

PREMESSA



I temi trattati in questo volume si collocano in quella classe di procedure, definite metodi e modelli per l'analisi di variabili categoriche, i cui dati discreti si distribuiscono in tavole di contingenza bidimensionali e multidimensionali. Come sottolineano Y.H.M. Bishop, S.E. Fienberg e P.W. Holland nella prefazione al volume del 1975, benché i metodi per l'analisi di tavole di contingenza siano stati oggetto di studio nella letteratura statistica fin dai tempi di K. Pearson e di R.A. Fischer, e benché il pionieristico articolo di M. Bartlett sull'assenza di interazione di secondo ordine in tavole tridimensionali sia datato 1935, lo sviluppo e la diffusione di tali metodi hanno dovuto attendere che gli strumenti di calcolo automatico fossero nella piena disponibilità dei ricercatori. Dalla metà degli anni Sessanta, le riviste scientifiche, nelle aree della statistica, delle scienze biologiche, sociali e della medicina, hanno dedicato sempre maggiore spazio al lavoro di studiosi che si sono occupati di variabili categoriche. I nomi più autorevoli che hanno contribuito alla messa a punto e alla diffusione di metodi e modelli per l'analisi di tali variabili sono, tra gli altri, quelli di M.W. Birch (1963; 1964a; 1964b; 1965), di L.A. Goodman (1964; 1968; 1971), di R.L. Plackett (1962; 1964; 1965; 1971). L'interesse sia dal punto di vista teorico sia applicativo nei confronti dei modelli log-lineari è particolarmente dovuto al fatto che essi si adattano all'analisi di tavole di contingenza multidimensionali, in quanto consentono di definire la struttura delle relazioni tra più variabili, tenendo conto della complessità che caratterizza le loro interazioni. I contributi più autorevoli e più recenti per lo sviluppo dei modelli log-lineari si trovano nei volumi di Agresti (1990), Christensen (1990; 1997), Jobson (1991; 1992), Fahrmeir e Tutz (1994).

Lo scopo di questo volume consiste nell'introdurre gli aspetti formali

dei modelli log-lineari fondamentali e alcune loro applicazioni. La trattazione si limita infatti ai modelli per l'analisi bivariata e trivariata. Le forme specifiche che tali modelli assumono, nel rappresentare le strutture che caratterizzano le relazioni tra i dati empirici, sono presentate per mezzo di un linguaggio formale semplice e con dimostrazioni essenziali e sintetiche. Analogamente, gli aspetti statistici, necessari in particolare nel processo di validazione dei modelli, è reso facile sulla base di esempi illustrativi. Tali esempi, caratteristici della ricerca psicosociale, descrivono tutte le fasi di modellazione dei dati e di validazione, tipiche dell'analisi log-lineare.

Gli argomenti qui trattati richiedono per la loro comprensione conoscenze elementari di matematica, di statistica descrittiva e di statistica inferenziale. Il lavoro di Robusto e Cristante (2001) rappresenta inoltre la premessa naturale ed essenziale per affrontare lo studio dei temi qui trattati. La parziale sovrapposizione di alcune tematiche fra il presente contributo e quello citato consente al lettore più esperto di affrontare senza ulteriori rimandi la lettura di questo volume.

Gli argomenti presentati si articolano nel modo seguente.

Il *Capitolo 1* ha carattere introduttivo. In esso sono definiti alcuni concetti di base che consentono di affrontare lo studio dei modelli log-lineari per l'analisi di tavole bidimensionali e tridimensionali. In particolare sono introdotti argomenti quali le distribuzioni di frequenze osservate, di probabilità, di frequenze attese e di frequenze attese stimate in tavole di contingenza. Viene data la definizione di indipendenza e introdotta la statistica «rapporto di verosimiglianza», entro la cui distribuzione di probabilità viene effettuato il controllo dell'indipendenza tra variabili categoriche. Il caso particolare della tavola di contingenza 2×2 e il «rapporto di associazione» vengono trattati allo scopo di definire i concetti di associazione parziale e di associazione marginale nell'analisi di tavole di contingenza tridimensionali.

Il *Capitolo 2* riguarda i modelli log-lineari per tavole bidimensionali. Vengono presi in considerazione il modello di indipendenza e il modello di dipendenza, la procedura di validazione dei modelli, la stima dei parametri e la verifica della loro significatività.

Il *Capitolo 3* è dedicato all'analisi degli effetti in tavole bidimensionali e al controllo della loro significatività.

Nel *Capitolo 4* viene definito il concetto di struttura in tavole tridimensionali e vengono presentate le ipotesi relative alle strutture caratteristiche delle relazioni tra le variabili. Sono inoltre introdotti i modelli log-lineari gerarchici in tavole tridimensionali, la procedura di validazione dei model-

li, la stima dei parametri che rappresentano le relazioni tra le variabili e la verifica della significatività delle stime.

Il *Capitolo 5* riguarda l'analisi degli effetti e la relazione tra effetti e modelli log-lineari gerarchici in tavole tridimensionali. In tale contesto si inserisce il tema relativo alla procedura per la scelta del modello che meglio rappresenta la struttura delle relazioni tra le variabili.

Il *Capitolo 6*, a completamento degli argomenti trattati nei capitoli 4 e 5, è dedicato ad un'applicazione dell'analisi log-lineare trivariata nel contesto di una ricerca con variabili di rilevanza psicosociale.

Questo volume è rivolto in particolare a studenti e ricercatori di psicologia, e più in generale delle scienze sociali, interessati allo studio delle variabili categoriche. Gli argomenti trattati sono un primo passo verso l'approfondimento dello studio dei modelli log-lineari per l'analisi multivariata di tavole di contingenza e verso lo studio di procedure che consentono di rilevare le strutture caratterizzanti le relazioni tra variabili categoriche.

Per quanto concerne i programmi di calcolo automatico per l'elaborazione dei modelli trattati, riteniamo utile dare un'indicazione riguardo al programma BMDP Dynamic v. 7.0. Si tratta di un programma statistico che più di altri si trova in linea con l'approccio teorico e applicativo qui adottato. I capitoli del manuale intitolati «4F – Part 2 – Analysis of two-way tables» e «4F – Part 3 – Analysis of multiway tables: log-linear models» (BMDP, 1992) contengono definizioni, spiegazioni ed esempi per l'elaborazione dei modelli log-lineari per tavole di contingenza bidimensionali e multidimensionali che, con modalità semplice e rigorosa, conducono il ricercatore all'individuazione delle strutture gerarchiche e non gerarchiche dei dati e alla scelta dei modelli più adeguati.

Padova e Roma, ottobre 2001

F.C. E.R. S.M.