

# Prefazione

## all'edizione italiana

Una delle caratteristiche più rilevanti del volume che state sfogliando è la minore enfasi su formule e formalismi, data dal tradizionale approccio all'insegnamento della Statistica, in favore di un approccio guidato dai concetti. Il presente manuale prende atto di questa evoluzione che si è registrata di recente nell'insegnamento di base e rende l'approccio concettuale più interessante e più accessibile per gli studenti impegnati nei primi corsi universitari di Statistica. A loro il manuale di Alan Agresti, anche nella versione da noi curata, offre la possibilità di apprendere concetti utili per tutto il resto della vita, non solo professionale. Al termine del percorso di studio e apprendimento, lo studente formato su questo volume guarderà ai problemi che deve risolvere dalla prospettiva di chi analizza i dati, con la consapevolezza che la Statistica, oltre a essere una Scienza, è anche un'Arte. E non gli sarà difficile accettare che, nella pratica operativa, le assunzioni non sono mai puntualmente soddisfatte, i modelli mai esattamente corretti, le distribuzioni mai perfettamente normali e che, quando si conduce un'analisi statistica, può essere necessario prendere in considerazione ogni sorta di fattore.

L'approccio seguito dal volume che state sfogliando segue le linee guida e le raccomandazioni approvate nel 2005 dall'American Statistical Association (ASA) per l'insegnamento nei corsi di base di Statistica<sup>1</sup>. Nel suo report l'ASA fornisce fondamentalmente sei suggerimenti per tali attività didattiche:

- diffondere la cultura statistica e sviluppare il ragionamento statistico;
- utilizzare dati reali;
- privilegiare la comprensione dei concetti piuttosto che la sola conoscenza delle procedure;
- favorire l'apprendimento attivo in aula;
- ricorrere alla tecnologia per sviluppare concetti e analizzare dati;
- valutare l'apprendimento degli studenti anche al fine di capirlo e migliorarlo.

---

<sup>1</sup> Tali linee guida sono contenute nel documento *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) scaricabile, nella sua versione rivista nel 2012, dal sito [www.amstat.org/education/gaise](http://www.amstat.org/education/gaise).

Pur senza conoscere nei dettagli il report dell'ASA, anche noi da anni siamo impegnati nella diffusione della cosiddetta *statistical literacy* onorando, di fatto, molte delle raccomandazioni prima elencate. La versione in lingua italiana del volume di Agresti e Franklin nasce proprio dalla convergenza di queste due azioni indipendenti, entrambe volte ad accrescere la cultura statistica: Alan (assieme a Christine) da una parte, con il loro meticoloso rispetto dei suggerimenti dell'ASA; noi, dall'altra, liberi divulgatori animati, nelle nostre esperienze di insegnamento e di redazione di manuali, dallo spirito di rendere familiare la Statistica agli studenti per fornire loro un bagaglio culturale irrinunciabile e uno strumento fondamentale per vivere in modo attivo nella società della conoscenza.

L'incontro fra uno dei più grandi divulgatori della Statistica al mondo, Alan, e noi ha prodotto questo manuale che, anche nella sua versione italiana, è molto diverso da quelli tradizionali. Cercheremo ora di illustrare perché.

Il libro risulta innanzitutto molto leggibile grazie all'ampia varietà di esempi, basati su dati reali, che propone. La chiave della loro lettura è comune: nel trovare la risposta alle domande formulate, si approciano i dati e si individuano le analisi più appropriate. In questo processo di apprendimento, si dà maggiore enfasi alle situazioni più comuni nelle applicazioni pratiche: grandi campioni, varianza incognita, test bidirezionali e campioni indipendenti. Gli altri casi non vengono certo trascurati, ma sono trattati in modo meno capillare e solo all'interno di alcuni esempi circoscritti. Tutti gli esempi presenti nel manuale hanno la stessa struttura e articolazione. Fa eccezione l'esempio di apertura di ciascun capitolo (solo il Capitolo 2 ne è privo), che ha lo scopo di catturare l'attenzione dello studente sui temi trattati nel capitolo stesso e di guidarlo nella lettura. La struttura comune degli altri esempi è composta da quattro elementi preceduti da un'introduzione in cui si inquadra la situazione oggetto di indagine e si presentano, spesso in collegamento con pagine precedenti, i dati da analizzare. I quattro elementi sono costituiti dalle seguenti voci:

- **Domande per esplorare:** si riferiscono alle informazioni fornite nell'introduzione all'esempio e pongono le questioni incentrate sulle metodologie che lo studente apprenderà dalla lettura dell'esempio stesso. Contestualmente insegnano al lettore che tipo di domande è appropriato formulare in base ai dati da analizzare.
- **Risposte e approfondimenti:** vi si trovano le risposte, fornite usando metodi statistici appropriati, alle domande di cui sopra.
- **Spunti di riflessione:** vi si chiariscono le principali idee indagate nell'esempio con un linguaggio non tecnico e con uno sguardo a contesti più ampi di quelli trattati.
- **Risolvi l'esercizio/gli esercizi...:** quasi sempre gli esempi si concludono indirizzando lo studente a risolvere uno o più esercizi (simili all'esempio) riportati alla fine del capitolo.

Altri strumenti di cui il libro è dotato ne facilitano la lettura. Fra questi, i box **In Pratica** e **Riepilogo**. I primi mostrano come gli statistici effettivamente analizzano i dati nella pratica operativa. I secondi, come anche il **Riassunto del Capitolo**, sono prospetti riepilogativi in cui si fissano le questioni più rilevanti che è bene cercare di memorizzare, magari attraverso qualche semplice formalismo.

Per non dimenticare mai da dove veniamo, e per contestualizzare storicamente alcuni degli argomenti proposti, gli scorci **Nelle mani di...** parlano al lettore di quei ricercatori che hanno fornito contributi fondamentali allo sviluppo delle scienze statistiche.

Infine, le pagine di questo volume sono arricchite dai seguenti box:

- **A parole:** con termini semplici spiegano definizioni e simbologia matematica.
- **Attenzione:** segnalano temi che necessitano di particolare attenzione per evitare errori o assunzioni mal poste.
- **Ricorda:** richiamano concetti presentati nelle pagine precedenti.
- **Lo sapevi?:** forniscono informazioni che servono a contestualizzare le questioni statistiche trattate.
- **Laboratorio con **: si tratta di prospetti originali rispetto all'edizione americana, che indirizzano lo studente a specifici laboratori monotematici sviluppati con il software **R** e raccolti alla fine del volume. L'argomento è importante, caratterizza il volume che state sfogliando e merita lo spazio conclusivo di questa Prefazione.

Riprendendo il discorso iniziale sull'approccio seguito, la caratteristica senz'altro più rilevante del volume di Agresti è che alcune questioni teoriche anche molto complesse di calcolo delle probabilità e d'inferenza (per esempio il teorema centrale del limite e la filosofia e il comportamento degli intervalli di confidenza e dei test di ipotesi) sono presentate senza formalismi affiancandole con simulazioni, quasi sempre proposte attraverso degli *applet*. **R** è “naturalmente” orientato alla simulazione, grazie alla “vettorialità” di gran parte delle sue funzioni, e si presta a eseguirla in modo semplice ed efficace. Di qui la nostra decisione di introdurre l'uso di un pacchetto che fosse, almeno in Italia, più diffuso del software Minitab<sup>®</sup> utilizzato nell'edizione americana. Fra i tanti a disposizione, la nostra scelta è caduta su **R**, non solo per la sua vocazione naturale a essere impiegato nelle simulazioni. Tra i vari motivi della nostra scelta, ne indichiamo alcuni qui di seguito.

**R** è *open source*. La sua diffusione è ormai capillare nelle università e negli istituti di ricerca, ma non solo: **R** sta progressivamente soppiantando i più famosi software statistici commerciali anche fuori dal settore della ricerca. Nel mondo del lavoro ci si rivolge sempre più volentieri a un pacchetto completo, documentato, aggiornato, i cui moduli sono sviluppati dai maggiori specialisti del mondo e, perché no, completamente e per sempre gratuito. Nel tempo abbiamo maturato una grande esperienza sulla didattica di base e avanzata con **R**. Siamo autori di tre manuali di Statistica con **R**: uno di analisi esplorativa dei dati, uno di esperimenti di probabilità e inferenza e l'ultimo sui metodi statistici nella ricerca sperimentale. Inoltre dirigiamo la collana di volumi “Statistica con **R**”, composta già di 3 manuali, uno dei quali di finanza quantitativa.

Appositamente per questo volume abbiamo sviluppato un pacchetto **R** che contiene le funzioni e i *data set* utilizzati nel libro e tanto altro ancora. Il pacchetto si chiama *sasld*, acronimo di *Statistics: The Art & Science of Learning from Data*. Un uso compiuto del manuale prevede di utilizzare questo pacchetto all'interno della piattaforma online associata a questo manuale, o eventualmente di scaricarlo per l'installazione sulla propria macchina. Nel testo il simbolo  contrassegna gli output (grafici o tabellari) che possono essere riprodotti eseguendo uno script disponibile nella piattaforma online. Anche per lo svolgimento di alcuni esercizi è raccomandato l'uso di **R**.

Come anticipato, il manuale è corredato da una serie di laboratori con **R** collocati alla fine del libro. Alcuni di questi sono pensati in sostituzione degli *applet* inseriti nel volume americano e riguardanti principalmente problemi di simulazione (il

campionamento, il lancio di un dado, le distribuzioni campionarie, la performance degli intervalli di confidenza e dei test di significatività ecc.). Per tali laboratori sono stati mantenuti in uso i *data set* originali. Tutti gli esempi proposti nel libro e tutte le applicazioni che prevedono, a vario livello, il disegno di simulazioni e l'esecuzione di esperimenti Monte Carlo sono perfettamente riproducibili nella forma proposta dal manuale in lingua italiana: è sufficiente, a tale proposito, fissare il numero di innesco del generatore dei numeri pseudo-casuali utilizzando lo stesso “seme” che abbiamo utilizzato noi negli script presentati. Altri laboratori illustrano come, impiegando **R**, si possano eseguire i test di significatività e calcolare gli intervalli di confidenza impiegando le tecniche illustrate nel testo (in particolare, nei Capitoli 10, 11 e 12).

Infine, alcuni laboratori con **R** illustrano più in dettaglio, e a volte estendono, alcuni argomenti trattati marginalmente nel testo. A questo proposito, vale la pena di ricordare:

- la proprietà di minimo della media aritmetica (questo stesso laboratorio potrebbe essere impiegato, con pochissime modifiche, per illustrare la proprietà di minimo della mediana);
- l'introduzione dei concetti di *odds* e di *odds ratio*, ai quali il testo non fa riferimento (si parla solo di rischio relativo), ma che stanno diventando di impiego sempre più comune anche nelle analisi descrittive e sono fondamentali per comprendere i modelli di regressione per l'analisi di dati qualitativi (modelli che sono argomento di corsi di alcune lauree magistrali e, anche, di corsi avanzati per lauree triennali);
- alcuni approfondimenti sul significato di correlazione lineare, illustrando con esempi numerici quanto è citato nel testo;
- il concetto di correlazione parziale, che non è affrontato nel testo;
- l'impiego di **R** e, in particolare, di alcune funzioni inserite nel pacchetto *sas1d*, per illustrare la distribuzione binomiale e la distribuzione normale;
- alcuni approfondimenti sugli intervalli di confidenza per una proporzione, sia simulando la performance dei metodi descritti nel testo quando i requisiti per il loro impiego non sono soddisfatti (Laboratorio 8.3) sia introducendo il concetto di intervallo di confidenza “esatto” (Laboratorio 8.4);
- la simulazione della performance degli intervalli di confidenza per una media, in particolare quando la distribuzione della popolazione non è simmetrica (Laboratorio 8.2), e l'illustrazione dell'uso del *bootstrap*, una tecnica della quale nel testo si fa semplicemente cenno, ma che è del tutto “naturale” impiegando un software come **R** (Laboratorio 8.5);
- il Laboratorio 9.1, dedicato all'illustrazione dettagliata del test binomiale per piccoli campioni, al quale nel testo originale sono dedicate alcune righe e un esercizio;
- il Laboratorio 12.6, dedicato alla simulazione di un modello di regressione lineare, che “costruisce” i dati (ri)proponendo gli assunti che stanno alla base del modello; i docenti che volessero adottare il testo, inoltre, impiegando lo stesso script proposto in questo laboratorio potrebbero trattare, in modo simulativo, l'inferenza relativa ai parametri associati al modello.

L'augurio che insieme ad Alan ci facciamo è che l'uso di questo manuale aiuti a maturare nei nostri studenti la consapevolezza del contributo fondamentale che l'Arte e la Scienza della Statistica forniscono a chi deve analizzare dati e prendere decisioni.

Ci piace pensare che i nostri sforzi di docenti e divulgatori possano insegnare ai nostri studenti a:

- produrre informazioni che consentano di rispondere a quesiti posti correttamente;
- comprendere quanto la probabilità aiuta a fare i conti anche con le casualità della nostra vita;
- apprezzare il ruolo cruciale della distribuzione di campionamento nei metodi inferenziali;
- scegliere in modo pertinente metodi di analisi descrittiva e di inferenza statistica per analizzare dati e trarre conclusioni;
- comunicare in modo chiaro ed efficace i risultati delle analisi;
- capire i limiti che molte ricerche presentano, o perché basate su uno studio osservazionale piuttosto che su una indagine campionaria o su un esperimento randomizzato o perché alcune variabili omesse non sono state rilevate dallo studio e in realtà potrebbero spiegare le associazioni osservate.

Nel dare alle stampe il volume cogliamo l'occasione per ringraziare chi ha contribuito, in vario modo, al nostro lavoro e alla sua realizzazione.

Primo e speciale ringraziamento va ad Alan che ci ha trasmesso il suo entusiasmo con grande simpatia ed umanità. Grazie mille a te, autentico Superman della Statistica!

Poi dobbiamo ringraziare coloro che, con preziosi consigli e suggerimenti, ci hanno consentito di rendere questo manuale più efficace e soprattutto, senza snaturarne l'approccio, più compatibile con le modalità con cui si insegna nelle aule universitarie italiane: Giuseppe Arbia, Davide Piacentino ed Emanuele Taufer. Ad Andrea Mazzitelli, Claudio Fronterre e Flavio Santi siamo grati per aver rilevato alcune imprecisioni contenute nelle prime versioni del volume. Ci preme inoltre ricordare Michele Andreas, Ivano Bison e Andrea Di Nicola per averci sempre sostenuti e incoraggiati.

Ringraziamo, infine, l'editore Pearson e, in particolare, Giordano Vecchi per aver creduto nel nostro progetto e in altri che stiamo concependo. Maria Diletta Strumolo ci ha seguito con professionalità (e, a volte, con pazienza) in alcune fasi editoriali del nostro lavoro di curatori di questo volume.

Da ultima, e non certo in ordine di importanza, va ringraziata l'onnipresente Silvia Maschio, e la CompoMat s.r.l., alla quale ci lega una affettuosa e ormai quinquennale collaborazione, per il suo insostituibile "dietro le quinte" editoriale.

Questo libro è alla prima edizione e potrebbe contenere refusi, imperfezioni o errori sfuggiti al processo di revisione. Ci scusiamo in anticipo per questo e saremo grati a chi volesse segnalarceli all'indirizzo e-mail [giuseppe.espa@unitn.it](mailto:giuseppe.espa@unitn.it), in modo da predisporre un'eventuale *errata corrige*.

Il nostro ringraziamento più grande va comunque a persone speciali che, con la loro presenza, ci ricordano costantemente compiti più importanti e più piacevoli di scrivere un libro: Massimo, Guido, Luisetta, Giovanna, Maria Rosaria, Annamaria, Eleonora. Il nostro lavoro è dedicato a loro.

*Giuseppe Espa  
Rocco Micciolo  
Diego Giuliani  
Maria Michela Dickson*

Trento, 24 dicembre 2015