analisi multivariata dei dati. Introduzione all'applicazione per la ricerca sociale*, Roma, Seam.
- Di Franco, G., 2003, *L'analisi multivariata nelle scienze sociali. Modelli log-lineari e variabili categoriali*, Roma, Carocci.
- Di Franco, G., 2005, *EDS: esplorare, descrivere e sintetizzare i dati. Guida pratica all'analisi dei dati nella ricerca sociale*, Milano, FrancoAngeli.
- Di Franco, G., 2006, *Corrispondenze multiple e altre tecniche multivariate per variabili categoriali*, Milano, FrancoAngeli.
- Di Franco, G., 2009, *L'analisi dei dati con SPSS. Guida alla programmazione e alla sintassi dei comandi*, Milano, FrancoAngeli.
- Di Franco, G., 2010, *Il campionamento nelle scienze umane. Teoria e pratica*, Milano, FrancoAngeli.
- Di Franco, G., 2011a, *Dalla Matrice dei dati all'analisi trivariata. Introduzione all'analisi dei dati*, Milano, FrancoAngeli.
- Di Franco, G., 2011b, *Tecniche e modelli di analisi multivariata*, Milano, FrancoAngeli.
- Di Franco, G., 2016, *I modelli di equazioni strutturali: concetti, strumenti e applicazioni*, Milano, FrancoAngeli.
- Di Franco, G., Marradi, A., 2003, *Analisi fattoriale e analisi in componenti principali*, Roma-Catania, Bonanno.
- Di Franco, G., Marradi, A., 2013, *Factor Analysis and Principal Component Analysis*, Milano, FrancoAngeli.
- Fabbris, L., 1997, *Statistica multivariata. Analisi esplorativa dei dati*, Milano, McGraw-Hill.
- Galtung, J., 1967, *Theory and Method of Social Research*, Londra, Allen & Unwin.
- Goode, W. J., Hatt, P. K., 1952, *Methods in Social Research*, New York, McGraw Hill; tr. it., *Metodologia della ricerca sociale*, Bologna, Il Mulino, 1962.
- Hyman, H. H., 1955, *Survey Design and Analysis*, New York, The Free Press; tr. it., *Disegno della ricerca e analisi sociologica*, Padova, Marsilio, 1967.