

# Introduzione

Negli ultimi anni ho avuto la possibilità di tenere vari seminari di introduzione all'analisi dei dati, rivolti prevalentemente a laureandi e giovani ricercatori in sociologia, scienza della politica, economia, psicologia. In questi seminari mi sono reso conto che un pubblico di questo genere è molto più attratto da una didattica orientata ai problemi della ricerca piuttosto che dall'illustrazione delle tecniche di analisi. In questo testo cerco di mettere a frutto queste esperienze; presento infatti percorsi di analisi simili a quelli che si seguono in una ricerca sociale. In particolare, dopo aver illustrato le procedure di costruzione della matrice dei dati (CAP. 1), esamino la relazione tra genere e lettura (CAP. 2), quella tra ambiente di residenza e identificazione territoriale (CAP. 3), quella tra differenze territoriali e comportamento di voto (CAP. 4). Nel presentare tali percorsi seguo un ordine che rispetta la natura delle variabili: prima si analizzano le variabili dicotomiche (CAP. 2), quindi quelle categoriali (CAP. 3) e infine quelle cardinali (CAP. 4).

Piuttosto che presentare una rassegna di programmi informatici prendo in esame SPSS (*Statistical Package for the Social Sciences*), il più diffuso programma per l'elaborazione dei dati nelle scienze sociali – in particolare, farò ricorso alla versione 10 in italiano (che propone peraltro alcuni *outputs* in inglese) – e fornisco una serie di riferimenti ai lettori interessati ad approfondire.

SPSS venne ideato negli anni sessanta dal politologo Norman Nie e dai suoi collaboratori (cfr. Nien *et al.* 1970, p. XXII) adottando l'ottica di politologi e sociologi piuttosto che quella di un informatico. Questo, insieme alla semplicità del programma, fu alla base del suo successo anche nella ricerca di *marketing* e nelle scienze naturali (in particolare tra gli epidemiologi). In ambiente Windows il programma è divenuto molto più *user-friendly* rispetto alla precedente versione per DOS, anche grazie all'ausilio della guida in linea; restano peraltro alcune difficoltà per l'utente, specie se non ha molte conoscenze metodologiche e statistiche. Il testo intende fornire un supporto agli utenti della versione-base di SPSS:

- nella scelta della tecnica più appropriata per affrontare un certo problema cognitivo analizzando dati organizzati in matrice;
- nella lettura degli *outputs* (tabelle e grafici).

Esso si rivolge quindi a studenti e ricercatori che intendono approfondire le loro conoscenze metodologiche e informatiche (a tal fine è stato inserito un piccolo *Glossario*); può risultare utile anche a studiosi già competenti in metodologia e/o statistica che però hanno scarsa familiarità con i programmi di elaborazione dei dati.

Per permettere al lettore di acquisire familiarità con il programma, nella versione *on line*, consultabile sul sito Internet [www.carocci.it](http://www.carocci.it), sono inseriti file di esercitazione.

La versione *on line* è articolata in quattro parti (descritte nel file allegati.doc):

- matrici dei dati usate nel testo;
- esercitazioni;
- approfondimenti e integrazioni ai temi affrontati nel testo (in particolare, si illustrano le principali tecniche per costruire indici; si presenta un'applicazione della correlazione parziale e della regressione multipla. Nel testo sono indicati come allegati);
- collegamenti a siti da cui è possibile accedere a versioni *freeware* e *shareware* di alcuni programmi di elaborazione dei dati.

Il testo non dovrebbe essere considerato dal lettore un'alternativa ai manuali di statistica descrittiva e/o di analisi dei dati ai quali si rinvierà tutte le volte che è opportuno. Inoltre, trattandosi di un testo elementare, l'attenzione sarà rivolta all'analisi monovariata, bivariata e trivariata (quest'ultima limitatamente ai controlli per stratificazione). È anche opportuno sottolineare che qui si illustra solo una forma di ricerca sociale, quella basata sulla matrice dei dati. I programmi di elaborazione dei dati possono giocare un ruolo di un certo rilievo anche nella cosiddetta ricerca qualitativa; ma questo sarebbe il tema di un altro volume.

Per concludere, fornisco alcune informazioni sulle matrici impiegate nel testo, riportate anche nella versione *on line*. I dati utilizzati nei capitoli 2 e 3 derivano da una ricerca sulla cultura politica toscana diretta da Mario Caciagli e Alberto Marradi. La fase sul campo è stata condotta tra il gennaio e il marzo 1992. Sono stati intervistati 518 maggiorenni sull'intero territorio regionale, cercando di garantire la rappresentatività del campione sulle seguenti proprietà: pro-

vincia e fascia dimensionale del comune di residenza, sesso, fascia di età (alcuni risultati sono stati presentati in Barlucchi, Fideli, Rodolfi, 1993).

I dati sulle elezioni del 1994 e sui referenda del 1995 utilizzati nel CAP. 4 sono stati organizzati da ricercatori dell'Istituto Carlo Cattaneo di Bologna. Alcuni risultati della ricerca sono stati pubblicati in un volume della serie "Politica in Italia" (Uleri, Fideli, 1996). Le variabili censuali prese in esame negli allegati *on line* derivano invece da un'aggregazione di dati a livello di collegio, gentilmente messi a disposizione da Roberto D'Alimonte. L'aggregazione è stata effettuata da Nicola Malloggi (per un'elaborazione dei dati a livello di collegio cfr. Caramani, 1996).

Mi piace considerare questo piccolo volume un'opera collettiva, anche se le responsabilità per eventuali errori sono interamente di chi scrive. Desidero anzitutto ringraziare Alberto Marradi, che ha insistito molto affinché adottassi l'ottica descritta all'inizio piuttosto che quella consueta nei manuali informatici, e che mi ha ricondotto sulla retta via ogni volta che scivolavo nell'approccio tradizionale. Nicola Malloggi e Andrea Pirni hanno generosamente collaborato alla costruzione delle appendici e degli allegati *on line*. Ringrazio per gli utili suggerimenti Ciro D'Amore, Giovanni Di Franco, tutti gli studenti dei corsi di SPSS (elementari e avanzati) che ho tenuto a partire dal 1999 presso il Dipartimento di Scienza politica e Sociologia dell'Università di Firenze, i ricercatori degli istituti IRIS di Prato e IRRES di Perugia, presso cui ho avuto occasione di tenere corsi analoghi (rispettivamente, nel 1996 e nel 2000).