
Prefazione

Viviamo nella società dell'informazione. Non passa giorno senza che questa frase ci sia ricordata da giornali, televisioni, Internet. Non passa giorno senza che veniamo bombardati di cifre riguardanti gli elementi più disparati del nostro sistema sociale: andamento dei prezzi, livello di disoccupazione, gradimento nei confronti di questo o quel partito politico. Non passa giorno senza che ci siano forniti i risultati di “sondaggi” riguardanti i più vari aspetti del nostro vivere sociale.

Questa formidabile mole di numeri che ci vengono forniti in dosi sempre più massicce, però, genera spesso scetticismo. Come stabilire quando l'informazione fornita può definirsi “corretta”? Come evitare di fornire numeri molto lontani dalla realtà che si pretende di rappresentare, e su cui si pretende di informare i cittadini?

Questo libro si occupa della corretta acquisizione e dell'uso efficiente di un tipo molto importante di informazione: l'*informazione statistica*. Con una certa semplificazione, e con qualche imprecisione, l'acquisizione dell'informazione statistica ha luogo tramite l'*osservazione* di una o più caratteristiche di interesse (*status* occupazionale, partito politico per cui si intende votare, etc.) sulle unità di una *popolazione* di riferimento.

Un elemento critico del processo di acquisizione dei dati è costituito dal fatto che molto spesso le popolazioni di riferimento sono composte da un numero molto elevato di unità, la cui osservazione completa richiederebbe costi e tempi proibitivi. Per tale ragione si ricorre a processi di acquisizione dei dati basati sull'osservazione di una parte delle unità della popolazione, la quale costituisce un *campione*.

L'idea-guida del presente libro è sostanzialmente una: sia il processo di selezione del campione che l'uso dei dati corrispondenti devono essere volti ad ottenere la massima *efficienza*. È questo obiettivo a spingere la trattazione verso regole di selezione del campione di tipo “scientifico”, fondate sul calcolo delle probabilità. Solo in questo caso, infatti, è possibile studiare in modo corretto cosa si intenda per “uso efficiente dei dati statistici”.

Questo volume è dedicato in larga misura agli aspetti di base del campionamento da popolazioni finite. Vengono stabilite le basi logiche del campionamento e ne vengono studiati i principali sviluppi elementari. L'unica fonte di errore statistico è dovuta al fatto che non si osservano tutte le unità della popolazione, ma solo una parte. Ad esso seguirà un successivo volume dedicato a tematiche più avanzate, generate (prevalentemente ma non esclusivamente) sia dal fatto che a volte può essere impossibile osservare alcune delle unità del campione e/o è possibile effettuare solo osservazioni affette da errori, sia dal fatto che spesso l'interesse verte sulla costruzione di *modelli statistici* per i dati concretamente osservati.

L'approccio espositivo seguito consiste nel procedere dal particolare al generale, partendo da aspetti il più possibile elementari, che vengono poi "complicati" per renderli via via più aderenti alle concrete situazioni applicative. Molti ragionamenti vengono prima basati su esempi concreti e poi su aspetti teorici che, se forniti in prima battuta, potrebbero rendere un po' ostica la comprensione di aspetti chiave. A questo proposito si sottolinea l'importanza degli esempi basati su dati (tutti disponibili sulla pagina web <http://extras.springer.com>), i quali non sono solo un utile complemento alla teoria, ma una vera e propria chiave di accesso al "ragionamento statistico". Lo stesso ruolo è anche svolto da molti degli esercizi proposti.

Il libro si rivolge sia a studenti universitari di corsi di laurea con una robusta componente quantitativa (Scienze Statistiche, Economia, etc.), sia a ricercatori in campo economico o sociale che utilizzano il metodo statistico. Il livello complessivo della trattazione è, per quanto possibile, elementare. L'unico pre-requisito richiesto per accedere a gran parte degli argomenti trattati è un corso elementare di statistica con elementi introduttivi di calcolo delle probabilità e inferenza statistica, come quelli impartiti nelle lauree triennali di Economia e Scienze Politiche. Non strettamente necessario, anche se utile, è un corso elementare di matematica, modellato sui corsi di matematica generale dei corsi di laurea in Economia.

Questo volume è prevalentemente destinato alle lauree triennali. Nella nostra esperienza, i Capitoli 1–7, la Sezione 8.6, i Capitoli 9–11 costituiscono il materiale per un corso di 32–36 ore in una laurea triennale di taglio statistico. L'aggiunta del Capitolo 12, di ampie parti del Capitolo 14 (lo stimatore di Horvitz-Thompson, in sostanza) e di alcune sezioni del Capitolo 15 (disegni di Poisson, di Poisson condizionato e di Pareto) copre un corso *standard* di campionamento di 48 ore, sempre per una laurea triennale di taglio statistico. Lo stesso materiale è stato anche usato, a vari livelli, in corsi di laurea e di Master in Facoltà di Economia. Alcune parti dei Capitoli 14 (stimatori di tipo calibrazione), 15 (disegno bilanciato) e gran parte del materiale sugli errori non campionari e sui modelli statistici per dati campionari (oggetto di trattazione in un successivo volume) sono invece destinati alle lauree specialistiche.

Le parti (sezioni o interi capitoli) con asterisco sono più avanzate, e non vengono utilizzate, in generale, nell'ambito delle lauree triennali. Alcune di

esse sono usate in lauree specialistiche, mentre altre sono state incluse in quanto di diretto interesse applicativo.

Molti degli esempi proposti nel volume si basano su *dataset* disponibili presso la pagina web <http://extras.springer.com>.

Nello scrivere il presente libro abbiamo contratto parecchi debiti di gratitudine. Il Dott. Mauro Scanu ha letto l'intero volume, indicando parecchi errori e fornendo utili suggerimenti. I Proff. Ludovico Piccinato, Marco Riani, Paola Vicard hanno letto ampie parti del volume, fornendo spunti e considerazioni di grande interesse. Utili suggerimenti sono stati anche forniti dal Prof. Francesco Battaglia. Naturalmente, di errori e omissioni residui sono responsabili i soli autori.

Roma, febbraio 2012

Pier Luigi Conti
Daniela Marella