

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 Quattrociochi, 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".

Giovanni Di Franco

DALLA MATRICE DEI DATI ALL'ANALISI TRIVARIATA

Introduzione
all'analisi dei dati

La cassetta degli attrezzi
Strumenti per le scienze umane/7

FrancoAngeli

Progetto grafico di copertina di Maria Teresa Pizzetti

Copyright © 2011 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*

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: dalla teoria alla 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 Guida per ricercatori, giornalisti e politici*

Volumi in preparazione:

Alberto Marradi, *Come evitare gli errori tipici in un questionario*

Indice

1. Introduzione	pag.	9
2. Dall'informazione ai dati: progettare e costruire una matrice dei dati	»	19
2.1 La matrice dei dati e l'algebra matriciale	»	19
2.2 Progettare una matrice dei dati	»	24
2.3 Unità di analisi e relative proprietà	»	27
2.4 Costruire una matrice dei dati	»	28
2.5 Altre matrici utili per l'analisi dei dati	»	44
2.6 Cosa leggere per saperne di più	»	47
3. Preparare i dati per l'analisi	»	49
3.1 Etichettare le variabili e le modalità	»	49
3.2 Il trattamento dei dati mancanti	»	51
3.3 Ricodificare i dati di una variabile	»	54
3.4 Costruire indici a partire dalle variabili presenti nella matrice	»	56
3.5 Ponderare i dati	»	58
3.6 Normalizzare e standardizzare i dati	»	61
3.7 La deflazione	»	64
3.8 Cosa leggere per saperne di più	»	72
4. L'analisi monovariata	»	73
4.1 Variabili categoriali semplici	»	77
4.2 Variabili con categorie ordinate	»	83
4.3 Variabili quasi-cardinali e cardinali	»	90
4.4 Cosa leggere per saperne di più	»	98

5. L'analisi bivariata	»	99
5.1 Variabili categoriali semplici	»	104
5.2 Variabili categoriali dicotomiche	»	119
5.3 Variabili con categorie ordinate	»	121
5.4 Variabili quasi-cardinali e cardinali	»	128
5.5 Cosa leggere per saperne di più	»	145
6. L'analisi trivariata	»	147
6.1 Introduzione all'analisi trivariata	»	147
6.2 Variabili categoriali	»	153
6.3 Variabili quasi-cardinali e cardinali	»	159
6.4 Cosa leggere per saperne di più	»	165

1. Introduzione

Paul Felix Lazarsfeld definì le operazioni di analisi dei dati nelle scienze umane con l'espressione 'il linguaggio delle variabili'. Con tale espressione volle indicare che l'interesse del ricercatore è rivolto allo studio delle relazioni fra le variabili ritenute rilevanti in connessione con l'oggetto di studio. Al metodologo austro-americano dobbiamo una prima sistematizzazione delle operazioni finalizzate allo studio delle relazioni fra le variabili, insieme a molti altri contributi ancora oggi usati nelle operazioni di ricerca.

Prima di studiarne le relazioni, occorre chiarire quali caratteristiche presentano le variabili tipiche nelle scienze umane. Nella mia esperienza di docente ho constatato che molti studenti hanno difficoltà a comprendere le differenze fra le variabili usate nelle scienze umane e ciò pregiudica gravemente la comprensione dell'intera analisi dei dati.

Cominciamo dalla definizione di variabile: è uno strumento attraverso il quale rileviamo stati diversi su una proprietà. La variabile è quindi un dispositivo costruito dal ricercatore per rilevare i dati sui casi della ricerca. Per costruire una variabile occorre una procedura che chiamiamo 'definizione operativa'. Essa consiste in un insieme di regole, procedure, decisioni, attraverso le quali a partire da un concetto di proprietà definiamo una variabile. L'applicazione di una definizione operativa si dice 'operativizzazione'. Ad esempio, se la proprietà che vogliamo rilevare è la temperatura di un corpo, la variabile progettata è il termometro che, sfruttando le proprietà del mercurio, è in grado di rilevare il grado di temperatura di un corpo perché produce una dilatazione del mercurio su una scala graduata. Se la proprietà

è il peso, la variabile sarà una bilancia che è in grado di misurare il peso adottando una unità di misura convenzionale (il kg).

Nelle scienze umane di solito si rilevano altre proprietà. Ad esempio, il genere dei soggetti, il livello di istruzione, lo status professionale, gli atteggiamenti, e così via. Anche queste proprietà possono essere definite operativamente ma quasi sempre le variabili costruite presentano caratteristiche piuttosto diverse rispetto a quelle delle discipline fisico-naturali. Quindi per comprendere le differenze fra le variabili occorre risalire alla loro definizione operativa. In generale possiamo considerare sei diverse definizioni operative che producono cinque tipi di variabili:

1. classificazione politomica;
2. classificazione dicotomica;
3. classificazione ordinata;
4. ordinamento;
5. misurazione;
6. conteggio.

Definire operativamente un concetto di proprietà significa fissare un insieme di regole che consentono di rilevare concretamente per ogni caso il suo stato sulla proprietà in considerazione e trasformarlo in un codice posto in una cella della matrice dei dati. Ad esempio, volendo rilevare gli stati sulla proprietà 'confessione religiosa' in un campione di italiani, dobbiamo stabilire:

1. lo strumento attraverso il quale raccogliere le informazioni (ad esempio una domanda diretta ai soggetti contenuta in un questionario, o una rilevazione da un registro amministrativo, etc.);
2. il modo più conveniente di prevedere la variabilità degli stati sulla proprietà; tornando al nostro esempio, per gli scopi della nostra indagine potrebbe essere sufficiente prevedere tre soli stati sulla proprietà confessione religiosa: a) cattolico; b) di confessione non cattolica; c) ateo (in questo caso adottiamo una classificazione tricotomica, cioè composta da tre classi). Ovviamente sono possibili delle classificazioni molto più dettagliate, o sensibili, e si valuterà caso per caso a quale convenga ricorrere, oppure si potrà decidere di individuare due soli classi (credente/non credente) e quindi adottare una classificazione dicotomica;

3. le regole di trasferimento degli stati sulla proprietà in dati sulla matrice. Questa fase viene normalmente etichettata come codifica e consiste nell'assegnare dei codici (di solito numerici) per ogni diverso stato rilevato (vedi capitolo due).

Abbiamo visto un primo esempio di progettazione di una variabile che prevede un'operazione di classificazione semplice. Indipendentemente dal numero di classi individuate, la classificazione semplice distingue una classe da tutte le altre. I criteri di una corretta classificazione sono tre: 1) unico criterio della divisione; 2) mutua esclusività delle classi; 3) esaustività complessiva delle classi.

Quando è possibile individuare un ordine o una gerarchia fra le classi la classificazione si dice ordinata. Anche in questo caso valgono i tre criteri sopra elencati. Ad esempio, volendo rilevare i titoli di studio in un campione possiamo individuare le seguenti classi: a) nessun titolo; b) scuola elementare; c) licenza media; d) diploma; e) laurea triennale; f) laurea magistrale; g) post-laurea. Come si può notare i sette titoli sono in ordine crescente: si comincia da nessun titolo e si arriva al post-laurea.

L'ordinamento (*scaling*) è una procedura che per certi versi è simile alla classificazione ordinata. La differenza è che nella classificazione le classi le individua il ricercatore, di solito prima della rilevazione dei dati; l'ordinamento è invece il risultato di un procedimento che consiste nella somministrazione ai soggetti di scale per la rilevazione di atteggiamenti. Dopo la rilevazione queste sono elaborate per scalare i soggetti e/o gli *items* della scala lungo una o più dimensioni non direttamente rilevabili.

La misurazione è un'operazione che confronta l'ammontare di una proprietà posseduta da un caso con un'unità di misura convenzionale. L'esito di tale operazione è una cifra che esprime l'ammontare della proprietà di un soggetto in multipli o sotto multipli dell'unità di misura. Ad esempio, se vogliamo sapere quanto è lungo un tavolo dobbiamo prendere un metro e confrontare quante volte è contenuto nella lunghezza del tavolo. Ovviamente, per essere più precisi dovremo rilevare anche i decimetri e i centimetri. In ogni caso dovremmo stabilire un livello di approssimazione perché in effetti la misurazione può regredire all'infinito (ad esempio si potrebbero rilevare i millimetri, i micron e così via). In generale le precondizioni per effettuare

una misurazione sono: a) la proprietà deve essere concepita come continua; b) deve esistere un'unità di misura convenzionale a partire dalla quale si costruisce uno strumento di misura; c) deve esserci un rilevatore che confronta lo strumento di misura con la proprietà di un soggetto o di un oggetto. Sottolineiamo la terza caratteristica perché spesso si fa confusione. Se chiediamo ad una persona quanto è alta e la persona risponde un metro e ottanta centimetri, a rigore non abbiamo effettuato una misurazione, ma abbiamo raccolto una informazione che potrebbe essere scorretta, per motivi sia intenzionali sia involontari. Sarebbe diverso se avessimo preso un metro e rilevato l'altezza di quella persona. Altri esempi per chiarire il problema sono: quando chiediamo l'età di una signora e otteniamo risposte che indicano un'età inferiore a quella effettiva; quando chiediamo quante ore al giorno si guarda la televisione: anche in questo caso la tendenza delle persone è quella di sottostimare il tempo effettivo passato davanti allo schermo.

Il conteggio è un'operazione che consiste nella rilevazioni diretta degli stati su una proprietà discreta. Ad esempio, il numero di figli, il numero di matrimoni, il numero di appartamenti di proprietà, il numero dei componenti di una famiglia, il numero di viaggi effettuati in un anno, e così via.

Da queste sei diverse operazioni derivano variabili con caratteristiche diverse: dalla classificazione politomica variabili categoriali semplici; dalla classificazione dicotomica variabili dicotomiche; dalla classificazione ordinata variabili categoriali ordinate; dall'ordinamento variabili categoriali ordinate o variabili quasi-cardinali (vedi oltre) a seconda di quale tecnica di *scaling* si è adottata; dalla misurazione e dal conteggio variabili cardinali. Indipendentemente da quale sia il livello di analisi dei dati, si devono saper scegliere le procedure compatibili con la diversa natura delle variabili.

Nelle ricerche le proprietà che si rilevano si possono classificare in continue o discrete. Una proprietà si definisce discreta se i suoi stati variano per incrementi finiti (ad esempio il numero dei figli di una coppia di coniugi), continua se i suoi stati variano per intervalli infinitesimi (ad esempio le misure fisiche come il tempo, lo spazio, etc.). Evidentemente una proprietà continua per essere misurata deve essere resa discreta individuando una unità di misura (convenzionale)

e costruendo uno strumento di misura che esprima i diversi stati rilevati con un numero finito di suoi multipli e sottomultipli. A loro volta le variabili che ne derivano si possono distinguere in cardinali e categoriali. Una variabile si definisce cardinale, in senso stretto, quando è esprimibile con un'unità di misura convenzionale, che può essere, specie nel caso di proprietà di tipo psico-attitudinali, anche sottintesa (come avviene nella costruzione delle scale Thurstone con la tecnica degli intervalli che appaiono uguali. Questa è una delle procedure di *ordinamento* o *scaling*); in ogni caso il *continuum* che rappresenta il campo di variazione della proprietà viene diviso in intervalli uguali (o supposti tali) a partire da un punto arbitrario (detto anche zero convenzionale). Quando una proprietà è concepibile come continua ma non esiste un'unità di misura convenzionale si può applicare una definizione operativa che prevede l'adozione di tecniche particolari dette 'scale autoancoranti'. Queste scale hanno la caratteristica di avere etichette solo ai due estremi: l'inizio e la fine della scala. Ad esempio, le scale di autocollocazione sull'asse destra-sinistra, i termometri dei sentimenti, il differenziale semantico, etc. Tutti gli altri punti compresi fra i due estremi sono privi di etichetta e sono rappresentati o da simboli grafici o da numeri in ordine crescente. I soggetti devono autovalutare l'ammontare della proprietà da loro posseduta scegliendo di collocarsi in un punto entro la scala. Le tecniche autoancoranti producono un tipo di variabili che si definiscono quasi-cardinali; nell'analisi dei dati tali variabili sono trattate come cardinali (dopo che in sede di pre-analisi sono state sottoposte ad una procedura di deflazione; vedi capitolo tre, paragrafo 3.7).

Un caso diverso è rappresentato da quelle proprietà i cui stati sono rilevabili non con una unità di misura o attraverso l'ordinamento, ma con una unità di conto (ad esempio il numero dei figli, l'ammontare dei depositi bancari posseduti da una famiglia non si rilevano misurando, ma attraverso il conteggio); in questo caso il valore zero rappresenta l'assenza della proprietà. Dal punto di vista dell'analisi dei dati questa differenza viene superata dalla standardizzazione (vedi capitolo tre, paragrafo 3.6), operazione necessaria per rendere comparabili variabili rilevate con diverse unità di misura o di conto, che permette di esprimere gli stati sulle proprietà in unità di scarto-tipo (scarto quadratico medio) dalla media che viene posta a zero. Quan-

do, invece, una proprietà non è rilevabile né con una unità di misura, né con l'ordinamento, né con una unità di conto, allora si dice categoriale. A partire dalla celebre classificazione di Stevens (1946; tr. it., 1991), il termine "misurazione" è stato eccessivamente esteso, riferendolo a qualsiasi procedura che permetta di attribuire dei numeri a degli oggetti o eventi seguendo determinate regole. La tabella 1.1 riporta il quadro sinottico delle definizioni operative possibili in relazione alla natura della proprietà.

Tabella 1.1 – Tipi di proprietà e operazioni di rilevazione connesse

tipo proprietà	variabile	cardinale	categoriale
continua		<i>misurazione o ordinamento</i>	<i>classificazione ordinata</i>
discreta		<i>conteggio</i>	<i>classificazione ordinata o non ordinata</i>

Per l'analisi dei dati interessa sottolineare che le operazioni eseguibili sulle variabili dipendono dalla loro natura: se le variabili sono cardinali o quasi-cardinali (frutto di misurazione, ordinamento o conteggio) possiamo applicare tutte le procedure statistiche compatibili con questo livello; se le variabili sono categoriali si devono applicare altri strumenti statistici.

Tabella 1.2 – Tipi di proprietà e tipi di variabili nell'analisi dei dati

tipi di proprietà	corrispondenti tipi di variabili rilevanti nell'analisi dei dati	
	<i>monovariata e bivariata</i>	<i>multivariata</i>
	considerate discrete:	
- categoriali non ordinate	categoriali	categoriali
- categoriali ordinate	ordinali	(trattate come categoriali)
- con stati enumerabili	cardinali	cardinali
	considerate continue:	
- non misurabili	ordinali	(trattate come categoriali)
	quasi-cardinali	(trattate come cardinali)
- misurabili	cardinali	cardinali

Fonte: Marradi 1993, 15.

Considerando la natura delle variabili in relazione alle operazioni di analisi dei dati, Marradi (1993) sottolinea come in sede di analisi trivariata e multivariata la distinzione tra le variabili tende a diventare dicotomica, mentre per i livelli di analisi monovariata e bivariata

essa è più articolata, come mostrato nel quadro sinottico riportato in tabella 1.2.

L'analisi dei dati è costituita da un variegato insieme di operazioni statistiche attraverso le quali si producono asserti e spiegazioni che riguardano i dati stessi. Il modo tipico della ricerca 'standard', ossia la ricerca che costruisce e analizza una matrice dei dati con opportune tecniche statistiche, nella produzione di asserti e di nessi tra asserti consiste nell'adozione di procedure di calcolo 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.

Possiamo definire i principali obiettivi dell'analisi dei dati nel modo seguente:

1. l'esplorazione, la descrizione, la sintesi e la classificazione dei dati relativi all'oggetto di studio;
2. l'analisi delle relazioni fra le variabili che in ipotesi sono ritenute rilevanti in connessione con l'oggetto di studio;
3. la previsione di stati futuri che riguardano l'oggetto di studio.

Fondamentalmente si individuano quattro livelli di analisi dei dati:

- 1) la pre-analisi dei dati;
- 2) il livello monovariato;
- 3) il livello bivariato;
- 4) il livello multivariato.

Prima dell'analisi dei dati è necessario preparare i dati, e questo di solito comporta diverse operazioni. In realtà, alcune di queste sono fra la pre-analisi e l'analisi monovariata. Per comodità espositiva, nel capitolo tre presenteremo solo le operazioni effettivamente preliminari all'analisi dei dati, rinviando ai capitoli successivi la presentazione delle operazioni di analisi dei dati vere e proprie.

A livello monovariato si studia l'andamento delle variabili rilevate una per una; si esamina la plausibilità dei valori (per individuare eventuali errori presenti nella matrice dei dati); si descrivono le loro distribuzioni calcolando alcuni valori caratteristici, quali quelli di concentrazione, di dispersione, di posizione e della forma della distribuzione (vedi capitolo quattro).

I procedimenti di analisi monovariata sono spesso sottovalutati. Bisogna però sottolineare che, se ben condotta, l'analisi monovariata consente di svolgere in maniera più efficace le successive elaborazioni bivariate e multivariate. Una buona analisi statistica dei dati comincia proprio da una buona analisi monovariata; solo dopo aver dato un'efficace descrizione delle distribuzioni delle singole variabili si può passare all'analisi delle relazioni fra di loro.

A livello bivariato (vedi capitolo cinque) si analizza la distribuzione congiunta di due variabili, in modo da evidenziare se esista o no, che forma e che forza abbia, una dipendenza funzionale o relazione statistica fra loro. In realtà, prima di poter controllare un'ipotesi di relazione causale tra due variabili, si devono accertare almeno quattro aspetti:

1. l'esistenza (e generalizzabilità) della relazione: la relazione analizzata differisce in modo significativo da quella che si sarebbe potuta ottenere per puro effetto del caso?
2. la forza della relazione: questa può essere quantificata entro una gamma definita di valori?;
3. la forma della relazione: la relazione è di tipo lineare o di quale altra forma? È diretta o inversa?
4. la genuinità della relazione bivariata: anche se la relazione analizzata non è imputabile agli effetti del caso, rimane la possibilità che essa sia dovuta all'effetto di una terza variabile che non stiamo prendendo in considerazione, ossia che si tratti di una relazione definita 'spuria'.

A queste domande rispondono degli strumenti logico-statistici diversi:

1. i coefficienti di significatività statistica;
2. i coefficienti di associazione statistica (anche detti di connessione);
3. l'analisi di opportune rappresentazioni grafiche della relazione bivariata;
4. l'analisi trivariata (multivariata).

A livello trivariato (multivariato; vedi capitolo sei) si prendono in considerazione simultaneamente tre o più variabili. In questo caso gli obiettivi principali sono due:

- l'esplorazione, descrizione, sintesi e classificazione dei dati;

- l'individuazione di relazioni tra un insieme di variabili secondo determinati modelli di ipotesi teoriche.

In questo volume prenderemo in considerazione le procedure di analisi trivariata finalizzate al controllo di una relazione bivariata. Rinviamo i lettori al nostro volume *Tecniche e modelli di analisi multivariata* per avere una rassegna dei principali strumenti di analisi multivariata.

Dopo aver discusso gli aspetti di cui si deve tenere conto prima di procedere nell'analisi dei dati che influenzano la scelta delle strategie più idonee e coerenti con gli obiettivi del ricercatore, possiamo descrivere gli effettivi procedimenti di tale analisi.

Dopo la fase della raccolta/costruzione delle informazioni, il primo passo consiste nella costruzione di una matrice dei dati (matrice casi per variabili) dove questi vengono organizzati disponendo per riga i casi (che possono essere, ad esempio, individui cui è stato somministrato un questionario con risposte standardizzate o enti territoriali – comuni, province, regioni, etc. – su cui si sono rilevati dei dati) e per colonna i loro stati sulle proprietà rilevate (variabili). Alla costruzione della matrice dei dati è dedicato il capitolo due.

I dati vengono di solito distinti in due tipi:

a) dati primari: sono costruiti direttamente nell'ambito della ricerca dopo aver predisposto opportuni strumenti di rilevazione. Nella ricerca sociale il più diffuso strumento di rilevazione diretta dei dati è il questionario, che può avere un grado di strutturazione (ovvero un livello di chiusura delle risposte alle domande presenti nel questionario) più o meno elevato. Nella pratica della ricerca sociale si somministrano spesso questionari semi-strutturati nei quali le domande hanno un certo numero di risposte precodificate e una modalità aperta (di solito denominata 'altro') per registrare le risposte diverse da quelle previste;

b) dati secondari: sono costruiti da altra fonte (come ad esempio i censimenti dell'Istat) e vengono acquisiti nell'ambito di una ricerca per fini spesso diversi da quelli per cui sono stati costruiti.

Comunque siano stati costruiti i dati, per essere elaborati si devono organizzare in una matrice.

Dopo aver costruito la matrice, si svolgono delle operazioni che possiamo etichettare come pre-analisi dei dati. In questa fase si predispongono le variabili per le successive analisi (vedi capitolo tre).

L'ultima fase di un procedimento tipico di analisi dei dati consiste nella predisposizione di un rapporto di ricerca. In esso si dovrebbe rendicontare in modo esauriente tutte le operazioni svolte nella ricerca (la definizione dell'unità di analisi e dell'ambito spazio temporale della ricerca; il piano di campionamento; lo strumento di rilevazione; il piano di codifica e di ricodifica delle variabili; i procedimenti di analisi dei dati impiegati). Si dovrebbe cioè avere la massima cura nel fornire ai lettori tutti gli elementi necessari per metterli nelle condizioni di valutare con piena consapevolezza i risultati della ricerca.

Naturalmente, la parte più consistente del rapporto di ricerca sarà dedicata alla presentazione dei principali risultati avvalendosi di tabelle e di grafici appositamente preparati. Nei capitoli successivi, nelle sezioni dedicate alla presentazione di esempi di ricerche, forniremo numerosi esempi di come si presentano i dati. L'attenta lettura di rapporti di ricerca, quando essi siano stati curati con rispetto verso i lettori, costituisce un prezioso strumento di formazione per i giovani e meno giovani ricercatori.

2. Dall'informazione ai dati: progettare e costruire una matrice dei dati

2.1 La matrice dei dati e l'algebra matriciale

Le ricerche della famiglia standard si caratterizzano idealmente per almeno tre aspetti:

- a) il primo consiste nel fatto che le proprietà interessanti per la ricerca sono definite operativamente e che l'unità di analisi è definita in maniera univoca, in modo da poter stilare un elenco esaustivo dei casi che determinano il risultato;
- b) il secondo consiste nella costruzione di una matrice dei dati in cui i casi sono disposti sulle righe e le variabili sono disposte sulle colonne;
- c) il terzo riguarda l'uso di tecniche statistiche nell'analisi dei dati raccolti.

Muoversi all'interno di questa famiglia di ricerche significa accettare gli assunti.

Il complesso di assunti fondamentali di questa famiglia è talmente interconnesso che può essere designato con un'etichetta globale (assunto atomista). Ma analizziamo separatamente ciascun componente il complesso:

- A1) ogni caso (sia esso un individuo, uno Stato, o un esemplare di qualsiasi altro genere di unità d'analisi) è perfettamente scindibile nei suoi stati sulle proprietà rilevate;
- A2) ogni stato su una proprietà, una volta trasformato in dato nella matrice, è perfettamente separabile dal suo detentore;
- A3) ogni stato su una proprietà, una volta trasformato in dato nella matrice, è perfettamente indipendente dagli altri stati del suo detentore su tutte le altre proprietà (rilevate e non);